For the Love of Teaching

UNDERGRADUATE STATISTICS

Edited By:
Alisa Beyer, PhD and Janet Peters, PhD
FEEDBACK

Feedback regarding the editorial content of this book or any of its essays should be directed toward the individual authors or the book’s editors. They (authors and editors) are solely responsible for the substance of the text. Feedback regarding technical matters of formatting or accessibility of this text via the online environment of the Internet should be directed to the Internet Editor. If you have any complaints or difficulties in accessing these materials, be sure to provide as detailed a description of your problem(s) as you can; you should include information about the browser you are using and the type of computer you are using.

COPYRIGHT AND OTHER LEGAL NOTICES

The individual essays and chapters contained within this collection are Copyright © 2020 by their respective authors. This collection of essays and chapters as a compendium is Copyright © 2020 Society for the Teaching of Psychology. You may print multiple copies of these materials for your own personal use, including use in your classes and/or sharing with individual colleagues as long as the author’s name and institution, and a notice that the materials were obtained from the website of the Society for the Teaching of Psychology (STP) appear on the copied document. For research and archival purposes, public libraries and libraries at schools, colleges, universities and similar educational institutions may print and store in their research or lending collections multiple copies of this compendium as a whole without seeking further permission of STP (the editors would appreciate receiving a pro forma notice of any such library use). No other permission is granted to you to print, copy, reproduce, or distribute additional copies of these materials. Anyone who wishes to print, copy, reproduce, or distribute copies for other purposes must obtain the permission of the individual copyright owner(s). Particular care should be taken to seek permission from the respective copyright holder(s) for any commercial or "for profit" use of these materials. ISBN: 978-1-941804-57-5

SUGGESTED REFERENCE FORMAT

We suggest that the overall text be referenced in this fashion:


Individual chapters may be referenced in this fashion:

Table of contents

Introduction of the book from the Editors ..................................................................................5

Part One: Considerations for Course Design ..............................................................................12

1. **Show Your Students that Statistics are Everywhere**
   Jessica Hartnett ..........................................................................................................................13

2. **Teaching Statistics Online: Advice for Being Your Best No Matter the Modality**
   Georjeanna Wilson-Doenges ..................................................................................................19

3. **Flipping the Script on Stats Education: Presenting Evidence for and a Guide to Flipping the Psychology Stats Classroom**
   Donelle Posey and Amy T. Nusbaum ......................................................................................30

4. **Encouraging Reflective Learning in Introductory Statistics**
   Ashley Waggoner Denton .........................................................................................................41

Part Two: Course project ideas ....................................................................................................54

5. **Passion-Driven Statistics**
   Kristel M. Gallagher, Kristin R. Flaming, and Lisa C. Dierker ..............................................53

   Charles R. Collins .....................................................................................................................83

7. **News Students Use: Critical Evaluation Assignment for Undergraduate Statistics Course**
   Alisa Beyer ..................................................................................................................................96

8. **Statistics and Writing Together at Last: Dual Skills of the Critical Thinker**
   Connor P. Principe ....................................................................................................................114

Part Three: Learning and Engagement Activities ........................................................................127

9. **The Important Yet Difficult Task of Making Statistics Engaging**
   Janet M. Peters ..........................................................................................................................128

10. **Puzzling it Out: A Collaborative Review Activity**
    Miranda M. McIntyre ................................................................................................................141

11. **Engaging Students in Statistics: The Power of Manipulatives**
12. Using StatHand to Improve Students’ Statistic Selection Skills
Peter J Allen, Jessica L Fielding, Ryan H S Kay, and Elizabeth C East

13. Feeling Good About Teaching the New Statistics
Tamarah Smith

14. Meta-Analysis as a Tool for Increasing Students’ Scientific Thinking
Jennifer V. Fayard

15. Advice and Activities for Teaching Replication in Statistics Courses
Nathaniel R. Greene, Mike Corcoran, and Lisa M. Bauer

16. Reject Rejecting the Null Hypothesis: Using Confidence Intervals to Encourage Meta-analytic Thinking
Thomas Cavanagh

17. Using Jamovi in an Undergraduate Psychological Statistics Course: A Free Alternative that Promotes Conceptual Understanding and Active Learning
Abby Heckman Coats and Andrew Mienaltowski

18. Yes, Beginning Statistics Students Can Use R!
Jennifer E. Samson

19. Compendium
Alexander J Bies

Contributors
Overview

Introduction of the book from the Editors

Having met at the Society for Teaching Psychology’s (STP) annual conference, we immediately bonded over our shared love for teaching statistics. We agreed on the potential for statistics to meaningfully shape the knowledge, skills, and experiences of our students and the importance of living up to that potential. Thus, after much discussion, reflection, and shared enthusiasm for supporting other teachers of statistics, we decided to embark on this new adventure.

The heart of this book originates from our shared sentiment that statistics is a crucial component in the curriculum of undergraduate psychology programs, yet it remains one of the most challenging courses for students to take and for faculty to teach. Such difficulty stems from several aspects of the course. For students, the course can be daunting due to the emphasis on quantitative analyses, low math self-efficacy, challenges with learning new computer software, heavy workload, and the perception that statistics will have little value in their personal and professional lives. For faculty, the course can be daunting due to the high workload, potential interference of negative student attitudes, and limited resources. The number and complexity of these factors can make teaching statistics feel overwhelming, but it doesn’t have to be.

This book is a labor of love designed to provide readers with advice, best practices, and fun ideas for teaching an introductory statistics course to undergraduate psychology students. We solicited contributing authors that have used research and best practices from the scholarship of teaching and learning to provide an outstanding resource for instructors of statistics.

An additional consideration in creating this book relates to changes in higher education that have created the need for new approaches to delivering content. More non-traditional and first-generation students are attending college than ever before, and technology has transformed course formats to include complete or partial online components. Moreover, the increasing cost of education has created a new demand for high enrollment courses and free or low-cost course materials. Thus, we encouraged authors to include activities that can be used for multiple teaching modalities (online, in-person, small or large classes, etc.), or to include ways to adapt the activity depending on the course environment.

Finally, we felt it important that the content in the chapters reflect important student learning outcomes. To this end, we used guidelines from the 2014 American Psychological Association statistical literacy task force (APA 2014) and the Guidelines for Assessment and Instruction in Statistics Education (GAISE) from the American Statistical Association as a framework for the book (Table 1). The GAISE and APA goals include suggested practices for instructors to use an applied approach to teaching statistics, active learning activities, and technology in undergraduate statistics courses.

Table 1. APA (left) and GAISE (right) goals for teaching statistics to undergraduates

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Interpret basic statistical results</td>
<td>1. Teach statistical thinking.</td>
</tr>
<tr>
<td>2. Apply appropriate statistical strategies to test hypotheses</td>
<td>2. Focus on conceptual understanding.</td>
</tr>
</tbody>
</table>
3. Apply appropriate statistical and research strategies to collect, analyze and interpret data, and report research findings

3. Integrate real data with a context and a purpose.

4. Distinguish between statistical significance and practical significance

4. Foster active learning

5. Evaluate the public presentation of statistics

5. Use technology to explore concepts and analyze data.

6. Use assessments to improve and evaluate student learning.

In addition, for those faculty interested in addressing specific APA and GAISE learning outcomes, we have identified how each chapter connects to the goals put forth by each taskforce (See Table 2).

**Table 2. Focal goals for each chapter in accordance with APA and GAISE**

<table>
<thead>
<tr>
<th>Chapter topic</th>
<th>APA</th>
<th>GAISE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>Part 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Stats Everywhere</td>
<td>X X X X X</td>
<td>X X</td>
</tr>
<tr>
<td>2 Teaching Online</td>
<td>X</td>
<td>X O</td>
</tr>
<tr>
<td>3 Flipping the Course</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Reflective Learning</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Part 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Passion Driven Statistics</td>
<td>O O</td>
<td>O O X O X</td>
</tr>
<tr>
<td>6 Service Learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 News &amp; Research</td>
<td>X</td>
<td>O X X</td>
</tr>
<tr>
<td>8 Writing in Stats Courses</td>
<td></td>
<td>X X O</td>
</tr>
<tr>
<td>Part 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Application &amp; Engagement</td>
<td></td>
<td>X X X</td>
</tr>
<tr>
<td>10 Puzzles &amp; Games</td>
<td>X</td>
<td>X X</td>
</tr>
<tr>
<td>11 Manipulatives Activities</td>
<td>X O</td>
<td>O X</td>
</tr>
<tr>
<td>12 Using StatHand</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td>Part 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 The New Statistics</td>
<td>O O</td>
<td>X O</td>
</tr>
<tr>
<td>14 Meta-analysis</td>
<td></td>
<td>X O O X</td>
</tr>
<tr>
<td>15 Replication</td>
<td>O O</td>
<td>X O X O O</td>
</tr>
<tr>
<td>16 Confidence Intervals</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Part 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 Jamovi</td>
<td>O</td>
<td>X X</td>
</tr>
<tr>
<td>18 R</td>
<td>O O</td>
<td>O O O X</td>
</tr>
<tr>
<td>Additional resources</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 Compendium</td>
<td>X X X X X X X X X X X</td>
<td></td>
</tr>
</tbody>
</table>
Organization of the Book

The book is organized into five main considerations for teaching undergraduate statistics. Part 1 examines possibilities for course design, Part 2 shares semester long projects or semester-based activities for the course, Part 3 shares fun and engaging activities that can be integrated into your statistics course, Part 4 shares advanced topics to be educating your students on, and Part 5 expands on students using statistical software, highlighting use of two free statistics software programs for teaching statistics. We end the book with a compendium chapter full of further resources to check out.

Part One: Considerations for Course Design

Creating a well-structured, well-executed course is one of the many challenges of teaching. This can be compounded with a high workload, high demand course such as statistics. Thus, Part 1 examines possibilities for various approaches to course development, from the philosophical approach to course design to platform considerations such as online and flipped classroom experiences.

In Chapter 1, Hartnett leads with an important message: while fewer than 5% of psychology majors go on to earn Ph.Ds. in psychology, over 80% of college students indicate that getting a better job after graduation is a “very important reason” to go to college. In this opening chapter, Harnett provides ways we can increase student buy-in to undergraduate statistics by sharing why and how to teach it at a practical level for a novice, applied statistician as well as highlighting the essential jobs skills gained from taking the course.

Chapter 2, by Wison-Doenges, shines a spotlight on online teaching. Techniques for teaching statistics online, connected to best practices in instructor behavior and technical content delivery are shared. The advice is divided into two general categories: teacher behavior and communication which includes communicating frequently in all aspects of the course, being engaged, and providing prompt feedback; and the technical delivery of content which includes consulting with campus instructional designers to maximize use of the Learning Management System, organizing content to align with course objectives, making short videos, developing meaningful discussions, allowing students to turn in pictures of their handwritten work, and holding online office hours.

Chapter 3, by Posey and Nusbaum, presents evidence for the efficacy of active learning using a flipped classroom approach. Their experience flipping a statistics class is presented, with a discussion of students’ impressions, logistics, and benefits as well as drawbacks.

In Chapter 4, Waggoner Denton explains how reflective learning can strengthen students’ understanding of statistics. In a challenging course such as introductory statistics, students learn about themselves, and learn statistics through specific, strategic reflection. The potential benefits of reflection and strategies for incorporating reflective activities into the statistics course are reviewed.

Part 2: Course Project Ideas

Once you’ve addressed the overall design of your class, you will want to consider course projects to encourage student development and demonstration of knowledge. To this end, Part 2 shares semester long projects or semester-based activities for the course, such as service learning and writing about statistics.

To begin, Gallagher, Flaming, and Dieker (Chapter 5) share the concept of Passion-Driven Statistics, a project-based approach to teaching statistics that involves students in the rich and complicated decision-making process of real-world statistical inquiry. Core features of the approach include providing opportunities for students to apply their knowledge flexibly in the
context of real data, the use of technology as a window to core statistical concepts, supporting students with varying levels of preparation, and attracting and inspiring students from underrepresented groups.

In Chapter 6, Collins’ course project takes on a social justice theme by incorporating service learning into the teaching of statistics. This chapter provides a guide to implementing a community-based partnership into a social science statistics course. The chapter outlines the process of project implementation and provides insights into future successful implementation.

Chapter 7, written by Beyer, focuses on scaffolding a final writing assignment that assesses students’ critical thinking skills and application of statistical knowledge. Students analyze news article coverage connected to a primary research article, explain course concepts connected to the research article, and critically evaluate how the research findings are presented.

Finally, Chapter 8, written by Principe, expands on making statistics a writing-focused course. Although writing and data analysis are rarely combined in courses, both represent forms of critical thinking. This chapter argues that students’ quantitative and writing skills are both likely to improve when the professor emphasizes these skills as complementary to the primary skill of thinking critically.

Part 3: Learning & Engagement Activities

After establishing learning goals and course projects, you can begin to add some depth and character to your course. You might consider what types of activities, games, and assignments might serve the dual purpose of teaching and engaging students. To this end, Part 3 shares fun and engaging activities that can be integrated into your statistics course.

In Chapter 9, Peters provides readers with class activities that will engage students with the content, such as using pop culture references, encouraging student application of the material, and increasing student awareness of professional skill development. Activities discussed in this chapter vary between single examples for class concepts and more major investment semester activities, such as long projects and course design elements.

For Chapter 10, McIntyre shares how puzzles are a fun, flexible tool that can spark students' sense of discovery in statistics. This chapter offers advice for designing and implementing statistical puzzles, including evidence-based tips for maximizing student engagement and learning outcomes.

Chapter 11, by Fetterman and Kneaval, is full of activities based on manipulatives (tangible items that can be used to demonstrate mathematical concepts such as food or Legos). Manipulatives are an effective method of illustrating statistical concepts and also help students cope with anxiety they have when taking statistics and research methods. The authors present manipulatives that address typical concepts covered in a psychology statistics course, along with lists of materials and suggestions for how to instantiate them in classrooms.

In Chapter 12, Allen, Fielding, Kay, and East describe the rationale, implementation, and evaluation of a brief training activity to improve students’ statistic selection skills. The activity is built around StatHand, a free iOS and web application. The chapter appendices include the authors’ training materials and over 40 research scenarios that can be freely adapted by instructors for formative and summative learning purposes.

Part 4: Advanced Topics

As the field of psychology moves away from p-values and null hypothesis significance testing, you might be interested in how to more formally or fully incorporate other concepts into your
course. Part 4 shares additional topics to include in your course, including replication, meta-
analysis, and confidence intervals.

In Chapter 13, Smith provides an overview of the “new” statistics. Following a brief review of
how statistics apprehensions negatively impact psychology statistics students, a concrete
example of a lesson using the new statistics is provided. This lesson, with emphasis on the new
statistics, both teaches cutting edge methods while simultaneously helping students learn and
love statistics.

Chapter 14, by Greene, Corcoran, and Bauer, shares how teaching replication in an
undergraduate statistics class can enhance student learning of statistics and make students
more critical thinkers about research in psychology. With a range of example activities that vary
in time commitment from semester-long to single-day, this chapter provides important resources
for instructors interested in introducing replication in their statistics courses.

In Chapter 15, Fayard provides activities for introducing meta-analysis in the introductory
statistics curriculum. The focus is on conceptual aspects of meta-analysis that are intended to
increase students’ understanding of psychology as science and integrate and reinforce
concepts from statistics and research methods. The author also shares resources for instructors
who want to learn more about meta-analysis.

Chapter 16, written by Cavanagh, is on the topic of confidence intervals. Cavanagh discusses
how null hypothesis significance testing (NHST) encourages researchers and consumers of
research to think in dichotomous terms, which often highlight differences rather than agreement
in statistical results. The author reviews the perspective of advocates of the “new” statistics, and
recommends the use of confidence intervals to encourage meta-analytic thinking.

Part 5: Open Data Analysis Software

Part 5, the final section, expands on students using statistical software, highlighting faculty use
of two free statistics software programs for teaching statistics.

In Chapter 17, Coats and Mienaltowski share their experiences of teaching Jamovi. This free
software, which runs on the R programming language, is a user-friendly alternative to SPSS.
The authors describe how Jamovi is used to teach important statistical concepts such as t-tests,
ANOVAs, correlations, and regression. Discussion includes pitfalls to avoid and evidence of
effectiveness.

Samson, in Chapter 18, shares more about teaching R to students in introductory statistics
courses. This chapter provides assurance that, with the appropriate support, even beginning
statistics students can use R successfully. Specifically, Samson demonstrates one way that
support might be provided with sample materials included. Statistics instructors are encouraged
to consider using R for reasons of access as well as promotion of deep conceptual
understanding and preparation for graduate study and the workforce.

Additional Resources

Bies, in Chapter 19, builds a strong compendium of resources for teaching statistics. This
chapter provides a set of resources that map onto previous chapters from the book. The
resources are organized by chapter order and topic. For continuity and ease of use, entries are
duplicated in cases where they are relevant for more than one chapter. Following the list, a
summary table maps onto the APA and GAISE goals for teaching statistics to undergraduates
for each of the resources.
Editors’ closing

We appreciated the timeliness, flexibility, and cordiality from the contributors of this book. Contributor biographies are listed at the end of the book. We also wish to thank Journey Norton, a former student at Chandler-Gilbert Community College, for her hard work on the cover designs. We hope you find a new love for teaching statistics from the joy we had compiling this book. Thank you and happy reading.

Alisa Beyer

Janet M. Peters
References


http://www.amstat.org/education/gaise.

Statistical Literacy Task Force. (2014). *Statistical literacy in the undergraduate psychology curriculum.*  
Part One: Considerations for Course Design
Show Your Students that Statistics are Everywhere

Jessica Hartnett, PhD
Gannon University

Summary

Only 4% of psychology majors go on to earn Ph.D.s in psychology. Meanwhile, 84% of college students indicate that getting a better job after graduation is a “very important reason” to go to college. As such, I think that we can increase student buy-in to undergraduate statistics if we teach it at a practical level for a novice, applied statistician, highlighting examples that use statistics in the real world.

Why we need to change how we teach statistics

I have taught Psychological Statistics (aka Introduction to Statistics) 60+ times over the last ten years, teaching both psychology and non-psychology majors the basics of statistics. Over the years, I have modified my course, searching for a way to introduce and explain statistics that felt authentic for me as an instructor and relevant and worthwhile for my students. I think I have found a method that works: Over the semester, I show my students that statistics are everywhere, and data can inform complicated (not just psychology) questions. I teach my students that statistical literacy is a flexible, adaptable, necessary skill for any career.

To teach this way, you cannot just tell your students that statistics and data are everywhere. You must show them. I will use this chapter to describe two strategies I use to convince my students that statistical literacy is a sought after and valuable skill.

1. Starting on the very first day of class, I present students with three facts to argue that statistics are necessary in the modern and future workplace and that I want to help my students to succeed in that workplace.

2. I reinforce this message during the semester by teaching in a practical, applied manner, emphasizing computing, communication, and examples from different disciplines.

Three facts for the first day of class

To make statistics relevant for students, I use three facts to convey to the students that I believe that they want to succeed, I understand the sorts of jobs they want upon graduation, and that my class can help my students develop the skills necessary to succeed.

Fact #1: Most college students indicate that getting a better job is “a very important” reason to attend college.

The National Association of Colleges and Employers (NACE) surveys incoming first year college students every year, asking them questions about college preparedness, why they chose their college, and what they expect to get from their college education. On this survey, 84% of students indicated, “To get a better job” as a very important reason for attending college (Stolzenberg et al., 2019).
Why not align how you teach with your students’ goal of getting a better job? Teaching content is necessary but teaching your students a new, specific skill set is invaluable. Which leads to the next question: What skills do employers want in job applicants?

**Fact #2: Potential employers want to hire people with skills they can develop in statistics class.**

NACE also asked hiring managers what skills they look for directly on potential hires resume. 71% of hiring managers want analytic/quantitative skills, and 82% want candidates with written communication skills (NACE Staff, 2018).

Analytic/quantitative skills and written communication skills are already necessary components of a statistics course. As we teach our students to conduct statistical analyses and then communicate the results of these findings, they gain content knowledge in the field, but they also become more employable. I suggest that you structure your course so you nurture your students’ analytic skills with the core course content, and strengthen their communication skills by emphasizing their statistical literacy and data reporting skills. However, statistical skills cover broad level abilities, from individuals who can generate bar graphs for their Luddite boss to the most cutting edge, Bayesian-thinking, Python-using analysis in the business. So, at what level should an undergraduate psychology instructor teach to best serve their students? I think that depends on what our students are doing with their psychology degrees. Which takes us to fact #3:

**Fact #3: Most psychology majors will apply for bachelor and master degree level jobs. We should teach statistical skills that will serve these students, not just students who are PhD bound.**

According to the APA Center for Workforce study (CWS Data Tool, n.d.), there are 3.5 million Americans with Bachelor’s degrees in psychology. Fifty-six percent (56%) of them did not pursue any graduate degree. Four percent (4%) obtained PhDs in psychology, and 13% obtained master’s degrees in psychology. The rest of them pursued graduate work outside of psychology. The majority of our students are not going into quantitatively rigorous doctoral programs. As such, I believe we need to re-think what we emphasize in our Introduction to Statistics courses to best serve students who may never take another statistics course. All of our students should be able to compute and interpret ANOVA, t-test, chi-square, regression, and correlation. All of our students also need to learn necessary data communication skills and how to read and evaluate data.

To reiterate: Our students want to get jobs. What kind of jobs? Mostly jobs that do not require a doctorate. The gate-keepers to those jobs, the hiring managers and employers, want employees with analytic and communication skills. We should use these facts to inform our course structure and content. Specifically, I think we should teach an applied introduction to statistics class that emphasizes the computing and communication of statistics that many of our students will encounter in the workforce.

**Teaching Statistics So Your Students Can Use Statistics**

I think we need to teach statistics the way it is used in real life. One part of this, which may be a substantial departure for many of you who teach statistics, is to stop calculating test statistics by hand and instead teach your students to conduct statistical analyses using only software and computers. I believe this for two reasons. One, no professional outside of a classroom has done this work by hand in decades. Two, we calculate test statistics to determine p-values. And then, if we are doing our job right, we have to explain to our students that they just spent a lot of time calculating a test statistic to get a p-value but p-values are growing out of favor and being augmented with effect sizes and confidence intervals. By teaching our students to communicate
findings both formally and informally, depending on the audience. I strongly believe we need to teach students very applied, practical, useful statistics classes filled with engaging, applied examples.

I am suggesting a big change, teaching statistics with minimal equations. Many of us teach introductory statistics the way we learned it: With a heavy emphasis on by-hand calculation in order to teach theory. You, a statistics instructor, were probably taught this way. Most statistics textbooks spend dozens and dozens of pages explaining and working through formulas. The main argument for this approach is that if students cannot understand the guts of statistics, they cannot understand WHAT the statistics are doing, without working through the underlying the math.

**I believe you can teach theory without formulas.** One counter argument I hear against this approach is that you cannot teach statistical tests without working through the formula for these tests by hand. I disagree. If you want to teach your student t-test theory, explain that t-tests were created by Gosset to accommodate small sample sizes, and they only ever compare two groups using one dependent variable. Provide your students with a research question and ask them to design different research studies that correspond with the different kinds of t-tests (Hartnett, 2013). You can explain paired t-tests using the example of the twin brothers (and astronauts) Mark and Scott Kelly (Hartnett, 2016). One was in space for over a year, the other for a few weeks. NASA is studying these identical twins in order to better understand the long term side effects of weightlessness and exposure to solar radiation.

Explain ANOVA and the concept of between and within group differences not with dry equations but with polling data. For example, look at how religious groups differ from one another on beliefs about climate change, and how individuals within one religion can have various opinions about climate change ([https://notawfulandboring.blogspot.com/2016/01/explaining-between-and-within-group.html](https://notawfulandboring.blogspot.com/2016/01/explaining-between-and-within-group.html)).

Explain how expertise is required to interpret effect sizes by describing the simultaneously small and large effects related to a malaria vaccination that is “only” 30% effective (Hartnett, 2019).

Using these applied, conceptual examples will also provide repeated evidence of the initial, first day of class argument that data is everywhere and statistical literacy is required across a wide array of disciplines, like biology, space travel, religious studies, and environmental studies.

One point of clarification: If you are teaching a student who falls in the 4% who will be going to graduate school, perhaps you should increase their math knowledge by advising them to earn a minor in computer science, statistics, or mathematics (depending on the structure of the university). I believe that such a student will still benefit from the pedagogy I have described but also need additional statistical education. Further information in this volume will describe using R and Jamovi to teach undergraduate statistics, which are beneficial and free programs for more advanced statisticians but can also be used in the Intro Stats classroom (see chapters 17 & 18 for Jamovi and R).

**Spend precious instructional time teaching your students how to use software to perform statistics. And then show them again.** I know we talk a lot about how our iGen students are digital natives, but they are still intimidated by math and learning new software to perform math. Showing your students how to use new software only once, in a lab section, is not sufficient for mastery. Additionally, walking through the logic of interpreting a p-value and an effect size once, or twice, or three times, is not enough. Reinforce these lessons with every computer-based example. Students need ample examples to understand this, especially how to handle statistical tests that result in a significant p-value but a small effect size.

While I emphasize teaching via computers, actually teaching in a physical computer lab helps but is not necessary. There are different ways to go about providing support for student work on
computers. You should work with your Information Technology Support to make sure that the
proper software is available, although most computers provide students with Excel, and JASP, a
free point-and-click software program, can be used via internet browser. If you are interested in
teaching your student coding, you could use another JASP-based, also free program Jamovi,
that uses both the JASP and R interfaces. You could also teach your students R as well. If you
work at a larger university, devote lecture time to applied theory and working through examples,
and lab time practicing software. You should also work with your university to see if there are
any laptop or tablet lending programs your students can use. If your university provides
mandatory laptops or tablets, you may want to determine if the software that you want to use is
compatible with those university initiatives. If none of that works, you can offer ample in-class
examples from one of these programs connected to your personal laptop. You can create
videos using Kaltura or other specialty desktop recording software. You can teach students to
close analysis on tablets or mobile devices.

In addition to teaching our students to be confident and fluent when performing statistical
analyses, I also believe we must train our students to write about data (for more on this, see
chapter 8). By improving their written communication skills, we can help them find and keep
jobs, and I believe that we should dedicate a substantial amount of our face to face time with
our students to these goals. This emphasis serves both workforce preparation goals as well as
graduate school preparation goals. If you are in a position to select a graduate student or
graduate RA, what would you prefer: A student who can perform a regression by hand with a
calculator, or a student who can conduct a regression using software and interpret and report
that regression using APA formatting basics?

In addition to emphasizing and living The Three Facts of Computing and Communication, I
believe it helps students to see multiple examples of the statistics they are using. I also think
that exclusively using psychology research examples may not benefit students who are more
interested in and drawn to applied examples that are not necessarily based in psychology
research. I think there is a real advantage to couching your statistical test examples in good
science reporting, as to emphasize the role of statistics in all kinds of research without dealing
with jargon and specialty education barriers you encounter when trying to teach with actual
research articles. Finally, an additional benefit to this type of teaching may be to reduce overall
anxiety in the classroom (Chew & Dillon, 2014) by reducing the emphasis on mathematics.

**Applied and varied statistical examples**

To emphasize the fact that statistics are used across all disciplines and workplaces, I do the
hard sell described in the first portion of this chapter. Over the semester, I continue to promote
the message by using applied examples from varied occupations. Here, I will detail two types of
applied examples: Conceptual and concrete. Conceptual examples, which are described in the
previous section, can replace formula when we teach our students the guts of statistical tests.
The second type of example is more concrete, and use real life examples paired with data to
provide students with computational examples.

These examples all use real-life, high-quality science reporting from popular sources (e.g., NPR,
Time Magazine). I take the popular report about real science, generate fake data that mimics
the actual finding, and then the students analyze the data. After listening to or reading the
popular story, students learn about the bigger context for the research project and know the
results of their data analysis before conducting the data analysis. Researchers had success
using popular source articles about problems involving statistics and found that students were
able to transfer statistical thinking to critical writing about the popular source articles (Daniel &
Braasch, 2013).
You might think that this takes the fun or challenge out of the analysis. I don’t disagree with you. I don’t think your students should know all of the outcomes for all of the examples they use in your class. However, the first time that your students conduct a new statistical test, I believe that by providing them with the results before they analyze their data, we allow them to concentrate on the process of data analysis, and not the end results. Indeed, the results are important in practice, but not as important during the learning process.

Here are two popular science examples from NPR I use to introduce ANOVA and chi-square. The first tested whether or not silence, music, or audiobooks could decrease pain in pediatric surgery patients. I play the news report by Patty Neighmond (n.d.), “To Ease Pain, Reach For Your Playlist.” Her report describes the original research (Suresh, De Oliveira, & Suresh, 2015) and includes an interview with two of the co-authors. The study found that both audiobooks and music participants experience a one-point decrease on the Wong-Baker Pain Scale, and both of those groups are significantly different than kids in the control group. I created a fake data set that mimics the actual findings, and that data set as well as more information on the activity is available at my blog: https://notawfulandboring.blogspot.com/2015/08/anova-example-using-patty-neighmonds-to.html.

This example hits on numerous statistical and research methods lessons: In a world stymied by click-bait news stories, this example uses high-quality science reporting about research. It also demonstrates statistics used in medical research. As a class, you can even circle back around and discuss the shortcomings of this research (e.g., very small n-size) and how the study could be improved and replicated. Hypothesis generation is also described: The research was inspired by the father-daughter co-authors’ mother and grandmother’s stay in a hospital.

My second example introduces the chi-square test of independence with research that studied empathy in very young children (Cirelli, Wan, & Trainor, 2014). Researchers divided children into two groups: Children who danced in sync or out of sync with a stranger. Next, they created a situation in which the child could assist or not assist that stranger. The researchers found that the children were more likely to provide assistance if they had danced in sync with the stranger. I turned this in a 2 x 2 chi-square. You can access my fake data that mimics real findings at my blog: https://notawfulandboring.blogspot.com/2018/06/chi-square-example-via-dancing.html.

Again, this example features high-quality science reporting. It also describes the inspiration for the original research: Before graduate school, the principal investigator worked at a daycare center. Working with children sparked her interest in developmental psychology. This example also demonstrates how research can inform and improve intimate parts of our lives, like parenting and encouraging bonding with our children.

To generate data sets, I use two different sources: Richard Lander’s Dataset Generator for Learning Introductory Statistics (https://rlanders.net/dataset-generator/) or Andy Luttrell’s Data Generator for Teaching Statistics (http://andyluttrell.com/datagen.html).

Wrap up

I believe that statistics instructors are in a unique position to teach our students a specific skill set in addition to a basic content area for psychology. By using the methods I describe, you can even advise your students to add this line to their resume at the end of the semester:

Special skills: Novice data analysis using JASP software, including descriptive statistics, t-tests, ANOVA, regression, correlation, and chi-square.

Coda: If you have any other questions about my techniques, feel free to email me at hartnett004@gannon.edu. If you are interested in seeing my other ideas for the teaching of statistics, feel free to visit https://notawfulandboring.blogspot.com or follow me on Twitter at @notawful.
References


NACE Staff. (2018, December 12). *Employers want to see these attributes on students’ resumes.* [https://www.naceweb.org/talent-acquisition/candidate-selection/employers-want-to-see-these-attributes-on-students-resumes/](https://www.naceweb.org/talent-acquisition/candidate-selection/employers-want-to-see-these-attributes-on-students-resumes/)


Teaching Statistics Online: Advice for Being Your Best No Matter the Modality

Georjeanna Wilson-Doenges, PhD
University of Wisconsin-Green Bay

Summary
Teaching statistics is certainly challenging no matter the delivery method, but the course can be especially daunting when taught completely online (Breneiser, Rodefer, & Tost, 2018). Particularly with a course like statistics where many students feel overwhelmed, anxious, or that they are “bad at math,” instructors may struggle with how to provide the same level of support in the online environment, especially when students vary widely in their preparedness (Grandzol, 2004). Regardless of the delivery method or course content, we know what it takes to be a good teacher (Kirby, Busler, Keeley, & Buskist, 2018). Using what we know are behaviors that model teachers use, this chapter will provide advice for teaching statistics online using these model teaching criteria as a guide. The advice is divided into two general categories: teacher behavior and communication which includes communicating frequently in all aspects of the course, being engaged, and providing prompt feedback; and the technical delivery of content which includes consulting with campus instructional designers to maximize use of the Learning Management System, organizing content to align with course objectives, making short videos, developing meaningful discussions, allowing students to turn in pictures of their handwritten work, and holding online office hours. Being mindful of both to use best practices regarding instructor behavior as well as the technical delivery of content can provide excellent online statistics courses that create the conditions for student success.

Introduction
Teaching statistics is certainly challenging no matter the delivery method, but the course can be particularly daunting when taught completely online (Breneiser, Rodefer, & Tost, 2018). Higher education is changing rapidly and one of the strategies for adapting is offering courses in a completely online format. Online courses are growing in popularity as enrollments continue to increase year after year with recent data showing that 15.4% of all students are studying exclusively online (Lederman, 2018). One of the best-selling points for online education is access: delivering course content asynchronously online increases access for students who, with work and family obligations, would not otherwise be able to go to college (Darby, 2019). But regularly matriculated students are also flocking to online courses because of convenience and preference: online classes fit well into their lifestyle and provide an opportunity to take a class when the in-person (FTF) version is offered at the same time as another class they need to take. In 2017 nearly 18% of students mixed online and in-person classes with one third of all students taking at least one course online (Lederman, 2018).

However, as Darby (2019) reports from a 2017 Educause survey, only 9 % of those teaching in higher education actually prefer to teach online. This shocking statistic highlights a challenge in the mismatch between what students seem to want and what faculty prefer. Part of the problem may be that many current academics did not grow up teaching or taking online courses, and they may not be 100% committed to improving their online skills (Darby, 2019). Teachers may
feel that online courses are inherently less effective than the quality education we believe we provide in the face-to-face classroom. Particularly with a course like statistics where many students feel overwhelmed, anxious, or that they are “bad at math,” instructors may struggle with how to provide the same level of support in the online environment, especially when students vary widely in their preparedness (Grandzol, 2004). In addition, for the type of students who are taking their coursework completely online, statistics may be particularly troublesome. According to the National Center of Education Statistics, nearly half of undergraduates enrolled in online courses work full-time and 45% have children (ClassesandCareers, 2018). If the student has many obligations outside of school like work and family, it may make a course like statistics overwhelming, especially at the beginning. And then, of course, there is the inherent fear factor associated with a math-type class in the first place (Dunn, 2014). If it has been five or 10 years since the student has taken a math class, this may make the learning curve steeper. Although students may be increasing their enrollment in online courses, one study found that the majority of students preferred FTF as opposed to online delivery for an introductory statistics course because of the difficulty of the subject matter, procrastination can be particularly damaging in a statistics course, and the lack of perceived access to the instructor to ask questions (Johnson, Dasgupta, Zhang, & Evans, 2009). The challenge of offering statistics online from both the teacher and the student perspective is clear, and instructors of online statistics courses must be particularly mindful of ways to overcome these obstacles.

Regardless of the delivery method or course content, we know what it takes to be a good teacher. In his classic work, Lowman (1995) observed that excellent university teachers have similar qualities that can be divided into two general categories: the teacher’s personality and communication skills, and their technical skills (as cited by Kirby, Busler, Keeley, & Buskist, 2018). Buskist and colleagues (2002) operationalized these qualities with 28 items named the Teacher Behavior Checklist (TBC). In their research, they found that, of the top 10 most important qualities according to faculty and students, both agreed on six: being knowledgeable, being enthusiastic, being respectful, having realistic expectations, being approachable and personable, and being creative and interesting. Further research employing factor analysis on the TBC yielded two subscales: “Caring and Supportive,” and “Professional Competency and Communication Skills,” supporting Lowman’s (1995) two category model of outstanding teaching (as cited in Keeley, Smith, & Buskist, 2006). In other seminal work, Richmond, Boysen, and Gurung (2016) enumerated six Model Teaching Criteria which include training, syllabi construction and course planning, instructional methods, course content, and assessment of students learning. All of these assessments of excellent teaching can serve as a guide for excellent teaching online as well. The delivery modality should not dictate whether the teacher is excellent or not, but rather the teacher should use the delivery modality to deliver the best student learning opportunities possible.

Using what we know are behaviors that the model teachers use, the rest of this chapter will be dedicated to advice for teaching statistics for psychology in a fully online format using these model teaching criteria and checklists as a guide. This advice is divided into the two general categories generated by Lowman (1995): advice about the teacher including behavior and communication, and advice about the course including structure and the technical delivery of content.

**Teacher’s Behavior and Communication**

Using the TBC (Buskist, Sikorski, Buckley, & Saville, 2002; Keeley et al., 2006) as a guide, we know that both students and teachers agree that being approachable, enthusiastic, creative, fair, knowledgeable, and respectful are model teacher behaviors, as well as having the traits of
being caring and supportive. How do these behaviors translate into the online environment? What are some ways that teachers can emulate model teaching criteria to enhance the learning of their online statistics students? The following is some advice for how teachers of online statistics can embody these behavioral and communication characteristics.

**Communicate frequently in all aspects of the course**

Most excellent teachers I know who teach FTF show up to class a few minutes early and talk casually with students before class begins, and they interact with students during and after class as well. Online courses should be no different (Darby, 2019). Keeping an ongoing dialog with students in all aspects of the course helps students succeed through their interactions with the instructor where they bolster students’ confidence, guide them when they are off-target, encourage them as they struggle to learn the material, correct them when they make errors, and share in the joy of learning. Developing rapport through communication is a model behavior from the TBC that leads to the teacher being perceived as accessible, approachable, and personable (Buskist et al., 2002). Chatting with students about tangential or unrelated information can also help humanize instructors in the online setting. When teaching a course like statistics that produces a lot of anxiety in students, it can be the casual, human interactions that can help students persevere when they are struggling most (Gazioglu, 2012). One of my colleagues accomplishes this in her online course by having students send her pictures of their pets. She shares these by posting the “featured pet of the week” on the Home Page, which may help reduce anxiety and help her connect in a real way with students. Another important form of communication is contacting students who have not logged into the course Learning Management System (LMS) for a few days (or longer) to check in on them. Asking the simple question, “Are you ok?” could be the difference between a student dropping out versus trying to catch up or get the help they need. When you cannot visibly see that a student is struggling, reaching out with concern can help you understand the circumstances of a student’s absence online. These behaviors show that you encourage and care for students in meaningful ways, another of the behaviors of excellent teachers (Buskist et al., 2002; Keeley et al., 2006).

**Be engaged**

Learning in an online environment can sometimes feel disconnected and isolated (Mocko, 2013), and in the example of a statistics course can also be scary. Your students need you to be an active participant and leader of the class by striving to improve student–teacher rapport and engagement (Richmond, Boysen, & Gurung, 2016). In an online course it is important to give the sense that the students are part of a group and that they are learning from the same instructor and completing the same assignments (Mocko, 2013). As Darby (2019) suggests, the instructor should be visibly and meaningfully engaged including posting weekly announcements, being a regular contributor to discussions, or posting a quick video or other instructions to clarify content or assignments that students have expressed trouble comprehending. In these ways, you show the students that you are active in the learning process and are there to support them in learning this challenging content (Mills & Raju, 2011). One example of engagement can be in discussion posts (Everson & Garfield, 2008). Students who are learning statistics for the first time have not developed a fluent understanding of the vocabulary and can misuse terms without realizing it. As an engaged instructor, you can catch these misunderstandings early and respond with a kind, but firm correction so that others reading the post do not become confused as well. In the online environment, students have the flexibility of time and space to reflect on their answers and your comments and corrections to help develop their critical thinking skills (Aloni & Harrington, 2018).
Provide prompt feedback

One best practice for online (and all) teaching is to provide timely constructive feedback (both summative and formative) to students (Boysen, Richmond, & Gurung, 2015). If you want your students to submit assignments on time that are well thought through, model behaviors that do not leave things to the last minute: grade assignments within 24 to 48 hours after submission whenever possible. In a course like statistics, getting quick feedback and correcting misconceptions promptly can help as students move forward and build on concepts they have already learned (Gazioglu, 2012). Statistics relies on understanding of prior topics and so feedback on students’ understanding of those topics can help them succeed moving forward. For example, if students do not understand the concept of the mean in the first few weeks of class, then how will they apply that concept when learning about the independent-samples t test that relies on testing differences between means? Timely feedback can communicate to students that they understand the material and help them build confidence moving forward, and can redirect students who are a bit off-target in their understanding before it is too late. Providing constructive feedback is one of the behaviors that makes for an excellent teacher and the timeliness of that feedback can communicate to students that you respect them and truly want them to learn and succeed (Kirby et al., 2018).

Technical Delivery of Content

Just as important as teaching persona and communication with students is the structure and technical delivery of the content of the class. Employing excellent skills in content delivery meets the TBC characteristic of competency and communication skills and supports Lowman’s (1995) second category of outstanding teaching. Being well-versed in the more structural and technical aspects of the course can show students that you are confident, effective, knowledgeable, technically competent, and prepared (Kirby et al., 2018). The following is some advice for how teachers of online statistics can exhibit the skills of excellent teachers in the technical delivery of statistics online.

Consult with your campus instructional designer and learn the ins and outs of your Learning Management System (LMS)

One of the keys to being a great teacher is utilizing the technology at hand to provide the best educational experience for your students regardless of topic. Your instructional design team has the expertise to train you to maximize your LMS (Darby, 2019). According to the Model Teacher Criteria, excellent teachers should constantly seek training and knowledge on pedagogical theory and practice (Richmond et al., 2016). Ensuring your training in online pedagogy and practice is up to date will help you provide an excellent learning experience for your students. One of the greatest tips I received from one of my university’s instructional designers was to use colored emojis to demarcate different types of content in Canvas (our LMS). When I was first learning how to use Canvas, I was disappointed with the lack of color on the Home page and its long list of files and assignments in black ink with a white background. I realize that this is best for accessibility for all students (Yang, 2017), however, adding a small emoji in color made a big difference while also not putting students with ability differences at a disadvantage. This simple trick of putting a green book emoji in front of required readings or a purple movie projector emoji in front of videos made my Home Page on my LMS cleaner and a bit more engaging for the students. I coordinated those emojis and colors with my syllabus as well: what the students see online matches what they see in the syllabus. This is a simple trick that impacts on clarity and reduces confusion for students.
Organize content to align with course objectives

**Course Learning Objective 1:** Calculate and interpret basic descriptive statistics.  
**Course Learning Objective 2:** Calculate and interpret basic inferential statistics using hypothesis testing.  
**Course Learning Objective 3:** Enter and analyze data using descriptive and inferential statistics using SPSS.  

**Module Learning Objectives**  
By the end of each module (week) students will achieve the following:

<table>
<thead>
<tr>
<th>Module (week)</th>
<th>Module Learning Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Calculate and interpret basic descriptive statistics: frequency tables and graphs by hand and SPSS</td>
</tr>
<tr>
<td>2</td>
<td>Calculate and interpret basic descriptive statistics: measures of central tendency and variability</td>
</tr>
<tr>
<td>3</td>
<td>Calculate and interpret basic descriptive statistics: correlation.</td>
</tr>
<tr>
<td>4</td>
<td>Calculate and interpret basic inferential statistics using hypothesis testing: 1 &amp; 2-sample t-test</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Course Learning CLO 1</th>
<th>CLO 2</th>
<th>CLO 3</th>
<th>Learning Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>Homework 1, Computer Lab 1, Quizzes 1 and 2, Exam 1</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>Homework 2, Computer Lab 2, Quizzes 3 and 4, Exam 1</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>Homework 3, Computer Lab 3, Quiz 5, Exam 1</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>Homework 4, Computer Lab 4, Quizzes 6, 7 and 8, Exam 2</td>
</tr>
</tbody>
</table>

**Figure 1. Course Learning Objectives linked to Course Modules and Assessments**

In their enumeration of criteria of model teachers, Richmond et al. (2016) focused on syllabi which serve to define the student-teacher relationship by providing a cognitive map and learning tool for the course as well as course transparency. In the online environment, the entire course structure on the LMS could also serve this same purpose. Online teachers of statistics should provide a clear road map of what students must accomplish and expectations for how assessments will be graded. This could include tips for success, common “bottleneck concepts” where student often get stuck, and campus resources to help students in the class or with issues outside of class that could hinder their ability to learn. One of the best tips for how to accomplish this in the online environment is to develop learning objectives for each module of the course and then link content delivery (e.g., readings, videos, etc.) and assessments (e.g., assignments, quizzes, and exams) to these module-specific learning objectives (see Figure 1) (Richmond et al., 2016). Teachers could also provide brief videos or announcements that clarify an issue with which students have been struggling that week and place it within the module for just-in-time teaching of troubling concepts.

**Make short videos**

In an earlier study of online statistics students, Sebastianelli and Tamimi (2011) reported that students found audio-video clips produced by their instructor to be the most helpful in learning quantitative content. However, I hear from students all the time about how they choose not to listen to audio PowerPoint presentations that are “long” (that means more than 10-15 minutes) or they speed up the audio to “get through it” more quickly. Today’s students are so used to a fast and furious style of information presentation, that a narrated slide show may not be modern enough to hold their attention. However, video presentation in lieu of voice-over presentations may be more effective, especially in the teaching of statistics. In their 2018 study, Breneiser, Rodefer, and Tost found that teaching a psychology statistics course online presented unique challenges that were met by providing tutorial videos that mimicked classroom instruction particularly in learning SPSS (IBM Corp., 2013) and APA-style reporting. Through a process of scriptwriting and editing, they produced 28 tutorial videos that ranged in time from about three
minutes to 10 minutes in length. Much like the Breneiser et al. (2018) study, I have been using very brief (four minutes maximum) videos to illustrate concepts in my online statistics class and have found that students watch them over and over again. For some of the more difficult concepts (like the Central Limit Theorem, for example) students in my online class watched the video 3.25 times, on average. I title each short video with the concept (e.g., the difference between the mean and median) and teach that concept with a brief example in just a few minutes. For a typical weekly module in my class, I would have between five and eight short videos averaging about 25 minutes of video content in total. What I have learned by using these short videos is that students will actually watch them as opposed to posting the same content in a single 25-minute-long video. As an aside, no matter the length or style of videos or PowerPoint presentations with audio that you use, students value being able to replay presentations at faster speeds, so be sure that students are able to adjust the speed as desired (Ried, 2010). In accordance with Breneiser et al. (2018), online students who were provided short video content were just as successful on assignments and exams as were FTF students, which is great news for online statistics instructors who fear their online students are not learning as well.

**Figure 2. Photo generated by a student assessing shapes of histograms.** This “histogram” is positively skewed because the highest frequencies are at low values on the graph and the smallest frequencies are at high values on the graph, so the tail is skewed positively.

**Develop meaningful discussions**

Interaction among students and the instructor in an online statistics class is important to help students feel connected to their fellow classmates and their teacher, to enhance the learning experience, and to motivate students to stay on top of things and not fall behind in the course (Gazioglu, 2012). In their research on asynchronous online discussions, Aloni and Harrington (2018) reviewed the empirical evidence that discussions can foster critical thinking and writing skills. They suggest that online discussion boards should have a clearly stated purpose with clear expectations about how students will post and why, as well as a clear structure, effective prompts, and quality facilitation (Aloni & Harrington, 2018). One suggestion is developing questions that are open-ended where there are many possible responses, rather than a limited or right/wrong answer. I have used the discussion board in my online statistics class to give students the opportunity to apply what they are learning and creatively demonstrate that application to increase critical thinking. One example is asking students to use a set of household items to create a faux histogram and then describe the shape of the histogram they created (see Figure 2). Students upload a picture of the items and describe the shape they were
intending to create. This has been a very successful discussion post that has generated a lot of conversation about the creativity of the items pictured and given students a chance to correct each other if they believe a classmate has incorrectly identified the shape of their pictured graph. By having to create the shape themselves, students must go one level beyond just answering a multiple-choice question to actually develop the graph with their own hands. An added benefit of asynchronous online discussion boards is involving more introverted students who would not normally volunteer to participate in a FTF class environment (Aloni & Harrison, 2018). In this way, online discussion boards are for everyone and can aid in more equitable learning for the entire class.

Allow students to submit pictures of handwritten work

With advancements in technology, students are able to submit pictures and videos of assignments which can be particularly beneficial in statistics and other math courses, where the content often includes various symbols and formulas (Foster, 2003). I have found that allowing students to hand-write calculations using formulas and then take a picture with their phone or scan the hand-written sheets affords them the ability to spend their time learning the content rather than learning how to word-process mathematical equations. The technical savvy needed to use an equation editor can sometimes be overwhelming to students who are already confronted with learning the foreign language of statistics how to manage an online platform and computer software like SPSS. In Canvas, my LMS, I can circle and make notes on students’ submitted photos of their calculations so I can point out exactly where things may have gone awry. In support of the TBC characteristics (Kirby et al., 2018), allowing students the ability to post pictures of their work shows that you are a technically competent teacher who is invested in staying current and using the most up-to-date information and methods to ensure student success.

Hold online office hours

Some students in this challenging course need to have contact and a conversation in order to understand a difficult concept or to get “unstuck” when calculating an equation. Holding synchronous online office hours is one way to provide that support to students (Gazioglu, 2012). Using the technology available to you, you can video chat online with one or multiple students where you can work through problems or lab assignments and you can visually see what they are doing and where they are struggling. Of course, this requires that you are technologically competent, and if you do not know how to run online office hours you should strive to be a better teacher by attending a workshop or training module to learn how (Kirby et al., 2018). You could also tell students when you will be “live” and run office hours on your Discussion Board of your LMS if video chatting is not your preference. I often hold synchronous office hours for my statistics class for an hour in the late evening on the night before assignments are due. I find that students are happy to have the “just-in-time” help that fits their schedule of when they typically do homework. Of course, this means I must stay awake later than I normally would, but for one night per week it is feasible.

Conclusion

As online course enrollment continues to rise in the foreseeable future, it is up to instructors to make sure we are doing our best to ensure the success of all students, regardless of modality of instruction. Recognizing that doing our best online teaching may not include the same strategies as FTF teaching, we must embrace the differences and retool ourselves to be great in both formats. This is particularly important when teaching the challenging quantitative content of statistics. Being mindful of best practices regarding instructor behavior, communication, and
technical delivery of content can provide excellent online statistics courses to create the conditions for student success.
References


https://escholarship.org/uc/item/596195sg
Appendix: Sample Video Clip for Students

https://youtu.be/-HZuWSvoJr0
Flipping the Script on Stats Education: Presenting Evidence for and a Guide to Flipping the Psychology Stats Classroom

Donelle Posey, PhD and Amy T. Nusbaum
Washington State University

Summary

There has been a recent focus on shifting away from traditional lecture-based instruction, in which the instructor is the “sage on the stage,” to active learning formats, in which the instructor is the “guide on the side.” Activity learning has been shown to increase performance in a variety of college courses. A common active learning format is the flipped classroom. In this format, students interact asynchronously with course content between classes and interact with the content in a hands-on fashion during the class period. In this chapter, we use our experience in implementing a flipped classroom format in a psychological statistics course to provide tips and guidance for instructors considering flipping their courses. We include tips for designing activities, strategies for group formation, suggestions for classroom management, and tips for getting student buy-in, among others. In our experience, we have observed a decrease in failing grades and a generally positive response from students about the flipped format. In a course that can often be anxiety-inducing, using a flipped format increases students’ positive perceptions of and performance in a psychological statistics course.

Introduction

There has been a recent effort to move college classrooms from the “sage on the stage” model, or the standard lecture format, to one of the “guide on the side,” an active learning format – recognizing that traditional lecture-based content delivery may not be the best way to help students learn. The notion of these distinctions is not new; in 1993, Alison King defined the sage of the stage as, “the one who has the knowledge and transmits that knowledge to the students, who simply memorize the information and later reproduce it on an exam” (p. 30). This approach has been the preeminent method of instruction in college courses. On the other hand, the guide on the side approach is often referred to as active learning, a wide umbrella term for a course design in which focus is on students as active participants in their own learning and lecture-based content delivery is minimized. There are many forms of active learning that can take place in the classroom. These can include quick response technologies, such as clickers, that allow for real-time feedback of students learning (Poirier & Feldman, 2007), class discussions, simulations, and demonstrations (Yoder & Hochevar, 2005), and problem-based activities (Richmond & Hagan, 2011). In general, research on active learning has been positive (Richmond & Hagan, 2011; Yoder & Hochevar, 2005; Poirier & Feldman, 2007). A 2014 meta-analysis of 225 studies examining active learning in STEM courses found that students in courses using active learning techniques had exam scores that were 6% higher than those in traditional lecture courses, and students in traditional lecture courses were 1.5 times more likely to fail the class than those in active learning classes (Freeman et al., 2014). Active learning also has long-term positive impacts on student skill development outside of the classroom. Kilgo, Sheets, and Pascarella (2015) found that students who engaged in active learning during their
college careers displayed higher levels of critical thinking, need for cognition, and intercultural effectiveness.

One specific form of active learning that has gained traction in recent years is the flipped classroom. In the generic version of a flipped course, the bulk of the content transmission occurs asynchronously outside of the classroom. This is often, though not always, in the form of video lectures that the students watch prior to a given course day. As described above, the in-class work requires students to actively engage in applying the course concepts to an in-class activity. The flipped classroom format has been tested in many types of courses. The format has been found to improve student attitudes and performance in a general statistics course (Wilson, 2013), increase student perceptions of learning in a nutrition course (Gilboy, Heinerichs, & Pazzaglia, 2015), and enhance test performance in graduate-level health sciences classes (Tune, Surek, & Basile, 2013; Pierce & Fox, 2012).

In psychology-specific courses, the effects of flipped approaches are also generally positive. In a study examining a flipped format in a psychology statistics class, Winquist and Carlson (2014) found that students who had taken part in the flipped instruction performed better than the control group on a statistics exam over one year after taking the course. Notably, they did not have students watch videos prior to coming to class, but instead had them completing assigned readings. There also appear to be differences in flipped efficacy based on course level. Roehling, Luna, Richie, and Shaughnessy (2017) flipped an introductory psychology course and found mixed results; while students reported they found the flipped days more interesting, they also reported that typical lectures were more helpful to them in learning material and wanted a mix of both types of instruction. Additionally, there were inconsistent findings on exam performance, with flipped lectures seeming to benefit performance on topics that were addressed with the in-class activities, but acting as a detriment to topics not covered by the active learning activities (Roehling, Luna, Richie, & Shaughnessy, 2017). Borchardt and Bozer (2017) reported on a unique type of flipped design that may address some of the problems considered in Roehling et al. (2017). They used a “micro-flipped” format in a general psychology class, which incorporates the standard flipped course components of pre-class engagement with the material and in-class activities with some degree of lecturing in class as well. By the end of the semester, they found that the students in the micro-flipped format were performing better on exams (Borchardt & Bozer, 2017). This kind of format has similar benefits to the fully flipped formats described above, while also providing students with some structured lecture time.

Notably, many of these studies, particularly those done in psychology courses, have been conducted in courses that are relatively small. It can be logistically challenging to flip larger courses, and course size is certainly a consideration when deciding whether and how to flip. The few studies specifically conducted with large-enrollment courses have also generally shown positive results. Kishimoto, Anderson, and Salamon (2018) reported on a flipped physics course with an enrollment of over 100 students. They found that not only did students in the flipped classroom outperform those in the traditional classroom, but the flipped classroom format also led to a reduced gender gap (that is, men were answering questions correctly more often in the traditional classroom, and this gap between men and women was reduced in the flipped format; Kishimoto, Anderson, & Salamon, 2018). Studies of the flipped format with high-enrollment introductory chemistry class found that a flipped format leads to improved grades compared to the traditional format (Yestrebsky, 2015; Eichler & Peeples, 2015). Balaban, Gilleskie, and Tran (2016) found that grades in a flipped Principles of Economics course were higher than those in a traditional course.

Knowing the benefits of active learning generally and the flipped classroom specifically, the authors undertook the task of flipping a large psychological statistics course. This is an ideal course to flip because the principles, steps, and answers are concrete and objective with little
room for subjective interpretations. Furthermore, because statistics can be aversive and anxiety provoking (Onwuegbuzie & Wilson, 2003) using a more interactive format can make the class more enjoyable and less fear-inducing to students. This case study represents our experience in flipping a large-enrollment ($n > 160$) psychological statistics course. Over the course of three semesters, we transformed the course from an entirely lecture-based course to a fully flipped format. The first semester was entirely lecture-based, the second semester was half lecture-based and half flipped format, and the third semester was the fully flipped format. All content and materials were the same across semesters. In the semesters in which the flipped format was used, students had access to the same recorded lectures and activity worksheets posted in the course learning management system (LMS). Additionally, exams were identical for all iterations of the course.

**Methods/Approach**

During a typical in-class activity, students would access the activity worksheet posted in the course LMS. We presented necessary information, such as a complete data set, to the students during the in-class activity so as to control pacing of the activity. In some cases, we provided supplies or materials for students to use during in-class activities. We did not require students to submit completed activities for a grade, but did require short graded clicker quizzes to assess their understanding of the material following the in-class activity. Based on the answers to the clicker quiz, we provided additional instruction if needed.

**Preparing the flipped classroom**

Many faculty are wary of the time and effort involved in designing a flipped classroom (Miller & Metz, 2014). To be sure, careful preparation is the key to implementing a successful flipped classroom, but preparation also allows for more confident delivery of the flipped activities and leads to reduced workload in the future. Preparation needed to design a flipped classroom includes developing or finding materials to be delivered asynchronously to students prior to the class period in which the flipped activity will occur, designing the in-class activity, and creating or assembling any supplemental materials that will be presented as part of the in-class activity, among others.

First to consider is what information students will need to access prior to the in-class activity and how it will be delivered. Approaches to consider include recording lectures, mining YouTube for existing lectures (e.g., Khan Academy, Crash Course), having students read relevant chapters from the textbook, or using publishers’ ancillary materials. The technology and software needed to record lectures ranges from basic to advanced. Choosing the appropriate approach to meet your needs depends on your comfort level with recording or screen capture technology. To ensure students’ preparation for the in-class activity, you may consider having students complete a short comprehension quiz either prior to the class period in which the in-class activity will occur or at the beginning of the class period.

Another preparatory step is to develop an in-class activity. Developing an effective in-class activity is essential to the success of the flipped classroom. To accomplish this, you will need to consider the format, content, and pedagogical approach of the in-class activity. Activities can take the form of worksheets, handouts, or problem sets, or students can produce the results of the activity in their own notes. One approach is to use inquiry-based activities that are designed so that students uncover the principles of a concept, challenge their assumptions, then revise their understanding depending on the outcome of the challenge (e.g., Lee, 2012). Another approach is to provide a case study or simulation in which the principles are presented such that students can apply their understanding of course concepts (e.g., Herreid & Schiller, 2013). You may choose to demonstrate the concepts in a clever way using props or games, or show students how to work through problems. In preparing the in-class activities it is beneficial to
Exploring Descriptive Statistics: Types of Measures of Central Tendency (The Mean, the Middle, and the Most) and Their Graphical Representation

The purpose of this activity is to introduce students to types of measures of central tendency. To accomplish this, students will identify differences between types of measures and how they are represented graphically.

**Prerequisites:** Students must have an understanding of
- Scales of measurement.
- Orders of operation and the symbols associated with mathematical operations including $\Sigma$.
- Knowledge of basic statistics symbols including $X$, $M$, and $n$.
- Frequency distributions, relative distributions, and normal versus skewed distributions.

**Supplies:** Students will need
- Their iClicker.
- A calculator.
- A writing utensil.

**Primary Learning Objectives:** After completing this activity, students will be able to
- Define measures of central tendency.
- Identify measures of central tendency from frequency tables, frequency distributions, and raw data sets.
- Identify the relative positions of the measures of central tendency in graphs.

**Secondary Learning Objective:** Completing this activity will prepare students to
- Appreciate that scores in a distribution have a physical location on an X-axis.

**Model 1: Meanie...**

Model 1 depicts a set of data that represents scores on a self-esteem measure that used a 7-point scale in which lower scores reflect lower self-esteem and higher scores reflect higher self-esteem.

<table>
<thead>
<tr>
<th>X</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

**Formula 1:** $M = \frac{\Sigma X}{n}$

Use Model 1 and Formula 1 to address the following questions:

1. How many 6’s are in this data set?
2. How many 3’s are in this data set?
   a. Looking only at the frequency table, how do you know this?
3. The symbol $\Sigma$ means “sum all of…”
   a. Using words rather than symbols, what is the numerator of Formula 1 telling you to sum all of?
4. What is $n$ of this data set?
   a. Is it the same as $\Sigma f$?

Figure 1. Excerpt of in-class activity introducing measures of central tendency.
Prerequisite knowledge, supplies needed, and primary and secondary student learning outcomes are listed. Model 1 data are intentionally missing. This information is presented in class and incentivizes students’ attendance and participation.
consider the student learning outcomes (SLOs) you wish to achieve then work backward to design the activity. Another strategy for designing in-class activities is to carefully review the exam and design the activity to achieve mastery of the material. Figure 1 is an excerpt of an activity we used in the flipped version of the class.

The final step in preparing the in-class activity is to plan its pacing. One method of anticipating the pace at which students will work through the activity is to ask teaching or research assistants to time their completion of the activity. This can help you to prepare for the pacing of the class period.

Implementing the flipped classroom

Implementing the flipped classroom begins on the first day of class. From the beginning, model the behavior of interpreting data by explaining the results of studies of active learning and explain how active learning will be used in the class. Buy-in from students is an essential component to implementing the flipped classroom, but earning buy-in can be challenging. Some groups of students are averse to this approach because it does not conform to the traditional lecture-style classes they are familiar with, and they may also believe that the flipped classroom format is more work on their part, particularly if they have been in classes previously which required little more than rote memorization (McNally et al., 2016). Students may believe that they won’t learn the necessary material to do well on exams, for example, by being worried that case study activities may not prepare them for exam questions about terminology. Research shows that this is a common misconception of the flipped classroom format. Indeed, the flipped classroom format is no more work than the standard format (He, Holton, Farkas, & Warschauer, 2016). Ways to assuage students’ aversions include being transparent about the rationale for using the active learning format, using it immediately from the start of the semester, and mapping SLOs onto the in-class activities.

Next to consider is how you wish to disseminate the in- and out-of-class content. Out-of-class material can be posted to a course’s LMS which allows students to easily access preparative material and content. You should consider how students will access the in-class activities. To minimize the use of paper and costs to the department by photocopying more than 100 activities to distribute in class, you could consider posting materials in the course’s LMS and asking students to print a copy of the activity and bring it to class. You could also consider presenting the activity via projector during class and having students write answers on their own paper. If amenable to the use of laptops and mobile devices, you could allow students to access the activity electronically. Some flipped learning advocates argue that integrating technology into the flipped experience can make them more successful and integrate the course more fully into students’ lives (Hwang, Lai, & Wang, 2015). While there are often negative perceptions of allowing technology in the classroom, research suggests that the flipped classroom would be a place where distraction is minimal – students report that they are less likely to use their phone when engaged in a group activity or discussion (Berry & Westfall, 2015).

Some flipped classroom activities may require the formation of groups. While some students dislike group work (Barfield, 2010) the ability to work in teams is a skill that employers value (Stewart, Wall, & Marciniec, 2016). Options for assigning groups include defining rules for group formation (e.g., one non-major must be included in the group), randomly assigning students to groups, allowing students to self-select into groups, or assign students to groups based on skill-levels or personality characteristics. The latter option requires a pretest measure of a skill set or personality and, based on the scores on the pretest, forming groups such that a mix of skill levels or personalities are represented in the group. There are numerous group optimization algorithms that combine characteristics of students and instructor-preferred constraints, or group-formation rules, to determine which combination of students will result in effective group outcomes in team-based learning situations (Isotani, Inaba, Ikeda, & Mizoguchi, 2009; Ounnas,
David, & Millard, 2009). In addition to maximizing group success, these algorithms relieve the time and effort needed by the instructor to manually form groups. In one study on the formation of groups, students who were organized into heterogeneous groups on three characteristics, leadership, communication, and subject knowledge, outperformed students who were randomly assigned or self-selected into groups in both group-level objective outcomes and individual-level grades (Moreno, Ovalle, & Vicari, 2012).

Potential problems may arise when forming groups. Students may be unhappy with their group. You could consider reassigning groups periodically during the semester to promote variety of group membership. Two challenges you may face in group activities exist on opposite ends of the continuum. These include the student who overpowers or dominates the group and the student who does not participate. In a case in which a student dominates the discussion, consider ground rules about contributions from all members such as requiring all other group members to contribute before the dominating student may contribute. In a case in which one student isn’t contributing to the group, you will need to uncover the underlying reason for the student’s lack of contribution. In some cases, the reason may be because of a lack of preparation by the student. Requiring knowledge checks prior to the in-class activity could increase students’ contributions to the group work. If a student suffers from anxiety about communicating in a group, you might consider offering the student another way to participate, such as being the record keeper or having the student contribute written rather than oral answers for incorporation into the group’s final product. Another issue that may arise is the lone student who is reluctant to work in a group, for a variety of reasons. To address this issue, it is often simplest to have a contingency plan in place for students who want to work on their own (i.e., make sure the activity is capable of being completed by one person).

The flipped classroom is easily implemented with small class sizes but may present challenges in large-enrollment classes. The most obvious challenge is classroom management. In small classes, classroom management is more straightforward as your attention can be individualized and more equally distributed among groups. In large-enrollment courses, distributing your attention and individualizing your instruction is more difficult. To ensure students are receiving adequate attention, consider utilizing additional help from undergraduate teaching assistants. The management of off-task behavior can be a challenge to instructors in any type of classroom. Off-task behavior in flipped classrooms may occur for a number of reasons. It can indicate that a group has completed the activity or is waiting to proceed to another part of the activity. This can be addressed by increasing the pacing of the activity, or providing additional activities for groups that are prone to finishing early. Off-task behavior could indicate that students are disengaged from the activity or are not contributing to the product. To incentivize on-task behavior, grade students’ final product or otherwise require that they demonstrate their learning in minute-papers or clicker quizzes, or to report muddiest points. Another method of minimizing off-task behavior is to utilize classroom monitors, such as your teaching assistants, to monitor and facilitate on-task behavior.

Room characteristics and configuration may facilitate or hinder flipped classrooms. Ideally, the flipped classroom should have flexible seating and tables that can be arranged to support group work. Large rooms, however, tend to have fixed seating. If your classrooms have fixed seating, try to schedule a room with built in tables that students or groups can use to share materials and facilitate participation of all group members. The least conducive room configuration is fixed seating with tablet desks that lift and fold over the students’ seats. Quite simply, this configuration, while acceptable for standard lecture-based classes, does not allow for easily grouping students or provide adequate work space. In these situations, try to be flexible about where students work, e.g., think about allowing students to get up from their desks and sit on the floor or in the aisles while they complete the group work. Ultimately, the more flexible the classroom space, the better it is to implement the flipped classroom model.
Pacing is important to the successful completion of flipped classroom activities, but it is one of the more difficult components to plan. Students work at different paces, instructors may misjudge the amount of time it takes students to complete the activity, and answering questions or addressing confusion can interrupt pacing. One way to address this is to design deliberate stopping points in the activity. This allows for better pacing because students can work at their own pace but cannot proceed without additional information. Thus, throughout the class students are brought back in unison. Another way to manage the pacing of the activity is to be prepared with additional examples, mini-lectures, summaries, or quizzes if students complete the activity in less time than anticipated. Consider including advanced questions for students who finish early. These questions may not be required for the students to complete, but can keep them engaged with the material while other students complete the required activity. Finally, we highly recommend keeping notes following each in-class activity on the pacing of the activity, the elements that were successful, and the revisions needed for future iterations of the course.

A common SLO that is emphasized in college classrooms is communication, either in written or oral forms. Requiring students or groups to report results of the activity can achieve this goal. Reporting out can take a number of forms. It can include having groups merge to share their work and arrive at consensus answers. You can consider having each group select a spokesperson who is prepared to answer questions for the group. Some active learning approaches, such as Process-Oriented Guided Inquiry Learning (POGIL; Moog, Creegan, Hanson, Spencer, & Straumanis, 2006), assign group members to roles such as speaker, manager, and record keeper. You may ask select groups to work problem or write answers on the whiteboard or document camera, then engage in a class-wide discussion of the answers. In large-enrollment classes, you can call on a section of the room to provide answers. Potential problems that can occur in reporting results include students’ general reticence to answer questions in front of others, students’ fear of appearing unintelligent, and features of the activity or pacing that don’t allow for completion of the activity. To address these potential problems, prepare students for reporting out. Knowing in advance that they will be required to report out can allow students to prepare answers more confidently and complete the activity more efficiently.

To incentivize attendance and engagement in activities, consider grading students’ participation or learning. The most straightforward approach is to require students to submit their completed in-class activity. You may choose to grade on completion or accuracy of answers. A consideration in requiring students to submit their completed activities is that, depending on your method of providing feedback to students on their performance, you may need to return students’ completed work in a timely fashion so that they are able to use the activities to prepare for more formal assessments. Another approach is to administer a quiz to assess students’ understanding of the concept demonstrated in the in-class activity. Quizzes could take the form of a paper-pencil quiz or a clicker quiz. The benefit of a clicker quiz is that you can engage in “just-in-time” teaching wherein you can assess students’ understanding of the concept and address any unresolved misunderstanding on the fly. This method allows you to tailor your teaching to what your students still need to learn. A potential problem with requiring completion of the in-class activity occurs if you misjudge the pacing of the in-class activity. If the activity takes longer than anticipated to complete, students may not be able to submit a completed activity or be prepared to answer all the questions on a quiz.

Providing feedback in the flipped classroom

Formative feedback is a natural form of feedback in the flipped classroom. Indeed, well-designed activities in themselves can provide immediate formative feedback to students about their (mis)understanding of the course content. The goal of effective formative feedback is to
provide students with frequent, immediate assessment of their knowledge (Fink, 2013). Approaches such as clicker quizzes have the benefit of providing immediate feedback if the answer to the quiz questions are provided immediately. If you choose to use a paper-pencil quiz or require that students submit completed activities, returning the graded quizzes or activities in a timely manner is important so that students are able to integrate this information into their understanding of the course content and be better prepared for future summative assessments. If you choose not to require students to submit their completed activities, you should, at the very least, provide them with the correct answers to the in-class activities. This could occur after the completion of the activity while students are still present or following the class period in the course LMS. It may also be beneficial to review with students the SLOs for the activity and map them onto the larger unit or course SLOs.

**Results/Outcomes**

To measure student learning following in-class activities, we implemented a post-activity quiz using a quick response system. Unlike activity-style questions which tended to be open-ended or short answer questions, the quiz questions were exam-style multiple choice questions. We chose this approach for several reasons. First, it allowed us to observe whether students could generalize what they learned in the activity to multiple choice exam-style questions. Additionally, we were able to address obvious misunderstandings using just-in-time teaching. Second, it incentivized engagement in the activity, gave students practice with exam-style questions, and served as formative assessment for the students so they could gauge their understanding of the material.

The true test of the effectiveness of the flipped classroom approach came about with the unit exams. Unit exams were entirely multiple-choice questions that required conceptual and applied understanding of the course material. To determine the overall effectiveness of the flipped classroom format, we compared the aggregated lecture-based outcomes to the aggregated outcomes of the fully flipped version of the course. In our experience flipping the psychological statistics course, students generally performed better in the course, with a 6% decrease in failing grades, from 20% to 14%, and a 4% increase in C’s, from 26% to 30%. D’s, B’s, and A’s remained approximately equivalent between the two formats. Additionally, when asked to report their perceptions of the flipped format, students agreed or strongly agreed that they believed they learned more about course topics with the flipped format, enjoyed the flipped format, and preferred the flipped format to the traditional lecture-style format. Students also disagreed with the statement that there was too much work outside of class for the course.

**Benefits and Drawbacks**

Benefits of the flipped classroom are many. Research shows that the flipped classroom can improve student performance, and in our case, the lower failure rate in the flipped format compared to the standard format appears to support the research. Additionally, not only are students active participants in the creation of their own knowledge, they report enjoying the active learning format over the traditional lecture-based format. Indeed, the in-class activities are enjoyable as an instructor, too. Once materials are created, reusing the flipped content from semester to semester is straightforward. Finally, the topic of statistics can be less aversive to students when flipped activities are incorporated into the course.

Drawbacks include securing the initial buy-in from students. Students may approach a flipped class with skepticism and doubt of its effectiveness and the instructor must assure them that this approach is effective and will not jeopardize, and may in fact improve, their grades. Another drawback is the time and effort it requires to prepare the materials and implement the flipped format the first time. The preparation, however, is worth the successful implementation and
reusing of the course and its materials in the future. There may be elements of the course that require the use of new technologies, especially if you are recording or screen casting your lectures or using a quick response system. Learning new technologies is time consuming in itself and contributes to the time and effort necessary to develop and implement the flipped classroom. Depending on the materials used or developed, the flipped format allows for more resources available for students to utilize when learning and studying material, but without lecturing directly to students, it is difficult to feel confident that students are learning what you intend for them to learn. Though you may plan periodic knowledge checks, you may not be sure that students have mastered the material until exam time.

It is also important to remember that our students are not merely ID numbers on the grading roster, but instead complex individuals who exist in an inequitable world. While a flipped classroom offers many benefits, if done insensitively it can harm the most marginalized students in our classrooms. Anna Perry wrote a blog post on their experience in a flipped physics classroom that we should all consider as we structure active learning experiences (2019). They discuss how being forced to work with peers who were both implicitly and openly discriminatory with no alternate eventually led them to change their major and leave physics altogether. Most powerfully, they say:

There is inequality in our society and there is inequality on this campus. There is a silent conflict between the members of our classrooms, whether we are aware of that or not. Unless a professor actively works to dismantle this inequality within their classroom, they are reproducing it. I do not leave my identity at the door of the physics classroom. I carry it with me—whether I want to or not—when I interact with my peers, when I listen to my professor speak, when I read the textbook. I am a political being because I live in a world that politicizes my gender, my race, my looks, my socioeconomic status. Society likes to ignore this, and thus we let asymmetrical power dynamics continue to thrive. Professors ignoring this results in more of the same (Perry, 2019).

So, while the flipped classroom can be a tool to improve the learning experience of our students and help them harness their own power as learners, it can also be an unintentional tool to further widen educational inequities. We as educators must be careful in the ways we think about, structure, and reflect on the active learning experience to ensure this does not occur. How we handle these potential problems will depend on our institution, teaching styles, and student populations. However, some things to think about include allowing students to change groups if you have assigned them (and making that option clear to students), creating behavior contracts with your students (either as a whole, or within each group if groups are permanent), and setting standards for respect in all classroom activities. While no instructor can prevent all forms of discrimination, we can set up our classrooms in ways that reduce the likelihood of those behaviors.

The flipped classroom format has been shown to generally improve student performance in a variety of disciplines and our experience flipping psychological statistics bears out those results. It has been our experience that the key elements to successful implementation of the flipped classroom include the preparation necessary to execute the flipped format, the management of in-class components, and the provision of formative feedback to the students. The guidance we provide in this chapter arises from this experience.


Encouraging Reflective Learning in Introductory Statistics

Ashley Waggoner Denton, PhD
University of Toronto

Summary

Reflection has the power to transform an ordinary learning experience into an extraordinary one. According to Fink’s (2003, 2013) taxonomy of significant learning, when we move beyond content to focus on other aspects of learning, we can create lasting change in a learner. In this chapter I argue that the incorporation of reflective learning into a statistics course can enhance the student experience and promote lifelong learning. Reflection can take many forms and there is no “correct” way to incorporate reflection into a course. However, different reflection activities may promote certain forms of learning more than others. For example, reflecting on study strategies may help students become better learners, whereas reflecting on lecture content may help increase the perceived value of the course material. Importantly, because the enhancement of any form of learning is likely to enhance other forms as well, any reflection activity has the ability to affect numerous aspects of the learning experience. Here I describe my use of a reflective learning journal in a statistics course and the impact it can have on students. I also review other reflective strategies that have been used in statistics (or similar) courses. By providing an overview of these various methods, it is my hope that instructors will be able to identify a strategy that can be successfully adopted or adapted for use in their own course.

Why? The Benefits of Reflection

For many psychology students, the statistics course is nothing more than an anxiety-provoking hurdle that must be overcome – a hurdle filled with numbers and equations and that bears little relevance to things they actually care about (Conners, McCown, & Roskos-Ewoldsen, 1998; Schutz, Drogosz, White, & Distefano, 1998). For instructors, these misperceptions and attitudes are commonly viewed as challenges to teaching statistics, and rightly so (e.g., statistics anxiety can lead to numerous maladaptive learning strategies; Gonzalez, Rodriguez, Failde, & Carrera, 2016). But because of these challenges, the statistics course also provides an opportunity that is ripe for students to gain critical insight into themselves as learners. Students can walk away from their statistics course with a deeper understanding of how they coped with anxiety, developed new skills, overcame set-backs, and made connections across topics (to quote a former student, “not only did I learn statistics, but also lifelong learning strategies”). However, these insights won’t just happen spontaneously. They require reflection.

As many have pointed out, reflection is a powerful tool (e.g., Dewey, 1933; Boud, Keogh, & Walker, 1985; Dietz, 1998; Mezirow, 1991; Costa & Kallick, 2008). A successful learner might be able to calculate a t-statistic and interpret what it means. A successful and reflective learner not only knows how to do this, but also understands why they did it and how they learned it (which are important components of statistical thinking, see goal 1 of GAISE, 2016). They might recognize, for example, how putting the formula into their own words, rather than using the symbols, helped them understand the steps involved in the computation or helped keep their
anxiety at bay. With this insight, the student can then take this knowledge and transfer it to new learning experiences (see Brockbank & McGill, 1998). Reflection also enables the learner to make connections between what was learned, what they already know, and what they are learning in other contexts (Dietz, 1998; O'Rourke, 1998). Returning to our t-test example, a reflective learner might think about how they could use this new knowledge to answer a question they have always wondered about (does Dominos deliver faster than Papa John’s?) or recognize how it has been applied in their other courses (so this is how they determined that dog people are more extraverted than cat people!). As a result of reflection, new mental maps may form, problems may be solved, issues may be clarified, and meta-cognitive skills may be developed (e.g., Boud, Keogh, & Walker, 1985). Thus, reflection also encourages increased conceptual understanding (goal 2 of GAISE 2016). Reflection can take many forms and have numerous consequences, but ultimately it has the power to transform the statistics course from a knowledge-acquisition event into a significant learning experience.

Fink’s (2003, 2013) taxonomy of significant learning includes six major categories: foundational knowledge, application, integration, human dimension, caring, and learning how to learn. In this model, each of form of learning connects to the others in a synergistic manner such that enhancing any category of learning (e.g., emphasizing how the subject matter integrates with other topics) tends to increase other forms as well (e.g., by making those connections, the learner may come to care more about the subject). When all six categories are promoted, a significant learning experience – defined as one that creates lasting change within an individual - is likely to result.

In an introductory statistics course, foundational knowledge is so important, and often so challenging for students, that it can be easy for instructors to spend the majority of their time focused on content. But without reflection, much of what is learned may be limited to the statistics classroom as students fail to realize how transferable their newfound statistical knowledge is (despite how often we may tell them). Students may also fail to realize how they managed to learn this knowledge in the first place. However, with a bit of reflection, any number of the other categories from Fink’s taxonomy may be enhanced (see Table 1). For example, being prompted to reflect on the most important thing that was learned during class each day might help a student recognize the value in the material they are learning (enhancing the caring dimension). By keeping a reflective notebook, a student may realize how much their confidence or statistical self-efficacy (see Finney & Schraw, 2003) has increased over the course of the semester (enhancing the human dimension).

However, for many reflection activities, the most relevant learning category is learning how to learn. One interpretation of learning how to learn involves learning how to become a better student. When students are encouraged to reflect on their learning it can improve their self-monitoring and goal-setting capabilities and lead to changes in study habits and other skills (Sweidel, 1996). Another form of learning how to learn involves helping students become “self-directing learners” (Fink, 2013, p. 59). Self-directing learners are able to recognize gaps in their understanding and formulate plans for filling those gaps. Rather than relying on an instructor to tell them what they should learn next, a self-directing learner is able to develop their own learning agenda, along with a strategy that puts their plan into action (Fink, 2013). These are precisely the skills that I have witnessed my students develop through the use of a reflective learning journal.

**How? The Methods of Reflection**

“The reflective journal entries allowed me to take a step back from number crunching and see how I, as a student, am processing and digesting the material. For each entry, I
got to see things from a larger perspective, such as how my studying habits were affecting my understanding of the concepts or how the textbook could come to be so
<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Benefits of Reflection</th>
<th>Potential Reflection Prompts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foundational knowledge</td>
<td>Understanding and remembering facts and ideas.</td>
<td>Enhanced recognition of what is and is not understood.</td>
<td>What is the most important thing you learned today? What are you still confused about?</td>
</tr>
<tr>
<td>Application</td>
<td>Engaging in a new action, including the development of skills and new forms of thinking.</td>
<td>Practice new ways of (critical, creative, and/or practical) thinking.</td>
<td>How are you using what you learn in the course?</td>
</tr>
<tr>
<td>Integration</td>
<td>Making connections between ideas, experiences, and realms of life.</td>
<td>Development of new or deeper connections.</td>
<td>How does this material relate to what you are learning about in your other courses?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>How is your approach to learning similar or different in this course compared to your other courses?</td>
</tr>
<tr>
<td>Human Dimension</td>
<td>Learning about oneself or others.</td>
<td>Improved self-understanding. Enhanced recognition of the personal and social implications of what is learned.</td>
<td>Why is it important to learn about statistics?</td>
</tr>
<tr>
<td>Caring</td>
<td>Developing new feelings, interests, and values.</td>
<td>Greater appreciation of the subject. Enhanced motivation and energy for learning.</td>
<td>How do you feel about statistics? Have your feelings about statistics changed over the course of the semester?</td>
</tr>
<tr>
<td>Learning How to Learn</td>
<td>Learning about the process of learning itself.</td>
<td>Improved study habits and self-monitoring skills. Development of learning agendas and strategies.</td>
<td>How are you learning in this course? What are your learning strengths and weaknesses? What are your learning priorities?</td>
</tr>
</tbody>
</table>
useful. As such, this journal served as a pushing hand, steering me in the direction toward becoming an efficient and knowledgeable student." (Fall 2018 student)

Reflective Learning Journals

In my own course, I encourage reflection through the use of a reflective learning journal (a practice commonly undertaken in nursing education programs; see Thorpe, 2004). Details surrounding the first iteration of this assignment can be found in Waggoner Denton (2018) (also see Chapters 8 and 9 for more information on the incorporation of writing in statistics courses). As described in the course syllabus:

A reflective learning journal is a record of the reflective thought and meaning you are making as you engage in a learning experience. Thinking about your thinking – or ‘metacognition’ – can help you become a better learner. Statistics in particular can be a challenging class for students because for many it is associated with a lot of anxiety. By reflecting on your feelings, you will understand them better and be able to manage them more effectively. Even if you have no anxiety, reflecting on your learning throughout the course will help you become a better learner (and who doesn’t want that?). In addition to reflecting on your learning, you can also use these journals to monitor your progress, problem-solve, and track your learning goals and priorities.

I introduce the journal assignment on the first day of class and explain its intention and the importance of reflective learning (see Appendix). I also highlight that a reflective learning journal is not intended to be a personal diary or a simple log of learning-related events. The journal is worth 5% of the course grade (enough to provide motivation to complete it) and is submitted online (through our course learning management system). Having the journal completed online is convenient for the students, and makes the assignment manageable in my large (200 student) class. Students are responsible for submitting a minimum of six or eight journal entries (see below) over the course of the 12-week semester, and all entries must be submitted at least five days apart from one another (since the purpose of the assignment is to encourage reflection throughout the course). While it would also be reasonable to assign specific deadlines for each entry (e.g., every other Friday), I prefer to let the students manage their journal entries themselves. This also saves me the trouble of dealing with extension requests and other issues that would arise if I assigned specific due dates.

The first time I used this assignment (Fall 2015), students were required to submit eight journal entries, and the journals were graded on completion (see Waggoner Denton, 2018). In the most recent version of the course (Fall 2018) I decided to place more emphasis on quality over quantity. I dropped the number of required submissions from eight to six, and instead of grading solely on completion, at the end of semester I randomly selected two entries from each student (by rolling a die) and factored the quality of their reflection into their final journal mark. Unlike some of the other methods that are reviewed below, I do not provide specific prompts for the journal entries. Rather, students are provided with general instructions (both in class and within the course LMS; see Appendix) on what the journals should look like (anything from one-paragraph to a single page) and what a reflective learner does. But ultimately, they have the freedom to write about whatever they chose, whenever they chose to do so. Potential benefits of the journals are reviewed in the section What? The Effects of Reflection.

Alternative Methods

Below I review some additional reflection strategies that others have used, highlighting the key features of each strategy as well as any notable outcomes for the learner. Each of these

---

1 Please feel free to use (verbatim or modified) in your own course, no credit necessary.
approaches can be viewed as an active learning strategy (goal 4 of GAISE, 2016) in the sense that they encourage students to think about some aspect of a classroom activity they have recently engaged in. Table 2 allows instructors to quickly compare and contrast each strategy along metrics that may aid in the process of deciding whether a particular strategy may work for their course or not. However, it is also important to keep in mind that all of these strategies are customizable and certainly do not need to be used “as is”. For example, although most of these activities were conducted using paper and pencil methods, there is no reason why they could not be moved online (see Chapter 3 for advice on online teaching strategies).

**Reflective Notebooks.** While the reflective learning journals emphasize critical reflection over diary-style exposition or rote record-keeping, Salinas (2004) reports on the use of a reflective notebook that incorporates all of these forms of writing. Student teachers in a mathematics course were asked to keep a notebook for the class that contained not only their class notes and assignments but also “a mixture of journal-writing, record-keeping, and diary-style noting of thoughts” (Salinas, 2004, p.316). Students were also encouraged to ask questions about the material. Every week the instructor would collect the notebooks and write back to the students. At the end of the semester, the students received one grade for maintaining the notebook and another grade for creating a portfolio that included 10 excerpts from the journal demonstrating their growth in understanding and/or confidence. By examining the students’ entries from the beginning to the end of term, Salinas (2004) concludes that the students appeared to develop new interpretations of mathematics and an appreciation of why it was important. They also began to reflect on and evaluate their own thinking and learning. Notably, the back-and-forth communication between the instructor and the students was also seen an integral part of the course. The students enjoyed the activity and perceived it as useful. However, it is impossible to know how many of the perceived benefits of the reflective notebook came from the conversational aspect of the notebook. There is no report of whether those who kept better notebooks performed better in the course or improved along any other dimension.

**Learning Check-In.** McGrath (2014) describes the use of a “learning check-in” in which the students in her statistics course were required to email her with two questions, meet with her for 20-30 minutes, and then complete a learning reflection form. She found that students who were randomly assigned to complete the learning check-in before the second test performed better on the test (controlling for scores on the first test). However, it is not clear whether it is the question generation, the meeting, the reflection, or some combination of these things that are responsible for the improved grades. The learning reflection form ask students to: reflect on their learning in the course, list behaviors that both help and hinder their learning, identify three behaviors they could begin, stop, or modify in order to help them succeed in the course, and finally, develop a study plan for the next test. It is not known whether the reflection form was graded or not, or how the students felt about the activity or meeting with the instructor.

**One-Minute Papers.** One-minute papers are a common way of encouraging students to engage in reflection at the end of a class. In their typical form, students are asked to take a few minutes to write down the most important thing they learned that day, as well as any questions that remain unanswered, although other variations exist (see Stead, 2005). Research has indicated that these papers can improve student performance, although they often take much longer to complete (up to 10 minutes of class time) than their name implies. Chiou, Wang, & Lee (2014) implemented a one-minute paper in an introductory statistics course (using the traditional two-part format) and found that compared to a class where the students instead spent the last 5 – 10 minutes reviewing a textbook exercise, students in the one-minute paper condition had significantly lower statistics anxiety and performed better on tests. Notably, the course instructor responded to each paper at the following class meeting. All of the students reported that they felt the one-minute papers were helpful and their comments indicate that writing the one-minute papers helped them reflect on their understanding and think more deeply about the material.
Table 2. Overview of reflective activities.

<table>
<thead>
<tr>
<th>Reflective activity</th>
<th>Amount of in-class time needed</th>
<th>Modality</th>
<th>Amount of work for instructor</th>
<th>Student Perceptions</th>
<th>Feasible for large classes?</th>
<th>Most relevant learning categories from Fink’s taxonomy</th>
<th>Key Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflective learning journals</td>
<td>None</td>
<td>Online</td>
<td>Minimal</td>
<td>Mixed</td>
<td>Yes</td>
<td>Learning how to learn</td>
<td>Waggoner Denton, 2018</td>
</tr>
<tr>
<td>Reflective notebooks</td>
<td>None</td>
<td>On paper</td>
<td>Significant</td>
<td>Positive</td>
<td>No</td>
<td>All</td>
<td>Salinas, 2004</td>
</tr>
<tr>
<td>Learning check-in</td>
<td>None</td>
<td>Online + in-person</td>
<td>Significant</td>
<td>Unknown</td>
<td>No</td>
<td>Learning how to learn</td>
<td>McGrath, 2014</td>
</tr>
<tr>
<td>One-minute papers</td>
<td>5 – 10 minutes</td>
<td>On paper</td>
<td>Minimal to moderate</td>
<td>Positive</td>
<td>Maybe</td>
<td>Foundational knowledge; caring</td>
<td>Chiou et al., 2014</td>
</tr>
<tr>
<td>Brief reflective responses</td>
<td>2 minutes</td>
<td>On paper</td>
<td>Minimal to moderate</td>
<td>Positive</td>
<td>Maybe</td>
<td>Caring; integration</td>
<td>Hamilton &amp; Mallet, 2018</td>
</tr>
<tr>
<td>Homework and exam wrappers</td>
<td>5-10 minutes</td>
<td>On paper</td>
<td>Minimal</td>
<td>Positive</td>
<td>Maybe</td>
<td>Learning how to learn</td>
<td>Chew et al., 2016</td>
</tr>
</tbody>
</table>

Brief Reflective Responses. Recently, Hamilton and Mallet (2018) reported on a modified version of the one-minute paper that they call brief reflective responses. Students can complete the activity at any point during the class and are given only two minutes at the end of the class to finish (or quickly write) their reflection. These responses are also meant to be more flexible, as the students are not given any specific prompts. To assess the potential benefits of this flexibility, Hamilton and Mallet (2018) randomly assigned students in a social psychology course to complete reflective responses throughout the semester that were either self-prompted or instructor-prompted. Although they did not find any differences in exam grades or enjoyment of the activity between the two groups, completion of the self-generated reflective responses led to greater perceived relevance of the course material, which in turn predicted exam performance. Hamilton and Mallet (2018) suggest that when students are not given a specific prompt to respond to, they may consider a wider-range of course material and make more connections to their personal lives. However, it is important to note that because this activity was tested in a social psychology course, where personal connections with the material often abound, it is unclear how well such an activity would transfer to a statistics course.

Exam and Homework Wrappers. An exam wrapper is a structured reflection activity that occurs after an exam is returned. Students are asked to reflect on their exam preparation, the errors they made, and the modifications they will make to their study habits moving forward (Lovett, 2013). While the reported effects of exam wrappers on student performance have been mixed (e.g., with null findings reported by Soicher & Gurung, 2017 and Stephenson et al., 2017), Chew and colleagues (2016) have used exam and homework wrappers with success in an
introductory statics (the study of forces without motion) course for engineering students. Students completed homework wrappers during class time after the return of two graded homework assignments. They also completed a pre and post-exam wrapper for the second exam in the course. The homework and post-exam wrappers included four components: student satisfaction with their performance, the types of mistakes they had made, their level of confidence in their knowledge, and open-ended questions regarding their study strategies. The pre-exam wrapper focused solely on study strategies. Students found the wrappers enjoyable and perceived them to be a useful learning activity. Due to the lack of a comparison group, it is not known if the use of the wrappers improved students’ performance in the course.

**What? The Effects of Reflection**

Here I will focus on the effectiveness of the reflective learning journals. Key outcomes for the other reflective activities have been included in the descriptions above and additional details can be found in the original articles.

**Performance.** As reported in Waggoner Denton (2018), students who completed more journal entries performed better on the final exam, \( r(171) = .29, \ p < .001 \), controlling for scores on the first test. Combined with qualitative data from the students (from Fall 2015), this suggested the possibility that students who wrote more journal entries were better able to manage their time and/or emotions and were perhaps utilizing more effective study strategies and learning from past mistakes. In Fall 2018, the quality of the students’ journal entries was factored into the journal grade (rather than the journals being marked solely for completion as was the case previously). Out of the 181 students in the class, 19 received zeros for failing to submit at least five entries. Marks for those who submitted at least five entries ranged from 1-5, with an average grade of 4.05 (or 81%) (SD = .71). An examination of course data from Fall 2018 indicates that the marks students received on the journal entries continued to be predictive of final exam grades even after controlling for test 1 scores, \( r(175) = .31, \ p < .001 \). Results of a multiple regression analysis predicting final exam grades from test 1 scores, scores on weekly problem sets, and marks on the reflective learning journal (all entered as percentages) indicated that the model explained 49.3% of the variance in final exam scores, \( F(3, 173) = 57.95, \ p < .001 \). All three predictors were significant, with the journals’ contribution (\( B = .091, \ t = 2.98, \ p = .003 \)) comparable to that of the weekly problem sets (\( B = .186, \ t = 2.68, \ p = .008 \)). This is a notable finding, given that (unlike the problem sets) the journals do not provide the students with feedback or involve actual practice with the course content.

**Perceptions.** In Fall 2015, student perceptions of the journals taken at the mid-point of the semester were mixed. However, qualitative data from the end of the semester suggested that many students came to develop an appreciation of the journals (Waggoner Denton, 2018). This is a trend that I have continued to notice. As an example, here are two journal excerpts from Fall 2018:

However, what I realized is that these review journals were actually very helpful. It helped me to analyze my studying techniques and actually revise it so I could try something new. In the beginning i felt that the prof was giving out useless additional work for us to do. I thought that no one is going to be truthful in their responses like they are just going to "BS" on it. Basically, that was my strategy to get through it too. On the other hand, after my first midterm exam is when everything changed. I was crashed by the mark I got, so i wanted to make a change therefore, I went back into my review
journals and looked it over. I realized the mistakes I have made and tried to improve on it in every journal I did... I'm happy that our professor asked us to do this.

While I was rather unconvinced at first that writing these journals would particularly benefit me in any noticeable way, I realized that it has encouraged me to look back at my own learning processes, as well as how to improve them. By reflecting weekly on my habits, I have noticed things about my own actions that I was unconscious of before. For example, I realized that I have a habit of skimming through parts of the textbook without truly registering the information if I had gone over the material before. Therefore, instead of reading those sections again, I did practice questions related to those sections, and looked back at the chapters whenever I did not understand a concept or formula. I would not have noticed my lack of attention if I had not reflected on my past actions, and realized what areas of the material I usually struggled more on. I am thankful for having the opportunity to keep a reflective journal for this course, as I most likely would not have thought to do this otherwise. Keeping a reflective journal is a very effective learning tool for me, and in the future, I intend to continue to keep a reflective journal in order to further improve my studying habits and techniques.

While not all students benefit from the journals (of course not everyone is going to take it seriously), I receive numerous comments from students each semester on how transformative the journals have been for them. For example, the following is an excerpt from a Fall 2018 student course evaluation:

I would like to end this review by touching on one of the most effective teaching strategies that I've ever encountered during my educational journey: journaling. Professor Denton embedded into student evaluations the requirement to complete a weekly journal each week. In these journal entries, we were asked to reflect on the material covered in the past week, our thoughts about statistics and the impact that it has on our lives, and comments about our successes and challenges in statistics class and beyond. As a result of these entries, I was able to reflect on my learning not only in PSY201, but in other classes as well. This was a deeply empowering experience, and it led to me setting new goals for myself in all of my classes. As a result, I felt more confidence going into the final examination period, and I took active control of my academic performance.

It is this type of feedback that encourages me to keep the journal as part of my statistics course. In the future, I will be returning to a requirement of at least eight journal entries, as I do not believe that grading the quality of the journals led to any improvement in overall quality (see also Sumson & Fleet, 1996; Dyment & O'Connell, 2011). By completing more entries, the practice of reflection is more likely to become habit-forming. As with most things, the more practice the students get with reflection, the better.

**Conclusion**

Teaching statistics presents unique challenges for instructors, and although there is no silver bullet, I do believe that many of these challenges can be turned into strengths when students are encouraged to engage in reflection. The feedback that I have received regarding the reflective learning journals has been among the most valued feedback that I have received in my teaching career, because it indicates that the students have not just learned about statistics, but that they have learned things about themselves they will carry with them into the future. Whether you adopt one of the strategies outlined here, mix a few of them together, or end up
coming up with something completely new, I hope that you have been inspired to encourage more reflection in your students. Those who take it seriously will be the better for it.
References


Appendix:

- These are an example of the slides that I use when introducing the reflective learning journal to the class.

Reflective Learning Journals

- Learn more effectively
- Organize your thoughts and emotions
- Track and evaluate your progress

- A Reflective Learning Journal is NOT
  - A diary
  - A log
  - Your study notes
  - A discussion board

Example of Reflective Learning

“We do not learn from experience... we learn from reflecting on experience.”
—John Dewey

Throughout the course there have been alternative equations presented in the text and in lecture and until this point I chose to stick to one method in fear of creating more confusion. While studying for the final, I decided to try an alternative method presented in lecture for pooled variance and realized that it wasn’t more confusing but in fact the opposite. I guess I figured that if I understood one method why continue floundering by learning another. As I continued with this approach the problem I found that it actually gave me a more thorough understanding of how the numbers worked. So I’m thinking that I’ll go back and try do this with other equations while prepping for the final.”
Part Two: Course project ideas
Passion-Driven Statistics

Kristel M. Gallagher¹, PhD, Kristin R. Flaming², PhD, and Lisa C. Dierker³, PhD
Thiel College¹, Weslyan University²

Summary

How can we infuse passion into our students’ experience of introductory statistics? How do we show our students that statistics matter, not just in the realm of psychology but in their own lives? Our chapter describes a passion-driven, project-based curriculum aimed at reframing the traditional introductory statistics experience to allow students to tackle exciting questions within the field of psychology that matter to them. Known as Passion-Driven Statistics, and funded by the National Science Foundation, our approach specifically engages students in authentic projects with large, real-world data sets. Utilizing a flipped classroom approach, our students answer research questions of their own choosing with descriptive and inferential statistics. They also learn statistical programming in the pursuit of managing and analyzing data. At the end of the semester, our students present their original work at a research poster session in which they have the opportunity to share their statistical journey with their peers and campus community. Excitedly, our research shows that students who enroll in this project-based course are more likely to show an increase in their interest in pursuing advanced course-work in statistics compared to students who enroll in traditional statistics courses. Our chapter will provide you with the resources to engage your students in a passion-driven pursuit of statistical knowledge and understanding!

Overview and GAISE/APA Statistical Literacy Alignment

As instructors of introductory statistics, we have long recognized that a central challenge of this course is the development of a curriculum that not only serves diverse students with varying levels of preparation, but also sparks communication, reasoning, and collaboration that crosses traditional disciplinary boundaries. Many of our students arrive in our classroom with a fear of math, countless of whom self-select into psychology and related social science fields explicitly to avoid math. To that aim, Passion-Driven Statistics was developed as a multidisciplinary, project-based introductory statistics course that engages students in applied statistical projects that cross both divisional and departmental boundaries (Dierker, Kaparakis, Rose, Selya, & Beveridge, 2012). Funded by the National Science Foundation (grant #0942246, #1323084, and #1820766), our curriculum engages students in statistics in a format that is unique and refreshing for students and instructors alike. Before diving into the nuts and bolts of our passion-driven approach, we want to first highlight some of the core features of our course. Importantly, we provide this overview through the lens of the recommendations outlined in the Guidelines for Assessment and Instruction in Statistics Education (GAISE) report (GAISE, 2016) and the goals of the APA Statistical Literacy Task Force (APA, 2014).

For example, our course teaches students statistical thinking (GAISE recommendation #1) by asking them to choose their own research project and giving them time to think critically about statistical issues related specifically to their project. Notably, our students learn to recognize the
usefulness of data for answering questions of interest to them (and to society) and tackle complicated real-world questions that involve more than just one or two variables (GAISE recommendation #3). Further, by supporting students to make their own decisions about the most appropriate ways to visualize, explore, and, ultimately, analyze data, our course also emphasizes conceptual understanding over hand calculations (GAISE recommendation #2). Rather than focusing on rules associated with traditional lists of statistical tools (e.g., z-test, one sample t-test, two sample t-test, paired t-test, etc.), we organized our course to focus on the decisions and skills involved in statistical inquiry, such as the selection and interpretation of tests (see Chapter 12 on Using StatHand to Improve Students’ Statistic Selection Skills).

Our Passion-Driven Statistics course also covers basic themes such as measurement and descriptive and graphical representation (APA Statistical Literacy Goal #1), as well as more specific inferential methods needed to test hypotheses and/or explore the empirical structure of data. What makes our approach unique is that all of these topics are introduced as our students’ own research questions dictate. In this way, our students are provided with opportunities to learn to select the tools that are most appropriate to address their research questions and to engage in statistical decision making, a process that will likely be mirrored many times over for any student who continues on to graduate work in the field of psychology. Using leading statistical software, including SAS, R, Stata, Python or SPSS, (GAISE recommendation #5), our students are exposed to a wide variety of statistical methods and learn to choose and use them flexibly as they are needed (APA Statistical Literacy Goal #2).

Further, our course is project-based and uses a flipped classroom approach (see Chapter 3 on Flipping the Script on Stats Education), which we have found is an effective means of freeing class time for active problem solving (GAISE recommendation #4) because students view lecture videos outside of class. In our flipped project-based class, the majority of each class session is devoted to students actively working with instructors and peer mentors on their analyses and thinking about their data. In the first week, our students develop their own research question after an introduction to a number of large data sets. In the weeks that follow, students’ research questions evolve as they continue through the course and apply newly learned statistical techniques. This is an exciting time as instructors because students begin to take ownership of their research questions as the “big picture” of the course comes into focus for them.

All statistical analyses are conducted within the context of our students’ research questions, culminating with a poster presentation that allows for a meaningful summative assessment of student learning (GAISE recommendation #6). At this concluding event, students have the opportunity to describe to peers, instructors, and external experts their process of inquiry, including the different decisions made along the way and their premises, conclusions, and any barriers that they faced (APA Statistical Literacy Goal #5). Our project-based course is designed to take advantage of students’ natural curiosity and provide a common language for approaching questions across numerous disciplines (Cobb, 2007). The focus of the course is on the decisions and skills involved in asking and answering questions with data, with the overarching goal of telling accurate and engaging stories.

**Implementing Passion-Driven Statistics**

So, what does Passion-Driven Statistics look like “on the ground” and how can you go about infusing passion into your introductory statistics course? Passion-Driven Statistics can be applied flexibly to suit your teaching style, course goals, and constraints. Our approach has been implemented in statistics courses, research methods courses, individual research tutorials/studies, and summer research boot camps. We have engaged students from a wide variety of academic environments, ranging from high school-aged students to second year
medical students. Excitedly, many psychology departments across the United States have begun to adopt the Passion-Driven Statistics curriculum, in full or part, as the primary statistical teaching model for their major. However, this curriculum can be tailored to be taught in as little as 3 weeks or as long as a full semester.

How do we accomplish this? Regardless of the format or setting, all students choose from a variety of large, real-world, and publicly accessible data sets. Our psychology students, for instance, are often drawn to The National Longitudinal Study of Adolescent to Adult Health, the “Add Health” dataset, because of the diversity of social and behavioral variables (and more recently, biomarker data). The publicly available waves of data (Waves I – IV; https://www.cpc.unc.edu/projects/addhealth/documentation/publicdata) from this longitudinal data set allows students to choose a nationally representative sample of adolescents, young adults, or adults to answer their research questions. Some of the popular topics our students are drawn toward include (of course) sex, drugs, politics, religion, and relationships.

![Figure 1. Study Design of the “Add Health” Data Set](https://www.cpc.unc.edu/projects/addhealth/design)

Based on the type of data students choose, each generates testable hypotheses, prepares data for analysis, selects and uses descriptive and inferential tools, and evaluates, interprets, and presents research findings (orally, graphically, and as written text). Importantly, our students do no experience activities as distinct separate tasks. Rather, these activities happen as a series of ongoing interactive tasks with both in-person, one-on-one support and open access, on-line materials. Our course materials, provided freely to all instructors, are intensive enough to allow students to constantly move forward with their project, yet broad enough to force each student to creatively explore their questions and make the decisions involved in data analytic inquiry.

All of our students, regardless of the course format or setting, present their completed projects during a poster session. Even in our condensed versions of the curriculum, we are sure to include this culminating event as it provides our students with an empowering experience, allowing them to communicate methods, results, and insights not only to data experts but also to non-scientific audiences (see Chapter 8 on Writing to Improve Statistical Comprehension). Like the curriculum itself, the formality and structure of the poster presentations vary based on the goals and resources of the course instructor. While some instructors chose to conduct a rather
informal classroom presentation, using multimedia slides on a projector, other instructors organize a full-scale campus event with printed posters displayed on easels or poster boards. These poster sessions are often the highlight of the students’ semester, and for many the beginning of a deep passion for statistics (see Chapter 4 on Encouraging Reflective Learning in Introductory Statistics).

Figure 2. Students Presenting Completed Projects

Resources Needed
You need very minimal resources to begin infusing passion with our Passion-Drive Statistics approach into your statistics course. Below we highlight the two key resources needed to get started, both of which can be obtained by you and your students for free (either on your own or with our help; see Get Involved! section).

1. A large, accessible data set with a corresponding codebook or data dictionary. For this example (and because it is a popular choice of our students), we will be using the publicly accessible version of The National Longitudinal Study of Adolescent to Adult Health Wave I, a representative U.S. sample of adolescents in grades 7 through 12 (https://www.cpc.unc.edu/projects/addhealth/design/wave1).

2. Access to statistical software (e.g., SAS, R, SPSS, Python, Stata, or other). For institutions with limited resources and/or limited IT support, as well as those wanting students to be able to work from any computer on or off campus, we recommend SAS OnDemand for Academics: Studio (https://www.sas.com/en_us/software/on-demand-for-academics.html), a free, non-case sensitive, cloud based platform requiring only internet access (but see also Chapter 18 on Yes, Beginning Statistics Students Can Use R).

Time Requirements
Given that our curriculum is “data in the service of research questions”, projects can be designed to require as much or little time as desired. Based on our experiences, we do however recommend a minimum of 3 weeks of a regular semester schedule or a 1-week intensive schedule. In our own courses, we employ a full, 15-week semester of Passion-Driven Statistics. Figure 3 and 4 show two variations of an approximate weekly schedule that we have used, including a mapping of topics to the final course project. Please note that for instructors wishing to implement the full, 15-week semester using the Passion-Driven Statistics curriculum, we are happy to provide our own course materials for support!
Introduction

- Psychiatric disorders are a potent group of risk factors consistently implicated in the development of nicotine dependence (Rodriguez et al., 2003; 2004).
- While the association has been well established in the literature, less is known about the ways in which psychiatric disorders may play a role in the emergence of nicotine dependence.
- Most research on psychiatric disorders and smoking has focused on heavy smoking e.g. the self-medication hypothesis (Khantzian, 1997).
- Alternatively, however, psychiatric disorders may signal a greater sensitivity to nicotine dependence at low levels of smoking (i.e. individuals with psychiatric disorders may develop nicotine dependence symptoms at lower levels of smoking than those without psychiatric disorders).

Research Questions

- Which psychiatric disorders are associated with nicotine dependence? Which for comorbidity?
- Does the association between smoking quantity and nicotine dependence differ for smokers with and without these psychiatric disorders?

Methods

Sample

- Young adults (age 18 to 25) who reported daily smoking in the past year (n=1320) were drawn from the first wave of the National Epidemiologic Study of Alcohol and Related Conditions (NESARC).
- NESARC is a nationally representative sample of non-institutionalized adults in the U.S.

Measures

- Lifetime psychiatric diagnoses assessed using the NIAAA, Alcohol-Related Disorders Interview Schedule – DSM-IV (AUDADIS-IV).
- The tobacco module includes questions on symptom criteria for DSM-IV nicotine dependence.
- Current smoking was evaluated through quantity ("On the days that you smoked in the last year, about how many cigarettes did you usually smoke?").

Results

Univariate

- Fully 61% of daily smokers met criteria for DSM-IV nicotine dependence.
- A total of 55% met criteria for any psychiatric disorders.

Bivariate

- Chi-Square analysis showed that daily, young adult smokers with a psychiatric disorder were significantly more likely to meet criteria for nicotine dependence (OR = 6.4, 95% CI: 1.5 to 26.2) than those without a psychiatric disorder (OR = 1.01, 95% CI: 0.99 to 1.02).
- The most common diagnostic category was alcohol dependence (69%, 95% CI: 0.5 to 1.8).
- As expected, the number of cigarettes smoked per day was significantly associated with DSM-IV nicotine dependence, OR=1.04 (1.03-1.05).

Multivariate

- Major depression (MDD), specific phobia, alcohol dependence, and antisocial personality disorder (ASPD) were each associated with DSM-IV nicotine dependence after controlling for comorbidity.
- The interaction between number of cigarettes smoked per day and the presence of nicotine dependence was statistically significant (p<0.05).
- At each level of use, the probability of nicotine dependence was significantly higher among those with the disorder than those without (Figure 1).
- In contrast, the interaction between alcohol dependence and number of cigarettes smoked per day was statistically significant when predicting nicotine dependence (Figure 2).

Discussion

- Individuals with major depression, specific phobia, or ASPD may be more sensitive to nicotine dependence across levels of smoking.
- Individuals with alcohol dependence are more sensitive to nicotine dependence at low levels of daily smoking, but not at the highest levels when compared to those without a psychiatric disorder.
- Notably, the present findings are based on cross-sectional data and do not reflect the smoking levels at which nicotine dependence emerges among those with and without psychiatric disorders.
- Further research is needed to determine whether sensitivity to nicotine dependence is based on physical and/or psychological differences related to psychiatric disorders.

References

- 5.

3. Figure 3. Approximate Weekly Schedule (Variation #1)
In the United States, a survey study found that the name "religion" was used by adults to refer to the activities that they believed would improve or maintain their spiritual well-being. However, many adults were found to be less likely to endorse a strong agreement with the use of birth control, increasing their vulnerability to STDs (Johnson et al. 2017).

Current studies suggest that the relationship between religiosity and sexually transmitted infections is complex and moderated by factors such as age, gender, and education. Because an individual's opinion on sexual activity is often affected by religious affiliation and identity, additional research is warranted.

Here is the basic order of steps involved in the inquiry process. In our courses, students follow these same steps though we present them in greater detail and in combination with their learning of the conceptual understanding of the various statistics themes of our courses. Many of these steps also include additional assignments, for example a short literature review or annotated bibliography around step 2, and/or structured lab assignments.

1. Read through the codebook or data dictionary, focusing on topics or constructs that you find interesting, relevant, or exciting.
2. Choose a research question based on the available data. That is, choose a specific topic or construct of interest and then a second that you believe relates to the first. Prepare a personal codebook of your own by printing the pages that include only the variables that are important to your research.

Here are the steps involved in the project:

1. **Read through the codebook or data dictionary**, focusing on topics or constructs that you find interesting, relevant, or exciting.
2. **Choose a research question based on the available data.** That is, choose a specific topic or construct of interest and then a second that you believe relates to the first.
3. **Prepare a personal codebook of your own by printing the pages that include only the variables that are important to your research.**

Now, let's dive into Passion-Driven Statistics.
you have selected. Your codebook should contain two variables, minimum, however, it may contain more if you find a number that relate to a similar construct.

I am interested in gender (topic 1) and whether or not it is related to being older relative to one's grade level (topic 2). I have a hypothesis that males are more likely to start kindergarten later than females (something known as “academic red shirting”). I print the codebook pages that include the variables for gender, age, and current grade level.

3. Use your selected statistical software program to examine frequency tables for the variables you have chosen.

4. Conduct data management on your chosen variables. For example, you may need to set aside missing data, create a new variable by collapsing response categories, create a new variable by aggregating across more than one variable, and/or label your variables. The extent of this step will be dictated by your particular variables, research question(s), and statistical software program.

I will want to set response categories of “refused to answer” or “don’t know” to missing so that these are not included in my analyses. Given my interest in students being older relative to their grade level, I will also need to create a new variable based on both age and grade level that indicates whether or not a participant can be classified as “older for grade” (yes or no).

5. Describe frequency tables, as well as center, spread, and variability for each of your newly data managed variables (one variable at a time).

Figure 5. Graphing Decision Flowchart for Step 6
6. Using the graphing decision flowchart (above) to help you determine what type of graph is most appropriate for your data, graph your first variable by your second variable (an association between two variables). Determine if there seems to be a relationship between the two through a visual examination of the graph.

Are participants who are older for their grade more likely to be male or female?

I look at my bivariate bar graph and examine whether it seems like one gender (males or females) are more likely to be older for their grade.

7. Based on your reading of the data codebook, choose a third topic that you believe may be related to the first two topics. Add the variable(s) that measure this topic to your personal codebook.

Though I am interested in whether males or females (topic 1) are more likely to be older for their grade level (topic 2), I wonder whether being older for your grade is also related to feeling socially accepted (topic 3) and whether the answer to that question is different for males and females.

8. Conduct additional data management for the newly selected variable(s) [see step 4].
9. Graph the third variable by each of your first two topics [see step 6].

Are males or females more or less likely to feel socially accepted?

Do those who are older for their grade rate their feelings of social acceptance higher or lower than those who are not older for their grade?

10. Select a relationship of interest (two variables) and graph that relationship by different levels of a third variable.

Is there a relationship between being older for your grade and feeling socially accepted? Does that relationship look different for males versus females?

---

**Figure 6. Table for Helping Students Select the Correct Bivariate Test**

11. To this point, the analyses and graphical representations have been purely descriptive in nature. With these descriptive findings in mind, we transition to the topic of inferential statistics, laying the conceptual groundwork (course video:}
http://bit.ly/Inferential_Statistics), and introducing three statistical tests that will allow students to apply hypothesis testing to their research question(s). These statistical tests include Analysis of Variance (for questions including one categorical and one quantitative variable); Pearson Correlation (for questions including two quantitative variables); and Chi-Square Test of Independence (for questions including two categorical variables).

12. Bivariate relationships can then be examined based on subgroups (i.e., categorical third variables) by subsetting the sample and examining the bivariate relationships for each subgroup (i.e., testing for moderation). Thus, based on graphs that the students have generated at earlier steps in the project, we teach them how to evaluate observed differences in terms of whether they meet statistical significance and can be inferred to the larger population.

13. Finally, multivariate models (e.g., multiple regression and logistic regression) can also be introduced through the concept of evaluating relationships for potential confounders.

Assessment of the Curriculum

We are happy to report that the success of the Passion-Driven Statistics curriculum has been documented both in peer-reviewed publications and by instructors of individual courses. For instance, published research on Passion-Driven Statistics has shown that our project-based course attracts higher rates of under-represented minority (URM) students compared to a traditional math statistics course (Dierker, Cooper, Selya, Alexander, & Rose, 2015) and higher rates of female and URM students compared to an introductory programming course (Cooper & Dierker, 2017). Further, URM students are twice as likely as non-URM students to report increases in their interest in conducting research after completing the project-based course (Dierker et al., 2016).

When compared to a traditional statistics course, Passion-Driven Statistics students report finding the course more rewarding and more useful than other college courses. These students also report receiving more individualized support than other college courses and are most likely to report that they accomplish more than expected in the course. An especially exciting finding is that students enrolled in the Passion-Driven Statistics course are also more likely to report increased confidence in working with data and increased interest in pursuing advanced statistics course work compared to those from a traditional statistics course (Dierker et al., 2018a).

Our assessments have included students from liberal arts colleges, large state universities, regional college/universities, community colleges, and high schools where the passion-driven approach has been successfully implemented. When looking at students across these educational settings, 84% of them report interest in one or more follow-up courses, with an interest in statistical programming being endorsed by the largest number of students (53%) (Dierker et al., 2018b). For those of us who have taught statistics at any level, you recognize the magnitude of these finding. At the end of a statistics course, students are generally only relieved to have “checked the box” to show they have fulfilled their statistics requirements; our students are reporting a desire for more statistics!

Qualitative comments about the curriculum have also provided evidence that students perceive this model as effective in stimulating and supporting their continued interests and curiosity in statistics. As one student put it, “I absolutely loved this experience. Even though data analysis is not necessarily exciting to learn, it is incredibly useful and the uniqueness of the project that I was doing kept me interested. I am not sure I have ever taken such a useful class, nor been so proud of my work.” Other students have said, “This course taught me how to take initiative and start a scientific project that I can call my own” and “I have never felt so excited and motivated to be part of an academic environment as I have in this class.”
Indeed, students find Passion-Driven Statistics to be a rewarding experience. Based on more than 1,500 surveys completed anonymously following delivery of the passion-driven approach at 11 colleges and universities (Dierker, unpublished data), nearly 74% of students rated their work as rewarding or very rewarding. Further, more than a third of students exposed to the passion-driven curriculum rated the experience as more useful than other courses they have taken. More than half (60%) of students believed they were likely or very likely to use their data analysis skills in the future and 80% reported that they would recommend the course to others. We believe this course is helping to “change the face” of statistics at our institutions through positive student experiences such as those highlighted here.

Instructor Experiences

As instructors of statistics, we certainly recognize that teaching courses in introductory statistics at both the undergraduate and graduate level can be challenging. We know that there is no typical statistics student; all students come into our statistics courses with differing backgrounds, experiences, interests, and levels of preparation. The passion-driven approach to teaching introductory statistics allows us to reach our different students by putting the focus on their interests and their passions. Once students feel that they have a stake in the game, we cultivate their natural curiosity with real world data and a meaningful project-based classroom experience.

“My students and I derive incredible benefits from this project. When I ask students to reflect on the course, they often mention that their favorite part of the class was the Passion-Driven Statistics piece. They truly enjoy getting their hands dirty with real data.” — Laura Estersohn, Scarsdale High School, Scarsdale, NY

Indeed, Passion-Driven Statistics is about making statistical real for students. Our students learn to use statistics to understand real world problems and conduct research that actually means something (versus canned labs or exercises). Students are excited to learn that they are creating new knowledge, perhaps even asking questions with data that have never been asked before. Thus, when you teach using the passion-driven approach, you have the opportunity to engage students in real research in your classroom every semester. We love teaching statistics with this model because it transforms the classroom into a place where students are the stars of their own shows. It is not about what an instructor brings to the classroom. As instructors, we are there to guide students on a journey simply by providing them with the statistical skills they need to conduct and understand applied statistics. The rest is up to the students. It can be a magical experience for both students and instructors!

“I learned statistics by doing statistics and I had a 12-year career as a research scientist – so I was really excited to use this approach. It is one of my favorite classes to teach.” — Dr. Jennifer Willford, Department of Psychology, Slippery Rock University, Slippery Rock, PA

What we tell our students (and what we do) with this approach is that we take our students outside of their comfort zones and then support them through the fallout. Statistics in the real world can be messy and unpredictable. In Passion-Driven Statistics, our students are faced with decisions about their own data that, to them, may seem messy and unpredictable. As instructors, we use these moments as teachable experiences for students. We empower them by supporting their decisions and letting them fail, but never letting them become failures. Our experiences have taught us that you do not necessarily need to water-down statistics at the introductory level, nor do you need to get rid of the difficult parts. Instead, we need to make
those parts much more supportive to students. It is through this mindset that our students able to ask (and answer!) questions about their data that involve multiple layers. Can introductory statistics students conduct AND understand multivariate analyses? YES THEY CAN!

We should point out that the goal of Passion-Driven Statistics is to engage students by empowering them to ask and answer questions that matter to them using statistics. We want to drive interest and understanding in learning statistics that is organic, not forced or insincere. We accomplish this through our project-based approach that utilizes a flipped classroom. However, this is of course not the only way to accomplish this goal. We encourage instructors to adapt and modify our approach to a teaching model that best suits their needs, style, and goals. No two instructors who use the passion-driven statistics approach do it in the exact same manner. Some instructors have fully flipped their courses and rely heavily on our course videos outside of the classroom, while utilizing class time solely for the purposes of working students’ research projects. Others maintain a more traditional approach, lecturing in class and having students complete the research project during structured lab times. Another approach is to infuse small bits of the research project throughout the semester but focus heavily on it near the end of course as a culminating project. And, there are lots of other ways in between all of that. What matters most, it seems, is the desire and willingness of the instructor to engage their students in a passion-driven pursuit of statistical knowledge and understanding.

Get Involved!

Our goal in writing this chapter was to familiarize you with the passion-driven approach of teaching introductory statistics. If you would like to get involved in Passion-Driven Statistics, or simply desire to learn more about our project-based curriculum, we would welcome the opportunity to speak with you.

We are also happy to share our free, open-source materials, many of which are available on the Passion-Driven Statistics webpage (http://passiondrivenstatistics.com). Additional course materials from our project partners (e.g., sample syllabi and course schedules, model assignment instructions, public access data sets and codebooks, and statistical software resources) are available to instructors wishing to adopt the Passion-Driven Statistics curriculum from our Schoology account (https://www.schoology.com/).

You will need to set up a free account with Schoology and request an access code from Lisa Dierker (ldierker@wesleyan.edu). You may also contact Lisa Dierker to be included in future announcements for faculty workshops and opportunities to view the curriculum “in action” by visiting a poster presentation or class session. We welcome the involvement of new instructors who want to impart a passion for statistics to their students!

Author Note. This material is based upon work supported by the National Science Foundation (NSF) under Grant No. 1820766. Any opinions, findings, interpretations, conclusions or recommendations expressed in this material are those of its authors and do not represent the views of the National Science Foundation.
References


Appendix A

Suggested Grading Rubric for Student Posters

Reviewer (your name): ____________________________

Poster Title: _____________________________________________________

Presenter: __________________________________

Instructions to reviewer: Use these criteria to rate the poster presentation on a scale of 1-5. *(1=strongly disagree; 2=disagree; 3=neutral; 4=agree; 5=strongly agree)*

**Poster Visuals**

1. Overall, the poster was clear and easy to follow.  
   1 2 3 4 5
2. Table(s) and/or graphics on the poster enhanced the presentation.  
   1 2 3 4 5

**Presentation Content**

3. Background research (literature) assisted me in understanding the topic.  
   1 2 3 4 5
4. Research questions were stated in plain language.  
   1 2 3 4 5
5. There was enough detail about methods for me to understand the results.  
   1 2 3 4 5
6. Statistical results were explained in a clear and organized manner.  
   1 2 3 4 5
7. Results were interpreted in plain language, including the real world implications of the findings and important next steps for the research.  
   1 2 3 4 5

**Question and Answer**

8. Presenter’s response to questions demonstrated knowledge of subject matter.  
   1 2 3 4 5
9. Presenter’s response to questions demonstrated knowledge of the statistical approach taken with the project.  
   1 2 3 4 5
10. The presenter was excited about/engaged in his or her project.  
    1 2 3 4 5

**Other Comments** *(use back of page for comments):*
Appendix B
Handout Detailing Initial Steps in the Inquiry Process Using SAS

Passion-Driven Statistics

STEP 1: Reading the codebook

Read through the Adolescent Health Study (Wave 4), Behavioral Risk Factor Surveillance System (2016), or National Household Survey of Drug Use and Health (2016) codebook. Choose one based on description in the Schoology Codebook folder.

STEP 2: Choosing a research question

Based on your reading of the codebook, choose a specific topic of interest and then a second topic you believe is related to the first topic. Prepare a short codebook of your own.

For example, I am interested in gender (topic 1)

<table>
<thead>
<tr>
<th>1. Respondent’s biological gender</th>
<th>BIO_SEX</th>
<th>num 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>3147</td>
<td>1</td>
<td>male</td>
</tr>
<tr>
<td>3356</td>
<td>2</td>
<td>female</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>refused</td>
</tr>
</tbody>
</table>

and how it is related to being older relative to one’s grade level (topic 2).

<table>
<thead>
<tr>
<th>4. Current age of respondent</th>
<th>AGE1</th>
<th>num 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>4837</td>
<td>11.68 to 20.93 (calculated from date of birth)</td>
<td></td>
</tr>
<tr>
<td>1667</td>
<td>.</td>
<td>missing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>20. What grade {ARE/WERE} you in? (\text{If school doesn’t have grade levels of this kind, enter “99.”})</th>
<th>H1G120</th>
<th>num 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>979</td>
<td>1 (seventh grade)</td>
<td></td>
</tr>
<tr>
<td>992</td>
<td>8 (eighth grade)</td>
<td></td>
</tr>
<tr>
<td>1107</td>
<td>9 (ninth grade)</td>
<td></td>
</tr>
<tr>
<td>1144</td>
<td>10 (tenth grade)</td>
<td></td>
</tr>
<tr>
<td>1122</td>
<td>11 (eleventh grade)</td>
<td></td>
</tr>
<tr>
<td>993</td>
<td>12 (twelfth grade)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>96 (refused)</td>
<td></td>
</tr>
<tr>
<td>128</td>
<td>97 (legitimate skip (not in school?))</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>98 (don’t know)</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>99 (school doesn’t have grade levels of this kind or not applicable)</td>
<td></td>
</tr>
</tbody>
</table>
**STEP 3: Examining Data: Writing your first program**

Write a basic program. Add **variable names** that you have chosen to the **request for results** statements *(yellow highlight)*

**Example program**

```plaintext
libname mydata "/courses/d59155e5ba27fe300"
access=readonly;

data new; set mydata.addhealth_pds;

/*data management goes between "data" and "proc sort or proc freq" statements*/

proc sort; by aid;

/*statements for output/results are included after "proc sort"*/

proc freq; tables bio_sex age1 h1gi20;

run;
```
**STEP 4: Managing Data**

Add **data management statements** to your program. This will include **setting aside missing data** for each of our chosen variables and one or more of the following a) creating a new variable by collapsing response categories; b) creating a new variable by aggregating across more than one variable; and c) labeling your variables.

``` SAS
libname mydata "/courses/d59155e5ba27fe300"
access=readonly;

data new; set mydata.addhealth_pds;
/* Setting aside missing data*/
if BIO SEX=6 then BIO SEX=.;
if H1GI20 GE 96 then H1GI20=.;

/* Creating a new variable by collapsing response categories*/
if AGE1 EQ . then AGEGROUP=.;
else if AGE1 LE 16.51 then AGEGROUP=1;
else if AGE1 GT 16.51 then AGEGROUP=2;

/* Creating a new variable by aggregating more than one variable*/
if H1GI20 EQ . or AGE1 eq . then OLDER4GRADE=.;
else if (H1GI20=7 and AGE1 GE 14) or (H1GI20=8 and AGE1 GE 15) or (H1GI20=9 and AGE1 GE 16) or (H1GI20=10 and AGE1 GE 17) or (H1GI20=11 and AGE1 GE 18) or (H1GI20=12 and AGE1 GE 19) then OLDER4GRADE=1;
else OLDER4GRADE=0;

/* Labeling variables*/
label AGEGROUP='age group'
    BIO SEX='gender'
    H1GI20='grade level'
    OLDER4GRADE='older for grade';

proc sort; by AID;

proc freq; tables AGE1 BIO SEX H1GI20 AGEGROUP OLDER4GRADE;
proc means; var AGE1 H1GI20;
run;
```
STEP 5: Describing one variable at a time (Univariate)

Describe univariate (i.e. one variable) frequency tables for each of your newly data managed variables.

For each of your categorical variables, you will be examining the number or percent of observations in each category.

```
Proc freq; tables BIO_SEX OLDER4GRADE;
```

<table>
<thead>
<tr>
<th>BIO_SEX</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3147</td>
<td>48.39</td>
<td>3147</td>
<td>48.39</td>
</tr>
<tr>
<td>2</td>
<td>3355</td>
<td>51.81</td>
<td>6503</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Frequency Missing = 1

51.6% of the sample was female (48.4% male).

<table>
<thead>
<tr>
<th>OLDER4GRADE</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4025</td>
<td>84.99</td>
<td>4025</td>
<td>84.99</td>
</tr>
<tr>
<td>1</td>
<td>711</td>
<td>15.01</td>
<td>4736</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Frequency Missing = 1769

711 (15.0%) of the sample were older for their grade.
STEP 6: Examining one variable by another variable (an association)

Display and describe the association between your two topics (i.e. two data managed variables)

**Proc sgplot; vbar BIO_SEX /response= OLDER4GRADE stat=mean;**
**TITLE 'The Association between Sex and Older for Grade';**

This code requests a graph for OLDER4AGE and splits the sample by BIO_SEX (i.e. gender) creating two bars, one for males and one for females.

Males are more likely to be older for their grade then females. It appears about 18.5% of males are older for their grade compared to only 11.5% of females.
STEP 7: Choose a third topic

Based on your reading of the codebook, choose a **third topic** that may be related to the first two topics. Add the variable to your short codebook.

Though I am interested in whether males or females (topic 1) are more likely to be older for their grade (topic 2), I also wonder whether being older for your grade is related to **feeling socially accepted** (topic 3) and whether the answer to that question is different for males and females.

<table>
<thead>
<tr>
<th>35. You feel socially accepted.</th>
<th>HIPF35</th>
<th>num 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1857</td>
<td>1</td>
<td>strongly agree</td>
</tr>
<tr>
<td>3674</td>
<td>2</td>
<td>agree</td>
</tr>
<tr>
<td>667</td>
<td>3</td>
<td>neither agree nor disagree</td>
</tr>
<tr>
<td>241</td>
<td>4</td>
<td>disagree</td>
</tr>
<tr>
<td>40</td>
<td>5</td>
<td>strongly disagree</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>refused</td>
</tr>
<tr>
<td>15</td>
<td>8</td>
<td>don't know</td>
</tr>
</tbody>
</table>
**STEP 8: Manage data for your third topic**

Add **data management** to your program. This will include **setting aside missing data** for each of our chosen variables and one or more of the following a) creating a new variable by collapsing response categories; b) creating a new variable by aggregating across more than one variable; and c) labeling your variables. After the PROC SORT statement, **ask for a frequency table for your new variable**.

**New code to add to data management:**

```
/*setting aside missing data*/
if H1PF35 GE 6 then H1PF35=.;

/*third topic - feeling socially accepted*/
if H1PF35 =. then FEELSOCACCEPT=.;
else if h1PF35 le 2 then FEELSOCACCEPT=1;
else if h1PF35 le 5 then FEELSOCACCEPT=0;

proc sort; by AID;
proc freq; tables FEELSOCACCEPT;
run;
```
STEP 9: Examine the third variable by each of your first two topics

Display and describe the associations among your three topics.

1. Are males or females more likely to feel socially accepted?

```plaintext
Proc sgplot; vbar BIO_SEX /response=FEELSOCACCEPT stat=mean;
TITLE 'The Association between Sex and Feeling Socially Accepted';
```

Males are only slightly more likely to feel socially accepted than females.
2. Are those who are older for their grade compared to those who are not older for their grade more likely to feel socially accepted?

Proc sgplot; vbar OLDER4GRADE /response= FEELSOCACCEPT stat=mean; TITLE 'The Association between Older for Grade and Feeling Socially Accepted';

Adolescents who are older for their grade are only slightly less likely to feel socially accepted than adolescents who are not older for their grade.
STEP 10: Graph the relationship of interest by third variable

Display and describe the association of interest (two variables) and graph that relationship by different levels of a third variable.

Is the relationship between being older for your grade and feeling socially accepted different for males and females?

```
Proc sgplot; vbar BIO_SEX/*third var.*/ response=FEELSOCACCEPT/*response var.*/ group=OLDER4GRADE/*explanatory var.*/ groupdisplay=cluster stat=mean; TITLE 'The Association between Older for Grade and Feeling Socially Accepted across Sex';
```

Whether or not **males** are older for their grade they are similarly likely to feel socially accepted. **Females** who are older for their grade are less likely to feel socially accepted than females who are not older for their grade.
STEP 11: Analyze the bivariate relationship of interest

Examine the **bivariate relationship** using the appropriate inferential statistical analysis to test your hypothesis for your research question. Choose the correct test based on your variables. If your question includes one categorical and one quantitative variable use *Analysis of Variance*, for questions including two quantitative variables use *Pearson Correlation*, and for questions including two categorical variables use *Chi-Square Test of Independence*.

/*Chi-Square Test of Independence*/
**Proc freq; tables FEELSOCACCEPT*OLDER4GRADE/chisq;**

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Percent</th>
<th>Row Pct</th>
<th>Col Pct</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEELSOCACCEPT</td>
<td>OLDER4GRADE(older for grade)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>FEELSOCACCEPT</td>
<td>0</td>
<td>559</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11.84</td>
<td>2.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>81.73</td>
<td>18.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13.91</td>
<td>17.76</td>
</tr>
<tr>
<td>FEELSOCACCEPT</td>
<td>1</td>
<td>3459</td>
<td>579</td>
</tr>
<tr>
<td></td>
<td></td>
<td>73.25</td>
<td>12.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>85.66</td>
<td>14.34</td>
</tr>
<tr>
<td>Total</td>
<td>4018</td>
<td>704</td>
<td>4722</td>
</tr>
<tr>
<td></td>
<td>85.09</td>
<td>14.91</td>
<td>100.00</td>
</tr>
</tbody>
</table>

**Statistics for Table of FEELSOCACCEPT by OLDER4GRADE**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>1</td>
<td>7.1432</td>
<td>0.0075</td>
</tr>
</tbody>
</table>

Adolescents who are **older for their grade** (82.2%) are only slightly less likely to feel socially accepted than adolescents who are **not older for their grade** (86.1%).

\[ X^2 = 7.14 \text{ 1 df, } p = .0075 \]
STEP 12: Examine the bivariate relationship based on subgroups

Examine the bivariate relationship based on subgroups (i.e., categorical third variable) by subsetting the sample and examining the bivariate relationship for each subgroup. That is, determine if the third variable moderates the bivariate relationship.

/*Test for Moderation: Chi-Square Test of Independence*/
Proc sort; by BIO_SEX; /*third var.*/
Proc freq; tables FEELSOCACCEPT /*response var.*/
*OLDER4GRADE /*explanatory var.*/ /chisq;
where BIO_SEX ne.; by BIO_SEX;

Whether or not males are older for their grade they are similarly likely to feel socially accepted (86.3% vs. 88.0%).

$X^2 = 0.991 \text{ df, p = .3192}$
Females who are older for their grade are less likely to feel socially accepted (76.2%) than females who are not older for their grade (84.5%).

\[ X^2 = 12.20 \text{ df, } p = .0005 \]
STEP 13: Examine the bivariate relationship for potential confounders

To test a third variable as a potential confounder of a bivariate relationship use multiple regression (i.e., quantitative response variable) or logistic regression (i.e., categorical response variable. Recode variables to 1=Yes and 2=No prior to analyzing the data.

/*Logistic Regression: test association b/ predictor & response*/

**Proc logistic;**
class OLDER4GRADE;
model FEELSOCACCEP=OLDER4GRADE;

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>1.6777</td>
<td>0.0543</td>
<td>953.8035</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>OLDER4GRADE</td>
<td>1</td>
<td>-0.1449</td>
<td>0.0543</td>
<td>7.1120</td>
<td>0.0077</td>
</tr>
</tbody>
</table>

**Odds Ratio Estimates**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLDER4GRADE 1 vs 2</td>
<td>0.748</td>
<td>0.605 – 0.926</td>
</tr>
</tbody>
</table>

/*Logistic Regression: test third variable for confound of association b/ predictor & response*/

**Proc logistic;**
class OLDER4GRADE BIO_SEX;
model FEELSOCACCEP=OLDER4GRADE BIO_SEX;

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>1.6781</td>
<td>0.0545</td>
<td>947.4711</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>OLDER4GRADE</td>
<td>1</td>
<td>-0.1716</td>
<td>0.0548</td>
<td>9.8011</td>
<td>0.0017</td>
</tr>
<tr>
<td>BIO_SEX</td>
<td>1</td>
<td>0.1845</td>
<td>0.0423</td>
<td>19.0562</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

**Odds Ratio Estimates**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLDER4GRADE 1 vs 2</td>
<td>0.710</td>
<td>0.572 – 0.880</td>
</tr>
<tr>
<td>BIO_SEX 1 vs 2</td>
<td>1.447</td>
<td>1.226 – 1.707</td>
</tr>
</tbody>
</table>

Confirm the bivariate relationship is significant.

After adjusting for the potential confounding factor of sex, older for grade (OR 0.710, CI 0.572 – 0.880, p = 0.0017) was significantly and negatively associated with the likelihood of feeling socially accepted. Those who are older for their grade are 0.710 times less likely to feel socially accepted.
Additional SAS Code:

Bold font indicates code or other text that should be typed literally. Un-bolded should be replaced.

Examine response categories for categorical or quantitative variables

PROC FREQ; tables Var1 Var2 Var3;

Examine means and standard deviation for quantitative variables

PROC MEANS; var QuantVar1 QuantVar2 QuantVar3;

Graph one variable at a time (univariate)

PROC GCHART; VBAR CategVar / Discrete type=PCT Width=30;
PROC GCHART; VBAR QuantVar;

Graph association between two variables (both categorical)

PROC SGPLOT; vbar CategExplanatoryVar /response=CategResponseVar stat=mean;

Test association between two variables (both categorical)

PROC FREQ; tables CategResponseVar*CategExplanatoryVar /chisq;

Graph association between two variables (one categorical and one quantitative)

PROC SGPLOT; vbar CategExplanatoryVar /response=CategResponseVar stat=mean;

Test association between two variables (one categorical and one quantitative)

proc anova;
  class CategExplanatoryVar;
  model QuantResponseVar = CategExplanatoryVar;
  means CategExplanatoryVar;

Graph association between two variables (both quantitative)

Proc SGPLOT; scatter x=QuantExplanatoryVar y=QuantResponseVar;

Test association between two variables (both quantitative)

Proc corr; var QuantResponseVar QuantExplanatoryVar;

Add a third variable (moderator)

proc sort; by CategThirdVar; /*bivariate graphing or test code above*/ by CategThirdVar;
Teaching Statistics for Social Justice: A “Recipe” for Project Implementation

Charles R. Collins, PhD
University of Washington - Bothell

Summary

This chapter provides a case example of a community-based partnership undertaken in an undergraduate statistics course within the school of interdisciplinary arts and sciences at the University of Washington Bothell. In the spring quarter of 2015, my Understanding Statistics course partnered with Real Change News, an award-winning newspaper and homeless advocacy organization that gives low-barrier employment opportunities to under- and unemployed residents by selling their newspapers. The purpose of this project was to add to the existing capacity of Real Change by providing research and statistical expertise. Additionally, I sought to foster a social justice educational experience by adopting a project focused on issues of equity with an advocacy organization. The goal of the project was to develop a survey and collect data about the readers of Real Change’s newspaper. This chapter provides seven steps for creating and implementing a community-based project into an undergraduate statistics course. More specifically, I outline the weekly steps of project implementation, which included partnership development, survey development, data collection design, student orientation, data collection, course content, and data analysis and project presentation. I also present potential impacts of the project for students, Real Change, and the university. I conclude the chapter by suggesting future considerations for educators who may consider adopting a similar approach to their own statistics courses. These include: creating a memorandum of understanding, developing a timeline and clearly stating the project, considering the university system, and considering the class structure.

Introduction

Social justice is both a process and a goal in which social actors engage in democratic processes to strive for the goal of fair and equitable distribution of resources (Bell, 2016). To seek social justice, it is necessary to understand manifestations of power and oppression. As such, social justice education seeks to foster capacities of students (and others) to understand the systemic components of oppression; develop the knowledge and skills to critically examine issues of justice and participants’ role within those systems; and to build the skills, capacities, and agency for actors to resist and dismantle systems of oppression (Adams, 2016; Freire, 1970, 1973). Bell (2016) outlines a set of practices that can be utilized to guide social justice education, which include: raising awareness of the ways oppression manifests socially and politically; taking critical actions to resist unjust systems (Prilleltensky, 2003, 2008; Prilleltensky & Fox, 2007; Watts & Abdul-Adil, 1998; Watts, Diemer, & Voight, 2011; Watts, Griffith, & Abdul-Adil, 1999; Watts, Williams, & Jagers, 2003); being an accountable and responsible ally with those seeking mutual liberation (Kivel, 2017); and building connections and solidarity because, “eradicating oppression ultimately requires struggle against all forms, and that coalitions among diverse people who can offer perspectives from their particular social locations provide the most promising potential for creating change” (Bell, 2016, p. 16).
Although the majority of social justice education is not aimed at statistics courses (or other mathematics-based courses) a growing body of literature is highlighting teaching for social justice in statistics and mathematics (Lesser, 2007). Gutstein (2003, p. 45) offers some insight into teaching mathematics, and by extension statistics, for social justice: “reading the world with mathematics means to use mathematics to understand relations of power, resource inequities, and disparate opportunities between different social groups and to understand explicit discrimination based on race, class, gender, language, and other differences.” This body of literature offers guidelines on the ways that educators can integrate social justice pedagogy into statistics education. These include, for example, integrating conceptual examples and/or datasets into courses related to justice topics (e.g. discrimination, poverty, etc.), integrating an equity focus within the classroom and particularly among students from historically marginalized communities (Gutstein, 2003), enhancing culturally relevant pedagogy (Leonard, Brooks, Barnes-Johnson, & Berry III, 2010), or conducting research on a particular topic such as race or other student driven issues (see chapter 5 on Passion Driven Statistics; Martin, 2009). Related, community-based learning (or service-learning) has also been considered a powerful pedagogical tool to advance issues of social justice (Cipolle, 2010; Einfeld & Collins, 2008) and has been pressed for within statistics and mathematics education (Duke, 1999).

The purpose of this chapter is to provide a case example of an undergraduate social science statistics course that adopted an expressed goal of promoting social justice by combining real world examples in partnership with a local social justice organization. In partnership with a local homeless advocacy organization, Real Change, my BIS 315: Understanding Statistics course in the spring quarter of 2015 sought to utilize data and statistical analyses as a tool to promote social justice and build the power of Real Change to affect issues of homelessness. Below, I provide a brief synopsis of homelessness in Seattle, WA – Real Change’s home-base, describe the Real Change Project including planning and implementation, outline potential impacts, and provide lessons learned.

Homelessness in Seattle, WA

Seattle is a city located in the Northwestern United States (U.S.) with a population of a little more than 700,000 (~3,000,000 in the greater metro area) and has grown by about 100,000 residents over the past ten years (U.S. Census, 2010; 2017). Recently, an economic explosion in the technology industries (e.g. aerospace engineering, computer technology, biotechnology, etc.) has created high paying jobs for the well-educated. Since 2010 (to 2017) the number of residents over 25 years of age with bachelor's degrees grew by 31% and median household income grew by 30%. This growth has accompanied steep cost of living increases. For example, median home values increased by 18% and median rent by 44% (U.S. Census, 2010; 2017). Unfortunately, this economic boom has not been evenly distributed throughout the city. During this time, Seattle has experienced intense growth in homelessness. In fact, the “one-night-count” tallied 8,978 (2,800 unsheltered) homeless individuals in 2010 and 12,112 homeless residents with 6,320 unsheltered in 2018 – a near 35% increase in homelessness (Coleman, 2018). Fortunately, many local organizations are working diligently to address this issue. Real Change is one such organization.

The Real Change Project

Real Change News (or simply Real Change; www.realchangenews.org) is a non-profit news organization in Seattle that “produces an award-winning weekly newspaper that provides immediate employment opportunity and takes action for economic, social, and racial justice” (“About Real Change”, n.d.). Their goal is to empower un- and under-employed citizens by providing low-barrier job opportunities through selling Real Change newspapers. These
individuals are *Real Change* newspaper “vendors”. Vendors purchase papers for $0.60, sell them for $2.00, and keep the difference. *Real Change* vendors often provide a face and story to the issue of homelessness through one-on-one interactions with readers. *Real Change* also strives to enhance local awareness and social action around the topic of homelessness by covering news on local economic and social issues, printing stories of their vendors, and giving greater representation to some of the most vulnerable residents of Seattle. Additionally, Real Change’s “OutsideIN” initiative successfully helped earmark $1.2M for a new shelter in Seattle through a “campaign to pressure policymakers to reduce the numbers of unsheltered people in King County” (“Real Change” n.d.).

In the spring quarter of 2015, I partnered with *Real Change* to integrate social justice pedagogy into my BIS 315: Understanding Statistics course – an upper division undergraduate social science statistics course. BIS 315: Understanding Statistics is a core course for many majors in the School of Interdisciplinary Arts and Sciences at the University of Washington Bothell. Generally, learning goals of the course are to introduce an interdisciplinary audience to statistical thinking that enhances their conceptual understanding, ability to interpret basic statistical results, and apply statistical analyses to test hypotheses. Given the significant latitude my institution provides BIS 315 instructors, I also integrate active learning (see chapters 9-12 on active learning in statistics courses) into the course and frequently do this through community partnerships.

To integrate a community learning experience into BIS 315, *Real Change* and I developed a survey intended to understand various patterns and characteristics about their newspaper readers. The aim of the survey was to gather feedback regarding the quality of the publication, readership styles (e.g. other publications read), and to better understand readers’ demographics, for example. Below, I outline the steps our collaborative effort took in the development and implementation of the survey into my statistics course. My goal of this chapter is to provide a clear “recipe” so that a reader of this chapter may successfully implement a similar project into their own course.

**Project Planning**

About six months prior to the start of the Spring 2015 quarter and the commencement of BIS 315: Understanding Statistics, *Real Change* and I began forming our relationship and planning the project. Although the planning was not intensive in time or energy, it took several months due to scheduling and other factors. Below, I outline these processes. Table 1 provides a timeline of the project planning and implementation processes.

**Step 1: Partnership Development**

Any successful community-based educational experience requires strong partnerships. As a community psychologist, I spend a significant portion of my time building and maintaining relationships with community leaders. I am fortunate that a considerable portion of my job is being in community. For this particular project, however, the relationship between myself and *Real Change* occurred through fortuitous luck. A *Real Change* manager had posted on social media asking for assistance from someone with survey expertise. A colleague of mine forwarded the post to me, at which point I contacted *Real Change* to discuss a possible collaboration. After an email correspondence, I met with *Real Change* leaders at their offices to discuss a potential collaboration and integration of a survey project into a class. Once we agreed verbally to the terms, we moved forward with survey development.
Although my position requires that I build and maintain relationships, not all educators have the privilege of spending their time and resources doing this work. Fortunately, most universities have community-based and/or service-learning offices that are solely dedicated to doing this work. The mission of the Community-Based Learning and Research (CBLR) office at the University of Washington Bothell for example, centralizes “the advancement of mutually beneficial relationships between the university and the extra-campus community through community-based learning and research for the purposes of education and growth among all involved parties.” To do this, the CBLR fosters relationships between communities, faculty, and students to partner on projects “for the purpose of mutual beneficial exchange of knowledge and resources in a context of partnership and reciprocity.” For educators who may not have the individual capacity to build community partnerships, their home institutions likely have an office dedicated to this work.

Table 1: Timeline of Project Components

<table>
<thead>
<tr>
<th>Dates</th>
<th>Project Component</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project Planning</strong></td>
<td></td>
</tr>
<tr>
<td>&gt;6 Months Prior to Project Implementation</td>
<td>Step 1: Partnership Development</td>
</tr>
<tr>
<td>(Service learning offices can significantly expedite this process)</td>
<td></td>
</tr>
<tr>
<td>3-6 Months Prior to Project Implementation</td>
<td>Step 2: Survey Development</td>
</tr>
<tr>
<td>3-6 Months Prior to Project Implementation</td>
<td>Step 3: Data Collection Design</td>
</tr>
<tr>
<td><strong>Project Implementation</strong></td>
<td></td>
</tr>
<tr>
<td>Week 1 of Term</td>
<td>Step 4: Orientation</td>
</tr>
<tr>
<td>Weeks 2 – 8 of Term</td>
<td>Step 5: Survey Launch and Data Collection</td>
</tr>
<tr>
<td>Weeks 3 – 8 of Term</td>
<td>Step 6: Course Content</td>
</tr>
<tr>
<td>Weeks 8 – 10 of Term</td>
<td>Step 7: Analyses, Report Writing, &amp; Presentation Preparation</td>
</tr>
</tbody>
</table>

Step 2: Survey Development

Three years prior to this project, Real Change developed a survey of their readers. Real Change leadership utilizes their reader survey to assess a variety of topics about the paper, their vendors, and their readers and make organizational changes based on results, including informing their strategic plan. These include, for example, the quality of different paper sections, reader demographics, and the top reasons readers purchase the paper. Six months prior to the implementation of the survey into the statistics course, I met with Real Change leaders to negotiate implementation. Once an agreement was reached, Real Change leaders and I updated the survey. This iterative process was such that Real Change made suggestions, I would implement those edits and make my own suggestions, and Real Change would make more recommendations. Generally, Real Change leaders provided the context expertise (i.e. what questions are most useful and appropriate) and I provided survey design expertise (e.g. what scaling or response options are most effective). Considering Real Change had a survey developed, this process mostly included minor edits to the existing survey. In total, it took about
five iterations of conversations to finalize the survey for implementation. Once the survey was finalized, it was necessary to design the data collection method.

**Step 3: Data Collection Design**

Considering that readers only interact briefly with vendors (most interactions are only a few seconds), it was important to develop a data collection design that had the greatest potential to recruit survey participants – i.e. *Real Change* readers. *Real Change* understood that there were two important factors in recruiting participants – vendor motivation and convenience for participants to complete the survey. Although *Real Change* advertised the survey in the paper and on their website, Facebook, and other social media outlets, the paper heavily relied on its greatest partners – *Real Change* vendors. Vendors provided cards (similar to business cards) to readers that contained a weblink to the survey. The card also contained an ID code that was associated with the vendor from whom the reader received a card. Vendors were provided payment in the form of a free paper (a value of $0.60) as compensation for recruiting survey respondents. For each respondent who completed the survey, vendors received one free paper.

**Project Implementation**

Given the interdisciplinary nature of the faculty, instructors of BIS 315 are given significant latitude in the design of courses. The class met two days a week (Tuesdays and Thursdays) for two hours a day. Because many students commute to our campus, the need for collaborative and synchronous work required that much of the project took place during these four hours a week. In addition, some students do not have access to software needed to complete project tasks (e.g. Microsoft Publisher), which also required that work was conducted in the classroom space. The room itself was a lab-type room with 34 computers. BIS 315 typically has a maximum enrollment of 32 students, providing plenty of computer access to all enrolled students. For this class, I teach a combination of conceptual content (e.g. what are descriptive statistics and when should they be used?) and applied statistical content using Microsoft Excel. I have eliminated all but the most basic statistical hand calculations (e.g. central tendency) from the course. I opted to use Excel over other programs due to access (compared to SPSS and Stata) and ease of use (compared to R), although our classroom had licenses to these programs. All the analyses conducted in class can be done with Excel – albeit, some analyses are not quite as smooth compared to other statistics packages. Additionally, once our students graduate, they are much more likely to use Excel for work related projects. Once all data analyses were completed, the report was constructed with Microsoft Publisher and the presentation conducted in PowerPoint.

While the survey and data collection designs were conducted prior to the start of the term, the creation of an online survey, data collection, cleaning, analysis, and reporting were all conducted in-class throughout the Spring 2015 quarter. Each of these steps were conducted primarily by students. This 10-week course began on Tuesday March 31st and ended on Thursday June 4th with a final presentation given to *Real Change* leaders on Tuesday June 9th (during finals week). To assist with project flow, I constructed committees where each student was randomly placed into a committee that was responsible for a specific component of the project. The 32 students were split into eight groups of four students. See Table 2 for a description of each committee. Because there were ebbs and flows for each committee, students would move outside of their own committees if another team needed assistance.
Table 2: Description of BIS 315 Project Committees

<table>
<thead>
<tr>
<th>Committee Name</th>
<th>Responsibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey Construction Committee</td>
<td>Built the online survey tool in Qualtrics (<a href="http://www.qualtrics.com">www.qualtrics.com</a>)</td>
</tr>
<tr>
<td>Data Analysis Committees</td>
<td>Conducted various univariate and bivariate descriptive analyses on assigned components of the survey (e.g. reader demographics)</td>
</tr>
<tr>
<td>Report Committee</td>
<td>Communicated across data analysis committees and wrote the final report for Real Change</td>
</tr>
<tr>
<td>Presentation Committee</td>
<td>Communicated with the report committee and presented brief findings to Real Change leaders</td>
</tr>
</tbody>
</table>

**Step 4: Orientation (Week 1)**

During the first week of class students were oriented to both Real Change and the project. On the first day of class, Real Change leaders attended the class session and introduced students to the organization, their purpose and philosophy, and to the reader survey. On the second day of class, I outlined the steps of the project and the goals for the class. On that same day, the first committee – the survey construction committee – was created.

**Step 5: Survey Launch and Data Collection (Week 2-8)**

After the second class (end of week 1) the survey construction committee created and activated the online survey, which was completed by the end of week 2. At that point, the entire class piloted the online survey. On the first day of week 3 all students completed the online survey and noted any issues with it. Real Change leaders also completed the survey and provided feedback to our class. The survey construction committee then updated the survey to address any issues raised by the class or Real Change. Once the online version of the survey was finalized, it was officially launched on April 24th 2015. To give readers time to participate, the survey was live until May 23rd 2015 – the end of week 8 of the quarter. In total, the survey generated more than 1,100 responses from readers.

**Step 6: Course Content (Week 3 – 8)**

Between weeks 3 and 8 (while the survey was active) I taught standard course material such as sampling, probability, descriptive statistics, inferential statistics, etc. Surprisingly, because of the deep integration and the intensity of the first few weeks working to the launch the survey, I was able to center the Real Change survey to all course content. For example, when discussing measurement and scaling, we turned to the survey for examples of levels of measurement (i.e. nominal, ordinal, interval, and ratio). When discussing sampling, we examined the costs and benefits of our data collection method (i.e. convenience sampling) against other methods (e.g. random probability sampling). Once we moved into descriptive (e.g. central tendency, dispersion, etc.) and inferential (e.g. t-tests, ANOVA, etc.) statistics, we downloaded survey responses and used available data for analysis. The integration of the survey into all components of the course along with deep familiarity with the survey and associated data provided a strong foundation for the final steps of the process.
Step 7: Analyses, Report Writing, & Presentation Preparation

Once the survey was closed (May 23rd, the last day of week 8) and responses were downloaded from Qualtrics, the final two weeks of the quarter (weeks 9 and 10) were solely dedicated to data analysis, report writing, and presentation preparation. Over the weekend between weeks 8 and 9, I took time to clean the data in preparation for analysis. Students were grouped into committees (see above) – most students were members of analysis committees, each of which were responsible for analyzing data associated with specific survey sections. Because all analyses conducted were univariate or bivariate descriptive and the existing variables/scales needed some transforming, I worked hands-on with committees to guide them in conducting their analyses. After an analysis committee felt an analysis was complete, the team would send their results and associated figures/tables to the report and presentation committees to integrate into those two components of the project. Often, a member of the report or presentation committee would request a specific analysis to be conducted, at which point the associated analysis committee would complete the analysis.

On June 9th, 2015, during finals week, a team of six Real Change leaders came to our campus for a presentation given by the students of the presentation committee. Real Change leaders watched a 20-minute presentation. At the end of the presentation, Real Change leaders asked the class various questions about the analyses and engaged in a larger conversation about the survey, homelessness, and their readers generally. Much of this conversation included speculative questions that were not answerable from the data (e.g. “I wonder if so-and-so vendor has better relationships with readers since most surveys were completed in this neighborhood?”). Overall, Real Change leaders were impressed with the presentation and report. In fact, I have partnered with them (outside of a class context) to collect and analyze data for their 2018 reader survey.

Potential Impacts

A social justice education is one that raises awareness of social issues, takes critical actions to address those issues, creates accountable allyship with those most affected by them, and builds collaborative partnerships with communities. My goal for this particular class was to touch each of these components of a social justice education by utilizing institutional human (e.g. students) and resource (e.g. class time) capital to take action on a growing issue of importance in Seattle. Additionally, because data, and by extension statistics, is a form of knowledge, and as the cliché attests, “knowledge is power,” I sought to build power in partnership with Real Change to promote their capacity to address homelessness. Through our deep (and continuing) partnership and prolonged focus on homelessness, this class had the potential to raise students’ awareness of homelessness issues in Seattle; by utilizing the survey and data as a tool for power, our class took critical actions to address issues of homelessness; and by taking guidance from the leadership of Real Change, we were accountable allies to the organization. Although this chapter is not intended to evaluate the impact of the survey directly on Real Change, the students, or the university, there were several goals that had the potential to make an impact to these groups. Below, I outline these potential impacts for the students, Real Change, and the institution.

Student Impacts

Beyond typical undergraduate statistics learning goals (e.g. understand and calculate basic social science statistical concepts, enhance knowledge of analytical tools, etc.), this course sought to foster a social justice education through the application of statistics in a community partnership. Students had the opportunity to learn statistics, but to also apply that learning in a
real-world context. During the last two weeks of the quarter (Step 7), students applied their learning to data that they collected and that had the potential to positively affect a community partner. Through this process, they grappled with the nuances of data analysis and reporting that may not have been present without the partnership. For example, students had to answer questions such as: What components of the survey should be included and excluded in a report? How do we interpret and report a Likert-scale to partners? For bivariate analyses, what variables should be compared? Additionally, by partnering with an organization like Real Change, students were intimately exposed to issues of homelessness throughout a 10-week quarter. I believe this exposure and connection with Real Change increased the likelihood that students would become more critically aware of homeless issues.

Impacts for Real Change

By partnering with a university course, leaders at Real Change increased their capacity in several ways. First, by outsourcing a large project they were able to free up significant portions of time for their staff to work on projects that are more central to the function of the organization, such as writing the newspaper. Throughout the quarter, students estimated that they cumulatively volunteered more than 550 hours to the project and I personally dedicated about 80 hours. Second, the students and I provided content area expertise of survey methods, measurement, data collection, and statistical analysis that may not have been readily available to leaders at Real Change. And third, Real Change has utilized the findings of the study to inform their organizational strategy and strategic plan, particularly around access to the paper, diversity of readership, and even funding.

Impacts for the University of Washington Bothell

The university has also potentially benefited from the deep partnership created with Real Change. Specifically, students and faculty have both continued working with Real Change in a variety of capacities. Indeed, more than two dozen students have worked directly with Real Change in courses and on other projects. Additionally, the university recently applied for the Carnegie Classification for community engagement and featured the Real Change project (and other projects with Real Change) as a component of the application process. This deep partnership has been touted across the campus as one framework for conducting community-based teaching and scholarship.

Future Considerations

Although the project was successful, in reflection, there were areas that could have been improved upon for student learning and for potentially greater impact to the partnership. Below, I outline some “future considerations” from the partnership and provide recommendations for instructors who may consider developing such a project.

Create a Memorandum of Understanding & Student Contract

Fortunately, communication, mutual respect, and goals between myself and Real Change leaders were strong, so there were no conflicts during the course of the project. However, this may not be true for all partnerships. As such, in the future I would create a memorandum of understanding (MOU) to outline issues such as the purpose of the project, data ownership, and roles and responsibilities, for example. MOUs are agreements between faculty and community partners that can be referenced in the case of a conflict or disagreement. In addition, they can assist in creating a vision for the project and answering any questions that may arise in the
partnership. Most universities have MOU templates that faculty can adapt for their own purposes.

In addition to a MOU, I would also recommend creating (or adopting) a student behavioral and/or privacy contract. My institution requires such a contract for students participating in community-based learning within minors (e.g. at an elementary school), it does not for other community partnerships. Although I did not experience any behavioral or privacy issues in this course, there is always a possibility. For example, students had open access to the data, which could have been downloaded and shared. If working with more sensitive and/or identifiable data, a student contract is a necessity.

**Develop a Timeline and Clearly State the Project**

*Real Change* had a clear goal when pursuing the partnership – collect reader data via survey methods. Because there was a clear goal the project was able to proceed with very few (if any) hiccups. A MOA may be a good way to outline specific project deliverables and create a timeline. Without these, the project could have easily been derailed.

**Consider the University System**

The University of Washington has adopted the 10-week quarter system and the school of interdisciplinary arts and sciences does not have sequenced courses. This created additional pressures to implement the project within the short time constraints. A project like this may have a greater potential to impact student learning if created in a semester system or in a sequence of courses (e.g. research methods followed by statistics). If implemented in one of these formats, we could have taken more time to explore some of the nuances of data and statistics that we either skipped or I ended up taking on myself (e.g. survey development, data cleaning, etc.).

Service-learning is not a priority on all campuses, and even when it is, processes and outcomes for faculty and students are not always clear. At UWB, for example, very few BIS 315 courses include a service-learning component, and the system we currently use, flagging service-learning courses with a “S,” is not widely recognized by faculty and students. As such, many students who enrolled in this particular course were surprised by the project. Fortunately, most enjoyed and appreciated the experience, but some would have opted to take another section if they knew they would have to work on such a project. Additionally, there is not a system in place to formally recognize faculty who conduct service-learning projects nor for students who participate in them. My use of community partnerships for courses like statistics is a philosophical and pedagogical decision and may not directly translate into promotion and tenure. This is likely the same for others who adopt a similar approach.

**Consider the Class Structure**

For this project, I opted to create committees in which students conducted their project work. There were several issues with this structure: (1) there was often lags in workload where one committee would have a significant workload followed by very little work (e.g. the survey construction committee); (2) communication between committees was often difficult to foster; (3) and there was overlap between analysis committees such that, at times, two committees would conduct the same analysis. To address some of these issues, I would consider creating a similar committee structure but one that allowed for greater movement of students between them and for the formation and dissolution of committees where needed.
Conclusion

This chapter illuminates a social justice-oriented project, in partnership with a homelessness advocacy organization, and conducted in an introduction to social science statistics course. The purpose of the project was to utilize institutional resources and statistical knowledge as tools for social justice education. Throughout this project, I partnered with Real Change to conduct a survey of their newspaper readers. Throughout a 10-week quarter, students created and collected data from an online survey, analyzed survey data, wrote a report, and presented findings to Real Change leaders. Through this process, students had the opportunity to apply real world data to an issue of significant concern to the local region. This project had the potential to impact students, Real Change, and the institution. However, there were some issues that arose throughout the project that could have been avoided. As such, the purpose of this chapter is to provide clear directions on how to conduct a similar project and avoid some of the pitfalls I encountered. For another perspective in implementing service-learning, see part 3 of chapter 7 of this book.
References


Appendix: Research Project Committees Handout and Timeline

BIS 315: Understanding Statistics Spring 2015 Committee Descriptions

**Research Project Committees**

Students will have the opportunity to continuously apply course knowledge through a research project in partnership with the community organization Real Change (www.realchangenews.org). Real Change is conducting a survey of their newspaper readers to identify their reading demographic, the quality of their newspaper, and potential ways to improve the paper. Our class is tasked with building the online survey; collecting, cleaning, & analyzing the data; and preparing a report to present to organizational leaders.

This document outlines each committee that will be created to complete the class project. This document also provides a description of the work expected from each committee. Students should read the information below and sign up for the committee for which they would like to contribute. Spaces for each committee are limited so it is recommended that you sign up for a committee ASAP.

Along with committee work, groups are STRONGLY recommended to create roles and responsibilities for each group member. Roles and responsibilities are to the discretion of the group. Additionally, all committee communication should take place via canvas discussion boards – even for issues such as scheduling meetings. This is important so that I have all the information needed to mediate any potential group issues. Without this information, I will not be able to mediate any potential issues.

**Below is a description of each committee:**

**Online Survey Construction Committee:** Responsible for developing the Real Change reader survey using an online survey tool (e.g. Survey Monkey, Qualtrics, etc.). Students in this committee will identify the online survey tool and use the finalized reader survey to build the website. NOTE – Members of this committee may be reassigned into one of the committees below at the conclusion of responsibilities.

- **Timeline:** Weeks 1-3

**Data Download and Cleaning Committee:** Responsible for downloading the data collected by the online reader survey and “clean” the data. To clean data, the committee will analyze the data by running basic descriptive analyses to identify whether all data are within a feasible range. For example, the committee will ensure all ages reported are within a feasible range (e.g. 16-85 years old). If there are cases beyond feasible ranges, the committee will have to make decisions about what to do with the data. Essentially, the committee will be tasked to ensuring the data are usable and in a place to hand off to the committees below. NOTE – Members of this committee may be reassigned into one of the committees below at the conclusion of responsibilities.

- **Timeline:** Weeks 4-5

**Descriptive Statistics Committee:** Responsible for conducting and interpreting descriptive statistics (e.g. central tendency, dispersion, etc.) in preparation of writing the report.
• Timeline: Weeks 5-10

Inferential Statistics Committee: Responsible for conducting and interpreting inferential statistics (e.g. t-tests, ANOVA, chi-squared) in preparation of writing the report.

• Timeline: Weeks 5-10

Graphs & Charts Committee: Responsible for utilizing the analyses from the descriptive and inferential statistics committees above to create visual representations of the data in preparation for the final report.

• Timeline: Weeks 5-10

Report Committee: Responsibility for ensuring the quality and content of the final report. This DOES NOT mean writing the entire report. However, this committee will work closely with the descriptive statistics, inferential statistics, and graphs & charts committees to build the content of the report. Additionally, this committee will be responsible for editing and finalizing the report.

• Timeline: Weeks 5-10

Presentation Committee: Responsible for presenting the results of the report to organization leaders at Real Change. This committee will collaborate primarily with the report committee to build a presentation suitable for a public presentation.

• Timeline: Weeks 5-10

Qualitative Committee: Responsible for accumulating, analyzing (e.g. theming), and presenting the qualitative data collected from the online reader survey. In the reader survey, there will be a couple qualitative questions (e.g. “What do you think Real Change can do to improve the paper?”). The qualitative committee will collect these responses and make sense of them for Real Change.

• Timeline: Weeks 5-10
News Students Use: Critical Evaluation Assignment for Undergraduate Statistics Course

Alisa Beyer, PhD
Chandler-Gilbert Community College

Summary

In this chapter, I review the whys and hows in creating a summative end of term writing assignment that assesses critical thinking skills and application of statistical knowledge through analyzing a research article and how it is reported in the media. Students evaluate news article coverage connected to a primary research article, explain connected course concepts, and critically evaluate the presented research. I expand on a preliminary assignment that prepares students for the end of the semester paper. In the conclusion section, I identify ways to either expand upon or pare down the assignment.

Why did I Create this Assignment?

All students are continuing to be consumers of media regardless of the scientific training they will further receive. Students taking statistics also may not realize how statistics is relevant to their lives (unless you share that – see chapter 1 and 9 for inspiration). In designing this assignment, I wanted students to have a major assignment in the course that applied their knowledge of statistical concepts to news stories. Students may identify reputable news sources and perceive that information is being reported accurately, yet oftentimes news stories do not report the complete story due to brevity, knowledge base of the reader, and perspectives of reader interest. News stories may omit essential information for the reader to fully evaluate research conclusions and limitations. Students may not realize how much knowledge they have gained in the course to critically evaluate news stories sharing research. Thus, this assignment is aimed at connecting students’ in class learning to the real world.

In addition to an apprehension to learning statistics, students can be intimidated and overwhelmed by empirical research. At my college, it is uncommon for students to have read a primary research article in introduction to psychology. I sometimes hear students say that they had never read an empirical article, or if they had, they focus on the introduction and discussion sections. Before this course, students may fail to connect and critically evaluate the hypotheses, methods, results, and author’s conclusions. The proposed assignment helps students do all of this while practically applying their new statistical knowledge with an opportunity to write about psychology.

I made this into a formal writing assignment for students to continue working on their writing. To me (and many others), writing is a learned behavior that benefits from practice. Students can learn more about context and audience when writing – carefully considering how much information is needed for the reader as the audience is not present (Emig, 1992). Strong writing skills are desired in the workplace as well. The National Association of Colleges and Employers (NACE) (2019) identified written and oral communication skills as essential for career readiness. In fact, the NACE 2019 Job Outlook survey found that written communication skills was the
most wanted skill to see on student resumes (82% of respondents). This assignment met my college department’s strong preference for instructors of 200-level classes to include a minimum of a 5-page writing assignment. Additionally, this assignment was designed measure students’ application of knowledge, information literacy, and critical thinking skills for my college’s program assessment.

One of the essential learning goals for any undergraduate psychology program is to think critically about research (APA, 2003). I think this assignment assesses an important steppingstone for identifying students’ evaluation and analysis of research. A major portion of this assignment is critical evaluation and analysis of the research article and news story. Students gain experience and practice analyzing secondary source content and identifying primary sources of reading connected to secondary sources. This assignment creates a formal opportunity for students to recognize the knowledge they have in evaluating new sources, knowledge they have about statistics, and their ability to make connections between results and potential misinterpretations or biases of the reported results. This assignment connects to goals four and five of APA’s Statistical Literacy Taskforce (Addison, Bliwise, Green, Heinzen, Nolan, Posey, Wendorf, Wilson-Doenges, 2014). To be a good consumer of the science of psychology, students need to be able to identify when statistics are being used and reported appropriately— including secondary references of study results. Goal five is to evaluate the public presentation of statistics including recognition if statistical are presented in a misleading way, and to validate the conclusions made. Goal four is to distinguish between statistical and practical significance including interpreting meaning of statistical significance and effect size statistics. Evaluating if an effect is “real” and further identifying its practicableness are two important learning outcomes for applied statistics (Addison et al., 2014).

A final assignment objective is for students to gain more digital learning and information literacy skills. Students may better understand and identify why secondary sources, particularly from the media, are not appropriate sources for most term papers (Connor-Green, 2000). I work with students to use appropriate attributions of sourced information using APA citations and references. Students learn the basics of APA Style while completing the assignment.

Assignment in the Making

I created this assignment three years ago after wanting to create a summative assignment that tied into college and program learning outcomes. I was inspired by an article from the AACU that impacted my design of the assignment—and while I really want to cite the reading here for your reference, I can’t seem to locate the source. The article reviewed tips and strategies for students reading primary research articles, upon which I based some of my writing assignment prompts. Last year, I also took a more explicit approach in tasks and skills students needed to set them up to be successful for the assignment.

To give students as much guidance as possible, the paper is broken down step by step for the student to critically read, analyze, synthesize, and apply information learned from the course to their analysis of the paper (see Table 1 and appendix A for complete handout). Students are instructed that each item explained in the instructions should be a section of the paper. For the final paper assignment, students select from two or three news/research article picks (see appendix B for some popular selections from students over the past three years). You could have students find their own topics and papers, but to save yourself time reviewing those (and frustration for students searching for appropriate combination), an list of choices is best. I provide a handout based on Anisfeld’s critically reading a research article (Anisfeld, 1987; see appendix C). Appendix C has tips for students, step by step, how to critically read the research paper. Students are also given more practice and instruction in a mini-version of this assignment.
Table 1. Brief instructions for final paper (full version in appendix A)

1. Consider. First re-read your selected news article.
2. Read and connect. Now I want you to read the corresponding primary article. Connect the two articles.
3. Consider. Now I want you to summarize the key points for each section of the primary research article – in your own words.
4. Analyze & interpret the data. Now it is time to really dig into the results section of the primary research article. Connect terms and concepts from class to the results section (as well as methods if applicable).
5. Analyze & interpret the data part 2. Connect the results with the study research hypothesis.
7. Think of the next study.

Scaffolding the Final Paper

The basics of the assignment are introduced on the first day of class and are explicitly linked to the benefits of taking the class. Around week six of the course, I provide more detail for the final paper assignment (16-week term). For context, by week six students have been introduced to basics of inferential statistics and hypothesis testing. After we cover t-tests, we work on a mini-version of the paper. The mini-version of the paper is meant to introduce them to critically reading a research paper and scaffold them for the final paper.

For the mini-version, students first read a New York Times article (assigned as homework) and then during the next class period, students read the primary article and answer questions in groups. This activity allows student to develop a critical reading lens when reading news sources and apply their data skills from the course to real life. Students fill out a basic worksheet that identifies the types of information and supports the skills they need for writing the final paper (see Table 2 and appendix D). In appendix D, I also provide sample instructor responses specific to the assigned article I use (bolded).

Table 2. Example reading prompts for empirical article questions

1. What was XX interested in studying?
2. What were her/his methods? Connect to terms from the class (experiment, repeated measures, between subjects, etc)
3. What were her/his variables of interest? (IV, DV)
4. Review the results and statistical tables/figures. Talk about statistics findings in the tables and what they mean. Explain the terms and concepts as defined in class and then describe the context of that term in the results section of the research paper. Connect the results with the study hypothesis.
5. What is the conclusion that researcher makes? Can you think of other explanations? Also compare to news article.

The article I use for scaffolding the final paper activity is based on a study done by Karen Wynn (1992) and first published in Nature. The content of the article relates to infant cognition, and typically about one-third of the class has taken or is enrolled in developmental psychology. Even if students have not yet taken a developmental course, all students have had some introduction
to developmental psychology in an introduction to psychology course (a pre-requisite at our institution). If you are worried that students may not be familiar enough with infant cognition, the brevity of the article means you could provide a quick mini-lecture on infant cognition or the study design (see appendix E). The news article selected for this paper is also unusual in that it is written by Daniel Goleman (1992). I like the corresponding New York Times article that compliments it as we have a short discussion about the news article being written by a psychologist verses a journalist. Furthermore, it is unusual for this news and research combination that Goleman provides more interpretations for the findings than Wynn does. I selected the Wynn study because with such a short research article, there is an opportunity for deeper student discussion. The Wynn research article provides a summary table of the t-tests (helps students sift through the results). The Wynn study also uses multiple t-tests (we converse about this decision). I also have an opportunity to talk more about bigger ideas connected to the new statistics (see Part 4 of this book on Advanced Topics). I also bring in a discussion about using the new statistics (see chapter 13) and I remind students to consider examining use of new statistics (i.e., effect size) - even if the selected research article does not.

The first two semesters that I used this assignment, I was able to do two practice in-class assignments, but timing was off these past two years which left me cutting out one of the practice activities. The additional practice activity was the same set up, but we did a news article and research article tied with using ANOVAs. I kept the t-test so students have a scaffolded assignment for the final paper earlier in the semester (most articles options I pick for the final paper report ANOVA, or correlation and regression). Although I call this activity a mini-paper, the work submitted was the appendix D worksheet with important aspects being covered in class discussions.

If we have time during the semester (usually around week 14), I set aside at least 30 minutes of class time for a final paper check-in. For this session I ask students to bring in their responses from the research article reading tips handout and any work they have done on their paper. During the 30 minutes session, I first divide up the class into working groups of 3-4 students based on where they are on their paper and what topics they were using. I have students work in groups (based on topics or where they are working on in the paper) for future guidance, or occasionally students are ready for peer feedback for a section of their paper.

Outcomes and Future Considerations

The rubric I developed was for assessment of application, critical thinking, and attribution (APA citations and references). It is based on the college and department level assessment practices. I have included a sample rubric in appendix F which includes evaluation of the:

- summary of the news article
- summary of the empirical article
- application of concepts from statistics to the readings
- analysis and evaluation the readings including big picture components for statistics
- use of appropriate APA style (focus for APA is citations and references)
- writing skills (e.g., organization, grammar)

This assignment identifies students’ knowledge and application of statistics. Students do well on this aspect of the paper. They can share more about what they know for sampling and statistics reported in the paper. For the most part students can differentiate and synthesize between the news and research article. Students still struggle with evaluating the research and thinking more critically about the statistics. Each semester I strive for stronger papers as I try to give more scaffolding, however, I have not seen across the board improvement in grades. I always get a
handful of students who turn in papers in the D range, and a handful of excellent papers and then many that are satisfactory (B – C range). I did include a sample student analytical section from a student who did well on the paper (appendix G). Even with small changes made over the past few years, I still feel some disparagement with the analytical quality of some of the papers.

I think a key to student success for the student writing quality was clear guidance. The majority of students were engaged during the t-test mini-paper discussion, and most submitted the mini-paper worksheet. Providing the mini-version of the final paper assignment before the end of semester was crucial in helping students gain a better understanding of the assignment expectations. The few students who did not do the mini-version struggled with the final paper (note that the mini-paper was an in-class assignment and I allowed to submit late, but few who missed class turned the mini-paper in).

I did not see a difference in the quality of papers after dropping the second practice opportunity (mini-paper) – but I also improved upon the instructions and guidelines for the assignment. Although grading summative papers is timely and adds more work at the end of the semester, I find satisfaction reviewing students’ applications shared and evaluation of the articles. Most of the students demonstrate understanding of the results and can synthesize the news and research articles. Honestly though, each semester I debate having this paper verses having smaller assignments that mirror the paper for more practice of concepts (but less writing about it). (Grading wise, I usually spent about 12-15 minutes per paper.) Perhaps with more mini-papers, students would develop stronger analytical skills for those bigger picture ideas.

I do not allocate time for peer review for the final paper, but that is something that could be incorporated. If the schedule allows, I end class early to answer questions about the final paper. I typically get about one-quarter of the students coming to office hours for input, advice, or questions about the final paper. If you did peer review, I still suggest the mini-paper activity. Even without class time for peer review, some students worked in pairs (on their own) to evaluate the readings for their final papers. I do not have an issue with this as I emphasize using their own wording and to keep separate notes as they are working together to answer the guiding questions for the paper. A few of the students did take advantage of the writing center, but I did not collect information if students did any peer editing of each other’s papers. While I do not leave class time for peer review, I do share tips for using the two-level pass through approach for peer review (or self-review) (see Beyer, 2018 for more guidance). I do not worry too much about student plagiarism as I use the college plagiarism software, and I switch out articles each time I teach the class. Each semester, I select at least two different articles from previous semesters to cuts down on students’ susceptibility to recycle previous students’ work.

Although many students don’t want to be writing papers at the end of the semester, the final paper gives me one last chance to examine their statistical thinking, interpretation of secondary and primary sources, and their understanding of research. Each semester, a few students mention the final paper was their favorite assignment in the student course evaluation comments. This positive feedback helps, but I do have some suggestions for how to scale down the assignment.

If the final paper is too overwhelming for you (or your students), I suggest keeping the idea of synthesis, analysis, and application experiences with news media and research but have the assignments be similar to the mini-paper (worksheet) with discussions. Students could be given a news article to review and answer questions for homework and then have the students work in groups to read over the research study ending class with a synthesis discussion. This would get students applying, synthesizing, and analyzing the texts without the inconvenience/timing of grading a lengthier paper. Connor-Greene (2000) also outlines an in-class activity for students to evaluate a news article and highlights sections of the primary research article. Her activity
includes a section for students to come up with recommendations for journalists’ approach to reporting science. Connor-Greene additionally followed up with students to write a critique identifying their own news and primary research comparison. You could have this be a capstone-like activity at the end of each statistic covered where students do a mini version of this assignment for each statistic.

Wrap-Up

This assignment is useful for any student, regardless of their major. Students develop skills in writing, understanding research, information literacy, applying their knowledge of statistics, and synthesizing across information sources. These are needed skills for the 21st century educated citizen. I encourage you to give some version of this assignment a try!
References


Statistics are all around us – so is psychological research. We hear about studies all the time in the news. How accurately are these studies summarized in news articles? This assignment gives you a chance to read over some of the psychological studies that have been reported in the news and then apply your knowledge of statistics to these new stories.

**You have 3 news story options for the project.** First review each one to decided which topic you will dig further into. The news articles (several from *New York Times* (NYT)) articles are available in Canvas.

**New article options (pick 1):** Articles/links are posted on Canvas!

Then you will read the primary research article that connected to the new story you selected. Links to the articles are in Canvas (also available via CGCC library). Each news article as a set primary article (empirical study) to review.

**Primary article options (pick one that corresponds the news article chosen above):** Articles/links are posted on Canvas!

**Note:** You may also wish to review the primary research article before your final selection of the NYT article. If the research article has more than 1 study, you will be asked to select 1 of the studies discussed in the article (not more work if you pick one that has several studies – all articles have been set up to be equal in terms of paper responses).

**You will then write up a paper answering the following questions (please follow this organization for your paper)**

8. **Consider.** First re-read your selected *NYT* article. Provide a summary in your own words of the key points to the article as well as important connections to the statistics course. What key findings were shared about the research? How were the research methods described? What was the key point of the article?

9. **Read and connect.** Now I want you to read the corresponding primary article. I want you now to connect how the NYT article connects to the primary research article. The research article has an introduction, method, results, and discussion section whereas new stories intermingle all of these things. Connect the two articles. If your primary article has more than 1 study summarized – choose study 1 or study 2 (be sure to indicate which was chosen).
10. Consider. Now I want you to summarize the key points *for each section* of the primary research article – in your own words. This is NOT paraphrasing the published abstract but your own summary and takeaways *for each section* of the paper (introduction, method, results and discussion). If your primary article has more than 1 study summarized – keep working with the same study you selected for question 2.

11. Analyze & interpret the data. Now it is time to really dig into the results section of the primary research article. **Connect terms and concepts from class to the results section (as well as methods if applicable).** You will explain the term and concept as defined in class and then describe the context of that term in the results section of the research paper. If your primary article has more than 1 study summarized – keep working with the same study you selected for question 2. If your study had more than 1 outcome measure, focus only on two of the outcome measures. I expect that you will make at least 7 connections – defining the class concept while connecting to the news or research article.

12. Analyze & interpret the data part 2. Connect the results with the study research hypothesis. Critically discuss issues including sampling, sample size, how results were presented, implications draw from the researcher in the conclusion section.

13. Synthesize. How well did the journalist (NYT article) connect to the research paper? What do you think should have been added? Was there anything misleading presented in the NYT article? If you were to be the journalist, how would you have integrated the findings from the primary research paper (in a way for someone who has not taken this class to understand the study findings)?

14. Think of the next study. Given what you have read in these two articles, what would be your next step research wise? Describe your research question and possible study design. What is your research question? What is your study hypothesis? Identify your key variables (including defining how to measure the key variables) and the test-statistic you would use to test your hypothesis.

**Key assignment information:**

- you pick the news article and use the corresponding research article to answer the questions listed above
- you need to identify and share related content from the class and apply to the articles
- you should review major course concepts as you analyze, interpret, and synthesize the articles
- you paper should address all 7 prompts, while I don't want you to number and respond, you can organize your paper with subheaders for each prompt as a section of the paper
- paper should be at least 7 pages, but not more than 10
- you must use APA format and citation style

**Grading for write up:**

This will be graded based on how well you apply, analyze and evaluate the course material as connected to the readings. Points will be deducted for poor grammar/unclear writing.
Exemplary papers will use the textbook as a reference as well as connect to class topics. Exemplary responses will also integrate the course material with the readings using your own words/voice (not heaving lifting or quotations from your sources).

Remember to cite and reference textbook, news article and research article as appropriate. See Canvas for information on citations and references with APA.

The paper is worth 50 points.

Due XXX – upload to XXX
# Appendix B

Some of the most popular articles used over the past 3 years

<table>
<thead>
<tr>
<th>News article</th>
<th>Primary article</th>
<th>Inferential statistics involved</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>NYT</em> article, Picture a Leader. Is She a Woman? (3/18/2018), <a href="https://nyti.ms/2Dwi1xJ">https://nyti.ms/2Dwi1xJ</a></td>
<td>McClean, E., Martin, S. R., Emich, K., &amp; Woodruff, T. (2018). The social consequences of voice: An examination of voice type and gender on status and subsequent leader emergence. <em>Academy of Management Journal</em>, 61 (2), 1869-1891.</td>
<td>2 studies: Correlations, Cronbach alphas, multilevel modeling (study 1), 2x2 ANOVA (study 2), Confidence intervals <em>had my students focus on study 2</em></td>
</tr>
</tbody>
</table>

An amazing resource to check for student reader friendly research articles (then find media pairing): [https://www.everydayresearchmethods.com/2019/11/updated-list-articles-that-are-student-friendly.html?fbclid=IwAR2nA2_jg_5MvArrickyUychGu1j0fWNga5ox742-lH4n6eA1sX9U_egXM](https://www.everydayresearchmethods.com/2019/11/updated-list-articles-that-are-student-friendly.html?fbclid=IwAR2nA2_jg_5MvArrickyUychGu1j0fWNga5ox742-lH4n6eA1sX9U_egXM)
Appendix C

Tips and guide for critically reading empirical articles (for this class but can generalize)

1. Read the abstract.
2. Then read over the Method and Results sections of the paper. This request is for you to be independently considering the design, method and results before reading the author’s perspective and interpretation (as well as what is likely given the story in the introduction). Write down 8-12 sentences of your concluding thoughts about what method was used – and what information can we gain from that method; what analyses were run and why; what were the findings.
3. Review any tables or figures that summarize the results. Are the results consistent (is what is reported for the statistics the conclusion you would make given your knowledge learned from the course?) Is there a table or figure that you can draw to help you conceptualize the findings?
4. How strong are the findings? Consider evidences of positive results (rejecting the null), power, effect size, how many statistics reported were reject the null, etc. Remember to consider p-value, 1-tail vs 2-tail, etc.
5. Read the entire paper now – examining previous research and suspected results in the introduction, introduction of the constructs that were measures and how measured in the past…for the conclusion, examine the interpretation of the findings and limitations noted by the author.
6. Consider ways the author might have biased the findings – for example, examine coding and how variables were measured, if there was coding reliability, how subjective are the data recorded? Biases could also be interpretations – compare the finding and then how the author interpreted the finding.
7. Distinguish between theoretical constructs and operational constructs (how each variable was measured), did the authors include reliability and validity measures (could appear in introduction as well as method section)
8. Examine the information gained in the results and the methods and measures used to get them. Identify the process of how each works together. For example, if a researcher ran a regression, you would expect relationships between the variables reported AND the measures to have I&R data.
9. Contrast limitations the study has with possible limits in generalizability (sample, sampling, etc.)
10. Evaluate the study. Avoid nit-picking. Evaluate the method in terms of does it answer the questions proposed in the study. The most valuable critical analysis connects the results and conclusions to the methods – does these all fit well together? Can you identify another method or approach that could have been used given the data collected?
11. Remember the author is an expert, but you have objectivity, unbiased common sense, and general knowledge of statistics from this course – know your assets and don’t give up!
Appendix D

Scaffolded assignment and general feedback I give students

Psychology 230 Dr. Beyer t-test article reading


1. What was Dr. Wynn interested in studying?

Examining if babies know math – making argument have innate math abilities for simple addition and subtraction examining preferential looking for possible and impossible math out comes

2. What were her methods? Connect to terms from the class (experiment, repeated measures, between subjects, etc)

This is technically a mixed design but treated like multiple parts to study instead of all in one -- being published in the 90s – multiple t-tests were run. So it is repeated in that there was a pre-post (but treated separately). The focus of the study was comparing the post test comparisons. Babies were in 1 of 2 conditions, the correct math answer (plausible) and the incorrect math answer (implausible). Testing was done for addition and subtraction. This was an experiment where babies were assigned to 1 of the 2 conditions. Being assigned into 1 of 2 conditions is a between subjects design.

3. What were her variables of interest? (IV, DV)

*IV = math conditions

*DV = looking time

4. Review table 1. Talk about her statistics findings and what they mean.

All pre-tests showed no significance between the 2 conditions -- this is important to set up that babies are similar prior to manipulation. All posttest comparisons you see babies look longer at the incorrect answer (so for math 1+1 = 1 had higher score for LT than 1+1 = 2). Her groups refer to answer outcome so LT(1) is where they saw 1 as final answer (this is the correct answer for 2-1 but the incorrect answer for 1+1). Higher looking times mean the baby is spending more time processing the information (it is unexpected). Could also discuss and interpret M & SDs.

Explain the terms and concepts as defined in class and then describe the context of that term in the results section of the research paper.
Could connect multiple terms such as t, p, significance, pre-post, experiment, etc – you would define the term and discuss in context of study. For your paper, you would cite course materials as you connect to class materials.

Connect the results with the study hypothesis.

Found support – there were differences in the 2 groups. In both addition and subtraction, babies looked longer at the incorrect/implausible math answer.

5. What is the conclusion that Dr. Wynn makes? Can you think of other explanations?

Dr. Wynn states that math is innate – babies know math. Dr. Gelman in the NYT article brings up other explanations such as spatial learning and general idea of violation of expectancy (outcome is unexpected but not necessarily that this is evidence of mathematical literacy/learning). Can expand this section – also you can add in here other considerations you know from this class – like multiple t-tests are no good, power, effect size, replication, etc.

6. Review the NYT article – contrast what was share there to what your read in this article….

So here would discuss alternative explanations for findings. Could also discuss if level of detail was enough for the NYT, perhaps the Gelman summary was easier to understand than the academic language of the Nature article, etc – this is your chance to compare and contrast the 2. Consider what information in the empirical article (actual study) was missing from news article that would be more clear or if the news article appropriately shared findings, etc.
Appendix E
Example of some of the supplementary slides reviewing the Wynn study

Wynn study
NYT & Nature readings

Concepts of Number
• When do children begin to understand concepts of quantity and number?

Adding and Subtracting in Infancy?
• Karen Wynn (1995)
• Present addition and subtraction in a perceptual display

Adding and Subtracting in Infancy?

Adding and Subtracting in Infancy?

Adding and Subtracting in Infancy?
• 5-month-olds show increased looking to the “unexpected” result
• Is this evidence for adding and subtracting?
• Maybe...
# Appendix F

Sample rubric used

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NYT summary</strong></td>
<td></td>
</tr>
<tr>
<td>exemplary</td>
<td>Rating</td>
</tr>
<tr>
<td>own words, substantive summary</td>
<td></td>
</tr>
<tr>
<td>but not overly detailed</td>
<td></td>
</tr>
<tr>
<td><strong>research article summary</strong></td>
<td></td>
</tr>
<tr>
<td>exemplary</td>
<td>Rating</td>
</tr>
<tr>
<td>own words, substantive summary</td>
<td></td>
</tr>
<tr>
<td>but not overly detailed include</td>
<td></td>
</tr>
<tr>
<td>key aspects of study including</td>
<td></td>
</tr>
<tr>
<td>design, methods and results</td>
<td></td>
</tr>
<tr>
<td><strong>application</strong></td>
<td></td>
</tr>
<tr>
<td>exemplary</td>
<td>Rating</td>
</tr>
<tr>
<td>cite test/course materials</td>
<td></td>
</tr>
<tr>
<td>answered questions</td>
<td></td>
</tr>
<tr>
<td>comprehensively</td>
<td></td>
</tr>
<tr>
<td>competent</td>
<td>Rating</td>
</tr>
<tr>
<td>not all appropriate material</td>
<td></td>
</tr>
<tr>
<td>referenced but had majority</td>
<td></td>
</tr>
<tr>
<td>not comprehensive</td>
<td></td>
</tr>
<tr>
<td>developing</td>
<td>Rating</td>
</tr>
<tr>
<td>little references to course</td>
<td></td>
</tr>
<tr>
<td>material</td>
<td></td>
</tr>
<tr>
<td>failed to answer all questions</td>
<td></td>
</tr>
<tr>
<td>or not enough effort shown</td>
<td></td>
</tr>
<tr>
<td>No Marks</td>
<td></td>
</tr>
<tr>
<td>missing responses, no references</td>
<td></td>
</tr>
<tr>
<td>to course materials</td>
<td></td>
</tr>
<tr>
<td><strong>analysis</strong></td>
<td></td>
</tr>
<tr>
<td>exemplary</td>
<td>Rating</td>
</tr>
<tr>
<td>fully examined conceptions</td>
<td></td>
</tr>
<tr>
<td>super minor or no inaccuracies</td>
<td></td>
</tr>
<tr>
<td>strong connections to validity,</td>
<td></td>
</tr>
<tr>
<td>accuracy and relevance</td>
<td></td>
</tr>
<tr>
<td>competent</td>
<td>Rating</td>
</tr>
<tr>
<td>examined conceptions</td>
<td></td>
</tr>
<tr>
<td>minor inaccuracies,</td>
<td></td>
</tr>
<tr>
<td>connected to validity,</td>
<td></td>
</tr>
<tr>
<td>accuracy and relevance</td>
<td></td>
</tr>
<tr>
<td>developing</td>
<td>Rating</td>
</tr>
<tr>
<td>limited response in analysis</td>
<td></td>
</tr>
<tr>
<td>with limited connections to</td>
<td></td>
</tr>
<tr>
<td>validity, accuracy and relevance</td>
<td></td>
</tr>
<tr>
<td>inaccuracies that interfere</td>
<td></td>
</tr>
<tr>
<td>with meaning</td>
<td></td>
</tr>
<tr>
<td>poor</td>
<td>Rating</td>
</tr>
<tr>
<td>limited</td>
<td></td>
</tr>
<tr>
<td>to no analysis</td>
<td></td>
</tr>
<tr>
<td><strong>evaluation (Q 6 &amp; 7)</strong></td>
<td></td>
</tr>
<tr>
<td>exemplary</td>
<td>Rating</td>
</tr>
<tr>
<td>connect with to summaries and</td>
<td></td>
</tr>
<tr>
<td>recommendations, clear</td>
<td></td>
</tr>
<tr>
<td>understanding of data</td>
<td></td>
</tr>
<tr>
<td>competent</td>
<td>Rating</td>
</tr>
<tr>
<td>some recommendations,</td>
<td></td>
</tr>
<tr>
<td>reasonable understanding of data</td>
<td></td>
</tr>
<tr>
<td>developing</td>
<td>Rating</td>
</tr>
<tr>
<td>little or no recommendations</td>
<td></td>
</tr>
<tr>
<td>and/or less clear understanding</td>
<td></td>
</tr>
<tr>
<td>of data</td>
<td></td>
</tr>
<tr>
<td>poor</td>
<td>Rating</td>
</tr>
<tr>
<td>missing evaluative comments</td>
<td></td>
</tr>
<tr>
<td><strong>APA and writing</strong></td>
<td></td>
</tr>
<tr>
<td>exemplary</td>
<td>Rating</td>
</tr>
<tr>
<td>follow APA, 0-1 errors,</td>
<td></td>
</tr>
<tr>
<td>appropriately cite using APA,</td>
<td></td>
</tr>
<tr>
<td>have complete reference section,</td>
<td></td>
</tr>
<tr>
<td>logical, sophisticated, organized</td>
<td></td>
</tr>
</tbody>
</table>
Appendix G
Example analysis section of student paper

Analysis and Interpretation of the Data (Procrastination option by Wohl et al)

Two of the outcome measures from the study were of positive and negative affect, and procrastination before the second exam, using separate MMRs to analyze both. In the MMRs, beta (β) represents the slope of the line and how close the values are to the regression line when plotted (Aron et al, 2013).

To begin with, negative affect was regressed on procrastination level for the first midterm, self-forgiveness for that procrastination, and the interaction between the two, meaning they used negative affect for the independent/y variable, against procrastination and self-forgiveness as dependent/x variables, and then measured the interaction between the two using a graph. The main effect (the effect the independent variable has on the dependent variable) between of procrastination on the first exam and self-forgiveness predicted a negative affect with a correlation of β=.22 at p=.05 and β=.41 at p<.001, meaning the correlation between the procrastination and negative affect is positive and weak, though still there at a significant level, while the correlation between self-forgiveness and negative affect was negative and of moderate strength, at a very significant level. The interaction between the two was also significant at p=.05 with β=.22 which is a negative and weak correlation, meaning that then when self-forgiveness increased, on average, procrastination decreased.

Procrastination before the second exam underwent a similar analysis using an MMR. The main effects for procrastination before the first exam and self-forgiveness for that procrastination predicted procrastination before the next exam, with a correlation of β=.59 at p<.001 and β=.20 at p=.02, meaning that procrastination for the first exam correlated moderately positively with procrastination on the second exam at a very high significance level, and self-forgiveness correlated weakly negatively to procrastination on the second exam at a high significance level. The interaction between those independent variables was also significant and had a β=.23 at p=.02, meaning that there was a weak negative correlation between the two, indicating that as self-forgiveness increased, procrastination on the second exam lessened, though this was not significant one standard deviated below the mean for procrastination prior to the first exam.
Connecting the results to the study hypothesis and critically looking at sampling, results, and implications

The researchers’ hypothesis before the study was that “self-forgiveness would interact with procrastination prior to the first midterm to predict lower levels of negative affect, and this in turn would predict lower levels of procrastination for the second midterm.” They found support for their hypothesis and were able to reject the null hypothesis that self-forgiveness and negative affect would have no effect on procrastination on the second exam compared to the first. Their sample was taken from one section of an intro psychology class in college, because they noted college students are known for the procrastinatory behaviors. The number of students who completed both sessions of the study were 134, well above the minimum 30 required for statistical research. However, since their sample was only made up of psychology students, there could be some problems there because perhaps psychology students are more or less inclined to procrastinate than other students or introspect better than other students because of the nature of their curriculum. The study would have been better if it at least included more than one type of class in its sample. While this issue is not specifically addressed in their limitations section, potential confounding caused by the fact that the students were all first-year was, in addition to the fact that they only looked at one type of academic task, a multiple-choice test, and other potential unknown variables effecting procrastination behaviors that aren’t self-forgiveness.
Chapter 8

Statistics and Writing Together at Last: Dual Skills of the Critical Thinker

Connor P. Principe, PhD
Pacific University

Summary

Research has repeatedly demonstrated that writing assignments increase the critical thinking skills of students (e.g., Kuh, 2008; Quitadamo & Kurtz, 2007; Stephenson & Sadler-McKnight, 2016); however, instructors may be reluctant to include time-intensive writing assignments in quantitative courses (e.g., Bahls, 2012). In this chapter, I argue that students and faculty create a false dichotomy when writing and statistics are viewed as separate skills. Although initially daunting for students, those who learn to view both verbal and quantitative analyses as tools of critical thought (and not as separate tools of language and math, respectively) are more engaged and often arrive at a deeper understanding and appreciation of statistical material. A majority of this chapter describes some of the activities and assignments of a writing-intensive statistics course that have helped students to learn, better understand, and (in some cases) even like statistics. Despite their variety, all assignments attempt to fulfill three objectives of critical thinking in line with the GAISE goals of the American Statistical Association. 1) The assignments teach statistical thinking by emphasizing skill building (i.e., learning how to think about data) over knowledge acquirement (e.g., memorizing formulas). 2) Active learning is fostered when students collaborate to learn by working in pairs and small groups. Writing activities are frequently peer-reviewed, and analyses are agreed upon before they are conducted. 3) As analytical skills develop, instructor scaffolding is slowly removed, and students are given the opportunity to integrate real data with a context and a purpose, thereby engaging in meta-cognition (i.e., thinking about their own learning). The chapter closes with advice for faculty seeking to include similar assignments in their classrooms.

Why You Should Include Writing in Your Statistics Course (Even Though You Don't Want To)

Writing in the Discipline: A Personal History

Several semesters ago, I found myself as a participant in a writing-in-the discipline (WID) workshop provided by my university. The goal of the workshop was to provide faculty with resources, ideas, and advice regarding how to implement our new college-wide curriculum that would require all graduating students to have writing-intensive coursework both inside and outside of the major. Inspired, I began to plan writing assignments and other activities for my developmental psychology course to encourage critical thinking such as peer review, journal keeping, and online discussion groups. Knowing me as my department’s statistics professor, a colleague from computer science asked me whether I had considered incorporating any of these techniques into my Behavioral Statistics course. My reply came faster than she could finish her question.

No.
Actually, my reply was preceded by an expletive and accompanied by what I can only imagine was a look of combined horror and disgust. What a question to ask! No, I would not be integrating additional writing assignments into a statistics course. I explained to my colleague that (perhaps unlike her students), most of my social science students enter statistics terrified that it will require mathematical reasoning and computation. Often, students tell me that they changed their major or career path to avoid courses that feature math. I roughly estimate that only 50% of the energy I devote toward statistics takes the form of actual teaching; the other 50% is directed toward soothing frayed nerves, offering pep talks, and reminding students that statistics are an important tool of deductive thought, not merely arcane numbers in a results section. Adding more writing to this mix seemed like madness. So, no: I wouldn’t be doing that.

Perhaps it was because I was so horrified of the idea that I couldn’t stop thinking about it. Maybe it was the fear of changing a well-established course. In fact, I consider myself to be a skilled statistics professor: Not only have (most of) my students survived, a few have even thrived, discovering a love for data analysis and acceptance into graduate quantitative fields. Nevertheless, compared with other social science courses, quantitative courses are typically associated with lower course evaluations (e.g., Uttl & Smibert, 2017) and therefore likely include higher percentages of frustrated, disinterested students. Usually, my most favorable statistics course evaluations usually took some form of “It wasn’t as bad as I thought it would be.” Few social science students, if any, look forward to taking statistics. Therefore, any change to such a course would be high risk.

As with many things that are high risk, however, a successful change would likely yield a high reward, and deep down, I knew the course could be improved. In particular, I was haunted by the idea that many students might be passing statistics without having a firm sense that they truly understood the material on a fundamental level. Occasionally, a student would comment that although they enjoyed the class and could now use various forms of statistical software, they still felt like they did not “get” statistics. These comments ate at me, so I set out to redesign my course to ensure that students really “got” statistics. Memories of my writing-in-the-discipline workshop convinced me that I needed them to engage in critical statistical thinking; and for that to happen, I needed them to write more than I had in the past.

Writing and Critical Thinking within a Statistics Course

Having taught at a small private liberal arts college for nearly a decade and having attended a 4-year peer institution as an undergraduate student (many) years before that, I am easily convinced by the argument increasingly made by university faculty, administrators, and accreditors that students need more writing instruction. Scholarship of teaching and learning (SoTL) research has repeatedly demonstrated that writing assignments increase the critical thinking skills of students (e.g., Kuh, 2008; Quitadamo & Kurtz, 2007; Stephenson & Sadler-McKnight, 2016). In fact, the connection between critical thinking and writing is so well accepted that hundreds of universities have adopted or are currently adopting writing-across-the-curricula (WAC) and WID courses, and several excellent texts have been written (e.g., Bean, 2011; Bahls, 2012) encouraging the integration of more writing into all college courses.

Writing fosters and improves critical thinking in a multitude of ways; for the sake of brevity, however, this chapter focuses on how I have used writing to foster critical thinking in three ways that are particularly relevant to my statistics course. Importantly, although my course typically includes 15-20 sophomore-level psychology majors, most of the techniques listed below apply to any college-level course.

After describing these assignments, I discuss how they have been received by students as well as aspects of the assignments that have worked and others that remain under construction.
Ultimately, I provide advice for faculty members, ranging from those who seek to create writing-intensive statistics courses to those merely looking to incorporate one or two new assignments into an existing unit. Admittedly, introducing or increasing the amount of writing in a course is not without challenge, but it need not be backbreaking. In fact, I have found the improvement in student outcomes to be well worth the effort.

**Examples of Low- and High-stakes Statistical Writing Assignments**

**Low-stakes Labs**

I have found that one of the most effective ways to encourage critical thinking with regard to data analysis is through the use of low-stakes writing assignments administered once or twice each week throughout the semester (I refer to these assignments as “labs”). In most cases, a full class period is devoted to these labs, although the assignments are posted online at least 24 hours before the class periods and due 1 week after the class period. Thus, between 1/3 and 1/2 of all in-class time is devoted to these assignments. All labs are open-book and intended as group assignments. Although I usually teach statistics in a classroom computer lab, I encourage students to bring their own laptops if they contain their class notes. Nearly all students do so; that, combined with the fact that I use the readily downloadable and free software R to teach data analysis (see Chapter 18), frees us from requiring a specific on-campus space. Therefore, the course design emphasizes to students that they are in class to learn thinking skills, rather than memorize facts and formulas. While the students are engaged in the assignment, I typically patrol the room offering my thoughts on their ideas and answering questions. However, I explicitly instruct the students to first think about how to answer each question out loud with each other before they ask me for guidance.

The assignments themselves combine the data manipulation and interpretation tasks common across statistics courses with brief writing-to-learn (WTL) tasks. Students are given a dataset (either publicly sourced or one that I have collected over the years). Students must manipulate and visualize the data to complete the assignment; however, emphasis is placed on the meaning of the numbers via the WTL questions (i.e., gaining numeracy skill vs. learning rote calculation steps). Importantly, I never use—and am opposed to using—the sort of neatly manicured datasets often found in introductory statistics texts because can lead students to believe that statistical solutions are synonymous with straightforward conclusions. In particular, I favor the hundreds of datasets archived by the University of Florida taken or simulated based on summary statistics from the experiments with dependent variables such as number of hen egg pores, UFO sightings in the US, and average florescent lamp life (see: [http://users.stat.ufl.edu/~winner/datasets.html](http://users.stat.ufl.edu/~winner/datasets.html)). I want students to learn that real data are messy and often lead to ambiguous results. Because I want students to become comfortable with ambiguity and shades of gray, I try not to pull punches: Nearly every item on the lab requires that students engage in some sort of critical thinking. After each lab is due, I post a key to the quiz that include full explanations for each answer. Some answers include walking through various calculations, whereas others include my musings on my own difficult questions. Specifically, I want students to see me struggle with my own analytical decisions. For example, “Does a 7-point Likert scale provide continuous or categorical data?” and “Under what circumstances is it permissible to remove an outlier from your dataset?” are not easy questions with easy answers. Over the years, I have learned that students actually gain confidence in both themselves and me as the instructor when they see that I occasionally struggle to put ideas into words as well.

Assignment 1 in the Appendix demonstrates the type of lab that I often use early in the semester. In this example, students are asked to code and interpret a sample of Beck
Depression Inventory (BDI) scores (in this case, I collected anonymous data in a prior experiment). In addition to asking students to discuss data distribution and shape, they must place their answers in the context of the world as they know it. For instance, after taking each participant’s total BDI score and labeling it with its clinical cut-off (i.e., 21-31 = “moderate depression”), students are asked a much broader question about the strengths and weaknesses of using continuous versus categorical scales to define mental illness. In addition, I typically include at least one “trick” question on all labs to emphasize the point that not all datasets can be used to answer all questions. For instance, after finding that the per capita wine consumption of California is significantly higher than that in Alaska, I might ask “A friend of yours becomes an alcohol abuse advocate. What do your data tell you about where she should direct future resources?” The answer of course is “nothing” (or at least very little that is likely to be directly relevant). Students in my classes learn, relatively quickly in most cases, that statistics are both a powerful tool and inherently limited by the data themselves.

In Example 2, students later in the semester use their improving critical thinking skills to conduct inferential statistics. One of the key analytical skills that I hope all of my students come away with is being able to correctly match research questions with their appropriate analyses (i.e., the primary skill of a statistical consultant). Thus, in the second example, I ask students to turn research questions into null hypotheses and apply the correct statistical technique. Crucially, I never tell students what test they should conduct (e.g., “Please perform an independent-samples t-test…”)—even though they would love to be told this information—because no one will ever give them those instructions in the real world. Rather, students learn to think like detectives, using the question, the dataset, and their gained statistical knowledge to piece together their analytical plan. Moreover, I have students write why they chose to use a particular test. For example, a student answering Question 1 in Example 2 might reply, “I should conduct a repeated-measures ANOVA because 16 dogs were submitted to four conditions over time (four trials). The outcome is mean cortisol level. Repeated-measures ANOVA allows me to look at four dependent means over time. Because I plan to make two post hoc (paired-samples t) tests, I will apply the Bonferroni correction to control for Type I error inflation. Only p-values = .05/3 will be considered as significant.” Questions and answers such as these allow the professor to walk through the thought process of the student, and they allow the students to reflect on how much they have learned up to that point in the course.

High-stakes Final Project

In addition to the low-stakes writing assignments discussed above, I include at least one high-stakes writing project to emphasize to students the value of writing and data analysis as tools of the sort of critical thinking that they will need for the careers that they will one day have. This project builds upon previous low-stakes assignments in that it also asks students to match research questions and hypotheses with the appropriate analytical technique; however, it also expands on previous low-stakes assignments by offering far more freedom. In the example provided in the Appendix, students are given a very large dataset. In longer (i.e., two-semester) and/or more advanced courses (i.e., senior capstone), I have students create their own questions and collect their own data; in briefer or introductory courses, I collect and provide these data myself. Using the data, students write their own hypotheses that can be tested by applying the statistical techniques covered throughout the semester. In these briefer courses, I typically require that students write three hypotheses and test each one using a different analysis. Furthermore, I group the analyses by level of complexity (that is, when it is introduced in the course according to the syllabus). Thus, in the example that I have provided, a student might use a t-test for the first hypothesis, a chi-square test for the second, and a ANOVA with a post hoc test for the third. Importantly, for each hypothesis, students write miniature data analysis, results, and discussion sections that focus on their analytical decision making.
Students usually receive this assignment halfway through the semester; although they have not covered all of the necessary course material yet, they have enough knowledge and confidence to begin this project.

To relieve stress and improve the quality of students’ final submissions, I set aside class time (anywhere from 30 min to the full period, depending on students’ needs) for peer review. I also use this strategy to appropriately pace my students and avoid procrastination. Specifically, I assign students low-stakes mini outcomes throughout the second half of the course so that they work on one question at a time. During the initial peer review sessions, student might only seek to confirm with each other that their research question is clear and that their analysis plan makes sense; at a second peer-review session, students provide feedback on each other’s writing via response-centered reviews (see Bean, 2011). Because I have found that student often (mis)interpret “peer review” as evaluative as though they were assigning grades, I provide guidance in the form of worksheets (see Appendix) that walk each student through the specific aspects of their writing that must be clear to receive a high score. After this session, I ask students to turn in an edited (but not final) draft upon which I return feedback no more than 7 days later. Ultimately, students submit all three of their completed research questions as a final project feeling confident because their papers contain few elements that their peers and professor have not already read and approved. Furthermore, students understand exactly how they will be scored because they are provided with a rubric for all high-stakes writing assignments (see Appendix for an example). Lastly, I often have students present one of their analyses in the form of a brief oral defense; although the idea is initially terrifying to students, they have received so much feedback from their peers and their professor by that point that their presentations become more of a celebratory act than the nightmare they might have imagined. Importantly, the oral defense also provides me with one last opportunity to assess students’ critical thinking and their understanding of data analysis.

**Student Work and Reaction**

After implementing low- and high-stakes writing assignments in my Behavioral Statistics course, I have found that both writing and statistical performance have improved. With regard to the former, this result is no surprise, as much literature has shown that any student who practices writing and receives feedback from peers and instructors, particularly within a discipline, will improve over time (e.g., Brownell, Price & Steinman, 2013; Hohenshell & Hand, 2006; Pelaez, 2002). With regard to the latter, I have found that performance has improved even on assessments that do not employ writing-based inquiry (i.e., multiple-choice and computational exams) compared with previous courses that did not include writing. Furthermore, short-answer assessments have revealed better understandings of classical and experimental probability, the differences between descriptive and inferential techniques, and the limits of what statistics can tell a researcher. In addition, course evaluations have revealed that students are aware of the amount of knowledge and skills that they have learned; in many cases, students describe knowing next to nothing at the beginning of the course but feeling confident in their abilities to analyze data, read and understand results sections in their field, and explain the results of a study to those with less statistical knowledge. I have also noticed a marked decline in the number in-class questions that begin “What’s the point of…” since I began to use writing in my course.

Regarding high-stakes assignments, I have found that students become deeply invested in “choose-your-own-adventure”-style writing assignments because they allow them to flex their creativity muscles and show off what they have learned. Furthermore, most students (especially those who have played a role in data collection or survey design) end the class with the feeling of having accomplished something important and meaningful. Thus, although writing in statistics
may at times feel like (and actually be) more work than statistics without writing, assignments that emphasize critical thinking and decision making relieve some of the pressure that math-phobic students have of being either right or wrong.

Recently, I compared the student evaluations of two statistics courses that I taught during the same semester: a writing-intensive course and course where the writing was limited to a few short-answer questions on tests. With regard to the item “Rate your increase in skills/understanding as a result of taking this course”, those who took the writing-light course provided a mean rating of 5.52 out of 7 (n = 21, Mdn = 5, SD = 1.21); those who took the writing-heavy course provided a mean rating of 6.4 (n = 15, Mdn = 7, SD = 0.74). Because these courses also differed with regard to student composition (i.e., general education vs. majors) and class size, I hesitate to give all of the credit to writing instruction; however, these data provide preliminary support for my anecdotal conclusions above.

It should be noted, however, that not all students respond to this class with love. Although the overwhelming tone of student course evaluations (both formal and informal) has been one of appreciation, many students comment on the perceived high workload of combining data analysis with writing. Students report understanding that data analysis and writing are both tools of critical thinking, but many (perhaps most) state that they must work extremely hard to keep up with the material given that they are constantly engaged throughout the course. Several students have noted that they put in significant hours between classes to ensure that they know what questions to ask in class and to be prepared for peer-review assignments. A minority of students have expressed frustration that this course demands so much time given their busy loads (e.g., 3-4 additional courses, athletics, work, and family life). Nevertheless, given the observable improvement in student performance and knowledge base, I am confident that incorporating writing into statistics has been a success.

Implementing Similar Assignments in Your Course

Given these modest successes, I am happy to have added writing to all of my Behavioral Statistics assignments; however, I am occasionally reminded of why I was so reluctant to consider the idea in the first place. First, writing can be time-intensive, particularly if one is not used to providing meaningful feedback. Earlier in my career, I found myself editing my students’ writing as I would my own. That is, I spent an inordinate amount of time commenting on and editing clarity of thought via sentence construction and diction. This method takes a tremendous amount of time (dozens of hours in some cases) and ends up doing very little to actually provide student with what they need. However, more writing need not be more work. My strongest piece of advice, then, is to increase the amount of student writing peer review. In particular, the sort of guided peer review instructions that I provide to students (see Appendix) has been a tremendous time-saver with regard to grading student responses. Importantly, the peer-review that I want student to apply is highly delimited. The general instructions, “Peer review each other’s work”, are not helpful for students. However, specific instructions such as “Identify the author’s research question” is because students are often shocked to learn how difficult this task is in their peers’ (poor) writing and how embarrassing it is to watch their peers struggle with their own. Although peer review requires additional time in class, I have found it to cut down on the time I spend providing comments on writing, in some cases dramatically. In fact, in many cases, grading writing becomes a simple matter of following the rubric.

Following peer review, I try to limit myself to commenting on only 2-3 major points (e.g., Does the student clearly describe what the required scatterplot shows? Does the student correctly identify the influence of outliers?), even when additional errors are present. I have found that
when I correct everything that I can find, students are less likely to use critical thought (i.e., use the autocorrect feature of their word processing software and turn in essentially the same draft a second time). Providing only a few—but meaningful—comments or questions helps students not to become overwhelmed and encourages them to critically think about what their answer might be missing. Furthermore, I find that limiting my comments in papers helps remind students what I tell them all the time in class: Writing is a process, not an outcome; like thought, it (hopefully) does not stop just because something has been in.

If you are interested in attempting something similar to my high-stakes assignment, I strongly recommend having students collect their own data. Although I have allowed students to use archived datasets such as those employed for my low-stakes assignments to some degree of success, I worry that students are also learning that it is acceptable to look at data before one creates testable hypotheses. I spend a great deal of time explaining to students that deductive reasoning means creating a theory before it can be tested but that class limitations engender a more expedient solution (i.e., do as I say and not as I do). I have only allowed students to collect their own data over full semester (or longer) courses. I would still employ a similar assignment but use archived datasets for partial semester or online courses.

Lastly, I recognize that many aspects of a writing-intensive course are likely to work best under “ideal” classroom conditions (e.g., small faculty-to-student ratio) and less so under others (e.g., large classes, online courses). Nevertheless, a course need not be writing intensive to employ writing in an effective manner. Past research across disciplines, including the quantitative, physical, and social sciences, as well as my own experience have taught me that writing not only has a place in the statistics curricula, it advances the critical thinking of students in ways that perhaps other techniques cannot.
References


Appendix

Sample Assignment #1

Levels of Measurement Lab

Please refer to the BDI dataset on Moodle. As always, these data come from a real study and were provided by real (but anonymous) participants. Work with your classmates to think through all of your answers. Upload your completed lab by Friday at 8:00 am.

1) According to Beck, each item on the BDI should be scored from 0-3, where higher scores reflect MORE depression. Based on the KEY on the Excel file, do any of items on the BDI need to be reversed scored? Explain your decision.

2) Recode the BDI (i.e., ensure that all response values range from 0-3 and the appropriate questions are reverse scored), sum the values for each participant, and report the number of people who could be classified as

<table>
<thead>
<tr>
<th>Range</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-10</td>
<td>These ups and downs are considered normal</td>
</tr>
<tr>
<td>11-16</td>
<td>Mild mood disturbance</td>
</tr>
<tr>
<td>17-20</td>
<td>Borderline clinical depression</td>
</tr>
<tr>
<td>21-30</td>
<td>Moderate depression</td>
</tr>
<tr>
<td>31-40</td>
<td>Severe depression</td>
</tr>
<tr>
<td>&gt; 40</td>
<td>Extreme depression</td>
</tr>
</tbody>
</table>

Given what you know about statistics (so far!) and what you know about the real world, comment on the appropriateness of using the above cut-off values to define disparate groups of depression severity. What are the strengths and weaknesses of this approach to defining mental health?

3) Use R to create a histogram of the total depression score. Is the data normally distributed, uniformly distributed, positively skewed, or negatively skewed? Interpret what the shape of this distribution means in the context of these participants. Do you think that these participants represent a normative population or a clinical population? Explain your answer.

4) Argue for or against using the BDI as a valid and reliable measure of depression. What do the data tell you about this scale’s validity and reliability?

5) Are the 21 items of the BDI continuous or categorical?
Sample Assignment #2

Repeated-Measures ANOVA Lab

Refer to potentiallyexplodingdogs.xlsx to answer the questions below. These real data come from Sherman, B. L., Gruen, M. E., Case, B. C., Foster, M. L., Fish, R. E., Lazarowski, L., ... Dorman, D.C. (2015). A test for the evaluation of emotional reactivity in Labrador retrievers used for explosives detection. Journal of Veterinary Behavior, 10, 94-102. The data have been modified slightly to enable a repeated-measures ANOVA. Please see the key in the Excel file to interpret the variables.

1. Cortisol level is used to measure physiological stress. Does the stress of detecting bombs significantly increase the cortisol levels of bomb-sniffing Labradors across four trials, after accounting for individual differences? (Note that you plan to make two post hoc comparisons: Time 1 vs. Time 2 and Time 3 vs. Time 4).
   a. Write null and alternative hypotheses.
   b. What test(s) should you conduct? Explain why you believe these statistics best test your hypotheses.
   c. Conduct the relevant test and report your results in APA format.
   d. Calculate effect size(s) for the above analyses and interpret them.
   e. Make a decision about your null hypothesis and report your conclusion.
   f. Graph your results. In no more than three sentences, explain what your plot shows to a person with no knowledge of statistics.
   g. If the result of your omnibus test was significant and you have the necessary power, perform your planned post hoc comparisons. Report your results in APA style and report your conclusion.
   h. List any aspect of your data or analytical decisions that might bias your conclusions.

2. Is there is difference in the average open-field anxiety scores between black and yellow Labs?
   a. Write null and alternative hypotheses.
   b. What test(s) should you conduct? Explain why you believe these statistics best test your hypotheses.
   c. Conduct the relevant test and report your results (including effect size) in APA style.
   d. Make a decision about your null hypothesis and report your conclusion.
   e. Graph your results. In no more than three sentences, explain what your plot shows to a person with no knowledge of statistics.
   f. List any aspect of your data or analytical decisions that might bias your conclusions.
Sample Assignment #3

Choose-your-own-adventure Analysis

Refer to FinalSurvey.xlsx; be sure to use the key on the second worksheet to interpret the variables and verify coding. This dataset includes the data compiled from the questions that have been asked throughout the course by myself and your peers. In total, there are 360 variables (!).

1. Create a research question related to your topic of interest that can be answered using one of the following tests: independent-samples t-test, one-way ANOVA, or simple regression.

2. Create a research question related to your topic of interest that can be answered using one of the following tests: chi-square (either goodness of fit or test of independence), ANCOVA, or logistic regression.

3. Create a research question related to your topic of interest that can be answered using one of the following tests: factorial ANOVA (with one post hoc test) or repeated-measures ANOVA (with one post hoc test).

For each question above, write succinct Data Analysis, Results, and Discussion sections (no more than two pages each). Students receiving the highest scores will clearly explain why their analysis is appropriate for their research question, provide all relevant results regardless of significance, and describe the limitations of their conclusions.

Remember that you will orally defend one of these three analyses in class (5 minutes of presentation, 5 minutes of questions from your peers). The students receiving the highest scores on their oral defense will demonstrate that they have clearly thought through and can justify their analytical decisions.
Peer review guide for Results and Discussion section(s).

You have been randomly assigned to a group of three or four students, and you have each received copies of your rough drafts. Be sure to have read each draft before class begins. When class begins, follow these steps.

Decide whose paper will be reviewed first. Each “review” will actually be an informal conversation between/among the non-authors. The non-authors will attempt to answer the following questions based only on what they have read.

1) What was the author’s research question?
2) What statistical analyses were employed to answer this question?
3) Were the statistical analyses appropriate for this question?
4) What were the results?
5) Are the results surprising or unexpected?
6) Are the results explained in the discussion section?
7) Do you think there is a better way to ask the research question or analyze the data?
8) Did the author provide convincing explanations for their findings (or lack of findings)?
9) Are the future directions appropriate for this study?
10) Are the graphs/tables appropriate? Do they help people make sense of the results?

During this conversation, the author will listen and take notes. Only when these ten questions have been discussed may the author include themselves in the conversation. At this point, the author may ask questions of their reviewers about ways to help improve their paper with regard to the items above.

Note that you are not evaluating quality of writing directly, nor are you scoring the paper in any way.

Repeat the process until all papers have been reviewed.

NEXT STEP: Use your notes from this review process and conversation to continue to write and improve your results and discussion section. A revised draft is due end of day Monday April 15.
# PSY 301 Results/Discussion Grading Rubric

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Unacceptable (≤ D)</th>
<th>Acceptable (~ C)</th>
<th>Good (~ B)</th>
<th>Excellent (~ A)</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyses</td>
<td>Incorrect or inappropriate analyses. Analyses do not match RQ.</td>
<td>Correct but limited analyses. Analyses match RQ.</td>
<td>Correct and complete analyses. Analyses match RQ.</td>
<td>Thoughtful and complete analyses. Analyses match RQ.</td>
<td>30</td>
</tr>
<tr>
<td>Statistics</td>
<td>Statistics are missing or incorrectly reported.</td>
<td>Statistics are presented but incorrectly reported.</td>
<td>Statistics are presented and correctly reported. Occasionally provides too much data.</td>
<td>Statistics add important detail to the description of the analyses but do not overwhelm.</td>
<td>30</td>
</tr>
<tr>
<td>Discussion</td>
<td>Results are not discussed, merely restated.</td>
<td>Results are described but not placed in context of real world.</td>
<td>Results are described and placed in context of real world.</td>
<td>Results are described as making a contribution to existing knowledge.</td>
<td>20</td>
</tr>
<tr>
<td>Grammar &amp; Mechanics</td>
<td>Grammatical/spelling/punctuation errors substantially detract from section.</td>
<td>Very few grammatical errors interfere with reading the section.</td>
<td>Grammatical errors are rare and do not detract from the section.</td>
<td>The section is free of grammatical errors.</td>
<td>10</td>
</tr>
<tr>
<td>APA Style &amp; Writing Conventions</td>
<td>Errors in APA style detract from the paper. Word choice is informal. Writing is choppy with many unclear passages. Headings are incorrect.</td>
<td>Errors in APA style are noticeable (i.e., wordiness, lack of brevity). Word choice is occasionally informal. Writing has multiple awkward or unclear passages.</td>
<td>Rare APA errors that do not detract from the paper. Scholarly writing style. Writing has minimal awkward or unclear passages. Headings are correct.</td>
<td>No APA errors. Writing flows and is easy to follow. Headings are correct.</td>
<td>5</td>
</tr>
</tbody>
</table>
Part Three: Learning and Engagement Activities
The Important Yet Difficult Task of Making Statistics Engaging

Janet M. Peters, PhD
Washington State University – Tri-Cities

Summary
Statistics can be a high stress, high demands course for students and faculty alike. One way to improve engagement is to create well-structured assignments that incorporate fun, personal relevance, and professional development. In Part 1, I start with simple, low investment ideas as ways to spark engagement through fun and intellectual curiosity. In Part 2, I share my pre-lecture activity assignments, which encourage engagement through student ownership and creativity. Finally, in Part 3, I discuss engagement through a semester-long service-learning project that develops job related skills.

The importance of student engagement
For the last several years, I've had the privilege of connecting with colleagues at teaching conferences, over social media, and in virtual committees. I’ve spent many social hours commiserating over the workload of teaching statistics. I’ve read countless blogs and social media posts that describe the anxiety of teaching a class that can sink student evaluation scores, which many faculty in teaching-focused positions depend on for raises and promotions. I’ve borrowed and shared more resources for my statistics courses than all my other courses combined. From all of these interactions, one thing has become clear: teaching statistics is hard work. However, despite all the difficulties associated with teaching the course, I’ve yet to hear an instructor doubt the importance of the subject matter.

Yet this mutually agreed upon importance among faculty does not seem to translate into student perceptions of the course. In my experience, many students are hesitant to take the course and often delay it as long as possible. As Principe (2020) points out in Chapter 8, faculty end up devoting a lot of energy to managing student perceptions of the course. Student concerns include anxiety over their performance in the class, stress about the workload, and skepticism regarding the utility of what they will learn. Thus, my goal for this chapter is to explore the ways in which we can make statistics more engaging for students so that they, too, can appreciate the crucial role statistics play in our personal and professional lives.

Considerations
As you read this chapter, there may be times when you disagree with me or find my ideas untenable for your class. You might have different interests, a different student population, different institutional support, different class sizes, or any number of other factors that influence the way we teach. That's okay. This chapter is not meant to be a manual for how people should teach their classes. Further, nothing in this chapter is proposed as a singular solution to addressing all your learning outcomes. Instead, it is intended to be an overview of ideas to help spark engagement in your own students. Thus, as you read this chapter, I implore you to keep in mind that student learning matters most. So, before you change a single homework problem or adopt a new project, make sure you know your desired student learning outcomes. All coursework, fun or not, should be derived from what students should be getting out of the class. Once you’ve established what they should be learning, then you can move towards making the process more engaging. That is, you can have the most interesting class in the world, but none
of that matters if you aren’t achieving your learning outcomes. Once you’ve built the framework, then you can add the depth and excitement discussed on this chapter.

Overview

The rest of this chapter is divided into three parts, dedicated to sharing my ideas, approaches, and lessons with you. In Part 1, I start with simple, low investment ideas as ways to spark engagement through fun and intellectual curiosity. In Part 2, I share my pre-lecture activity assignments, which encourage engagement through student ownership and creativity. Finally, in Part 3, I discuss engagement through a service-learning project that develops job related skills.

Part 1: Distinct, simple ideas for engaging content

Amusing examples, interesting worksheets, and relevant scenarios can help engage your students with course content as they work through homework and lab assignments. When I think about the easiest ways to engage students in my own statistics courses, I think of pop culture, interesting/relevant examples, and food. I wrote about many of these in a blogpost for the Graduate Student Teaching Association (GSTA) that was aimed at incorporating fun into statistics courses (Peters, 2019). In it, I describe different ways to make statistics fun. Here, I expand on some of those examples. These activities are typically the simplest ways for me to implement engaging content because I don’t have to commit to a large-scale project or semester-long approach. I can do one or two of these activities and leave it at that, or I can include them for each unit. That is, if you are interested in easily adding activities into your existing course structure, this section will help you achieve that goal.

Pop-culture

Your students live rich, full lives outside of the classroom. Many of them have interests in books, movies, podcasts, music, and more. During the semester, students will often informally discuss pop-culture events such as celebrity marriages or the finale of an iconic TV series. Just yesterday I heard two students bonding over a murder mystery podcast and discovering their shared interest in forensic psychology. You can leverage these moments and the popularity of pop culture events and characters to create interest in the topic you are covering. For example, to capitalize on the zeitgeist for comic book movies, you could make a superhero themed worksheet to help students practice the scales of measurement (nominal, ordinal, interval, and ratio). Practice questions might include "Tony Stark’s annual income" or "The degree of Loki’s mischievousness rated on a scale of 1 to 7." Both questions include elements from superhero lore, but answering the questions is not dependent upon that knowledge. That is, even if a student is unfamiliar with Tony Stark and/or Loki, it does not matter because they can still answer the question. The key is that your items should be inclusive and based on class content. Anyone should be able to answer the questions, even if they have never seen or heard of the pop culture reference you are making. Conversely, if students are familiar with the examples, the homework and assignments are more engaging. For a sample of pop culture examples, please see Table 1 (Note: I do not provide the data for these examples, they represent conceptual examples for which you can create your own data; see compendium at end of book for ideas on data development).

Table 1. Examples of using pop-culture to illustrate statistical concepts

<table>
<thead>
<tr>
<th>Concept/Content Area</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale of Measurement</td>
<td>Characteristics of Superheroes, Disney characters, or any other references you or your students might enjoy</td>
</tr>
</tbody>
</table>
### Comparing the success of the Harry Potter movie franchise to the Harry Potter book series

Comparing scores on different scales can be accomplished by comparing attributes of one fictional character to another:

- **Superhero**: Is the Joker more evil ($\mu = 12, sd = 2.00, X = 14$) than Batman is brave ($\mu = 100, sd = 15, X = 120$)?
- **The Wizard of Oz**: Is the lion braver than the scarecrow is intelligent?
- **The Office**: Is Dwight lower on agreeableness than Jim is on conscientiousness?

### Calculating probabilities with z-scores

- **Famous people**: If Beyoncé (musician) has a $z$-score of 2.07 on a scale of fabulousness, what is the likelihood that someone will be more fabulous than her?

### Related/Matched/Paired samples t-tests

- **Star Wars**: Leia is a general in the army and in charge of training soldiers. She hopes that her new training program will improve soldier performance. To assess this, she compares each soldier’s performance score from before and after the training.

### Independent samples t-tests

- **Friends (TV show)**: Ross has a job as a professor of archaeology at the local university. To assess whether or not his students are significantly more stressed than other students, he takes a sample of his own students and compares it to a random sample of students not enrolled in his archaeology class.

### Comparing Samples

- **Harry Potter**: At Hogwarts School of Witchcraft and Wizardry, teams earn points for good behavior and lose points for bad behavior. Did Gryffindor earn significantly more points than the other three teams?
- **Hunger Games**: Imagine that a country is divided into 12 different districts. Some of the districts are rich and well-educated, while others experience poverty and debt. If we were to randomly sample 100 people from each district, would there be a significant difference in annual income?

### Correlation

- **Schitt’s Creek (TV)**: Rose Apothecary, a new general store, advertises itself as a boutique shop and carries some very expensive items. David, the owner, uses the results from one of his customer satisfaction surveys to assess the relationship between item price and customer perceptions of item quality ($r = .38$). Interpret the strength and direction of this relationship.
- **Game of Thrones (TV)**: In watching the hit TV series Game of Thrones, you notice that the longer someone sits on the throne, the more corrupt they become. Identify the direction of this relationship.
Interesting & relevant examples

If you find it difficult to include pop culture, you can use other fun, quirky examples to engage your students. For example, some of my institution’s more interesting historical research is that of Dr. Grover Krantz, an anthropologist who dedicated most of his career to exploring the existence of Bigfoot. As a fun example to learn about z-scores, we analyze data compiled by the Bigfoot Field Researchers Organization (this is real data from a real organization). We use the total number of Bigfoot sightings by county and calculate z-scores, including the number of Bigfoot sightings in the county in which our university is located. We look at other counties that have far higher z-scores and discuss what that means for Bigfoot sightings. Students find this example engaging because it is a fun way to learn about z-scores and provides some historical trivia about our university.

You might not be lucky enough to have a notable bigfootologist, but there are plenty of other examples from which you can draw upon. Below, I list some potential ideas you might consider.

- Discuss or analyze interesting research findings from your home institution that have made their way into popular press or been misrepresented in the media.
- Look into the history of your geographic region or identify famous people from the town or university. Find something fun or notable about your town and use it as the basis of an example or assignment.
- You might also use controversy from local events. For example, our town experiences some controversy over the mascot for one of the local high schools (their mascot is the “bombers”, so named because of the critical role our city played in the Manhattan Project and developing nuclear weapons). To use this as a teaching tool in statistics, we correlated perceptions of the mascot with age (correlation) and assessed for any differences in perception of the mascot between males and females (independent samples t-test).
- Create examples based on rivalries with other schools, such as comparing school spirit, graduation rates between the universities, etc.
- Use publicly available data on your university’s homepage, such as demographic information, number of students, campus size, student to faculty ratio, etc.
- To provide more nuanced data, you can request internal institutional data such as graduation rates by major, GPA by college/department, etc.

The key to creating engaging work for your students is to find examples that are unique to you, your students, university, or community. If you’re having trouble finding examples that are directly related to your students, try finding examples that would be meaningful to your students. The abstract nature of statistics can lead students to perceive the course as not only difficult, but also detached from their everyday lives and professional future. Based on my experience, students understand why statistics are important to research, but rarely understand why statistics are (or should be) important to them personally and professionally. This presents an opportunity for us, as instructors, to create examples and assignments that foster a personal connection to the course concepts by demonstrating that statistics are a pervasive part of our lives – from the way companies use data in marketing to committing a Type I or Type II error in checking an infant’s diaper.

For example, to show my students how different markets exploit statistics to alter consumer perceptions and behavior, I use examples from the wedding industry, housing market, and job
websites. We use national data to examine student loan debt. Depending on where we are in the semester, we might use z-scores, t-tests, or correlations to analyze these data.

You can also use less serious examples. For homework assignments, I’ve been able to find user data on companies such as Netflix, Facebook, and Tinder. The degree to which students use or are familiar with these companies varies, but they provide real data that comes from the type of news articles the average consumer is likely to encounter. See Table 2 (below) for examples of how you might use these applications in your own class.

<table>
<thead>
<tr>
<th>Concept/Content Area</th>
<th>Examples</th>
</tr>
</thead>
</table>
| z-scores and probability | Regional Cost of Living Data  
  • Comparing scores on different scales can be done by comparing cost of living in your area to employment rates – in which metric does your city score better?  
  • We also use this data to calculate probabilities with z-scores. For example, we use real estate websites to pick a house and calculate its z-score. We then calculate the probability that a house selected at random is more or less expensive. As evidence of engagement, students sometimes want to look up their own house and calculate their z-score.  
  Average Temperatures  
  • Because my university is located in a city which everyone assumes gets as much rainfall as Seattle, we use data from the National Weather Service to calculate our z-score for annual rainfall and compare it to Seattle, the state average, and the national average.  
  Graduate school and the GRE  
  • By doing some online digging, you can find out the mean and standard deviation of the GRE. You can also find this information by major. We use this information to compare samples of students to the national average and discipline specific averages. We also look up the scores preferred by masters and doctoral programs and calculate probabilities. I try to pick programs suggested by the students to make it more personal.  
  Student Debt  
  • A salient concern for college students, we look up the average student loan debt for universities. Students then compare the average cost of our university and have the option to calculate their own z-score if they are curious how their debt compares to the national average. |
| t-tests | One sample t-tests  
  • Netflix, Facebook, and Tinder all have published data that can be realistic examples for students to use. These make for great examples because companies almost never publish the standard deviation. For example, in 2015 Netflix released company data that suggested the average subscriber views 93 minutes of content each day, which we use to compare to our class sample data. We also use this as an opportunity to discuss confounding variables in data collection (i.e., are average Netflix subscribers actually watching 93 minutes of television each day? What else might be happening?) |
Food service in almost any franchise restaurant. A former student gave an example of needing to measure six ounces of meat for each burrito; another recounted having to learn how to scoop three ounces of ice cream for each serving size. We used those student examples to create practice problems in class.

Regional Cost of Living Data
- Since our geographic area is comprised of three distinct cities, we use real estate websites to compare the cost of housing in each city and analyze whether or not the difference is significant.
- I look up the state test scores for ten schools in each city’s school district and we compare test scores across the three cities.

Annual Salary (conceptual examples)
- I use salary data to compare the average salary for three different types of psychologists (school counselor, clinical psychologist, and organizational psychologist).
- I use the same dataset to show students that we can also compare income based on education level (bachelors, masters, and doctorate).
- If you are a member, the Society for Industrial and Organization Psychology (SIOP) compiles and publishes salary data from their members. I show students the difference in annual salary between academics, internal consultants, and external consultants. We could also compare salaries based upon which industry (tech, retail, etc.) or sector (government, non-profit, for-profit) in which the consultants complete their work.

Using food

If you have small class sizes or a budget for classroom expenses, permission to have food in the classroom, and an allergy-free classroom, food can be a fun way to teach students about statistics. For example, I’ve used M&Ms to demonstrate sampling with replacement and Skittle flavors to demonstrate ANOVAs (see chapter 10 for M & M/Skittles worksheets). Once I used miracle berries to demonstrate paired t-tests (a miracle berry is a small plant that, when eaten, can make sour foods taste sweet). To do this, I had students compare the sweetness of lemon slices before and after eating the miracle berry (which was voluntary). We collected and analyzed the class data together, after which students completed a small APA style lab write-up of the results. After the experiment, I brought other foods the students could try for fun (pickles, grapefruit, etc.). It was an enjoyable way to learn about t-tests and remains a student favorite.

Another way I have utilized food in my statistics course is to use pizza to demonstrate the difference between samples (a single slice) and the population from which they are drawn (the whole pizza). I showed the class a pizza with pepperoni toppings and asked them which slice they would select. Different students had different strategies for selecting the “optimal” slice and we discussed how the pizza analogy related to sampling issues in statistics. They came up with some impressive parallels, many of which had not even occurred to me.

If you are interested in more ideas and thorough instructions on using food to teach statistical concepts, check out Chapter 11 in this book on using manipulatives, such as food, in the classroom.
Part 2: Using pre-lecture activities to support student engagement

Implementing the activities described above are enjoyable ways for me to foster engagement, but one trait all those activities have in common is that I, the instructor, created them. Some of the examples allow or require student input, but the framework is still my creation, not theirs. Although this approach is useful for introducing concepts and providing opportunities to practice, I also want to create an opportunity for students to make their own connections between their life and course content.

One tool that encourages my students to connect class concepts to their own lives is to assign pre-lecture activities (PLAs). For the PLAs, students complete the reading/lecture listed on the syllabus and then have complete freedom to create whatever product they want, in an effort to demonstrate that they have read or watched the lecture and tried to apply the content. Yes, you read that right – they can create whatever they want. The main grading criteria for the assignment is that they create unique, applied examples, not just regurgitated definitions or copied examples from the book/lecture. As long as they apply the content, they can be as creative or boring as they want (creativity is not part of the grade). Thus, not all students produce creative submissions, but ALL submissions are applied examples of the concepts. Many students just put together practice exam scenarios or basic notecards (with examples instead of definitions), and I am okay with that. They come to class having read the chapters and at least attempted to understand how the concepts might apply to their lives. But the part that brings me the most joy is that a lot of students get excited for the opportunity to create something special. I have had students develop comic book characters, write and sing songs about t-tests (while also playing ukulele), bake cookies in the shape of distributions, write a murder mystery on the premise of solving the replication crisis, and create an animated cartoon series. Other students take the assignment personally and write about their lives; one student explained how the scales of measurement related to her time as a beauty pageant queen, another student who was a volleyball coach applied z-scores to help decide the starting lineup, and others have written about their pets, children, jobs, and family.

I like the PLAs because they provide a consistent structure for students to move beyond memorization and look at their world through the lens of statistics; to take everyday occurrences and turn them into sources of data for learning. It takes a little getting used to for the students, but I’ve had some amazing student work come out of this. In additional to the examples of excellent student work mentioned above, I’ve had students submit a script for a Parks & Recreation (TV show) episode where the characters explained t-tests, an essay on how Eminem’s album Kamikaze demonstrated different scales of measurement, and an artwork series featuring Jackson Pollock-esque scatterplots for correlations. I am absolutely stunned at the talent of my students and I find myself looking forward to grading these assignments. However, as noted throughout this chapter, I tend to have small class sizes, which makes using PLAs more feasible than my counterparts with large class sizes.

In general, students seem to really enjoy the PLAs (see sample images from student work, below). For some, the enjoyment comes from the opportunity to be creative, and for others, the greatest benefit is helping them to prepare for class and exams. Either way, I love PLAs because they inject an element of fun while also meeting my desire to have students apply the material and come to class prepared.
Part 3: Professional Development Opportunities to foster engagement

As you can tell from the above section, I have spent considerable effort to find and create content that engages students. However, even after adding several semesters worth of examples, I still felt my course was missing something. Over time, I realized that although the examples were making statistics more interesting and personal, students still had little idea why they needed the class for a degree in psychology. A 2013 Gallup poll echoes these sentiments, as the majority of students (59%) reported that they learned most of their job skills outside of school. Thus, to make the skills students were gaining explicitly clear, I was inspired to provide them with more opportunities for professional development. To address this need, I began looking for a course project that would not only meet our content related learning objectives, but to help students understand how to apply statistics to realistic problems and develop professional skills that would be marketable post-graduation.

To begin the process of emphasizing marketable skills, I sought to identify which skills employers were seeking by using literature from both academic research (e.g., Landrum & Harrold, 2003) and popular press (e.g., Forbes and Harvard Business Review). Based on the accumulated evidence, I created a list of skills that my class should address. This list included oral communication (being able to give strong presentations, articulate thoughts, vacillate between technical terminology and language for clients), written communication (both formal and informal), personal skills (time management, accountability, motivation, stress, and coping), interpersonal skills (maturity, teamwork, leadership), and technical skills (quantitative literacy, familiarity with software, etc.). Since my elementary effort at compiling a list of desirable skills, a 2018 APA taskforce has since created a far more helpful document that outlines the transferable skills psychology students should have as they enter the workforce (Naufel et al., 2018). Their work culminated in five broad competencies that employers value: cognitive, communication, personal, social, and technological skills. I encourage you to check out their work, as each competency lists specific skills students should develop and provides operational definitions of each skill. For example, specific skills under social competencies include collaboration, inclusivity, leadership, management, and service orientation. For ease of use, the paper is formatted as a user-friendly PDF handout.
Once I identified the skills I wanted to help students develop, I needed to determine the platform to deliver them. I decided the best platform would be a service-learning project. Each semester, I partner with a local organization and our class analyzes data and provides feedback. To date, I have worked with a youth mentoring program, health clinic, and homeless shelter. Regardless of which organization we are partnering with, the process is generally the same each semester: someone from the organization visits our class to introduce themselves and the mission of the organization, we take a tour of the facility, and we spend one day volunteering for the organization. In lab, we spend time each week analyzing data from the organization and writing up results. At the end of the semester, students present their findings to the organization and work together to write a white paper. However, since a thorough overview of implementing service-learning in statistics is available elsewhere in this book (see Chapter 6), I will instead focus on \(a\) the types of assignments that cultivate the desired skills (see Table 3, based on Naufel et al., 2018) and \(b\) how I ensure students know exactly what job related knowledge, skills, and abilities they are gaining in my course (see Table 4).

### Table 3. Demonstrating student skills based on class activities

<table>
<thead>
<tr>
<th>Competency</th>
<th>Skill</th>
<th>Class Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive</td>
<td>Analytical Thinking</td>
<td>Write-up of lab results</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Weekly Lab</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Decide on appropriate analyses</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Conduct and interpret analyses</td>
</tr>
<tr>
<td></td>
<td>Critical Thinking</td>
<td><strong>Homework Assignments</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Evaluate appropriateness of data collection method</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Identify statistical concepts in data (which variables represent which scales of measurement, etc.)</td>
</tr>
<tr>
<td></td>
<td>Creativity</td>
<td>Create &quot;end product&quot; community partner could distribute</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Examples include video, fact sheet, brochure, etc.</td>
</tr>
<tr>
<td></td>
<td>Information Management</td>
<td><strong>Final Paper and Final Presentation</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Synthesize information about community partner from multiple sources (results, website, client interactions, etc.)</td>
</tr>
<tr>
<td>Communication</td>
<td>Oral Communication</td>
<td>Meet with community partner directors, clients</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Active listening and conversation skills</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Final oral presentation for community partner</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Communicating to diverse audience</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Final oral presentation at university research symposium</strong></td>
</tr>
<tr>
<td></td>
<td>Written Communication</td>
<td>• Communicating to scientific community</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Create public facing document for community partner</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Brochure and/or final report for diverse audience</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>APA style final lab report</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Includes all methods, results, and conclusions</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Peer Evaluations</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Provide written feedback to group members</td>
</tr>
<tr>
<td>Personal</td>
<td>Integrity</td>
<td><strong>Complete FERPA Training</strong></td>
</tr>
</tbody>
</table>
However, despite my best efforts to create opportunities for students to develop professionally relevant skills, I noticed that many students were not making the connections. That is, even with the service-learning project and the close partnership with a community organization, students seemed to still feel like this was “just” a statistics course and not relevant to their future careers. Consequently, I created a structure that provided explicit connections between our coursework and the bevy of skills students were acquiring throughout the semester. The end result is that I take every opportunity I can find to highlight the many, varied competencies we are developing (see Table 4 below and Hartnett’s [2020] discussion in Chapter 1 of this book).

### Table 5. Opportunities to explicitly connect statistics to marketable skills

<table>
<thead>
<tr>
<th>Class Activity</th>
<th>Implementation Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1st day of lecture</strong></td>
<td>On the first day of class, I immediately introduce two important concepts. First, that statistics relates to every aspect of our lives and I give a LOT of examples (healthcare decisions, housing, parenting, career choices, etc.). Second, the course is going to give them MANY skills that employers are looking for and that by acquiring these skills, they can be more competitive applicants. I show them several articles that list relevant skills and provide an overview of how the course will develop those marketable skills.</td>
</tr>
<tr>
<td><strong>Introduction to Service Learning</strong></td>
<td>Students read a very short article in class that outlines the benefits of service learning (Astin, Vogelgesang, Ikeda, &amp; Yee, 2000). They then complete a reflection assignment that connects service-learning to professional goals and development. We follow up with an in-class discussion about the importance of developing transferable job skills while earning a degree.</td>
</tr>
<tr>
<td><strong>Community Partner Visit: Class Discussion</strong></td>
<td>In the class period before representatives from our community partner visit, we have a class discussion on professional etiquette, developing and maintaining client relationships, and the ethics associated with our project and information privacy. Students then connect those skills to the skills necessary for their future profession.</td>
</tr>
</tbody>
</table>
Student Reflections

Once every three weeks, students spend a few minutes reflecting on their future professions, the skills required by those professions, and which of those skills they are gaining in our class. To give this activity structure, I provide them with a "menu" of skills I think they are developing (otherwise they always say the same skill - time management).

Self & Peer group evaluations

To ensure students are gaining insight into the group work process and have the opportunity for development, students evaluate both the quantity and quality of each group member twice throughout the semester (including themselves); once at the beginning and once at the end. They must identify areas of strength and development for everyone. This gives each student the chance to adjust their behavior and develop their interpersonal skills.

Final Presentations

Most students are extremely nervous to be presenting to the organizational representatives. Thus, I use this opportunity to highlight the importance of public speaking. I dedicate 20 minutes of class time to clearly identify oral communication as an essential work skill, examples of how it is used in a variety of psychology related jobs, the idea that strong presentation skills can distinguish applicants, and that presentations skills can be developed (followed by helpful tips and tricks).

Last Day of Class

By now, we have already wrapped up the course content, so I use this last day to review all the amazing skills they have learned.

- First, I remind students the top skills employers are looking for in new job applicants. I then show them a slide that lists all the skills they have learned throughout the semester (similar to the table presented above, Table 3). We review each skill and how it was developed.
- Next, I show students the most common interview questions (e.g., Tell me about a challenge or conflict you have faced and how you dealt with it; How do you deal with stressful situations?). We discuss how they could use their experiences in statistics and with the service-learning project to answer those questions.
- Finally, we go over how to add their service learning projects to their resumes and how to highlight their classroom experiences as relevant and viable job skills. Out of all my slides, these slides are the most requested by students.

Of course, the table above is not exhaustive. If you want your students to engage in meaningful application of course content, it takes daily reminders. Thus, I like to complement the skills learned in the service-learning project with the fun, practical applications from earlier in the chapter. Taken as a whole, I find that students are emotionally and cognitively engaged in my statistics course.

Final Thoughts

I love teaching statistics. The opportunity to make a demanding course into something that will engage and excite students is a challenge I relish. Well-designed examples, activities, and course projects keep me accountable to my student learning outcomes and I hope I’ve successfully shared some of these ideas with you. If you’re looking for more inspiration, there are countless websites, blogs, and teaching forums that can help you find what you’re looking for. You can also check out the compendium at the end of this book (Chapter 19) that includes many of the resources I use for inspiration (and special acknowledgment to the STP Facebook page, Jessica Hartnett’s Not Awful and Boring blog about stats, and STP’s Annual Conference on Teaching for all the inspiration to my teaching over the years). No matter which resources
you use, pick examples that you find exciting and engaging. You won’t be able to convey the fun and intrigue to students if you don’t enjoy it for yourself.

Post Script: If you are curious about any of my activities or would like to learn more, please feel free to email me at janet.peters@wsu.edu or follow me on Twitter at @ProfessorJanet.
References


Chapter 10

Puzzling it Out: A Collaborative Review Activity

Miranda M. McIntyre, PhD
California State University – San Bernardino

Summary

Puzzles are a fun, flexible tool that can inspire students to recognize the joy of discovery in statistics. An emerging area of research supports the value of game-based collaborative learning. When used effectively, educational games increase student performance and attitudes toward learning. Instructors can harness these benefits by introducing puzzles to “gamify” their statistics curriculum. This chapter offers advice for designing and implementing statistical puzzles, including evidence-based tips for maximizing student engagement and learning. Suggestions for time-strapped and tech-savvy instructors are provided to demonstrate how this technique can be adapted in nearly any classroom. Instructors are encouraged to consider adopting puzzle-based activities to improve students’ quantitative performance and motivation.

Why Use Puzzles?

Traditional review sessions can be boring and ineffective. Often, these reviews involve instructors reciting the same material presented during class or students churning through dozens of practice examples. This chapter introduces a puzzle-based game that can liven up review sessions and increase their impact. Rather than aimlessly flipping through their notes, students apply the course content to decipher codes, reveal patterns, and discover that statistics can be fun!

The puzzle approach to reviewing class material is inspired by the popular “escape room” trend. In escape room games, players work together to find hidden clues and solve riddles that allow them to escape a locked room. While locking students in study hall is not recommended, the same model can be applied to reviewing course content. Using this technique, teams of students solve a series of puzzles that require statistical knowledge. Each puzzle must be solved to “unlock” the next problem. The game-like design of this activity stimulates engagement with the material while encouraging students to practice and hone their quantitative skills.

Beyond capturing students’ interest, this activity makes use of effective pedagogical strategies. Students must pool knowledge with their peers to achieve a common goal, which tends to boost learning outcomes when compared with individual work (Johnson, Johnson, & Smith, 2007; Prince, 2004). Research on collaborative learning finds that this achievement benefit stems from several mechanisms, including explanation and error correction (Krause, Stark, & Mandl, 2009; Nokes-Malach, Richey, & Gadgil, 2015). To solve a puzzle correctly, students must arrive at a common answer by assessing the group’s ideas. The act of justifying or refuting a solution leads students to critically evaluate their approach and resolve gaps in their knowledge. This process promotes conceptual understanding and active learning; two strategies that are consistent with American Statistics Association teaching recommendations (GAISE College Report ASA Revision Committee, 2016). Collaborative learning generally improves performance, but group cooperation can also be costly. The learning process can be disrupted in poorly-coordinated
groups (Nokes-Malach et al., 2015), highlighting the importance of instructor facilitation during the activity.

Further research has demonstrated the benefits of game-based activities as a specific type of collaborative learning. Students who participate in an educational game perform significantly better on follow-up assessments, compared with randomly-assigned control groups (Cardozo, Miranda, Moura, & Marcondes, 2016; Luchi, Montrezor, & Marcondes, 2017). The instant feedback provided by solving a puzzle or completing a game may contribute to this performance boost. Students tend to learn more when they receive immediate feedback and can re-answer until they are correct (Attali & Powers, 2010; Epstein, Epstein, & Brosvic, 2001; Krause, Stark, & Mandl, 2009). The importance of timely assessment feedback is also emphasized by the American Statistics Association (GAISE College Report ASA Revision Committee, 2016). In addition to the performance benefits, game-based learning increases students’ motivation and attitudes toward learning (Plass et al., 2013; Sung & Hwang, 2013). The positive impact of games on student attitudes can occur even when there is not a direct performance boost (Ebner & Holzinger, 2007).

Taken together, the research on collaborative game-based learning demonstrates that puzzles are functional as well as fun. The escape room technique can be used to review material in a way that is dynamic and engaging, but also beneficial to learning. Statistics instructors are encouraged to consider adopting this activity to improve students’ quantitative performance and motivation.

**Puzzle Design and Implementation**

This activity can require extensive preparation depending on how elaborate or customized an instructor chooses to make the puzzles. For those who prefer a simpler set-up, it is possible to use standard practice problems from a textbook or existing review materials. Because each puzzle is a self-contained unit, this technique is highly adaptable. Instructors can adjust the length of the activity and the topic coverage to suit their particular curriculum. See the final section of this chapter (“Wrap-Up”) for suggestions on adapting and extending the game. Sample puzzles that cover a range of introductory topics are included in the appendix.

**Planning Ahead**

What makes an effective statistical puzzle? Each part of the activity should teach or review a specific quantitative skill while adding an element of fun. An ideal puzzle should rehearse a bite-size piece of knowledge, such as “Which type of comparison is appropriate for these groups?” or “How do you make a significance decision based on test values and critical values?” A beginning-to-end problem is not well-suited to creating a puzzle – for example, “Using this dataset, generate a testable prediction and conduct the statistical test to evaluate your hypothesis.” The end objective in a course may be for students to perform the entire process from hypothesis generation to reporting, but this technique is designed for practicing the building blocks in that process.

To create a puzzle from scratch, start with a concrete learning objective. Again, this should be a relatively specific skill rather than a multi-step process. For instance, students may be expected to interpret the strength and direction of correlation coefficients. Next, decide how students will encounter the material. Will the correlations be presented as output in a table? Or in text, as an article-style results section? At this stage, the mechanics of the puzzle must also be considered. Students could be asked to arrange the coefficients along a number line, sort the coefficients into categories depending on strength, or identify which coefficients would be significant at a given sample size. Finally, specify what the solution for the puzzle will look like. What will students need to submit to continue on or finish the game? This could be a sequence of
numbers, a code word, a visual pattern, or other answer depending on how the puzzle is constructed. Several examples are summarized in the following section and additional puzzles are available in the appendix.

Ideally, each puzzle includes a fun twist that will capture students’ interest and play into the game-based nature of the activity. Simple code systems work well, such as Morse code or letter-to-number cyphers (A=1, B=2, C=3, etc.). Visual elements are also stimulating—consider incorporating not just graphs and tables, but also shapes, colors, or even objects (see Fetterman & Kneaval's Chapter 11 on manipulatives). Depending on the classroom environment and location, it may also be possible to play with physical space. The answer to a puzzle could identify a row of chairs in the lecture hall or another room in the building (where the next puzzle is pinned to the seat/door). Variety is the key, as it keeps the game engaging and rewarding. However, it is important to keep the focus on learning rather than allowing the game itself to become a distraction. Instructors should aim for simplicity so that students are practicing course content instead of struggling to decipher a code system. The puzzle mechanics should be relatively simple to keep learning outcomes at the forefront.

Once the sequence of puzzles has been designed, the materials should be prepared in advance of the class session. A set of puzzle materials should be assembled for each anticipated team. The instructor should also prepare an answer key to keep track of the correct solutions and the order of puzzles. Groups will work through the activity at various paces—a well-prepared group may fly through five puzzles while another group struggles to complete their second task. If each solution must be turned in or checked by the instructor, it can be tricky to stay organized. This process can be streamlined by setting up the activity so that each solution gives students access to the next puzzle (opening an envelope, going to a physical location, etc.).

Puzzle Examples

A few puzzles are summarized below to provide inspiration for puzzle construction. A list of the full sample puzzles are provided in the appendix.

<table>
<thead>
<tr>
<th>I</th>
<th>O</th>
<th>I</th>
<th>N</th>
<th>R</th>
<th>N</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>N</td>
<td>R</td>
<td>O</td>
<td>I</td>
<td>O</td>
<td>I</td>
</tr>
<tr>
<td>I</td>
<td>O</td>
<td>I</td>
<td>N</td>
<td>R</td>
<td>N</td>
<td>R</td>
</tr>
<tr>
<td>R</td>
<td>N</td>
<td>R</td>
<td>START</td>
<td>O</td>
<td>I</td>
<td>O</td>
</tr>
<tr>
<td>O</td>
<td>I</td>
<td>O</td>
<td>I</td>
<td>N</td>
<td>R</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>R</td>
<td>N</td>
<td>R</td>
<td>O</td>
<td>I</td>
<td>O</td>
</tr>
<tr>
<td>O</td>
<td>I</td>
<td>O</td>
<td>I</td>
<td>N</td>
<td>R</td>
<td>N</td>
</tr>
</tbody>
</table>

*Figure 1. Sample puzzle materials for levels of measurement (left) and factorial designs (right).*

- **Levels of Measurement** – A list of variables is provided (e.g., “Distance travelled in meters”) along with a maze grid (see image on the left, below). Students identify whether each variable is measures using a nominal, ordinal, interval, or ratio scale. The correct answers identify a series of steps that create a path out of the maze.
- **Main Effects and Interactions** – Students receive a blank Venn diagram and sample graphs from 2X2 factorial designs (for a completed puzzle see image on the right, below). They interpret the results of each graph to identify any main effects or interactions. The puzzle is solved when the correct graph is matched with each area of the Venn diagram.

- **Critical Values** – Students receive descriptions of study designs along with their sample sizes and alpha levels. Using a critical values table, they identify the critical value(s) for each test. When the correct values are identified, they connect to form a shape. Students open an envelope marked with that shape (among several decoys), which leads to the next puzzle.

### In the Classroom

To implement this activity, an instructor plays the role of “Puzzlemaster” while students work in teams to solve the puzzles. Students should first be divided into teams. Many students prefer self-selected groups rather than randomized or instructor-assigned groups. Research indicates that enjoyment and teamwork are higher in self-selected groups, but randomly assigned groups tend to be more efficient and lower in conflict (Chapman, Meuter, Toy & Wright, 2006). Instructors should also consider inclusiveness when determining groups. Randomly assigning students to teams will likely produce groups that are more diverse and varied than if students choose their own teams. Group size is another consideration – teams of three to five students are ideal for this activity. It can be difficult for large groups to coordinate effectively and the risk of social loafing increases as groups get bigger (Davies, 2009).

Instructors should introduce the game and clarify any ground rules before beginning. Explain the purpose of the activity to students and how it supports their learning. If the activity will count toward a grade or extra credit, describe how this will be assessed. For best results, instructors should explicitly emphasize teamwork during the activity. Students’ goal should be to refine their quantitative skills, not to race through the puzzles as quickly as possible. Competitive students have a tendency dominate their group while struggling students are left out and unable to benefit from the activity. This issue may be prevented by stating that all team members should be involved and that everyone should understand the solution before moving on.

Once the first puzzle is revealed and the teams begin working, the instructor should actively monitor their progress. Depending on puzzle complexity and student skill levels, the puzzles tend to take between 5 and 10 minutes to solve. Two common issues in this activity are groups that become stuck on a puzzle and groups that do not share the work evenly. When a group struggles find a solution, the instructor can provide hints to nudge them in the right direction. If notes or textbooks are permitted during the activity, they can also direct students to the relevant material. If team members participate unevenly, they may benefit from a reminder to emphasize teamwork. When students submit a solution, the instructor may ask the group whether everyone understands why it is correct. Students who are reserved or struggling with the material may need direct encouragement.

After the puzzles are completed, a debriefing session can be used to conclude the activity. Students can be asked to share which puzzles were trickiest to stimulate a discussion of the most difficult material. Instructors can also provide follow-up individual exercises that rehearse the skills used to solve the puzzles. Students can then assess their own competency after the group activity.

### Outcomes

This activity was piloted in an introductory statistics course with approximately 50 students. The game was conducted in two lab sections that each comprised half of the class. Students’
experiences and responses were extremely positive. Most teams participated eagerly even though the activity was not part of their grade. In one particularly efficient group, the students powered through six puzzles in ~25 minutes and were disappointed to find out that there were no more puzzles! Students who had difficulty with the game seemed to become more motivated to learn the content. Many stayed after class to discuss the solutions and requested a copy of the puzzles so that they could practice on their own. Due to the small sample size and lack of a control group, no quantitative performance data is available from the pilot test. However, students subjectively reported that the game was more helpful and more enjoyable than a typical review session.

The biggest challenge that arose during the pilot was ensuring that all students were actively involved. While the game lends itself to a competitive atmosphere, instructors should be careful not to encourage students to finish quickly—for example, by rewarding bonus points to the team that finishes first. This promotes speed rather than competency, and shy or struggling students are likely to be left out. One adjustment that could encourage inclusive participation is a Jigsaw Classroom approach. In this strategy, team members are provided with different pieces of information that must be combined to find the correct answer (Social Psychology Network, 2019). Thus, the group’s success requires all members to contribute during the activity. As noted in the previous section, it may also be helpful to include individual follow-up exercises for self-paced practice.

Wrap-Up: Adaptations and Conclusions

The puzzle approach described in this chapter is just one way to incorporate collaborative game-based learning in the classroom. Instructors may discover other ways to adapt elements of this game in their curriculum. For some, the time investment can seem prohibitive. This technique can be preparation-heavy, especially when designing a sequence of multiple puzzles and conducting the full game in class. Alternatively, an individual puzzle can be used as a warm-up activity at the beginning of class or a check-in exercise after new material is introduced. A game-like twist, such as a code word materializing from the correct answers, can turn a regular practice example into a more stimulating activity. If class time is limited, the puzzles also work well as take-home exercises—students may be more likely to complete voluntary practice problems when they are framed as a game. Another time-saving approach is to flip the activity around. Students can be asked to design novel puzzles that demonstrate their understanding of the material. In sum, instructors can capitalize on the benefits of game-based learning without dedicating hours to puzzle-making.

For the technology-savvy, this technique can also be extended into digital learning environments. While the game works well with simple paper-and-pencil, online tools greatly expand its possibilities. For example, a puzzle could lead students on a scavenger hunt for QR codes that link to the next puzzle when scanned. Custom URL strings can also be used to validate correct solutions. Online classes are an ideal medium for these technology-based extensions (see Wilson-Doenges’ Chapter 2), but digital tools can also improve the game for in-person classes. In courses that are especially large, it could be difficult to coordinate groups and manage printed puzzle materials. Online materials can be used to streamline the game and process teams’ answers automatically (for an impressively elaborate example, see the annual MIT Mystery Hunt: https://www.mit.edu/~puzzle/).

Whether used as a multi-stage game, a lecture launcher, or an online activity, instructors can customize this approach to suit the needs of their own course. Ultimately, puzzles are a flexible tool that can spark interest in statistics and deepen learning. The thrill of puzzle-solving is in the act of discovery—a sudden moment of clarity when the answer is revealed. By engaging with puzzles in the classroom, students may come to recognize the same joy of discovery in statistics.
References


Appendix
Puzzle Index in McIntyre’s Complete Puzzle Materials

<table>
<thead>
<tr>
<th>Puzzle</th>
<th>Topic/Learning Outcome</th>
<th>Pages</th>
<th>Preparation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Identifying variable level of measurement</td>
<td>Multiple</td>
<td>None</td>
</tr>
<tr>
<td>B</td>
<td>Calculating measures of central tendency (mean, median, mode)</td>
<td>Single</td>
<td>None</td>
</tr>
<tr>
<td>C</td>
<td>Identifying statistical notation (statistics vs. parameters)</td>
<td>Single</td>
<td>None</td>
</tr>
<tr>
<td>D</td>
<td>Understanding frequency distributions, normality, modality, and skewness</td>
<td>Multiple</td>
<td>None</td>
</tr>
<tr>
<td>E</td>
<td>Calculating z-scores from raw scores, using the z distribution</td>
<td>Multiple</td>
<td>None</td>
</tr>
<tr>
<td>F</td>
<td>Distinguishing between null and alternative hypotheses</td>
<td>Multiple</td>
<td>None</td>
</tr>
<tr>
<td>G</td>
<td>Identifying Type I and Type II error</td>
<td>Single</td>
<td>None</td>
</tr>
<tr>
<td>H</td>
<td>Determining statistical significance, interpreting test values and critical values</td>
<td>Multiple</td>
<td>None (Color print helps)</td>
</tr>
<tr>
<td>I</td>
<td>Identifying t-tests, calculating degrees of freedom</td>
<td>Single</td>
<td>Cutouts</td>
</tr>
<tr>
<td>J</td>
<td>Calculating degrees of freedom, locating critical values using a t-table</td>
<td>Multiple</td>
<td>Separate envelopes</td>
</tr>
<tr>
<td>K</td>
<td>Identifying types of group comparisons</td>
<td>Single</td>
<td>None</td>
</tr>
<tr>
<td>L</td>
<td>Interpreting correlation coefficients and scatterplots</td>
<td>Multiple</td>
<td>Cutouts</td>
</tr>
<tr>
<td>M</td>
<td>Interpreting post hoc results for one-way ANOVA</td>
<td>Multiple</td>
<td>None</td>
</tr>
<tr>
<td>N</td>
<td>Identifying main effects and interactions for two-way ANOVA</td>
<td>Multiple</td>
<td>Cutouts</td>
</tr>
<tr>
<td>O</td>
<td>Determining statistical procedures based on research questions</td>
<td>Single</td>
<td>None</td>
</tr>
</tbody>
</table>

Note: Materials for this project were supported by a grant from the Social Science Research and Instructional Center, and available at https://ssric.org/trd/exercises (Direct file link: http://www.ssric.org/files/2019-12/StatPuzzles-AllFiles.pdf). The materials are under a CC-BY-NA-SA 4.0 copyright per the terms of the grant.
Engaging Students in Statistics: The Power of Manipulatives

Joshua D. Fetterman¹, PhD and Meredith E. Kneavel², PhD
Chestnut Hill College¹, LaSalle University²

Summary
Manipulatives are concrete objects that students can touch and feel (and in some cases eat) that instructors use to demonstrate statistical concepts. Students find manipulatives engaging and fun, and copious research attests to their efficacy as a teaching tool. One of the many reasons for their efficacy is their ability to help students overcome “statistics anxiety,” which if left unchecked can lead to a host of negative outcomes for students. For these reasons we recommend manipulatives for psychology statistics and research methods courses. In this chapter we review research on statistics anxiety and propose using manipulatives to eliminate that anxiety. We then provide suggestions for manipulative activities that address the topics that are typically covered in a psychology methods courses. For each activity we provide materials, instructions, and recommendations. Our experience has been that these activities increase student learning and enjoyment in psychology statistics classes, and they help mitigate statistics anxiety.

Statistics in Psychology
Most graduate and undergraduate psychology curricula require students to complete both a statistics and research methods course (Aiken, West, Sechrest, & Reno, 1990; Norcross et al., 2016). Both of these courses typically require students to perform some sort of statistical computations. Since psychology is one of the most popular college majors (Undergraduate Degree Fields, 2019), a substantial number of students must pass these courses in order to obtain their degrees. Furthermore, high achievement in these courses predicts performance on a broad measure of psychology knowledge as well as in subsequent psychology courses (Freng, Webber, Blatter, Wing, & Scott, 2011). Despite the emphasis placed on statistical and methodological training in psychology programs, both faculty and students believe that students leave psychology programs with deficits in their design and analysis abilities (Aiken et al., 1990; Huntley, Schneider, & Aronson, 2000), and evaluations of students’ statistical knowledge justify these concerns (Gonulal, 2019). How can it be the case that psychology students receive such extensive instruction in statistics, yet that instruction is ineffective? One explanation for these deficiencies is statistics anxiety.

Statistics Anxiety
Statistics anxiety affects a majority of graduate and undergraduate students (Onwuegbuzie & Wilson, 2003; Zeidner, 1991). This anxiety revolves around broad aspects of statistics courses, such as calculation and interpretation, asking for help, attending class, taking tests, and statistics instructors themselves. (Gonzalez, Rodriguez, Falde, & Carrera, 2016; Daley & Onwuegbuzie, 1997; Onwuegbuzie & Wilson, 2003; Williams, 2010). It is related to, although distinct from, mathematics anxiety, which is associated with using mathematical symbols and operations (Baloglu, 2004; Gonzalez et al., 2016). Older students and those who lack experience or feelings of mathematical efficacy are particularly susceptible to statistics anxiety.
(Pan & Tang, 2004; Wilson, 1997). Unsurprisingly, the negative effects of statistics anxiety are myriad and range from poor performance in statistics classes to physiological symptoms such as increased heart rate and watering eyes (Freng et al., 2011; Gonzalez et al., 2016; Malik, 2015; Onwuegbuzie & Wilson, 2003; Paechter, Macher, Martskvishvili, Wimmer, & Papousek, 2017; Roberts & Bilderback, 1980). In spite of their poor performance, students with statistics anxiety do not lack intelligence or the ability to succeed in a statistics course; the problem appears to be the anxiety itself (Baloglu, 2004). For these reasons, finding ways to help students cope with their anxiety is critical (Baloglu, 2004; Blalock, 1987). Fortunately, researchers have found many methods to help students cope with their anxiety (Dillon, 1982; Hendel & Davis, 1978; Macheski, Lowney, Buhrmann, & Bush, 2008; Pan & Tang, 2004; Schacht & Stewart, 1990; Sgoutas-Emch & Johnson, 1998; Waples, 2016; Williams, 2010). Our favorite method for helping students overcome their anxiety is incorporating manipulatives into the classroom.

**Manipulatives**

Manipulatives are tactile, physical objects that can be used to demonstrate mathematical concepts (Marley & Carbonneau, 2015). There is abundant evidence that students’ mathematics education benefits from the use of manipulatives and that they positively influence students’ attitudes towards math (Carbonneau, Marley, & Selig, 2013; Ojose & Sexton, 2009; Sowell, 1989). Science students (Schroeder, Scott, Tolson, Huang, & Lee, 2007) and students with disabilities (Bouck & Park, 2018) benefit from the inclusion of manipulatives in the classroom and their effects seem to be particularly potent for older students (Carbonneau et al., 2013). Students who use manipulatives do a better job of staying on task, believe manipulatives enhance their understanding of material, enjoy working with them, and appear to be engaged (Crockett & Kilgour, 2015; Lee & Ferrucci, 2012; Moyer, 2001; Swan & Marshall, 2010). Indeed, some teachers even report *taking away* their classes’ manipulatives as a punishment (Moyer, 2001). Given the benefits that accrue from the use of manipulatives in the classroom, it is not surprising that the American Statistical Association’s Guidelines for Assessment and Instruction in Statistics Education College Report recommends that teachers engage their students in active learning, perhaps through the use of manipulatives (GAISE College Report ASA Revision Committee, 2016). Most importantly for the current context, researchers have found manipulatives to be effective in reducing anxiety (Vinson, 1997).

Getting students engaged in class and perhaps even having fun is an important step to helping them overcome their statistics anxiety, and there is ample evidence that manipulatives accomplish this goal very effectively. Indeed, in our own teaching we have found manipulatives to be an effective tool for eliminating statistics anxiety and enhancing student engagement and learning. In the remainder of this chapter we outline exercises using manipulatives that are tied to each of the typical elements of an undergraduate psychology statistics course. Demonstrations that involve the calculation of descriptive or inferential statistics can all be done by hand, or by using SPSS or Excel, depending on the objectives of the course and instructor preferences. Detailed instructions, data collection sheets, and materials needed for each intervention are included in the appendix. See Table 1 for a list of topics and corresponding demonstrations.

**Table 1. List of Demonstrations by Statistical Concept Covered**

<table>
<thead>
<tr>
<th>Statistical Concept</th>
<th>Demonstration</th>
<th>Appendix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptives</td>
<td>Using Things that Students Have on Hand</td>
<td>A</td>
</tr>
</tbody>
</table>
Manipulative Interventions for Specific Elements of a Statistics Course

Descriptives

Descriptives Demonstration 1: Using Things that Students Have on Hand: For illustrating descriptive statistics, it is efficient to use items that students bring with them to class. For instance, instructors can record the number of pens and pencils each student has, tabulate these for the class, and ask students to calculate the measures of central tendency and variability for each. This could also be done with the number of items students have in their pockets, on their desks, in their backpack, or with the number of apps on their cell phone. For the very tech savvy (or those who want to learn from their students) one could even ask students to report the number of hours and minutes they spend looking at the screen of their phone (it is reported under settings and is broken out by ‘social media’, ‘productivity’, and ‘entertainment’). This can be adapted to online courses by having students send in their data. For large classes, it may be useful to have students provide their data in one class and analyze it in the next class meeting. See Appendix A for sample data recording sheet.

Descriptives Demonstration 2: Snack Pack Variability: Another way to illustrate descriptive statistics is to distribute “fun sized” packs of M&Ms or any small “fun size” candy and have students count the number of candies in each packet. “Fun sized” packs of M&Ms are typically

<table>
<thead>
<tr>
<th>Probability</th>
<th>Playing Card Probability</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampling</td>
<td>A “Mini” Sampling Distribution of Sampling Means</td>
<td>D</td>
</tr>
<tr>
<td>One-Sample t-tests</td>
<td>Does Mars™ Lie About How Many M&amp;Ms are in a Bag?</td>
<td>F</td>
</tr>
<tr>
<td>Independent Samples t-test</td>
<td>M&amp;Ms versus Skittles: Which Has More?</td>
<td>G</td>
</tr>
<tr>
<td>Repeated Measures t-test</td>
<td>Speed/Accuracy Trade-Off</td>
<td>I</td>
</tr>
<tr>
<td>Chi-Square Test for Goodness of Fit</td>
<td>Do Our M&amp;Ms Fit the Population Parameters?</td>
<td>J</td>
</tr>
<tr>
<td>Correlation</td>
<td>Distance and Accuracy</td>
<td>L</td>
</tr>
<tr>
<td>Analysis of Variance &amp; Mixed Factorial Analysis of Variance</td>
<td>The Importance of Binocular Vision</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>Speed, Accuracy, and Underhand Pitches</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Expansion of the Speed-Accuracy Trade-Off</td>
<td>O</td>
</tr>
</tbody>
</table>
available in the candy isle of grocery stores, or can be ordered online. After distributing the candy to students, the instructor can then compile and display a list of each student’s count for the class. An especially interesting comparison can be made between this value and the value listed on the packaging for the expected number of M&Ms in each package. This activity can be effectively adapted for group learning. Distribute “fun sized” packs of M&Ms to groups of four or five students and ask them to count the number of colors of each candy. After recording this data on a data collection sheet, students can calculate measures dispersion and central tendency for each color. In order to illustrate how central tendency and dispersion are differentially affected by data alterations, have students imagine changing the data in various ways (e.g., imagine that there are five additional candies of each color), and then redo their calculations. See Appendix B for materials designed for both individual and group activities.

**Probability**

Probability Demonstration: Playing Card Probability: Break students into groups and give each group a deck of playing cards. Students can determine the probability of drawing particular cards (e.g., a king) or categories of cards (e.g., a face card) from the deck. Students can also determine joint probability (e.g., drawing a card that is the Ace of Hearts) and conditional probability (e.g., drawing a face card that is not red). After this, have students draw cards from the deck and determine the probability of selecting each card. Importantly, after students have drawn a card, make sure that students set that card aside so that on subsequent trials the denominator in the probability function changes. This will illustrate the effect of sampling without replacement. Card players sometimes remove jokers from a deck of cards before starting their game, so be sure to specify whether jokers should be included in the deck to avoid confusion. In fact, one point of discussion could be how the probabilities would change depending on the inclusion of jokers. A sample data collection sheet is provided in Appendix C.

**Sampling**

Sampling Demonstration 1: A “Mini” Sampling Distribution of Sample Means: After defining the sampling distribution of sample means, present students with three important facts about this sampling distribution: 1) the sample mean is an unbiased estimator of the population mean, 2) the sampling distribution of sample means is normally distributed, and 3) the standard error of the mean (\(\sigma/\sqrt{n}\)) is the standard deviation of the distribution of sample means. Next, create a ‘mini’ sampling distribution of sample means. Do this by distributing M&Ms “fun sized” packets to students and have them calculate the percentage of blue M&Ms in their packet. Next, generate a small number of random samples from www.random.org (2019). Evaluate this “mini” sampling distribution of sample means to determine if the three important facts mentioned above are born out in the data. These values will probably not match up precisely (i.e., the way that they should be based on theory), but explain to students that this is to be expected because the sampling distribution that you have created is incomplete; if you had time to create a complete or larger sampling distribution they would match up precisely. By setting up an excel spreadsheet to calculate these values automatically (or using the spreadsheet provided) you can substantially reduce the amount of time and the number of calculations that are necessary to do this activity. A more detailed set of instructions is provided in Appendix D.

Sampling Demonstration 2: Sample Size – Why Bigger is Better: Read the article in Quarts by Purtil (2015) about how M&Ms are created in two different factories, one in Tennessee and one in New Jersey. It turns out that these factories produce slightly different distributions of M&Ms colors. Ask students to read it, as well. Hand out fun size bags of M&Ms to each student and have them calculate the percentage of each color in their bag. Present the percentages from the New Jersey and Tennessee factory to students and have them guess which factory their bag came from based on the distribution in their bags. Then have students merge their M&Ms with
individual classmates, and eventually the class as a whole, each time making a new guess about the distribution center of origination. You should find that as their samples become bigger, they more accurately reflect the distribution center of origination, and that students’ guesses become more accurate as this process continues. You can determine the factory of origination for your M&Ms by looking at the outer bag that contains the “fun sized” M&Ms packs, but this information is not printed on the “fun sized” packs themselves. See Appendix E for more detailed instructions and a data collection sheet that walks students through this process.

**t-tests**

One Sample *t*-test Demonstration 1: Does Mars™ Lie About How Many M&Ms are in a Bag? According to the Anne (n.d.), there are three calories in one regular M&M. The packaging of M&Ms “fun sized” packs indicates that a serving size is three packs and contains 190 calories. Therefore, 190/3 should equal 63.33 calories per packet, and at three calories a piece, there should be 21 M&M’s in each pack. Experience indicates that “fun sized” packs typically do not contain that many M&Ms. One interesting task that illustrates the use of a one independent sample *t*-test is to have students determine if Mars™ is lying about how many M&Ms they put in each “fun sized” pack. The value for calories per serving and serving size seem to change periodically, but the necessary calculations to do this activity are relatively straightforward, and can be redone if you happen to have “fun sized” M&Ms packs with different values. When having students calculate the *t*-test, set the comparison value to 21 (see Appendix F for detailed information). For online classes, give students the same problem but have them report the number they find in fun size bags from home. For large classes, have students gather into groups of 5-6 to count sample sizes.

Independent Samples *t*-test Demonstration 1: M&Ms versus Skittles: Which Has More? To demonstrate independent samples *t*-tests, students can compare M&Ms and skittles “fun size” packs to see which product contains more candy in each pack. Give each student one packet of each type of candy and ask them to count the number of candies in each pack and record the total (see Appendix G for data collection sheet). Conduct a two independent sample *t*-test to determine if there is a significant difference between the two types of candy.

Independent Samples *t*-test Demonstration 2: Altered Perception Ball Passing Accuracy: For this demonstration, instructors will need glasses that alter vision and small balls such as tennis balls. Vision altering glasses can be obtained from the reading glasses section of pharmacies, from games such as Googly Eyes (search for “Googly Eyes Game” on Amazon.com), or by taking a pair of sunglasses and scratching the lenses or wrapping them in plastic wrap or lace. Pair up students in the class and pass out glasses to half of the group. Have them pass the ball back and forth for 90 seconds either wearing, or not wearing, the vision altering glasses. Compare the two treatment conditions to determine differences. See Appendix H.

Repeated Measures *t*-test Demonstration 1: Speed/Accuracy Trade-Off: In this demonstration students pass a ball back and forth under two treatment conditions: speed and accuracy. Pair the students in the class and pass out a ball. Pass out the directions for either the speed condition or the accuracy condition (see Appendix I) and records students’ performance (the number of completed passes). Then have the students follow the second set of instructions and, again, record their performance. In order to illustrate the importance of rigorous methodology that eliminates confounds, instructors can briefly discuss order effects and employ a counterbalanced design for this activity. Calculate a related samples *t*-test to see if the instructions influence the number of completed passes. For online classes, have students employ the help of a friend at home or toss the ball against a wall, following the instructions for both the speed and accuracy conditions. Collate the class data and run the *t*-test for the class.
For large classes, have students separate into larger groups rather than pairs (i.e. groups of 6-8) to pass the balls and record the number of passes under both treatment conditions.

**Chi-Square Test for Goodness of Fit**

Chi-Square Test for Goodness of Fit Demonstration 1: Do Our M&Ms Fit the Population Parameters? In this demonstration, the instructor distributes a small cupful of M&Ms from a large bag to a small group of two to three students. For large classes, this can be modified by using two or three large bags and using larger groups of four or five students with slightly larger cupfuls. Have students count and record the number of each color of M&M. Provide the ‘factory output’ percentages so that students can calculate expected frequencies. The ‘factory output’ percentages that have been previously published are blue: 24%, orange: 20%, green: 16%, yellow: 14%, red 14%, brown 14%; although other reports have updated these percentages and may vary depending on the factory that produced them (Purtill, 2017). You should be able to check your bag of M&Ms to determine the factory that produced your bag of M&Ms, and then use the information from Purtill (2017) to determine the correct proportions (see Sampling Demonstration 2 above and Appendix J).

Chi-Square Test for Goodness of Fit Demonstration 2: Buzzfeed Oreo “Facts:” Buzzfeed has a brief video entitled *9 Things About Oreos You Didn’t Know* which, among other interesting facts about these cookies, claims that 84% of males eat Oreos whole and 41% of females twist Oreos apart. While students are eating the Oreos that have been distributed, conduct a poll to classify them as either “twisters” or “non-twisters.” Instruct them that they must choose one category or the other. Then, record the observed frequencies from the class data through a show of hands. Next, break this data up according to gender, and calculate Chi-Squares to see if the preferences of the class map onto the preferences reported in the video. See Appendix K for more detailed instructions. This can be done in large classes, online classes, and small classes with minimal modification.

**Correlation**

Correlation Demonstration 1: Distance and Accuracy: Create a poster board target and place it on the floor, using masking tape to secure it. Place a tape measure on the floor extended straight out from the target. Have the students throw the beanbag at the target and, after it lands, record the distance that the student stood from the target and the distance that the beanbag fell from the target. You can let students pick their own distance, use a random number generator to generate random numbers of inches that students must stand from the target, or assign them predetermined distances to stand from the target (see Appendix L for a procedure that assigns predetermined distances). After recording the data, create a scatterplot that can be inspected for linearity, and then calculate a Pearson (or if appropriate a Spearman) correlation to assess the relationship between distance from target and accuracy. For large classes, randomly select a certain number of students to do the activity in front of the class.

**Analysis of Variance & Mixed Factorial Analysis of Variance**

Analysis of Variance Demonstration 1: The Importance of Binocular Vision: For this demonstration, pairs of students will pass a ball back and forth. However, they will do so under three different conditions: with both eyes open, with only their right eye open, and with only their left eye open. Students should be instructed to make as many successful passes as possible. As mentioned in the section on repeated measures *t*-tests, this is an opportunity to discuss and demonstrate counterbalancing. However, a full counterbalancing procedure would require six unique orders, and consequently may be cumbersome in this context. If you choose to counterbalance, consider using a Latin Square procedure (where each condition appears exactly once in each serial position), as this would only require the use of three unique orders.
As described the proper analysis for this data would be a one-way within subjects ANOVA, but it could easily be adapted to be a one-way between subjects ANOVA by having each group of students pass the ball under only one set of conditions. This is also an opportunity to briefly discuss binocular vision with your students, and test the importance of this feature of human perception. Larger classes can employ larger groups, and online students may need a friend to complete the activity, or toss the ball against a wall (results can then be sent to the instructor for collation). See Appendix M for more detailed instructions.

Analysis of Variance Demonstration 2: Speed, Accuracy, and Both: This is an adaptation of the independent samples t-test activity described earlier, where students pass a ball back and forth using instructions for either accuracy or speed. Divide students into pairs, and the pairs into three groups, one group that will receive speed instructions, one group that will receive accuracy instructions, and one group that will receive instruction to maximize both speed and accuracy. Instruct students to count their successful passes, and then use a one way between subjects ANOVA to determine which condition made the most successful passes. See Appendix N for more detailed instructions. You may want to include a covariate as suggested by my students when we did a debriefing of this activity such as previous athletic experience, gender (!), or comfort with ball throwing. For large classes, you can modify this to have passes occur between larger groups of students of 6-8, and online students should ask a friend for help to complete the activity or toss the ball against a wall.

Mixed Factorial Analysis of Variance Demonstration 3: An Expansion of the Speed-Accuracy Trade-Off: In this demonstration, students are tested under one of two treatment conditions, either speed or accuracy (i.e., the between subjects factor), but are also given three trials in order to assess whether or not ball passing improves with practice (i.e., the repeated measures factor). Again, completed passes will be the dependent variable. Have students pair up into groups of two. Give each group either the speed or accuracy instructions (see Appendix O). Students should pass the ball according to their instructions, but repeat the passing test three times for the repeated measures portion of the mixed factorial part of the design. Each passing test should last for 60 seconds. For large classes, modify this to have passes occur between groups of 6-8 students instead of pairs (you will need to spread out). For online courses, students could be assigned to a treatment condition and asked to find a friend or toss a ball against a wall.

![Figure 1. Students participating in the ANOVA ball toss activities](image-url)
Conclusion

Research suggests that students learn statistics and probability through active participation rather than through lecture alone (Gnanadesikan, Scheaffer, Watkins, & Witmer, 1997). Additionally, activity-based learning improves outcomes, especially when they are well thought out and reinforced at multiple time points (Garfield & Ben-Zvi, 2007). Our own experiences in the classroom with these techniques suggest that students do find them enjoyable but the ‘fun’ of the interactive components must extend beyond the activity into deep learning (Garfield & Ben-Zvi, 2007). This is illustrated by students’ reaction to the mixed-factorial ANOVA activity. One student indicated, “[my understanding] increased through the hands on activity.” Another, indicated, “it created a better understanding because I was able to see it happen in motion.” In order to instill deeper learning, following the activity students answered questions about mixed factorial designs, described between-within designs, and used a between-within output table to extract data and record the $F$ values. Students performed these tasks much more accurately following the ‘hands-on’ activity than they had prior to the activity, thus reinforcing previous reports that active learning improved comprehension. In sum, students have found these activities engaging and enjoyable, and their understanding of statistics benefits from the inclusion of these activities in class. As is illustrated above, many of these activities do build upon one another such as the t-test activity, which expands into the ANOVA and then Mixed Factorial ANOVA demonstrations. Similarly, the categorical activities utilizing M&Ms build on one another and illustrate concepts of dispersion and categorical matching which will help reinforce inherent concepts that are woven throughout the statistical curriculum.
References


sample of undergraduates. Learning and Individual Differences, 45, 214-221. doi: 10.1016/j.lindif.2015.12.019


Appendix A

Descriptives Demonstration 1: Using Things that Students Have on Hand

Materials Needed: Whatever students have on hand, data collection sheet, computer and projector (if desired)

Instructions: Have students count the number of a particular items such as, pens, pencils, cell phone apps, that they have on their person. Tabulate each students’ count on the data collection sheet. Calculate descriptive statistics such as mean, median, mode, standard deviation, variance, and range. It may be helpful to project the data collection sheet in front of the class.

Data Collection Sheet example

<table>
<thead>
<tr>
<th>How many social media minutes did you use on your phone this week? (convert to minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>367</td>
</tr>
<tr>
<td>453</td>
</tr>
<tr>
<td>221</td>
</tr>
<tr>
<td>…</td>
</tr>
</tbody>
</table>
Appendix B

Descriptives Demonstration 2: Snack Pack Variability and Counting Colors of M&Ms

Materials needed: one M&Ms “fun sized” pack (or other “fun sized” candy pack) per student, data collection sheets, computer and projector (if desired)

Instructions for individuals: Hand out one M&Ms “fun sized” pack per student. Have students count the number of candies in their pack. Record student counts on the data collection sheet. Calculate descriptive statistics such as mean, median, mode, standard deviation, variance, and range. It may be helpful to project the data collection sheet in front of the class.

<table>
<thead>
<tr>
<th>Student Names</th>
<th>Number of Candies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student 1</td>
<td></td>
</tr>
<tr>
<td>Student 2</td>
<td></td>
</tr>
<tr>
<td>Student 3</td>
<td></td>
</tr>
<tr>
<td>Student 4</td>
<td></td>
</tr>
<tr>
<td>….</td>
<td></td>
</tr>
</tbody>
</table>

Instructions for groups: Divide students into groups of four or five. Give each student a “fun sized” pack of M&Ms and give each group a data collection sheet. Ask students to count how many candies of each color are in their packs and then record this information on the data collection sheet. After that, have students answer the questions on their data collection sheet.

**Bonus Activity**

CANDY DATA DAY

Get in a group of 4-5 students and each count the different colors of M & Ms/skittles in your bag. Record the data and then feel free to eat it. Show your work! COPY YOUR DATA AGAIN ON PART 2 (page 2).

Part 1. This activity will help us get some more practice computing central tendency measures.

<table>
<thead>
<tr>
<th>Person’s name</th>
<th># of Red</th>
<th># of Orange</th>
<th># of Yellow</th>
<th># of Green</th>
<th># of Blue</th>
<th># of Brown/Purple</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Σ (soms)

1. What is the mean for each color?
2. What is the most frequent color (mode)?
3. What is the median for each color?
4. If 5 candies were added to the reds, what is the new mean for red?
5. What is the proportion of yellows (proportion = # of yellow/total candies)?
6. What is the percentage of orange candies (percentage = (# of orange/total candies)*100)?

**Part 2 Measures of Dispersion.** This activity will help us get some more practice computing variability measures.

<table>
<thead>
<tr>
<th>Person’s name</th>
<th># of Red</th>
<th># of Orange</th>
<th># of Yellow</th>
<th># of Green</th>
<th># of Blue</th>
<th># of Brown/Purple</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Σ (sums)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. What is the standard deviation for yellow candies?
2. What is the variance for red candies?
3. If 3 yellow candies were added to every bag, what would be the new mean and standard deviation for yellow candies?
4. If the number of yellow candies in every bag was multiplied by 3, what would the new mean and standard deviation for yellow candies be?
5. If every bag had the exactly the same number for each color of candies, then the standard deviations for each color will have a value of _____.


Appendix C

Probability Demonstration: Playing Card Probability

Materials needed: one deck of playing cards for each group of students, one data collection sheet per group

Instructions: Break students up into groups of four or five. Give each group a deck of playing cards and a data collections sheet. Ask students to answer the questions on the data collection sheet.

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the probability of drawing a king?</td>
<td></td>
</tr>
<tr>
<td>What is the probability of drawing a face card?</td>
<td></td>
</tr>
<tr>
<td>What is the probability of drawing the Ace of Hearts?</td>
<td></td>
</tr>
<tr>
<td>What is the probability of drawing a card that is not red?</td>
<td></td>
</tr>
<tr>
<td>Draw any card and calculate the probability that you would draw that card. Do not replace it in the deck. Repeat this procedure 10 times, calculating the new probability each time.</td>
<td></td>
</tr>
</tbody>
</table>
Appendix D

Sampling Demonstration 1: A “Mini” Sampling Distribution of Sample Means

Materials needed: one “fun sized” bags of M&Ms per student, access to random.org, prepared excel spreadsheet and a computer that can project it in front of the class

Instructions: Have students count off and distribute one “fun sized” bag of M&Ms to each student. Ask students to calculate the percentage of blue M&Ms in their bags. Have students call out their percentage of blue M&Ms, and record this in excel. Tell students that these percentages will serve as a population. Next, go to www.random.org (2019) and go to the “integer set generator.” Request 20 sets of 5 integers and set the value for each integer to be in between 1 and the total number of students in the class. Tell students that these sets will be used to create an incomplete sampling distribution. Read off the numbers in each set, and tell students to call out their percentage of blue M&Ms when they hear their number. Record these values in excel, as well. It is best to enlist the help of a student from the class with the recording process, as this part can be tedious. Have excel calculate the mean of each set, and also display the means in a histogram. Tell students that the histogram is an incomplete sampling distribution of sample means. It is best to have a spreadsheet set up to do these things automatically (or use the provided excel spreadsheet). Evaluate the data to determine a) if the mean of the sampling distribution of sample means is equal to the population mean (i.e., is it an unbiased estimator), b) if the sampling distribution of sample means appears to be normally distributed (i.e., does it follow the central limit theorem), and c) is the standard error, as calculated from population data (σ/\sqrt{n}), equal to the standard deviation of the sampling distribution of sample means.
Appendix E

Sampling Demonstration 2: Sample Size – Why Bigger is Better

Materials needed: One “fun sized” pack of M&Ms per student, a computer that can project distribution center distributions in front of the class, one data collection sheet per student.

Instructions: Ask students to read the article the article by Purtil (2015) in Quartz (available at https://qz.com/918008/the-color-distribution-of-mms-as-determined-by-a-phd-in-statistics/). This can be assigned as homework to save class time. Next, distribute packs of M&Ms to students and have them calculate their percentage of M&Ms of each color and record that information on the data collection sheet. Project the distributions used by the New Jersey factory and the Tennessee Factory (see below) on the board in front of students and ask them to note the differences between those distributions. Ask students to make a guess about the factory of origination for their M&Ms and record it on the data sheet. Next, have them pool their M&M sample with the sample of their neighbor, recalculate the percentage of each color, make another guess, and record it on the data collection sheet. Finally, have the entire class pool their M&Ms, recalculate percentages, and make a guess about the factory of origination. Then have students answer the questions on the data collection sheet. Determine whether their guesses about factory of origination are correct and discuss the answers to the questions as a class.
Student worksheet

Read this article: https://qz.com/918008/the-color-distribution-of-mms-as-determined-by-a-phd-in-statistics/

What Rick Wicklin did is a type of sampling – he drew a sample (the 712 M&M’s), looked at the proportions that exist in the sample, and then drew a conclusion about which Mars™ Distribution location was supplying SAS with their M&M’s. You are now tasked with doing the same thing! You have been given a pack of M&M’s. Open the pack but do not eat any before you count the colors (so help me if you eat any before you count up the colors…)! Next, determine the percentage of each color of M&M that is in your pack. To do this take the number of each color, and divide by the total (for example, if you have 5 blue M&M’s, and 20 total M&M’s, divide 5/20 to get 25%). Then, make a guess about which distribution center distributed your M&M’s came from.

Blue percentage:
Orange percentage:
Green percentage:
Yellow percentage:
Red percentage:
Brown percentage:

**Distribution Center Guess:**

Next, pool your raw data (i.e., numbers of each color of M&M’s) with the person sitting next to you. Re-calculate your percentages and see if your answer changes.

Blue percentage:
Orange percentage:
Green percentage:
Yellow percentage:
Red percentage:
Brown percentage:

**Distribution Center Guess:**

Finally, as a class, pool all of your data (this may require each person reading off their raw data and one person tabulating it) and make another guess about the distribution center.

Blue percentage:
Orange percentage:
Green percentage:
Yellow percentage:
Red percentage:
Brown percentage:

Distribution Center Guess:

Questions:

1. How did your percentages change as you combined your sample with your classmates’ samples?

2. Did combining your sample with your classmates’ samples make you more or less confident in your guess about the distribution center?

3. Why do bigger samples lead to more reliable estimates?

4. Which M&M color do you think was the most predictive of the production facility? Why?
Appendix F

One Sample t-test Demonstration 1: Does the Mars™ Lie About How Many M&Ms are in a Bag?

Materials needed: one “fun sized” M&Ms pack per student, data collection sheet, computer and projector to display the data collection sheet in front of the class (if desired)

Instructions: Distribute a “fun sized” M&Ms pack to each student. Walk students through the calculations necessary to determine how many M&Ms should be in each package, according to the values provided by Mars™. Tell students to count the number of M&Ms in each of their packs, and record that information on the data collection sheet. Have students calculate the mean and standard deviation of the class data, and ultimately perform a one independent sample t-test, using 21 as the comparison value.

<table>
<thead>
<tr>
<th>Student</th>
<th># M&amp;Ms per pack</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

... ... ...

Mean =
Standard Dev. =
Standard error =
Appendix G

Independent samples t-test Demonstration 1: M&Ms versus Skittles: Which has More?

Materials needed: one “fun sized” pack of M&Ms per student, one “fun sized” pack of skittles per student, one data collection sheet

Instructions: Distribute one “fun sized” pack of M&Ms and one “fun sized” pack of Skittles to each student. Ask them to count the number of candies in each pack, and record their counts on a data collection sheet that is projected on a screen in front of the class. Ask students to calculate means and standard deviations, and ultimately perform an independent samples t-test to determine which product contains more candies in “fun sized” packs.

<table>
<thead>
<tr>
<th>Type of Candy</th>
<th>Total Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>M&amp;M</td>
<td></td>
</tr>
<tr>
<td>M&amp;M</td>
<td></td>
</tr>
<tr>
<td>M&amp;M</td>
<td></td>
</tr>
<tr>
<td>M&amp;M</td>
<td></td>
</tr>
<tr>
<td>....</td>
<td></td>
</tr>
<tr>
<td>Mean =</td>
<td></td>
</tr>
<tr>
<td>Standard Dev. =</td>
<td></td>
</tr>
<tr>
<td>Standard error</td>
<td></td>
</tr>
<tr>
<td>Skittles</td>
<td></td>
</tr>
<tr>
<td>Skittles</td>
<td></td>
</tr>
<tr>
<td>Skittles</td>
<td></td>
</tr>
<tr>
<td>Skittles</td>
<td></td>
</tr>
<tr>
<td>....</td>
<td></td>
</tr>
<tr>
<td>Mean =</td>
<td></td>
</tr>
<tr>
<td>Standard Dev. =</td>
<td></td>
</tr>
<tr>
<td>Standard error</td>
<td></td>
</tr>
</tbody>
</table>
Appendix H

Independent Samples t-test Demonstration 2: Altered Perception Ball Passing Accuracy

**Materials needed:** one small ball (such as a tennis ball*) for each group of two students, one pair of vision altering glasses for each member of half of the groups, stopwatch or clock that indicates seconds, one data collection sheet

**Instructions:** Divide students into groups of two. Give all groups a ball and half of the groups vision altering glasses. Have group members stand 10 feet apart (or as far apart as space allows). Tell them to pass the ball back and forth as many times as they can while you time them. Make sure to remind them to keep count of the number of times that they have successfully passed the ball. Record the number of successful passes of both group types on the data collection sheet. Have students calculate means and standard deviations, and ultimately preform an independent samples t-test to determine if the glasses impacted passing accuracy.

<table>
<thead>
<tr>
<th>Perception</th>
<th>Total Number of successful passes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-altered</td>
<td></td>
</tr>
<tr>
<td>Non-altered</td>
<td></td>
</tr>
<tr>
<td>Non-altered</td>
<td></td>
</tr>
<tr>
<td>Non-altered</td>
<td></td>
</tr>
<tr>
<td>Non-altered</td>
<td></td>
</tr>
<tr>
<td>....</td>
<td></td>
</tr>
</tbody>
</table>

Mean =
Standard Dev. =
Standard error =

<table>
<thead>
<tr>
<th>Perception</th>
<th>Total Number of successful passes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altered</td>
<td></td>
</tr>
<tr>
<td>Altered</td>
<td></td>
</tr>
<tr>
<td>Altered</td>
<td></td>
</tr>
<tr>
<td>Altered</td>
<td></td>
</tr>
<tr>
<td>....</td>
<td></td>
</tr>
</tbody>
</table>

Mean =
Standard Dev. =
Standard error =

*starter tennis balls are slightly larger and a little softer than regular tennis balls and have a foam like consistency. They can be purchased online or at any sporting goods store.
Appendix I

Repeated Measures t-test Demonstration 1: Speed/Accuracy Trade-Off

Materials needed: one small ball (such as a tennis ball*) for each group of two students, one set of speed instructions per group, one set of accuracy instructions per group, stopwatch or clock that indicates seconds, one data collection sheet

Instructions: Put students in groups of two and give each group a ball. Tell students that they will be passing a ball back and forth for 60 seconds, and that they will do this twice, with a slightly different set of instructions each time. Give half the groups the speed instructions and half of the groups the accuracy instructions. Have group members stand 10 feet apart (or as far as space allows). Make sure to remind them to keep count of the number of times that they have successfully passed the ball. Record each group’s number of successful passes on the data sheet. It would be helpful to project the data sheet in front of the class, if possible. After the first session, have the groups switch their instructions, and then do the second sessions (again, reminding them to count the number of successful passes). Again, record each group’s number of successful passes on the data sheet. Have students calculate difference scores, the standard deviation of the difference scores, group means, and ultimately do a repeated measures t-test to determine under which condition students made the most successful passes.

Speed Instructions: For this fun ball playing game, you will pass the ball back and forth as many times as possible, as quickly as possible, meaning with as much speed as possible, trying to make as many passes as possible in the time you have. Count your passes. Your instructor will time the trials for 60 seconds each.

Accuracy Instructions: For this fun ball playing game, you will pass the ball back and forth as many times as possible, as accurately as possible, meaning with as much precision as possible, trying not to drop the ball. Count your passes. Your instructor will time the trials for 60 seconds each.

Repeated measures t-test data recording sheet:

<table>
<thead>
<tr>
<th>Group</th>
<th>Speed Condition Passes</th>
<th>Accuracy Condition Passes</th>
<th>Difference scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

Mean =  
Standard Dev. =  
Standard error =  

*starter tennis balls are slightly larger and a little softer than regular tennis balls and have a foam like consistency. They can be purchased online or at any sporting goods store.
Appendix J

*Chi-Square Test for Goodness of Fit Demonstration 1: Do Our M&Ms fit the Population Parameters?*

*Materials needed:* one large bag of M&Ms (or several large bags for larger classes), small cups to hold M&Ms, one data collection sheet (or excel spreadsheet)

*Instructions:* Tell students that you are going to determine if the proportions of M&Ms from your bag match the proportions that Mars™ reports producing (and show them the below proportions). Place students into groups of two or three and give each group a cup. Fill each cup with some M&Ms, and instruct students to count the number of each color of M&M in their cup. Project the data collection sheet in front of the class (an excel spreadsheet would also work well for this demonstration), and record the data from each group. Calculate the total number of each color of M&Ms (these will be the observed frequencies). Next, have students calculate the expected frequencies by multiplying the percents that Mars™ claims to produce for each color by the total number of M&Ms (e.g., to calculate the expected frequency of blue M&Ms with 150 total M&Ms you would multiply 150 by .24). Ultimately, have students complete a Chi-Square Test for Goodness of Fit to see if the proportions in your bag of M&Ms match those reported by Mars™.

<table>
<thead>
<tr>
<th>Group</th>
<th>Blue</th>
<th>Orange</th>
<th>Green</th>
<th>Yellow</th>
<th>Red</th>
<th>Brown</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Factory Output percentages

<table>
<thead>
<tr>
<th>Blue</th>
<th>Orange</th>
<th>Green</th>
<th>Yellow</th>
<th>Red</th>
<th>Brown</th>
</tr>
</thead>
<tbody>
<tr>
<td>24%</td>
<td>20%</td>
<td>16%</td>
<td>14%</td>
<td>13%</td>
<td>13%</td>
</tr>
</tbody>
</table>
Appendix K

Chi-Square Test for Goodness of Fit Demonstration 2: Buzzfeed Oreo “Facts”

Materials needed: one Oreo Cookie for each student, one data collection sheet (or excel spreadsheet)

Instructions: Show the Buzzfeed video “9 Things About Oreos You Didn’t Know” (available from: https://www.youtube.com/watch?v=fV6QyzU5DsU or by searching for the name of the video in youtube.com). While students are watching the video, hand out Oreo cookies. After students have seen the video and eaten their cookies, record their gender and whether they twisted their Oreo cookie or ate it whole (tell them that they must classify themselves simply as a twister or non-twister – there is no middle ground). This data will be your observed values for men and women, and proportions from the video for men and women will be used to create your expected values (e.g., multiply the total number of women by .41 to get the expected frequency of female twisters). Have students calculate the expected frequencies, and then perform two Chi-Square Goodness of Fit Tests (one for each gender) to determine whether the proportions from class match those reported in the video.

Data Collection Sheet

<table>
<thead>
<tr>
<th></th>
<th>Male Frequencies</th>
<th></th>
<th>Female Frequencies</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Total</td>
<td>Observed</td>
<td>Total</td>
</tr>
<tr>
<td>Twister</td>
<td>Expected .16 x (total number of males)</td>
<td></td>
<td>Expected .41 x (total number of females)</td>
<td></td>
</tr>
<tr>
<td>Non-twister</td>
<td>Observed .84 x (total number of males)</td>
<td></td>
<td>Observed .59 x (total number of females)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>Total</td>
<td></td>
<td>Total</td>
</tr>
</tbody>
</table>
Appendix L

Correlation Demonstration 1: Distance and Accuracy

**Materials needed:** masking tape, two tape measures, a homemade target, one hat, prepared slips of paper (see below), a beanbag, one data collection sheet

**Instructions:** Create a homemade target by cutting an 18-inch diameter circular piece of poster board and drawing a red circle in the center. Put this on the floor and tape it in place with the masking tape to ensure that it does not move during the activity. Lay one tape measure extending straight out from the target. The tape measure should be extended to at least: 3 x (the number of students in the class) inches. For example, if there are 20 students in the class the tape measure should be extended to 60 inches. Prepare one small slip of paper for each student in the class. On each piece of paper, write a multiple of three (e.g., 3, 6, 9, 12,…). For larger classes it is recommended that you duplicate distances so that students do not end up standing too far from the target. Put the pieces of paper in a hat. Have students line up and each take a turn throwing the beanbag to try to hit the target. Before each student’s turn, have that student draw a slip of paper from the hat. The number on their slip of paper will determine how many inches that student must stand from the target (as measured by the tape measure). After students take their turn trying to hit the target, measure how many inches the beanbag fell from the target using the other tape measure. Record the distance that students stood from the target and the distances that beanbags fell from the target on a data collection sheet that is projected in front of the class. After all of the data has been collected, have students calculate a correlation coefficient to determine the relationship between distance and accuracy.

<table>
<thead>
<tr>
<th>Student Distance from Target</th>
<th>Beanbag distance from Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>6”</td>
<td></td>
</tr>
<tr>
<td>12”</td>
<td></td>
</tr>
<tr>
<td>18”</td>
<td></td>
</tr>
<tr>
<td>24”</td>
<td></td>
</tr>
<tr>
<td>30”</td>
<td></td>
</tr>
<tr>
<td>36”</td>
<td></td>
</tr>
<tr>
<td>……</td>
<td></td>
</tr>
</tbody>
</table>

\[ r = \]

Appendix M

Analysis of Variance Demonstration 1: The Importance of Binocular Vision

Materials needed: one small ball (such as a tennis ball*) for each pair of students, stopwatch or clock that indicates seconds, one data collection sheet (or excel spreadsheet)

Instructions: Place students into pairs and give each pair a ball. Have students stand 10 feet apart (or as far apart as space allows). Tell students that they are going to pass the ball back and forth three different times, once with their right eye closed, once with their left eye closed, and once with neither eye closed. If using a counterbalancing procedure, at this time tell each group the order in which it is to go through the conditions. Remind students to count the number of successful passes that they make. Record the number of successful passes made by each group immediately after each passing session on the data collection sheet, which could be projected in front of students. Have students complete a within subjects ANOVA on this data to determine the effects of binocular vision on ball throwing ability.

<table>
<thead>
<tr>
<th></th>
<th>Right eye open</th>
<th>Left eye open</th>
<th>Both eyes open</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>…</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>Mean</td>
<td>Mean</td>
</tr>
<tr>
<td>Standard error</td>
<td>Standard error</td>
<td>Standard error</td>
<td></td>
</tr>
</tbody>
</table>

*starter tennis balls are slightly larger and a little softer than regular tennis balls and have a foam like consistency. They can be purchased online or at any sporting goods store.
Appendix N

Analysis of Variance Demonstration 2: Speed, Accuracy, and Underhanded Pitches

Materials needed: a small ball (such as a tennis ball*) for each pair of students, printouts of passing instructions, stopwatch or clock that indicates seconds, one data collection sheet

Instructions: Place students in pairs and give each pair of students a ball. Divide the pairs of students into three different groups. Have students stand 10 feet apart (or as far apart as space allows). Give each group either speed instructions, accuracy instructions, or speed and accuracy instructions (see below). Remind students to count their number of successful passes, and then give them 90 seconds to see how many successful passes they can make. Record the number of successful passes from each group on the data collection sheet (it may be helpful to project this in front of the class). Complete a one way between subjects ANOVA to determine which set of instructions guides groups to make the most successful passes.

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of successful passes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed group 1</td>
<td></td>
</tr>
<tr>
<td>Speed group 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy group 1</td>
<td></td>
</tr>
<tr>
<td>Accuracy group 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Underhanded group 1</td>
<td></td>
</tr>
<tr>
<td>Underhanded group 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Speed Instructions: For this fun ball playing game, you will pass the ball back and forth as many times as possible, **as quickly as possible, meaning with as much speed as possible**, trying to make as many passes as possible in the time you have. **Count your passes.** Your instructor will time the trials for 90 seconds each.

Accuracy Instructions: For this fun ball playing game, you will pass the ball back and forth as many times as possible, **as accurately as possible, meaning with as much precision as possible**, trying not to drop the ball. **Count your passes.** Your instructor will time the trials for 90 seconds each.

Speed and Accuracy Instructions: For this fun ball playing game, you will pass the ball back and forth as many times as possible, **as accurately and as quickly as possible, meaning that you should maximize both speed and accuracy**, trying not to drop the ball. **Count your passes.** Your instructor will time the trials for 90 seconds each.

*starter tennis balls are slightly larger and a little softer than regular tennis balls and have a foam like consistency. They can be purchased online or at any sporting goods store.
Appendix O

Analysis of Variance Demonstration 3: An Expansion of the Speed Accuracy Trade-Off

Materials needed: one small ball (such as a tennis ball*) for each pair of students, stopwatch or clock that indicates seconds, data collection sheet

Instructions: Put students in pairs and give each pair a ball. Divide the pairs in half. Have students stand 10 feet apart (or as far apart as space allows). Tell students that they will be passing a ball back and forth under two different sets of instructions, and that they will be doing three different trials to see if their number of completed passes improves with practice. Next, give the pairs their instructions (see below), and remind students to count the number of completed passes that they made during each trial. Record the number of completed passes that each pair made in the data collection sheet at the after each trial. It may be helpful to project the data collection sheet in front of the class. At the conclusion of the three trials complete a mixed factorial ANOVA to see the effects of instruction on ball passing, the effects of practice on ball passing, and whether or not instructions and practice have interactive effects on ball passing. We also include below a brief assessment that we have used in the pass to determine the impact this activity has on student understanding of mixed factorial ANOVA.

<table>
<thead>
<tr>
<th></th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed group 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed group 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed group 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy group 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy group 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy group 3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Speed Instructions: For this fun ball playing game, you will pass the ball back and forth as many times as possible, **as quickly as possible, meaning with as much speed as possible**, trying to make as many passes as possible in the time you have. **Count your passes.** Your instructor will time the trials for 60 seconds each.

Accuracy Instructions: For this fun ball playing game, you will pass the ball back and forth as many times as possible, **as accurately as possible, meaning with as much precision as possible**, trying not to drop the ball. **Count your passes.** Your instructor will time the trials for 60 seconds each.

Assessment of speed-accuracy trade-off for mixed-factorial ANOVA

1. What is a mixed factorial design?
2. Describe a between-within design
3. How did your understanding of mixed factorial design change after the activity with the speed-accuracy activity?
4. What questions do you still have about ANOVA and between-within design?
5. Present a mixed factorial ANOVA table and have students write all of the results of the ANOVAs

*starter tennis balls are slightly larger and a little softer than regular tennis balls and have a foam like consistency. They can be purchased online or at any sporting goods store.
Using StatHand to Improve Students’ Statistic Selection Skills

Peter J Allen, PhD, Jessica L Fielding, PhD, Ryan H S Kay, and Elizabeth C East
University of Bristol

Summary

Psychology undergraduates find identifying appropriate analyses for common research designs difficult. Resources have been developed to aid this process, including decision trees commonly included in statistics textbooks. The use of such trees is supported by research demonstrating their efficacy and popularity. In recent years, decision trees to aid statistic selection have been adapted for digital media. One such adaptation is StatHand, a free iOS and web app that aids statistic selection by prompting users to focus systematically on each structural feature of their research design. Previous research has suggested that simply providing students with an app like StatHand is not enough to promote accurate statistic selection. Rather, students need to be trained in its use. In this chapter we describe a brief statistic selection training activity built around the use of StatHand. The development of the activity was informed by two sets of literature. The first suggests that accurate statistic selection is a consequence of ‘structural awareness’. The second pertains to the success of ‘wise’ psychological interventions across a range of contexts, including education. The students we have trained using our methods (N = 50) demonstrated substantially greater statistic selection proficiency than untrained students in previous research. Our training methods can be adapted for a range of contexts. The chapter appendices include our training materials and over 40 research scenarios spanning the range of analyses covered in StatHand. These can be freely adapted by instructors for both formative and summative learning activities.

Statistic Selection Skills

One of the five learning goals for an undergraduate psychology degree specified by the Society for the Teaching of Psychology’s Statistical Literacy Taskforce (2014, p. 2) is the ability to “apply appropriate statistical strategies to test hypotheses”. In meeting this goal, students should be able to “select …an appropriate statistical analysis for a given research design, problem, or hypothesis”. For most, this is a difficult task. Research indicates that psychology students are not good at recalling, recognizing or explaining how they would select appropriate statistical analyses for common research scenarios.

To illustrate the recall deficit, Gardner and Hudson (1999) gave students a series of typical research scenarios and asked them to specify an appropriate statistical analysis for each. Although the students were in third-year or above, they were unable to name an appropriate analysis for most scenarios. Indeed, even the highest performing student had an accuracy level of just 56%. To illustrate the recall deficit, in a multiple-choice selection task that Ware and Chastain (1989) administered at the end of a first-year psychology statistics unit, students averaged less than 45%. This was despite the researchers’ (and their colleagues’) initial beliefs that the task was easy enough for a typical first-year student to complete successfully. Finally, to illustrate the explanation deficit, Allen, Dorozenko, and Roberts (2016) asked undergraduate psychology students to describe how they would select appropriate statistical analyses for
research scenarios similar to those developed by Gardner and Hudson (1999) and Ware and Chastain (1989). These students, who had each completed an average of three research methods courses, described haphazard and inefficient selection strategies that were unlikely to reliably lead them to appropriate analyses.

Recognition of these deficits has led educators to develop resources to facilitate the statistic selection process. Foremost amongst these resources are decision trees, which are routinely included in statistics textbooks (e.g., Allen, Bennett, & Heritage, 2019; Nolan & Heinzen, 2017). The proliferation of such trees is supported by research demonstrating their objective efficacy and subjective appeal (Carlson, Protsman, & Tomaka, 2005; Protsman & Carlson, 2008). However, despite their popularity, traditional paper-based decision trees are not without limitations. The most obvious of these limitations is brevity. They typically need to fit onto a single sheet of paper, meaning that information that would aid their navigation (definitions, examples etc.) is either spatially separated from the tree, or entirely absent.

Figure 1. StatHand home screen on the Chrome mobile web-browser.

To overcome such limitations, decision trees to aid statistic selection have been adapted for digital media (e.g., Koch & Gobell, 1999). One recent adaptation is StatHand (Allen et al., 2016, 2017). StatHand is a free iOS (available via the iOS App Store) and web app (see https://stathand.net) that aids the process of selecting appropriate statistical analyses for a wide range of circumstances. It achieves this by asking a series of questions that prompt users to focus systematically on each structural feature of their research design. The questions are annotated with relevant definitions and examples, such that relative novices can navigate the app without needing to consult additional sources. In answering these questions, the user progressively homes in on a statistical analysis appropriate to their research design. The StatHand web app running on a mobile web-browser is illustrated in Figure 1, and the iOS version running on an iPad is illustrated in Figure 2.

In a recent experimental evaluation 217 psychology students were randomized to four different decision-making aids (StatHand, a familiar textbook, a familiar paper decision tree, or the textbook and decision tree combined) and asked to identify appropriate statistical analyses for five research scenarios (Allen, Finlay, Roberts, & Baughman, 2019). The students in the StatHand condition significantly outperformed students in the other three conditions ($d = .55$ to
However, in an absolute sense, their performance was underwhelming. On average, they identified appropriate analyses for just 1.74 ($SD = 1.19$) of the five scenarios. On a typical university marking scale this would be a firm ‘fail’. This suggests that simply providing students with a tool like StatHand may not be enough to promote accurate statistic selection. Rather, students need to be trained in its use.

The next section of this chapter describes a brief statistic selection training activity built around the use of StatHand. The development of the activity was informed by two sets of literature. The first suggests that accurate statistic selection is a consequence of ‘structural awareness’, which can itself be trained (e.g., Yan & Lavigne, 2014). Structural awareness refers to the ability to see past the topic area of a research problem, and focus on its structural characteristics (e.g., the number and nature of the independent and dependent variables) and the relationships between them (Quilici & Mayer, 2002). The second pertains to the success of ‘wise’ psychological interventions across a range of contexts, including education (Walton, 2014). Wise interventions are brief and targeted. They are designed to change specific behaviors (in both the short and longer term) by exploiting specific psychological processes. In this instance, those processes are meta-cognition and structural awareness.

**The Training Activity**

According to the wise framework proposed by Walton (2014), there are three elements to consider when developing a brief psychological intervention: (1) a clear and specific underlying theory/concept; (2) the recursive process that is being targeted or broken; and (3) context. Research has shown that students will enlist various cognitive and meta-cognitive strategies when learning new skills/processes (Somuncuoglu & Yildirim, 1999). However, it has become increasingly clear that when learning statistics, students are driven by a fear of failure and take a more tactical surface approach to gaining good grades (e.g., Asikainen & Gijbels, 2017; Diseth & Martinsen, 2003; Newble & Entwhistle, 1986). This contrasts with using deeper learning approaches which are considered to promote deeper semantic understanding and a lasting learning experience (e.g., Bilgin & Crowe, 2008). Meta-cognition is closely tied to deeper learning as it goes beyond simple cognition (*how we think*) and is characterized by our awareness of how we think. This deeper appreciation allows not only semantic understanding of concepts, but synthesis of ideas and the ability to adapt and apply this understanding in different situations (e.g. see Bayat & Tarmizi, 2010).

Our training focuses specifically on developing students’ structural awareness. That is, their ability to identify and focus on the structural (e.g., IV, DV etc.) characteristics of research designs, rather than the research topic or other surface level features. We designed the intervention to break students’ habits of utilizing surface level learning strategies and to encourage them to be more structurally aware, thus facilitating deep learning. The final element of a wise intervention is that it needs to be context specific – it needs to be important to the individual for it to be assimilated and have a lasting impact. Embedding this training within the curriculum automatically makes it inherently important to students. However, if not embedded within the curriculum, but perhaps used as a supplementary learning activity, the training has been designed to provide students with feedback on their progress. This ensures that they see the benefits of using a more structurally aware approach to statistic selection, which encourages this self-motivation and self-reflection (see chapter 5 on reflection).

There are two phases to the training with students; a teacher-guided training phase and then a self-guided practice phase. The teacher-guided training and self-guided practice each take around 20-25 minutes to complete. Depending on how these methods are adapted, we envisage they would be suitable for a standalone lecture or class lasting no longer than one hour.
Teacher-Guided Training Phase

The teacher-guided training begins with presenting one of the four research scenarios in Appendix A to students and asking if they know how the resultant data might be analyzed. Our experience (as well as previous research; Allen, Dorozenko, & Roberts, 2016) suggests that most undergraduate students will either indicate that they do not know or will guess the statistical test that they have used most recently or most frequently. This provides an opportunity to explain that there is a systematic process that can be followed to work out how a set of data could be analyzed, and that there are many resources (e.g., flow diagrams, websites etc.) that will ‘step you through’ this process. (On the occasions students do identify an appropriate analysis, these resources can be pitched as tools that can be used to ‘double check’ or confirm their thinking.)

At this point, we introduce students to the StatHand app and demonstrate how it can be used to identify an appropriate analysis for the study in the scenario. This process involves highlighting various features of the study’s design (e.g., number and nature of dependent and independent variables) and outlining how they correspond to the options and examples in StatHand. Once an appropriate statistical analysis has been identified, we use the StatHand History tool (see Figure 2) to reiterate why it was selected (e.g., because we had one interval or ratio level dependent variable and one independent variable with two independent levels). In doing so, we are highlighting the structural characteristics of the research scenario. The ability to identify such characteristics is key to the notion of structural awareness. Asking why also encourages students to reflect on the reasons behind particular choices, which is an important meta-cognitive skill and promotes the importance of enlisting a deeper approach to understanding as an effective decision-making process (Walton, 2014, see also chapter 5 on reflection). This can also be a good time to illustrate how changing one structural feature (e.g., from independent to related samples) changes the analysis (in Figure 2, from an independent-samples to a paired-samples t-test). Finally, we demonstrate how our research scenario is structurally equivalent to the example in StatHand, even though their topics are quite different. We call this a ‘mapping exercise’. For example, the ratio level dependent variable in the first scenario in Appendix A is words recalled. In the example in Figure 2, it is drinks consumed. Again, the ability to see past the surface/topic characteristics of a research scenario and focus on its structural characteristics is key to structural awareness.

Students are encouraged to work more independently on the remaining three scenarios in Appendix A, receiving reduced instruction and feedback as they progress (i.e., we will talk through earlier scenarios but only confirm answers for later ones). This approach creates a safe learning environment where students are encouraged to think and work independently but are
not afraid to make mistakes in the learning process. We use a handout to guide this training (see Figure 3), although PowerPoint should be similarly effective.

**Self-Guided Practice Phase**

After the teacher-guided training phase of our activity, we give students another four scenarios (see Appendix B) formatted in a handout like the one illustrated in Figure 3. Students are instructed to work through this handout independently and told they won’t receive feedback from the teacher on whether the statistics they identify are correct. This part of the wise intervention process encourages students to assimilate their learning from the training phase when there is no external motivation to find the correct answer. It is clear in wise interventions that the behavior change has to be salient for the individual to have a lasting and effective impact.

Note that we have deliberately restricted our activity to just four tests, which differ on just two structural characteristics (see Figure 4). This was because our intent was to highlight to students the importance of attending to the structural features of research designs, not to teach an extensive range of different tests and procedures. However, teachers may wish to swap the scenarios in Appendices A and B with any of the 41 scenarios in Appendix C, which cover the full range of analyses described in StatHand. The scenarios in Appendix C can also be freely used or adapted by educators for a range of additional formative and summative learning activities.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>Nominal</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-square test of contingencies</td>
<td>Independent samples t-test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>McNemar test of change</td>
<td>Paired samples t-test</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4. The tests used in our training activity.**

**Our Findings**

We ran our training with \(N = 50\) first-, second- and third-year psychology students, and coded their responses to each element of our self-guided practice handout. Answers were coded as correct, incorrect or absent, although incorrect and absent were merged for the inferential analyses.

Our students were able to correctly identify appropriate statistical tests for 81% (\(SD = 25\%\)) of the scenarios in the practice handout. This was significantly and substantially higher than the 34.8% accuracy level achieved by the untrained students in the Allen et al. (2019) sample, \(t(49) = 12.72, p < .001, d = 1.85\). Further analyses indicated that our trained students were significantly better able to identify appropriate statistics for some scenarios than they were for others, Cochran’s Q (3, \(N = 50\)) = 15.48, \(p = .001\). Specifically, they were significantly less likely to identify an appropriate statistic for the paired-samples t-test scenario (64%) than they were for the independent samples t-test (88%) and chi-square (88%) scenarios (Bonferroni corrected \(p = .004\)). Identification accuracy for the McNemar test scenario (82%) was also higher than that for the paired-samples t-test scenario, though not significantly so (Bonferroni corrected \(p = .065\)).

When asked “why did you choose this test?”, our students correctly identified an average of 85.4% of the relevant structural characteristics for each scenario. However, as illustrated in Table 1, performance levels for some scenarios and characteristics were higher than for others. In particular, students seemed less able to correctly identify the structural characteristics of the paired samples t-test scenario, with several of them confusing the number of levels of the IV (aka. the number of data sets) with the number of IVs, which led them to a factorial, rather than
a one-way design. This may be an artefact of the nature of the scenario we used, which was perhaps less ‘typical’ than the scenarios used for the other three tests. However, this possibility requires further investigation.

Table 1. Correctness of Students’ Responses to the Question, “Why Did You Choose This Test”, Split by Scenario and Structural Characteristic

<table>
<thead>
<tr>
<th>Percentage of Correct Responses</th>
<th>M</th>
<th>Q</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. DVs</td>
<td>DV Data Type</td>
<td>No. IVs</td>
<td>No. Data Sets</td>
</tr>
<tr>
<td>Independent samples t-test</td>
<td>96\textsuperscript{a}</td>
<td>82\textsubscript{ab}</td>
<td>94\textsuperscript{a}</td>
</tr>
<tr>
<td>Paired samples t-test</td>
<td>88\textsuperscript{a}</td>
<td>68\textsubscript{a}</td>
<td>76\textsubscript{ab}</td>
</tr>
<tr>
<td>Chi-square test of contingencies</td>
<td>94\textsuperscript{a}</td>
<td>86\textsubscript{ab}</td>
<td>94\textsuperscript{a}</td>
</tr>
<tr>
<td>McNemar test of change</td>
<td>92\textsuperscript{a}</td>
<td>86\textsubscript{b}</td>
<td>88\textsubscript{ab}</td>
</tr>
<tr>
<td>M</td>
<td>92.5</td>
<td>80.5</td>
<td>88.0</td>
</tr>
<tr>
<td>Q</td>
<td>3.18</td>
<td>10.77</td>
<td>14.09</td>
</tr>
<tr>
<td>p</td>
<td>.364</td>
<td>.013</td>
<td>.003</td>
</tr>
</tbody>
</table>

Note. Percentages on the same row with the same superscript letters do not differ significantly at a Bonferroni corrected $\alpha$ of .05. Percentages in the same column with the same subscript letters do not differ significantly at a Bonferroni corrected $\alpha$ of .05. These pairwise comparisons are McNemar tests of change. $M$ = Mean percentage of correct responses for the relevant scenario or structural characteristic. $Q$ = Cochran’s Q for differences between percentages of correct responses across scenarios or structural characteristics. $p$ = $p$-values associated with reported Cochran’s Q values.

Figure 5. Percent of correct, incorrect and absent answers in the mapping exercise for the (A) independent samples t-test, (B) paired samples t-test, (C) chi-square test of contingencies and (D) McNemar test of change scenarios.

Finally, in the mapping exercise, students correctly matched 64.5% of the Appendix B scenarios’ structural characteristics with their corresponding surface/topic characteristics and with the corresponding surface/topic characteristics of the StatHand examples.
As illustrated in Figure 5, there was relatively little variation in average correctness across the four scenarios, $F(3, 147) = 1.90, p = .131$, partial $\eta^2 = .04$. The larger number of ‘absent’ responses for the paired-samples $t$-test scenario can be attributed to our coding. We automatically converted relevant responses to absent in instances where an incorrect test had been identified.

Final Remarks

This chapter describes an activity we developed to help students use StatHand to select appropriate statistical analyses for different research designs. The data we collected from $N = 50$ undergraduate psychology students suggested that this activity is effective. Specifically, compared to the untrained students in the Allen et al. (2019) sample, our students were substantially better able to identify appropriate statistics for four common research designs. They were also proficient at declaring the structural characteristics that led them to the selection of particular analyses. Finally, the majority were capable of correctly matching the structural characteristics of different research scenarios with their corresponding surface/topic characteristics, as well as the surface/topic characteristics of the examples in StatHand. However, there was more room for improvement on this task.

We trained our students individually, as they were participants in a larger randomized controlled trial (RCT). However, we see no reason why this activity can’t be adapted for use in small group teaching and/or lectures. In such contexts, students could initially work individually, and then pair up to compare and discuss their answers prior to a class-level discussion. Furthermore, there are no compelling reasons why educators need to print handouts (PowerPoint should be equally effective, and far more environmentally conscious), restrict their lessons to just four statistical analyses (there are scenarios reflecting 10 times this number in Appendix C, including all the analyses in the Passion-Driven Statistics curriculum, see chapter 5) or even deliver these lessons face-to-face (guided online tutorials, see chapter 2, or homework activities with pre-programmed feedback for use in a flipped classroom context, see chapter 3, could be developed relatively easily). The important point is that students are prompted to focus systematically on the structural characteristics of the research designs they are exposed to, and the implications of these characteristics for selecting appropriate statistics.

Beyond their use in undergraduate research methods and statistics classes, StatHand and the accompanying training materials may be useful in a range of contexts where people need to consider (or double-check) how quantitative data could or should be analyzed. For example, given research demonstrating that even higher-level students have limited statistic selection skills and confidence (e.g., Allen, Dorozenko, & Roberts, 2016; Gardner & Hudson, 1999), students undertaking capstone or dissertation projects may find the resources we have developed to be useful. So might students undertaking research internships, UREs (undergraduate research experiences) or service-learning projects (see chapter 7), or students who have moved into psychology as mature learners and/or from non-science disciplines (e.g., students on MSc conversion courses in the UK). Finally, our experience suggests that some colleagues, particularly those with limited statistical training, may also benefit from StatHand and the opportunity to practice using it.

Although the findings reported herein are pleasing, they raise a number of new questions requiring investigation. First, how would students perform on such tasks without the aid of the application? Though there are relatively few circumstances where one would need to rely purely on memory to complete a statistic selection task, such circumstances do exist (e.g., when ‘put on the spot’ by a supervisor). How much training would be required to bring students to the level of ‘expert’, where they could reliably (a) identify the all the relevant structural features of a research design, (b) construct a conceptual model in which the relationships between structural
features are represented, and (c) integrate that model with existing knowledge to select an appropriate statistic? Second, the individual training we have run and evaluated represents a ‘proof of concept’. However, as an actual method of teaching students it is very inefficient. We have described how it could be adapted for a classroom or online context and have no reason to believe that such adaptations would be less effective than individual training. However, this claim should be tested empirically. Furthermore, efforts should be made to determine a minimally effective training dose. Ideally, these efforts will be experimental, such that causal links between training and student learning/performance can be established. Finally, research is needed to investigate the extent to which training of the nature described herein actually impacts on performance on tasks believed to be reflective of structural awareness. Several such tasks have been proposed in previous research, including triad judgement tasks (Rabinowitz & Hogan, 2008), explanation tasks (Yan & Lavigne, 2014) and scenario generation tasks (Quilici & Mayer, 2002). Indeed, this is the primary focus of the RCT that we alluded to earlier.
References


Appendix A: Scenarios Used in Teacher-Guided Training

A lecturer wants to know if listening to classical music when studying improves memory. She recruits a sample of first year students and asks half to memorize a word list whilst listening to classical music. She asks the other half to memorize the same word list in silence. She then records how many words from the list each student is able to recall. Answer: Independent samples t-test.

An environmental scientist is interested in factors that influence public acceptance of recycled water. He recruits a sample of home owners and shows each two pictures of sinks full of water. In the first picture, the water is clear. In the second picture the water is a light brown color. Following each picture, each resident is asked whether or not they would support building a new water recycling plant that would produce water of the color illustrated. Answer: McNemar test of change.

A psychology student is studying the effects of auditory interference on reaction time. On each experimental trial participants quickly press a keyboard spacebar in response to a flash of light. For half of the trials, spoken word poetry is played in the background. The remaining trials are completed in silence. Each participants’ average reaction time for both the auditory interference and silent trials is recorded. Answer: Paired samples t-test.

A drug company wants to assess customer satisfaction with a new headache medicine. They recruit a sample of regular headache sufferers and give half of them a packet of the new medicine. The other half are given a packet of the current market leading brand of headache medicine. After a period of time, the company contacts each participant and asks whether or not they were satisfied with their assigned medicine. Answer: Chi-square test of contingencies.
Appendix B: Scenarios Used in Self-Guided Practice

A researcher wants to know if imagining being in the presence of others influences charitable behavior. She asks half of her participants to imagine that they are alone and the other half to imagine that they are in a busy café. She then presents each participant with information about an animal welfare charity and asks them how much they would be willing to donate. Answer: Independent samples t-test.

An introductory psychology lecturer is interested in understanding whether his students’ perceptions of the scientific status of psychology are influenced by the topic they have most recently studied. Following a lecture on Freud’s theories he asks everyone in the class to indicate whether or not they consider psychology to be a scientific discipline. Following a lecture on psychopharmacology he asks them all the same question. Answer: McNemar test of change.

A head of department wants to know if Statistics 100 grades differ from Psychology 100 grades. All students in the department take both of these units at the same time. The head of department records each student’s grades for each of the units. Answer: Paired samples t-test.

An occupational psychologist wants to understand the effects of including the word ‘please’ in an email request. She composes two versions of an email asking for some brief information about daily work habits. The word ‘please’ is included in the first version, but not in the second. She sends out the first version to half of the people in her contact list, and the second version everyone else. She records whether or not each contact provides the information she requested within one week. Answer: Chi-square test of contingencies.
Appendix C: Additional Scenarios

This Appendix contains 41 research scenarios which map onto the statistics, tests and procedures covered in StatHand (see Table C1). They are organized according to the five broad data analysis objectives that are presented to users when first opening the application (or visiting https://stathand.net). For many of the scenarios, there is no definitive ‘correct’ technique for analyzing the data they would likely generate. Having said that, most tend to suggest one obvious technique, whilst opening up the possibility for alternatives on further consideration. Consequently, they might best be used as discussion starters rather than, for example, multiple choice items for which there can only be a single ‘correct’ answer.

Table C1. The Statistics, Tests and Procedures Described in StatHand, Grouped by Data Analysis Objective. The Objectives Listed Correspond with the Five Options Presented to Users on the StatHand Home Screen

<table>
<thead>
<tr>
<th>Objective</th>
<th>Statistics, Tests and Procedures Described in StatHand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Describe a sample</td>
<td>Bar graph; category count; histogram; interquartile range; Mean; median; mode; pie chart; range; standard deviation; stem-and-leaf plot.</td>
</tr>
<tr>
<td>Compare samples</td>
<td>ANCOVA (independent samples and mixed; one way and factorial); ANOVA (independent samples, repeated measures and mixed; one way and factorial); chi-square (goodness of fit and contingencies); Cochran’s Q test; Friedman two-way ANOVA; Kruskal-Wallis one-way ANOVA; Mann-Whitney U test; McNemar test of change; t-test (one sample, independent samples and paired samples); Wilcoxon signed-rank test (one sample and paired samples).</td>
</tr>
<tr>
<td>Analyze relationships or associations between variables</td>
<td>Chi-square test of contingencies (with Phi or Cramer's V); correlation coefficients (point-biserial, rank-biserial, Spearman’s and Pearson’s); eta; linear regression (bivariate and multiple; standard and hierarchical); logistic regression (binary and multinomial; standard and hierarchical); ordinal regression (standard and hierarchical).</td>
</tr>
<tr>
<td>Examine the underlying structure of a measure</td>
<td>Confirmatory factor analysis; exploratory factor analysis; principal components analysis.</td>
</tr>
<tr>
<td>Examine the reliability of a measuring instrument</td>
<td>Cohen’s kappa; Cronbach’s alpha; intraclass correlation; Kuder-Richardson 20; Weighted kappa.</td>
</tr>
</tbody>
</table>

Describe a sample

Scenario 1: Anwar and Sally have just finished collecting data for a large study examining the methods that people use to find reliable information on the internet. Prior to reporting their results, they need to describe their sample. From each participant, they collected the following demographic information: age, marital status, highest level of education and annual income. Which measures of central tendency and dispersion should they report for each of these variables, and how should they be graphed?

Marital status is a nominal variable. Consequently, its mode should be reported, along with the number of categories of marital status represented in the sample. The number of people endorsing each marital status category can be captured in bar graph. (Or, alternatively, the proportion of the sample endorsing each marital status category can be captured in a pie chart.)

Highest level of education is most likely an ordinal variable. Anwar and Sally can report the median highest level of education, along with either a range or interquartile range. A bar graph can be used to visualize the distribution of education in the sample.

Age and annual income are both continuous (ratio level) variables. Assuming they are normally distributed, a mean and standard deviation can be reported for each. If they are skewed, which seems particularly likely for income, a median can be reported as well. Graphically, the distributions of both age and income can be captured using histograms or stem-and-leaf plots.

Compare samples

Scenario 2: The manager of a voluntary extra tuition program wants to know whether or not ‘regular’ attendees (i.e., students who go to 5 or more sessions per semester) achieve higher end-of-semester grades than non-attendees. However, he suspects that his study may be confounded by the fact that regular attendees also tend to be smarter students! Consequently, he wants to use IQ as a control variable in his analyses. What statistical analysis would you advise him to conduct?

The manager has a between subjects (or independent groups) design with one ratio level dependent variable (end-of-semester grades) and one independent variable with two levels (regular attendee vs. non-attendee). He also has one covariate. Assuming he can meet the relevant assumptions, a one-way analysis of covariance (ANCOVA) can be used to analyze
his data. Alternatively, hierarchical multiple regression could be used. Both should lead to the same conclusions.

**Scenario 3:** A researcher is interested in whether female or male students who play Tetris, Call of Duty, or no computer games over 10 days have significantly different mental rotation speeds (measured in milliseconds). He randomly divides 60 participants into three groups: 10 female and 10 male participants were asked to play Tetris for 20 minutes a day, 10 female and 10 male participants were asked to play Call of Duty for 20 minutes a day, and 10 female and 10 male participants were instructed to play no video games over the 10 days. The researcher is interested in whether playing a video game influences participants’ mental rotation speeds, and if so, which game is most effective/deleterious. The researcher further wants to know if participants’ mental rotation speeds differ according to gender. Finally, the researcher believes that the participants average daily ‘screen time’ may also influence their mental rotation speeds and wants to control for this variable in his analyses. Which statistical analysis would you recommend to this researcher?

This researcher appears to have a between subjects (independent groups) design, with a ratio level dependent variable (mental rotation speed) and two independent variables (game and gender). He also has a ratio level covariate (screen time). Consequently, a factorial between groups ANCOVA can be used to analyse his data.

**Scenario 4:** A friend of yours is running an experiment which involves measuring the self-esteem of 40 children, and then randomizing them into two conditions. The children in the experimental condition are then praised after displaying good behavior, whereas the children in the control condition are not. After a period of time, the self-esteem of each child is measured for a second time. Your friend wants to know if any changes in self-esteem observed between the pre- and post-tests are influenced by the experimental manipulation and would like to include the children’s ages (which she has also recorded) as a control variable in her study. What statistical analysis would you recommend?

This design most obviously lends itself to a mixed model ANCOVA, as there is one repeated measures independent variable (time), one between subjects independent variable (praise vs control), a continuous covariate (age), and a (presumably) interval level dependent variable (self-esteem).

**Scenario 5:** Jake believes that the type of music you listen to while studying may have an impact on test scores. Jake randomly assigns 60 students to listen to either, rock, country or classical music while they study a passage of text. After an hour of study, the students are given
a test on the contents of the passage. What statistical analysis should Jake use to test his hypothesis?

In this scenario, there is one continuous (interval or ratio) level dependent variable (test scores), and one independent variable (music type) with three levels. It’s a between subjects (independent groups) design, as different students have been assigned to each type of music. Consequently, a one-way between subjects/groups ANOVA can be used to analyze Jake’s data. If significant, it can be followed by either planned comparisons or post-hoc tests. If Jake’s dependent variable is considered ordinal and/or he can’t meet the assumptions of the parametric ANOVA, a Kruskal-Wallis ANOVA can be used instead.

Scenario 6: You work in an animal laboratory and have been asked to investigate whether rats can be ‘bred’ to perform well on a T-maze task. (This is a commonly used task requiring that a rat learn which features of the maze identify where food is located.) As a secondary consideration, you’ve been asked to look at whether performance is also influenced by the nature of the environment in which the rats were raised. You have access to a group of rats that have been selectively bred to perform exceptionally well on this task (the ‘bright’ rats), a group that have been selectively bred to perform poorly on the task (the ‘dull’ rats) and a group who were bred without regard for their maze performance (the ‘control’ rats). Furthermore, half of each group has been raised in an ‘enriched’ environment, whilst the other halves have been raised in an ‘impoverished’ environment. Performance on the T-maze is measured as the time that it takes for the rat to find the food, averaged over five trials. There is one trial every 48 hours, and testing begins when each rat is exactly 60-days old. Your objective is to find out if and how breeding and environment influence T-maze performance.

This appears to be a between subjects (independent groups) design, as there are different subjects in each of the six breeding/environment groups. There are two independent variables (breeding and environment), and the dependent variable (time) is measured on a ratio scale. Consequently, a factorial between groups ANOVA can be used to analyse this data.

Scenario 7: You are interested in whether a short (2-week) course of mindfulness therapy is effective in reducing parental stress levels for young mothers with postnatal depressive symptoms, and whether any improvements are evident at two-month, six-month and one-year follow-ups. At each of the four testing sessions, stress is measured using the 15 item “Parental Stress Scale”. Participants complete this scale by responding to each item on a scale ranging from 1 (strongly disagree) to 5 (strongly agree). These responses are then summed to give a
total stress score between 15 and 75 for each participant. Which statistical analysis would you use here?

*If we assume that the dependent variable (stress scores) is at least interval level, a one-way repeated measures ANOVA can be used to analyse this data. If it is decided that stress is an ordinal level variable and/or the assumptions of the parametric ANOVA are not met, a Friedman two-way ANOVA can be used instead.*

**Scenario 8:** The National Cycling Association encourages members to wear fluorescent vests at all times, as it believes that doing so makes them more visible on the roads and thus safer. However, they have not yet tested this belief. To begin to do so, they hire a scientist who programs a driving simulator to drop a virtual cyclist into a driving simulation at random intervals. When a cyclist appears, the ‘driver’ must respond, as quickly as possible, by pressing a button located on the simulator steering wheel. The simulator automatically records the time (in milliseconds) that it takes the driver to react to the presence of the cyclist. In a complete testing session, a driver is required to respond to 30 cyclists: 15 wearing fluorescent vests, and 15 in normal, non-fluorescent clothing (the ‘control’ cyclists). Furthermore, five fluorescent cyclists and five control cyclists are dropped into the simulation during ‘daylight hours’; five of each type of cyclist are dropped into the simulation at ‘dusk’; and the remaining five of each type of cyclist are dropped into the simulation at ‘night’. The 30 trials are fully randomized, and reaction times are averaged across each type of trial (e.g., fluorescent at dusk; control at dusk etc.) for each participant. How might the data collected in this experiment be analyzed?

*In this experiment, there is a ratio level dependent variable (reaction time) and two repeated measures independent variables (cyclist type and time of day). Consequently, it can be analyzed using a factorial repeated measures ANOVA.*

**Scenario 9:** A friend of yours is running an experiment which involves measuring the self-esteem of 40 children, and then randomizing them into two conditions. The children in the experimental condition are then praised after displaying good behavior, whereas the children in the control condition are not. After a period of time, the self-esteem of each child is measured for a second time. Your friend wants to know if any changes in self-esteem observed between the pre- and post-tests are influenced by the experimental manipulation. What statistical analysis would you recommend?

*This design most obviously lends itself to a mixed model ANOVA, as there is one repeated measures independent variable (time), one between subjects independent variable (praise vs control) and a (presumably) interval level dependent variable (self-esteem).*
Scenario 10: A vending machine manufacturer wants to know which brands of cola students prefer. He sets up a machine in a busy university hallway, and stocks it with an equal amount of three cola brands: Coca-Cola, Pepsi and Royal Crown. He then returns 48 hours later and counts up the number of cans of each cola that have been sold. Which statistical test should the manufacturer use to determine whether or not the students prefer some brands more than others?

In this scenario, the manufacturer is seeking to compare observed category membership frequencies (i.e., the number of cans of each type of cola sold) with a set of expected category membership frequencies (i.e., the number of cans of each type of cola that would be sold if students made their selections randomly, and expressed no true preferences). This can be achieved with a chi-square test for goodness of fit.

Scenario 11: You are interested in whether Star Trek fans or Star Wars fans are more likely to be married or not married. Which statistical analysis would you use?

Here, we wish to compare samples. There is one nominal ‘dependent variable’ (marital status), and one nominal ‘independent variable’ (series preference) with two levels. The obvious analytic technique here is the chi-square test of contingencies, which can be used to determine whether or not marital status is contingent on (or related to) the preference for Star Trek vs. Star Wars.

Scenario 12: The Dean of Psychology at City University suspects that the ‘fail rate’ for Statistics 100 is higher than that for the other two compulsory first semester psychology classes (Behavioral Science 100 and Human Biology 100). There were 200 students enrolled in these three classes last semester. The Dean has a record of whether each student passed (coded as ‘1’) or failed (coded as ‘0’) each class. How should she test her hunch?

This is a repeated measures design, with a dichotomous dependent variable (passed vs. failed). There are three levels of the independent variable (Statistics 100, Behavioral Science 100 and Human Biology 100). A Cochran’s Q test could be used by the Dean to determine whether the proportion of fail grades differs significantly across the three classes. If the Cochran’s Q test was statistically significant, it should be followed up with a series of McNemar tests of change.

Scenario 13: The City Council are facing financial difficulties and need to make some cuts to the services they provide to residents. They have come up with four possibilities, each of which will result in approximately the same reduction in expenditure: (a) halve the amount of money spent on maintaining public recreation spaces; (b) close the Emergency Room at one of the City’s three public hospitals; (c) sell 40% of City Park to private developers; or (d) reduce the
number of police patrolling the streets by 20%. Because they’re facing an election in 12-months, the Council are keen to pursue the option likely to cause the least amount of outrage amongst their constituency. Consequently, they have asked a sample of residents to rank the four possibilities outlined above from 1 = ‘most preferred option’ to 4 = ‘least preferred option’. They now need to analyze this data. What would you recommend?

_This is a repeated measures (within subjects) design with one ordinal dependent variable (ranked preference) and one independent variable with four levels (the four cost reduction options). It lends itself most obviously to a Friedman two-way ANOVA. If statistically significant, it should be followed with a series of Wilcoxon signed rank tests to identify which pairs of options differ significantly._

**Scenario 14:** Your friend works in an animal laboratory and has been asked to find out which of three nutritional supplements produce optimal cognitive performance in rats. She sets up an experiment in which 40 rats are randomized to four groups. Members of one group are given a placebo, whilst the remaining groups are given supplements A, B and C. She then places all 40 rats in a maze, and gives each a score between 1(st) and 40(th) based on the order in which they successfully complete it (i.e., the first rat across the maze ‘finish line’ is given a score of 1, the second across gets a score of 2, and so on). What statistical analysis would you advise to your friend?

_This is a between subjects (or independent groups) design with one ordinal dependent variable (completion rank), and one independent variable with four levels (placebo, supplement A, supplement B and supplement C). It lends itself most obviously to a Kruskal-Wallis one-way ANOVA. If the ANOVA is statistically significant, the friend should be advised to follow it with a series of Mann-Whitney U tests to identify which pairs of supplements differ significantly._

**Scenario 15:** You have been hired to investigate whether a new vaccination impacts on subjective lethargy (tiredness). A sample of university students are randomized into two groups. The first group are given the vaccination, whilst the second group are given a placebo. Each student is then asked to rate their lethargy on 3-point scale, where 1 = not at all tired, 2 = somewhat tired, 3 = very tired. What statistical test would be appropriate for comparing the average levels of lethargy reported by each group?

_The Mann-Whitney U test can be used to compare two independent samples of ordinal data._

**Scenario 16:** Before implementing a new anti-bullying intervention program at the 12 primary schools under their authority, the local school district asks each school’s counsellor whether they currently believe their school has a ‘bullying problem’. The counsellors responded to this
question by answering either ‘yes’ or ‘no’. At the conclusion of the program, each counsellor was again asked the same question. What statistical analysis should be used to determine whether or not the counsellors’ perceptions of bullying at their schools changed between the two points in time?

*The McNemar test of change* can be used to determine whether or not category membership on a binary dependent variable (counsellors’ perceptions of whether or not a bully problem exists) changes between two points in time (before and after the intervention). Note that a McNemar test of change requires that both variables are dichotomous, and a repeated measures design.

**Scenario 17:** Lena believes that the Matching Figures Test (MFT), a visual identification test, is too difficult for children who are younger than five years old. The MFT consists of 24 items. For each item, a child is shown a picture for two seconds then, after a five second pause, is asked to select the same picture from a three-picture line-up. If a child is simply selecting at random (or guessing), we would expect him/her to select the correct picture on approximately 8 of the 24 trials \((24/3 = 8)\). Lena has tested 75 five-year-olds and would like to know if they are performing on the MFT at a level that is any better than chance. What statistical test should she use here?

*Lena wants to compare the mean of a sample of ratio level data against a predetermined value, 8, which represents a level of performance equivalent to chance. Assuming the MFT data are reasonably normally distributed, she should do this using a one-sample t-test. If normality cannot be assumed, a one-sample Wilcoxon signed-rank test may be more appropriate.*

**Scenario 18:** The AFL (Australian Football League) Commission is interested to know whether West Coast Eagles supporters and Dockers supporters differ in the average amount of money they spend per season on AFL merchandise. Which statistical analysis would you recommend?

*The commission want to compare two independent samples of ratio level data. Assuming these samples are reasonably normally distributed, an independent samples t-test should be used. If normality cannot be assumed, the Commission should consider using a Mann-Whitney U test instead.*

**Scenario 19:** You are interested to know if a new fitness program is effective. You measure the weight of 30 participants prior to starting the program and again after completing the program. Which statistical test would you use to compare the two sets of measurements?

*This is a repeated measures (within subjects) design, with a ratio level dependent variable and one independent variable with two levels (before vs. after completing the program). The*
resultant data could be analysed with a paired samples t-test. If the assumptions of this test cannot be met, a Wilcoxon signed rank test could be considered instead.

Scenario 20: In 2010, the residents of Capitol City were asked to indicate whether they were ‘very dissatisfied’, ‘dissatisfied’, ‘neither satisfied nor dissatisfied’, ‘satisfied’ or ‘very satisfied’ with the performance of the city council. The median response was ‘neither satisfied nor dissatisfied’. Earlier this year, residents were asked the same question. The mayor would like to know whether satisfaction with the council has improved since the previous survey.

The mayor wants to compare the median of a sample of ordinal data collected earlier this year against a specified value (i.e., the 2010 median). The appropriate statistic for doing this would be a one-sample Wilcoxon signed-rank test.

Scenario 21: The National Cricket Association is keen to test the effects of their new ‘subliminal advertising’ campaign. They recruit a sample of 30 community members and ask them to rank 10 sports from 1 = ‘most favorite’ to 10 = ‘least favorite’. The participants are then taken to a movie theatre and asked to watch an action film. Unbeknownst to the participants, positive subliminal references to cricket have been inserted throughout the film. At the conclusion of the film the participants are again asked to rank the 10 sports. The Association would like to know if the typical rank that participants assign to cricket improves between the two points in time. What statistical analysis could they use to determine this?

A Wilcoxon signed rank test can be used to compare two related samples or ordinal (or ranked) data.

Analyze relationships or associations between variables

Scenario 22: Is whether or not someone has a mortgage related to whether they prefer to watch commercial or non-commercial television news programs? What statistical test could be used to determine whether or not these two variables are related?

A chi-square test of contingencies, along with a phi coefficient could be used to examine the nature of the relationship between these two dichotomous variables. If either or both of the variables had more than two levels, Cramer’s V should be used in place of phi.

Scenario 23: A teacher would like to know whether there is an association between physics grades and gender amongst her 10th grade students. What statistical test would you recommend.

A point biserial correlation coefficient can be used to quantify the relationship between a dichotomous variable (gender) and a continuous (i.e., interval or ratio level) variable (physics
grades). Alternatively, if the grades can be considered the ‘dependent variable’ in this study, an **independent samples t-test** can be used instead. Both will lead the teacher to the same basic conclusion.

**Scenario 24:** A doctor would like to know whether there is an association between BMI (Body Mass Index) classification (underweight, normal, overweight, or obese) and dog ownership (yes or no) amongst his patients. What statistical test would you recommend?

The doctor is interested in the relationship between an ordinal variable (BMI classification) and a dichotomous variable (dog ownership). Such a relationship can be quantified with a **rank biserial correlation coefficient**. Alternatively, if BMI classification can be considered the ‘dependent variable’ in this study, a **Mann-Whitney U test** can be used instead. Both will lead the doctor to the same basic conclusion.

**Scenario 25:** The Banana Growers Association (BGA) want to know if there is a relationship between retailer size (categorized as ‘small’, ‘medium’ or ‘large’) and the banana wholesale prices the retailers are able to negotiate. What statistic should they use to address this research question?

Assuming that the relationship between size and price is monotonic, and that the BGA has price and size data for a range of different retailers, **Spearman’s rho** can be used to quantify the strength and direction of the relationship between these two variables.

**Scenario 26:** A doctor is interested in finding out whether his patients’ weight is related to how much TV they watched in the previous week. What statistical analysis would you recommend to him?

Assuming the relationship between these two variables is linear, it will be best captured by **Pearson’s product moment correlation coefficient**.

**Scenario 27:** The Computer Retailers Association (CRA) would like to quantify the strength of the relationship (if any) between annual income and preferred operating system (Windows, MacOS, Linux or Other) amongst their members. What statistic would you recommend?

**Eta** is a symmetric measure of association between one nominal variable (operating system) and one continuous (i.e., interval or ratio level) variable (income).

**Scenario 28:** A teacher wants to know if she can predict end-of-semester exam scores using mid-semester exam scores. What would you recommend?
The criterion variable here is continuous (end-of-semester exam scores), as is the predictor variable (mid-semester exam scores). **Bivariate regression** can be recommended to the teacher.

**Scenario 29:** You work at a university library and have been tasked with finding out which students accrue the largest ‘overdue fines’. The head librarian has provided you with a data file that gives you the total amount of fines (in dollars) accrued by each borrower during the previous 12 months, along with a range of additional information (e.g., each borrower’s course of study, age, gender, number of items borrowed etc.). What statistical analysis would you use?

*The intent here is to predict the size of fines (a continuous variable) using a range of continuous and categorical predictor variables. This scenario suggests **standard multiple regression**, following the dummy coding of any categorical predictors with more than two levels (e.g., course of study).*

**Scenario 30:** A teacher wants to know if, after controlling for students’ mid-semester exam scores, she can predict their end-of-semester exam scores with their written assignment marks. The students complete two written assignments in addition to the two exams during the semester. What statistical analysis should she use?

*In this scenario, there is a continuous criterion variable (end-of-semester exam scores), two predictor variables (the two sets of written assignment marks) and one control variable (mid-semester-exam scores). **Hierarchical multiple regression** appears appropriate.*

**Scenario 31:** You are interested in whether the following factors in combination can predict whether or not participants who have been sexually harassed in their workplaces reported the incident to their supervisors: (a) relationship status (coded as 0 = single and 1 = in a relationship); (b) feminist ideology (measured using a self-report scale); (c) frequency of the harassment (measured on a 5-point scale); and (d) offensiveness of the behavior (measured on a 10-point scale). What statistical analysis would be appropriate here?

*In this scenario, we have a dichotomous criterion variable (whether or not the incident was reported) and several predictor variables (relationship status etc.). There do not appear to be any covariates in the study, and thus an appropriate analysis would be a **standard binary logistic regression**.*

**Scenario 32:** You’re working at a travel agency and have already established that wealthier customers are more likely to travel internationally during their holidays. Your manager wants to know if she can improve her ability to predict whether or not customers travel internationally by incorporating a few more demographic variables (age, gender, number of flights booked in the
previous 12 months etc.) into her statistical model. The manager hopes that she will be able to use this model to refine her advertising materials. What statistical analysis would you consider using here?

This scenario appears to be suggesting a **hierarchical binary logistic regression** to predict whether or not customers travel internationally (a dichotomous criterion variable). Income would be added to this model on the first block, followed by the remaining demographic variables on the second block. The manager seems to be hoping that the full model will have significantly greater predictive utility than the model containing only income as a predictor.

**Scenario 33:** A large electronics retailer wants to use the information in their customer database to develop a model that can be used to predict which type of tablet device customers are most likely to purchase. Over the preceding 12 months, they have sold 10,000 tablets, which have been categorized by operating system: iOS, Windows, Android, and Other. In addition to the type of tablet purchased, the retailer also knows the following about each tablet customer: number of other mobile devices owned, annual income, age, gender and postcode. What statistical technique will you suggest to the retailer?

In this scenario, there is a nominal criterion variable that has four levels (iOS, Windows, Android and Other). There are several predictor variables (income, age etc.), though postcode will need to be transformed to be useful. (Perhaps postcodes can be grouped by socioeconomic level?) Once the issue with postcodes has been addressed, multinomial logistic regression seems to be an appropriate recommendation. If the retailer has a rationale for the order in which predictors should be entered into the model, **hierarchical multinomial logistic regression** can be used. If not, all predictors can be entered simultaneously into a **standard multinomial logistic regression**.

**Scenario 34:** The editor of an academic journal is interested the degree to which he can predict his reviewers’ recommendations (‘reject’, ‘accept with major revisions’, ‘accept with minor revisions’ or ‘accept’) based on the characteristics of the papers they are reviewing. These characteristics include (a) length; (b) the number of $p$-values reported; and (c) number of previous papers published by the author. What statistical technique would you recommend.

The editor has an ordinal criterion variable (recommendation), and multiple predictor variables (which are all continuous in this case), suggesting ordinal regression. If the editor has a rationale for the order in which predictors should be entered into the model, **hierarchical ordinal regression** can be used. If not, all predictors can be entered simultaneously into a **standard ordinal regression**.
Examine the underlying structure of a measure

**Scenario 35:** A researcher is developing a new eighty-item measure to assess mental ability and wants to know whether there is a small set of constructs underlying the questionnaire. Which statistical analysis would you recommend?

*Either exploratory factor analysis or principal components analysis could be used here. The researcher will need to decide which is more appropriate for his or her purposes.*

Examine the reliability of a measuring instrument

**Scenario 36:** Kate and Phil have video-recorded 100 interactions in a busy shopping center car park. Each has independently classified each interaction as either ‘aggressive’, ‘friendly’ or ‘neutral’. Now, they want to compute a statistic that will indicate the extent to which their classifications are consistent. What would you recommend?

*Cohen’s kappa can be used to as an index of inter-rater reliability here.*

**Scenario 37:** David presented 200 participants with a 10-item questionnaire that measured perceived social support. Participants responded to the items on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). David wants to make sure that the measure is internally consistent. Which statistical analysis should David use?

*Cronbach’s alpha can be calculated to assess the internal consistency of a unidimensional Likert scale.*

**Scenario 38:** Researchers are interested estimating the test-retest reliability of a recently developed measure of the big five personality factors (openness, agreeableness, neuroticism, extraversion and conscientiousness). They have administered the measure to the same group of participants on two separate occasions, and now need some statistical advice.

*An intraclass correlation coefficient can be used here. Note that the researchers will need to compute a separate coefficient for EACH sub-scale/factor in the measure.*

**Scenario 39:** Two psychiatrists independently watched 50 video recordings of interviews with patients diagnosed with schizophrenia and counted the number of symptoms displayed by each patient. They now want to compute a statistic that will indicate the degree with which their symptom counts agree. What would you suggest?

*An intraclass correlation coefficient can be used to assess inter-rater reliability, when the ratings have been made on a continuous scale.*
Scenario 40: A team of researchers are examining social anxiety levels in first year university students. They administer a measure of social anxiety to a sample of 150 first-year university students and want to assess the reliability of the measure in this sample. They have come to you to ask how to do this. The measure is unidimensional, and participants responded to each item with either ‘true’ or ‘false’. Which statistical analysis would you recommend?

**KR20 (Kuder-Richardson 20)** can be calculated to assess the internal consistency of a unidimensional scale with a dichotomous response format.

Scenario 41: Kate and Phil have video-recorded 100 unambiguously aggressive interactions in a busy shopping center. They have then independently rated each interaction as ‘very aggressive’, ‘aggressive’ or ‘mildly aggressive’. Now, they want to compute a statistic that will indicate the extent to which their classifications are consistent. What would you recommend?

As the interactions have been classified into ordered categories, **weighted kappa** should be used as an index of inter-rater reliability.
Part Four: Advanced Topics
Feeling Good About Teaching the New Statistics

Tamarah Smith, PhD
Gwyedd Mercy University

Summary

Psychology students experience many apprehensions when studying statistics. They are often anxious, believe they are not “wired” to do math, and hold negative attitudes regarding the value of studying statistics. These apprehensions have profound effects on their ability to learn statistics. Problem based and active pedagogy can help students overcome these barriers and teach the material well; however, the inherent non-intuitive nature of null hypothesis significance testing (NHST) makes it challenging to hook students thereby allowing such good pedagogy to address apprehensions and lead to greater learning. The new statistics can help. The new statistics refers to the emphasis of effect sizes, confidence intervals, and meta-analysis over NHST. Not only is this good practice as it provides analysis that is better aligned with the types of research questions asked in psychology, but it is also more intuitive and lends itself to problem based active pedagogy. By teaching the new statistics, instructors can hook students and engage them in good pedagogy while simultaneously teaching statistical skills that are better equipped for answering psychological questions.

In this chapter, I will briefly review the apprehensions that psychology statistics students face, their impact, and how good active problem based pedagogy can help. I will provide a concrete example of a lesson using the new statistics and outline the way that the underlying concepts engage students and help them overcome statistics apprehensions all while deepening their statistical knowledge.

Statistics Attitudes and The New Statistics

Nearly all psychology undergraduates are required to learn statistics (Friedrich, Buday, & Kerr, 2000). Unfortunately, psychology students are particularly unhappy about this. Their expectations of being successful in learning statistics are low and they do not value statistical skills as something they will need (or want) for their profession (Ruggeri, Dempster, Hanna, & Cleary, 2008). For example, in focus groups with psychology majors, students reported that “there [is] no point” in studying statistics because it is “useless” (Ruggeri et al., 2008). Students in the study also expressed that understanding statistics material was challenging. One student stated that they felt they “need an interpreter” while another student commented, “I am running a t-test but what does that t-test actually tell you?” (Ruggeri et al., 2008). Together these comments suggest that even though students are learning some skills in statistics, as exemplified by the student who is able to actually run the t-test, they fail to understand how to use these skills and quite frankly see no use in doing so.

Researchers have been examining the effects of these type of negative attitudes towards statistics for some time, and data have consistently shown a relationship between such attitudes and statistics performance. Meta-analysis (Emmioğlu & Capa-Aydin, 2012) has shown that more negative attitudes such as a lower sense of competence with statistics, students’ affect when studying statistics as well as their value for learning statistics are all related to lower exam and course grades. Many activities and even interventions have been designed to help students
change these attitudes to be more positive and thereby increase their statistics performance. In this chapter, I provide an example lesson to address these attitudes that also teaches the “new statistics” (Cumming, 2014). I describe how this is yet another, but more comprehensive way to improve students’ attitudes towards statistics.

The new statistics refers to the increased focus on effect sizes, confidence intervals, and meta-analysis, and ideally a complete removal of $p$-values when interpreting data (Cumming, 2014). This approach addresses the problematic nature of $p$-values while simultaneously increasing our ability to better estimate the true nature of an effect or relationship. Cumming (2014) demonstrated this through a simulation study where a series of “trial studies” were run using data randomly drawn from two populations with known parameters. The trial studies were conducted to simulate the results of a series of standard independent samples $t$-tests. Because the parameters were set, the true effect was known and the series of trial studies could be examined for the extent to which the results of comparing two samples drawn from the populations revealed the true effect. Using $p$-values was not a reliable way to reveal this effect. The number of trial studies that resulted in a non-significant $p$-value ($p > .05$) was almost equal to the number of trials with a significant $p$-value (see Figure 1, Cumming 2014). As such, a single $p$-value alone provides poor insight. Effect sizes, and particularly confidence intervals, provide richer information. Confidence intervals provide a better estimate of the effect or relationship expected over repeated sampling; something the $p$-value fails to do. Effect sizes help us to interpret the data within context in a more qualitative way; rather than asking “is there an effect/relationship” we can ask “what is the magnitude of the effect/relationship”, a question that is more in line with our desire to see the impact of variables in the real world. Cumming found that across the trial studies, 83% of the time the mean of the subsequent trial study was in fact predicted by the prior trial studies’ 95% CI. This means that the 95% CI gives greater accuracy to what will be observed in other samples drawn from the sample population; or in other words, what is truly happening within the population at large.

From a teaching perspective, there are many advantages to teaching the new statistics: it aligns with the American Psychological Association (APA) goals for the undergraduate major such as explaining behavior using scientific evidence by interpreting complex findings not only in the context of significance but also by using effect sizes and common language (APA, 2013). This goes beyond simply stating “significant” or “not-significant”. It involves understanding the context and the way that multiple data points help us understand that context. The new statistics lends itself to this very approach, which is also in alignment with GAISE, Guidelines for Assessment and Instruction in Statistics Education (American Statistical Association, 2012). The GAISE recommendations emphasize focusing teaching on conceptual concepts in statistics using real data with context and purpose, and teaching statistical thinking. Teaching the new statistics provides a way to tap into students’ intuitive understanding of statistical logic. Yes, they do have an intuitive understanding of statistics, but we continually ignore this and instead teach from the inherently non-intuitive framework of $p$-values and NHST. Take for example the way in which $p$-values are interpreted in NHST: what the $p$-value “…tell[s] us is how surprised we should be to obtain the current effect, or one more extreme, if $H_0$ were to be true” (Chambers, 2017, p.24). Chambers provides one of the more accessible definitions of this, and yet, it would take me far more than the word limit of this chapter to help students wrap their heads around what this really means and how to use the $p$-value in NHST to make meaningful decisions from psychological research. Importantly, when students are presented with the non-intuitive NHST framework, it gives them all the confirmation they need of their preconceived notion that they will not be able to understand statistics.

Below I provide a sample lesson using the new statistics that gives students an introduction of how to use statistics to make meaningful decisions using scientific reasoning while capitalizing
on the inherent intuition they already have. Not only does the reliance on inherent logic help increase their sense of competence in statistics, but the new statistics lends itself to the type of lesson that follows that is steeped in problem based and active pedagogy, two methods known to positively address students’ negative attitudes towards statistics (Hilton, Schau, & Olsen, 2004; Lawson, Schwiers, Doellman, Grady, & Kelnhofer, 2003; Schoenfelder, Olson, Bell, & Tom, 2007; Wiberg, 2009).

A New Statistics Lesson to Improve Statistics Attitudes

The lesson is grounded in context—not content. The content of your statistics course is the actual statistical material: learning descriptive statistics, learning to run an analysis and write up the results, etc. The context is the story behind it. A story is more than a word problem. It is part of an overarching fiction with characters acting in an environment, and is such that individuals can become immersed within it (See Nilsson, Nordahl, & Serafin [2016] for more on fiction, story, and immersion in learning). In this example, I use the “story” of a university administrator needing to decide whether AlcoholEdu is worth requiring for incoming students. AlcoholEdu is an online education program that can be completed by incoming college student with the goal of reducing high risk drinking (and its negative consequences) while in college. A quick Google of the product produces plenty of information about it.

Step one: Explore AlcoholEdu with students and generate the information needed for the next steps.

To accomplish this step, talk about what AlcoholEdu is designed to do and the students’ opinions about that. Talk about it as if you are teaching a seminar on alcohol use in college, not as if you were teaching statistics. After you as a group have grasped the context including the role of students, administrators, drinking, negative consequences of drinking, and the effect learning about alcohol can have, talk about the evaluation methods that could be used to test the effects of AlcoholEdu. You could set this up many ways. For this example, I will use a scenario where two groups of students serve as samples with one group completing AlcoholEdu and one group not completing it. Ask students to determine what the underlying question is that is being tested and what types of outcomes would be meaningful. Table 1 provides prompts to help you engage in this conversation so that the group can develop the necessary information for the following analysis.

Table 1. Question prompts for step one of the lesson.

<table>
<thead>
<tr>
<th>Prompts</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>What do we really want to know here?</td>
<td>These questions are helping you get the students connected to the context. This may or may not yield specifics that are useful in the analysis; however, in all cases it helps them connect to the topic and to assign value to it, an important attitude correlated with learning.</td>
</tr>
<tr>
<td>Can you really change that with an online program?</td>
<td></td>
</tr>
<tr>
<td>What are other methods for reducing high risk drinking?</td>
<td></td>
</tr>
<tr>
<td>Why would we want to reduce this drinking?</td>
<td></td>
</tr>
</tbody>
</table>
What would make you believe AlcoholEdu was successful or not successful? What is important to us in this context?

Here you are helping to move students towards operationalizing a meaningful change and perhaps even how they define as “high risk drinking”. This is critical for our later interpretation of the results. We will no longer be looking for \( p < .05 \) or \( p > .05 \), so we will need new criteria to examine our data against. This conversation helps to create those criteria.

**Step two: Introduction of descriptive statistics and understanding what Cohen’s \( d \) is telling us**

Provide students with the basic descriptives and completed Cohen’s \( d \) formula using data for the lesson (example provided below). They do not have to be familiar with these statistics beforehand; you can simply introduce them now. For example: “We got a sample of 100 students who did not receive the AlcoholEdu intervention and asked them how many drinks they had the last time they drank alcohol. Here is the average for the group with the standard deviation: \( M=3.88(1.01) \). Not every student had exactly the average number of \( M=3.88 \) drinks. That would have been weird. So, the standard deviation tells us how much a student may have differed from that average number of drinks. That was a typical, or standard deviation of 1.01 drinks.” After describing the mean and standard deviation for the intervention group, describe Cohen’s \( d \). For example: “Here is a statistic that tells us what we call the ‘effect size’. In other words, it helps us understand the magnitude of the impact (or effect) that AlcoholEdu has on how many drinks students have. Take a look at these numbers. Can you tell me where they came from?”

Notice how there is no need for technical jargon or demonstration of calculations. You can certainly teach that to your students if you feel it is necessary, but wait. Wait for them to get hooked through the context and have an opportunity to build their confidence. In other words, we will address the attitude barrier and then teach more difficult content.

**Step three: Students evaluate the questions/criteria set up during Step One using the data provided**

The data provided in this lesson are limited. That’s okay. Just as you may want to start small with teaching NHST, you can start small when teaching the new statistics. In this step, ask students questions to get them thinking about how the data answer the research questions. Ask them to consider the criteria they developed and evaluate accordingly. Table 2 provides prompts for this discussion.

**Table 2. Outline, data, and prompts for steps two and three of the lesson.**

<table>
<thead>
<tr>
<th>Step 2 Prompt</th>
<th>Data Provided</th>
</tr>
</thead>
</table>
| Provide descriptive statistics (step 2) | AlcoholEdu Group: \( M=3.88(1.01) \)  
Comparison Group: \( M=4.98(1.02) \)  
Cohen’s \( d = \frac{4.98 - 3.88}{1.015} = 1.07 \) |
| Explain descriptives (step 2) | Explain mean and standard deviation in context (limit your discussion of what a mean and standard deviation are conceptually; there are no calculation demonstrations done here!) |
| Explain Cohen’s \( d \) (step 2) | Again, focus on context over concept whenever possible. See text for example. |
Activity: determine where numbers in Cohen’s $d$ came from (step 2)

Show the completed formula. Have students work in groups or individually to write down where each number in the formula came from.

Conversation: what is this telling us? (step 3)

This can be done in groups, as a class, or both. Here are some prompts. What do you think these numbers are telling us? What is your gut reaction? Is there a meaningful difference between the groups? How much of a standard deviation did the treatment group move? What is a standard deviation and how is it helpful here? In terms of context, what does this mean (reflect on the prior group decision of how much of a decrease in drinking would matter that you determined in step one)?

When asking students whether they feel there is a meaningful difference, it may pull an initial gut reaction, or decision based simply on the means. Students may disagree. All of this is great. Use prompt questions, such as those in Table 2 to help direct students towards using the effect size and a-priori criteria to evaluate the results. Remember when you ask them about the numbers (e.g., how much of a standard deviation did the treatment group move?), help them come to the answer using context not content. Think about how you can re-word your explanatory sentences to really be context based (e.g., “If we typically could expect a student in the comparison group to be one drink away from the mean of 4.98, is it a big deal that the AlcoholEdu group’s average is about 1 drink away from that average of 4.98?”)

Step four: Get students active with data analysis and a software program.

The purpose here is just to get them to be able to press the correct buttons. The interpretation is in the next step. I have students work with JASP, a free statistical analysis package that is similar to SPSS, but more straightforward and contemporary. You can certainly use SPSS, JAMOVI (see Chapter 18), or any other software program for this step (see Chapter 17, R), just adjust the step accordingly. The data and output for this lesson are available here: https://osf.io/muy6u/files/. First, display an error bar graph (or box plot graph) that shows the 95% CI for both groups (in this case for the AlcoholEdu group and the comparison group) and the Cohen’s $d$ value that you dissected in Step Three. See the Appendix for an example. This information is produced differently across software packages, but for the new statistics they are some of the critical statistics and graphs we want students to be able to create.

Next, ask students to open up the dataset provided using JASP (or preferred software). In JASP, I direct them to go under the $t$-test menu (not to “run a $t$-test”) and “click buttons” until they 1) duplicate the error bar graph you displayed and 2) a table that shows the same Cohen’s $d$ as you have in your display. Have students help each other. Provide scaffolding as needed.

Step five: Explain error bars and confidence intervals, and ask student to make a final decision.

At this point, the students should have a good sense of what means, standard deviations, and effect sizes can tell us (not necessarily how to calculate them; that’s okay for now). In step five of the lesson you will explain the confidence intervals for the example and ask students to come to a final conclusion with respect to the research question the group generated in step one.
Confidence intervals may feel like a higher-level concept, and the temptation might be there to go into explanations of concept, but don’t. Stay focused on the context. I prefer to do this lesson first and then do a lecture in at the following class meeting on central limit theorem and confidence intervals. Having the context first activates generation, a strong method for learning. Here’s what you want to cover when discussing the confidence intervals.

Using the error bar graph or box plot, show students how we can see the mean for the respective groups. Point out the limits on the confidence intervals (“whiskers”) and explain that this is demonstrating a prediction of what the mean number of drinks would be in other samples of students. Talk that through: “If we repeated this study, the new comparison group would not have an average of $M=4.98$...that would be very strange. We are not robots. We are people. New people will yield new numbers. BUT, they shouldn’t be that different if they are related, meaning they are all college students. So the graph shows us what we would expect a future comparison group’s average to be 95% of the time.” When thinking about it this way, it’s actually a very straightforward idea. Students can grasp this because it is intuitive. You do not need to get into theory or calculations. Just let them wrap their head around the general idea and then move straight into the context. For example, “looking at this graph for the comparison group, what would we expect the mean to be for another comparison group that we sample in the future? Would there ever be a time when the comparison group and the AlcoholEdu group could have the same average number of drinks? So, what do you think then regarding our research question? How helpful is AlcoholEdu? Is it worth the money? How much money might it be worth?”

A follow up to this lesson might be to have students write up their findings. I will leave that to you. Different departments have various approaches to this that require customization in how the assignments are done.

**Lesson Impacts on Learning and Attitudes**

Since teaching the new statistics, I have noted several differences in my students’ outcomes. In the past I may have been able to gauge students’ understanding of the material on how well they execute an analysis or calculation and then chose a correct interpretation of the results. Most students are able to master the execution of the analysis or calculation, but a much smaller group would master the ability to interpret the results. This latter skill requires that students understand the purpose of the analysis/calculation. With teaching the new statistics, the majority, if not all of my students now have the ability to not only to execute the analysis/calculation but also to interpret what the results mean and apply it to context. I see this reflected in increased exam and course grades as well as in their narrative summaries written in their final statistics portfolio. This ability is hinged on their new understanding of why we are doing what we are doing, something explicitly discussed throughout the new statistics material, such as with this lesson.

An important secondary outcome of this is that the students feel more confident in their work because they understand what is going on. Because the lessons are more intuitive, they rightly feel that they are using skills they already have and experience less anxiety because, by tapping into their intuition, they seem less likely to believe that are engaged in something that they are not prepared to do. Finally, because the lessons are context driven, it is easier to peak their interest and by seeing the way in which statistics are used in context, they begin to draw parallels as to how these methods will be valuable for them in their careers.
Conclusion

The APA Guidelines for the undergraduate major are clear in the need to teach students to think scientifically using multiple types of data, including confidence intervals and effect sizes. The GAISE are direct in the need to engage students in active problem solving activities when teaching statistics and to reduce the amount of rote calculations. Many researchers have identified the need to address students’ attitudes when studying statistics (Ramirez, Schau, & Emmioglu, 2012; Schau, 2003), and the data are quite convincing that such attitudes are related to student learning outcomes (Emmioglu & Capa-Aydin, 2012). This lesson, with a structure that can be adapted for teaching all of the statistical analysis designs taught at the introductory level, addresses the objectives of these guidelines as well as the recommendations found in the literature on addressing students’ attitudes towards statistics to help our psychology students learn (and love) statistics.
References


Appendix

Below is the output for JASP Independent Samples t-test showing descriptives, effect size, and error bar plot.

Resources for teaching the new statistics, including these plots, original data, and output can also be found at [https://osf.io/muy6u/files/](https://osf.io/muy6u/files/)

### Independent Samples T-Test

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>df</th>
<th>p</th>
<th>Cohen's d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinks</td>
<td>7.630</td>
<td>198</td>
<td>&lt;.001</td>
<td>1.079</td>
</tr>
</tbody>
</table>

*Note.* Student's t-test.

### Group Descriptives

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinks Comparison Group</td>
<td>100</td>
<td>4.98</td>
<td>1.02</td>
<td>0.10</td>
</tr>
<tr>
<td>AlcoholEdu Group</td>
<td>100</td>
<td>3.88</td>
<td>1.01</td>
<td>0.10</td>
</tr>
</tbody>
</table>
Chapter 14

Meta-Analysis as a Tool for Increasing Students’ Scientific Thinking

Jennifer V. Fayard, PhD
Ouachita Baptist University

Summary

With the recent call for reforms in psychological science, researchers have advocated for practices like meta-analysis becoming more prevalent and taught as early as undergraduate statistics (e.g., Cumming, 2014; Funder et al., 2014). However, the vast majority of instructors do not teach meta-analysis at all (Friedrich, Childress, & Cheng, 2018) due to lack of time or unfamiliarity with the method. Instruction in meta-analysis can have many benefits for introductory statistics students, both practically and in terms of their larger understanding of science and statistics. Practically, meta-analysis integrates and reinforces concepts from statistics and research methods and help students consider what different types of research questions are answered with different statistical tests. On a larger scale, introducing meta-analysis in introductory statistics can help us train a new generation of students to think in terms of effect size rather than statistical significance, increase students’ understanding of psychology as a science, encourage them to value replication, and help them take a broader view of research. The present chapter outlines several activities to introduce conceptual aspects of meta-analysis in the introductory statistics course that range from reading a meta-analysis and considering the results of conflicting studies to conducting a mini meta-analysis. It also includes resources for instructors to educate themselves about meta-analysis before presenting the material to students.

Introduction

Many professors are familiar with students who come into their first statistics course with a pronounced lack of interest (Rajecki, Appleby, Williams, Johnson, & Jeschke, 2005), or even an intense fear of math. Often, when statistics is paired with a research course, the context of using math to answer a question about human behavior helps them understand what those numbers mean, and if we are lucky, their fear turns to interest or even excitement. But is the reverse true—can understanding statistics help students understand how science works and how to do better research? Incorporating a meta-analysis unit in introductory statistics is an excellent way to reinforce basic concepts, provide a rich context for understanding how statistics and research design fit together, think critically about how to interpret statistics, and encourage students to value psychology as a science. In this chapter, I discuss the benefits and challenges of incorporating a meta-analysis unit in the undergraduate classroom and provide suggestions for activities and lectures, whether you have one day or one week to spend on the topic.

Meta-analysis is both a set of statistical techniques and a research method designed to estimate the overall strength of a relationship in the population by combining all of the existing data on that relationship in one analysis. Following the “replication crisis,” concerns about publication bias, and criticisms of null hypothesis significance testing, the way we conduct and evaluate psychological science is changing, and meta-analysis is at the forefront of those changes. The field is moving toward greater emphasis on effect sizes over NHST, increased reporting of confidence intervals and statistical power, publishing interesting null findings, and focusing on
cumulative methods such as replication and meta-analysis (Cumming, 2014; Eich, 2014; Stanley & Spence, 2014; Vazire, 2016). In light of these changes, several researchers have highlighted the need for meta-analysis to be more widely taught at both the graduate and undergraduate level (e.g., Funder et al., 2014). However, according to a recent national survey of psychology programs, most introductory statistics courses only cover effect size and confidence intervals for two days or less, and the overwhelming majority—between 73 and 84 percent—do not cover meta-analysis at all (Friedrich, Childress, & Cheng, 2018).

Why would we recommend that such an advanced technique be introduced at the undergraduate level? On a purely practical note, one reason we should introduce meta-analysis early on is that students will inevitably find meta-analyses in their literature searches. In my experience, most do not understand what they are, and thus discard them. But the benefits of this instruction go far beyond teaching students how to cite meta-analytic results. Besides training students in line with best practices in the field, a meta-analytic mindset prepares students to be better scientific thinkers by reframing their approach to research and data early in their academic careers. Below, I review a few of the specific ways meta-analytic instruction can benefit introductory statistics students.

**Meta-Analysis Encourages Stronger Scientific Reasoning and Appreciation of Psychology as Science**

Scientists build theories based on an accumulation of evidence rather than on the basis of a single study. Even if students understand this, they rarely have opportunities to watch scientists engage in this process. Further, despite content knowledge in psychology, many psychology majors do not develop a sense of psychology as science or understand the value of empirical studies (Holmes & Beins, 2009). There are several approaches to remedying this problem; specifically, meta-analysis not only offers an opportunity to emphasize the cumulative nature of our field, but also to explain why empirical studies are valuable, and provides an example of a concrete method in which scientists build on previous knowledge. In statistics and methodology courses, most examples center around a single study rather than asking students to make connections across multiple studies. This is not necessarily a bad thing, but we should take advantage of opportunities to show students how we practice our craft. This approach can help them understand how much weight to ascribe to single studies, as without this training, students often overemphasize the importance of a single study. We are all familiar with papers that rely on one article alone to serve as evidence for a phenomenon, sometimes even invoking the dreaded p-word (prove), despite our repeated insistence that they erase it from their vocabularies. Since meta-analysis considers an entire body of literature, it naturally emphasizes the need for replication before drawing conclusions. However, it can also challenge students’ misconceptions and show them a more nuanced understanding of science in general and statistics specifically.

Meta-analysis encourages students to think in terms of the size and precision of an effect rather than a simple significant/not significant dichotomy (Cumming, 2014). When statistics courses focus on NHST, students often think of non-significant results as not meaningful. However, many real effects may not be statistically significant for one reason or another. In a meta-analysis, even studies that were not significant may contribute to a sizeable overall effect that has real-life consequences (Borenstein, Hedges, Higgins, & Rothstein, 2009; Braver, Thoemmes, & Rosenthal, 2014; Cumming, 2012). As with replication, a discussion of meta-analysis necessitates coverage of effect size and confidence intervals. Covering these topics in introductory statistics may help students think about effect size from the beginning and inoculate them against the all-or-nothing thinking associated with NHST.
Meta-Analysis Can Enrich and Integrate Statistical and Methodological Concepts

In addition to increasing scientific thinking, instruction in meta-analysis can help students understand the concepts they have learned in statistics more deeply. One way this can happen is through integrating statistics and research methods. There is evidence that taking statistics and research methods courses concurrently (Stranahan, 1995) or as integrated courses (Barron & Apple, 2014) increases student performance both immediately and long-term. Meta-analysis, by nature, forces researchers to think about statistics methodologically, so that even if they are not currently enrolled in a research methods course, they must consider which tests work with which methods and which tests are comparable. For example, if students find a correlational study and an experiment that uses a one-way ANOVA, they must think about whether the two statistics are answering the same fundamental research question and thus can be considered in the same analysis.

By requiring students to read and think about research, they see theory, research design, and statistics working together and gain practice reading, interpreting, and reporting statistics. Further, in meta-analysis, researchers must think critically about the statistics and methodology they encounter and make active decisions about how to handle it. This process can help students begin to gain confidence and a sense of agency in research and statistics and assist in their transition from novice to scholar.

Activities to incorporate meta-analysis concepts in the classroom

In an ideal world, every department would be able to offer an advanced statistics course that provides an in-depth coverage of topics like meta-analysis. Realistically however, you may have only one day in which to build an advanced topic into your introductory statistics course. I argue that even one day gives you an opportunity to introduce students to some of the principles associated with meta-analytic approaches and encourage them to begin thinking differently about data. Below I present some activities and materials that you can implement if you have one day or up to one week to spend covering meta-analysis. These activities are designed to be consistent with several APA and GAISE goals for teaching introductory statistics (Statistical Literacy Task Force, 2014; GAISE, 2016): focus on conceptual understanding (GAISE 2), distinguish between statistical significance and practical significance (APA 4), and teach statistical thinking (GAISE 1). In order to increase students’ positive attitudes and self-efficacy for statistics and research methods and help them grasp the usefulness of meta-analysis, each activity is intended to be interactive, either in small groups or as larger class discussions, and provide a great deal of scaffolding (Ciarocco, Lewandowski, & Van Volkom, 2013). Further, each of these activities assumes that your course has already covered a few topics: caveats to null hypothesis significance testing, effect sizes, and how to read a journal article. Finally, especially if you have very limited time in your course, keep in mind that a discussion about meta-analysis will dovetail easily with an introduction to replication (Chapter 16), effect size, and confidence intervals (Chapter 17).

In addition to time constraints, one of the primary reasons statistics instructors do not cover meta-analysis is that they are not confident in their knowledge. If you are new to meta-analysis and would like to brush up before teaching it, a wonderful place to start is Cumming (2014). After that, Borenstein, Hedges, Higgins, and Rothstein’s (2009) textbook provides brief chapters on various aspects of meta-analysis, and Cumming’s (2012) text provides several user-friendly example exercises that pair with his (2016) ESCI Excel-based program, along with a step-by-step guide to conducting a larger-scale analysis. Chan & Arvey (2012) briefly review the advantages meta-analyses offer over single studies.
If you have only one day

With one day, you cannot teach a student how to perform a meta-analysis, but you can give them a conceptual understanding of how a meta-analysis works, ask students to think critically about familiar statistics in a real-life context, and present an example of ways meta-analysis can help scientists draw conclusions about behavioral phenomena.

**Activity: Introduction to meta-analysis and making sense of conflicting data.** It is an exciting moment when a student understands a results section they are reading for the first time, almost as if they have suddenly learned to speak a new language! However, it can be confusing or even upsetting for students when they first encounter inconsistent or conflicting results across different studies. For example, imagine that a student reads an article on power posing (Carney, Cuddy, & Yap, 2010) and implements posture changes in their daily lives, only to later read an article calling into question the validity of these studies (Ranehill, Dreber, Johansson, Leiberg, Sul, & Weber, 2015). This raises many questions for them: Why did one study find statistically significant results while another did not? Which one should they believe? A student who has invested in the idea that standing up straighter will make them more confident and successful might simply ignore the conflicting findings and dismiss the failed replication attempt. However, for some, it might shake the foundation of what they have just learned in research methods and statistics. If all the experts don’t agree, how will we ever find the true answer? Is all of our research just futile? In the worst case scenario, they may conclude that the field is not as scientific as we make it out to be! How can we equip students to grapple with inconsistency and uncertainty and prevent this scientific crisis of faith? The goal of this activity is to introduce meta-analysis as a tool for looking at the literature as a whole and showing students how to find trustworthy answers to their uncertainties and big questions.

**The materials.** You may choose any two articles for this activity, but I suggest a few in the appendix. These sets of articles were chosen for two reasons. First, both articles in each set have been included in a meta-analysis. I have referenced these meta-analyses so that, following the discussion about conflicting results and the introduction to meta-analysis, you can show students the meta-analytic effect of the relationship they read about in their homework. This activity could be done without an accompanying meta-analysis, but showing students the final result will make the activity more impactful. Second, the studies with null results were direct or conceptual replication studies, which may be helpful if you plan to blend instruction on meta-analysis and replication.

**The homework.** After students are familiar with a few types of basic statistics (e.g., Pearson correlations, t-tests, one- or two-way ANOVA), have students read two journal articles that show conflicting results and ask them summarize the main findings of both articles. Then, they should answer the following questions: What would they say if a relative read about this effect on the news and asked them, as a psychology major, if it was really true? Why do they think the studies found a discrepancy? Was one of them right and one of them wrong? What are the differences between the studies that might explain the difference in results? Depending on students’ familiarity with research methodology, they may or may not be well-equipped to answer the latter question. If not, you may need to prompt them with questions asking them to compare the sample’s demographics, materials and measures sections, and procedure sections and search for any differences they find. Students should bring a copy of both articles and their notes to class for the discussion.

**The discussion.** If you have a small class and the students are willing, it would be preferable to have a full class discussion. However, if you have a large or particularly shy class, you might break them up into groups of 4-5 students. When students arrive, ask them to talk about what they thought when they read the two articles. What conclusions, if any, did they walk away with about the variables in the study? As you guide them through discussing their homework
questions, you may ask them to look at specific items such as differences in sample size between the two studies, as this may have contributed to statistical significance, and the effect sizes each article found regardless of statistical significance.

**The lecture.** After the discussion, instructors can introduce the basic goals and mechanics of a meta-analysis. There are a few important points to include. First, this type of analysis circumvents the issue of which of two conflicting studies to believe because meta-analyses examine all the existing studies on a relationship together to give an estimate of the true size of the effect in the population. Since one of the goals of theory building is to accumulate a great deal of evidence and replicate studies, meta-analysis is a more comprehensive and powerful way to examine the literature than looking at a collection of single studies individually. Second, meta-analysis emphasizes the magnitude of an effect and the precision with which we estimate it over statistical significance. Third, by attempting to find and include unpublished data, meta-analyses are less biased than single studies and traditional review articles. Finally, since many students are unsure how to interpret meta-analyses when they read them, brief instruction on how to read a meta-analysis would be useful.

**If you have one week**

With one week, you and your students can delve more deeply into the mechanics of a meta-analysis. The activity above is a great place to start introducing students to general concepts in meta-analysis, while the activities in this section get into mechanics and other issues. The first two activities below are designed to be done in one day, and either may be done in place of the activity on conflicting results if they would be better suited to your class.

**Activity: Read and analyze a published meta-analysis.** In order to give students a sense of what makes a meta-analysis different from and more powerful than a single study, students can read and analyze a published meta-analysis. In order to complete this activity, students should be familiar with the typical format for individual studies. Choose a meta-analysis and have students reflect on the method and results sections. In particular, have them note what elements make the analysis rigorous, detailed, and transparent. Ask them to compare the method sections of the meta-analysis and a traditional single study. What makes them different? Then, ask them to compare and contrast the results sections of a meta-analysis and a single study. What are the implications of both? Are the conclusions and certainty in the meta-analysis stronger than in the single study? If so, why?

In addition, students can evaluate the meta-analysis to see if it meets established guidelines for quality. There are a few systems for evaluating meta-analyses and systematic reviews, such as the Meta-Analysis Reporting Standards (MARS) in the 6th edition of the APA manual, Cochrane Collaboration, and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). The MARS and PRISMA systems are user-friendly to undergraduates and include a checklist for the minimum standards for reporting a meta-analysis (American Psychological Association, 2010; Moher et al., 2009). Students can review the criteria on the checklist and locate where the meta-analysis meets each.

**Activity: Comparing meta-analyses and narrative reviews.** Another useful exercise would be to have students compare a meta-analysis and a traditional narrative review article. An advantage of systematic reviews is that they examine the entire set of literature on a topic, and in the case of meta-analysis, offer a numeric summary effect. Because authors of narrative reviews choose which articles are important, do not typically justify why those particular articles were reviewed, and offer their own synthesis, they have a greater chance of being biased. Students may compare these two types of articles using a similar method as above, or you may even ask them to compare meta-analyses to single studies and narrative reviews simultaneously. Regardless of whether you choose articles that come to the same or different conclusions, you can still have a great discussion about the different methods they used to
come to those conclusions, and which is most trustworthy and why. Article Set 2 in the Appendix gives examples of a narrative review and meta-analysis to compare.

**Activity: Conduct a mini meta-analysis.** The goal of this activity is to give students a taste of what is involved in actually conducting a meta-analysis. Students will have the opportunity to think critically about the literature and consider some of the questions scientists have to ask while performing this type of study. Dodd (2000) published a similar activity that could be used as an alternative to the activity below, but there are a few key differences between the two. Here, students will conduct their own literature search, establish and justify inclusion and exclusion criteria, organize and code their data, and analyze their data themselves. The present activity does not involve hand calculation or transformation, but if you wish, students may follow Dodd’s (2000) procedure and convert effect sizes to Fisher’s z.

**The background:** In order to successfully complete this activity, students should have had an overview of meta-analysis and experience with basic keyword searches in PsycINFO.

**The setup:** If you have two days or more to devote to this activity, you can very briefly take students through almost all the steps involved in conducting a meta-analysis! The ideal amount of time to complete this activity would be two full days, but it may be completed in less time with some modifications. Regardless of class size, this activity would work best in small groups of 3-4 students. In larger classes in which you cannot easily visit with every group, groups might do some work outside of class and bring their questions to class, email them, or post them on a discussion board.

The first step is to establish a research question. I recommend that the entire class decide on a topic together so that each small group is working on the same problem—managing one meta-analysis is challenging enough, and juggling students’ questions about multiple topics would be quite cumbersome. Once you determine a relationship to study, each group should establish their own inclusion and exclusion criteria—which articles they want to use and which they want to exclude. One criterion that you may want to establish with all groups is type of effect size. Although meta-analysis software can transform multiple effect sizes into a single type, you may be unlikely to have the time or resources, and students may not have enough knowledge of comparable effect sizes, to do this. Thus, for ease of use, I recommend that you restrict studies to one type of effect size. Correlation is an excellent candidate, as it is straightforward to understand, requires no additional transformation, and is compatible with the free software I reference below. However, this would be a good time to ask students to think about what types of research questions different statistics answer. Should correlations, t-tests, and ANOVA go in the same meta-analysis? Why or why not? Aside from effect size, in order to practice transparency and encourage students to think actively about the decisions they are making, students should attempt to explain why they chose an exclusion criterion, particularly if they exclude studies based on demographic characteristics, sample size, quality of the study, measures, operational definition, or another methodological element.

Next, walk students through setting up their database. In Excel or another spreadsheet program, students should have columns for study name (authors, year), sample size, correlation, measures used, and sample characteristics (e.g., undergraduate students, community sample, members of an online cat forum). They should also keep lists of keywords and combinations of keywords they search and notes on articles they strike from the analysis based on their exclusion criteria.

**The literature search.** Each group should aim to find five articles that meet their inclusion and exclusion criteria. I strongly recommend doing a preliminary search to assess how difficult it is likely to be to find appropriate articles—some topics will be easier than others. One way to considerably shorten the time required would be to provide a set of articles for students to review rather than have them conduct their own search. This might be especially helpful if your
students have not had much experience searching for articles. If you do this, I recommend giving them some articles that are likely to be excluded based on their criteria so that students get the experience of justifying rejecting an article.

**The analysis:** Once your groups have found five articles, you may analyze the meta-analytic effect. There are two free options for running a basic meta-analysis. The easiest to use is Cumming's (2016) ESCI software, especially if you are unfamiliar with R (but see Chapter 18 if you are interested in learning more about R). This program runs in Excel and gives a summary effect, confidence interval, and forest plot for up to 30 studies. Kovalchik (2013) published a thorough guide to conducting meta-analysis in R that includes examples and a comparison of the features and graphics from three different R packages. These packages also present basic output and forest plots. Students should examine the basic statistics and discuss how to interpret the magnitude of the effect size and the width of the confidence interval. Due to the nature of this exercise, it would not be very informative to conduct a publication bias analysis. However, you may demonstrate changes in the effect by adding or removing studies to simulate including or excluding unpublished data (see Cumming, 2012 for an example in ESCI).

**Activity:** Apply a meta-analysis. In order to demonstrate the utility of meta-analysis, you may ask students to use the results of a meta-analysis to develop and justify a policy recommendation. This policy may be at the organizational level, such as a club; at the institutional level, such as your university; or at the national level. In their justification, students can explain what meta-analysis is, what it tells us, and why meta-analytic findings are strong enough to warrant a recommendation for action. Alternatively, to emphasize meta-analysis’s role in theory building, students might use the results to design a new study on that topic.

**Outcomes: A New Generation of Scholars?**

I developed these activities for introductory statistics and research methods based on my experiences teaching a full-semester undergraduate course on meta-analysis. All the activities except the Mini Meta-Analysis and Apply a Meta-Analysis are the same or a variation on activities students complete early in the full course to get the lay of the land before diving into a full meta-analysis. I have had the opportunity to teach the course twice, and both times I have seen a dramatic change in my students’ understanding of and attitude toward research and statistics, as well as an increase in their level of scientific thinking. They transformed from passive recipients of facts to active, creative scholars. They increased in their self-confidence about statistics, which has increased their excitement about research. In feedback, students have reported that they have a greater understanding of psychology as a science and that the “bird’s eye view” of the field they saw helped bring together ideas they encountered across the psychology curriculum. While a shorter period of time may not yield such pronounced changes, these activities may plant the seeds for more growth as they advance in their coursework.

**Advice and Conclusions**

Students have a lot to gain from even a brief unit on meta-analysis. However, these are challenging concepts to teach. In my experience, it is best to plan to go slow and be prepared to go even slower. It is easy to be ambitious when teaching meta-analysis, but students must have a good grasp of the prerequisite material before they can grapple with the more advanced concepts. But if you are patient, encouraging, and meet them at their ability level, most of them will rise to the challenge. Of course, meta-analysis is not perfect, and it is not immune from bias (Borenstein, Hedges, Higgins, & Rothstein, 2009; Chan & Arvey, 2012), but it is a valuable tool and has the potential to change students’ perspectives. After this unit, I hope you will be rewarded with students who are interested in how science works and who, regardless of whether they go on to graduate school, will be better equipped to evaluate scientific data.
References


Appendix: Articles to Use in Activities

Set #1: Does Ego Depletion Lead to Self-Regulation Failure?

Significant effect: Food consumption


Null effect: Food consumption


➢ Note: Two meta-analyses have been published on ego depletion, one that did not include unpublished data (Hagger et al., 2010) and one that did (Carter et al., 2015), and the papers came to different conclusions about whether ego depletion is a real phenomenon. If you wish, you could use these analyses to demonstrate the impact of publication bias.

Meta-analysis 1:


Meta-analysis 2:

Set #2: Does Consuming Sugar Restore Self-Regulation Capacity Following Ego Depletion?

Significant effect:


Null effect:


Note: Here, we have a narrative review (Galliot & Baumeister, 2007) and a meta-analysis (Dang, 2016) that come to different conclusions. These articles may be useful for comparing narrative reviews to meta-analyses.

Narrative review:


Meta-analysis:

Set #3: Does Ovulation Affect Women’s Mate Preferences?

**Significant effect:**


**Null effect:**


Note: Two meta-analyses on this relationship were published the same year and found different results. If you have time, you might examine the differences between these meta-analyses with students and ask them to think about why two meta-analyses might have reached two different conclusions.

**Meta-analysis 1:**


**Meta-analysis 2:**

Advice and Activities for Teaching Replication in Statistics Courses

Nathaniel R. Greene, Mike Corcoran, and Lisa M. Bauer, PhD

University of Missouri - Columbia

Summary

Psychological scientists have become increasingly aware of the need to replicate published findings to build a trusted knowledge base. Replication is an essential part of scientific progress, and can be used as a valuable learning tool about statistics. This chapter provides advice and activities tailored to teaching replication in undergraduate statistics courses. The primary aims of these activities are fostering an understanding of the importance of replication in psychology, enhancing student learning of statistical topics, and encouraging students to think more critically about research in psychology and science, more generally. These activities range from course-long projects to shorter activities that can assess student learning of replication and related statistical topics. Many of these activities are based on the Collaborative Replications and Education Project (CREP), a crowdsourced replication project available on the Open Science Framework that can provide hands-on learning and classroom activities centered on replication. We outline one approach to teaching replication using resources available from CREP to engage students in a semester-long replication project. We also provide examples of shorter activities that can be used to compare the results of an original study with those from a replication study. These activities and resources promote student learning of statistical concepts and analyses and enhance their understanding of how to communicate statistical findings.

Scientific Replications: What Are They, and Should I Teach Replication in an Undergraduate Statistics Course?

One of the most essential steps of the scientific method involves replicating previous findings to determine whether they are generalizable. Recently, psychology has been faced with a harsh realization: many of the field’s most well-cited studies have failed to replicate (Klein et al., 2018; Open Science Collaboration, 2015). In light of this “replication crisis,” psychological scientists have become increasingly aware of the need to conduct replications of previously-published studies (e.g., Lindsay, 2015). Several researchers have suggested that undergraduate students can conduct these replications as part of research methods courses (Frank & Saxe, 2012; Grahe et al., 2012; Standing, Greiner, Lane, Roberts, & Sykes, 2014; Wagge et al., 2019), thereby allowing them to contribute to the field in a meaningful way by producing research projects that are often publishable in replication repositories (e.g., PsychFileDrawer.org) or journals that accept replication attempts (e.g., Psi Chi Journal of Psychological Research). Encouraging undergraduate students to conduct replications should serve another equally important purpose: teaching replication should promote student learning of psychological research and statistics, thereby expanding students’ knowledge base in psychology and enhancing their scientific inquiry and critical thinking skills (i.e., connecting the teaching of replication to the APA Guidelines for the Undergraduate Psychology Major, 2013).
One way in which replication can be used to enhance student learning and promote critical thinking skills is by incorporating replication into undergraduate statistics courses, where it can be used to complement teaching of both foundational topics (e.g., p values, sampling theory, Type I error) and more advanced topics (e.g., effect size, power analysis, confidence intervals). Teaching replication in an undergraduate statistics course can promote statistical thinking (GAISE, 2016, Goal 1) and foster conceptual understanding of statistics (GAISE, 2016, Goal 2) by incorporating real data from published studies and from replication studies with a focus on analyzing and comparing results (GAISE, 2016, Goal 3). All the activities considered in this chapter encourage students to interpret basic statistical results (Statistical Literacy Task Force, 2014, Goal 1) and to evaluate statistics presented in published reports (Statistical Literacy Task Force, 2014, Goal 5). In addition, some of the activities considered here provide opportunities for students to collect and analyze data, which can provide hands on learning (GAISE, 2016, Goals 4) and equip students with skills to carry out appropriate statistical analyses (GAISE, 2016, Goal 5; Statistical Literacy Task Force, 2014, Goal 3).

Replication can be used as a powerful tool to promote student learning of statistics, but the decision to incorporate replication into an undergraduate statistics course should be balanced with the time commitment it will take to teach students about replication. Therefore, this chapter provides a range of activities, which vary from semester-long research projects to single-class activities, to provide instructors with a variety of options for implementing replication into their statistics courses.

Where Do I Begin?

Many instructors may find that they do not have the time or resources to devote to supervising student-led replication studies (see Activity 3) or introducing lectures and assignments centered on replication. Therefore, before designing a replication activity, instructors should consider the time commitment the activity will require of both the instructor and the students and the availability of resources to conduct the activity. Table 1 presents a checklist that instructors can consult before beginning a replication activity.

**Table 1. Guidelines for Designing a Replication Project**

<table>
<thead>
<tr>
<th>Part 1: Initial Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>How important is it to include replication as a topic in this course?</td>
</tr>
<tr>
<td>Do I have the resources (e.g., time, computers) to devote to a semester long project?</td>
</tr>
<tr>
<td>Are there alternative ways I can introduce topics such as replication in a statistics course?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part 2: Choosing a Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where can I find a study to replicate? (CREP, PsychFileDrawer, psychology journal)</td>
</tr>
<tr>
<td>What types of statistical analyses will be used in this study?</td>
</tr>
<tr>
<td>Will collecting data for this study be manageable, for me and my students?</td>
</tr>
<tr>
<td>What are the ethical considerations of conducting this study?</td>
</tr>
<tr>
<td>Should I get IRB approval for the study, and require ethics training for my students?</td>
</tr>
</tbody>
</table>
**Part 3: 8 Steps for Carrying Out the Replication**

<table>
<thead>
<tr>
<th>Step 1</th>
<th>Submit an IRB proposal prior to the semester (if you plan to make the data available beyond the classroom)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 2</td>
<td>Students read the original article; design an activity to assess student learning (e.g., What were the hypotheses and results of the original study?)</td>
</tr>
<tr>
<td>Step 3</td>
<td>Students and instructor discuss the importance of replication and how to collect data.</td>
</tr>
<tr>
<td>Step 4</td>
<td>Students collect data (individually or in groups, or via an online survey) – data collection depends on resources available and the type of study selected</td>
</tr>
<tr>
<td>Step 5</td>
<td>Instructor (or students, as part of an activity) enters the data into a file (e.g., SPSS, JASP, R, Excel, SAS)</td>
</tr>
<tr>
<td>Step 6</td>
<td>Students analyze the data as an assignment, and submit results to the instructor (Note: Instructor can ask for different types of analyses to assess student learning of statistical concepts, e.g., t-tests, ANOVAs, regression, correlation)</td>
</tr>
<tr>
<td>Step 7</td>
<td>Instructor goes over the analyses in class and asks students to answer questions (e.g., effect sizes, statistical significance, whether the replicated findings support or contradict the original findings)</td>
</tr>
<tr>
<td>Step 8</td>
<td>Instructor asks students to summarize the findings in an APA-style report or in a table or figure. This final assignment can also include submitting the findings to a replication repository such as CREP, if IRB approval was obtained in Step 1.</td>
</tr>
</tbody>
</table>

For instructors who are interested in teaching replication, we recommend the Noba module by Diener and Biswas-Diener (2019) as a brief overview of the problem. This module, available at [https://nobaproject.com/modules/the-replication-crisis-in-psychology](https://nobaproject.com/modules/the-replication-crisis-in-psychology), can be used to introduce the topic of replication to undergraduate students.

We recommend that instructors begin planning their replication activities before the start of the semester. Instructors who wish to proceed with a semester-long project, such as the one below based on the Collaborative Replications and Education Project (CREP; [http://osf.io/wfc6u](http://osf.io/wfc6u)), are encouraged to obtain IRB approval before the class begins or during the first weeks of the semester. In addition to the resources provided below, we recommend instructors read the article by Wagge et al. (2019), which provides a good overview of the CREP process in their Figure 1.

**Activity 1: Replication Vignettes and the “Replication Recipe”**

This activity is intended to be completed as in-class or take-home assignment. We recommend instructors complement this activity with a lecture on replication, which should highlight the importance of replication and link it to statistical topics covered in the course. For example, instructors could relate replication to sampling theory by demonstrating how each sample drawn from a population of interest will have sample statistics that differ from other samples of the same size. The focus on replication could demonstrate whether these differences are merely due to chance. These “Replication Vignettes” include brief summaries of an original study and a replication study and several questions designed to assess student learning. An example
A vignette is provided in Appendix A (Figure 1), along with some sample questions. To create these vignettes, instructors can write a brief summary of an original study and a replication study. We selected studies available on the CREP (https://osf.io/m8et3/wiki/home/) to create the vignette in Figure 1. Questions can also be borrowed from the “Replication Recipe” (Brandt et al., 2014), which includes a 36-question guide to conducting replications (e.g., “The replication effect size [is / is not] (circle one) significantly different from the original effect size?”). Notably, the “Replication Recipe” may also be used in companion with a CREP or other replication project. The full 36-question “Replication Recipe” is available in Table 1 of Brandt et al. (2014), but we reproduced some sample questions in Appendix B.

Activity 2: Replication Lab Activity

For instructors who have access to a computer lab, a brief lab on replication can be a great way to teach students about replication. We recommend instructors familiarize themselves with one of the following statistical packages: SPSS, R, JASP, or Jamovi (see Chapters 17 and 18 on R and Jamovi). Instructors should also ensure that students have at least a working knowledge of the statistical program used for one of the activities below.

Activity 2.1: JASP and the Open Science Framework (OSF)

JASP (available for free download at https://jasp-stats.org/) is a great alternative to SPSS and includes many appealing features, such as user-friendly methods to import data directly from an OSF repository (for a video tutorial see https://www.youtube.com/watch?v=M5bEkKD8KYM). With this feature, students or the instructor can download data directly from a study available on OSF, such as one of the CREP replication studies (see Activity 3 below). Using the data from a replication study, students can be instructed to carry out a variety of statistical analyses on the dataset (e.g., t-tests, ANOVA, regression) and compare the results they obtain to the published results of an original study. This assignment would teach students how to conduct statistical analyses and how to read and interpret results in published papers.

Activity 2.2: Repeat Sampling Studies

A great alternative to using data available on the OSF is to create data yourself or have students gather data. One way this can be accomplished is by creating two similar datasets measuring a variety of psychological constructs. These datasets can be created for any statistical program and should include the same (or similar) measures with roughly the same number of observations. Students can be instructed to perform the same statistical analyses on each dataset to determine whether results obtained in one sample are replicated in another sample of the same size. This assignment could be related to variety of topics introduced in the course, such as p values and inferential errors. For example, if one dataset provides statistically significant results, but the other dataset (measuring the same constructs with the same sample size) does not, this may suggest that the first results were merely a Type I error.

A second possibility is to have students collect data on the same measures from two different samples. Students (or the instructor) can compile a list of questionnaire items and sample other students on campus or send a survey link out on social media to collect data from friends (for purely educational purposes). The survey can be completed on two different samples at the same time point, or students can be asked to collect data at two different time points, obtaining roughly similar-sized samples each time. Students may also be divided into two or more groups, and each group may be asked to collect data from a small sample of participants, with the same measures across groups. Using any of these methods, students can enter the data they collect into a statistical program and perform the same analyses on each dataset. This could be a very hands-on and fun way to teach students about replication.
Activity 3: CREP Project

Instructors who wish to include a more substantial project centered on replication may consider conducting a replication project with students. This approach may be especially suited for classes that combine research methods and statistics, at universities where students are taught research methods prior to their statistics course, or where students complete multiple statistics courses.

What is CREP?

The Collaborative Replications and Education Project (CREP) is a replication repository available on the OSF at https://osf.io/wfc6u/. The CREP includes a list of highly-cited studies (https://osf.io/flaue/) from different sub-disciplines of psychology available for student-led replications. Each study includes materials and instructions to prepare either a Direct or Direct+Plus replication. With the latter option, students can add new variables, in addition to those required of a direct replication. With CREP, students can design and carry out a replication project and submit the project to the OSF, where it can be included in a meta-analysis or other publishable form.

Could CREP Work for Your Class?

There are several considerations instructors should weigh carefully when deciding whether a CREP project is suitable for their courses (see Table 1). First, the instructor should consider the size of the class. A replication project could work for classes of varying sizes, but for smaller classes, each student may have to collect data from more participants to meet the required sample size needed to successfully carry out a CREP replication project (information about sample size requirements for CREP projects are available on the “CREP instructions and workflow procedures” page at https://osf.io/srh4k/, see the file titled “CREP step by step”).

Second, the instructor should consider the time it will take to conduct a replication project. The “CREP step by step” guide includes a printable checklist that outlines the approximate amount of time each step of the CREP process will take. For instructors who wish to share the results of their courses’ replication studies on the CREP repository, IRB approval and ethics training for all students will be required prior to submission. Many universities allow IRB exemptions for data collection if the research is purely for educational purposes and will not be presented outside of the classroom. However, IRB approval provides many benefits, including the opportunity for students to submit their replication study to the CREP or to present at university-sponsored research events, regional conferences, or even larger psychological conferences.

Third, instructors should consider whether there are adequate resources to conduct a replication project, and which study available on CREP would be most ideal for their classes. Instructors are encouraged to carefully select a CREP study based on the following considerations:

- how feasible it would be to conduct that study,
- the types of statistical analyses and number of participants required,
- the lab space available to conduct the replication (if the study requires in-person data collection).

Getting Started

The CREP provides a number of resources for instructors planning to design a replication project, including a general workflow and step-by-step guide (https://osf.io/srh4k/), example projects, and a presentation on how to do a CREP project (https://osf.io/stdgm/). Before the semester begins, instructors should decide which study, from the list of available studies on CREP, to use in their class and whether the data collected from that study will be submitted to
CREP. If so, IRB approval will be necessary, and the instructor will need to create an OSF page and contact a CREP administrator to be added as a contributor. As a CREP contributor, the instructor will need to provide the materials, IRB proposal and approval, an electronic recording of an example of the data collection procedure, information about the analyses (such as a codebook, raw data, syntax, and output), and a completion pledge.

Instructors should introduce the CREP replication project to students during the first week of class, to allow enough time for ethics training (if applicable) and data collection. As an initial assignment, students can be instructed to read the original study and identify the hypotheses, the research design, the materials used, the sample size, and the statistical analyses that were conducted. Over the next few weeks, the results of the original study can be discussed as they relate to statistical topics that are introduced throughout the semester. Instructors also need to outline the data collection procedure to students and supervise students during data collection.

**Collecting and Analyzing Data**

Instructors should decide whether students will collect data individually or in groups. Some research suggests that group projects can help reduce students’ anxiety about statistics (Standing et al., 2014), and it may help students collect data in a more timely manner. Data collection should be informed by the available resources and the design of the study. If the study is to be shared on the CREP, the data collection method should follow the CREP guidelines for the particular replication study. If the study is purely for educational purposes, we still encourage instructors to follow the data collection methods of the original study as closely as possible. However, in some cases, this may not be feasible, and instructors may wish to collect data using more convenient methods. Perhaps the most straightforward method to collect data is through a campus subject pool. Another option involves students recruiting in-person convenience samples, as by administering study surveys to random participants on campus. However, instructors who intend to submit the replication project to the CREP should consult with the CREP administrators to determine whether this is an accepted form of data collection. Data may also be collected online, as through Amazon’s Mechanical Turk, but one caveat to this approach is that participants must receive financial compensation for participation.

Instructors must also carefully consider how the data will be analyzed. Analyses for the replication project should be connected to the design of the study (i.e., the analysis plan of the original study should be followed as directly as possible). However, instructors may include additional analyses to teach students a variety of statistical tests related to the course content. The availability of resources and the instructor’s learning objectives will largely dictate what analyses will be conducted and how involved students will be in the analysis plan. Instructors who have access to a computer lab on campus may be able to reserve the lab for a data analysis day. If you do not have access to a computer lab, it may be feasible to ask students to bring their own laptops for an analysis day, but this should involve group work as some students may not be able to bring a laptop to class. If you choose this option, some statistical packages, such as JASP, Jamovi, and R, are freely available for download (reminder Chapters 17 and 18 are on using Jamovi and R in introductory statistics classes).

Beyond considering how the data will be analyzed, instructors should also consider how involved students should be in the data analysis process. This should be related to the content goals and specific learning objectives of the course. For example, if a learning objective is tailored to a more conceptual, as opposed to a more applied, understanding of statistics, or if students are just being introduced to data analysis, instructors may conduct the analyses with the students and relate the types of analyses conducted to conceptual foundations discussed in
the lecture. If students are more advanced and able to synthesize their statistical knowledge, instructors may instead provide a more general roadmap for students to complete the analyses on their own. Regardless of the level of student involvement, there are many opportunities for intermediate assignments that can test student understanding and ensure adequate progress. For example, students can be instructed to suggest analyses for their study or write out a data analysis plan (see Chapter 5 on Passion Driven Statistics).

Tying It All Together

After completing a replication project, instructors are encouraged to evaluate student learning of the importance of replication in psychology. Instructors should include a final assignment related to the replication project, which may include a student write-up of the results. If the study received IRB approval and all the steps of the CREP checklist were successfully achieved, the final assignment may involve submitting the replication project to the OSF and linking it with the CREP pages. Finally, instructors may wish to discuss the replication project and, more generally, the replication movement in psychology and ask for student feedback.

Alternative Sources for Replication Studies

There are alternative sources for finding studies that can be easily amended to a replication study. Instructors can search through journals such as Psychological Science, Current Directions in Psychological Science, and the Journal of Experimental Psychology: General, which are written for more general audiences and may therefore be easier for undergraduates to read. Instructors may find a study in a recent issue of one of these journals that may be feasible for students to engage in a direct replication, a replication plus extension, or a conceptual replication study.

Instructors may also consult PsychFileDrawer (https://psychfiledrawer.org/), which contains a “Top-20 list of studies that users would like to see replicated.” PsychFileDrawer also contains a search function that instructors can use to search for a particular study, research topic or design, or type of statistical analysis. Furthermore, the website keeps track of replication studies that have been both successful and unsuccessful, which may help when searching for a study to replicate. As with the CREP, the data from student-conducted replication studies that have received IRB approval can be shared to PsychFileDrawer.

General Advice for Incorporating Activities on Replication

Besides introducing an assignment on replication, instructors should also consider how a lecture and assignment on replication relates to other topics introduced in the course, and how student learning will be assessed. Ultimately, these decisions are up to the instructor’s discretion, but we offer some general suggestions below.

To What Other Statistical Topics Can I Relate Replication?

Replication should be used to promote critical thinking about statistics as they are commonly used in psychology, therefore satisfying the APA’s goals to promote critical thinking in undergraduate students. To this end, replication can be connected to a variety of different statistical topics. Replication can be used to emphasize which statistical models are most appropriate for specific research designs, as well as assumptions and limitations of those models. A lecture on replication can demonstrate that, when the assumptions of models are violated, the results of an original report may be misleading, and a replication attempt can be used to highlight both the robustness of the effect in question and the limitations of a statistical model employed to test that effect. Second, replication can be connected to topics such as Type
I error and \( p \) values, thereby enhancing student understanding of these foundational topics. In particular, a replication study that fails to reproduce the statistically significant results of an original study may draw into question whether the findings of the original study were “false positives.” This can also relate to a lecture on effect size, where the true size of an effect may be hard to glean from a single study, and thus replications are necessary to more directly estimate the effect size (see also Chapter 14 on meta-analysis). Further, because replication studies traditionally take into consideration the sample size needed to detect a predicted effect with adequate statistical power, a lecture on replication can also be used to discuss the importance of power analysis in planning sample sizes. In all instances, replication can be used as a powerful tool to promote critical thinking about statistics.

How Should I Assess Student Learning about Replication?

Student learning can be evaluated in a number of ways. Our “Replication Vignettes” can be used as an in-class or take-home assignment with questions that evaluate student learning about replication and related topics. Instructors may also include a written assignment in which students are instructed to compare the results of an original study with a replication study. A paper can allow students to discuss the importance of replication in their own words, allowing them to evaluate statistical findings and think critically about the reported results. If instructors elect to evaluate student learning via a written assignment, they should include enough information to help students find a replication study (e.g., by including a list of possible replication studies). One way this may be accomplished is by consulting the list of attempted replication studies from the CREP (at https://osf.io/m8et3/wiki/home/). There, previous teams who have attempted replications of well-cited studies share their data and written summaries of the data, which can be accessed freely. Finally, instructors may include questions about replication on an exam, which can include a mix of multiple-choice (e.g., “Did the replication study successfully replicate the results of the original study?”) and short-answer questions (e.g., “Given that a sample size of 100 is necessary to obtain a power of .8 to detect the effect reported in the original study, discuss why the results of the original and replication study differed.”)

Conclusions

One of the APA’s broad goals for the undergraduate psychology major is to enhance students’ scientific inquiry and critical thinking skills by expanding their knowledge base in psychology (American Psychological Association, 2013). As psychology continues to evolve through its “reproducibility movement,” it is imperative that the core tenets of replication are taught to undergraduate students, to broaden their understanding of the field. In this chapter, we proposed that replication can be taught to students in statistics classes, thereby fostering a deeper understanding of statistics and encouraging students to be more critical consumers of scientific research. We outlined several activities that instructors can use to teach students about the importance of replicating well-cited studies to better understand the generalizability of an effect. By incorporating these activities into undergraduate statistics classes, instructors can strengthen students’ understanding of statistics and its importance to psychological research.
References


Appendix A

**Vignette on Replication and Replication Recipe**

The color red has been linked to enhanced female attraction to males in many species of vertebrates (e.g., Setchell, 2005, *International Journal of Primatology, 26*). Does the color red have a similarly enhancing effect on women’s attraction to men?


Elliot et al. tested the idea that women find photographs of men to be more attractive when those photographs contain red borders. In their Experiment 3, heterosexual women (n = 33), with a mean age of 19.54 years (range from 18 to 25) viewed black-and-white photos of a Latino-American, college-aged man on either a red or gray background and rated the man’s attractiveness. Perceived attractiveness was assessed with a 3-item questionnaire (e.g., “How attractive do you think this person is?”; α = .94), for which participants rated each item on a 9-point scale (from 1 = “not at all attractive” to 9 = “extremely attractive”). An independent-samples t test was conducted to test whether perceived attractiveness differed between the two color conditions. Results revealed a significant color effect, t(32) = 2.44, p < .05, d = 0.86, with participants in the red condition (M = 6.69, SD = 1.22) rating the man as more attractive than those in the gray condition (M = 5.27, SD = 2.04).


Nine student-led replication projects of Elliot et al.’s (2010) Experiment 3 were carried out and synthesized in a meta-analysis, with a total sample of n = 581 heterosexual or bisexual women. The median sample size for each replication study was n = 59. The mean age of the participants was 20.53 (SD = 3.18). Participants viewed black-and-white photos of a similar Latino-American, college-aged male as in the study by Elliot et al., and then rated their perceived attractiveness using the same 3-item questionnaire (α = .89). The procedures of the replication studies were almost identical to those of Elliot et al.’s (2010) Experiment 3, with a couple minor differences. The original study tested participants individually, but two of the replication studies tested participants in pairs (although none of the participants could see each other’s responses). One study also added a yellow condition to the original materials. In all 9 replication studies, there was no significant difference in perceived attractiveness based on color condition. The average (unweighted by sample size per study) attractiveness ratings, across studies, for the man with a red background (M = 5.72) and with a gray background (M = 5.80) did not differ, p = .53, d = -0.07 (95% CI: -0.31, 0.16).

**Question 1.** What is the effect size of the original study? Of the replication studies?

**Question 2.** Considering the difference in sample sizes between the original study and the median sample size of the replication studies, which effect size estimate is more likely believable?

**Question 3.** Calculate the critical value for the independent samples t statistic at an alpha of .05 for Experiment 3 of Elliot et al. (2010).

**Question 4.** Are these replication studies direct replications, conceptual replication, or direct with extensions? Explain.
Appendix B

Sample Questions from the “Replication Recipe” (Brandt et al., 2014, p. 219, Table 1)

“The effect size of the effect I am trying to replicate is _______”
“The sample size of the original effect is _______”
“My target sample size is _______”
“The rationale for my sample size is _______”
“What differences between the original study and my study might be expected to influence the size and/or direction of the effect?”
“My analysis plan is” (justify differences from the original)
“I would define a successful replication as _______”
“The effect size of the replication is _______ and its confidence interval is _______”
“The replication effect size is significantly different from the original effect size?” YES or NO
“The limitations of my replication study are______________”
Reject Rejecting the Null Hypothesis: Using Confidence Intervals to Encourage Meta-analytic Thinking

Thomas Cavanagh, PhD
Dominican University of California

Summary

Whereas most the chapters in this book focus on the use of pedagogical approaches to teaching statistics, this chapter will instead focus on the teaching of a specific statistical concept, namely the value of interpreting confidence intervals as an alternative to null hypothesis significance testing (NHST). Despite being criticized essentially since its inception for convoluted interpretations which create the appearance of clear, dichotomous results, NHST remains dominant across the scientific literature. Recently, individual researchers and scientific organizations have encouraged the use of the “new” statistics, including confidence intervals, to inspire more nuanced thinking about empirical data. I will begin this chapter by critiquing the false dichotomy of NHST and discussing how interpreting confidence intervals can encourage meta-analytic thinking. Next I will introduce an activity that demonstrates how using confidence intervals, as opposed to NHST, can lead to very different interpretations of the same study results. Finally, I will discuss the outcomes I hope this activity will achieve for students.

Researchers are rejecting null hypothesis significance testing

Null hypothesis significance testing (NHST), driven by the interpretation of p-values, has dominated scientific research for nearly a century (Cumming, Fidler, Kalinowski, Lai, 2012). Recently, though, many researchers, organizations, and journals have begun advocating instead for the application of the “new” statistics; that is, interpreting confidence intervals as opposed to p-values and embracing meta-analytic thinking as opposed to hypothesis testing (Cumming et al., 2012). Below, I describe NHST, p-values (and how they are often misinterpreted), and the benefits of reporting and interpreting confidence intervals in their stead.

Null hypothesis significance testing tells us when there is not not an effect

Imagine that I designed a new childcare initiative to increase job satisfaction at my university, and I want to assess whether or not it has a positive impact. I measure pre- and post-childcare job satisfaction scores and then, just like I was taught in graduate school, I devise two hypotheses:

H0: The “null” hypothesis. There is no significant difference between pre- and post-childcare job satisfaction scores.

H1: The “alternative” hypothesis. There is a significant difference between pre- and post-childcare job satisfaction scores.

I run a paired-sample t-test and calculate the following results:

Post-childcare job satisfaction scores ($n=15$, $m=4.40$, $sd=2.29$) were significantly higher than pre-childcare job satisfaction scores ($n=15$, $m=3.80$, $sd=2.15$), $mean\ difference = 0.60$, $t(14)=2.201$, $p=.045$. 

Because my \( p \)-value is below .05, I can reject the null hypothesis. And because I can reject the null hypothesis, the common thinking goes, I then accept the alternative hypothesis and say, yes, my intervention worked— it increased job satisfaction by 0.60 points on average.

There is, however, a severe misstep in logic here. With a \( p \)-value, the only thing I am testing is whether or not I can confidently reject the null hypothesis (Inman, 1994). The literal interpretation of the \( p \)-value above, for example, would be that I am 95.5% confident there is not a difference between pre- and post-childcare job satisfaction scores. Although the \( p \)-value allows me to estimate how confident I am that the null hypothesis is incorrect, it does not provide any evidence that the alternative hypothesis is correct (Inman, 1994). Of course, the central focus of research is usually the alternative hypothesis, not the null hypothesis, and researchers interpret their \( p \)-values accordingly. They reject the null hypothesis, and then, like an adept slight-of-hand magician, discuss the results as if they support the alternative hypothesis.

**Confidence intervals provide a range of possible values for an effect**

In my interpretation above I did not just use the \( p \)-value as justification to reject the null hypothesis, I argued the \( p \)-value provided evidence in support of the alternative hypothesis, and I provided a point value for its effect (the mean difference of 0.60 job satisfaction points). My \( p \)-value, however, is not telling me that job satisfaction in my sample increased by an average of 0.60 points. It is only telling me that I can be pretty confident that there was not no change in job satisfaction. The \( p \)-value is in no way confirming my alternative hypothesis. Researchers know this, of course, and never come right out and say that their alternative hypothesis is correct. Instead they use vague language and say things like, “provides evidence in support of.”

Let us imagine a different scenario. This time, instead of reporting a \( p \)-value, I report a 95% confidence interval. Confidence intervals and \( p \)-values are intimately connected, and calculated, using essentially the same information. Instead of results being reported as a single point estimate, however (as they often are with \( p \)-values), confidence intervals provide a range of plausible values within a given confidence level for the effect in the population given the information from our sample. For example:

Results revealed an average mean difference of 0.60 points between pre- and post-childcare measures of job satisfaction, 95% CI [.015, 1.185].

Those values mean that if I were to collect 100 samples from the same population, 95 of those samples would show a mean difference between pre- and post-childcare job satisfaction scores of between .015 and 1.185 points. This is a much more nuanced way of thinking about our results compared to the dichotomous, accept-or-reject logic of NHST (Coulson, Healey, Fidler, & Cumming, 2010).

This nuance encourages researchers, and consumers of research, to focus on aggregating common data to form a coherent picture of scientific phenomena, whereas hypothesis testing highlights artificial differences (Coulson et al., 2010). For example, say a rival professor sees my study on childcare and decides to run his own study. He is not quite as good a researcher as me, of course, so the study is not quite as successful. The effect is just a tad smaller, and his results fail to attain statistical significance:

Post-childcare job satisfaction scores \((n=15, m=4.33, sd=2.19)\) were significantly higher than pre-childcare job satisfaction scores \((n=15, m=3.80, sd=2.15), mean\ difference = 0.533, t(14)=2.086, p=.056.\)

He writes a long article lambasting my results and calling me a fraud. Because we are nuanced followers of the new statistics, however, we go back to his data, calculate his 95% confidence interval, and find the following:
Results revealed an average mean difference of 0.53 points between pre- and post-childcare measures of job satisfaction, 95% CI [-0.015, 1.08].

By displaying the confidence intervals instead of the p-value, we actually see agreement between his study and mine, not contradiction (Coulsen et al., 2010). He estimates a mean population effect of .53 with a 95% confidence interval of -0.015 and 1.08; I estimated a mean population effect of 0.60 with a 95% confidence interval of 0.015 and 1.185. If we look only at p-values we reason that one study was significant, the other was not, and we see disagreement. If we look at the confidence intervals, however, we see two studies pointing us towards the same conclusion, that measures of job satisfaction increased after the childcare intervention. It also clearly illustrates what we don’t know: the exact size of the effect.

This is the peril of p-values. They don’t test the alternative hypothesis, but by employing bad logic we use rejection of the null hypothesis in support of the alternative hypothesis (Inman, 1994). Researchers provide point estimates of means, mean differences, or correlation coefficients, and then use rejection of the null hypothesis to act as if those point estimates are accurate. In actuality, p-values are not very useful for estimating the size of an effect in a population, whereas confidence intervals are (Finch & Cumming, 2009).

Followers of the “new” statistics advocate for the use of confidence intervals and meta-analytic thinking

If this all seems fairly ridiculous to you, you are not alone. Pearson and Fisher, two of the researchers who first developed and popularized p-values (also enthusiastic eugenicists, may their names be stained forever), strongly opposed using p-values for NHST (Inman, 1994). Both of them, however, continued to use NHST in their own research, setting a precedent to this day where, as a field, we condemn the fetishizing of p-values and then continue to use them as the ultimate criterion of a “successful” study (Inman, 1994).

Recently, there has been a serious effort to fundamentally change how researchers think about and use p-values, NHST, and confidence intervals. Led by charming Australian grandfather Geoff Cumming and scientific organizations like the American Psychological Association, supporters of the new statistics advocate for the use of confidence intervals as opposed to point estimates, and essentially ignoring p-values all together in favor of what they call “meta-analytic” thinking: that is, the awareness that no single study can prove or disprove the existence of an effect, but that each study represents one data point in an infinitely large population of data points that give us information about our phenomenon of interest (Cumming et al., 2010). By rejecting the dichotomous thinking of NHST, we can focus on agreement in our data instead of false differences, and think about statistical results, and scientific phenomenon generally, in a more nuanced and useful way.

The “new” statistics provide a more nuanced view of statistical results and scientific phenomena

In sum, using p-values for NHST encourages dichotomous thinking that creates the appearance of conflict when there is largely agreement, and gives researchers a false sense of confidence in their point estimates. Confidence intervals, on the other hand, provide researchers with a range of possible values for the impact of effects, and encourage them to see a single study as just that- one data point in a theoretically infinite distribution of data points concerning the phenomenon of interest. Below I will describe an activity that can be used to bring this contradiction to life visually and experientially for students.
Methods

The appendices of this chapter provide materials to facilitate an activity demonstrating how the interpretation of NHST and confidence intervals can lead to very different conclusions. For each sheet, questions for the students are in normal font, whereas expected responses/teaching notes are italicized.

The activity is based on a fictional experiment in which a CEO investigates whether hiring death metal bands to play in her cupcake factories increases bakers’ productivity, as estimated by pre- and post-death metal measures of the average number of cupcakes made per baker per day. The CEO runs her study in two factories: the Red Velvet Factory and the Funfetti Factory (both factories make the same assortment of cupcakes, they’re just identified using cupcake names).

Part I: p-values

Hand out the Red Velvet Factory results (Appendix A: Red Velvet, p-values) to half the students, and the Funfetti Factory results (Appendix B: Funfetti, p-values) to the other half. Give students approximately five minutes to review the results and answer the discussion questions. Then, pair students together with other students who received results from the same factory, and have them discuss their answers for another five minutes or so.

Red Velvet discussion: p-values

After students are finished discussing their answers in pairs, draw a line on the white board at the front of the room, starting at zero and ending at seven, with tick marks placed every half-unit (i.e., 0, 0.5, 1.0, 1.5, etc.). Ask a student who received the Red Velvet Factory results to come up to the board and mark the mean difference from their study on the line. Go through the discussion questions provided on the sheet, getting answers only from students in the Red Velvet group first.

Specific teaching notes are provided in the appendices, but, essentially, students familiar with interpreting p-values in light of NHST should report that a significant difference was found between the pre- and post-death metal measures of productivity. They should interpret this as evidence of the efficacy of the intervention.

Students in the Red Velvet group are also given a point estimate of the mean difference (1.13) and information about the logistics of the cupcake factory, including number of bakers (1,000), how much profit the company earns from each cupcake ($2), and how much the death metal band costs ($2,200 per day). Using this information, students are asked to decide, based purely on financial outcomes, whether the factory should continue to invest in the death metal intervention.

Using the above information, students who received the results for the Red Velvet Factory should calculate a net profit increase of $60 per day (see teaching notes for calculations). From a purely financial perspective, their advice should be to continue with the intervention.

Funfetti discussion: p-values

Go through the same questions with the students who received the Funfetti Factory results, being sure to have them mark their point estimate for the mean difference on the line on the board as well. In this case, students should report that no significant difference was found between the pre- and post-death metal measures of productivity, and they should interpret this as lack of evidence of the efficacy of the intervention.

Students who received the Funfetti Factory results are given the same logistical information about the factory as the Red Velvet students, and asked the same question about the financial
impact of the death metal intervention. Students might refuse to provide an estimate because of
the non-significant p-value, but if they do calculate an estimate, it should be a net change of -
$66 (i.e., a $66 net loss per day; see teaching notes for calculations). Given the above
information, students should recommend that, from a purely financial perspective, the factory
should discontinue the death metal intervention.

**Group discussion: p-values**

After both groups have reviewed their answers, lead a group discussion with the students as a
whole group. Although many interesting points may come up, the emphasis should be on
whether students see more agreement or disagreement in the findings of the two studies.
Through its accept-or-reject logic, NHST encourages dichotomous thinking, and students will
likely highlight the perceived differences in the results. Some discussion questions you can use
include:

- Are you surprised by the other group’s findings?
- Does finding out about the other group’s results make you more confident or less
  confident than you were having only read your own results?
- Did the different results influence your opinion regarding whether or not the death metal
  intervention should continue?
- Was there more agreement or more disagreement between the results of the two
  studies?
- Does this situation ever arise in scientific research?

**Part II: confidence intervals**

Next, provide students with “Appendix C: Red Velvet, CI” or “Appendix D: Funfetti, CI” handouts
(students should receive the results from the same factory they used in Part I). This handout
provides the same information as before, but this time presented as confidence intervals as
opposed to p-values. Again, give students five minutes to answer the questions on their own
and five minutes to discuss with a partner who received the same results.

**Red Velvet discussion: Confidence intervals**

After students are finished discussing their answers, ask a student who received the Red Velvet
Factory results to come up to the board and add the high and low estimates from the 95%
confidence intervals onto the line which already has the mean differences. Then, once again, go
through the discussion questions provided on the sheet, getting answers only from students in
the Red Velvet group. Ensure that they interpret the confidence intervals correctly.

Given this new information, ask students whether, from a purely financial perspective, they
would recommend that the intervention should continue. This time, instead of calculating a
single number for the financial impact of the intervention, students who should calculate a range
of possible numbers, from a net loss of $2,124 to a net gain of $2,266 (see teaching notes for
calculations). Students should now be much less confident about recommending that the
intervention continue. There is a large likelihood that the effect is small enough that the
increased productivity will not offset the cost of the band. They should also, however,
recommend that the intervention continue to be tested; at the high end of the confidence interval
the factory could be increasing net profits by over $2,000 per day.

**Funfetti discussion: confidence intervals**

Next, have students from the Funfetti Factory group draw their confidence interval on the board
on top of the Red Velvet Factory results. Go through the discussion questions and ensure that
they interpreted the confidence intervals correctly. Then ask them if they thought that, from a purely financial perspective, the intervention should continue. Students in this group should question their original decision, as well. Although the point estimate for the mean difference provided in the previous results did not financially justify continuing with the intervention, the new 95% CI they calculated shows that there could potentially be significant financial upside (an additional $1,186 in net profit per day; see teaching notes for calculations). Like their Red Velvet classmates, the Funfetti students will likely say that the results are inconclusive but the potential upside merits more research.

**Group discussion: Confidence intervals**

After both groups have reviewed their new answers, lead a group discussion with the students as a whole group. The emphasis, again, should be on whether students see more agreement or disagreement in the findings of the two studies. As compared to their earlier interpretations using NHST, students should now see the results as fundamentally in agreement, and a potential opportunity for more research. It might be helpful to review the same questions listed in the first group discussion above, and to note how using confidence intervals changed students’ answers to these questions. Additionally, you should explore students’ opinions on the relative strengths and weaknesses of NHST and confidence intervals. Questions might include:

- What are the strengths of NHST?
- What problems do you see with NHST?
- What are the strengths of confidence intervals?
- What problems do you see with confidence intervals?
- Are proponents of the new statistics correct? Should we encourage reporting confidence intervals as compared to p-values in scientific research? Why or why not?

**Confidence intervals encourage more nuanced thinking about statistical results**

The focus of this chapter was to critique the use of p-values and the false-dichotomy logic of NHST, to encourage the greater use of confidence intervals to encourage more nuanced scientific thinking, and to describe an activity that illustrates the value of these concepts to students. Ultimately, the goal of the chapter, and the activity, is to motivate students to question the accept-or-reject logic of NHST, and embrace the ambiguity of results as presented by confidence intervals.
References


Appendix A

Red Velvet Factory Results

p-values

The CEO of a cupcake empire loves death metal, and is convinced that if she hires death metal bands to play live music eight hours a day in her factories, productivity will skyrocket. Before instituting her idea companywide, she decides to pilot test it in two of her factories: the Red Velvet Factory and the Funfetti Factory (all the factories make the same assortment of cupcakes, they’re just identified using cupcake names). She selects a random sample of 30 bakers from each factory, measures current levels of productivity (how many cupcakes a given baker makes per day), institutes the death metal intervention for six weeks, and then measures productivity again. She posits the following two hypotheses:

- **H₀**: The difference between the average number of cupcakes made per baker per day before and after the death metal intervention is 0.
- **H₁**: The difference between the average number of cupcakes made per baker per day before the death metal intervention and after the death metal intervention is not 0.

She collects the data and runs a paired-sample t-test for each of the factories. The results from the Red Velvet Factory are described below:

The difference between the pre-death metal average number of cupcakes made per baker per day (n=30, m=40, sd=6.716) and post-death metal average (n=30, m=41.13, sd=7.253) was significant, mean difference=1.13, t(29)=2.108, p=.044.

**Using the information above, please answer the following questions:**

Interpret the null and alternative hypotheses in light of the p-value.

- **H₀**: Given our significant p-value, we **reject the null hypothesis**.
- **H₁**: Given our significant p-value, we **accept the alternative hypothesis**.
We can be 95.56% confident that the death metal intervention did not affect the average number of cupcakes made per baker per day.

Interpret the meaning of the findings. Was the death metal band successful at increasing the number of cupcakes made each day? Explain your answer.

Students will likely interpret the significant p-value as providing evidence that playing death metal significantly increased the average number of cupcakes made per baker per day.

Review the information in the table below. From a purely financial standpoint, would you recommend the CEO continues to pay for death metal bands to perform every day?

<table>
<thead>
<tr>
<th>Number of bakers in the factory</th>
<th>1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profit per cupcake</td>
<td>$2</td>
</tr>
<tr>
<td>Cost of band per day</td>
<td>$2,200</td>
</tr>
</tbody>
</table>

Students will likely use the point value estimate of the mean difference to calculate how much extra profit the organization made:

\[
1,000 \text{ bakers} \times 1.13 \text{ extra cupcakes per day} = 1,130 \text{ extra cupcakes per day} \\
1,130 \text{ cupcakes} \times $2 \text{ profit} = $2,260 \text{ in extra gross profit per day} \\
$2,260 \text{ extra gross profit} - $2,200 \text{ for band} = $60 \text{ extra net profit per day}
\]

Given the above calculations, from a purely financial standpoint it makes sense to continue to hire death metal bands to play every day.
Appendix B

Funfetti Factory Results

p-values

The CEO of a cupcake empire loves death metal, and is convinced that if she hires death metal bands to play live music eight hours a day in her factories, productivity will skyrocket. Before instituting her idea companywide, she decides to pilot test it in two of her factories: the Red Velvet Factory and the Funfetti Factory (all the factories make the same assortment of cupcakes, they’re just identified using cupcake names). She selects a random sample of 30 bakers from each factory, measures current levels of productivity (how many cupcakes a given baker makes per day), institutes the death metal intervention for six weeks, and then measures productivity again. She posits the following two hypotheses:

\[
H_0: \text{The difference between the average numbers of cupcakes made per baker per day before and after the death metal intervention is 0.}
\]

\[
H_1: \text{The difference between the average numbers of cupcakes made per baker per day before the death metal intervention and after the death metal intervention is not 0.}
\]

She collects the data and runs a paired-sample t-test for each of the factories. The results from the Funfetti Factory are described below:

The difference between the pre-death metal average number of cupcakes made per baker per day (n=30, m=40.20, sd=6.87) and the post-death metal average (n=30, m=41.17, sd=7.30) was significant, mean difference=0.967, t(29)=1.84, p=.077.

Using the information above, please answer the following questions:

Interpret the null and alternative hypotheses in light of the p-value.

Students’ answers will hopefully include the following interpretation:

\[H_0: \text{Given our non-significant p-value, we fail to reject the null hypothesis.}\]
H₁: Given our non-significant p-value, we find no evidence in support of our alternative hypothesis.
We are not confident that death metal music impacted employee productivity.

Interpret the meaning of the findings. Was the death metal band successful at increasing the number of cupcakes made each day? Explain your answer.
Because the p-value failed to achieve significance, we don’t find any evidence that the death metal intervention increased productivity.

Review the information in the table below. Given these values, would you recommend the CEO continues to pay for the death metal band to perform every day?

<table>
<thead>
<tr>
<th>Number of bakers in the factory</th>
<th>1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profit per cupcake</td>
<td>$2</td>
</tr>
<tr>
<td>Cost of band</td>
<td>$2,200</td>
</tr>
</tbody>
</table>

1,000 bakers X 0.967 extra cupcakes per day = 967 extra cupcakes per day
967 cupcakes X $2 profit = $1,934 in extra gross profit per day
$1,934 additional gross profit - $2,200 for band = -$266 net profit change per day
(i.e., a net loss of $266 per day)

Given the information above, from a purely financial perspective it does not make sense to continue with the intervention.
Appendix C

Red Velvet Factory Results

Confidence Intervals

One of the CEO’s managers is skeptical of her results, and decides to look at them using alternative statistical techniques, specifically, confidence intervals. His findings for the Red Velvet Factory are reported below.

There was an observed mean difference of 1.132, 95%CI [0.038, 2.233], between number of cupcakes made per day before the beginning of the death metal intervention (n=30, m=40, sd=6.72) and six weeks after the intervention was instituted (n=30, m=41.13, sd=7.25).

Use the information above to answering the following questions:

Interpret the confidence interval.

We are 95% confident that the true population mean difference in cupcake productivity before and after the death metal intervention is between 0.038 and 2.233.

Interpret the meaning of the findings. Was the death metal band successful at increasing the number of cupcakes made each day? Explain your answer.

The findings do provide some evidence that the death metal intervention did increase the average number of cupcakes each employee made per day. The effect could potentially be as low as .038, or as high as 2.233. We are uncertain where the true population mean difference falls within that range.

Review the information in the table below. From a purely financial perspective, would you recommend the CEO continues to pay for the death metal band to perform every day?

<table>
<thead>
<tr>
<th>Number of bakers in the factory</th>
<th>1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profit per cupcake</td>
<td>$2</td>
</tr>
<tr>
<td>Cost of band per day</td>
<td>$2,200</td>
</tr>
</tbody>
</table>
Students should calculate the change in profit based on the lower 95% confidence interval estimate:

\[
1,000 \text{ bakers} \times 0.038 \text{ extra cupcakes per day} = 38 \text{ extra cupcakes per day} \\
38 \text{ cupcakes} \times 2 \text{ profit} = 76 \text{ extra gross profit per day} \\
76 \text{ extra gross profit} - 2,200 \text{ for band} = -2,124 \text{ net profit change per day (i.e., a net loss of } 2,124 \text{ per day)}
\]

Students should then calculate the change in profit based on the higher 95% confidence interval estimate:

\[
1,000 \text{ bakers} \times 2.233 \text{ extra cupcakes per day} = 2,233 \text{ extra cupcakes per day} \\
2,233 \text{ cupcakes} \times 2 \text{ profit} = 4,466 \text{ extra gross profit per day} \\
4,466 \text{ extra gross profit} - 2,200 \text{ for band} = 2,266 \text{ extra net profit per day}
\]

Given this wide range, it’s not clear whether the intervention is worth the investment or not. Students should come to the conclusion that the results are ambiguous, and more research needs to be done before we can decide whether this was a valuable intervention or not.
Appendix D

Funfetti Factory Results

Confidence Intervals

One of the CEO’s managers is skeptical of her results, and decides to look at them using alternative statistical techniques, specifically, confidence intervals. His findings for the Funfetti Factory are reported below.

There was an observed mean difference of 1.133, 95%CI [-.110, 2.043] between number of cupcakes made per day before the beginning of the death metal intervention (n=30, m=40.20, sd=6.87) and six weeks after the intervention was instituted (n=30, m=41.17, sd=7.302).

Use the information above to answering the following questions:

Interpret the confidence interval.

Student answers will hopefully approximate the following interpretation:

*We are 95% confident that the true population mean difference in cupcake productivity before and after the death metal intervention is between -0.110 and 2.043.*

Interpret the meaning of the findings. Was the death metal band successful at increasing the number of cupcakes made each day? Explain your answer.

*The findings do provide some evidence that the death metal intervention did increase the average number of cupcakes each baker made per day. The effect could potentially be as low as -.110, or as high as 2.043. We are uncertain where the true population mean difference falls within that range.*

Review the information in the table below. Given these values, would you recommend the CEO continues to pay for the death metal band to perform every day?

<table>
<thead>
<tr>
<th>Number of bakers in the factory</th>
<th>1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profit per cupcake</td>
<td>$2</td>
</tr>
<tr>
<td>Cost of band</td>
<td>$2,200</td>
</tr>
</tbody>
</table>
Students should calculate the change in profit based on the lower 95% confidence interval estimate:

1,000 bakers X -0.110 extra cupcakes per day = 110 fewer cupcakes per day
-110 cupcakes X $2 profit = $220 in lost gross profit per day
-$220 loss gross profit - $2,200 for band = -$2,420 net profit change per day (i.e., a net loss of $2,420 per day)

Students should then calculate the change in profit based on the higher 95% confidence interval estimate:

1,000 bakers X 2.043 extra cupcakes per day = 2,043 extra cupcakes per day
2,043 cupcakes X $2 profit = $4,086 extra gross profit per day
$4,086 extra gross profit - $2,200 for band = $1,886 extra net profit per day

Given this wide range, it’s not clear whether the intervention is worth the investment or not. Students should come to the conclusion that the results are ambiguous, and more research needs to be done before we can decide whether this was a valuable intervention or not.
Part Five: Open Data Analysis Software
Using Jamovi in an Undergraduate Psychological Statistics Course: A Free Alternative that Promotes Conceptual Understanding and Active Learning

Abby Heckman Coats¹, PhD and Andrew Mienaltowski², PhD
Westminster College¹, Western Kentucky University²

Summary
Software is often an integral part of psychological statistics courses. As undergraduates learn how to interpret, conduct, and apply statistical techniques, hands-on experience with the software that psychologists use is beneficial. Many textbooks include instruction on specific software packages, such as SPSS. However, SPSS has not been substantially updated in several years, and it is becoming increasingly expensive each year. Thus, many universities and psychology organizations have stopped using SPSS (Muenchen, 2019). This chapter describes our experience with switching from SPSS to the free alternative Jamovi. We describe how we use Jamovi to teach important statistical concepts relevant to t-tests, ANOVAs, correlations, and regression. We describe pitfalls to avoid so that others can learn from our experience. We share questionnaire and exam performance data that shows that Jamovi is user-friendly and associated with good learning outcomes.

Why Should You Use Jamovi in Statistics Courses?
Despite the importance of undergraduate statistics courses to the development of a conceptual foundation for hypothesis testing and data visualization, many students do not realize how important statistics will be in their professional lives. Students are often anxious about performing mathematical operations, and few have been asked to do so with a purpose in mind. Indeed, the math skills of undergraduates have declined in recent years (Carpenter & Kirk, 2017). Students who have negative attitudes about their cognitive competence or who have poor statistics self-efficacy at the beginning of the semester have poorer final exam performance (Dempster & McCorry, 2009; Salim, Gopal, & Ayub, 2018). Nevertheless, statistical analysis is an essential tool for psychology majors to learn (Perlman & McCann, 1999). All branches of psychology are linked to study design and carefully planned interventions or observations. Thus, it is not surprising that completing statistics early in one’s undergraduate years predicts higher GPAs in upper-division psychology courses, even after controlling for ACT scores and prior GPA (Freng, Webber, Blatter, Wing, & Scott, 2011). Programs requiring at least two statistics and methodology courses lead to earlier completion of statistics requirements (Lauer, Rajecki, & Minke, 2006). This is especially useful for students who actively take part in data collection via independent study or lab projects.

Yet, statistics courses for psychology majors are often a hard sell for students. To make the course more enjoyable, active learning and problem-based learning can be used. Active learning improves motivation and leads to improved outcomes. Application exercises promote far transfer of statistical knowledge to novel situations (Daniel & Braasch, 2013). For example, students who identified whether statistical concepts were correctly used in news articles were able to also use their statistical knowledge to critically analyze the use of websites such as ratemyprofessor.com. Problem-based learning is another effective technique for improving motivation and learning in statistics courses (Karpiak, 2011). Students using statistics software
as they work in cooperative learning groups to solve complex real-world problems performed better on a statistics assessment administered the following semester compared to students in traditional lecture-based courses. Using statistics software such as Jamovi to accomplish these goals helps students achieve GAISE learning outcomes #4 (Foster active learning) and #5 (Use technology to explore concepts and analyze data).

Statistics software is an important tool for implementing these active learning techniques. In the real world, psychologists and other professionals use software to do their jobs. Quantitative literacy and software skills are among the most sought-after job skills among employers (National Association of Colleges and Employers, 2014). Using a software package that will always be available to them (such as Jamovi) gives students an important advantage in the workplace. In addition, statistical software improves student engagement, motivation, and learning compared to paper-and-pencil based tutorials (Hartnett, 2013; King, Malcolm-Smith, Jaftha, Louw, & Tredoux, 2013). Statistics software can be an effective tool for flipped classrooms and sandwich classrooms (See Chapter 3 for more information on flipped classrooms). In these types of classes, “homework” consists of reviewing information that traditionally would be presented in lectures, whereas in-class work is devoted to applied problems (Akcayir & Akcayir, 2018; Hoepner & Hemmerich, 2018; Wilson, 2013). Best practices for flipping statistics courses include using videos, interactive tutorials, and narrated slides, rather than relying solely on readings (Peterson, 2016). Using a free, user-friendly software package like Jamovi makes this easier because students are comfortable exploring Jamovi on their own. Many excellent free tutorials and videos are available online (See Appendix D).

Statistics software application is an important job skill, connects to capstone courses, and facilitates experiential learning experiences such as student research. But where are students learning to use statistics software? Many students have their first experience with statistics software in their undergraduate introductory statistics course. Even nearly 20 years ago, roughly 70% of undergraduate psychology programs utilized statistical software, with most opting to use SPSS (Bartz & Sabolik, 2001). In fact, students report a preference to use both software-based and problem-based tutorials (King et al., 2013). We have also found that pairing conceptual hands-on calculations with statistical software training from lesson to lesson is an effective way to increase student engagement as they learn course material. In the past, these statistical packages were expensive and were available to students on time-limited licenses that were not renewed once the class ended. In particular, SPSS has ended most price-discounting programs for students. As the price has risen and the parent company (IBM) has focused on corporate sales, many universities and psychology organizations have stopped using SPSS (Muenchen, 2019). Jamovi, JASP, Vassar, PSPP, and R are free open-source alternatives to SPSS. Each has advantages and disadvantages (see Chapter 19 for more information about R and Chapter 20 Compendium for resources for other options). In our opinion, Jamovi has the greatest potential for teaching students basic statistics skills while also providing options for incorporating more advanced topics (Jamovi runs on the R programming language). With the emergence of powerful, free statistical tools like Jamovi, we can incorporate statistical software training into our classrooms, ask students to download and use this software, and then further foster the growth of students’ skills after the statistics course via independent study.

**How to Use Jamovi in Statistics Classes**

Although it takes planning to incorporate Jamovi into a statistics course, students cultivate an understanding of tools that they can use on their own when the course is over. From a pedagogical standpoint, embedding statistics software training into one’s course empowers students by deemphasizing the tedious mathematical operations that professional psychologists rarely perform and instead emphasizes the application of concepts taught in the course. Moreover, if brought into the picture when introducing tests, the statistical output generated from
the software might help to keep their focus on the research question being addressed instead of losing the students to their math insecurities.

In what follows, we will provide some examples for how you can use Jamovi to introduce common statistical concepts. Appendix A contains example slides that you can use to teach students how to download and begin using the program. When you download Jamovi for your own use, you will find that there are datasets that you can use. We often build sample datasets in our courses by generating random data in a spreadsheet and importing the data into Jamovi, which readily accepts .csv (or comma separated values) files. Most spreadsheets have functions that will allow you to generate random numbers within specific ranges. Careful manipulation of these ranges can establish data that will illustrate significant test outcomes to students. If you have a hard time generating examples, you could also adapt those found in textbooks. You could also use archival data sets that you find online or can adapt from your own research projects. Another option is to use an established dataset generator such as: http://rlanders.net/dataset-generator/ (listed in online resources shared in Appendix D)

T-tests

Independent samples and paired samples \( t \)-tests mark a transition into applied analysis that should lend well to generating examples. Underlying these two types of tests is the notion that the calculated test value reflects a ratio of between-conditions variability to within-conditions variability, or error. The larger the ratio is, the more likely it is that between-conditions variability surpasses error and a significant statistical difference emerges between the two means being compared. For both \( t \)-tests, the numerator of the test is easily calculated. However, calculating the denominator of each test becomes a nightmare of mathematical operations – with pooled variance for independent samples \( t \)-tests and calculating sums of squares using mean difference scores for paired-samples \( t \)-tests. After conceptually introducing either test, discussing a Jamovi example of the test enables us to discuss the key components of the statistical comparison before delving into the math.

Consider the following example for an independent samples \( t \)-test. A clinical psychologist compares her clients' depression symptoms that have been measured using the Center for Epidemiological Studies Depression (CES-D) Scale. Some clients were assessed before treatment could begin, and other were assessed after 8 weeks of Cognitive Behavioral Therapy (CBT). The hypothesis is that individuals who receive CBT for 8 weeks will have lower CES-D scores than those who have not yet begun treatment. The data file within Jamovi appears above in Figure 1. Note that the two groups do not have the same sample size, an issue that complicates the mathematical operations for the \( t \)-test.

By selecting "Independent Samples T-Test" from the T-Tests section of the Analyses menu (Figure 1), you will notice how similar Jamovi’s command window is to the graphic user interface of other statistics packages. Set Treatment as your Grouping Variable and CES-D as your Dependent Variable, and Jamovi calculates the output for the test. You can specify additional statistics that you may also discuss, like confidence intervals and effect size. Figure 2 displays the command window for this test. The user chooses what options to select within the Independent samples \( t \)-test options. The Grouping Variable box is where the independent variables goes.
Figure 1. Independent samples t-test example data in Jamovi.

Figure 2. Independent samples t-test command window in Jamovi (split box in two).

The output for the study is depicted in Figure 3. Clients who received CBT have significantly lower CES-D scores than those who had not yet been treated. By asking your students leading questions, you can encourage them to think critically about the outcome of the test and to develop their understanding for what the output reflects.

**Independent Samples T-Test**

<table>
<thead>
<tr>
<th>statistic</th>
<th>df</th>
<th>p</th>
<th>Mean difference</th>
<th>SE difference</th>
<th>95% Confidence Interval</th>
<th>CES-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student's t</td>
<td>4.21</td>
<td>14.0</td>
<td>&lt;.001</td>
<td>7.83</td>
<td>1.66</td>
<td>3.84</td>
</tr>
</tbody>
</table>

**Group Descriptives**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CES-D</td>
<td>7</td>
<td>17.7</td>
<td>18.0</td>
<td>4.95</td>
<td>3.12</td>
</tr>
<tr>
<td>CBT</td>
<td>9</td>
<td>9.89</td>
<td>9.00</td>
<td>3.41</td>
<td>1.14</td>
</tr>
</tbody>
</table>

Figure 3. Jamovi output for independent samples t-test.
Questions to ask may include:

- Which group had lower CES-D scores? Where do you see this information? Is the difference consistent with expectations?
- How large was the difference in scores between the groups? Aside from calculating the mean difference in your head, do you spot the difference in the output?
- What does the statistic value reflect? Which values in the output reflect the numerator and the denominator of the *t*-test calculation?
- How does the statistic value listed compare to the critical value found in a *t*-distribution table? Would we reject the null hypothesis?
- How is the test statistic different from the Cohen’s *d* value?

Although it is important to show students how to perform independent samples *t*-tests step by step, students will appreciate hearing that psychologists in the field use software instead. Appendix B contains an example homework assignment to help students practice doing *t*-tests in Jamovi.

**ANOVA**

As with *t*-tests, one main objective for analyses of variance (ANOVA) is to explain to students how *F*-tests allow one to calculate the ratio of between conditions variability and the within conditions variability (or error). In Figure 4, green segments reflect between conditions variability, and red segments reflect within conditions variability. The numerator of the calculation, between-conditions variance (or MS\text{between}), is calculated using the same steps for
between-subjects (or independent samples) and within-subjects (or repeated measures) ANOVAs. It reflects the degree to which condition means differ from one another. Note in Figure 4, panels a and b compare three conditions with distributions that are roughly equivalent in width and differ by the degree of differences between the distributions’ means. Essentially, a larger numerator means that the F-test value is larger. Also note that, in panel c, the degree of variability within the conditions is less than in panel b, reducing the denominator of the F-test and increasing the resulting F-value. Once the basis for the F-test is discussed, one might launch into describing the source table and equations used to calculate sums of squares, degrees of freedom, and the mean square values. These equations are key to performing ANOVAs, but your students may benefit from first seeing an ANOVA performed in Jamovi. This will allow you to build their conceptual grasp of the test before they are bogged down in sums of squares calculations that often generate confusion and math insecurity. Appendix C contains example slides for teaching students how to conduct between-subjects ANOVAs in Jamovi.

Figure 5 depicts a dataset in which \( n = 7 \) adults participated in an experiment looking at the impact of working memory load on memory accuracy. All participants completed a working memory task with 5 trials for set sizes of 2, 3, 4, and 5 items. The 2-item trials were presented first as practice to introduce the task. The remaining trials were randomly mixed. Memory accuracy was recorded using the percentage of trials out of 5 for each set size where all of the items in the memory set were remembered with concurrent accuracy on the intervening distractor task. The cognitive psychologist managing the study wants to know if memory accuracy declines with increasing set size. The depicted study calls for a repeated measures ANOVA in which the set size is an independent variable with 3 levels.

![Figure 5. Working memory sample ANOVA data in Jamovi.](image)

Performing a repeated measures ANOVA in Jamovi requires that the data file is set up so that each participant's scores for each level of the independent variable fall in the same row. Each column captures performance for a specific set size within the task (wide file set up). You will note that this format is identical to what one would use within SPSS. In the Analyses menu, one selects ANOVA and then chooses Repeated Measures ANOVA from the available tests, shown in Figure 6.

References
Within the command window, depicted in Figure 7, one needs to specify a name for the repeated measures factor (set size) as well as a descriptor for each of its levels (set3, set4, set5). As descriptors are entered, Jamovi creates a repeated measures cell to which one would direct the appropriate data column. Within the command window, you can select a measure of effect size, choose post-hoc tests, and plot estimated marginal means.

The output of the analysis is generated as options are selected, so you can see how different choices impact the information presented to you. The ANOVA source table is shown in in Figure 8 and includes the most pertinent details for performing the hypothesis test. Under within subjects effects are the between-conditions (Set Size) and error (Residual) sources of variance. Under between subjects effects is the subjects source of variance that makes repeated measures ANOVAs so powerful. This layout is helpful because it allows you to emphasize that the variance within the model created by differences between the participants themselves has been removed from the unaccounted variance, reducing the error variance. As with the earlier t-test example, presenting this source table to the students facilitates discussion about the test itself and allows you to have time to address the original research question being proposed (i.e., did the participants’ performance did vary as a function of memory set size?) before delving into the mathematical formulas used to generate each number. Students will use knowledge gained from prior tests to ask questions and draw inferences about interpreting ANOVA output. For instance, you might ask the following questions based on the source table in Figure 8, post-hoc tests in Figure 9, and estimated marginal means in Figure 10:

- What are the three sources of variance that Jamovi includes in the source table for this one-way repeated measures ANOVA?
When we set up a source table for an ANOVA, we often include a row that tracks the total sums of squares and degrees of freedom. Jamovi does not do this. Can you still calculate them?

If you retrieved the critical F value for this ANOVA, what is it and how does it compare to the calculated F value?

Given that the F-value shows a significant difference between the set sizes, according to the descriptive statistics, which set size led to the best performance? The worst? Do the post-hoc tests confirm of these differences?

What proportion of the variance in memory performance is accounted for by the set size of the task? How might you use the sums of squares (a reflection of the sources of variance) to calculate the effect size for the manipulation?

From examining the estimated marginal means graph, can you identify possible group differences?

**Repeated Measures ANOVA**

<table>
<thead>
<tr>
<th>Within Subjects Effects</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>( \eta^2_p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set Size</td>
<td>3543</td>
<td>2</td>
<td>1771.4</td>
<td>46.5</td>
<td>&lt;.001</td>
<td>0.886</td>
</tr>
<tr>
<td>Residual</td>
<td>457</td>
<td>12</td>
<td>38.1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Type 3 Sums of Squares

<table>
<thead>
<tr>
<th>Between Subjects Effects</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>( \eta^2_p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual</td>
<td>5714</td>
<td>6</td>
<td>952</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Type 3 Sums of Squares

**Figure 8. Source table for repeated measure ANOVA in Jamovi.**

**Post Hoc Tests**

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Set Size</th>
<th>Set Size</th>
<th>Mean Difference</th>
<th>SE</th>
<th>df</th>
<th>t</th>
<th>Bonferroni</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 3</td>
<td>Set 4</td>
<td></td>
<td>11.4</td>
<td>3.30</td>
<td>12.0</td>
<td>3.46</td>
<td>0.014</td>
</tr>
<tr>
<td>Set 4</td>
<td>Set 5</td>
<td></td>
<td>20.0</td>
<td>3.30</td>
<td>12.0</td>
<td>6.06</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

**Figure 9. Bonferroni post-hoc tests for repeated measures ANOVA in Jamovi.**
Correlation and regression are vital to addressing research questions for students who are going on to sub-disciplines of psychology that involve observational designs, like clinical, counseling, community, or industrial-organizational psychology. Introducing these tests involves teaching your students about covariation. To conduct correlation and regression analyses in Jamovi, you set up your database so that each participant’s responses are recorded in a row, with separate columns reflecting the factors of interest.

Consider the data in Figure 11. Twenty participants completed three intelligence measures – facial emotion recognition, verbal ability, and inductive reasoning. Each test consisted of 100 items, so the data reflect each participant’s score on each test. A social psychologist is interested in testing whether emotion recognition ability is associated with verbal ability and/or inductive reasoning ability.
To examine the correlations between these factors, select Regression from the Analyses menu and then select Correlation Matrix. Within the command window, select those factors to add to the correlation matrix, and Jamovi generates the matrix, as in Figure 12. You can (a) specify which type of correlation coefficient you want to estimate, (b) ask Jamovi to report and flag significance values for each correlation, and (c) choose a scatterplot plot to display for the factors in your analysis.

After introducing the concept of a correlation coefficient, walking students through a Jamovi example will allow them to build their test outcome interpretation skills. For instance, for the Figure 12 output correlation matrix, you might ask the following questions:

- Which factor shares a strong association with emotion recognition - verbal ability or inductive reasoning?
- The verbal ability and inductive reasoning tests are both intelligence sub-tests. How would you expect them to be related to one another? Is this confirmed in the correlation matrix?

Within an introductory statistics course, correlation analysis serves as a launch pad for regression. The correlation coefficient reflects the standardized relationship between two factors, X and Y, whereas regression involves generating an unstandardized linear equation to estimate what one should observe for factor Y given one’s performance on factor X. Regression analyses yield two critical outcomes: (1) the linear equation including the y-intercept and unstandardized regression coefficient, and (2) a source table characterizing the variance in Y is accounted for by the predictor X. When transitioning from correlation to regression, we discuss the importance of shifting the way that we describe the relationship between the two factors. Correlation coefficients describe expected outcomes for Y given X in terms of standard deviations of each factor (e.g., when \( r = +1 \), scoring 1 SD above the mean on factor X should...
co-occur with scoring 1 SD above the mean on factor Y). We emphasize to students that individuals who are interested in their project’s data but are unfamiliar with statistics will not understand what this means. The average person does not make comparisons on the metric of standard deviation. Consequently, a regression analysis explains the relationship between two factors using the actual metrics for each factor. A specific score on X is linked with a predicted score on Y. With respect to the intelligence tests introduced a moment ago, regression analyses allow one to predict what one’s score on the emotion recognition test should be given a verbal ability score.

Figure 13. Regression command window in Jamovi (view cut in half).

Within Jamovi, a regression equation showing how verbal ability predicts emotion recognition ability is generated by choosing Linear Regression from the Regression option in the Analyses menu. Within the command window, as shown in Figure 13, Emotion is the dependent variable, and Verbal is the Covariate. By default, Verbal is entered into Block 1 as a predictor. If you have multiple predictors, then you can specify their entry in a hierarchical fashion by adding additional blocks. The output, shown in Figure 14, only includes the minimum information needed to create the linear equation, but you can request a source table for the regression model as well, shown in Figure 15.

Figure 14. Jamovi regression output.
As with ANOVAs, the mathematical formulas used to calculate the regression equation and source table sums of squares obfuscate students’ initial attempts to understand regression analyses. By discussing the purpose for this test, setting up an example in Jamovi, relating the correlation coefficient to the unstandardized regression coefficient, and introducing how one might use $R^2$ to understand the predictive value of $X$ to $Y$, you will improve your students’ conceptual understanding of regression. Given the Linear Regression output in Jamovi, you might ask your students the following questions:

- Given that the correlation between verbal ability and emotion recognition is $r = .609$, how might you interpret this in your own words using standard deviation as a metric?
- What does $R^2$ represent? How much variance in emotion recognition is not explained by verbal ability? Do you have any thoughts about what else might predict emotion recognition?
- What is the linear equation that represents the relationship between verbal ability ($X$) and emotion recognition ($Y$)?
- How do you know if verbal ability significantly predicts emotion recognition?

We reviewed introducing Jamovi, and using Jamovi for t-tests, ANOVA, correlation, and regression. These are just a few of the ways Jamovi can be used to teach introductory statistics concepts.

**Lessons learned and pitfalls to avoid when using Jamovi**

Although Jamovi is an excellent tool for teaching statistics, there are pitfalls to avoid and lessons we learned in adopting it for the first time.

1) In Jamovi, the analyses are saved along with the data. If you want students to do the analyses on their own, you’ll want to delete your own analyses before sharing the file with students.

2) Some email platforms flag Jamovi files as malware, so you may need to share exported data (.csv files and .pdf files). Alternatively, you can post on a learning management system (Canvas works well) or a cloud-based system (e.g., Google drive or OneDrive).

3) Ensure that students have access to a computer that can run software. Many netbooks do not allow for Jamovi installation. Ideally, the college or university IT department would install the software on computer lab/classroom machines. This has been easy at our institutions but may be a challenge depending on how IT services are managed. Students should also be encouraged to install Jamovi on their own computer (if they have one). This will be convenient for them and is a big advantage over expensive proprietary software such as SPSS.

4) If you upgrade Jamovi during the term, keep the version your students use. If you share your files with the students, they might not be able to open them due to compatibility problems between different versions of Jamovi. If IT upgrade requests...
take time at your institution or if students install Jamovi on their personal computers, be aware of the exact version you want them to install.

5) Error messages in Jamovi are rather cryptic. When problems occur, the professor may need to engage in a trial and error process to figure out what went wrong. Because the software is new, searching for the meaning of error messages on the Internet is not always successful. Having some familiarity with the R programming language (which Jamovi runs on) helps.

6) Incorporate Jamovi into lessons rather than simply running an example or two after you have covered your content.
   - Although adding Jamovi to a lesson does not take much time, you want to ensure that you leave enough time to talk with students about the tests.
   - Students may have questions about how to connect the structure of the data file to the research question being addressed as well as questions about how to interpret the output.
   - Students may also have basic questions about saving Jamovi files and exporting data/results. Allow plenty of time for trouble-shooting.

7) Make it count!
   - Assign Jamovi activities as formative assessments and include output on summative assessments for practice with interpretation.
   - Consider using an entrance/exit ticket system involving answering questions about using Jamovi to track knowledge growth and to identify students who are struggling with the software.

8) Follow up with students who consistently fail to work on Jamovi-related activities.

9) Jamovi is lagging on data visualization (e.g., figures, plots, etc.), so you may need to work out graphics in Excel using Jamovi-provided output.

10) Recoding variables and creating subscales require different steps in Jamovi compared to SPSS. Recent Jamovi versions have great features for data management (see Appendix D). Once students learn how to do this, it goes quickly.

Assessment results and student reactions

Although there are pitfalls to avoid, overall, we found Jamovi to be an effective tool for teaching introductory statistics. Near the end of the semester, students completed a survey about their reactions to and impressions of Jamovi. Twenty-one students enrolled at a medium-sized public university and twelve from a small private liberal arts college completed the surveys. One student who reported not completing any of the assigned Jamovi-related course work was excluded. Students responded to each item on a 5-point scale, ranging from 1 = strongly disagree to 5 = strongly agree. Survey results are displayed below.

Table 1. Student satisfaction ratings for Jamovi.

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jamovi was easy to install</td>
<td>4.43</td>
<td>1.07</td>
<td>5</td>
</tr>
<tr>
<td>Opening data files in Jamovi is straightforward</td>
<td>4.06</td>
<td>0.97</td>
<td>4</td>
</tr>
<tr>
<td>The menu structure of Jamovi is easy to navigate.</td>
<td>4.15</td>
<td>0.83</td>
<td>4</td>
</tr>
<tr>
<td>I know how to choose specific statistical tests in Jamovi.</td>
<td>4.18</td>
<td>0.68</td>
<td>4</td>
</tr>
<tr>
<td>I am confident that, when I choose a test in Jamovi, I am choosing the right test.</td>
<td>3.70</td>
<td>0.95</td>
<td>4</td>
</tr>
<tr>
<td>The stats test output provided by Jamovi is easy to understand.</td>
<td>4.06</td>
<td>1.00</td>
<td>4</td>
</tr>
<tr>
<td>I choose different features/options when running tests to better understand the outcome of my test.</td>
<td>3.67</td>
<td>0.99</td>
<td>4</td>
</tr>
</tbody>
</table>
I could explain how to use Jamovi to a classmate or to a student who takes this class next semester. 3.59 1.30 4
I will continue to use Jamovi to run statistics tests after this class is over. 3.76 1.37 4

Students reported high satisfaction with the basic features of Jamovi, such as installation, opening files, navigating menus, and interpreting statistical output. Lower ratings were observed for more advanced features such as tailoring options to one’s analysis goals and confidence in teaching Jamovi to others.

The survey also asked for open-ended comments about Jamovi. Sample comments included,

- “A lot easier/simpler to use in comparison to SPSS. Easier to navigate and choose different tests,”
- “Jamovi is much more user friendly than SPSS and I would recommend it before advising someone to use SPSS,”
- “I found Jamovi very user friendly and easy once I had used it a couple times. There were never any times that I felt like I didn’t know what I was doing with it,” and
- “very easy to comprehend and utilize, however the software does not explain mistakes well enough to help understand what was done wrong.”

The last comment is consistent with a pitfall we described in the previous section.

It is interesting that several students compared Jamovi favorably to SPSS. To follow up on these comments, we compared exam performance in a spring 2018 section of the course at the liberal arts college (taught using SPSS) to a spring 2019 section at the same institution (taught using Jamovi). On the exam about ANOVAs (including one-way, factorial, and repeated measures), the average exam score in the SPSS semester was 84.21%, compared to 92.46% when the course taught Jamovi. For an exam about linear, multiple, and hierarchical regression, the mean exam score in the SPSS semester was 91.86%, compared to 93.62% in the Jamovi semester. It should be noted that three students withdrew due to failing grades when the course was taught with SPSS. No students withdrew when the course was taught with Jamovi. Perhaps the more user-friendly nature of Jamovi allowed students to focus on conceptual learning rather than being distracted by interpreting complex SPSS output. This is backed up by a small survey of six graduating liberal arts college seniors, who learned both SPSS and Jamovi across their coursework. All six preferred Jamovi and thought it made statistics clearer and more straightforward than SPSS. Of course, this data is preliminary, and future research will be needed to see if these results are obtained with a larger sample size from a more diverse number of institutions.

Discussion

Jamovi is an easily accessible, free statistics program that works well in introductory statistics courses. Students can install it on their own computers, and our information technology departments have happily installed it in computer classrooms. Our tips for using Jamovi are based on our experience teaching small- to medium-sized, in-person courses that introduce statistics to psychology majors. The program also could be used in other settings. For example, Jamovi is helpful in getting students involved in research, either in working in faculty labs or on their own independent research projects. Because it is a free download, students can install it on their own computers and use it anytime and anywhere. This is especially helpful when research resources are limited. Statistics software is helpful for encouraging and supporting student research (Mendoza & Martone, 2019; Reavis & Thomas, 2019).
Because of the wealth of online resources about Jamovi, the recommendations in this chapter could easily be adapted to teaching online. Past research has demonstrated the importance of video tutorials in learning statistics. Videos showing tutorials on how to use SPSS led to equal performance in online and face-to-face classes (Breineiser, Rodefer, & Tost, 2018). This could easily be adapted to Jamovi, and in fact many videos for Jamovi already exist (See Appendix D). Screencast tutorials also enhance student learning of statistics and can be used in large and small classes (Lloyd & Robertson, 2012).

Jamovi is an effective and easy way to implement active learning into the introductory statistics course. It helps students develop job-relevant skills related to quantitative reasoning and data management. It is free and available for install in computer labs and on students’ personal computers. Because there are no site licenses, students will always be able to access it, even after graduation when they are in the workforce or furthering their education (although upgraded versions may be required). We recommend using the many free resources that are available for teaching Jamovi. The payoff is large as students improve their job-ready skills.
References


Appendix A

Slides to show students how to download & set up Jamovi on their computer

Downloading & Setting Up Jamovi

Download the latest version of Jamovi here (choose "solid")

Your computer may warn you that the developer is unknown. Go ahead and download it; it’s safe.
The User Guide has lots of helpful info as you learn Jamovi. Please read this whole page.

Click “modules”, then “Jamovi library” to add important modules. This will let you do additional analyses, such as:
- power analysis (to see how many participants you’ll need)
- Mediation & moderation
- Learning R, which is commonly used in grad school and jobs
You can open data files here, that have been created in other stats programs or that you or your instructor have previously saved.

You can set up a data file in Jamovi. Click “Data,” then setup to enter variable names, descriptions, etc.
Under "Data," clicking "Setup" will let you enter descriptions of your variables, the levels & scale of measurement, etc.
Appendix B
Example t-test homework assignment to practice Jamovi

Homework assignment: t-tests

Save your responses to all questions in one Word document and put in the Canvas dropbox no later than the beginning of class on the due date (10% late penalty per calendar day, starting at the beginning of class on the due date). Please cut and paste all relevant output from Jamovi and/or Excel into the Word document.

1. You are alarmed by research demonstrating rising rates of depression in adolescents. You wonder if getting a smartphone at a young age is associated with more depression symptomology. You ask 14 adolescents how old they were when they received a smartphone. Seven adolescents received a smartphone before the age of 13 and seven received a smartphone after the age of 13. The DV is a score on a depressive symptomology scale, in which higher scores indicate more symptoms of depression. Scores for the 14 adolescents are below.

<table>
<thead>
<tr>
<th>Adolescents who received a smartphone after the age of 13</th>
<th>Adolescents who received a smartphone before the age of 13</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>16</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>0</td>
<td>32</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>45</td>
</tr>
</tbody>
</table>

a) What are the null and alternative hypotheses?
b) What is the "IV" in this example? Is it a within-subjects or between-subjects variable? What is the DV?
c) What level/scale of measurement is the "IV"? What level of measurement is the DV?
d) Enter the data in Jamovi. Make sure you set up the data properly, including variable labels and the values of the "IV" levels. Refer to your notes from earlier in the semester on setting up Jamovi files. Perform the necessary statistical test. (Hint: Use your answers to questions b) and c) to guide your work.) Do you reject the null hypothesis or fail to reject the null? Why?
e) Write a complete APA-style results section.(include ALL needed info – descriptive stats, inferential stats, etc.)

Use Excel to make a graph illustrating the means of each group. Make sure you use proper graphing technique (labeling axes with proper capitalization, include error bars, etc.)
Appendix C

Slides to show between-subjects factorial ANOVA

Between-Subjects ANOVA in Jamovi

[Image showing Jamovi interface with a note: Jamovi comes with example data files that you can access here]
Move over DV & IV ("Fixed factors")

Choose effect size

Move over any IV with >2 levels that you want to do Tukey on

Output of your ANOVA shows up here. It will change in real time if you change your data. You can copy & paste from here into Word.
Scrolling down lets you ask for graph

It's very easy to ask for Standard Error bars on your graph

Make sure you ask for Descriptives (you'll need them for APA-style write-up)
## Appendix D

### Online resources

| Jamovi textbooks (open-source) | [https://www.learnstatswithjamovi.com/](https://www.learnstatswithjamovi.com/)  
<table>
<thead>
<tr>
<th></th>
<th><a href="https://www4.uwsp.edu/psych/cw/statistics/textbook.htm">https://www4.uwsp.edu/psych/cw/statistics/textbook.htm</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Video tutorials on Jamovi (free)</td>
<td><a href="https://datalab.cc/tools/jamovi">https://datalab.cc/tools/jamovi</a></td>
</tr>
</tbody>
</table>
| Example Jamovi blog posts that provide step-by-step instructions for common data organization tasks such as reverse-coding and creating subscales. | [https://blog.jamovi.org/2017/11/28/jamovi-formulas.html#mean-scores](https://blog.jamovi.org/2017/11/28/jamovi-formulas.html#mean-scores)  
|                              | [https://blog.jamovi.org/2018/10/23/transforming-variables.html?fbclid=IwAR3-c1jB-vjSxLY42HRz3cMIPk9etQ5FVP2In8urBkvm-hHYzbb4kf-Nwxc](https://blog.jamovi.org/2018/10/23/transforming-variables.html?fbclid=IwAR3-c1jB-vjSxLY42HRz3cMIPk9etQ5FVP2In8urBkvm-hHYzbb4kf-Nwxc) |
| Dataset generator from Richard Landers | [http://rlanders.net/dataset-generator/](http://rlanders.net/dataset-generator/) |
Yes, Beginning Statistics Students Can Use R!

Jennifer E. Samson, PhD
Queens University of Charlotte

Summary
Incorporation of statistical software is a vital part of a quality educational experience, but it is often left up to the department or even the individual instructor to determine which software to include and how. In this chapter, I encourage the reader to consider using R with beginning statistics students in social sciences and provide some guidance for introducing R. The biggest advantage of R over SPSS or other common software is student access and therefore opportunities for hands-on practice, but development of deeper conceptual understanding of statistics, ownership of learning, and preparation for graduate school or workforce experience are also considerations. R may present an additional challenge to students, especially due to its more intimidating programming interface (as compared to the point-and-click Guided User Interface in SPSS). However, carefully designed instructions and attention to scaffolding, especially in building foundational skills such as data import and cleaning, can mitigate these difficulties. Therefore, I include in the text and as appendices multiple example activities from throughout the semester, demonstrating scaffolding of R skills and movement from learning software to using the software as a tool to practice statistical analyses. Ultimately, I conclude that good pedagogy is good pedagogy, regardless of software choices, but the advantages of teaching primarily with R outweigh the disadvantages.

Introduction: Why R?
A quality statistics education includes the incorporation of appropriate technology (GAISE, 2016; Statistical Literacy Task Force, 2014). However, it is often left to the department or even to the individual instructor to determine what technology to introduce and how. In this chapter, I ask you to consider the idea of using R (https://www.r-project.org/). I share my experiences to try to convince you that yes, it is possible to use R with beginning statistics students and that there are specific advantages to doing so. Then, I provide some guidance and concrete examples of activities I’ve used to introduce R.

Before I begin to describe my experiences, I want to emphasize that there are a multitude of other resources out there for using and teaching with R, with more appearing every day. Don’t spend your valuable time creating materials until you’ve confirmed that what you need doesn’t already exist. It’s a lot easier to adapt something that currently exists than to write a new activity from scratch. My hope is that the current chapter provides more behind-the-scenes, “how do I use these materials” advice than many of the available texts so that you’ll be able to combine the benefit of my experiences with whatever materials work best for your specific students, class structure, and preferences.
Background

My experiences teaching introductory statistics for social sciences took place at a mid-sized state institution, where the majority of students are first-generation and a large percentage are non-traditional, juggling school with families, and full-time jobs. At that time, only 31% of students, university wide, lived on campus (ATU Institutional Research, 2015). Many of my students chose a major in social sciences because they experienced varying degrees of math anxiety and many commented that they specifically feared statistics. I taught 2-3 sections of introductory statistics (regular, 3-credit course) each semester, with each section including between 25 and 35 students. Between my third and fourth years, I transitioned from teaching with SPSS to using R and eventually R Studio (https://www.rstudio.com), first for one guinea pig section, and then for everyone.

For those who are unfamiliar with this software, R is an open-source programming environment, available by free download, for statistical computing (R Foundation). It is constantly being improved through the publication of packages for various procedures, contributed by users, which allows it to reflect the latest methodological advances in statistics. R Studio works with R to create an easier-to-navigate console which, among other advantages, allows for the quick import and formatting of data files and keeps data, syntax, and documentation (help window) organized and readily available. The syntax window adds color and autocomplete suggestions to the syntax file to aid coding. R Studio is also available as a free download (although commercial versions can be purchased).

Pros and Cons of Using R

I made the switch to R in my classes for multiple reasons. The most immediate, pressing reason was access; most of my students could not afford even a rental license for SPSS and therefore were limited to completing assignments during times they could access campus computers. Especially for commuters, this severely limited their opportunities for practice outside of class, a vital component for mastery (e.g. Brown, Roediger, & McDaniel, 2014; Doyle, 2011). By switching to R, I estimate that I went from 90% of my students not having their own copy of the software to 90% having one. No more, “I couldn't get to campus to access SPSS” or “the library was closed” excuses. No more conducting analyses but then not being able to access the results because students forgot to save their results in a non-SPSS file format. The use of a software program that most everyone could access at will increased opportunities for active learning, a vital component of a quality statistical education (GAISE, 2016).

If access were the only issue, however, there are other, low-cost or free alternatives out there I could have used. For instance, I experimented with Rcmdr, a package which introduces a point-and-click style GUI within R (also see chapter 17 in this edition about Jamovi). But what I discovered were advantages beyond just access. If you click a wrong button in SPSS or forget to change a default setting, you can get very wrong analyses and beginners especially may blindly copy and paste these results without ever knowing anything is wrong. R won't let users guess at analyses; if you type a command incorrectly, it usually returns just an error message. Relatedly, the process of dissecting and typing commands allowed many students to develop a deeper understanding of what they were actually doing. One student (who later served as a departmentally funded statistics grader/tutor) commented to me that,

“understanding what the code and what the program is doing specifically rather than ‘this gives you this’ ended up adding a great deal of depth in understanding…With SPSS I've noticed it sometimes took more effort for me to understand sometimes what is going on. Not because the mechanisms were complicated, but because it did not have the code included that directly stated what it was doing.”
In addition, I’ve noticed (just anecdotally) that more and more graduate programs and workplaces are moving to R, so exposure at an undergraduate level gives these students an advantage. Finally, teaching with R has forced me to learn it more thoroughly!

There were several obstacles to teaching with R that I considered before I made this transition. Because many of my colleagues prefer using SPSS, students would still need to be exposed to that software as well to prepare them for later courses. There was less support at my institution from IT for troubleshooting and maintaining R and R Studio than for SPSS, although they did install R and R Studio on all of our computer lab stations and in the library. Finally, in many ways I was learning R only a little ahead of my students. I spent a considerable amount of time that first year creating, testing, and troubleshooting activities, and I sent a lot of homework hints after the fact as students made me aware of some of R’s quirks or we found easier ways to code a particular procedure. (On the other hand, this learning together could be seen as an additional advantage; according to Bain, 2004, the best college teachers present themselves as fellow learners working with students to solve problems and demonstrating how they approach challenges).

Probably the biggest obstacle was the use of a programming interface rather than a point-and-click GUI. One student grader told me, the “format of the R software was rather intimidating in comparison [to SPSS].” Because the primary objectives of the class were to learn basic statistical principles (i.e. the software is just a tool, in line with guidelines from the APA statistical literacy task force, 2014), I specifically did not want the whole class to be about struggling with R. Therefore, teaching with R required much more scaffolding on the front end to get to a point where students could use it successfully to practice statistical concepts. And actually, I’d argue that the additional difficulty, when sufficient scaffold is provided, is itself another advantages of teaching with R; lightbulb moments are the best part of teaching and working with R afforded me more than my share. I saw over and over students doing the work to get through the challenge of using the software and, as a result, owning their learning in a new way (Doyle, 2011).

Methodology: How I Used R

Establishing the Basics First

The first thing I learned about R, both as a user and in teaching others to use it, was that the hardest part of the process is often getting data into the program and then into a usable form. R Studio, especially recent versions which allow direct imports from Excel, have mitigated this somewhat. However, data import and cleaning were still one of the biggest challenges of the semester, every semester. And, of course, you can’t do any statistics until the data is in the program and in the correct form. Therefore, one successful adaptation I made early in this process was adjusting my course calendar to include extra time at the beginning of the semester for students to learn the basics of R, in class. Specifically, the objectives for my first in-class R activity were for students to open R, load a data set, and create a new data set (see Appendix A). This took up the entire 50-minute period!

In subsequent labs, students learned to recode, compute (see Appendix B), and perform basic descriptive statistics (see Appendix C). About a third of the way into the semester, right before we dove into hypothesis testing, I assigned a R review lab (see Appendix D). The review lab was a check, for the students and for me, that they had mastered basic skills such as importing,

---

2 We were fortunate to meet in a computer lab for the entire semester, but this could also be accomplished with a few class periods spent on a field trip to the lab, or by having students who can do so bring their computers and then working in small groups.
recoding, computing, and running descriptive statistics. Completing the lab gave students the opportunity to check their own knowledge (meta-cognition) and to ask questions about lingering issues before these slowed them down on future assignments. Taking the time to review these basics at this point in the semester saved a lot of headaches later on!

As the semester went on, the initial time investment paid off, and lab instructions slowly removed support until students were able to use the software as a tool to practice different types of hypothesis tests. For example, compare the instructions for starting the lab in Appendix B with those in later assignments. At the beginning of the semester, Appendix B lays out the steps for importing data, opening a new script, and attaching the data file.

**Example from Appendix B**

“Open RStudio (start a new session if needed) and import the data set classdata (see instructions from Monday’s class)…

On the top left of your screen click file-new file-R script. This opens a script window as a top-left pane on your RStudio console. You’ll type commands into the script file, and then save this file so you have a record of what you’ve done. Notice that you can use the tabs at the top of the window to toggle between your script file and your data…

In the script window, type `attach(filename)` . Highlight this command and click Run. This tells RStudio to use this data set until you tell it otherwise!”

By the time students are using R to complete t-tests and chi-square analyses, the instructions assume they can open the data and get started with no instructions (specifically, the t-test review activity in Appendix E just asks students to, “use ‘classdata’” and the chi-square activity in Appendix F includes just a quick reminder to, “Import classdata, open a new script file, and attach your data.”

**Scaffolding R Use**

I employed basic scaffolding principles throughout the semester. So, although we worked through the first few labs entirely together, by halfway through the semester, we started labs together and then students completed them independently while I was available to answer questions (with a short debrief at the end of the class session). For instance, when we did the t-test review lab (Appendix E), I began the class by giving students a few minutes to decide which test they needed for each question (letter a), went over these answers, and then turned them loose to run the tests. About 10 minutes before the class period ended, we discussed the first couple of tests (mostly so I could check their hypothesis test decision-making process), and then students completed the assignment independently to turn in during the next class period. The chi-square lab (Appendix F), near the end of the semester, was completed entirely for homework; we spent that class period discussing usage and interpretation of chi-square tests in general.

A couple of additional tips:

- I made a point to increase student buy-in by using their own data, from a survey completed on first day of class, for the majority of examples in class and on homework. I found it important to make sure this survey contained different types of questions, such as “how many pairs of shoes do you own?” (count data), “how tall are you, in inches?” (continuous), and “are you a cat person or a dog person?” (categorical).
- I made sure that we celebrated progress. During an activity about halfway through the semester, I stopped the class as soon as everyone had their data loaded and reminded them that just that step took the whole class period not very long ago.
**Learning to Write Code**

The trickiest part of the whole endeavor was making lab handouts complete, but concise enough for students to follow. This becomes especially important when students are completing assignments independently as opposed to together during class time. One former student commented to me that, “I think the handouts that you used giving instructions with examples were helpful because they were brief and to the point.” I found that including certain details in my instructions removed some confusion. One detail that helped was introducing the notation early in the semester where I underlined file names in my code, which might be different in students’ own examples. I also used different fonts and italics to separate general instructions from code that the students were supposed to copy exactly. For instance, in Appendix B, Courier font tells students to type those commands exactly as they appear, while italics cue them to write an answer on their homework which will eventually be submitted for credit.

**Example from Appendix B**

```
“In the script window, type attach(filename)…

Hint: Replace filename with your date set name. It should be “classdata” unless you’ve renamed it.”

“Use the command below to save this new copy of your data set to a new file called, “classdata 1-25-17”. write.table(filename, file = "classdata 1-25-17.csv", sep = ",", quote = FALSE, row.names = FALSE)

9. Why should you save all your changes as new files instead of writing over the existing files? (1 point)

10. When you close R, save your script file. (You don’t need to save the workspace, etc. as long as you have your script file and your data file.)”
```

Another strategy I found to be especially helpful for students was annotating code. When I took the time to not only give them example code, but to describe in detail what each piece tells the computer, students were better able to translate commands to new situations. (I purposefully used very lay terms for describing that the code was doing, to encourage students to remember in terms they could understand rather than memorizing.) For instance, in the descriptive statistics lab (Appendix C), students saw this command to return the mean of shoes for each group designated by the variable, live (whether the participant lived on campus or off):

`aggregate(shoes, by = list(live), FUN = mean, na.rm = TRUE).` We took the time in class to break it up, so they saw that aggregate was the command, shoes was the variable they wanted the computer to tell them about, by = list(live) was the command to separate it by the variable live, FUN = mean tells the computer to return the mean of each group, and na.rm = TRUE tells it to ignore any missing data. This way, when they later needed the mean of weight by gender, they were much more likely to see what parts of the command remain the same and what they have to substitute (i.e. weight for shoes and gender for live). An extension of this would be to ask students to notate their own code as they move through examples.

**Introducing SPSS**

I did introduce students to SPSS, near the end of the semester, for a couple of reasons. First, as described above, students needed some familiarity with SPSS before they embarked on follow-up courses at my institution and I expected that many of them would also encounter SPSS in graduate programs or in the workforce. More importantly, however, I believe that exposure to a second statistical software program helped to cement students’ understanding of
statistical concepts (a priority according to GAISE, 2016) and increase their confidence by showing them they could transfer knowledge to a new situation. In other words, I wanted them to realize it didn’t matter which software they used; they could successfully do statistics. After introducing SPSS, I offered students to option to complete remaining assignments (e.g. Regression, see Appendix G) with either software.

Results: Students’ Success with R

Over the years, all of my statistics classes have shown significant growth from pretest to posttest. In that first semester when I introduced R in one section, I tested for differences by software (SPSS vs. R) and found no significant differences in final course grade, but a marginal \((p < .10)\) trend towards students in the R section scoring higher on the final exam than those in SPSS sections, after controlling for pretest scores. From these analyses, and my own observations, I concluded that there was no evidence the use of R impeded beginning statistics undergraduate students’ learning. In other words, students learned statistics at least as well, if not better, with R as with SPSS.

Observationally, the types of errors students made changed; instead of getting wrong analyses from SPSS, I’d get, “I can’t get this to work” with R. I believe having no results to turn in, rather than a wrong analysis from SPSS, made some students more likely to persist when they encountered challenges. Often the issue was a misplaced punctuation mark or misspelled variable name (the case sensitive thing throws a lot of students off at first). Just as often, students found the error in their own or their classmates’ code before I did.

Conclusion

Especially since I teach statistics and research methods, I would be remiss not to mention that my observations about the benefits of teaching introductory statistics with R are primarily anecdotal and at best correlational; controlled experimental trials are needed to determine whether software choice actually affects students’ opportunities for practice, ownership of their learning, deeper understanding of what the software is doing, and ultimately, their learning and retention in introductory statistics and preparation for further study. In general, however, I found that – big surprise – whether I taught with SPSS or R, good pedagogy was still good pedagogy. Providing lots of examples, varied and repeated practice, and scaffolding still served students well. Ultimately, I found the move to R to have more advantages than disadvantages, and I would make the same decision again.
References


https://www.atu.edu/ir/docs/cds/CDS_fall_2014.pdf


http://www.amstat.org/education/gaise.


R Studio. https://www.rstudio.com/

Appendix A

Introduction to Data In-Class Activity (Students’ first exposure to R)

These directions assume you’re working on a computer where R and RStudio are already installed. See directions on the back of this page and assigned readings to download and install R and RStudio on your computer. This activity also assumes you are working on a computer with Excel or an equivalent program installed.

1. Open Excel and create a new file call “Practice”. Data sets should always be set up so each row is a participant and each column is a variable. Label the first two columns “gender” and “height”. (Note that you cannot have spaces or other special characters in variables names in R.) Record the gender (as “m” or “f” – remember R is case sensitive) and height (in inches) of 10 people sitting around you. Save as an excel (xlsx) file.

2. Open RStudio. You should see a three-pane window. (In the future, you can click session-new session to open a new window that clears out everything you did previously.)

3. In RStudio, click import dataset-from excel (on the top right pane). Click Browse on the top right. Find Practice.xlsx on your computer and select it. You should now see a preview of the data set on the bottom. When the data looks correct in the window, click Import. You should then see your data in the top-left pane.

4. Download the file, olddata, from the class web site. Follow the directions in #3 above to import olddata into RStudio. Make sure it appears correctly on the top-left. Notice that you can use the tabs above the data sets to toggle between them.

R and RStudio are already installed on lab and library computers. Follow these directions to download and install them on your computer. (More information about R and RStudio is available in your assigned readings.)

1. Go to http://www.r-project.org/
2. Click download R on the top in blue.
3. Choose a CRAN near us (I’ve had good luck with either Kansas or Texas.) The CRAN is local server with a copy of the program and information stored on it; it does not matter which you choose. I do recommend you use one of the “https” links rather than an older “http” link to the same place.
4. Choose Mac or Windows, click the appropriate link at the top of the page (in blue) and follow the prompts to install the program.
5. You have now installed the base program R. You do not need a shortcut to open R; we will not open it directly because RStudio makes it easier to use (but it has to be installed before RStudio will run).
7. Under the first column (RStudio Desktop Open Source license), click the green “download” button.
8. Choose your operating system (probably either Windows or Mac, at the top of the list) and follow the prompts to download and install the program.
Appendix B

Introduction to R Lab (completed in class soon after Introduction to Data Activity)

As you complete each step, write your answers to the numbered, italicized questions, neatly, on another piece of paper. Turn in your answers for Lab 1 credit. Keep this page for your notes!

Note: In RStudio directions, underlined sections may vary depending on what you’ve named your files and/or variables. RStudio has an autocomplete that will help you.

Open RStudio (start a new session if needed) and import the data set classdata (see instructions from Monday’s class). This is your class’ survey results.

1. How many variables are in this data set? How many participants? (1 point)

On the top left of your screen click file-new file-R script. This opens a script window as a top-left pane on your RStudio console. You’ll type commands into the script file, and then save this file so you have a record of what you’ve done. Notice that you can use the tabs at the top of the window to toggle between your script file and your data.

2. List at least 2 reasons keeping a record of your analyses is useful. (1 point)

In the script window, type attach(filename) Highlight this command and click Run. This tells RStudio to use this data set until you tell it otherwise! (Hint: Replace filename with your date set name. It should be “classdata” unless you’ve renamed it.)

There is a variable in the data set (miles) which gives the number of miles each participant drives in an average week. Follow the directions below to compute a new variable called milesday which gives the number of miles each participant drives in an average day.

- Type the following into your script window: filename$milesday <- miles/7
  o When you copy a command, take a few minutes to think about what you’re asking R to do. For instance, this command tells R to make a new variable in the file filename, called milesday, that is created by taking the old variable miles and dividing it by 7. Thinking about the commands this way will help you apply the command structure to new problems.

- Highlight the line you just typed, and click Run. You should see the command appear in the console pane (bottom left) and the new variable appear in your data set (click the tabs to toggle between the data set and the script).

3. What is the value of milesday for participant #32? For participant #50? (1 point)

4. Apply what you learned and write the command to compute a new variable called “clothes” that is the sum of “shoes”, “jeans”, and “hats”. Hint: try running the command to check yourself. (1 point)
Type the commands below to recode the variable “live” into a new variable called “oncampus” where “on” is coded 1 and “off” is coded 0. (Note that what these commands are doing is first creating a new variable called “oncampus” and then filling in the values of that variable for each participant based on their response to live.)

```r
filename$oncampus <- live
filename$oncampus <- ifelse(live == "on", 1, ifelse(live == "off", 0, NA))
```

5. What is the value of oncampus for Participant #25? For participant #2? (1 point)

You’ve just created a “dummy variable”. You should always name a dummy variable so that 1 = “yes” and 0 = “no”, e.g. in this case 1 represents, “yes this person lives on campus” and 0 represents, “no this person does not live on campus”.

6. Why do you think this naming convention would be useful? (1 point)

7. Apply what you learned and write the command to recode the variable “gender” into a new variable called “male” where male = 1 and female = 0. (1 point)

8 (BONUS): Write the command to recode the variable “class” into a new variable called “classnum” where Freshman = 1, Sophomore = 2, Junior = 3, Senior = 4, and Graduate = 5. (up to 1 point possible)

Use the command below to save this new copy of your data set to a new file called, “classdata 1-25-17”. write.table (filename, file = "classdata 1-25-17.csv", sep = ",", quote = FALSE, row.names = FALSE)

9. Why should you save all your changes as new files instead of writing over the existing files? (1 point)

10. When you close R, save your script file. (You don’t need to save the workspace, etc. as long as you have your script file and your data file.)
Appendix C

Descriptive Statistics Lab

Open RStudio, and import “classdata 1-30-17”. Open a new script file and attach your data set (see directions from lab 1).

1. Type the command below to create a histogram of height.
   hist(height)
   Sketch the results on your paper. Is it normal, right-skewed, or left-skewed? (2 points)

2. Create a histogram of sibling. Sketch the results on your paper. Is it normal, right-skewed, or left-skewed? (2 points)

3. What type of variable (discrete or continuous) is shoes? Why? On what level of measurement (nominal, ordinal, interval, ratio) is shoes? How do you know? (2 points)

4. Type the commands below to get the min, max, quartiles, mean, and sd of shoes. (Hint: the “0th percentile is the min and the “100th percentile is the max.). Report the results in your homework. (3 points)
   quantile(shoes, na.rm = TRUE)
   mean(shoes, na.rm = TRUE)
   sd(shoes, na.rm = TRUE)

5. Repeat #3-4 with the variable, weight. (5 points)

6. Next you’ll look at differences in the mean of shoes between students living on campus and those living off campus. Before you begin – who do you think will own more shoes – on campus or off campus? Why (any reasonable explanation is acceptable)? (1 point)

7. Type the commands below to get separate means, standard deviations, and quartiles (including min and max) for on and off campus living. Report your results in your homework. Was your theory in #6 supported? (2 points)
   aggregate(shoes, by = list(live), FUN = mean, na.rm = TRUE)
   aggregate(shoes, by = list(live), FUN = sd, na.rm = TRUE)
   aggregate(shoes, by = list(live), FUN = quantile, na.rm = TRUE)
Appendix D

Software Review Lab (completed independently about 1/3 of the way through the semester, just before we really begin using R for inferential statistics)

The purpose of this lab is to review the software skills you’ve learned so far. Future assignments will assume you have mastered these skills, which include:

- Import data in RStudio
- Compute new variables from existing ones
- Recode variables
- Save a data set created in RStudio
- Use a script file to keep a record of your work
- Use RStudio to create histograms
- Use RStudio to compute descriptive statistics (mean, quartiles, including min/max/median, standard deviation) both for a whole sample and sub-groups
- Copy and paste RStudio output and graphs into a Word document

You are encouraged to use your notes from previous labs and to work with classmates to help you complete this assignment (but remember, each person should turn in a unique paper).

Please write/type your answers on a separate sheet of paper and copy/paste relevant software output as requested (should NOT be hand-written for this assignment).

1. Import “classdata 1-30-17” (available on class web site) into RStudio. Attach the data set. Open a new script file.
2. Compute a new variable, called kg, that gives each participant’s weight in kg. (Hint: see Lab 1 for compute instructions; to convert pounds to kilograms, you would multiply pounds by 0.45.) Copy and paste the command from your script file into your homework. (2 points)
3. Recode a new variable, called morning that is coded 1 if the participant is a morning person (see variable alert) and 0 if they are more alert in the afternoon, evening, or late night. (See Lab 1 for recode instructions.) Copy and paste the command from your script file into your homework. (2 points)
4. Attach your data set again. (You need to do this because you made changes to it)
5. Create a histogram of “kg” (see Lab 2 for instructions). Copy and paste the histogram into your homework (for this assignment, it should not be hand-drawn). Write a sentence describing the shape of “kg”. (2 points)
6. Have RStudio calculate the mean, standard deviation, and quartiles (including min and max) of “kg” (see Lab 3 for instructions). Copy and paste the results into your homework. (2 points)
7. Have RStudio calculate the mean and standard deviation of “kg” for students who are morning people and non-morning people separately (see Lab 2 for instructions). Copy and paste the results into your homework. Just from “eye-balling” these results (no
statistical test yet), do you think weight differs between morning-people and non-morning people? (2 points)

8. Use the write.table command in lab 1 to save a new copy of your data called “classdata 3-6-17”. Copy and paste this command from your script file into your homework. (2 points)

9. Save your script file. You do not need to turn it in, but keep it in case I ask to see it.
Appendix E

Mixed T-Tests Lab (notice that, at this point in the semester, I can provide much less instruction for basic R commands and focus on the hypothesis tests we are learning!)

For all tests, assume our class is a random sample of students and use “classdata” to answer the questions. Use $\alpha = .05$ for all tests.

1) I believe males will differ from females in the average number of shoes they own. Conduct a hypothesis test to test my theory.
   a. What kind of $t$-test do you need? How do you know? (1 pt)
   b. Write the null and alternative hypotheses. (1 pt)
   c. Use R to run the test. Write $t$, $df$, and $p$ from your output in the correct format and state your decision about the null hypothesis. (2 pt)
   d. State your conclusion about ATU students’ shoe ownership. (1 pt)

2) Test whether the average student endorsement of the statement “I am generally a happy person.” (variable: happy) differs from 3 = neutral.
   a. What type of $t$-test do you need? How do you know? (1 pt)
   b. Write the null and alternative hypotheses. (1 pt)
   c. Use R to run the appropriate test. Write $t$, $df$, and $p$ from your output in the correct format and state your decision about the null hypothesis. (2 pt)
   d. State your conclusion about ATU students’ happiness. (1 pt)

3) Do students report spending different amounts of time working (work) vs. studying (study)?
   a. What type of $t$-test do you need? How do you know? (1 pt)
   b. Write the null and alternative hypotheses. (1 pt)
   c. Use R to run the appropriate test. Write $t$, $df$, and $p$ from your output in the correct format and state your decision about the null hypothesis. (2 pt)
   d. State your conclusion about ATU students’ time usage. (1 pt)

4) I want to know if students living on campus vs. off campus differ in their endorsement of the statement, we need better food options on campus, on average.
   a. What type of $t$-test do you need? How do you know? (pt)
   b. Write the null and alternative hypotheses. (1 pt)
   c. Use R to run the appropriate test. Write $t$, $df$, and $p$ from your output in the correct format and state your decision about the null hypothesis. (2 pt)
   d. State your conclusion about ATU students’ residence status and opinion of food options. (1 pt)
Hints: T-Tests in R

**One-Sample T-Test** (from Lab 7, to test the null hypothesis that the population mean of weight is equal to 182.5):

t.test(weight, mu = 182.5)

**Dependent-Samples/Matched-Pair T-Test** (from Lab 7, to test whether there is a difference between moon day and nonmoon day aggression scores):

t.test(moon, nonmoon, paired = TRUE)

**Independent-Samples/Two-Sample T-Test** (to test whether males or females own more shoes, on average. Note that the first variable is what you’re measuring and the second tells R who’s in which group):

t.test(shoes ~ gender)
Appendix F

Chi-Square Lab (completed independently between class periods)

Import classdata 1-30-17, open a new script file, and attach your data.

Test whether there is a relationship between drink (prefer milk/juice, coffee/tea, other/none) and live (live on or off campus).

a. What type of chi-square test do I need? Why? (1 point)
b. Write the null and alternative hypotheses for my test. (1 point)
c. Run the chi-square test with the following commands:
   table1 <- table(live, drink)
   chisq.test(table1)
d. Write the chi-square, df, p-value in the correct form. (2 points)
e. Just for practice, use the chi-square table in your book to find a critical value if α = .05. (1 point)
f. If α = .05, what is my decision about the null hypothesis? What does this mean for the relationship between living arrangement and drink preference in the population? (2 points)

2. Now, test whether an equal number of Freshman, Sophomore, Junior, and Senior students take statistics.

a. What type of chi-square test do I need? Why? (1 point)
b. Write the null and alternative hypotheses for my test. (1 point)
c. Run the chi-square test with the following commands:
   table2 <- table(class)
   chisq.test(table2)
d. Write the chi-square, df, p-value in the correct form. (2 points)
e. Just for practice, use the chi-square table in your book to find a critical value if α = .05. (1 point)
f. If α = .05, what is my decision about the null hypothesis? What does this mean for the distribution of classes taking statistics? (2 points)
Appendix G

Regression (after introduction of SPSS near the end of the semester)

You may use either R or SPSS for this assignment (see software instructions on next page).

Open classdata 4-19-17 in SPSS (or classdata 1-30-17 in R). Use classdata to answer the following questions.

1. Examine the correlation between shoes and jeans (use \( \alpha = .05 \)).
   a. Write the null and alternative hypothesis for the correlation hypothesis test. (1 point)
   b. Use SPSS or R to get \( r \) and \( p \); write these in the correct form (\( r = xx, p = xx \)). (1 point)
   c. State your decision about the null hypothesis and your conclusion about shoes and jeans. (2 points)
   d. Create a scatterplot with shoes on the y axis and jeans on the x. Copy and paste your scatterplot (with correct title and labels) into your homework. (1 point)

2. Regress shoes on jeans (use \( \alpha = .05 \)).
   a. Write the null and alternative hypothesis for the hypothesis test on the slope of jeans. (1 point)
   b. Use SPSS or R to get \( b_1 \) and \( p \); write these in the correct form (\( b_1 = xx, p = xx \)). (1 point)
   c. State your decision about the null hypothesis and your conclusion about shoes and jeans. (2 points)
   d. Write the regression equation to predict shoes from jeans. If someone owns 4 pairs of jeans, how many shoes do we expect them to own? (2 points)

3. Now, regress shoes on jeans AND oncampus (\( \alpha = .05 \)).
   a. Write the null and alternative hypothesis for the hypothesis test on the slope of jeans, controlling for oncampus. (1 point)
   b. Use SPSS or R to get \( b_1 \) and \( p \); write these in the correct form (\( b_1 = xx, p = xx \)). (1 point)
   c. State your decision about the null hypothesis and your conclusion about shoes and jeans, controlling for oncampus. (2 points)
   d. Write the regression equation to predict shoes from jeans and oncampus. If a student living off campus owns 4 pairs of jeans, how many shoes do we expect them to own? (2 points)
   e. Ignoring the p-value for oncampus for a moment (in other words, whether the slope of oncampus is significant or not), who do we expect to own more shoes – students living on or off campus? How do you know (hint: from the slope of oncampus)? (1 point)
Correlation, Regression, and Scatterplot Software Instructions for Regression Lab

SPSS:

Click Analyze-Correlate-Bivariate

Move your variables (in example 1, shoes and jeans) to the right-side window.

Click Paste and then, in the syntax window, highlight and run.

Click Analyze-Regression-Linear

Place the y variable (what you’re trying to predict, in example 2 shoes) in the top box and the x variable (the predictor, in example 2 jeans) in the second box. For multiple x variables (as in example 3), just add the additional variables to the second box.

Click Paste and then, in the syntax window, highlight and run.

Click Graphs-Legacy Dialogs-Scatter/Dot

Click “simple scatterplot” and Define

Place the y variable in the top box and the x variable in the second box

Click Paste, then, in the syntax window, highlight and run.

In the output window, double-click on the graph to open an editing window and add a title (click the button at the top that looks like a title).

RStudio:

Open a new script file and attach your data set.

Use this command to run a correlation between the variables x and y:\texttt{cor.test(x, y)}.

So, for example 1, you would type:\texttt{cor.test(shoes, jeans)}.

Notice that R gives you more information than you need – you should report the r (correlation) from the bottom of the output (under sample estimates: cor) and the p-value (but not the whole t-test results) from the top.

Use this command to regress y on one x:\texttt{summary(lm(y ~ x))}
So, for example 2, you would type: `summary(lm(shoes ~ jeans))`

For additional x's, just add the additional x's, separated by a +: `summary(lm( y ~ x1 + x2))`

Type the following command to make a scatterplot of shoes x jeans:

`plot(jeans, shoes, main = "Jeans x Shoes")`
Compendium

Alexander J Bies, PhD
Gonzaga University

Summary
The following list provides a set of resources that map onto the chapters from the book. The resources are organized by topic. Entries are duplicated in cases where they are relevant for more than one chapter to minimize scrolling away from a particular topic. Following the list, a summary table maps the goals for teaching statistics to undergraduates which each of the resources.

Overview of the Compendium
This compilation of resources is intended to provide a survey of the best and most recent resources available for teaching statistics to students in Psychology (or social behavioral sciences).

Resources Organized by Book Chapter

Ch 2: Teaching stats online

- Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) provides a variety of online resources - online analysis tools, datasets, fun materials for instructional use, lab activities, lectures, homework, projects, etc.
  https://www.causeweb.org/cause/resources
- The Rice Virtual Lab in Statistics provides a number of web-based simulations (requires JAVA) on a variety of statistical topics (http://onlinestatbook.com/stat_sim/) in association with their book.
  http://onlinestatbook.com/
- Lowry wrote an online statistics textbook (http://vassarstats.net/textbook/) and online statistical calculators for a variety of statistical tests.
  http://vassarstats.net/
- Lane wrote an ad-supported online statistics textbook (http://davidmlane.com/hyperstat/index.html) with exercises and problems.
  http://davidmlane.com/hyperstat/questions/index.html
- Desrochers and Margolin (2010) developed an interactive workbook to facilitate learning about factorial research designs which includes a completion certificate.
  http://www.acs.brockport.edu/~mdesroch/Factorial3/
- See also the OERs where datasets are linked by JOPD and APA, along with two statistics books, and the article by Winquist and Carlson which references the Carlson and Winquist book from sage.
- Cohen’s D and effect size visualization https://rpsychologist.com/d3/cohend/
- Confidence Interval interpretation visualization https://rpsychologist.com/d3/CI/
- Correlation strength guessing game http://guessthecorrelation.com/
- Correlation visualization https://rpsychologist.com/d3/correlation/
- P distribution visualization https://rpsychologist.com/d3/pdist/
Ch 3: Flipped classroom

- The University of North Carolina Center for Faculty Excellence provides an introductory resource for flipping a class. [https://cfe.unc.edu/teaching-and-learning/introduction-flipping-class/](https://cfe.unc.edu/teaching-and-learning/introduction-flipping-class/)
- Winquist and Carlson (2014) summarize the results of students' statistics knowledge 1 year after taking a traditional lecture or flipped classroom psychology statistics course. Their description of class activities is brief but clear. They asked students to use Carlson and Winquist's *An introduction to Statistics: An Active Learning Approach* ([https://edge.sagepub.com/carlson2e](https://edge.sagepub.com/carlson2e)) for reading and activities prior to class. [http://jse.amstat.org/v22n3/winquist.pdf](http://jse.amstat.org/v22n3/winquist.pdf)

Ch 5: Passion driven statistics

- Ciarocco, Strohmetz, and Lewandowski (2010) provide bibliographies for studies with varied research designs in developmental, health, neuropsychology, sensation and perception, and social psychology. Each includes discussion prompts and in-class activities. [http://teachpsych.org/resources/Documents/otrp/resources/ciarocco10.pdf](http://teachpsych.org/resources/Documents/otrp/resources/ciarocco10.pdf)
- Spencer (2015) provided a number of prompts for research design using a consulting model, which could easily be adapted for analysis plans and statistical tests. [https://teachpsych.org/Resources/Documents/otrp/resources/spencer15.docx](https://teachpsych.org/Resources/Documents/otrp/resources/spencer15.docx)
- The Data and Story Library provides interesting data with descriptions on a variety of topics. [https://dasl.datadescription.com/](https://dasl.datadescription.com/)

Ch 6 (and part of Ch 9): Service-learning


Ch 8: Integrating writing and statistics

- Phrasebank.manchestster.ac.uk has a page of example phrasing for writing about statistical results. [www.phrasebank.manchester.ac.uk/reporting-results/](http://www.phrasebank.manchester.ac.uk/reporting-results/)
- Bradfordx produced a video on reporting NHST in APA style. [https://www.youtube.com/watch?v=o_LMGReW_NE](https://www.youtube.com/watch?v=o_LMGReW_NE)

Part 3: Active learning ideas in the classroom (canned activities)
• Statisticsteacher.org is an online journal published by the American Statistical Association. Numerous articles provide ideas for activities, particularly the grades 9-12+ section. [https://www.statisticsteacher.org](https://www.statisticsteacher.org)
• Causeweb.org provides some online computer activities. [https://www.causeweb.org/cause/resources](https://www.causeweb.org/cause/resources)
• Holmes and James (2008) developed a set of 9 activities that facilitate student learning in descriptive statistics, z-scores, probability, experimental design, measurement and sampling. [http://teachpsych.org/resources/Documents/otrp/resources/holmes08.pdf](http://teachpsych.org/resources/Documents/otrp/resources/holmes08.pdf)

**Ch 9: Engaging students in statistics**

• The Teaching of Psychology Idea Exchange (ToPIX) has a list of instances of statistics in the news relevant for teaching of statistics for psychology and of pages that reference such information. [http://topix.teachpsych.org/w/page/49255463/Statistics%20in%20the%20News](http://topix.teachpsych.org/w/page/49255463/Statistics%20in%20the%20News)
• Notawfulandboring.blogspot.com is a blog offering interesting examples for teaching statistics. [http://notawfulandboring.blogspot.com/](http://notawfulandboring.blogspot.com/)
• Tyler Vigen’s blog presents strong correlations in theoretically unrelated data. [https://www.tylervigen.com/spurious-correlations](https://www.tylervigen.com/spurious-correlations)
• Statsland.wordpress.com provides amusing discussions of statistical concepts that will engage students. [https://statsland.wordpress.com/](https://statsland.wordpress.com/)
• Hackathorn, Ashdown, and Rife (2016) developed and curated a set of comics and other humorous material related to statistics. [https://teachpsych.org/resources/Documents/otrp/resources/HACKATHORN.zip](https://teachpsych.org/resources/Documents/otrp/resources/HACKATHORN.zip)
• Lesser and Pearl (2008) provide suggestions for how to implement humor in a way that is reflective of teaching goals. [https://www.tandfonline.com/doi/full/10.1080/10691898.2008.11889572](https://www.tandfonline.com/doi/full/10.1080/10691898.2008.11889572)
• Segrist and Hupp (2017) developed an annotated list of resources about using humor in the classroom and course materials. [https://teachpsych.org/Resources/Documents/otrp/resources/segrist15.pdf](https://teachpsych.org/Resources/Documents/otrp/resources/segrist15.pdf)

**Ch 11: Manipulatives**

• Sledjeski (2016) provides instructions for activities using poker chips that represent people (by writing demographic information onto each chip) and using them to draw samples to address questions in classical topics in statistics classes (sampling, graphing, descriptive statistics, probability and sampling, and a variety of hypothesis tests). Could be adapted to have students analyze their data, as well. Instructor manual, male and female face labels, and data files here: [https://teachpsych.org/page-1603066](https://teachpsych.org/page-1603066) with the resource linked below.
  - Male face labels [https://teachpsych.org/resources/Documents/otrp/resources/male%20faces%20labels.pdf](https://teachpsych.org/resources/Documents/otrp/resources/male%20faces%20labels.pdf)
  - Female face labels [https://teachpsych.org/resources/Documents/otrp/resources/female%20faces%20labels.pdf](https://teachpsych.org/resources/Documents/otrp/resources/female%20faces%20labels.pdf)
  - Data file [https://teachpsych.org/resources/Documents/otrp/resources/poker%20chip%20data.xlsx](https://teachpsych.org/resources/Documents/otrp/resources/poker%20chip%20data.xlsx)
Holmes and Jemes (2008) developed a set of 9 activities that facilitate student learning in descriptive statistics, z-scores, probability, experimental design, measurement and sampling. [http://teachpsych.org/resources/Documents/otrp/resources/holmes08.pdf](http://teachpsych.org/resources/Documents/otrp/resources/holmes08.pdf)

**Ch 13: The new statistics**

- Open science framework (OSF) page for teaching the New Statistics. [https://osf.io/muy6u/](https://osf.io/muy6u/)
- Rpsychologist provides a visualization tool for effect size in terms of Cohen’s d. [https://rpsychologist.com/d3/cohend/](https://rpsychologist.com/d3/cohend/)
- Rpsychologist provides a visualization tool for the interpretation of confidence intervals. [https://rpsychologist.com/d3/CI/](https://rpsychologist.com/d3/CI/)
- Engaging posts about confidence intervals. [http://notawfulandboring.blogspot.com/search/label/confidence%20intervals](http://notawfulandboring.blogspot.com/search/label/confidence%20intervals)
- Engaging posts about effect sizes. [http://notawfulandboring.blogspot.com/search/label/effect%20sizes](http://notawfulandboring.blogspot.com/search/label/effect%20sizes)

**Ch 14: Meta-analysis**


**Ch 16: Confidence intervals**

- Grodofsky (2007) provides links to video clips of a game from *The Price is Right* that provides a non-computational introduction to confidence intervals and poses discussion questions to probe and increase students’ understanding. [http://www.teachpsychscience.org/files/pdf/13201425907PM_1.PDF](http://www.teachpsychscience.org/files/pdf/13201425907PM_1.PDF)
- Confidence interval simulations are provided at multiple websites. Good, interactive demonstrations can be found here:
  - [http://www.rossmanchance.com/applets/ConfSim.html](http://www.rossmanchance.com/applets/ConfSim.html)
  - [https://shiny.rit.albany.edu/stat/confidence/](https://shiny.rit.albany.edu/stat/confidence/)

**Part 5: Data analysis tools**

In addition to Excel and SPSS, there are several free complete statistical analysis software packages. One that stands out is JASP:

- JASP is free, open-source statistical analysis software that works on Windows, Mac, and Linux. Teaching resources associated with JASP, including a data library, are linked at the homepage. [https://jasp-stats.org/](https://jasp-stats.org/)
  - Open source teaching materials for JASP have been created by Buchanan, Hopke, and Donaldson (2018). [https://osf.io/t56kg/](https://osf.io/t56kg/)
- PANDAS is a Python-based statistical analysis suite. While free, it is not suitable for students who do not have a programming background. [https://pandas.pydata.org/](https://pandas.pydata.org/)

There are also many applets and online resource for data analysis or determining particular statistical values.
• A variety of descriptive statistics and inferential statistical test values can be performed at Social Science Statistics' page, which even has a “which test should I use?” wizard (https://www.socscistatistics.com/tests/what_stats_test_wizard.aspx). This resource even includes one-way ANOVA for independent or repeated measures and multiple regression, p, effect size, and confidence interval calculators. https://www.socscistatistics.com/descriptive/ and https://www.socscistatistics.com/tests/

• Mathportal.org provides a more limited list of descriptive and inferential statistics calculators, but provides symbolic notation of the calculations involved in computing the test statistics unless the explanation is turned off. https://www.mathportal.org/calculators/statistics-calculator/

• In contrast to the preceding calculators, danielsoper.com provides a variety of statistical calculators that rely on descriptive statistics or inferential statistics as input. https://www.danielsoper.com/statcalc/default.aspx

Part 5: OER resources for data analysis, etc.

• The Journal of Open Psychology Data (JOPD) provides many datasets for reuse, contextualized by the authors with an article briefly covering the background, thorough methods, and a description of the dataset. https://openpsychologydata.metajnl.com/

• The American Psychological Association provides links to datasets and repositories for a variety of resources from census data to data from large scale studies on developmental, health, aging, and other topics. https://www.apa.org/research/responsible/data-links

• Illowsky and Dean have written an open textbook Introductory Statistics that covers typical statistics topics with practice problems and homework sets. https://openstax.org/details/introductory-statistics?Book%20details

• Shafer and Zhang have written another open textbook also called Introductory Statistics that covers similar topics with good practice problems and links to large online data sets. https://open.umn.edu/opentextbooks/textbooks/135

• Tagler (2010) developed laboratory exercises (with answers) that develops students' skills in Excel. https://teachpsych.org/page-1603765

• Gibson, Klatzkin, and Littlefield (2012) developed conceptual and laboratory exercises using Excel or SPSS. These exercises are consistent with the goals of passion-driven statistics. https://teachpsych.org/page-1603793


Tables of Resources

The following table is organized in the general order of the contents of the book as presented above. The second column has a list of keywords related to the type of content, its format, aim and topic. The third column presents the chapter numbers within this book with which each resource listed below is associated.

The fourth and fifth columns present numbers which correspond to the APA (Statistical Literacy Task Force, 2014) and GAISE (2016) goals which each link may achieve. The APA goals are 1) Interpret basic statistical results, 2) apply appropriate statistical strategies to test hypotheses, 3) apply appropriate statistical and research strategies to collect, analyze, and interpret data, and report research findings, 4) distinguish between statistical significance and practical significance, and 5) evaluate the public presentation of statistics. The GAISE goals are 1) teach statistical thinking, 2) focus on conceptual understanding, 3) integrate real data with a context
and a purpose, 4) foster active learning, 5) use technology to explore concepts and analyze data, 6) use assessments to improve and evaluate student learning.

<table>
<thead>
<tr>
<th>Content and hyperlink</th>
<th>Type/Format and Keywords</th>
<th>book section</th>
<th>APA goals</th>
<th>GAISE goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service learning project in stats - timeline and outline</td>
<td>Peer-reviewed paper, service-learning</td>
<td>2</td>
<td>1,2,3,4</td>
<td>1,2,3,4,6</td>
</tr>
<tr>
<td>Approaches to service learning in stats courses</td>
<td>Peer-reviewed paper, service-learning</td>
<td>2</td>
<td>1,2,3,4</td>
<td>1,2,3,4,6</td>
</tr>
<tr>
<td>Stats in the news</td>
<td>Blog, fun</td>
<td>3</td>
<td>1,2,5</td>
<td>1,2,3,4,5</td>
</tr>
<tr>
<td>Interesting stats examples</td>
<td>Blog, fun</td>
<td>3</td>
<td>1,2,5</td>
<td>1,2,3,4,5</td>
</tr>
<tr>
<td>Interesting (spurious) correlations</td>
<td>Blog, fun</td>
<td>3</td>
<td>1,2,5</td>
<td>1,2,3,4,5</td>
</tr>
<tr>
<td>Comics and humorous materials (Hackathorn et al. (2016))</td>
<td>Peer-reviewed paper, supp. materials, fun</td>
<td>3</td>
<td>1,5</td>
<td>1,2,4,6</td>
</tr>
<tr>
<td>Suggestions for humor in classroom and materials</td>
<td>Peer-reviewed paper, fun</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annotated list of resources about using humor</td>
<td>Peer-reviewed paper, fun</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poker chip people activities</td>
<td>Peer-reviewed paper, supp. materials, activity</td>
<td>3</td>
<td>1,2,3</td>
<td>1,2,4,5,6</td>
</tr>
<tr>
<td>9 Hands-on (in-class) activities and demonstrations</td>
<td>Peer-reviewed paper, supp. materials, activity</td>
<td>3</td>
<td>1,2,3</td>
<td>1,2,4,5,6</td>
</tr>
<tr>
<td>Math anxiety scale</td>
<td>Webpage, activity</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Math anxiety reduction techniques</td>
<td>Webpage, resource</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Phrase bank for statistical reporting</td>
<td>Webpage, resource, writing and reporting</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>APA style reporting video</td>
<td>You-tube video, writing and reporting</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Reducing math anxiety through writing exercises</td>
<td>Peer-reviewed paper, activities, writing and reporting</td>
<td>2 &amp; 3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Increasing statistical thinking skills through writing exercises</td>
<td>Peer-reviewed paper, activities, writing and reporting</td>
<td>2 &amp; 3</td>
<td>2,3</td>
<td>1,2,4,5,6</td>
</tr>
<tr>
<td>Open Psychology Data datasets</td>
<td>Real datasets</td>
<td>5</td>
<td>2,3</td>
<td>3,4,5</td>
</tr>
<tr>
<td>APA data links</td>
<td>Real datasets</td>
<td>5</td>
<td>2,3</td>
<td>3,4,5</td>
</tr>
<tr>
<td>Open textbook – Illowsky &amp; Dean</td>
<td>Open textbook</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open textbook – Shafer &amp; Zhang</td>
<td>Open textbook</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lab exercises in Excel</td>
<td>Practice data, Excel</td>
<td>5</td>
<td>1,2,3</td>
<td>1,4,5</td>
</tr>
<tr>
<td>Conceptual and lab exercises in Excel and SPSS</td>
<td>Practice problems, practice data, Excel, SPSS, writing</td>
<td>5</td>
<td>1,2,3</td>
<td>1,4,5,6</td>
</tr>
<tr>
<td>Graphing literacy and ability exercises</td>
<td>Peer-reviewed paper, Practice problems, Excel</td>
<td>5</td>
<td>1</td>
<td>1,4,5,6</td>
</tr>
<tr>
<td>Flipped classroom introduction</td>
<td>Active learning</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource</td>
<td>Description</td>
<td>Version(s)</td>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>Hussy, Richmond and Fleck's flipped psychology class primer</td>
<td>Peer-reviewed paper, Active learning</td>
<td>2</td>
<td>1,2,4</td>
<td></td>
</tr>
<tr>
<td>Winquist and Carlson’s flipped classroom study</td>
<td>Peer-reviewed paper, Active learning</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Carlson and Winquist’s textbook used for flipped class</td>
<td>Active learning</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>CAUSE (causeweb.org) - resources for teaching</td>
<td>Practice data, practice problems, activities, lecture slides, fun</td>
<td>2 &amp; 3</td>
<td>1,2,3, 4,6</td>
<td></td>
</tr>
<tr>
<td>Rice Virtual Lab simulations</td>
<td>Online activities, online simulator</td>
<td>2 &amp; 5</td>
<td>1,2,3, 4,5</td>
<td></td>
</tr>
<tr>
<td>Rice Virtual Lab book</td>
<td>Open textbook</td>
<td>2 &amp; 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VassarStats online statistical calculators</td>
<td>Online calculator</td>
<td>2 &amp; 5</td>
<td>3, 5</td>
<td></td>
</tr>
<tr>
<td>VassarStats online textbook</td>
<td>Open textbook</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane’s exercises and problems</td>
<td>Practice problems, practice questions</td>
<td>2 &amp; 5</td>
<td>3, 5, 6</td>
<td></td>
</tr>
<tr>
<td>Lane’s ad-supported online textbook</td>
<td>Open textbook</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interactive workbook about factorial designs</td>
<td>Practice problems, practice questions</td>
<td>3</td>
<td>2, 3, 1, 2</td>
<td></td>
</tr>
<tr>
<td>Statisticteacher.org activities</td>
<td>Active learning, canned activities</td>
<td>3</td>
<td>2, 3, 1, 2, 4</td>
<td></td>
</tr>
<tr>
<td>Causeweb.org activities</td>
<td>Active learning, canned activities</td>
<td>3</td>
<td>2, 3, 1, 2, 4</td>
<td></td>
</tr>
<tr>
<td>Holmes and Jemes activities</td>
<td>Peer-reviewed paper, active learning, canned activities</td>
<td>3</td>
<td>2, 3, 1, 2, 4</td>
<td></td>
</tr>
<tr>
<td>Intro to teaching meta-analysis</td>
<td>Peer-reviewed paper, meta-analysis, canned activities</td>
<td>4</td>
<td>1, 2, 3, 4, 1, 2, 3, 4, 5</td>
<td></td>
</tr>
<tr>
<td>JASP – free data analysis software</td>
<td>JASP, free software</td>
<td>5</td>
<td>3, 5</td>
<td></td>
</tr>
<tr>
<td>JASP – teaching resources</td>
<td>JASP, teaching resources</td>
<td>5</td>
<td>3, 5</td>
<td></td>
</tr>
<tr>
<td>JASP – user guide</td>
<td>JASP, user guide, lab resource</td>
<td>5</td>
<td>3, 5</td>
<td></td>
</tr>
<tr>
<td>JASP – OSF folder of teaching materials</td>
<td>JASP, Open science framework</td>
<td>5</td>
<td>1, 2, 3, 3, 4, 5</td>
<td></td>
</tr>
<tr>
<td>JASP – You-tube tutorials</td>
<td>JASP, You-tube</td>
<td>5</td>
<td>3, 4, 5</td>
<td></td>
</tr>
<tr>
<td>PANDAS – free Python-based data analysis software</td>
<td>Python, free software</td>
<td>5</td>
<td>3, 5</td>
<td></td>
</tr>
<tr>
<td>SocSciStatistics online descriptive statistic calculators</td>
<td>Online stats calculator, central tendency, charts (bar, frequency polygon, histogram), percentile, variability</td>
<td>3</td>
<td>3, 5</td>
<td></td>
</tr>
<tr>
<td>SocSciStatistics online inferential statistical test calculators</td>
<td>Online stats calculator, ANOVA, chi-square, confidence intervals, correlation, effect size, p-value, regression, standard error, t-test, z-test</td>
<td>3</td>
<td>3, 5</td>
<td></td>
</tr>
<tr>
<td>MathPortal descriptive and inferential calculators</td>
<td>Online stats calculator, central tendency, correlation, probability, regression, t-test, variability, z-test</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Statistical calculators relying on descriptive or inferential input</td>
<td>Online stats calculator, ANOVA, chi-square, confidence intervals, correlation, effect size, p-value, regression, statistical power, t-test, z-test</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>The Price is Right and Confidence Intervals</td>
<td>Peer-reviewed paper, activities, confidence intervals, discussion questions, fun, practice questions, You-tube</td>
<td>2 &amp; 3</td>
<td>3,4</td>
<td>1,2,4, 5,6</td>
</tr>
<tr>
<td>Rossman-Chance confidence interval simulator</td>
<td>Online simulator, confidence intervals, fun</td>
<td>4</td>
<td></td>
<td>2,4,5</td>
</tr>
<tr>
<td>Dudek confidence interval simulator</td>
<td>Online simulator, confidence intervals</td>
<td>4</td>
<td></td>
<td>2,4,5</td>
</tr>
<tr>
<td>Bibliographies for a variety of published studies and discussion prompts</td>
<td>Peer-reviewed paper, activities, discussion questions, passion-driven statistics</td>
<td>2 &amp; 3</td>
<td>3</td>
<td>1,2,4, 6</td>
</tr>
<tr>
<td>Prompts for consulting model discussion and activities</td>
<td>Peer-reviewed paper, activities, discussion questions, passion-driven statistics</td>
<td>2 &amp; 3</td>
<td>2,3</td>
<td>1,2,4, 6</td>
</tr>
<tr>
<td>Data and brief descriptions on a variety of topics</td>
<td>Passion-driven statistics, real datasets</td>
<td>2, 3 &amp; 5</td>
<td>1,2,3</td>
<td>1,3,4, 5</td>
</tr>
<tr>
<td>OSF folder for teaching the New Statistics</td>
<td>Open science framework, new statistics</td>
<td>4</td>
<td>1,2,3,4,5</td>
<td>1,3,4, 5</td>
</tr>
<tr>
<td>Cohen’s d effect size visualization tool</td>
<td>Effect size, new statistics, online visualization</td>
<td>4</td>
<td>3</td>
<td>1,2,4, 5</td>
</tr>
<tr>
<td>Visualization for the interpretation of confidence intervals</td>
<td>Confidence interval, new statistics, online visualization</td>
<td>4</td>
<td>3</td>
<td>1,2,4, 5</td>
</tr>
<tr>
<td>Correlation strength guessing game (old video game style)</td>
<td>Correlation strength, fun, online activities</td>
<td>3</td>
<td>1,3</td>
<td>1,2,4, 5</td>
</tr>
<tr>
<td>Rossman-Chance guess the correlation strength applet</td>
<td>Correlation strength, fun, online activities, online simulator, online visualization</td>
<td>3</td>
<td>1,3</td>
<td>1,2,4, 5</td>
</tr>
<tr>
<td>Correlation visualization</td>
<td>Correlation strength, online activities, online visualization</td>
<td>3</td>
<td>1,3</td>
<td>1,2,4, 5</td>
</tr>
<tr>
<td>P-distribution visualization</td>
<td>P-distribution, online activities, online visualization</td>
<td>3</td>
<td>1,3</td>
<td>1,2,4, 5</td>
</tr>
<tr>
<td>Standard normal distribution table visualization</td>
<td>Standard normal distribution, online activities, online visualization</td>
<td>3</td>
<td>1,3</td>
<td>1,2,4, 5</td>
</tr>
<tr>
<td>Statistical significance and power visualization</td>
<td>Standard significance, power, online activities, online visualization</td>
<td>3</td>
<td>1,3</td>
<td>1,2,4, 5</td>
</tr>
</tbody>
</table>
Other resources

There are far too many resources available to have listed them all here, and we recognize that there will continue to be innovations and new developments. A few search strategies for your own searches are to add visualization or video or data for particular topics (t-test, confidence intervals), search a teaching of statistics database (see below) and look for links in the above links.

Student attitudes (emotion, anxiety, math phobia)

- Freedman hosts a math anxiety survey, which students could fill out and score http://www.mathpower.com/anxtest.htm along with tips for reducing math anxiety http://www.mathpower.com/reduce.htm

Interactive Visualizations

- Cohen’s D and effect size visualization https://rpsychologist.com/d3/cohend/
- Confidence Interval interpretation visualization https://rpsychologist.com/d3/CI/
- Correlation strength guessing game (video-game style) http://guessthecorrelation.com/
- Correlation strength (Rossman-Chance applet) http://www.rossmanchance.com/applets/GuessCorrelation.html
- Correlation visualization https://rpsychologist.com/d3/correlation/
- P distribution visualization https://rpsychologist.com/d3/pdist/
- Statistical significance and power visualization https://rpsychologist.com/d3/NHST/

General websites

- Society for the teaching of psychology (Division 2 of the APA). https://www.apa.org/about/division/div2
- American Statistical Association’s Section on Statistics and Data Science Education. https://community.amstat.org/statisticaleducationsection/home
- UCLA’s Institute for Digital Research and Education page on Teaching Statistics. https://stats.idre.ucla.edu/other/mult-pkg/seminars/statteach/resources-for-teaching-statistics/
References


Contributors

Editors

Dr. Alisa Beyer is residential faculty at Chandler-Gilbert Community College. She has been teaching undergraduate statistics since 2005. She received her PhD in Psychology in 2007 from the University of Kansas. She has taught statistics as a standalone course and in combined statistics-research methods courses, with and without a lab component. She has taught full-time at a state college, liberal arts college, and community college.

Dr. Janet M. Peters earned her PhD in Industrial and Organizational Psychology from Colorado State University. As a Clinical Assistant Professor at Washington State University Tri-Cities, her scholarship focuses on student and employee engagement. As Director of Instructional Excellence & Innovation, she is passionate about implementing empirically supported teaching strategies and helping others find creative ways to include them into their various disciplines.

Authors in chapter order

Dr. Jessica Hartnett an associate professor in the Department of Psychology and Counseling at Gannon University in Erie, PA. She spends a lot of time teaching, researching, and blogging about the teaching of statistics (https://notawfulandboring.blogspot.com/). In her spare time, she plays Throw Throw Burrito her two sons and makes a lot of coffee to drink with her husband.

Dr. Georjeanna Wilson-Doenges is a Professor of Psychology at the University of Wisconsin-Green Bay. She received her PhD in Social Ecology from the University of California-Irvine. Her research interests involve benchmarks of SoTL research, best practices for online teaching, and sense of community and fear of crime in neighborhoods. She has taught statistics, research methods, conservation psychology, and environmental psychology for over 25 years. Throughout her career, Georjeanna has emphasized teaching and writing about data analysis in clear and accessible ways. Georjeanna is involved in STP, specifically in the SoTL Workshop where she serves as the director and statistical consultant.

Dr. Donelle (Dee) Posey is Clinical Associate Professor and Associate Chair for Undergraduate Education in the Department of Psychology at Washington State University. Her primary teaching responsibility is the program’s Psychological Statistics course. She has received grants, conducted research, and presented on the development and implementation of the flipped classroom model.

Amy Nusbaum is a Ph.D. candidate in experimental psychology at Washington State University. Her research interests include active learning, open education, and improving outcomes for marginalized students in higher education. When not in the classroom, she also run a student food bank, directs an undergraduate peer mentoring program, works with
Undocumented Initiatives, and serves on the executive board of the Commission on the Status of Women. Her favorite part of her job is helping students find and own their power.

Dr. Ashley Waggoner Denton is an Associate Professor, Teaching Stream at the University of Toronto. She received her PhD in Social Psychology from Indiana University in 2012. Ashley teaches over 2,500 Introductory Psychology students each year, as well as teaching courses in Statistics and Social Psychology. In collaboration with undergraduate students, Ashley conducts research on topics that lie at the intersection of social psychology and pedagogy and she has published in outlets including Social Psychology of Education and Psychology Learning and Teaching. She is passionate about creating opportunities for her students to learn and grow both inside and out of the classroom, and she has championed numerous initiatives within her department that provide students with opportunities for community-building and professional skill development. Ashley is the 2019 recipient of the Jane S. Halonen Teaching Excellence Award given by the Society for the Teaching of Psychology.

Dr. Kristel M. Gallagher specializes in social and health psychology, teaching a wide variety of undergraduate courses in these content areas, as well as being the primary instructor for statistics and research methods for psychology majors at her institution. She has received numerous institutional, regional, and national grants and awards for her innovative teaching methods and enthusiastic approaches in the classroom. Dr. Gallagher also maintains an active teaching-focused research program, examining various approaches to enhancing student academic success and personal well-being.

Dr. Kristin Flaming is a Research Associate for the National Science Foundation funded Passion-Driven Statistics project. Her academic training is in educational psychology with a specialization in research, evaluation, measurement, and statistics. Previously she taught for a psychology department at a regional institution primarily teaching Introductory Statistics, General Psychology, and graduate level Research Methods. She has implemented the Passion-Driven Statistics curriculum in her own classroom.

Dr. Lisa Dierker is the Walter Crowell University Professor of Social Sciences at Wesleyan University. Her research on addiction sparked in her a passion for introducing a wider and more diverse audience of learners to the skills needed to ask and answer the most challenging and important questions we face today. Through funding from the National Science Foundation, she is disseminating a welcoming, project-based model for teaching data analysis and applied statistics. At the heart of this initiative is the belief that access is more than the availability of a seat in the classroom. It is about a warm and welcoming seat at the table.

Dr. Charlie Collins is an assistant professor of community psychology at the University of Washington Bothell. His teaching is guided by one underlying philosophical principle – teaching must be paired with the practical application of theoretical ideas to promote a high level of learning, critical thought, empowerment, and active engagement among students. This framework strives to engage students in critical thinking and awareness of larger community and societal issues while simultaneously teaching theoretical processes that underscore social problems. He applies this philosophy to content courses (e.g. introduction to community psychology) and methods courses (e.g. statistics) and has resulted in deep partnerships with organizations throughout the Pacific Northwest.
Dr. Alisa Beyer is residential faculty at Chandler-Gilbert Community College. She has been teaching undergraduate statistics since 2005. She received her PhD in Psychology in 2007 from the University of Kansas. She has taught statistics as a standalone course and in combined statistics-research methods courses, with and without a lab component.

Dr. Connor P. Principe received his Ph.D. in psychology from the University of Texas at Austin, where he was also a Statistics and Scientific Computation Graduate Fellow. Dr. Principe is now Associate Professor of Psychology at Pacific University where he has received awards for his teaching and scholarship. Most recently, he integrated separate research methods and statistics course to create a writing-intensive combined course that has been well-received by students.

Dr. Janet M. Peters earned her PhD in Industrial and Organizational Psychology from Colorado State University. As a Clinical Assistant Professor at Washington State University Tri-Cities, her scholarship focuses on student and employee engagement. As Director of Instructional Excellence & Innovation, she is passionate about implementing empirically supported teaching strategies and helping others find creative ways to include them into their various disciplines.

Dr. Miranda McIntrye is an assistant professor of psychology at California State University, San Bernardino. She studies interests from the perspective of personality and social psychology, with an emphasis on understanding participation and representation in STEM fields. This research focus also informs her approach to teaching statistics, which capitalizes on interest to engage and motivate students.

Dr. Joshua D. Fetterman obtained his B.S. in Psychology and minor in Philosophy from York College of Pennsylvania in 2003, and his PhD in Psychology (social area) from the University of Pittsburgh in 2012, where he studied memory in groups and the effects of power on motivation, cognition, and affect. He is now an Assistant Professor of Psychology at Chestnut Hill College in Philadelphia, Pennsylvania where he teaches in both the undergraduate and graduate psychology programs. He teaches a variety of classes including research methods, statistics, social psychology, cognitive and affective psychology, and personality psychology. His research interests include power, motivation, personality, and innovative teaching techniques.

Dr. Meredith E. Kneavel is Professor in the Department of Urban Public Health and Nutrition and Associate Dean in the School of Nursing and Health Sciences at La Salle University in Philadelphia, Pennsylvania. Dr. Kneavel’s current research spans two main areas: student-athlete health and pedagogy. A main focus of student-athlete health has been understanding concussion and concussion reporting. Her and a colleague developed and assessed a novel peer concussion education program which was supported through grants from the NCAA and Department of Defense. As part of her work with concussions, she also serves as the Director of Assessment and Research for the Center for Concussion Education and Research at Chestnut Hill College. Her pedagogical research focuses on understanding effective techniques in the classroom and particularly ways to enhance learning and understanding in statistics. Dr. Kneavel has also studied gender differences and the effects of stress on learning and memory. Dr. Kneavel received her B.A. in Psychology from Loyola University in Baltimore and her Ph.D.
in Psychology with a concentration in Biopsychology from the Graduate School and University Center of the City University of New York in New York, NY.

**Dr. Peter Allen** is a Lecturer whose teaching is focused mainly on research methods and statistics. For the last several years, his research has centred on evidence-based learning and teaching in higher education, with a particular emphasis on the development of statistical literacy, the use of virtual patients for the training of fundamental clinical skills and peer learning. Peter is particularly interested in understanding the barriers that psychology students face when learning research methods and statistics, and strategies that can help them become better researchers and scientific thinkers.

**Dr. Jess Fielding** is a Lecturer whose teaching is primarily based on research methods and statistics. Her research background is seated within cognitive neuroscience, previously conducting research looking at the neurological pathways involved in pain perception, and at emotional resilience and individual differences in anxiety, in healthy participants. More recently she has been looking at reward based mechanisms in the brain and understanding underlying connectivity in decision making processes. Alongside this neuro-behavioural research she is also interested in pedagogical literature looking at understanding and improving learning and the student experience with particular emphasis on the impact of emotional resilience.

**Ryan Kay** is a Neuroscience BSc graduate and contributed to this work in addition to his study of Experimental Psychology (Conversion) MSc. His main research interests include the neurological processes of learning, and applications of psychology in the clinical workplace. Specifically, Ryan is interested in investigating factors involved in staff burnout and psychological distress in healthcare professionals.

**Elizabeth East** contributed to this research whilst completing an MSc in Experimental Psychology. Her research interests include teaching and learning, mental health, and philosophical psychology.

**Dr. Tamarah Smith** is an associate professor of education with ten years’ experience teaching quantitative methods courses in multiple disciplines. Her scholarship is focused on statistics education, particularly the way in which students’ anxiety, attitudes, and beliefs impact outcomes in statistics.

**Dr. Jennifer V. Fayard** teaches research methods, meta-analysis, social and personality psychology, and psychology of creativity at Ouachita Baptist University. One of her primary goals is to help students fall in love with research through emphasizing its practical significance and fostering students’ science self-efficacy.

**Dr. Lisa Bauer** teaches a variety of undergraduate classes including: Cognitive Psychology, General Psychology, Human Cognition Capstone, Human Memory, Research Methods in Psychology I, and Research Methods in Psychology 2 (Statistics). She also enjoys mentoring graduate students in the Teaching of Psychology practicum and collaborating with current and former students on teaching-related activities. She hopes that her passion for teaching inspires her students to discover their own passions and motivates them to follow their dreams.

**Mike Corcoran** is a Social and Personality psychologist with a background in Industrial and Organizational Psychology. His teaching emphasizes hands-on engagement with psychological
research and psychological research methods, and he thoroughly enjoys mentoring and collaborating with others in research projects. He hopes that his passion for psychology, teaching, and research inspire students to aspire to new heights and to be critical and engaged thinkers both inside and outside the classroom.

**Nathaniel Greene** is a doctoral student in cognitive psychology with a strong emphasis in statistical methodology. His teaching aims to promote critical thinking about statistical practices in psychology by encouraging students to think about the assumptions and limitations of statistical models. In addition to his teaching and research activities, he enjoys mentoring undergraduate students through their capstone research projects. He hopes that his teaching will inspire students to become critical consumers of knowledge, both in the classroom and beyond.

**Dr. Thomas Cavanagh** is an assistant professor of management who has taught statistics to diverse populations of students, including undergraduate honor students, mid-career MBA students, and PhD students in the social sciences. He strives to help students of all backgrounds understand statistics at a conceptual level so that they can use them as a framework for understanding complex relationships in their own lives.

**Dr. Abby Heckman Coats** is an associate professor and chair of psychology at Westminster College in Fulton, Missouri. She teaches a variety of courses related to Developmental Psychology and research methodology. Her research focuses on social-emotional development in adulthood.

**Dr. Andrew Mienaltowski** is an associate professor of Psychological Sciences at Western Kentucky University in Bowling Green, Kentucky. He teaches general education psychology to introduce undergraduate students to the discipline as well as courses for majors like cognition and statistics. His research interests include age differences in emotion perception and the neuroscience of emotion processing.

**Dr. Jennifer Samson** earned her B.S. in Elementary Education/Child Development and her M.S./PhD. in Psychological Science (Developmental/Quantitative Methods) from Vanderbilt University. She taught statistics and research methods at Arkansas Tech University for 5 years before moving to Queens University of Charlotte in 2017, where she is primarily responsible for the Psychology Research Methods course sequence. Dr. Samson’s research interests include children’s relationships and meta-analytic methods, which affords her the opportunity to involve students in a variety of projects.

**Dr. Alexander J. Bies** an Assistant Professor of Psychology at Gonzaga University and a section editor for the journal Neuroscience Insights. He earned his PhD from the University of Oregon, during which time he was awarded an American Psychological Association Dissertation Research Award. Alex enjoys using online tools to get his students engaged with statistics, and is deeply committed to helping students learn statistics through experiences that are less painfully dull than his own were.
For the Love of Teaching Undergraduate Statistics

AUTHORS

Allen  Bauer  Beyer  Bies  Cavanagh  Coats
Collins  Corcoran  Duerker  East  Fyzerd  Fetterman
Fielding  Flaming  Gallagher  Greene  Kneaval  Hartnett
Mcintyre  Mianaltowski  Nusbaum  Peters  Posey  Principe
Samson  Smith  Waggoner Denton  Wilson-Doenges