Psychology 359: Advanced Research Methods in the Behavioural Sciences
Tentative Syllabus: Fall 2023

Lectures: Tuesdays and Thursdays, 2:00–3:15 in BUCH A104.

Lab: Tuesdays 3:30–4:45 in BUCH A203 (bring your laptop).

Instructor: Jeremy Biesanz, Professor, 4351 CIRS.

Office hours Mon 1–2pm by appt.

Teaching Assistant: Hin-Ngai Fu. Office hours Wed 11–12 in Kenny 3512.

Course Information and Objective

Psychology 359 is an introductory course dealing with the basics of behavioral statistics, experimental design, and computer applications. Students should have had at least the equivalent of a one-semester undergraduate course in behavioral statistics—this means that no instruction in elementary descriptive statistics is included in Psychology 359.

We will examine in depth the theoretical underpinnings of inferential statistics and selected inferential procedures (e.g., correlation and *t*-test and introduction to the general linear model). In addition to the statistical content covered in the lecture part of the course, some topics in experimental design, current research techniques, and issues in behavioral research, along with use of available open-source statistical software will be covered in both the lab and lecture components of the course.

This course is an advanced introduction to statistical methods in psychology with the following goals.

1. An understanding of the current issues facing Psychology and other scientific disciplines with respect to replicability and reproducibility. How has science and inference often been practiced and how should it be conducted instead?
2. Detailed understanding of best practices in data science with respect to laboratory research. We will focus on sustainable research workflows from idea conception through manuscript submission. The primary focus will be on best methods and practices for planning studies before collecting data and how to preserve and document data and analyses in a manner that would be compatible with open science.
3. Understanding and familiarity with open science. We will briefly cover pre-registration, open data, open materials, and reproducibility. By the end of the course you will be fully familiar with all of these and be able to implement these in your own research if and when needed.
4. A general introduction to statistical inference. Some of this should be review, but many of you will have covered this at different levels in previous courses. Here we will cover parameter estimation, sampling theory, hypothesis testing, confidence intervals, effect sizes, causal inference, statistical power, and understanding p-values. These are the basic inferential tools for all statistical analyses and for understanding your data. We will generally use the simple two-group experimental design (i.e., *t*-tests) while covering these topics.
5. An introduction to the open-source statistical computing language R. Many of the techniques and methods that we will cover are easily implemented in R but would require specialized (and expensive) software otherwise.

Class Discussion Board

The class discussion board will be on Piazza. I encourage you to use the discussion board to ask questions and solicit advice from your fellow students as well as your instructors. Rather than emailing questions to the teaching staff, I encourage you to post your questions on Piazza. Find our class signup link at: Piazza Signup Link.

Text

Navarro, D. (2016). *Learning Statistics with R: A Tutorial for Psychology Students and Other Beginners (Version 0.6).* A pdf is available at LSR.

Evaluation

Course student evaluation will consist of the following:

1. Problem sets (40%). A number of problem sets will be assigned during the course. These will be separate from the work done in lab. Problem sets will be submitted online as pdf files through Canvass. Problem sets *will not* be equally weighted — problems sets with more points will carry more weight.
2. Midterm (25%) and final exam (35%)

The overall class average is likely to be curved in a manner that reflects that this is a selective course. Consistent with previous years, I anticipate that the class average will be ~80% with a standard deviation of ~7.

Software

You will need a laptop or computer. The following software will be used extensively this term:

> R will be the primary major statistical package that we will use. R Studio is an interface (and more) for R. TeX. Download site for Mac and for Windows. General information here. This is used to generate pdf files in R Studio.

Handouts, Additional Materials and Readings, and Datacamp

Class materials will be made available through Canvas including readings, lecture notes, problem sets, and other class materials. I will also make extensive use of the Piazza discussion board on Canvas and encourage you to post questions there as well.

Throughout this term we will also make extensive use of Datacamp. We will provide details on which courses (videos) to watch on the class Canvas discussion forum and in lecture/lab. Although completion of these courses on Datacamp will not explicitly be part of the course evaluation, they will provide the essential training and background to complete the problem sets and final project. In other words, they are necessary but not formally graded.

Strategy for the Course

It is critical to keep up with the course and the readings on a continuous basis. It is a good strategy to review your notes from the previous lecture before coming to class. In this way you will discover if parts of your notes are not clear can ask for clarification in class. I will look to you throughout the course for feedback about your level of understanding. *You should ask questions in class.* I highly encourage it! If you have a question, it is very likely that other students in the room have the same question. It helps to actively participate in class.

Topical Units

Best Data Science Practices (BDSP)

What is the ideal workflow? How can we achieve this? We will revew the entire workflow and discuss and practice best practices in version control, data cleaning, dataset organization (tidy data), codebooks and documentation, reference managers, and archiving.

Review of Statistical Inference (RSI)

Here we will cover stanard statistics distributions (e.g., normal, $t(df)$, $\chi^2(df)$, and $F(df_1, df_2)$), null hypothesis significance testing, and p-values.

Commonly used Statistical Tests (CST)

We will examine and focus on the t-test and correlation in depth. We will consider different versions of these tests (e.g., assuming that groups have different variances, correlations with categorical variables and rank-order associations). These statistical models will be used for the rest of the course for inference, effect size estimation, and study planning.

Best Reporting Practices (BRP)

How do we estimate effect sizes? How do we determine the uncertainty associated with effect sizes? We will cover standardized effect sizes and approaches to determining confidence intervals for effect size estimates.

Study Design and Planning (SDP).

What is statistical power? How can we estimate statistical power for a new study? Why is retrospective statistical power not a useful concept?

Open Science (OS)

How does open science relate to best data practices? We will cover preregistration, open data, and open materials and discuss each of these in depth.

Current Crisis in Science (CCS)

Although sometimes referred to as the crisis in Psychology (or more focused on specific areas), it is clear that science in general has a problem with how business as usual has been conducted. We will review the crisis, discuss problematic practices, and solutions to these problems.

General Linear Model (GLM)

We will examine how to analyze continuous as well as categorical independent variables in the multiple regression framework. This will be a basic introduction to the general linear model, focusing on the interpretation of regression coefficients, how to extract specific information from analytical models, as well as how to examine assumptions and diagnostics and understand and conduct robust analyses.

Week	Topics and LSR Chapter
1. Sept 7	Introduction and Overview
2. Sept 12–14	Inferences and hypothesis testing. (LSR 11)
3. Sept 19–21	Probability distributions, p-values, and quantiles. (LSR 9)
4. Sept 26–28	t-tests for independent groups and correlations. (LSR 13)
5. Oct 3–5	Measures of association. (LSR 12)
6. Oct 10	Inference, errors, and the fragility of p-values. No class Oct 12.
7. Oct 17–19	Effect size estimates and confidence intervals. (LSR 10)
8. Oct 24–26	Statistical power and study planning
9. Oct 31–Nov 2	Review (Oct 31) and Midterm (Nov 2)
10. Nov 7–9	Introduction to the general linear model. (LSR 15)
11. Nov 14–16	Introduction to the general linear model (cont.)
12. Nov 21–23	One-Way ANOVA (LSR 14)
13. Nov 28–30	Analysis of Covariance (ANCOVA)
14. Dec 5–7	Open science and best practices

Table 1: Tentative Outline for the Fall 2023 Term.
LSR refers to the Learning Statistics with R textbook.
Additional readings for each week will be made available in Canvas.