# Psychology 613: Multivariate Statistics
## Spring 2022, McKenzie 471, T/R 10:00 – 11:20

https://uoregon.zoom.us/j/95215888295?pwd=bHlCQm5pazdJdlBYbjlNbXpvYitHUT09

**Instructor Information**

Professor: Elliot Berkman
Office hours: Tuesdays 2-3pm in LISB 229 or https://calendly.com/prof-berkman/berkman-office-hours
Contact info: berkman@uoregon.edu

| TA Name | Office hours | Contact |
|---|---|---|
| Cameron Kay | Wednesdays, 2:00-3:00pm, Zoom https://uoregon.zoom.us/j/92940472480?pwd=MzZUcERNQnM3bmZ6SXRsM0wzeG51dz09 | ckay@uoregon.edu |
| Sara Lieber | Thursdays, 11:30am-12:30pm, Zoom https://uoregon.zoom.us/j/94795147166 | slieber@uoregon.edu |

**Course description**

The multivariate statistical methods used in psychology and allied fields are broad in range and deep in complexity. A thorough treatment of any one of these could encompass an entire course. However, most advanced multivariate statistical techniques share many underlying structures. The purpose of this course is to survey a number of these techniques, particularly those most popular among and useful to psychologists, while emphasizing a conceptual understanding of their underlying model. This will be accomplished through hands-on application of each technique using software packages.

This course has four educational objectives:

1. To survey a range of multivariate statistical methods of interest to empirical psychology.
2. To help you develop a strong conceptual understanding of multivariate models.
3. To introduce you to computational methods for each technique in **R**.
4. To give you sufficient understanding of each technique to self-teach for your own research.

**Course Organization and Requirements**

Lectures

The first principal component of the course is lecture. The goals of the lecture for a given statistical technique are **(a)** to teach you the conceptual logic of what the technique does and how it does it, when to use the technique, and what its assumptions are, **(b)** to show you the most common or popular computational method for the technique, and **(c)** to give you the opportunity to ask questions about how and when the technique might be used in your own research.

I strongly encourage discussion and questions. You are encouraged to participate in course discussions and to interrupt me when I lecture in order to ask a question or to share an insight.

Though attendance is not graded, I strongly recommend that you come to all lectures and obtain detailed notes for those that you are unable to attend. Notes will be posted on Canvas after each lecture.

Labs

The second principal component of the course are the lab and review sections. Each **Friday**, Cam and/or Sara will show you how to use the R software platform to apply the techniques for that week to real data sets. During the labs, you will work through example problems and may begin working on the homework for that week. You are encouraged to work in teams during the labs.

Problem sets

There will be **five problem sets** throughout the quarter. These assignments are intended to elucidate the concepts underpinning each technique, and as such will be more challenging conceptually than computationally. The problem sets will be assigned on *Thursday and due the following Thursday* at 5pm. You may begin working on the problem sets in the computer lab section on Friday. You may work in groups to generate the computer output, but ***the final product must be completed individually***.

Exams

There will be **one midterm** and **one final**. Both will be **take-home**. These will consist of a mix of computational and conceptual questions about the material. There will be no make-up exams and no extensions. The midterm will be distributed on **Thursday, April 28th** and due **Thursday, May 5th at 5pm**; the final will be distributed on **Thursday, June 2nd** and due **Thursday, June 9th at 5pm**. You ***may not work in groups*** on the exams, and the ***final product must be completed individually***.

## Grading

The midterm, final, and cumulative problem set score will each be worth 1/3 of your total grade. These scores will be combined and weighted to yield one score out of 100%. I will average the scores of the top 10% of the students and use that number to determine the cutoff for letter grades. To get an A- you will need to get 90% of the average top score, to get a B- you will need to get 80% of the top score, and so on. This system has the advantage of a curve in that if everyone does poorly on the exams because they are too hard, nobody suffers, but it is also possible for every single person to get an A (since you could all do as well as 90% of the average of the top 10% of students).

## Policies

Late/missed assignments. Due dates for each assignment are listed in the "Lecture/Assignment Schedule." Late assignments will not be accepted without an extension (which will usually be granted).

Cheating/plagiarism. Not allowed. This is graduate school; the purpose of the work is self-evident. If you find yourself tempted to cheat to get a good grade, simply ask for an extension instead.

Students with special needs. The UO works to create inclusive learning environments. If there are aspects of the instruction or design of this course that result in disability-related barriers to your participation, please notify me as soon as possible. You may also wish to contact Disability Services in 164 Oregon Hall at 346-1155 or disabsrv@uoregon.edu.

## Texts

There are no required readings. However, I will post selected readings in PDF form and excerpts from the following excellent sources on Canvas as *optional, contextualizing* content:

Abelson, R.P. (1995). *Statistics as Principled Argument*. Lawrence Erlbaum: Hillsdale, NJ.
Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer: New York, NY.
Kline, R.B. (2011). *Principles and Practice of Structural Equation Modeling* (3rd ed.). Guilford Press: New York, NY.
Navarro, D. (2020). Learning Statistics With R. https://learningstatisticswithr.com/lsr-0.6.pdf
Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Sage: Thousand Oaks, CA.

**Lecture/Assignment Schedule**

| Month | Day | Topic | Reading | Assignment |
|---|---|---|---|---|
| March | 29 | Introduction and overview | Abelson, Preface and Ch 1 | |
| | 31 | Introduction to matrix algebra | Woodward et al., 1990 Appendix A (Canvas) | PS1: Matrix Algebra |
| April | 5 | Programming in R | LSR Ch 3 and 4 (Canvas) | |
| | 7 | Matrix algebra in R | | PS2: Central Limit Theorem |
| | 12 | Interactions and moderation I | | |
| | 14 | Interactions and moderation II | UCLA reading on decomposing intx | |
| | 19 | Psychometrics | | |
| | 21 | Factor and components analysis I | Handouts on Canvas | PS3: FA/CA |
| | 26 | Factor and components analysis II | | |
| | 28 | Structural equation modeling I | Kline, Ch 2 | **Midterm out** |
| May | 3 | Structural equation modeling in R | Kline, Ch 5 | |
| | 5 | Logit analysis | | **Midterm due at 5pm** |
| | 10 | Logistic regression I | | |
| | 12 | Logistic regression II | | |
| | 17 | Machine learning<br>*Special guest lecture* | Statistical Learning Ch 2 (Canvas) | PS4: ML |
| | 19 | Bayesian approaches | Navarro, Ch 17 LSR (Canvas) | |
| | 24 | Multilevel modeling I | R+B, Ch 2 | |
| | 26 | Multilevel modeling II | R+B, Ch4 | PS5: MLM |
| | 31 | Multilevel modeling III + R | A+M, 2010, pp. 13-18, 28-32 + Winter tutorials | |
| June | 2 | Network analysis<br>*Special guest lecture* | TBD | **Final out** |
| **Finals** | **7** | | | |
| **Week** | **9** | **Take-home final due at 5pm** | | |

*Note*. A+M = Albright and Marinova (2010); Kline = Kline (2011); LSR = Learning Statistics with R by Navarro; R+B = Raudenbush and Bryk (2002).