PSYCHOLOGY 642

General Linear Modeling (Part 1)

Instructor: Patrick E. McKnight, Ph.D.
Email: pmcknigh at gmu dot edu
Official GMU faculty page
Office: David King Hall 2064/2065
Office Hours: Tues 10:30a-noon or appointment
Phone: (703) 993-8292
Course Website
Class Locations: ONLINE FALL 2023
Class Google Group

TA: Kevin Loo
Email: kloo at gmu dot edu
Lab Locations: Virtual/Online
ZOOM LINK: LAB MEETING
TA Office Hours: by appointment
Office: DK2032

Course Overview

Psychology 642 is the first course of a two-course sequence that serves to introduce psychology graduate students to the general linear model. Traditional graduate psychology statistics courses emphasize statistical techniques as a matter of declarative knowledge. Students are expected to know each procedure and its “appropriate” application. An alternative approach tends toward technical discourse (e.g., matrix algebra, formula memorization, and hand calculations) and requires greater attention to minute detail and mathematical vernacular. A less used but equally suitable approach treats statistics as a method of principled argument. The method I use for this course is a hybrid of the three approaches. You will be expected to know the statistical terminology, apply your knowledge in both carrying out the procedures as well as interpreting the results, and then you will be expected to use the results in a manner consistent with scientific discourse.

Course Objectives

The purpose of this course is to further your introduction to data analysis, research design, and measurement. Your course work to date ought to have prepared you well by covering measures of central tendency, measures of dispersion, measures of association, and measures of difference. Due to time constraints, I do not intend to review these terms or their purposes so I urge every student to review that material prior to this course. What I do intend to cover is a comprehensive view of univariate, bivariate and multivariate statistics - why we use statistics, why you should learn these tools, and what are the most important features to learn and understand. You will gain practical skills in interpreting, applying and explaining statistical procedures. The combination of didactic sessions (remote) and weekly laboratory meetings with students will offer each student the opportunity to see the procedures, conduct the procedures yourselves, and then teach one another what you learned. See one, do one, and teach one is a standard approach used in contemporary medical education and modern project-based learning.

Course Prerequisites

Students ought to recognize and understand the following concepts and terms: measures of central tendency (e.g., mean, median, mode), measures of dispersion (e.g., variance, standard deviation, range), tests of difference (e.g., t-tests, ANOVA), measures of association (e.g., correlation, covariance), tests of association (e.g., multiple regression, chi-square), and research design. **Students who are not well-prepared with previous statistical coursework will be strongly encouraged to take a more introductory course.**

Do yourselves a favor and watch this video. Stop at the 4:40 mark and we will discuss why videos are not helpful to those who wish to actually **learn** statistics. After watching the 1 minute and 50 seconds of one of the top rated YouTube videos on regression, you will appreciate the value added by buying my time and attention. Watch it. I promise you will appreciate your newfound knowledge in our discussions (private) on flip.com. Join in!

At the end of the semester, you will achieve the following:

Learning Objectives

1. **Learn** a statistical software package
2. **Generate** statistical results independent of others
3. **Interpret** results accurately
4. **Understand** of the limitations of inference

5. **Develop** a vocabulary that allows you to read and interpret the literature
6. **Gain** in ability to choose the appropriate analytic approach for a specific question

Reading Materials

Required Textbooks (links req GMU authorization)

- Iversen, G.R. and Norporth, H. Analysis of Variance (Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 07-001). Newbury Park, CA: Sage. (ISBN: 978-0803930018) - **ref "IN" below**
- Carmines, E.G. and Zeller, R.A. (1979). Reliability and Validity Assessment. (Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 07-017). Newbury Park, CA: Sage. (ISBN: 978-0803913714) - **ref "CZ" below**
- Lewis-Beck, M.S. (1976). Applied Regression: An Introduction. (Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 07-022). Newbury Park, CA: Sage. (ISBN: 978-1483381473) - **ref "L-B" below**
- Fox, J. (1991). Regression diagnostics. (Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 07-079). Newbury Park, CA: Sage. (ISBN: 978-0803939714) - **ref "Fox" below**

Optional Textbooks

- Bruning, J.L. and Kintz, B.L. (1997). Computational handbook of statistics (4th Edition). Allyn & Bacon. (ISBN: 978-0673990853).
- Vickers, A.J. (2009). What is a p-value anyway? 34 stories to help you understand statistics. Addison Wesley. (ISBN: 978-0321629302)
- *Abelson, R.P. (1995). Statistics as principled argument. Lawrence Erlbaum Associates, Hillsdale, NJ. (ISBN: 0-8058-0528-1).*
- *van Belle, G. (2002). Statistical rules of thumb. Wiley-Interscience: New York. (ISBN: 0-4714-0227-3)*

Additional Optional Readings - DIRECTIONS (please read)

Linked below in the Module Content section are optional readings. These readings are intended to challenge your understanding of the topic. Read them. Take some time to read them early in the module. Each one pushes your working statistical vocabulary. You will not understand all the points in each article right away, however, you will take from each an understanding of something you never knew before. What did you learn? Tell me what you

learned but do so by behaving in a way that exhibits that knowledge - not by false report (ahem...cheating) or rote form (i.e., regurgitate to me what I said to you) but rather in an accurate, succinct, and knowing manner. Permit me to contrast them. Exhibit your knowledge of effect sizes:

False Report: <via chatGPT or any other LLM> Effect sizes quantify the magnitude of relationships or intervention impacts in statistics. Common measures include Cohen's d, Pearson's r, and odds ratio. They help interpret results by providing standardized, practical significance regardless of sample size, aiding comparisons across studies.

Rote: Effect sizes are defined with the following equation (d = (x1 - x2)/sp). Large effects are around 1 and small effects are close to zero. Effect sizes are important for calculating statistical power. Cohen said that effect sizes (d) estimates of 1+ are large, .5, medium, and .2 small. Effect sizes are also important for statistical power.

Deeper: When we find something significantly different, we only know that the observed mean difference is not consistent with the null hypothesis of **no mean difference**. In short, we know $p(D \mid H_0) < 0.05$. What we do not know is how big of a difference is it and should it matter? The “big” and “mattering” questions can only be estimated. Estimating size or big comes in the form of effect size as an estimate of magnitude but not of importance. An effect size is a metric of magnitude (i.e., "how big") and can calculated in many ways (Yeaton & Sechrest, 1981); I prefer Cohen's d (x1 - x2 / s*). Regardless, most estimates are "standardized" in a way that they can be modestly compared. Effect size is also relevant to statistical power; larger effects can more readily be detected. Do you have any questions?

You can do it! Give me a deeper appreciation of your knowledge. Show me what you know.

Grading Criteria & Rubric

Grading Basis

Grades will be determined by performance ratings on **three** video presentations. Put together a script that demonstrates your knowledge of the material listed in the module notes. Recording videos requires you to organize your thoughts, select the material most indicative of your knowledge and articulate your ideas succinctly (in 15 mintues or fewer). Each student must independently demonstrate proficiency with the statistical procedures; the video format allows you to present everything you know in a concise and well-organized presentation. All videos are **limited to 15 minutes**; speed and fluency of your

performance will be indicative of your proficiency. Students may submit a video at any time but no later than 72 hours before grades are due at the end of the semester. The later you submit your videos, the less time (or chance) you have to redo it if you should fail; why is this detail important? Each student may **retake one and only one module after failure**. Due to limited resources, we (my TA and I) must limit the number of retakes to one and the retake must be completed within two weeks of the failed attempt. All retakes must be submitted prior to the last day of classes. Failure to send and receive a grade on your work results in a zero ("0") for that module. Please help us all avoid such problems.

Grading Rubric

Performance on the modules will be graded on ternary scale (3-levels) where level 0 represents a failure to complete the module, level 1 represents a passing grade for the module, and level 2 represents a passing grade with the optional reading well-integrated into your responses. Since these are recordings of your presentations, I will not have the opportunity to prompt you; thus, you must communicate whatever you know about the topics within 15 mintues.

Final Grades

Three module videos that range from 0 to 2 each combine to give you a total score that ranges between 0 and 6. I assign final grades as follows:

6 pts = A; 5 pts = A-; 3-4 pts = B; 2 pts = C; < 2 pts = F

Module Content

The course content breaks down in this course to three modules as listed below. Each requires you to learn a set of concepts, integrate them into a coherent story, and apply them to data (linked below). For each module, you will have different data. All data come from a random generation process that I can reproduce the answers.

MODULE 1: Introduction to Data Analysis

1. Distributions (Optional Readings: 1; 2)
2. Measures of central tendency
3. Measures of dispersion
4. Data integrity
5. Data transformations

BUT NOT IN THAT ORDER. Please. Make me believe you know what you are talking about with these terms.

MODULE 2: The Nuts and Bolts (Optional Readings: 1; 2; 3)

1. Measures of Association
2. Reliability and Validity (?)
3. Measures of Difference
4. What's in a p? (Deeper Dive: Vickers, 2009)
5. Statistical power

MODULE 3: The Basic GLM (Optional Readings: 1; 2)

1. Bivariate regression & one-way ANOVA
2. Multiple regression & ANOVA (only main effects)
3. Sums of Squares

Human Subjects Training

One final requirement for PSYC 642 is the successful completion of the Collaborative Institutional Training Initiative (CITI) program. The program consists of an online exam that ensures that all persons who engage in work with human subjects understands the inherent risks you may expose those subjects to and how to avoid those risks. Additional information is available here. All students must show documentation that CITI course was completed to fulfill the PSYC 642 requirements. **Please submit your CITI training information to the following Google Form:** https://forms.gle/NVrJgUFDvsUibgT66

Lecture Format

I link videos of my own and of others (who offer them via a permissive license) throughout the notes. Often, the links change as the resources change. Please consult the website often and report any dead links or needless paywalled material; I shall fix posthaste. Thank you in advance.

Lab Format

The statistics lab content strictly parallels the course content. Please consult the course web site for changes in topics covered each week. Attendance at the lab is essential and required for you to master the skills discussed in the lecture and it serves as an excellent opportunity to test yourselves on the modules.

Academic Honesty

I must state for the record that cheating of any kind will be dealt with by rules set forth in the University Honor Code. I prefer never to have any academic integrity problems arise during the semester. The aim of graduate education is to learn material that many others have not learned and master this material to ensure your future success. The degree you receive reflects the hard work you put into your courses. Please do not cheat yourself by misrepresenting your effort. Do the work or accept the consequences. Spend your effort learning the material and avoid being overly grade conscious. With a concerted effort to learn, you will not be tempted to cheat.

ChatGPT (and other LLMs)

Use them. Learn how to use them wisely. Don't waste my time or your opportunity to learn by letting them represent your work and your thoughts. It's your money and time. It is our reputation. Let's work together to make your degree worth more than the paper it was printed upon. ***So, use them; do not abuse them (or you).***

Disability Accommodations

If you are a student with a disability and you need academic accommodations, please see me and contact the Disability Resource Center (DRC) at 703-993-2474. All academic accommodations must be arranged through that office.

Important Dates

8/28 (Last day to add classes), **9/5** (Final drop deadline with no tuition penalty), **9/13-9/26** (Web withdrawal period with 100% liability), **10/10** (Fall break - no Tues class), **11/22-11/26** (Thanksgiving recess), **12/2** (Last day of classes)