PSYC 754: Quantitative Methods III - Regression

SPRING 2019

Instructor: Patrick E. McKnight, Ph.D.
Office: David King 2064/65
Office Hours: Thurs 10:30-noon and by appointment
Phone: 703-993-8282
E-mail: pmcknighXX@gmu.edu (remove XX)
TA: Ze Zhu (Mia): zzhu5@gmu.edu
Office hours: Thurs 2:00-3:00pm or by appointment
Location: David King Hall 3074
Lab section 201: F 8:30-10:20am Innovation 317
Lab section 202: F 10:30am-12:20pm Innovation 317
Aziz Abubshait: aabubsh2@gmu.edu
Office hours: Thurs 2:00-3:00pm or by appointment
Lab section 205: Thurs 4:30-6:20pm Innovation 129

Class Time/Location: Section 001: Tues/Thurs 9-10:15am Planet 126
Section 003: Tues/Thurs 3-4:15pm ENGR 1110
Class Website: https://sites.google.com/view/gmupsyc754-new

Course Overview

The purpose of this course is to fully immerse graduate students into regression as an applied statistical tool. You - the student - will learn standard linear regression. What I refer to here as linear regression includes all parts of the general linear model (GLM) along with extensions (e.g., generalized linear models, generalized estimating equations, linear mixed effects models, path models, etc.). The first three-quarters of the semester focuses solely on the GLM while the final quarter shifts to the aforementioned extensions.

Due to time constraints along with a daunting work load, every student needs to devote no fewer than 10 hours each week to the course. Please consult students who completed the course and they will all tell you that my hour totals are probably underestimates. I specified a time because I want all students to be prepared. The huge course text requires your attention and every student must read the assigned readings **prior to class**. I say **must** because if you fail to read ahead of time, you will not be able to ask questions or understand **in class assignments**. Reading, therefore, takes place outside class and cannot be substituted by class attendance. If you devote the prescribed time and energy to the course, you will learn how to conduct a thoughtful analysis with these tools. Additionally, you will be able to appreciate the strengths and limitations of linear models.

Course Details

Required Texts

1. Cohen, J. Cohen, P, West, S.G., Aiken, L,S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). Mahwah, NJ: Erlbaum. (aka CC)
2. Berry, W.D. (1993). Understanding regression assumptions. pp. 13-83. Newbury Park, CA: Sage. (aka Berry; requires GMU login but FREE!)
3. Fox, J. & Weisberg, S. (2018). R companion to applied regression (3rd Ed.). Thousand Oaks, CA: Sage (aka Fox)

Course Requirements

I expect all students to have a good grasp of basic linear regression. You ought to be able to run both a bivariate and multiple regression - not just run them but interpret the results. My two-week review is fast and furious so if you have any reservations about your knowledge, please review your prior course material. Also, all of the assignments require a computer and a statistics program (more later). Lecture material will be available via youtube (or its equivalent) throughout the semester. If you do not own a computer with reliable internet access or do not have access to a statistical software program, I recommend you acquire them well before the course begins.

Statistical Software

Most psychologists use SPSS but I do not; I use R. You decide what tool you wish to use but please be aware that I am not able to help you with any software problems. Refer to the wealth of online resources to troubleshoot problems. I assume that you know how to use your computer and statistical software program according to the guidelines set forth in this class. For students who wish to learn R, I recommend you visit my online materials (see course website listing for more details). Additionally, those who are new to R and wish to learn it, I require you to read Fox's chapters 1 through 3 and chapter 10; he provides an excellent overview of R for the course.

Course Structure

I structure my graduate courses as hybrid flipped classrooms. A flipped classroom typically requires students to watch lectures at home and then come into class prepared to do what typically looks like homework. My course is a

hybrid in that I do lecture in class but only for brief periods. After each brief lecture, I ask that each student perform an activity. Given that structure, I strongly encourage every student bring in an easily portable computing device that allows you to run your software of choice. Some of you may not have such a device right now but I assure you that a fair bit of learning comes from the application. Do your best to find a notebook, laptop, tablet, or whatever you desire to perform the classroom activities. In class, I provide you an overview of the material relevant to your modules (more on that later) and you engage in actual data analysis pertinent to understanding that material.

Each student will be required to read assignments, attend class, complete the requisite tasks (see below), and perform three modules (see below). Readings consist of book chapters from the required texts as well as additional, content-relevant articles posted on the course website. By attending to the online materials, reading the assignments, attending class and contributing to class discussion, students get the best opportunity to master each topic.

Grading

Overview: As many of you know, I grade based upon demonstrated skill and knowledge via modules. You perform a pre-defined set of skills that I call a module. That performance is similar to a musical performance where you practice a piece prior to your performance. You perform in front of me or one of my astute TA's and you receive one of three module grades - a “0” for failure, a “1” for good job, or a “2” for exceptional performance. The first two grades are easy. You get a “0” if you fail to demonstrate adequate mastery of the checklist items below or a “1” if you satisfactorily perform each item. We assign a “2” to students who show a mastery of the material by seamlessly integrating optional reading into the module performance. Each grade depends upon your preparation and “game day” performance. If you rehearse your performance many times then you will do fine. Anything short of repeated rehearsals results in awkward negotiations between you and me (or a TA). I suggest we avoid those negotiations and have you prepare well in advance for your performance.

Modules: As I mentioned above, your grades are based upon 15 minute modules. I require you to perform three (3) modules - at least one in front of me and the other two may be submitted via youtube. Students have given me great feedback about the modules and many like the video option while others prefer the standard, in person option. I leave it to you to decide which option you prefer but remember you must perform at least one (1) in front of me (yes ME and not one of my TA's). Here is what I expect from each module:

Module 1: Bivariate and Multivariate Models (for each)

- ❑ Explain the purpose/differentiate between the two models
- ❑ Explain the parts of the model
- ❑ Conduct an analysis (SW)
- ❑ Explain ALL the results
- ❑ Conduct diagnostics (SW)
- ❑ Discuss implications

Module 2: Moderation and Mediation

- ❑ Explain moderated and mediated models (compare and contrast)
- ❑ Explain and demonstrate your knowledge of the different parts of each model
- ❑ Conduct a moderation model with one continuous and one binary predictor (SW)
- ❑ Conduct a mediation model with continuous predictors (SW)
- ❑ Explain ALL the results from each model
- ❑ Conduct diagnostics (SW)
- ❑ Discuss implications

Module 3: Coding & Generalized Linear Models (you choose which one)

- ❑ Explain the model and contrast it with a standard linear model
- ❑ Explain how coding categorical predictors can be accomplished
- ❑ Explain the model parts
- ❑ Conduct an analysis (main effects and at least one interaction with both categorical and continuous predictors) on your model and a standard model for comparisons (SW)
- ❑ Explain ALL the results from both your model and the standard linear model
- ❑ Conduct diagnostics (SW)
- ❑ Discuss implications

Tasks and Assignments

Prior to performing a module, I recommend you complete a set of tasks in front of a study partner or a random classmate. These tasks help each of you refine your skills by “seeing one,” “doing one,” and “teaching one” - the basis of the US medical education and the fundamental difference between my course and other courses you may take in graduate school. These tasks are not graded; instead, they serve as the foundational building blocks for each module. If you can

perform the tasks, explain them, and teach others, you will be well-prepared for your module. I strongly encourage you to keep track of your own task performance and redo tasks you feel relatively weak completing. Learn from me, my TA's, and your classmates. Learning is a cooperative - not competitive - endeavor. Work together and we will all learn from one another.

Academic Integrity

I must state for the record that cheating of any kind will be dealt with by the 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. Please note that academic dishonesty is not akin to studying with your classmates. I strongly encourage you to study together, exchange notes, and offer each other constructive feedback about your module preparation. My course is designed to eliminate any possibility of dishonesty. The only avenues to cheat yourself is by not doing the work. So, let me repeat myself; please study with one another but do your own work. I demand both.

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 and Information

Students are responsible for verifying their enrollment in this class. Schedule adjustments should be made by the deadlines published in the Schedule of Classes (see the semester calendar for important details).

Last Day to Add: January 29th, 2019

Last Day to Drop: February 5th, 2019

Topics and Timeline

Below lists selected topics with relevant readings. Note the abbreviations are included in the required text section above. Where I list a "c" before a number, you need to read the chapter. If I only provide a name for the reading, you must read the entire book. Use the following as a guide for our semester. I plan to take the first two weeks on the review (number 1) and then about a week for each topic following that review. In some cases, I might get ahead or slow down but use a rough rule of thumb that we will discuss 11 topics in 15 weeks. Each of the topics has a hyperlink to the specific video I created - both the theory material linked in the reading assignment and the technical details linked below the topic headings in the outline. Please read and watch the videos. Finally, please post your questions to this web form each week. My wish is that you submit your questions for each week by Sunday evening to allow me one full day to prepare my answers.

Welcome to Regression (PSYC 754)

1. Review of simple regression (CC c1 & c2 (parts 1, 2, and 3), Fox c1-c3 & c10 for R users, Fox c4 for all)
    a. Purpose of regression (part 1 & part 2)
    b. Line of best fit
        i. Errors and least squares criterion
        ii. Calculation of b and a
    c. Standard error of estimate
        i. Standard error of b
        ii. Standard error of a
    d. Meaning of b
    e. Standardized approach
    f. Line of best fit example from Pedhazur
2. Other purposes of regression (CC c3 (parts 1, 2, 3, and 4), Fox c5-c6)
    a. Contribution of multiple predictors
    b. Linear Interaction effects
    c. Simple Nonlinear relationships
    d. Categorical Predictors
    e. Nested data
    f. Nonlinear interaction effects
    g. Analysis of non-continuous dependent variables
3. Assumptions of regression (Berry & CC c4 , Fox c8)
    a. Why learn them? BLUE
    b. Fixed Variables

c. (Recti)linearity
d. Gauss-Markov overview
    i. Homoscedasticity
    ii. Normally distributed residuals
    iii. Independent residuals
    iv. Continuous DV
    v. Predictor variables
    vi. Reliability
    vii. Unique Variance
    viii. Expected residuals
    ix. Autocorrelation
e. In-class Exercise/Example

4. General multiple regression (CC reread c3 & c5 (parts 1, 2, and 3))
    a. Meaning of b
    b. Significance tests for R and B
    c. Multicollinearity
    d. Partial and semi-partial correlations
    e. Stepwise vs. Hierarchical (slides to help you)
    f. Alternative methods for handling multicollinearity
    g. In-class Exercise/Example
5. Regression diagnostics (Fox c8)
    a. Using diagnostics to identify violated assumptions
        i. Homoscedasticity (explained | R | SPSS)
        ii. Normally distributed residuals (explained | R | SPSS)
        iii. Independent residuals (explained | R | SPSS)
        iv. Continuous DV (explained | R | SPSS)
        v. Predictor variables (explained | R | SPSS)
        vi. Reliability (explained | R | SPSS)
        vii. Unique Variance (explained | R | SPSS)
        viii. Expected residuals (R | SPSS)
        ix. Autocorrelation (R | SPSS)
    b. Using diagnostics to identify bizarre occurrences
        i. Hat values
        ii. Leverage (explained | R | SPSS)
        iii. Cook's Distance (explained | R | SPSS)
        iv. Missing Data

NOTE: End of Module 1 Material (expected week of module 1: Feb 25th)

MODULE 1 Description

MODULE 1 Demonstration

6. Moderated Multiple Regression (CC c6 & c7)
    a. What is a moderator?
    b. Hierarchical test of significance for moderator
    c. Sums of Squares - revisited
    d. Simple slope and intercept method
    e. Plotting interactions
    f. Power considerations
    g. Non-linear effects
    h. Non-linear interactions
    i. Moderation in class exercise/example
7. Mediated Multiple Regression (see Additional Resources and Optional Readings)
    a. What is a mediator?
    b. Mediator vs. Moderator
    c. Testing mediation
    d. Problems with standard mediation tests
    e. Bootstrap tests with mediation
    f. Complex mediation models

    g. Additional Resources
        i. McKinnon article on mediation
        ii. Andrew Hayes' code for SPSS
        iii. Mediation resources for R users
        iv. Sobel Simple R code
    h. Optional Readings
        i. Original Baron and Kenny article
        ii. Mediation-Moderation details
        iii. Novel approaches to testing mediation
    i. Mediation in-class exercise/example

NOTE: End of Module 2 Material (expected week of module 2: March 25th)

8. Regression with categorical predictors (CC c8 & c9)
    a. Categorical predictors
    b. Dummy coding (R | SPSS)
    c. Interpreting dummy coded results (R | SPSS)
    d. Unweighted Effects coding
    e. Interpreting unweighted effects coding results
    f. Dummy and Effects in-class exercise/example (Week 1 of this section)
    g. Exercise above rehashed (don't peek until you try it yourself)
    h. A mid-section overview of the remaining topics
    i. Weighted effects coding
    j. Interpreting weighted effects coding results
    k. Contrast coding
    l. Interpreting contrast coding results
    m. Nonsense coding
    n. Summary of coding schemes
    o. Weighted effects and Contrast coding in-class exercise/example (Week 2 of this section)
9. Random Coefficient Modeling (CC c15, Fox c5)
10. Logistic regression (CC c13)
    a. Regression with dichotomous criteria
    b. Interpretation of output
11. Path Analysis (CC c12)
    a. Differences from Multiple regression
    b. Differences from structural equation modeling
    c. Parameter estimation and model fit.