POL228: Introduction to Quantitative Research Methods
Fall 2019

Instructor: Rabia Malik
Office Number: MHRC 163
Office Hours: Tues & Thurs: 10:30am-12:30pm
Email Address: rabia.malik@lums.edu.pk
TA: Waqas Shabir
TA Office Hours: Mon & Wed: 6-8pm
TA Email Address: waqas.shabir7@gmail.com

Course Description

Quantitative research methods are becoming increasingly popular in the social sciences, especially in political science and economics. Careful use of data and the right statistical techniques allow us to address research questions in systematic ways, and uncover relationships between interesting variables to better understand the world around us. This course will introduce students to the basics of quantitative research methods and data analysis with a particular emphasis on applications to the social sciences. The course will balance its focus between the theoretical and applied side of quantitative analysis so that students learn not just what various statistical methods entail but are also able to confidently analyze their own datasets using both descriptive and inferential statistical methods. Throughout the course, students will also learn how to implement statistical techniques using R so that they are able to write their own simple code for data analysis. By the end of the course, among other things, students will be able to:

- explain how and why we use quantitative research methods in the social sciences;
- choose what statistical methods are appropriate for different types of research questions and quantitative analysis;
- write simple code in R to conduct their own data analysis;
- critically analyze quantitative research that uses linear models;
- make arguments using data, and be comfortable conducting and interpreting statistical models

Course Requirements and Grading

The course does not expect you to necessarily have any prior knowledge of statistics or R but I do expect you to attend class regularly and participate actively. Note that asking questions when you don't understand something and helping your classmates also count as participation. The grading breakdown for the course is as follows:

- Class Participation: 10%
- Homeworks (4): 20%
- Quizzes (3, with 1 dropped): 10%
- Midterm Exam: 25%
- Final Exam (take-home): 35%

I understand that learning statistics and how to write code to perform statistical analysis can be intimidating at first. However, both the TA and I are here to help you along the way; as long as you continue putting in a good faith effort to learn and do your best, you have no reason to be scared of anything in this course. I will teach R in a very hands-on way, especially in the first few weeks, but there are also several very useful resources that I encourage you to use as reference books for help with R. David Dalpiaz's "Applied Statistics with R" is clear, easy-to-follow, and gives examples in code to help students understand what is going on. Best of all, it is available for free online: `https://daviddalpiaz.github.io/appliedstats/applied_statistics.pdf`. Especially if you have not used R before, I strongly encourage you to use both this book and RStudio's own primers (`https://rstudio.cloud/learn/primers`) to familiarize yourself with the basics of the program.

There will be a total of four homework assignments through the semester that will build on material covered in class or in the readings. Assignments will generally be due one week after they are given. You are allowed to work with other students in the class on these assignments. *However*, you are required to hand in an individual writeup of the results and explanations, which is not written in collaboration with anyone. In addition, if you do discuss the assignments in a group, I ask you to submit names of the other students who you worked with. Except for a documented emergency, in the case of which you will talk to me well before the deadline to discuss an alternative submission timeline, late assignments will be penalized by half a percentage each for each 12-hour slot of late submission. For instance, a score of 90% on an assignment (4.5% of the grade) will fall to 80% (4% of the overall grade) if the assignment is submitted 0 to 12 hours late; and so on.

I highly recommend that you use Rmarkdown to prepare your homework, which allows you to make reproducible documents that include both your code and your writeup. We will dedicate some time in one class in the first few weeks to understanding how to use Rmarkdown.

There will be a total of 3 unannounced quizzes in class, with the lowest grade dropped for each student since unforeseen circumstances can sometimes lead to missing a class. Since the lowest grade will be dropped, I will not allow makeup quizzes. Therefore, I highly recommend that you do not miss a class unless there is an emergency. Beyond quizzes, it will be difficult to catch up on the class material that you miss in such cases. If you do miss a class, it is your responsibility to borrow notes from a fellow student and go over them to catch up. Once you have done that, if you have any remaining questions or confusions, you are of course welcome to come and ask me or the TA for help.

The final exam will be a take home one where I will provide you with a dataset that you will have to analyze using the skills that you have acquired during the semester. We will discuss details of this closer to the end of the semester.

Finally, just a reminder that I reserve the right to change any of the readings or topics during the course of the semester based on student feedback or my own interpretation of how difficult or easy you are finding the material. In the case of any such change, I will inform you in good time.

Course Schedule

Session 1
Introduction and course overview: Why use quantitative methods in the social sciences? What is statistics? Descriptive versus Inferential statistics

1. . **Chapter 1: p13-20**.
2. . **Chapter 1: p3-5**.

Session 2
The Basics: Variables and their types; Populations versus samples; Introduction to R

1. . **Chapter 2, Sections 2.1-2.3: p23-33**.

Sessions 3 and 4
Describing variables I: Range; Interquartile range; Histograms; Frequency tables; Density plots

1. . **Chapter 2: p8-13**.
2. . **Chapter 3, Section 3.1: p41-46**.

Sessions 5 and 6
Describing variables II: Measures of central tendency (mean, median, mode); Variability of data (standard deviation, variance)

1. . **Chapter 3, Section 3.2: p47-53**.
2. . **Chapter 3: p43-51**.

HW 1 will be distributed and will be due in a week.

Sessions 7 and 8
Probability: Basic Rules; Discrete Probability Distributions

1. . **Chapter 4, Sections 4.1 and 4.2: p79-84**.

Session 9
Continuous Probability Distributions: Normal distribution; Z-scores

1. . **Chapter 4, Sections 4.2 and 4.3: p81-92**.
2. . **Chapter 5: p73-88**.
3. . **Chapter 8: p149-165**.

Sessions 10 and 11
Confidence Intervals; Standard Errors; Plotting CIs in R; Uncertainty of data

1. . **Chapter 7: p112-136**.

Sessions 12 and 13
Hypothesis Testing: Small and large samples; Type I and II errors

1. . **Chapter 8: p137-156**.
2. . **Chapter 6: p151-183**.

HW 2 will be distributed and will be due in a week.

Sessions 14 and 15
Understanding and analyzing the relationship between two variables: T-tests; Correlations; Visualizing the relationship; Scatterplots; Independent and dependent variables

1. . **Chapter 7, Sections 7.1-7.4: p191-205**.

2. . **Chapter 5: p63-77**.

Sessions 16 and 17
Linear regressions with continuous dependent variables; Ordinary Least Squares; Assumptions

1. . **Chapter 9, Sections 9.1-9.5: p259-284**.

Sessions 18 and 19
Interpreting linear regression output; Predicting outcomes; Visualizing linear models; Control variables; Statistical and substantive significance

1. . **Chapter 11: p187-210**.

2. . **Chapter 4: p37-51**.

3. . **Chapter 10, Section 10.1-10.2: p299-305**.

HW 3 will be distributed at the end of Session 18 and will be due in a week.

Sessions 20 and 21
Threats to inference: Outliers; Reverse causality; Other violations of OLS assumptions and how to check for them

1. . **Chapter 9, Section 9.6: p284-289** and **Chapter 10, Section 10.3-10.4:p306-313**.

2. . **Chapter 13: p209-242**. **Online Resource**.

Sessions 22 and 23
Binary dependent variables: Logistic regressions; Running and interpreting logit models

1. . **Chapter 15, Sections 15.1-15.3: p471-484**.

2. . **Chapter 7:, Section 7.1: p99-110**.

Session 24
Other types of non-linear relationships: Quadratic terms; Interaction terms

1. . **Chapter 14, Section 14.5: p451-456**.

2. . **Chapter 11, Section 11.2: p164-171**. **Online Resource**.

3. One reading may be added by instructor here.

Sessions 25 and 26
Simple data manipulation techniques in R: Cleaning datasets; Changing variable types; Renaming variables; Creating variables et cetera

1. . **Chapter 2:, Sections 2.2-2.5: p19-30**.

HW 4 will be distributed at the end of Session 25 and will be due in a week.

Session 27
Putting it all together: How do we choose data sources that will help test our hypotheses? How do we pick variables and model specifications that will convince our reader of our argument?

1. . **Chapter 16: p313-338**. **Online Resource**.

Session 28
Final exam review, and discussion of the final exam format.