POL 682: Quantitative Methodology II

Day/Time: M 1:00-3:30 PM **Professor:** Frank Gonzalez
Room: Social Sciences Rm 332 **Office:** 337 Social Sciences Bldg.
Office Hours: Tues. 12-2 (or by appointment, or **Email:** fgonzo@email.arizona.edu
anytime I am in my office) **Semester:** Spring 2019

Course Description and Objectives

In this course, you will develop a basic yet solid understanding of linear regression. Linear regression is the primary workhorse of quantitative research across disciplines, and allows researchers to test an incredibly wide range of complex research questions with *relatively* simple statistical techniques.

Despite its robustness, linear regression comes with a set of assumptions, and these assumptions are rarely (if ever) completely supported. As such, a thorough understanding of these assumptions as well as your data and the underlying statistics of your model is *essential* in order know how to properly interpret your results or when to move on to another type of model. Further, in order to evaluate the work of others (which is a major part of academia), you need to be able to apply your knowledge to the modeling decisions of other researchers.

The format of this course should be fairly straightforward. We will start slow with the basic matrix algebra necessary for linear regression. We will go on to actually derive an OLS (Ordinary Least Squares) estimator, which can get a bit intimidating but will seem much simpler with time and practice. Then, we will go over the assumptions of OLS regression, how they may be violated, and what may be done to address these violations. Specific topics during this section will include: heteroskedasticity, autocorrelation, (multi)collinearity, model specification, measurement error, outliers, and endogeneity. Finally, as time allows, we will discuss more advanced, non-linear models involving categorical dependent variables.

Course Objectives and Expected Learning Outcomes

1. Develop a solid and thorough understanding of the math underlying basic linear regression.
2. Understand the assumptions of linear regression, when they are violated, and potential ways of addressing violations.

Course Format and Expectations

This class will be similar to your statistics courses in that it will involve more lecturing by the instructur than discussion. However, questions, comments, and discussion during class are not only encouraged but also expected. Further, the course will often be interactive. As such, students are expected to have done all readings for the week *prior* to class, and to

come to class prepared to critically evaluate each reading.

There will be quite a bit of work to do. Specifically, you will be responsible for 8 assignments (3 involving mathematical problems and 5 involving programming in R), a midterm, a final, *and* a final paper (I know, and I am sorry but not sorry ☺).

Required Texts

- Gujarti, Damodar & Dawn Porter (2009). *Basic Econometrics, Fifth Edition.* New York: McGraw Hill.

  The majority of course readings will be research articles available through JSTOR or Google Scholar. I will post any readings not available on D2L (let me know if you cannot access an article and it is not on D2L).

Optional, Recommended Readings

- Berry, William & Stanley Feldman (1985). *Multiple Regression in Practice.* Newbury, CA: Sage.

- Dalgaard, Peter. (2008). *Introductory Statistics with R.* Springer-Verlag New York.

- Fox, John, & Weisberg, Sanford. (2010). *An R Companion to Applied Regression* (2nd Edition). Sage Publications.

- Gelman, Andrew & Jennifer Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press.

- Spector, Phil. (2008). *Data Manipulation with R.* Springer-Verlag New York.

- Also, www.rseek.org is Google (indeed power by Google) for R only.

Grading and Assignments

- Participation and Attendance - 10% each for a total of 20%: Attendance is required, and participation is as well when called for. Students who miss class due to illness or emergency are required to bring documentation from their health-care provider or other relevant, professional third parties. Failure to submit third-party documentation will result in unexcused absences.

  The UAs policy concerning Class Attendance, Participation, and Administrative Drops is available at: http://catalog.arizona.edu/policy/class-attendance-participation-and-administrative-drop. The UA policy regarding absences for any sincerely held religious belief, observance or practice will be accommodated where reasonable, http://policy.arizona.edu/human-resources/religious-accommodation-policy. Absences pre-approved by the UA Dean of Students (or Dean Designee) will be honored. See: https://deanofstudents.arizona.edu/absences.

- Problem Sets ($3 \times 5\%$ each for a total of 15%): There are 3 assignments that will require you to solve mathematical problems. You <u>must</u> show ALL of your work (partial credit will be given if work is shown but you simply arrive at the wrong answer)! If analysis in R is necessary, you must also submit your script/syntax file.

You may complete the assignments electronically (using Word, Latex, or some other word processor that can write out math equations/symbols) or by hand. I presume most of you will complete them by hand because unless you are particularly handy with Latex it can be a pretty huge pain to use a word processor on the computer, and so if/when you complete the assignments by hand, please **please** PLEASE(!) (please) try to have neat/legible handwriting so that I can understand what you are doing.

- Programming Assignments ($5 \times 2\%$ each for a total of 10%): Programming assignments will be assigned similar to those from the "Programming in R" course last semester. You should submit your scripts and any relevant output from these assignments on D2L.

- Midterm Exam (25%): Your midterm will be **in-class** and will require you to answer several short essay questions related to linear regression, the OLS estimator, Gauss-Markov theorem, and interpreting results. The exam will be "closed-book" but you may create *your own* (as in not one downloaded from the internet or a fellow classmate's) one-page notecard using a standard 8.5×11 piece of paper (both sides may be used). R is not required for the exams although you will be required to interpret R output that I provide.

  MAKE SURE TO BRING A PEN/PENCIL, PAPER, AND A <u>CALCULATOR</u> WITH YOU TO THE EXAM!

- Final Exam (25%): The rules and format for the final exam are the same as the midterm exam, except the questions will have to do with material from the midterm *as well as* new material related to when assumptions are violated and how to address those violations. Also, the final exam will be **take-home** and you will have the final week of the semester to complete it. You <u>will</u> need to use R.

- Final Paper (25%): For the final paper, you need to write up the Theory (partial), Data, Methods, Results, and Discussion (partial) sections for an academic research paper. You will need to develop hypotheses aimed at answering a research question (or questions), describe the data, variables, and analyses you will use to test those hypotheses, lay out the results of your analyses, and interpret the results.

  You are free to choose your own topic and research question(s) to address, but you must also find **your own** data set to analyze to address your question(s). You do not need to collect original data (and probably should not unless you are willing to put in a substantial amount of extra work) - you can find data sets through Dataverse, ICPSR, or other public data repositories. I encourage you to search for a data set and consult with me EARLY!

  Some guidelines for the final paper:

  - The dependent variable must be continuous (or at least arguably continuous). We will not get to using categorical dependent variables until late in the semester. Make sure to examine univariate statistics of the variable you are interested in to ensure it will reasonably work in an OLS regression.

  - Make sure your model is theoretically informed! You should <u>not</u> simply be throwing random variables into a model to see which ones work. Use relevant literature.

– Do not cut and paste R output! Put your results in a relatively "nice-looking" table, and include a figure or two if you feel so compelled!

Grade Scale

Letter Grade	%
A	90-100
B	80-99.9
C	70-79.9
D	60-69.9
E	59.9 and below

- **Requests for incomplete (I) or withdrawal (W)** must be made in accordance with University policies, which are available at http://catalog.arizona.edu/policy/grades-and-grading-system#incomplete and
http://catalog.arizona.edu/policy/grades-and-grading-system#Withdrawal respectively.

- Should you wish to discuss your grade, you must meet with me in person. I am not able to discuss grades by email.

Academic Integrity

Academic dishonesty is taken *extremely* seriously in this course. **Anyone found guilty of fabrication, falsification, or plagiarism will, <u>at a minimum</u>, receive an F for this course, and will be referred to university authorities.** Fabrication means altering information dishonestly, falsification means inventing information dishonestly, and plagiarism means presenting someone else's work as your own, either by not giving proper acknowledgement of the source or by presenting in whole or in part any work that has been prepared by or copied from another person.

Students are encouraged to share intellectual views and discuss freely the principles and applications of course materials. However, graded work/exercises must be the product of independent effort unless otherwise instructed. Students are expected to adhere to the UA Code of Academic Integrity as described in the UA General Catalog. See: http://deanofstudents.arizona.edu/academic-integrity/students/academic-integrity.

The University Libraries have some excellent tips for avoiding plagiarism, available at http://www.library.arizona.edu/help/tutorials/plagiarism/index.html.

Selling class notes and/or other course materials to other students or to a third party for resale is not permitted without the instructors express written consent. Violations to this and other course rules are subject to the Code of Academic Integrity and may result in course sanctions. Additionally, students who use D2L or UA e-mail to sell or buy these copyrighted materials are subject to Code of Conduct Violations for misuse of student e-mail addresses. This conduct may also constitute copyright infringement.

Class Conduct

This course will consist of a substantial amount of discussion and debate, and I look forward to class discussions that are energized and thought-provoking, but this is not an excuse for incivility in the classroom. Your fellow classmates will undoubtedly often hold different

opinions than your own, and I expect you to handle this disagreement with civility and respect for differing viewpoints. To foster a positive learning environment, students and instructors have a shared responsibility to do so. We want a safe, welcoming, and inclusive environment where all of us feel comfortable with each other and where we can challenge ourselves to succeed. Failure to show such respect will reflect in your final grade (specifically, points will be deducted from your participation grade).

It is also worth noting that laptops are acceptable in the classroom. However, they should be used strictly for note-taking.

Threatening Behavior Policy: The UA Threatening Behavior by Students Policy prohibits threats of physical harm to any member of the University community, including to oneself. See http://policy.arizona.edu/education-and-student-affairs/threatening-behavior-students.

UA Nondiscrimination and Anti-harassment Policy

The University is committed to creating and maintaining an environment free of discrimination; see http://policy.arizona.edu/human-resources/nondiscrimination-and-anti-harassment-policy.

Incompletes and Late Papers

Course incompletes will not be given except in extreme situations where all of the appropriate documentation is provided. Late assignments will be docked one letter grade if not handed in at the beginning of class on the due date (i.e., assignments handed in during the middle or at the end of class will already be docked one letter grade). Grades on the assignment will be lowered one additional letter grade for each additional day late.

Additional Resources for Students

UA Academic policies and procedures are available at http://catalog.arizona.edu/policies. Student Assistance and Advocacy information is available at http://deanofstudents.arizona.edu/student-assistance/students/student-assistance.

Confidentiality of Student Records

For more information on the confidentiality of student records at UA, please see: http://www.registrar.arizona.edu/personal-information/family-educational-rights-and-privacy-act-1974-ferpa?topic=ferpa.

Accessibility and Accommodations

Our goal in this classroom is that learning experiences be as accessible as possible. If you anticipate or experience physical or academic barriers based on disability, please let me know immediately so that we can discuss options. You are also welcome to contact the Disability Resource Center (520-621-3268) to establish reasonable accommodations. For additional information on the Disability Resource Center and reasonable accommodations, please visit http://drc.arizona.edu.

If you have reasonable accommodations, please plan to meet with me by appointment or during office hours to discuss accommodations and how my course requirements and activities may impact your ability to fully participate.

Please be aware that the accessible table and chairs in this room should remain available for students who find that standard classroom seating is not usable.

Subject to Change Statement

Information contained in the course syllabus, other than the grade and absence policy, may be subject to change with advance notice, as deemed appropriate by the instructor.

Course Schedule and Readings

All readings should be done *prior* to the date for which the readings are assigned.

<u>Date</u>	<u>Topics and Readings</u>

<div align="center">

Intro and Major Concepts

</div>

Jan. 14	**Intro**

<div align="center">

no readings

*****Assignments:** none.

</div>

Jan. 21	**Martin Luther King Jr. Day**

<div align="center">

no class

</div>

Jan. 28	**What is this "regression" you speak of?**

<div align="center">

Gujarti & Porter *Introduction*, Ch. 1, and
<u>skim through</u> Appendices A & B

Gujarti & Porter Appendix 3A (pp. 92-96)

*****Assignments:** Programming Assignment #1 due
before class begins (Final from POL596A)

</div>

Feb. 4	**Single Variable Regression**

<div align="center">

Gujarti & Porter Ch. 2

*****Assignments:** none.

</div>

Feb. 11	**The OLS Estimator/** **Gauss-Markov Assumptions**

<div align="center">

Gujarti & Porter Ch. 3

Gujarti & Porter Appendix 5A (pp. 143-146)

*****Assignments:** Programming Assignment #2 due
before class begins

</div>

Feb. 18	**Model Fit, Inference, Multivariate Regression**

<div align="center">

Gujarti & Porter Ch. 5

</div>

Date	Topics and Readings

Matrix Algebra Refresher (available on D2L)

Lewis-Beck, Michael and Andrew Skalaban. 1990.
"The R-Squared: Some Straight Talk." Political Analysis,
2, 153-171.

King, Gary. 1990. "Stochastic Variation: A Comment
on Lewis-Beck and Skalabans The R- Squared". Political Analysis,
2, 185-200.

Achen, Christopher. 1990. "What does Explained Variance
Explain?: Reply." Political Analysis, 2, 173-184.

(Optional) King, Gary. 1986. "How Not to Lie with Statistics:
Avoiding Common Mistakes in Quantitative Political Science."
American Journal of Political Science, 666-687.

*****Assignments: Problem Set #1 due before class begins.

Feb. 25   **Interpretation or Regression Results**

Gujarti & Porter Ch. 7

*****Assignments: none.

Mar. 4   **Spring Break**

no class

Mar. 11   **Matrix Derivation of Multivariate Models**

Gujarti & Porter Ch. 8

Gujarti & Porter Appendix C

*****Assignments: none.

Regression Extensions

Mar. 18   **Variable Transformations and**
Interactions

Gujarti & Porter Ch. 6

Date	Topics and Readings

Brambor, Thomas, William Clark and Matt Golder. 2006.
"Understanding Interaction Models: Improving Empirical
Analyses." Political Analysis 14, pp. 63-82.

*****Assignments: none.

Mar. 25 **Dummy Variables and
Interactions (cont.)**

Gujarti & Porter Ch. 9

*****Assignments: Programming Assignment #3
<u>and</u> Problem Set #2 due before class begins.

Regression Assumptions

Apr. 1 **Midterm Exam (in-class) and
Regression Diagnostics**

Gujarti & Porter (part of) Ch. 11 (pp. 365-389)

(optional) Berry, William and Stanley Feldman. 1985 Multiple Regression
In Practice. Thousand Oaks, CA: Sage. Chapter 5.

*****Assignments: none.

Apr. 8 **Multicollinearity and
Heteroskedasticity**

Gujarti & Porter Ch. 10 (pp. 365-389)

Hayes, Andrew and Li Cai. 2007. "Using heteroskedasticity-
consistent standard error estimators in OLS regression: An
introduction to software implementation." Behavior Research
Methods, 39: 709-722.

(optional) Berry, William and Stanley Feldman. 1985 Multiple Regression
In Practice. Thousand Oaks, CA: Sage. Chapter 6.

*****Assignments: none.

Apr. 15 **Autocorrelation**

Gujarti & Porter (part of) Ch. 11 (pp. 389-401)

Date	Topics and Readings

Gujarti & Porter Ch. 12

Gujarti & Porter Ch. 17

(optional) Berry, William and Stanley Feldman. 1985 Multiple Regression
In Practice. Thousand Oaks, CA: Sage. Chapter 4.

*****Assignments: none.

Apr. 22	**Model Specification/** **Measurement Error**

Gujarti & Porter (part of) Ch. 13 (pp. 467-513)

Ansolabahere, Stephen, Jonathan Rodden and
James Snyder. 2008. "The strength of issues: Using multiple
measures to gauge preference stability, ideological constraint, and
issue voting." American Political Science Review, 102, 215-232.

Asher, Herbert. 1973. "Some consequences of measurement
error in survey data." American Journal of Political Science, 18,
469-485.

(optional) Berry, William and Stanley Feldman. 1985 Multiple Regression
In Practice. Thousand Oaks, CA: Sage. Chapter 3.

*****Assignments: Programming Assignment #4 due
before class begins.

Apr. 29	**Endogeneity**

Gujarti & Porter Ch. 18, 19, & (optional) 20

Gabel, Matthew and Kenneth Scheve "Estimating the effect
of elite communications on public opinion using instrumental
variables." American Journal of Political Science, 51, 1013-1028.

Green, Donald Philip and Jonathan Krasno. 1988. "Salvation
for the Spendthrift Incumbent: Reestimating the Effects of
Campaign Spending in House Elections." American Journal
of Political Science, 32, 884-907.

Date	Topics and Readings

(optional) Berry, William and Stanley Feldman. 1985 Multiple Regression
In Practice. Thousand Oaks, CA: Sage. Chapter 3.

(optional) Bartels, Larry. 1991 "Instrumental and Quasi-Instrumental
Variables." American Journal of Political Science, 35, 3, 777-800.

*****Assignments: Programming Assignment #5 &
Problem Set #3 due before class begins.

FINAL EXAM AND PAPER
both due by 11:59 PM on May 6th