EDUC 1110: Introductory Statistics for Education Research and Policy Analysis
Fall 2019

Lectures: Tues & Thurs, 10:30 pm - 11:50 pm.
Location: Smith-Buonanno Hall G13

Section: TBD
Location: TBD

Instructor: Pierre de Galbert
Email: pierre_degalbert@brown.edu
Office: 164 Angell Street, Office 266
Office Hours: TBD

Course Website: https://canvas.brown.edu/courses/1079319
Please check the course website frequently for important announcements, readings and assignments.

Course Objectives

This course provides an introduction to the use and interpretation of statistics for conducting quantitative research in the social sciences, with a focus on education policy. Students will become acquainted with descriptive and summary statistics, tabular and graphical methods for displaying data, statistical inference, analytic methods for exploring relationships between both categorical and continuous measures, and multivariate regression.

The primary goal of this course is for students to become more sophisticated consumers of basic quantitative research and capable data analysts who can begin to design and carry out their own quantitative research. In other words, the course aims to provide students with necessary skills at two levels:

1. As critical **consumers** of quantitative research, students should leave the course with:
    - An understanding of the central methods researchers use to describe data and to explore relationships between variables.
    - An ability to differentiate high-quality research from lower-quality research and to see where research does (and does not) support the inferences that authors make.

2. As beginning **data analysts**, students should leave the course with:

    - An understanding of the principles behind data and data analysis.
    - The ability to conduct their own original quantitative analyses using a range of core statistical methods.
    - The ability to talk and write about statistical analyses in a precise, clear, and accessible way.

Course Overview

The class takes as its fundamental assumption that you can best learn statistical methods by doing statistical analysis. Concepts and methods will be taught using real-world examples with multiple opportunities for students to apply these methods in practice. Even if you do not plan to enter the world of research, I believe that you will understand these methods best if you have used them in practice.

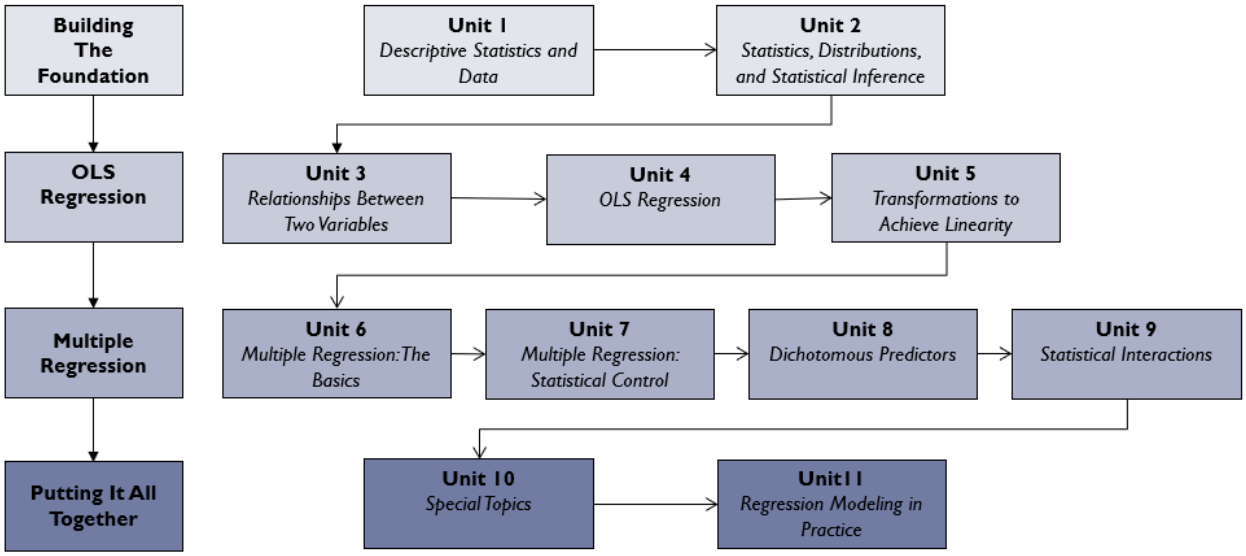
Unlike many other statistics classes, we will focus on *interpretation and analysis* rather than *calculation*. We will focus our attention on how to design an analysis, how to use the right analysis to address a particular question, and how to interpret results. It is much more important to me that you have an intuitive understanding of these methods than to be able to apply them by hand – that's what the computer is for!

Particularly later in the semester, lectures will be supplemented with research articles that exhibit the application of statistical methods in current educational and policy research. Class meetings will include lecture, small-group activities, large-group activities, and discussion of readings.

The course will be structured into four main parts:

1. Descriptive Statistics and Statistical Inference
2. Examining the Relationships Between Variables and OLS Regression
3. Multiple Regression and Statistical Control
4. Putting it All Together

The first weeks will set the stage, providing a brief overview of quantitative research, data, descriptive statistics, and statistical inference. In part 2, will examine how we can understand relationships between variables. We will spend the bulk of our time here on OLS regression, the main machinery behind most quantitative analysis. Part 3 will focus on multiple regression and statistical control. Finally, in part 4 we will examine some additional topics and walk through a complete data analysis together. The following roadmap illustrates the arc of the course.



Course Textbooks

Unfortunately, statistics textbooks tend to be written in ways that do not align directly with the way we cover material in this class. Thus, instead of spending a lot of time reading a statistics textbook, I would rather you spend your time (a) preparing ***carefully*** for class, (b) participating actively in class discussion, (c) reviewing and studying class materials, (d) completing the assignments, (e) working collaboratively to master the material, and (f) accessing supplementary materials as needed. We will focus attention largely on class slides. I will post these slides on the course website the evening before each class. **You should download the slides (and print them if you prefer) before class.** If printing is an issue, let me know and we can make alternative arrangements.

That said, textbooks can be particularly helpful in providing background material before we explore a topic together in class and can offer another perspective on the material, a different voice/approach, and additional topics (or topics in more depth) than we have time to do in class. As a result, I **strongly recommend** that you read about topics in the following text before class:

Recommended Text: (available on 3-hour loan at the Sciences Library)

- David Moore, George McCabe & Bruce Craig, *Introduction to the Practice of Statistics*, 9th Edition (2017)
    - *NOTE*: The 7th or 8th Edition is fine if that is much cheaper.

If the Moore, McCabe, and Craig text does not work well for you, you are certainly encouraged to find a different textbook that you like and consult it regularly. I have placed the following textbooks on reserve in the Sciences Library and recommend them (particularly the first two):

- Alan Agresti & Barbara Finlay. (2009). *Statistical Methods for the Social Sciences*, 4th Edition. Addison-Wesley.
- James Stock & Mark Watson. (2010). *Introduction to Econometrics*, 3rd Edition. Addison-Wesley.
- David Kleinbaum, Lawrence Kupper, Azhar Nizam, & Eli Rosenburg. (2013). *Applied Regression Analysis and Other Multivariable Methods.* Cengage Learning.
- Peter Kennedy. (2003). *A Guide to Econometrics*, 5th Edition. MIT Press.

For some classes, I will ask you to read research articles related to the topic of discussion. Please read these articles carefully and come to class prepared to discuss them.

Statistical Software

One of the goals of the class is to introduce you to a statistical software package. While Excel or other spreadsheet programs can do some analysis, for more sophisticated work you will need to use a program like Stata, SAS, SPSS, R, etc.

This year, we will be using Stata. You can download the software from Brown at:
https://www.brown.edu/information-technology/software/catalog/stata-se-1.
If you prefer to use another program to complete your assignments, that is fine but you will need to learn the program on your own.

You might find the following books useful resources in learning Stata. They are also on reserve in the Sciences Library:

- Alan C. Acock. (2012). *A Gentle Introduction to Stata*, Revised 3rd Edition. Stata Press.
- J. Scott Long. (2009). *The Workflow of Data Analysis Using Stata*. Stata Press.

The Acock book provides a general overview to using Stata. The Long book is an excellent resource for how to organize your analysis files and data.

In addition, there are many great on-line resources for learning Stata, including:

- Stata Cheat Sheets
  - http://geocenter.github.io/StataTraining/portfolio/01_resource/
- Stata Tutorial
  - http://data.princeton.edu/stata/default.html
- Official Stata Manual for "Getting Started with Stata"
  - https://www.stata.com/manuals/gsw.pdf
- E-book available through Brown library:
  - *The Stata Survival Manual* by David Pevalin and Karen Robson
- UCLA Website has excellent annotated Stata output and step-by-step guides
  - http://www.ats.ucla.edu/stat/stata/
- This Stata website has a variety of resources
  - http://www.stata.com/links/resources-for-learning-stata/
- Stata YouTube Channel: Very helpful for visual learners
  - http://www.youtube.com/user/statacorp?feature=watch

Course Requirements

Class Preparation:
Successfully completing this course will require you to read and review materials in advance of class each week. Given the scope of the course, there will not be time to present all material during classes. Students will be expected to prepare for class by reading assigned texts, reviewing class slides, watching short tutorials, and completing data exercises.

Class Participation:
The success of this class depends on the thoughtful, engaged participation of all class members. As a result, **attendance is expected for every single class**. Please notify me in advance if you will need to miss class.

Effective class participation is a matter of quality, not quantity. It requires that you come to class well-prepared to discuss all of the assigned readings and that you actively engage in class by posing questions and contributing to small- and large-group discussions. Asking clarifying questions and addressing points of confusion are highly encouraged.

You are not expected to be perfect; you are expected to try. One of the central skills I hope you learn from the class is how to talk about statistical concepts. In many ways, this is like learning a foreign language. This class will be a safe space to practice your understanding (rather than practicing for the first time in front of your boss).

Computers

To engage in in-class exercises, you will need to bring a laptop to class. If you do not have a laptop computer, that is not a problem. Please let me know and I will coordinate access to one for the class. However, while you are welcome to use laptops in class for taking notes, I strongly suggest you keep your computer closed during lecture and focus on engaging with the material. Much of the content covered in class will be illustrated through figures and graphs. I find it easier to take notes on these images using paper and pencil. If you are considering using a laptop please read this first:
For better learning in college lectures, lay down the laptop and pick up a pen.

If you do choose to use a laptop, please refrain from any other uses (e.g. checking email, web surfing, etc.) – these activities are distracting to me and (more importantly) to others trying to pay attention. Respect your classmates.

Problem Sets:

There will be five problem sets over the course of the semester. These assignments are designed to help you solidify your understanding of the key content and to give you the opportunity to apply statistical methods in practice. Each problem set will guide you through a real analysis of a dataset. You will need to conduct the required analyses in Stata, interpret your findings, and write up your results. **Two problem sets (PS 2 and PS 5) will be done individually**. For the remaining three problem sets, **you may choose to complete this work on your own or with a partner.** Pairs should turn in one assignment. Rather than dividing up the work on each assignment, each partner should take an active role in completing all analyses. In the past, students have found it helpful to complete the work on their own first before comparing strategies and refining. **You may only discuss your work with your partner**.

Applied Research Project:

Students will conduct an original research project that will require them to integrate a range of skills learned in the course. **Students are encouraged, but not required, to work in pairs to complete the applied research project in order to facilitate peer learning**. Groups of three or more will not be permitted. Students can obtain their own data or can use one of the datasets that I will make available on the course website for the project. The task will be to find a research question that is of interest and can also be answered with the data at hand. Project partners will turn in a one-page proposal in order to receive early feedback about their project ideas. Details on the paper will be provided in class later in the semester.

Grading:

Class Preparation and Participation will be graded holistically based on the quality, not the quantity, of students' contributions to class. There are three main components to successful class participation: 1) attending all classes and arriving on time, 2) coming prepared to discuss all of the required readings, and 3) actively engaging in class by posing questions, contributing to discussions, and being an active participant during in-class exercises.

Final Grade: Course grades will be calculated using the following **approximate** percentages

Class Participation	15%
Problem sets (5)	50%
Applied research project	35%

Important Policies

- Please submit all assignments via the course website.
- In order to be equitable to all students, I do not provide make-up assignments or extra credit.
- If you need an extension, don't hesitate to ask for it. They key is planning in advance and good communication. You must submit a request for extensions due to extenuating circumstances 24 hours before the assignment is due. Assignments turned in late without prior approval will lose one letter grade each day the assignment is late.

Academic Code:

I expect all students who enroll in my course to read, understand and abide by the Academic Conduct Code at Brown. In particular, please pay close attention to the section on "use of sources" (p. 6) to be sure to appropriately credit outsides sources and avoid any potential issue of plagiarism.

Credit Hours

The total of in-class hours and out-of-class work for all full-credit courses at Brown is approximately 180 hours for a 14-week semester. In this course, students can expect to spend ~34 hours in class (2h 40min per week for 13 weeks), ~26 hours preparing for class by reviewing assigned materials (2 hours per week), ~13 hours in section (1 hour per week), ~45 hours on problem sets (9 hours for 5 PS), ~60 hours researching and writing the final research paper.

Additional Support

I am committed to providing you with the resources necessary to meet your objectives in this class. I will hold regular office hours and am available to meet outside of those times as well. One note in this regard – it is incumbent upon you to let me know if you have areas of confusion. Ideally, you will raise these in class, but you can also let me know if you are struggling with concepts or have questions outside of class.

Similarly, I strive for an open, inclusive, and supportive class environment. If I do or say something that you perceive deviates from these goals, please let me know as soon as possible – ideally in the moment, but if not then after class (either directly or anonymously).

Section:
There will be a weekly section throughout the semester to support your learning process. This section will focus on applying methods learned in class using *Stata*, as well as reviewing concepts from lecture. Although this section is not required, it is ***strongly recommended***. Statistical computing techniques taught in section will not be covered in class.

If you cannot attend section, the practice problem sets that are done in section will be made available. I strongly suggest that you do these on your own to keep up learning the material and the statistical software.

Study Groups

I strongly encourage you to form study groups. Collaborative learning is an important component of this class, and you will get more out of the class if you engage with your peers outside of class time. You will also find that you come to class better prepared to share ideas and engage in discussion. **However, you should not discuss your work on the problem sets other than as described above.** Study groups are not required, but should you form one, I suggest that you include members with experiences, abilities, and career plans different from your own. Effective study groups typically have between three and four members.

Overview of Schedule

The following overview provides a tentative timeline for the course and assignments. There are 3 classes called "catch-up days" built-in the schedule. Catch-up days are built to allow for the fact that some units will take more time than others, depending on the level of interest and difficulty from students. Any changes to the timeline will be clearly communicated. The readings below are linked to Units, not to particular dates.

Class	Unit	Assignment	Due Date
Sep 5	Introduction		
Sep 10	Unit 1		
Sep 12	Unit 1	Stata Exercise	Sep 12 – 10 PM
Sep 17	Unit 2		
Sep 19	Unit 2	PS 1	Sep 23 – 10 PM
Sep 24	Unit 3		
Sep 26	Unit 3		
Oct 1	Unit 4		
Oct 3	Unit 4	PS 2	Oct 7 – 10 PM
Oct 8	Catch-up Day 1		
Oct 10	Unit 5		
Oct 15	Unit 5		
Oct 17	Unit 6	PS 3	Oct 23 – 10 PM
Oct 22	Unit 6		
Oct 24	Unit 7		
Oct 29	Unit 7		
Oct 31	Unit 8		
Nov 5	Unit 8	Project Proposal	Nov 7 – 10 PM
Nov 7	Catch-up Day 2	PS 4	
Nov 12	Unit 9		
Nov 14	Unit 9		
Nov 19	Unit 10		
Nov 21	Unit 10	PS 5	Nov 27 - **Noon**
Nov 26	Unit 11		
Nov 28	Thanksgiving Break		
Dec 3	Unit 11		
Dec 5	Catch-up Day 3		
Dec 10	Course Wrap-up		
		Applied Research Project	??

Readings and Materials

Unit 1: Introduction to Statistics, Distribution & Data

Required Reading/Videos/Stata Resources

- Read the course syllabus
- Review unit slides
- Moore, McCabe & Craig
  - Version 6/7/8 – Chapters 1.1, 1.2, 3.2
  - Version 9 – Chapters 1.1, 1.2, 1.3, 3.3
- The Age of Big Data – *New York Times* http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html
- What Data Can't Do – *New York Times* http://www.nytimes.com/2013/02/19/opinion/brooks-what-data-cant-do.html
- Big Data Transforms Education Research – Education Next http://educationnext.org/big-data-transforms-education-research-can-machine-learning-unlock-keys-to-great-teaching
- For better learning in college lectures, lay down the laptop and pick up the pen – Brookings https://www.brookings.edu/research/for-better-learning-in-college-lectures-lay-down-the-laptop-and-pick-up-a-pen

Recommended / Supplemental

- *A Quick Tour of the Stata Interface* by StataCorp:
  - https://www.youtube.com/watch?v=nV5WfR92LIM
- *Descriptive Statistics in Stata* by StataCorp:
  - http://www.youtube.com/watch?v=kKFbnEWwa2s
- *Descriptive Statistics* by Khan Academy:
  - https://www.khanacademy.org/math/statistics-probability/displaying-describing-data

Unit 2: Statistics, Distributions & Statistical Inference

Required Reading/Videos/Stata Resources

- Review unit slides
- Moore, McCabe & Craig
  - Version 6 – Chapters 1.3, (skim 5.1), 5.2, 6.1
  - Version 7/8 – Chapters 1.3, 5.1 (skim 5.2), 6.1
  - Version 9 – Chapters 1.4, 5.1, 5.2 (skim 5.3), 6.1
- Strategic Data Project Human Capital Diagnostic – LAUSD
  - http://sdp.cepr.harvard.edu/files/cepr-sdp/files/sdp-lausd-hc.pdf
- Standardizing a Variable (from UCLA):
  - https://stats.idre.ucla.edu/stata/faq/how-do-i-standardize-variables-in-stata/

Recommended / Supplemental

- Page, L. C., Fullerton, J., Cohodes, S. R., West, M. R., Bacher-Hicks, A., Owens, A., & Glover, S. (2013). The Strategic Data Project's Strategic Performance Indicators. http://www.revproxy.brown.edu/login?url=http://www.mitpressjournals.org/doi/pdf/10.1162/EDFP_a_00105

Unit 3: Relationship Between two Variables

Required Reading/Videos/Stata Resources

- Review unit slides
- Moore, McCabe & Craig
  - Version 6/7/8 – Chapter 2 (especially 2.1-2.3), 6.2, 7.1, 7.2
  - Version 9 – Chapter 2 (especially 2.1-2.4), 6.2, 7.1, 7.2
- Interpreting T-test output in Stata (from UCLA)
  - https://stats.idre.ucla.edu/stata/output/t-test/
- *One Sample t-test* by StataCorp
  - http://www.youtube.com/watch?v=HwzCyqW-0dc
- *Two group means comparison* by StataCorp
  - http://www.youtube.com/watch?v=by4c3h3WXQc
- *Hypothesis Testing* by Khan Academy
  - https://www.khanacademy.org/math/probability/statistical-studies/hypothesis-test/v/simple-hypothesis-testing
- *What Is a p-value?* By JBstatistics
  - https://www.youtube.com/watch?v=HTZ8YKgD0MI
- *Hypothesis Testing and P-values* by Khan Academy
  - https://www.youtube.com/watch?v=-FtlH4svqx4
- *Pearson's Correlation Coefficient in Stata* by StataCorp
  - http://www.youtube.com/watch?v=o7ko844ff-g
- *Bar Graphs in Stata* by StataCorp:
  - http://www.youtube.com/watch?v=jNjAdtQwW6M
- *Introduction to Graphs in Stata (from UCLA)*
  - https://stats.idre.ucla.edu/stata/modules/graph8/intro/introduction-to-graphs-in-stata/
- *Visual Overview for Creating Graphs in Stata:*
  - http://www.stata.com/support/faqs/graphics/gph/stata-graphs/

Recommended / Supplemental

- Hedges, L. V., Greenwald, R., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of educational research*, *66*(3), 361-396. http://www.revproxy.brown.edu/login?url=http://www.jstor.org/stable/1170528
- Hanushek, E. A. (1996). A more complete picture of school resource policies. *Review of Educational Research*, *66*(3), 397-409. http://hanushek.stanford.edu/sites/default/files/publications/Hanushek%201996%20RevEdRes%2066(3).pdf

- Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). Interpreting research on school resources and student achievement: A rejoinder to Hanushek. *Review of Educational Research*, 66(3), 411-416. http://www.revproxy.brown.edu/login?url=http://www.jstor.org/stable/1170530

Unit 4: OLS Regression

Required Reading/Videos/Stata Resources

- Review unit slides
- Moore, McCabe & Craig
  - Version 6/7/8/9 – Chapter 10
- *An Introduction to Linear Regression* by "statisticsfun":
  - https://www.youtube.com/watch?v=zPG4NjIkCjc
- *Simple Linear Regression in Stata* by StataCorp:
  - https://www.youtube.com/watch?v=HafqFSB9x70
- *Assumptions of Linear Regression* by George Ingersoll:
  - https://www.youtube.com/watch?v=kxq2jX4qC28

Recommended / Supplemental

- *Regression Module* by Khan Academy*:*
  - https://www.khanacademy.org/math/probability/regression

Unit 5: Transformations to Achieve Linearity

Required Reading/Videos/Stata Resources

- Review unit slides
- Descriptive analysis in education: A guide for researchers
  - https://ies.ed.gov/ncee/pubs/20174023/pdf/20174023.pdf
- Moore, McCabe & Craig supplemental material: Transforming data, decision analysis. Available on course website.

Recommended / Supplemental

- *Simple Linear Regression: Transformations* by Jeremy Balka (jbstatistics) (7 mins)
  - https://www.youtube.com/watch?v=HIcqQhn3vSM
- *Log-Level Regression & Interpretation* by Economicurtis (6 mins)
  - https://www.youtube.com/watch?v=wXC2kViEGz8
- Peet, E. D., Fink, G., & Fawzi, W. (2015). Returns to education in developing countries: Evidence from the living standards and measurement study surveys. *Economics of Education Review*, *49*, 69-90.
  - http://www.revproxy.brown.edu/login?url=http://www.sciencedirect.com/science/article/pii/S0272775715001065
- Heckman, J. The Case for Investing in Disadvantaged Young Children
  - http://heckmanequation.org/content/resource/case-investing-disadvantaged-young-children

- Reininger, M. (2012). Hometown Disadvantage? It Depends on Where You're From Teachers' Location Preferences and the Implications for Staffing Schools. *Educational Evaluation and Policy Analysis*, *34*(2), 127-145. [NOTE: we have not learned logistic regression yet. Focus on the predictor variables and big picture]

Unit 6: Multiple Regression: The Basics

Required Reading/Videos/Stata Resources

- Review unit slides
- Moore, McCabe & Craig – Chapter 11.1
- Rockoff, J.E., Jacob, B.A., Kane, T.J., & Staiger, D.O. (2011). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy*, *6*(1), 43-74. http://www.revproxy.brown.edu/login?url=http://www.mitpressjournals.org/doi/pdf/10.1162/EDFP_a_00022
- *Statistics 101: Multiple Regression (Part 1), The Very Basics* by Brandon Foltz
    - https://www.youtube.com/watch?v=dQNpSa-bq4M

Unit 7: Multiple Regression: Statistical Control

Required Reading/Videos/Stata Resources

- Review unit slides
- Moore, McCabe & Craig – Chapter 11.2
- Papay, J.P., Murnane, R.J., & Willett, J.B. (2015). Income-based inequality in educational outcomes: Learning from state longitudinal data systems. http://www.revproxy.brown.edu/login?url=https://journals.sagepub.com/doi/10.3102/0162373715576364

Recommended / Supplemental

- Skiba, R. J., Poloni-Staudinger, L., Simmons, A. B., Renae Feggins-Azziz, L., & Chung, C. G. (2005). Unproven links: Can poverty explain ethnic disproportionality in special education?. *The Journal of Special Education*, *39*(3), 130-144. **[Focus on correlations and OLS results]**
    - http://files.eric.ed.gov/fulltext/EJ722287.pdf
- *Partial Correlation Practice Problem* by Statistics (PSY 210 and ECON 261 at Nevada State) (7 mins)
    - https://www.youtube.com/watch?v=8i0h98chSHU
- *Multicollinearity* by Ben Lambert (5 mins)
    - https://www.youtube.com/watch?v=O4jDva9B3fw

Unit 8: Dichotomous Predictors

Required Reading/Videos/Stata Resources

- Review unit slides

- Fryer Jr, R. G., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *The Review of Economics and Statistics*, *86*(2), 447-464.
    - http://www.jstor.org.revproxy.brown.edu/stable/3211640

Recommended / Supplemental

- *Dummy Variables* by INCAE Business School
    - https://www.youtube.com/watch?v=9yTui_LoSOc
- *Introduction to Factor Variables in Stata, part 1: The basics* by StataCorp
    - https://www.youtube.com/watch?v=Wa1Nd9epHmY

Unit 9: Statistical Interactions

Required Reading/Videos/Stata Resources

- Review unit slides
- Kraft, M.A. & Papay, J.P. (2014). Do supportive professional environments promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*. *36*(4), 476-500. http://journals.sagepub.com.revproxy.brown.edu/doi/abs/10.3102/0162373713519496
- *Main Effects and Interactions* by Jim Grange
    - https://www.youtube.com/watch?v=OE46w0RqmQA

Recommended / Supplemental

- *Introduction to Factor Variables in Stata, Part 2: Interactions* by StataCorp
    - https://www.youtube.com/watch?v=f-tLLX8v11c

Unit 10: Special Topics

Required Reading

- Review unit slides
- Moore, McCabe & Craig
    - Versions 6/7/8/9 Chapter 14
- Ladd, H. F. (2011). Teachers' Perceptions of Their Working Conditions How Predictive of Planned and Actual Teacher Movement? *Educational Evaluation and Policy Analysis*, *33*(2), 235-261.
    - http://www.revproxy.brown.edu/login?url=http://epa.sagepub.com/content/33/2/235.full.pdf+html
- Johnson, S.M., Kraft, M.A., & Papay, J.P. (2012). How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement. *Teachers College Record*, *114* (10), 1-39. [Canvas]

Recommended / Supplemental

- *Logistic Regression – Introduction* by INCAE Business School
    - https://www.youtube.com/watch?v=gNhogKJ_q7U
- *Binary Logistic Regression* by Methodology LSE

- https://www.youtube.com/watch?v=0C_Hlh_jNq8

Unit 11: Regression modeling in Practice

Required Reading

- Review all unit slides