

ED 795 - CAUSAL INFERENCE WITH QUANTITATIVE DATA IN EDUCATION

Spring 2020

Instructor:	Steve Porter	Time:	Tues 16:30 – 19:15
Email:	srporter@ncsu.edu	Place:	Online
Office hours:	By appointment		

Office Hours: By appointment.

Objectives: The main issue we face when studying human behavior is that it is difficult to randomly assign individuals to control and treatment conditions for many program and policy issues (e.g., being held back a grade in elementary school, or obtaining a college degree). The goal of this course is to supplement your statistical training with additional approaches used in education to estimate causal effects on observational, versus experimental, data.

By observational, I mean any kind of data where units (students and their families, teachers, principals, schools, etc.) choose to engage in behaviors, programs or policies, rather than being randomly assigned by researchers. As should be obvious, these kinds of data are prevalent throughout education, and you will likely analyze such data in your dissertation. Although many educational researchers are loath to admit it, in general OLS, HLM and other versions of the general linear model can't be used to analyze observational data in any meaningful way, other than to generate reams of correlational results.

Learning objectives:

- Understand the potential outcomes approach to causality
- Understand and implement the following techniques to estimate causal effects with observational data in education:
 - Difference-in-difference and fixed effects models (panel data)
 - Propensity score analysis (also known as matching)
- Estimate and interpret logistic regression models
- Have a much deeper understanding of issues around research design and internal validity when analyzing observational data in education

Prerequisites: To succeed in this course you must have taken ED 711 or a similar semester-long course on multiple regression, as well as having received a grade of A- or above. If this does not apply to you, we should talk immediately.

Readings: There are no assigned texts. Each week we will read a selection of articles and book chapters from education that illustrate the methods for that week. I have also recorded lectures for each topic that you will view before class.

Software: We will be using Stata version 15 or 16 for this course. Please let me know if you have an earlier version of Stata. Stata may be available for the semester for doctoral students in the College of Education, please contact me if interested. Stata is available for purchase at a discounted rate for graduate students (<https://www.stata.com/order/new/edu/gradplans/student-pricing/>).

I recommend purchasing either IC or SE. SE allows the use of larger datasets, where “large” is defined by the number of variables, not observations. If you think you might end up using a national dataset, go with SE.

Tentative Course Outline:

Introduction; why we don’t want to use multiple regression	Week 1
Causality	Week 2
Difference-in-difference models	Weeks 3-4
Difference-in-difference with unit fixed effects	Weeks 5
Fixed effects models	Weeks 6-8
Introduction to logistic regression	Week 9
Propensity score analysis	Week 10-12
Synthetic control method	Week 13-14
TBD	Week 15

Evaluation: Your grade will be determined as follows:

1. Participation (20%) – Participation in class discussions is vital for this class, and part of your grade will be the frequency and quality of your contributions to our discussions. You should come to class prepared to discuss in detail the readings. I will also hand out non-credit assignments for you to work on your own time, and we will discuss your findings in class. Your completion of these assignments and discussion is also part of your participation grade.
2. Take-home exams (80%) – You will have three or four assignments during the semester that cover various topics in the course; you will have around one week to complete each assignment. These assignments will generally involve analyzing datasets and reporting results from your analyses, as well as reading and critiquing a journal article using the topic we are studying. You may use your notes, Moodle and assigned readings to do the assignments, but you may not consult with any other people or sources (e.g., the Internet). I refer to these short papers as exams to stress that you should complete these assignments completely on your own, without any computer or statistical assistance from anyone else (other than me, of course). Late assignments will not be accepted, because I review the assignment answers during class on the day the assignment is due.

Final grades will be assigned using this scale:

$97 \leq A+ \leq 100$	$73 \leq C < 77$
$93 \leq A < 97$	$70 \leq C- < 73$
$90 \leq A- < 93$	$67 \leq D+ < 70$
$87 \leq B+ < 90$	$63 \leq D < 67$
$83 \leq B < 87$	$60 \leq D- < 63$
$80 \leq B- < 83$	$0 \leq F < 60$
$77 \leq C+ < 80$	

Important Dates:

Last class	April 16th
Last take-home exam due	April 30th (by 5:00 PM)

Disabilities: Reasonable accommodations will be made for students with verifiable disabilities. In order to take advantage of available accommodations, students must register with the Disability Resource Office at Holmes Hall, Suite 304, Campus Box 7509, 919-515-7653. For more information on NC State’s policy

on working with students with disabilities, please see the Academic Accommodations for Students with Disabilities Regulation (REG02.20.01) (<https://policies.ncsu.edu/regulation/reg-02-20-01/>).

Academic Integrity: Students are required to comply with the university policy on academic integrity found in the Code of Student Conduct found at <http://policies.ncsu.edu/policy/pol-11-35-01>. Your submission of any test or assignment indicates “I have neither given nor received unauthorized aid on this test or assignment.”

Please note this syllabus is subject to change; any changes will be discussed in class.