I. Course description

This course introduces the application of basic statistical methods to economic analyses. Econometrics uses statistical methods to support or reject hypotheses from economics using empirical observation. It covers the basic concepts behind linear regression models by studying cases where the dependent variable is continuous and is a linear function of the parameters of interest. The aim is to improve students’ ability to conduct economic analysis using observational data, as economic studies rarely benefit from the availability of controlled experiments. The course will rely on exercises designed to provide hands on experience in implementing well-crafted empirical analysis. The emphasis of the course is on applied skills. We focus on correct and informative analysis rather than sophisticated and obscure methods. By the end of the course, students will be able to employ tools and methods and compare the results with respect to those obtained from initial estimations based on very restricted assumptions.

Module I— Basic principles for good applied work

In this module, we will first review the guiding principles required to conduct good applied research using observational data. We then move to reviewing the five main
assumptions behind multiple regression and its limitations. We will cover in detail the
topic of omitted variable bias and how to deal with this limitation in empirical work. We
will review fundamental model specifications used in the literature. Following this, we
will cover the use and interpretation of dummy variables as covariates of interest in research.

Module II— Using multiple regression to answer questions of interest

This section moves from the basic multiple regression analysis and introduces extensions
to the model. We will then study the use of OLS techniques to understand issues of
inequality, sources of it, and the socio-economic gradient in outcomes.

Module III— Expanding your toolkit to conduct good analysis

In the third module we expand the use of OLS and ask ourselves, what type of analysis
could we use when one has multiple cross sectional databases or panel databases? How
can one differentiate between changes within individual versus social changes over time?
We will learn how to estimate first-differences models, growth models and sibling
models. Then, we will study changing effect models, convergence models and cohort
replacement effects using multiple cross sectional databases. We will end the module by
covering linear decomposition methods. At the closing of the term, we will raise caution
about how far one can get using the proposed methods. OLS is powerful but it is a good
practice to understand its limits.

II. Learning objectives

1. Apply methodological principles and statistical concepts as they relate to the
field of health economics.

2. Conduct linear regression analyses of observational data to reach conclusions
relevant for decision-making processes in both national and international
settings.

3. Use the Stata computer software package to conduct applied empirical
research.

III. Reading material

This course will be mainly based on class notes. Different texts and papers will be
used during the development of the course. The following seven books are
recommended:

to effect. Princeton University Press.


Additional readings will be required during the term. The professor will provide online copies of any additional material in advance.

**IV. Format**

The course includes class lectures and lab activities. Every week, we will have one day of lecture, and one day of lab. The second part of each lecture will follow a seminar format. In this section, we will discuss a published paper that uses the technique covered in class. During the lab section, we will review the main steps to complete the assignment. Students are expected to complete all reading assignments before class as they may be called upon to review the material. Students who do not complete the reading will be unable to participate in class discussions or activities.

The course evaluation is based on 5 assignments, and a final exam.

**V. Determination of course grade**

Each student's grade will be calculated based on the following criteria:

**I. Weekly empirical assignments**

- *Assignment 1*: How to use and interpret dummy variables 12%
- *Assignment 2*: Using different model specifications 12%
- *Assignment 3*: Oaxaca Blinder decomposition approach 12%
- *Assignment 4*: Measuring individual changes 12%
- *Assignment 5*: Linking individual and social change 12%
II. Final Exam 35%

III. Slack participation 5%

(a) Weekly assignments

Five problem sets will be completed during the semester. The assignments are intended to build on the lectures of the module and illustrate key principles using applied examples. Each student must hand in his or her own problem set via Drop Box on the Courseplus webpage. Problem sets will not be accepted after deadlines.

(b) Final exam

The final exam will be closed notes and closed book. Students are encouraged to work together to answer the review questions that will be handed out one week before the exam and to learn from one another. During the exams, however, all work is to be done individually.

The criteria for the final grading of the course will be:

\[
\begin{align*}
A &= 90-100 \\
B &= 80-89 \\
C &= 70-79 \\
D &= 60-69 \\
F &= 0-59
\end{align*}
\]

There are no make-up examinations for unexcused absences. The instructor must be notified personally in case of an absence. Leaving a voice-mail or e-mail will not be a valid notification. The final letter grade for this course will be based on total points and not on an average of letter grades. If the student cannot attend the class, the assignment could be turned in by e-mail or dropped in my mailbox.

VI. Use of Computer Software

We will use the statistical software STATA 14 (or any other version) to learn the main statistical techniques used during each module.

VII. Pre-requisites.

Recommended pre-requisites for this course are 140.623 or 140.653

VIII. General notes

(a) Academic dishonesty
Students will adhere to university academic ethics guidelines as defined in: http://www.jhsph.edu/schoolpolicies/policy_academic_ethics.html#aone

(b) Changes on the Syllabus

This syllabus is subject to change based on the evolution of the class during the term.

(c) Class attendance

Class attendance is expected and sign-in will be conducted at the beginning of each class. The following criteria will apply to evaluate class attendance:

If a student is in doubt as to whether he/she has been given credit for attending a specific class session, it is the responsibility of that student to consult the instructor during or immediately after the class session. Changes in attendance will not be permitted after the date of a scheduled class session.

(d) Disability Access Statement

All reasonable efforts will be made to accommodate students. Please see disability coordinator for special needs.

VIII. Schedule of course meetings and assignments

MODULE I: Basic principles for good applied work

Learning objectives

- To provide a framework with the guiding principles to conduct good applied research using observational data
- To review the scope and limitations of causal inference using observational data.
- To introduce the classical linear regression model, its main assumptions, and the implications of violating the main assumptions.
- To discuss the issue of omitted variable problem in empirical work.
- To gain knowledge about the use and interpretation of dummy variables.

WEEK 1: Guiding principles to conduct good applied research

Session 1
Date: 01/21/2020. Lecture and seminar session
Topic: Guiding rules for good applied research.
Required Reading: [F] Chapters 1-2; [W] Chapter 19
Assignment: None
Session 2
Date: 01/23/2020. Lab session
Topic: More on guiding rules for good applied research.
Required Reading: [F] Chapters 1-2; [W] Chapter 19
Assignment: None

Additional reading:

WEEK 2: Fundamental assumptions behinds OLS and implications of its violations

Session 3
Date: 01/28/2020. Lecture and seminar session
Topic: Multiple regression and main assumptions
Required Reading: [G] Chapters 1-8; [K] 1-3; [W] 1-6; [F] Chapters 3-4
Assignment: None

Additional reading:

Session 4
Date: 01/30/2020. Lab session.
Topic: More on multiple regression
Required Reading: [G] Chapters 1-8; [K] 1-3; [W] 1-6
Assignment: None

Required paper:


WEEK 3: How to use and interpret dummy variables

Session 5
Date: 02/04/2020. Lecture and seminar session
Topic: Single dummy variable, interaction of dummies, splines and OLS weighted regression.

Required Reading: [G] Chapter 9; [K] Chapter 14; [W] 7

Assignment: 
Assignment 1: How to use and interpret dummy variables
Due date: 2/11/2020

Additional reading:

Session 6
Date: 02/06/2020. Lab session.

Required Reading: [G] Chapter 9; [K] Chapter 14; [W] 7
Assignment: None

Required paper:

Module II: Using multiple regressions to answer questions of interest

Learning objectives

- To enhance the students’ understanding of using different model specifications.
- To use OLS to explore inequalities in health and income: The Oaxaca Blinder decomposition approach; the concentration index and other measures of inequality.
- To acquaint the student with models to measure education gradients in health.

WEEK 4: Using different model specifications

Session 7
Date: 02/11/2020. Lecture and seminar session
Topic: Four fundamental functional forms useful in applied research: y-x; logy-logx; y-logx; logy-x.

Required Reading: [G] Chapter 6

Assignment: Assignment 2: Using different model specifications
Due date: 2/18/2020

Session 8
Date: 02/13/2020. Lab session
Topic: More on the four fundamental specifications useful in applied research: y-x; logy-logx; y-logx; logy-x.

Required Reading: [G] Chapter 6
Assignment: None

Required paper:


WEEK 5: Measuring inequalities in health

Session 9
Date: 02/18/2020. Lecture and seminar session
Topic: The Oaxaca Blinder Decomposition approach, the Concentration Index and the Gradient effect.

Required Reading:

Assignment: Assignment 3: Oaxaca-Blinder Decomposition.
Due date 2/25/2020

Additional reading:

Session 10:
Date: 02/20/2020. Lab session
Topic: More on The Oaxaca Blinder Decomposition approach, the Concentration Index and the gradient effect.

Required Reading:
Assignment: None

Required paper:
Module III: Expanding your toolkit to conduct good analysis
Learning objectives

- To apply different methods for analyzing multiple cross sectional databases.
- To highlight the difference between changes within individuals and social change
- To learn how to use the decomposition methods to differentiate covariates effects

WEEK 6: Measuring individual’s changes

Session 11
Date: 02/25/2020. Lecture and seminar session
Topic: First-differences models; Growth models and Siblings models
Required Reading: [F] Chapter 5.
Assignment: Assignment 4: Measuring individual changes
Due date 03/03/2020

Additional reading:

Session 12
Date: 02/27/2020. Lab session
Topic: First-differences models; growth models and siblings models
Required Reading: [F] Chapter 5
Assignment: None
Required paper:


WEEK 7: Studying Social Changes

Session 13
Date: 03/03/2020. Lecture and seminar session
Topic: Changing effect models; Convergence Models and Estimating Cohort Replacement Effect.
Required Reading: [F] Chapter 6.
Assignment: Assignment 5: Linking individual and social changes
Due date: 3/10/2020

Additional reading:

Session 14
Date: 03/05/2020. Lab session
Required Reading: [F] Chapter 6.
Assignment: None

Required paper:


WEEK 8: Linear Decomposition Methods

Session 15
Date: 03/10/2020. Lecture and seminar session
Topic: Linear Decomposition Methods
Required Reading: [F] Chapter 7
Assignment:
Additional reading:

Session 16
Date: 03/12/2020.
Topic: Closing arguments; Final Exam
Required Reading: [F] Chapter 7.

FINAL EXAM 3/12/2020