

GRADUATE SCHOOL OF PUBLIC AND INTERNATIONAL AFFAIRS
UNIVERSITY OF PITTSBURGH

PIA 2007/3000: Intermediate Quantitative Methods: Analysis of Policy Experiments

SPRING 2011

Instructor: Prof. Sera Linardi
E-mail /Phone: linardi@pitt.edu / 412.648.7650
Lecture hours: Wednesday 3 pm to 6 pm
Lecture hall: 3911 WWPH
Office Hours: Tuesday 11:00-12:00pm and by appointment
Office /Telephone: WWPH 3203 412.648.7650

TA: Juan Nicolas Hernandez
Email: jnh28@pitt.edu
Recitation: Thursday 12-2pm
Recitation hall: TBA

Secretary: Susan Sawyers
E-mail: suzanc@gspia.pitt.edu
Office: WWPH 3201

This course is a continuation of PIA 2001. Our goals are to:

1. Effectively use statistical software to explore data, find relationships between variables, form hypotheses about causality, and test them.
2. Communicate the result of your analysis concisely (through reports and presentations)
3. Critically consume empirical research
4. Identify opportunities to integrate your statistical skills into future career paths.

This class will proceed at a rapid pace, so you must be comfortable with PIA 2001 concepts. If you are a little rusty, spend sometime during break with your preferred data analysis software.

The randomized testing of social policy is now a global phenomenon, influencing decision makers in international development, domestic social policy, and political/ corporate strategy. In this class you will gain exposure to natural and field experiments related to MIT's Poverty Action Lab (J-PAL), Housing and Urban Development (HUD), Department of Health and Human Services, and other institutions.

Software

For in-class exercises and lectures, we will use R, the leading statistical analysis **freeware**. (<http://cran.r-project.org/>) This will give you some exposure to the large pool of resources available in online statistical community -- you will no longer be dependent on any corporation (and software costs) to do statistics. You need to download and install R during break. Right from the first day of class we will be conducting in-class exercises using R. To get you started, here is the R reference card: <http://cran.r-project.org/doc/contrib/Short-refcard.pdf> to print and bring with you on the first day of class.

I am aware that many of you are more familiar with SPSS. The optional **M&V** textbook below is a step by step workbook for SPSS that may be useful for the first few weeks. In the recitation, Nicolas will help you if you have any questions about translating concepts from one software to another. Homeworks and exams can be done in any statistical software that you choose, but we will only provide help for SPSS and R. I will provide bulletin boards for each software group so users can find and support each other.

We will be downloading data and writing scripts in every class. Please always bring your laptop and check that it is connected to the network. Needless to say, please resist the siren call of internet browsing while class is in session.

Books:

Required:

(MHE) Angrist, Joshua and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. 1st ed. Princeton University Press, 2008.

(IA) Ayres, Ian. (2008) *Super Crunchers: Why Thinking-By-Numbers is the New Way To Be Smart*, Bantam Books.

Optional:

(M&V) Mertler, Craig A. and Rachel A. Vannatta. *Advanced and Multivariate Statistical Methods*. 4th ed. Los Angeles, CA.: Pyrczak Publishing, 2005

(DH) Huff, Darrell. *How to Lie with Statistics*. Penguin, 1991.

Activities and Grading:

Since the focus of the class is to develop creative statistical thinking and communication skills, you will have plenty of opportunities to conduct your own data analysis and debate research design. Most activities will be collaborative in nature.

We will have several short presentations (10 minutes presentation, 5 minute discussion). For data analysis, there will be one group presentation on the process of finding, cleaning, and exploring a data set (10%) and an individual presentation on how you have used your group's shared data set to answer your own research question (10%). Groups will also choose a randomized policy experiment to present and discuss (10%). Your data analysis group and policy experiment group need not be the same.

There will be two homeworks (12.5%) and three exams (15%). Homeworks can be discussed with anyone, but must be written up on your own. Exams are conducted in-class. The first 75 minutes will be a written exam focusing on applying concepts you have learned in class. You will then get a 15 minute break before the software portion. In the second part of the exam, you will be given a question and a data file. You will have 90 minutes to perform the analysis necessary to answer the question.

In short, the grade for this course will be determined from :

2 homeworks (12.5% each). Total: 25%

3 exams (15%) each: Total: 45%

3 presentations (10% each): Total 30 %

You will be given one freebie this semester, which means that you can turn in one homework set late by 5 days, no question asked. For the freebie: turn in the late homework by Monday noon to Susan Sawyer's office. Use this wisely and only for emergencies such as illness. The Monday noon deadline will be enforced strictly since I plan to return graded homeworks and post solutions on Monday afternoon. Apart from the freebie, late assignments will not be accepted. Graded freebies are returned a week later.

When you feel that mistakes have been made in determining the grade of your homeworks or exams, we are happy to regrade them. Here are the steps to take:

1. Please compare your answers to the posted solutions
2. Please submit a written request stating your reasons for a regrade. If there are specific questions/answers that you want to explain, please do so. Submit the requests to me (Sera) at class or office hours, or you can drop it off with the administrative assistant for the class (Susan Sawyers).
3. We will regrade the ENTIRE homeworks /exam (using the posted solutions as before). The new grade may be higher or lower than your original grade.

Abbreviated Schedule (detailed schedule with reading to follow).

Week 1 (1/12) Introduction (AY Intro, Ch 1)

Motivation. Data exploration with R. Descriptive statistics (averages, frequency, N) table. Missing values. Correlation. Graphing. Standards for tables and figures in reports (examples from papers).

Week 2 (1/19) Hypothesis Testing (MHE Ch 1)

Also: http://www.americanprogress.org/issues/2007/04/pdf/data_driven_policy_report.pdf

Brainstorm for policy issues and data sources for case study. Data screening: outliers, normality, heteroskedasticity. Transforming data. Hypothesis testing. T-test. Wilcoxon, Chi-squared, F.

HW I posted. Form groups for case study through courseweb bulletin board.

Week 3 (1/26) Randomization and Causal Inference (MHE Ch 2, AY Ch 2)

The selection problem. Regression analysis for experiments: OLS and logistic regressions. Discuss policy experiment paper 1.

HW I due.

Week 4 (2/2) Multiple Regression I (MHE 3)

Measures of fit, interpreting output.

Short Presentation (first 3 groups): policy issue, descriptive statistics of data source, research questions

Week 5 (2/9) Multiple Regression II (MHE 3)

Interaction terms, dummy variables, probit and logit.

Short Presentation (next 3-4 groups)

Week 6 (2/16) Exam I

Week 7 (2/23) Panel Data and Fixed Effects (MHE 5.1)

Controlling for unobservables through fixed effects. If time permits: Controlling for observables and matching (propensity scores).

HW 2 posted.

Week 8 (3/2) Instrumental Variables (MHE 4.1, 4.4.1-4.4.2)

2-stage least squares (2SLS). Civil conflict (Miguel et.al 2004), compulsory schooling (Angrist and Krueger 1991) and foster care (Doyle 2008)

HW 2 due.

Week 9 (3/9) Spring Break

Week 10 (3/16) Difference in Differences (MHE 5.2-5.3)

Housing vouchers (Jacob 2004), China's one child policy (Qian 2009), old age pension in South Africa (Duflo 2009)

Randomized Experiment Discussion (first 3 groups): causal questions, identification, and results

Week 11 (3/23) Exam II

Week 12 (3/30) Standard Error Estimation (MHE 8.1-8.2)

Robust standard error, biases, clustering, and serial correlation.

Randomized Experiment Discussion (next 3-4 groups)

Week 13 (4/6) Lies, Damned Lies, and Statistics (DH, AY Ch.4,5, 7)

Discuss misuse of statistics.

Week 14 (4/13) Exam III

Week 15 (4/20) Individual Case Study Presentations I

Week 16 (4/27) Individual Case Study Presentations II