Rev 01/05/2012

METHODS IN HEALTH SERVICES RESEARCH

HPA 583b (Spring 2012)

Instructor: Doug McKee, 310-266-2438; douglas.mckee@yale.edu

**Instructor
Office Hours:** LEPH 315, TBD

Teaching Assistants: Nick DeVito (nicholas.devito@yale.edu)
Leo Quigley (leo.quigley@yale.edu)

TA Office Hours: TBD

Lecture Day/Time: Monday, 10:00-11:20am, LEPH 101
Lab Day/Time: Wednesday, 10:00-11:20am, LEPH 101
(Some labs will be in 47 College computer lab)

Course Description:
This is a course about the concepts, tools and approaches used to answer questions empirically. We will consider four types of questions. *Description*: what is the value (or distribution) of outcome variable Y? *Association*: does Y vary by explanatory variable X? *Evaluation*: what is the causal effect of X on Y? *Forecasting*: what will Y be in the future or under some alternate scenario?

Course Objectives:
The objectives of the course are to develop students' abilities to:
(1) Conduct surveys to collect original data,
(2) Perform basic data management and analysis tasks using SAS,
(3) Understand, estimate and interpret ordinary least squares and logistic regression models,
(4) Evaluate and conduct quantitative policy evaluations,
(5) Apply regression models to generate forecasts, and
(6) Conceptualize and execute empirical research projects.

Prerequisites:
One semester of biostatistics or permission of the instructor.

Readings:

You can get away without purchasing any books for this class, but you will learn a lot more if you work through the Allison book and the Angrist and Pischke book as we cover individual topics. These will both also be excellent references if you see yourself doing any data analysis at all in the future. And if you're going to continue writing SAS programs, I highly recommend the two SAS books. The Ayres book just shows how data analysis can be fun—it's great casual reading. The articles assigned are either overviews of methods or good examples of methods in action. All the books will be on reserve at the medical library and the articles are posted to classes*v2.

- Allison PD. *Multiple Regression: A Primer*. Thousand Oaks, CA: Pine Forge Press, 1999.
- Angrist JD, Pischke JS. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 2009.

- Delwiche, LD, Slaughter, SJ. *The Little SAS Book: A Primer, Fourth Edition*. Cary, NC: SAS Institute Inc., 2008.
- Cody RP, Pass R. *SAS Programming by Example*. Cary, NC: SAS Institute Press, 1995.

- Ayres I. *Super Crunchers: Why Thinking-by-Numbers Is the New Way to Be Smart*. New York, NY: Bantam, 2007.

- articles (available online through classes*v2)

Assignments:

Lab assignments. There will be 6 lab homework assignments involving SAS programming and application of the material taught. So the lab answers can be posted in a timely fashion, **no labs will be accepted for credit after the due date**. Labs will each be graded on a 0 to 2 point basis, and all labs should be turned in via the Classes V2 website.

In-class quizzes. **The first quiz will be given on February 15 and the second on April 4**. The quizzes will cover material presented in lecture and lab.

Group research project. In randomly-assigned groups of 4 or 5, students will conduct a multi-step research project that consists of: (1) developing a health-related (broadly considered) research question, (2) developing a brief survey instrument to collect data to answer that question, (3) cleaning the collected survey data and creating analytic variables, (4) analyzing the data (using SAS), (5) presenting the project in class, and (6) writing up a research report that describes the question, analysis and results. Part of the 1/11 lab will be devoted to a discussion of this assignment and its many deadlines/milestones.

Final individual research proposal. Students will be given a request for proposals (RFP) based on a ripped-from-the-headlines health care topic, and each student will develop and write up

his/her own research proposal. Students will turn in a write-up that includes a brief background, details on the data they would use (including how they would collect them), a description of the analytic approach, and a discussion of how their analysis addresses the issue and its strengths and limitations. Note that students need not actually execute their analyses. Part of the 4/9 lecture will be devoted to a discussion of this assignment. **Write-ups are due by 5pm on Friday, 5/4.**

Requirements and Grading:

Lab assignments	10%
In-class quizzes	30%
Group research project	35%
Individual research proposal	25%

Honor Code: All students are expected to comply with the Yale School of Public Health Code of Academic and Professional Integrity.

Laptop Use Policy: Laptops (and other computing devices) may be used only for class-related purposes, and laptop users are encouraged to sit near the back of the room to minimize class disruption.

LECTURE AND LAB SCHEDULE:

Mon 1/9: Course introduction and overview
Wed 1/11: Group project overview & SAS fundamentals

NO CLASS MONDAY 1/16 (Martin Luther King day)
Wed 1/18: Data checking and creating variables (SAS)

Mon 1/23: Survey design
Wed 1/25: Combining datasets & χ^2 test (SAS)
  * HW #1 due

Mon 1/30: Model interpretation for linear regression
Wed 2/1: Linear Regression (SAS)
  * HW #2 due

Mon 2/6: Model interpretation for logistic regression
Wed 2/8: Logistic Regression (SAS)
  * HW #3 due

Mon 2/13: Causality and evaluation
Wed 2/15: QUIZ #1 (Covers Survey Design through Logistic Regression)
  * HW #4 due

Mon 2/20: Model building and selection
Wed 2/22: Model building and selection (SAS)
    * HW #5 due

Mon 2/27: Difference-in-differences
Wed 2/29: Project Help Session

SPRING BREAK (3/5 – 3/16)

NO CLASS MONDAY 3/19
Wed 3/21: Linear and Logistic Regression in Excel (Optional) (47 College Computer Lab)
Fri 3/23: Instrumental Variables (3/23) (FRIDAY 10-11:20 in LEPH 101)

Mon 3/26: Model-based prediction and policy simulation
Wed 3/28: Difference-in-differences and IV (SAS)

Mon 4/2: Analyses for group projects
Wed 4/4: QUIZ #2 (Covers Causality & Evaluation to Model-based Prediction & Simulation)
    * HW #6 due

Mon 4/9: Discussion of individual research proposal assignment
Wed 4/11: Project help session (47 College Computer Lab)

Mon 4/16: First set of student presentations of group research project
    * Final group research project write-up is due in class
Wed 4/18: Second set of student presentations of group research projects

Fri 5/4: Final individual research proposal write-up is due by 5pm

READING:

1.   **Course introduction and overview (1/9)**

Eisenberg JM. Health services research in a market-oriented health care system. *Health Affairs*, 17(1):98-108, 1998.

Angrist JD, Pischke JS. Chapters 1-2 (Questions About Questions and the Experimental Ideal) in *Mostly Harmless Econometrics*. 2009.

2.   **Survey design (1/23)**

Burns KEA, Duffett M, Kho ME, et al. A guide for the design and conduct of self-administered surveys of clinicians. *Canadian Medical Association Journal*, 179(3):245-252, 2008.

SurveyMonkey. *Smart Survey Design*. Downloaded from: http://s3.amazonaws.com/SurveyMonkeyFiles/SmartSurvey.pdf.

Epstein AJ, Rathore SS, Alexander GC, Ketcham JD. Primary care physicians' views of Medicare Part D. *American Journal of Managed Care*, 14(11):SP5-SP13, 2008.

3.   **Model interpretation for linear regression (1/30)**

Kennedy P. Chapter 3. The classical linear regression model. In: *A Guide to Econometrics, 5th ed*. Cambridge, MA: The MIT Press, 2003. (**pp. 48-50 only**)

Allison PD. Chapter 1. What is multiple regression? In: *Multiple Regression: A Primer*. Thousand Oaks, CA: Pine Forge, 1999.

Allison PD. Chapter 2. How do I interpret multiple regression results? In: *Multiple Regression: A Primer*. Thousand Oaks, CA: Pine Forge, 1999.

Reference
Allison PD. Chapter 5. How does bivariate regression work? In: *Multiple Regression: A Primer*. Thousand Oaks, CA: Pine Forge, 1999.

Allison PD. Chapter 6. What are the assumptions of multiple regression? In: *Multiple Regression: A Primer*. Thousand Oaks, CA: Pine Forge, 1999.

Angrist JD, Pischke JS. Chapter 3 (Making Regression Make Sense) in *Mostly Harmless Econometrics, 2009.*

4.   **Model interpretation for logistic regression (2/6)**

Allison PD. Chapter 2. Binary logit analysis: basics. In: *Logistic Regression Using the SAS System: Theory and Application*. Cary, NC: SAS Institute, 1999. (**pp. 5-21, 28-30**)

Rubelcava, LN et al. The Healthy Migrant Effect: New Findings from the Mexican Family Life Survey. *American Journal of Public Health*, 98(1):78-84, 2008.

Reference
Allison PD. Chapter 3. Binary logit analysis: details and options. In: *Logistic Regression Using the SAS System: Theory and Application*. Cary, NC: SAS Institute, 1999. (**pp. 51-58**)

5. **Causality and evaluation (2/13)**

Taubes G. Do we really know what makes us healthy? *New York Times*, September 16, 2007.

Parker SW, Teruel GM. Randomization and Social Program Evaluation: The Case of Progresa. *Annals of the American Academy of Political and Social Science*, 599:199-219, 2005.

Veloski J, Tai S, Evans AS, et al. Clinical vignette-based surveys: a tool for assessing physician practice variation. *American Journal of Medical Quality*, 20(3):151-157, 2005.

6. **Model building and selection (2/20)**

Allison PD. Chapter 3. What can go wrong with multiple regression? In: *Multiple Regression: A Primer*. Thousand Oaks, CA: Pine Forge, 1999.

Allison PD. Chapter 7. What can be done about multicollinearity? In: *Multiple Regression: A Primer*. Thousand Oaks, CA: Pine Forge, 1999.

Allison PD. Chapter 8. How can multiple regression handle nonlinear relationships? In: *Multiple Regression: A Primer*. Thousand Oaks, CA: Pine Forge, 1999.

7. **Difference-in-differences (2/27)**

Dowd B, Town R Does x really cause y? Changes in Health Care Financing and Organization brief, 2002. (**pp. 1-10**)

Reference
Angrist JD, Pischke JS. Chapter 5 (Parallel Worlds: Fixed Effects, Differences-in-Differences, and Panel Data) in *Mostly Harmless Econometrics*, 2009.

Malluccio JA et al. The Impact of Improving Nutrition During Early Childhood on Education Among Guatemalan Adults. *The Economic Journal*, 119(April):734-763, 2009.

Meyer BD. Natural and quasi-experiments in economics. *Journal of Business and Economic Statistics*, 13(2):151-161, 1995.

8. **Instrumental Variables (3/23)**

Hildebrandt N, McKenzie DJ. The Effects of Migration on Child Health in Mexico. *Economía*, 6(1):257-289, 2005.

<u>Reference</u>
Angrist JD, Pischke JS. Chapter 3 (Instrumental Variables in Action: Sometimes You Get What You Need) in *Mostly Harmless Econometrics*, 2009.

SPRING BREAK

9. **Model-based prediction and policy simulation (3/26)**

Helft M. Google uses searches to track flu's spread. *New York Times*. November 12, 2008.

<u>Reference</u>
McKee DM, Todd PE. The Longer-Term Effects of Human Capital Enrichment Programs on Poverty and Inequality: Oportunidades in Mexico. *Estudios de Economía*, forthcoming.

Additional reading for the course

Baker S. Math will rock your world. *BusinessWeek*, January 23, 2006

Motheral B, Brooks J, Clark MA, et al. A checklist for retrospective data studies—report of the ISPOR Task Force on retrospective databases. *Value in Health*, 6(2):90-97, 2003.

SurveyMonkey. *SurveyMonkey User Manual*. Downloaded from: http://s3.amazonaws.com/SurveyMonkeyFiles/UserManual.pdf.