Group Independent Study in Epidemiologic Analysis

UC Berkeley School of Public Health
Syllabus for PH 298 (22) CC# 76673
Spring 2005, Modified 2005-02-14
http://www.idready.org/

Instructor

Tomás Aragón, MD, DrPH
Director & Medical Epidemiologist
UC Berkeley Center for Infectious Disease Preparedness
Tel: 510-643-4935; Fax: 510-643-4926
Email: aragon@berkeley.edu

Course Description

Dr. Tomás Aragón will supervise a group independent study of Dr. Steve Selvin's textbook *Epidemiologic Analysis: A Case-Oriented Approach*. Dr. Selvin is on sabbatical this year and will not be teaching this well-received and popular spring course based on this textbook. (His course assumes familiarity with the S language [R/S-Plus] and follows his fall course based on his textbook Modern *Applied Biostatistical Methods: Using S-Plus*). Dr. Selvin will be teaching this course next spring 2006. However, epidemiology students wishing to cover this material this spring 2005, can do so as a group independent study with Dr. Aragón. We will meet weekly to discuss each chapter and analytic approaches using the S language. Students not familiar with the S language can use this as an opportunity to learn and implement the S language for analysis of epidemiologic data. The course will meet on **Thursdays, 9am-11am, at the Center for Infectious Disease Preparedness at 1918 University Avenue, 4th Floor**.

Target audience: This course is intended for students wishing to cover the material presented in Dr. Selvin's book in a independent study group format.

Course prerequisite: Completion of introductory epidemiology and statistics courses. Familiarity with the S language (R/S-Plus) is very helpful, but not required.

Course objectives

Upon completion of this course, participants will be able to:

- Implement statistical methods for epidemiologic data from Selvin 2001.
- Familiarize themselves with the S language (or other statistical package) to implement statistical principles
- Lead or co-lead discussions from chapters from Selvin 2001.

Course format: Weekly group discussions

Course enrollment and fee: N/A

Course credit and grading: Units: 2-3; Grading: P/NP

Course location and schedule: Thursdays, 9:00am-11:00am, 1918 University Avenue, 4th Floor.

Course books

1. Steve Selvin (2001). *Epidemiologic Analysis: A Case-Oriented Approach*. Oxford University Press; ISBN: 0195144899
2. R Developers (2004). *An Introduction to R*, available at http://cran.r-project.org/doc/manuals/R-intro.pdf.

Course requirements and evaluation

1. Class attendance and participation
2. Lead or co-lead 2 or more class discussions
3. Complete 5 short analytic assignments (see below)

Course schedule outline

Date	***Wk***	***Topics***	***Facilitator***
01/18/05	1	Measurement of trend [R code (text)] | [R notes (PDF)]	Tomás Aragón (aragon@berkeley.edu)
01/27/05	2	Odds ratio and relative risk [R code (text)] | [R notes (PDF)] [R code (text): factors coded correctly]	Tomás Aragón (aragon@berkeley.edu)
02/03/05	3	Group discussion	Group (Tomás out of town)
02/10/05	4	Group discussion	Group (Tomás out of town)
02/17/05	5	Randomized trial [R code (text)]	Charlotte Nussbaum (cshalini@berkeley.edu) Sunaina Dowray (sdowray@berkeley.edu)
02/24/05	6	Goodness of fit [R code (text)]	Erin Bray (ehbray@yahoo.com) Wayne van Gemert (wvg@berkeley.edu)
03/03/05	7	open	TBA
03/10/05	8	Group discussion	Group (Tomás out of town)
03/17/05	9	Cluster analysis	Erin Bray (ehbray@yahoo.com) Wayne van Gemert (wvg@berkeley.edu)
03/22/05	--	SPRING RECESS	TBA
03/31/05	10	Matched analysis (one case and two controls)	Erin Bray (ehbray@yahoo.com)
04/07/05	11	Linear logistic regression	Charlotte Nussbaum (cshalini@berkeley.edu)

Date	Wk	Topics	Facilitator
04/14/05	12	Non-parametric regression analysis	Erin Bray (ehbray@yahoo.com)
04/21/05	13	Poisson regression	Nitika Pai (nitika@berkeley.edu)
04/28/05	14	open	TBA
05/05/05	15	open	TBA

Student assignments

There are 5 short analytic assignments at http://socrates.berkeley.edu/~biostat/Courses/Spring/Ph29632/HW/index.html.

Each assignment requires the following brief sections (one paragraph): Introduction, Methods, Results, and Discussion.

Grading assignment: Passing this course requires completion of this course requires (1) leading or co-leading 2 or more study group discussions, and (2) completing at least 3 analytic assignments.

Software and data sets

S language

This course will emphasize and cover use of the S language (R/S-Plus) to implement the statistical methods from Selvin 2001; however, students are welcome to conduct the analyses in their own favorite package. The important goal is to understand the statistical principles and methods, so whatever statistical package gets you there more efficiently is fine.

R is available for free download at Windows, Mac OS, and Linux at http://www.r-project.org.

S-Plus is available for purchase at http://www.insightful.com.

If you are not sure which to get, I highly recommend R (its free, powerful, and has plenty of free documentation).

Data sets

Data sets for each chapter are at http://socrates.berkeley.edu/~biostat/Courses/Spring/Ph29632/Data/index.html.

I have also put the data sets at http://www.medepi.net/selvin/.

Epidemiology tools ('epitools') and R help listserv

Some basic epidemiology tools ('epitools') written in R are available at http://www.epitools.net. At this site there is a link to join CIDP's R help group. We will use this help group to pose and answer questions as a group.

Text editors

A good text editor will make your programming and data processing easier and more efficient. A text editor is a program for, you guessed it, editing text! The functionality I look for in a text editor are the following: (1) toggle between wrapped and unwrapped text; (2) block cutting and pasting (also called column editing); (3) Easy macro programming; (4) search and replace using regular expressions; and (5) ability to import large datasets for editing

If you do not want to install a text editing program then just use the default text editor that comes with your computer operating system (for Windows use Notepad). A very user-friendly text editor designed for R for Windows is called Tinn-R and is freely available at http://www.sciviews.org/Tinn-R/ (installation instructions on this page).

My favorite text editor for R is X-Emacs (http://www.xemacs.org). This program is more powerful but has a higher learning curve. For Windows, just download and run this executable: http://www.xemacs.org/Download/win32/setup.exe. For "Installation method" select "Install from the Internet." For "Local Package Directory" it finds where you downloaded 'setup.exe'. For "Installation type" select "Native." For "Installation method" select "Direct Connection." For "Select Download Site" select "ftp://ftp.us.xemacs.org." Just allow default installation; it installs many useful packages. For free tutorials do a Google search. I maintain some shortcuts at http://www.medepi.net/xemacs/. Installing this package require a high-speed connection.

Recommended books

- Steve Selvin (2003). Biostatistics: How It Works. Prentice Hall; 1st edition; ISBN: 0130466166
- Steve Selvin (2004). Statistical Analysis of Epidemiologic Data (Monographs in Epidemiology and Biostatistics, V. 35). Oxford University Press; 3rd edition; ISBN: 0195172809
- Steve Selvin (1998). Modern Applied Biostatistical Methods: Using S-Plus. Oxford Press; 1st edition; ISBN: 0195120256

Supplementary readings

- Official R manuals at http://cran.r-project.org/manuals.html (FREE)
- Contributed R tutorials at http://cran.r-project.org/other-docs.html (FREE)
- Dalgaard P. Introductory Statistics with R. Springer Verlag; 1st edition (2002); ISBN: 0387954759
- Verzani J. Using R for Introductory Statistics. Chapman & Hall/CRC (November 29, 2004); ISBN: 1584884509
- Ripley BD, Venables WM. Modern Applied Statistics with S. Springer Verlag; 4th edition (2002); ISBN: 0387954570
- Maindonald J, Braun, J. Data Analysis and Graphics Using R. Cambridge University Press; 1st edition (2003); ISBN: 0521813360
- Venables WM, Ripley BD. S Programming. Springer Verlag; 1st edition (2000); ISBN: 0387989668