

## Advanced Demographic Analysis

### DEMOGRAPHY 211

TIME AND PLACE Tuesday and Thursday, 11am - 12:30 pm, 2232 Piedmont Ave., Rm. 100

WEB SITE <http://demog.berkeley.edu/~jrw/211>

TELEPHONE 510-642-9688

E-MAIL [jrw@demog.berkeley.edu](mailto:jrw@demog.berkeley.edu)

OFFICE HOURS Wednesday, 3-5 pm, 2232 Piedmont Ave., Rm. 210

### COURSE ORGANIZATION

A calendar for the course is given on the following pages. There will be weekly homework assignments – typically distributed on Tuesday of the week when the material is being covered, and then due the following Tuesday. The midterm falls in the ninth week of class, just before Spring break. This course will not meet on Thursday April 17 due to the PAA Annual Meeting. The final exam has been scheduled (tentatively) for Friday, May 16, from 9am until noon.

### READINGS

The course includes a collection of readings from various sources. All readings are available in hard copy in the Demography library and electronically through the course web page.

There are two main textbooks for the course:

- (1) Jay L. Devore, *Probability and Statistics for Engineering and the Sciences* (5<sup>th</sup> ed.), Duxbury Press, Pacific Grove, CA, 1999.
- (2) Daniel A. Powers and Yu Xie, *Statistical Methods for Categorical Data Analysis*, Academic Press, 2000.

I recommend that you purchase your own copies of these two books, as both are very useful references on statistical methods. In addition, we will be using both of the following reference manuals for the R and S-Plus statistical programming languages:

- (3) W.N. Venables, D.M. Smith, and the R Development Core Team. *An Introduction to R*. Version 2.5.1 (2007-06-27).
- (4) *S-Plus 5 for Unix Guide to Statistics*, Data Analysis Products Division, Mathsoft, Seattle, 1998.

These are available via links on the website for this course. I recommend that you print out only the parts of these manuals that you need for this course.

### GRADES

Course grades will be a weighted average of the midterm exam (25%), the final exam (25%), and all problem sets (50%).

### IMPORTANT DATES

Midterm Thursday, March 20, 2008 (in class)  
Final Friday, May 16, 2008, 9am – noon

## CALENDAR

Summary measures (January 22, 24, and 29)

## Topics:

percentiles, quantiles  
 mean, median, mode, trimmed mean  
 variance, standard deviation, coefficient of variation  
 interquartile range  
 outliers  
 correlation coefficient, rank correlation coefficient  
 rates, ratios, proportions  
 probabilities, odds

## Readings:

- (1) Devore, Chap. 1 "Introduction and descriptive statistics."
- (2) Regina Elandt-Johnson and Norman Johnson, *Survival Models and Data Analysis* (John Wiley & Sons, New York, 1980), Chaps. 1 "Survival data" and 2 "Measures of mortality and morbidity. Ratios, proportions, and means."

Graphical methods (January 31 and February 5)

## Topics:

bar graphs, histogram  
 boxplot, stem-and-leaf  
 scatter plot  
 smoothing  
 logarithmic scales

## Readings:

- (1) John M. Chambers et al., *Graphical Methods for Data Analysis* (Duxbury, Boston, 1983), Chaps. 2 "Portraying the distribution of a set of data" (pp. 9-32 only) and 3 "Comparing data distributions" (pp. 47-60 only).
- (2) Howard Wainer, "How to display data badly," *The American Statistician* 38(2):137-147.
- (3) Devore, Chap. 2 "Probability."

Probability (February 7, 12, and 14)

## Topics:

events, Venn diagrams  
 conditional probabilities, independence  
 random experiments, random variables  
 expected values, variances  
 probability distributions  
 linear combinations of random variables, law of large numbers  
 central limit theorem, normal approximation

## Readings:

- (1) Devore, Chaps. 3 "Discrete random variables and probability distributions" and 4 "Continuous random variables and distributions."

Sampling (February 14 and 19)

## Topics:

types of random samples (simple, stratified, cluster)  
 sample means, sample proportions  
 finite population correction  
 sample weights

## Readings:

- (1) Earl Babbie, *The Basics of Social Research* (Wadsworth, Belmont, CA, 1999), Chap. 8 “The logic of sampling.”
- (2) David Freedman et al., *Statistics* (2<sup>nd</sup> ed., W.W. Norton, New York, 1991), Chap. 22 “Measuring employment and unemployment.”

Estimation (February 21 and 26)

## Topics:

point estimates, confidence intervals  
 bias, efficiency, consistency  
 least squares, weighted least squares  
 maximum likelihood, Fisher information  
 robustness

## Readings:

- (1) Devore, Chaps. 6 “Point estimation” and 7 “Statistical intervals based on a single sample.”

Inference and uncertainty (February 28 and March 4)

## Topics:

hypothesis tests, statistical significance, statistical power  
 sensitivity, specificity  
 t test, chi-squared test, F test  
 Wald test, likelihood ratio test, score test  
 direct assessment, bootstrap, jackknife

## Readings:

- (1) Devore, Chaps. 8 “Tests of hypotheses based on a single sample” and 9 “Inferences based on two samples.”

Two-way tables (March 6 and 11)

## Topics:

additive models, multiplicative models  
 residuals, transformations, weights  
 sums of squares, variance explained  
 ANOVA, log-linear models, singular value decomposition

## Readings (optional):

- (1) Devore, Chaps. 10 “The analysis of variance” and 14 “The analysis of categorical data.”
- (2) Powers and Xie, Chap. 4 “Loglinear models for contingency tables.”

Regression (March 13 and 18, April 1 and 3)

## Topics:

simple regression, correlation  
 residuals, transformations, weights  
 continuous variables, categorical variables, dummy variables  
 multiple regression, design matrix  
 interpretation, interaction terms  
 variances, confidence intervals, hypothesis tests, bootstrap  
 goodness-of-fit, model selection

## Readings:

- (1) Devore, Chaps. 12 “Simple linear regression and correlation” and 13 “Nonlinear and multiple regression.”
- (2) *S-Plus Guide to Statistics*, Chap. 8 “Regression and smoothing for continuous response data.”
- (3) *An Introduction to R*, Chap. 11 “Statistical models in R” (pp. 52-58 only)
- (4) Harold V. Henderson and Paul F. Velleman, “Building multiple regression models interactively,” *Biometrics* 37(2):391-411.
- (5) Powers and Xie, Chap. 2 “Review of linear regression models.”

Midterm (March 20)Generalized linear models (April 8, 10, and 15)

## Topics:

logistic (or binomial) regression  
 probit regression  
 Poisson regression  
 multinomial regression

## Readings:

- (1) Powers and Xie, Chap. 3 “Logit and probit models for binary data.”
- (2) *S-Plus Guide to Statistics*, Chap. 10 “Generalizing the linear model.”
- (3) *An Introduction to R*, Chap. 11 “Statistical models in R” (pp. 58-61 only)
- (4) Steve Selvin, *Practical Biostatistical Methods* (Duxbury, Belmont, CA, 1995), Chaps. 10 “Logistic regression analysis” and 12 “Poisson regression analysis.”

Survival models (April 22, 24, and 29, and May 1)

## Topics:

notation  
 Kaplan-Meier, Fleming-Harrington  
 variances, confidence intervals  
 exponential, Gompertz, Makeham, Weibull  
 proportional hazards  
 Cox regression  
 accelerated failure time

Survival models (cont.)

## Readings:

- (1) Elisa T. Lee, *Statistical Methods for Survival Data Analysis* (John Wiley & Sons, New York, 1992), Chaps. 1 “Introduction,” 2 “Functions of survival time,” 3 “Examples of survival analysis,” 4 “Nonparametric methods of estimating survival functions,” and 6 “Some well-known survival distributions and their applications”
- (2) Powers and Xie, Chap. 5 “Statistical models for rates”
- (3) Selvin, Chap. 12 “Survival data analysis”
- (4) *S-Plus Guide to Statistics*, Chaps. 25 “Overview of survival analysis,” 26 “Estimating survival,” and 27 “The Cox proportional hazards model.”

Review and special topics (May 6 and 8)