Survey Sampling, Fall, 2006, Columbia University (31 Aug 2006)

lectures: W F 9–10:30am, 903 Social Work Bidg.
Prof. Andrew Gelman, 1016 Social Work Bldg. (851-2142) and 731 International Affairs Bldg. (854-7075), `gelman@stat.columbia.edu`, `www.stat.columbia.edu/∼gelman`
Feel free to email, call, or just stop by if you have any questions.

teaching assistant: Chia-Hui Huang, `ch2342@columbia.edu`, 1023 Social Work Bldg. (851-2158)

course description: Goals are to (1) learn how sample surveys are conducted and why these designs are used, (2) learn how to analyze survey results, and (3) be able to derive from first principles the standard results and their generalizations.

textbooks: Lohr, *Sampling: Design and Analysis* (available at Labyrinth Books at 536 W 112 St). We shall cover Chapters 1–8 and topics from the remaining chapters.

Groves et al., *Survey Sampling* (available at Labyrinth Books).

When discussing regression modeling and imputation, we will also use some chapters from *Data Analysis using Regression and Multilevel/Hierarchical Models* by Gelman and Hill.

software: We will use the open-source statistical package R; for information on downloading and setting it up, see this page: `www.stat.columbia.edu/∼gelman/bugsR/`. (You can ignore the stuff about the software package Bugs; just do what is needed to get R working.)

R has online help; also useful is the book *An R and S-Plus Companion to Applied Regression* by Fox.

In addition, Stata is a commercial statistical package that is particularly useful for analyzing surveys. Stata is on the campus Macs and PCs, and you can also purchase it inexpensively through Columbia's Academic Information Systems (AcIS). A good book for Stata help is *A Handbook of Statistical Analyses using Stata* by Rabe-Hesketh and Everitt.

coursework and grading: Homework assignments will be due every week or so. You may work with other students on the homework assignments, but you must write up and hand in your own solutions.

There will also be frequent quizzes and a final exam. The grade will be 50% homework, 25% quizzes, and 25% final exam.

What you should be able to do once the course is over:

- Design of sampling

  - Sampling from a list
  - Stratified sampling
  - Cluster sampling

- Sample size calculations

- Analysis of data from samples

    - Computing and displaying estimates and standard errors
    - Poststratification
    - Ratio and regression estimation
    - Linear and logistic regression

- Practical issues

    - Data collection
    - Nonresponse
    - Costs of surveys
    - Total survey error

Some examples we may discuss in class:

- Fish in a lake

- Postal employees and mail pieces

- Insurance claims

- New York City Social Indicators Study

- National Election Study and General Social Survey

- Commercial pre-election polls and exit polls

- Homeless women

- Child service providers

- Alcoholics Anonymous

- Radon in houses

- Rodents and roaches in apartments

- Internet hits

- Fragile Families study

- Social networks

- Australian schoolchildren

- Incentives for survey participation