

**Survey Data Reduction and Analysis**  
**PA 588 Syllabus**  
**Spring 2007**  
**(03-13-07)**

**Class Meeting:** Wednesdays, 3-5:50pm  
Room 101 Stevenson Hall

**Instructors:** Richard T. Campbell, Ph.D.  
Institute for Health Research and Policy  
1747 W. Roosevelt Road (West Campus)  
Chicago, IL 60608  
312-413-0480  
E-mail: [dcamp@uic.edu](mailto:dcamp@uic.edu)

Timothy P. Johnson, Ph.D.  
Survey Research Laboratory  
Room 629  
CUPPA Hall (East Campus)  
412 S. Peoria Street  
Office: 312-996-5310  
Residence: 630-469-8649 (between 7-9pm please)  
E-mail: [timj@uic.edu](mailto:timj@uic.edu)

**Office Hours:** By appointment.

**Required Texts:**

1. Lee and Forthofer (2005) *Analyzing Complex Survey Data, Second Edition*. Note: This text is available in the East Campus Bookstore. An earlier draft is available, with the author's permission, on the course Blackboard site.
2. Allison (2002) *Missing Data*. Thousand Oaks, CA: Sage Publications. Note: This text is also available in the East Campus Bookstore and an early version of this monograph is also available on the course Blackboard site.
3. Other readings as assigned.

## Course Overview and Objectives:

The purpose of this course is to provide an overview of current methods and issues in regards to the theory and logic of data editing, imputation, weighting, and special considerations in the analysis and reporting of data from complex sample surveys. Specific objectives include the following:

1. To develop an understanding of and appreciation of the complexities involved in the analysis of survey data.
2. To develop problem-solving skills relevant to the proper analysis of survey data.
3. To acquire experience using one or more of the available software packages that are appropriate for the analysis of complex survey data.

Because this course is concerned with current issues in survey analysis, the schedule and assignments may be revised as the course progresses and new material becomes available.

## Course Blackboard Web Site

Note that a Blackboard web site has been established for the course. Class assignments and background materials will be posted at this site on a periodic basis throughout the 8-week course. You should plan to visit the site every few days to check for relevant announcements and/or the availability of new materials.

## General Requirements and Grading:

1. Class attendance is *strongly* encouraged. Active involvement in class is also expected. If your final grade is borderline, my subjective assessment of your class participation may be used to make final adjustments.
2. It is essential that you do the assigned readings before they are discussed in class. Many of the writing assignments (see below) will be based on the reading assignments.
3. The computer software employed in many class examples is primarily SAS Stata, although some ~~paralle~~ examples will be available in SASStata. Both Stata and SAS are available on computers in the SPH and CUPPA computer labs and you should have free access to this software. Stata and SAS can also be purchased at discounted student rates. For SAS, contact [consult@uic.edu](mailto:consult@uic.edu). Stata is directly available from Stata Corporation at: [www.stata.com/info/order/new/edu/gradplans/gp3-order.html](http://www.stata.com/info/order/new/edu/gradplans/gp3-order.html).
4. Final grades will be based on four class assignments, class participation and the final exam.

**Examination:** There will be a final exam in this class. It goes without saying that cheating will not be tolerated and that all relevant University policies apply. Examinations will be open book.

## **Class Schedule and Reading Assignments:**

### Week 1 (3/14/07)

- **Topics:** Overview of issues in survey data analysis; Data reduction logic and methods
- **Readings:**
  - ICPSR (2005) *Guide to Social Science Data Preparation and Archiving* (link to document available on Blackboard site)

### Week 2 (3/21/07)

- **Topics:** Examples of complex survey data sets; Software for survey data analysis – SAS, STATA, SUDAAN, SPSS, BRRS
- **Readings:**
  - Lee & Forthofer, chapters 1-2.
  - Brogan (1998) Pitfalls of using standard statistical software packages for sample survey data. *Encyclopedia of Biostatistics*
  - Johnson & Elliott (1998) Sampling design effects: Do they affect the analyses of data from the National Survey of Families and Households? *Journal of Marriage and the Family* 60: 993-1001.
  - Korn and Graubard (1991) Epidemiologic studies utilizing surveys: Accounting for the sampling design. *American Journal of Public Health* 81: 1166-1173.
  - An & Watts. *New SAS Procedures for Analysis of Sample Survey Data*.

(3/28/07) – No class – Spring Break

### Week 3 (4/4/07)

- **Topics:** Review of sample designs; Basic issues in weighting – Purposes of weights, advantages and disadvantages; Complex topics in weighting – model-based inference
- **Readings:**
  - Lee & Forthofer, chapter 3 (on Blackboard)
  - Kalton & Flores-Cervantes (2003) Weighting methods. *Journal of Official Statistics* 19: 81-97 (on Blackboard)
  - Winshop & Radbill (1994) Sampling weights and regression analysis. *Sociological Methods & Research* 23: 230-257 (on Blackboard)
  - Little et al. (1997) Assessment of weighting methodology for the National Comorbidity Survey. *American Journal of Epidemiology* 146: 439-449

### Week 4 (4/11/07)

- **Topics:** Estimating design effects and variances
- **Readings:**
  - Lee & Forthofer, chapters 4-6
  - Kish and Frankel (1974) Inferences from complex surveys. *Journal of the Royal*

*Statistical Society* 36(B): 1-37 (on Blackboard)

- Krueger & Valliant. (2007) A survey on survey statistics: What is done and can be done in Stata. *The Stata Journal* 7: 1-21.
- Rao & Scott (1981) The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. *JASA* 76: 221-230.
- Xue et al. (2003) Building logistic regression models to analyze survey data with complex sampling design.
- An. Performing logistic regression on survey data with the new SURVEYLOGISTIC procedure.

#### Week 5 (4/18/07)

- Topics: Analysis of subsets of cases – Design effects must be considered when selecting subsets for analysis; Repeated cross sectional surveys – Trends of over time from surveys with constant frame; Special issues in longitudinal surveys; Combining multiple surveys
- Readings:
  - Kalton (1986) Handling wave nonresponse in panel surveys. *Journal of Official Statistics* 2: 303-314 (on Blackboard)

#### Week 6 (4/25/07)

- Topics: Introduction to missing data issues – Handling missing data properly; EM Estimation of Missing Data – Single imputation approaches
- Readings:
  - *Report of the Task Force on Imputation* (1996)
  - Kalton & Kasprzyk (1986) The treatment of missing survey data. *Survey Methodology* 12: 1-16.
  - Allison, *Missing Data*, chapters 104 (on Blackboard)

#### Week 7(5/2/07)

- Topics: Multiple Imputation – Why multiple imputation is important
- Readings:
  - Horton & Lipsitz (2001) Multiple imputation in practice. *The American Statistician* 55(3): 244-254
  - Allison, *Missing Data*, chapters 5-8 (on Blackboard)

#### Week 8 (5/09/07)

- Review and Final Exam

**Bibliography of Additional Recommended Readings (not required):**

Dorofeev, S and Grant P (2006) *Statistics for Real Life Sample Surveys: Non-Simple-Random Samples and Weighted Data*. Cambridge University Press.

Korn, ED and Graubard, BI (1999) *Analysis of Health Surveys*. New York: John Wiley & Sons.

Lehtonen, R and Pahhkinen, E (2004) *Practical Methods for Design and Analysis of Complex Surveys, Second Edition*. Chichester, England: John Wiley & Sons.

Levy and Lemeshow (1999) *Sampling of Populations: Methods and Applications, Third Edition*. New York: John Wiley & Sons.

Little, RJ and Rubin, DB (2002) *Statistical Analysis with Missing Data, Second Edition*. New York: John Wiley & Sons.

Lohr, SL (1999) *Sampling: Design and Analysis*. Pacific Grove, CA: Duxbury Press.

Rubin, DB (1987) *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.

Skinner, CH, Holt, D and Smith, TMF (1989) *Analysis of Complex Surveys*. New York: John Wiley & Sons,

Wolter, KM (1985) *Introduction to Variance Estimation*. New York: Springer-Verlag.