

BIOS 550 – Computer Analysis of Complex Survey Data

INSTRUCTORS: Paul Weiss

OFFICE: GCR 308

OFFICE HOURS: Tuesdays 10:00 - 11:30 am, or by appointment

LECTURES: Thursdays 10:00 am - 11:50 am, GCR P45

REQUIRED TEXTBOOKS:

- Heeringa, S.G., West, B.T., and Berglund, P.A. *Applied Survey Data Analysis*

REFERENCE TEXTBOOKS:

- Korn, E.L. and Graubard, B.I. *Analysis of Health Surveys*
- Cochran, W.G. *Sampling Techniques* (sample designs)
- Lehtonen, R. and Pakinen, E., *Practical Methods for Design and Analysis of Complex Surveys*
- Dorofeev, S. and Grant, P. *Statistics for Real-Life Sample Surveys*
- Wolter, K. *Introduction to Variance Estimation* (**BIOS students**)
- **Others to be determined**

EVALUATION:

Homework assignments will be assigned approximately bi-weekly. There will also be a short exam applying sampling concepts covered in the first half of the course, and a final exam covering analytical topics from the remainder of the course. There may be a computing component on the final exam. Both exams are open book and open notes, but use of the internet during the exams is not allowed. Homework comprises 20% of the course grade, the midterm comprises 40%, and the final exam comprises the remaining 40%.

GRADING:

This class is an elective for all departments. To reduce the number of audit requests, I have decided to require registration for the class as S/U rather than registration for the class as a letter grade. Due to restrictions set by the school this means that you may not take this class for a letter grade. Grades are assigned as follows:

- [70 – 100+) S
- [Below 70) U

HOMEWORK POLICY:

Homework sets will use real public-use data sets when possible, and students will apply computer analysis techniques to answer questions and develop tables based on these data. Data sets students may be using may include (but are not limited to) BRFSS, NHANES, NHIS and PRAMS. Some of these data sets will be analyzed in classroom exercises, and others may serve only as homework exercises or for a final project.

Homework sets will require the students to set up analysis programs in SAS, SUDAAN, and/or WesVar PC in an effort to produce publication-quality tables based on design-based estimates of survey data. Use of real-world data in this way is great experience for anyone who plans on using public-use data for secondary analysis at some point in the future. This class is ideal for masters' students intending to use a public-use survey data set for their thesis work.

Homework sets will be issued and collected on a semi-regular basis, and will reinforce many of the concepts covered in class.

EXAMS:

The midterm exam will cover sampling methods and concepts covered in the first half of the semester. Students will be allowed to use their notes and class slides for the exam. The final exam will cover materials from the second half of the semester, including analysis of survey data in SUDAAN and WesVar PC, survey costs and errors, and other topics. As with the midterm, the final will be open-note.

WEBSITE:

There will be a course website containing the syllabus, notes, assignments and important deadlines. The assignments will contain the due dates, details and the problem or problems you need to solve. Homework set solutions will also be posted here, along with code for producing the results.

Tentative Lecture Outline

Week	Topics	Readings from Heeringa et al.
1	Sample Designs I	1, 2.1 – 2.3
2	Sample Designs II	2.4 – 2.6
3	Sample Designs III	2.4 – 2.6
4	Finish Designs	2.8
5	Calculating Sampling Weights	2.7, 4
6	Non-Sampling Error	2.7, 4
7	Midterm Exam	
8	SUDAAN I	A4
9	SUDAAN II	A4
10	SAS Survey Procedures	A3
11	SAS Survey Procedures	A3
12	Replication / WesVar PC	A6.1
13	No Lecture	Final Exam (Cumulative, focus on material since midterm exam)

Note: Chapters 5 – 8 and 10 detail statistical methods we covered in Bios 500 and Bios 501. Students are encouraged to read these chapters in preparation for material following the midterm to review these concepts. Bios 550 will not re-teach statistical methods for analyzing data, but will use apply these methods to complex sample data using various software alternatives. These chapters use Stata as the software package of choice. As the class permits, Stata will be introduced as an additional software choice, given the growing usage of this package in public health settings.