ST640: Sampling Theory and Methods, Fall 2010

Meets: MWF 9:05 - 9:55 LGRT 219.
Instructor: John Buonaccorsi
Office: LGRT 1435K
Phone: 545-2809,
email: johnpb@math.umass.edu
website: http://www.math.umass.edu/~johnpb/s640.html

Text: (Required) Sampling Design and Analysis by Sharon L. Lohr, **1st Ed** (Duxbury). **NOTE: We are using the 1st edition, not the 2nd edition which just came out.**

Prerequisites: Calculus based Probability and Statistics (* see note below).

Grading: Homework: 50 %. Two tests worth 25% each.

DESCRIPTION: Numerous applications involve the sampling of units from finite populations. This includes, but is certainly not limited to,
- national, regional or local surveys of individuals or households for the purpose of assessing political or other opinions, labor status, spending behavior, medical status, behavior (e.g., physical activity), etc.
- sampling over spatial areas to assess habitat or other variables (estimating proportions of land in different types of usage, estimating crop yields or plant densities over regions, sampling the Quabbin reservoir to assess forest regeneration, etc.)
- sampling from a finite population of items for quality control purposes. This may be the producer sampling to assess quality before shipment, or the user/buyer assessing quality of a potential shipment.

This course provides an introduction to the theory and practice of sampling from such finite populations. This includes developing methods of analysis for a given design and contrasting the efficiency of different designs. Designs covered include simple random sampling, stratified sampling, systematic sampling, cluster sampling, sampling with unequal probabilities, multi-stage and double sampling, and general complex surveys. Methods covered also include ratio and regression estimators, the jackknife and other techniques for variance estimation, determination of samples sizes and optimal allocations and treatment of nonresponse. Much of the emphasis is on estimating means and proportions associated with populations (and subpopulations) but we also address other types of analyses (e.g., regression analysis) in the context of sampling from finite populations. The course includes an applied component, including data analysis. We will use SAS but student's proficient in another package/languages (e.g., R or STATA) can choose to program in that if they wish. No prior experience in SAS is assumed but some prior statistical computing experience is desirable.

* Prerequisites and level of the course. While there is a heavy emphasis on modeling, methods and applications, there is also a theory component of the course that assumes exposure to basic statistical theory, at least at the level of ST515-516 (which in turn assumes some calculus). Students who are interested in the applications but not the theory are welcome to do the course as pass/fail (and not do the theory part of homework and exams), audit the course (which will require completing the applied parts of the homework assignments) or to simply sit in.

Tentative Outline (Later parts will be refined as we move through the semester)

1. Introduction and overview (read Ch. 1)
2. Some Preliminaries. Read Appendix B for review, but can skip bottom half of p. 428 to end of B2 for now as we'll come back to this). Most important parts of this are properties of expectation and moments involving linear combinations of random variables (see bottom of page 427, top of page 428 plus class material) and use of conditioning (bottom of page 433, top of page 434).
3. Simple Random Sampling
   (a) Introduction (2.1,2.2)
   (b) Estimating the population mean, total or variance including approximate confidence intervals for the mean. (2.3,2.4,2.7, 428-429 of Appendix B.)
   (c) Confidence intervals for proportions (utilizing the hypergeometric).
   (d) Sample size determination. (2.5)
   (e) Optimality of the sample mean.
   (f) Ratio estimators. (3.1, p. 429-430 of Appendix B)
   (g) Regression estimators. (3.2)
   (h) Model-based approaches to inference. (2.8,3.4)
4. Stratified Sampling
   (a) Methods (4.1-4.3,4.6)
   (b) Allocation and Sample size (4.4,4.5)
   (c) Estimation with post-stratification. (3.3, 4.7)
5. Cluster Sampling
   (a) Single Stage Cluster Sampling (with systematic sampling as a special case) (5.0, 5.1,5.2,5.6)
   (b) Some general results under arbitrary probability sampling including Horvitz-Thompson estimation (6.4, 6.6 up to top of page 209).
   (c) Two Stage Cluster Sampling (5.3, 5.4, rest of section 6.6)
   (d) Designing Cluster Samples. (5.5)
6. Complex designs in general
   Estimation approaches, weighting, variance estimation, regression in complex sample surveys. Portions of Chapters 6, 7, 9 and 11.)
7. Nonresponse (parts of Ch. 8) and measurement error.
8. Two-phase studies (possibly) (12.1)