Introduction

Notes on Categorical Data Analysis and Survey Sampling

The NCSSM Statistics Leadership Institute was held on the campus of the North Carolina School of Science and Mathematics from June 26 to July 16, 1999. A grant from SAS Institute supported the Institute. Additional funding was secured from the College Board, Minitab, Inc., and Duxbury Publishers. Forty-eight experienced secondary statistics teachers met with outstanding statisticians for intensive study in four areas of statistical theory and application. The areas of concentration were

- The Theory of Inference
- Regression Analysis
- Experimental Design
- Categorical Data Analysis and Survey Sampling

The sessions on the theory of inference were led by Jon Cryer, Professor of Statistics at the University of Iowa, and by Jeff Witmer, Professor of Statistics at Oberlin College. These sessions were intensive and demanding, taking the first week of the Institute.

The sessions on Regression Analysis were led by Bob Stephenson, Professor of Statistics at Iowa State University. The sessions on experimental Design were led by Linda Young, Professor of Statistics at the University of Nebraska. These sessions shared the second week of the Institute.

The sessions on Categorical Data Analysis and Survey Sampling were led by Dick Scheaffer, Professor of Statistics at the University of Florida. These sessions took the first half of the last week of the Institute.

During each session, Chris Olsen and Gloria Barrett took careful and copious notes. These notes have been supplemented with homework assignments, computer and calculator simulations, and additional comments that came in evening discussions to give a broader and more complete story of the work of the Institute.

Each of the 4 major topics is presented as a separate section. Each begins with a short introduction, followed by from 3 to 8 sections containing the extended notes of the session. The notes for each topic should be read in order, since each section builds upon the previous sections. If possible, they should be supplemented with the references used in the Institute and noted in the introduction to each topic.

These notes cover the sessions presented by Dick Scheaffer of the University of Florida. They consist of 56 pages of notes in 2 sections. The reference for this section was Elementary Survey Sampling, 5th Edition, by Scheaffer, Mendenhall, and Ott, Duxbury Press., 1996.
Richard Scheaffer

Professor Scheaffer received his Ph.D. in statistics from Florida State University, whereupon he joined the faculty of the University of Florida and has remained there ever since. Now Professor Emeritus of Statistics, he was chairman of the Department for a period of 12 years. Research interests are in the areas of sampling and applied probability, especially with regard to applications of both to industrial processes. He has published numerous papers in the statistical literature and is co-author of four college-level textbooks covering aspects of sampling, probability and mathematical statistics.

In recent years, much of his effort has been directed toward statistics education throughout the school and college curriculum. He was one of the developers of the Quantitative Literacy Project in the United States that formed the basis of the data analysis emphasis in the mathematics curriculum standards recommended by the National Council of Teachers of Mathematics. He continues to work on educational projects at the elementary, secondary and college levels, and served as the Chief Faculty Consultant for the Advanced Placement Statistics Program for the 1997 and 1998 exams. Dr. Scheaffer is a Fellow of the American Statistical Association, from whom he has received a Founder's Award.

Chris Olsen has divided his college time between Iowa State University (undergrad) and the University of Iowa (grad). He has been teaching for 28 years, and has taught statistics at the high school level since 1976. He was a participant in the first Woodrow Wilson Summer Mathematics Institute on Statistics, and is currently a member of the AP Statistics Development Committee. He has been recognized with the Presidential Award for Excellence in Mathematics Teaching, and IBM Outstanding Computer Educator Award. Chris has been teaching at George Washington High School for the past 25 years, and is an active participant in the apstat-l listserv for AP Statistics teachers. With a daughter going off to college next year, he is in the throes of "empty-nest" and is very receptive to sympathy from any sources.

Gloria Barrett has taught at the North Carolina School of Science and Mathematics for 14 years where she has been a member of the writing team for two textbooks, *Contemporary Precalculus through Applications* and *Contemporary Calculus through Applications*. She is the author of the calculator workbook, *Statistics with the TI-83*. She was recognized with National Board Certification in 1998. In 1987 she attended the Woodrow Wilson National Fellowship Foundation summer institute on Mathematical Modeling and for ten years served as a member of the Woodrow Wilson outreach team that conducted one-week summer workshops for teachers at various sites across the country. Gloria served as a member of the development team for the Teachers Teaching with Technology (T³) institute in Modeling and Data Analysis and the institute in Advanced Statistics. She has been a presenter in these workshops in 1997 and 1998. She is a pioneer at NCSSSM in teaching Statistics via two-interactive TV through our Distance Learning Program.
Outline for Notes

Categorical Data Analysis

I. Analyzing Categorical Data: Chi-Square and Residual Analysis
II. Measuring the Strength of Association
   A. The Effect of Sample Size
   B. Concordant and Discordant Pairs
      1. Physicians Health Study
      2. Turtle Example Revisited
   C. Sensitivity and Specificity

Survey Sampling

I. Sample Surveys and Experiment
   A. Some Terminology
   B. Probability Samples
      1. Definitions and Comparisons Among Methods
      2. The Need for Probability Samples
      3. Sources of Errors
      4. Steps in Planning a Survey

II. Simple Random Sampling
   A. Correction Factor for Finite Populations
   B. Estimation of a Population Mean
   C. Estimation of a Population Proportion
   D. Estimation of a Population Total
   E. Sampling with Sub-Samples

III. Stratified Random Sampling
   A. Gallup Poll
   B. Estimating the Population Mean
   C. Problems of Sample Size and Allocation
   D. Comparison of Stratified Random Sampling to Simple Random Sampling

IV. Ratio Estimation
V. Cluster Sampling
   A. Estimation of a Population Mean and Total
   B. Comparison of Cluster Sampling and Stratified Sampling

VI. Systematic Sampling
VII. Estimating the Size of a Population
We present these material in hope that they will be useful to other experience teachers to help deepen their understanding of these fundamental aspects of statistics. Any errors, whether statistical, grammatical, or typographical are entirely my responsibility. I welcome and will make any necessary corrections that readers find.

Dan Teague
Director, NCSSM Statistics Leadership Institute
Instructor of Mathematics
NC School of Science and Mathematics