

Introduction

Notes on The Theory of Inference

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 Jon Cryer of the University of Iowa and Jeff Witmer, of Oberlin College on the Theory of Inference. They consist of 90 pages of notes, in 8 sections, with 2 appendices and this Introduction. The reference for this section was *Mathematical Statistics with Applications*, Wackerly, Mendenhall, and Scheaffer, Duxbury Press, 1996.

Jeff Witmer

Jeffrey A. Witmer holds a Ph.D. in statistics from the University of Minnesota and is a professor at Oberlin College. He has been a presenter in over two dozen workshops and was the principal investigator of an NSF-sponsored project to incorporate data analysis into middle school and high school science education. He has authored or co-authored half a dozen books. He has served as the editor of STATS magazine and on the editorial boards of The Journal of Statistics Education and The American Statistician.

Jon Cryer

Jon Cryer received his Ph.D. in statistics from the University of North Carolina - Chapel Hill in 1966. His minor was in mathematics. His undergraduate degree was in mathematics with a minor in physics from DePauw University. Jon was recognized as a Fellow of the American Statistical Association in 1996. He was elected to the Executive Committee: Section on Statistical Education of the American Statistical Association, 1998-2000, served as the Associate Editor for the Journal of the American Statistical Association, 1991-1996, and is a frequent speaker at NSF/MAA/ASA sponsored STATS Workshops. Jon is the author of numerous texts, including *Time Series Analysis*, PWS-KENT, Boston, Duxbury Advanced Series, 1986, *Statistics for Business: Data Analysis and Modeling*, second edition, with Robert B. Miller, Duxbury Press, and *Minitab Handbook for Windows, Release 12*, 1998, Hayden McNeil Publishing, Plymouth, MI. He has also created several multimedia statistical packages, including *Electronic Companion in Statistics*, with George Cobb, Cogito Learning Media, New York, 1997 and *Electronic Companion in Business Statistics*, with George Cobb, Cogito Learning Media, New York, 1997.

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 (T^3) 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 NCSSM in teaching Statistics via two-interactive TV through our Distance Learning Program.

Outline for Notes

Introduction to the Theory of Inference

- I Random variables
 - A. Discrete Random Variables
 - 1. Binomial Distribution
 - 2. Geometric Distribution
 - 3. Negative Binomial Distribution
 - 4. Hypergeometric Distribution
 - B. Sampling With and Without Replacement
 - C. Continuous Random Variables
 - 1. Gamma Distributions
 - 2. Chi-Square Distributions

- II. Moment Generating Functions and Their Properties

- III. Student's t Distribution
 - A. Normal as Limiting Distribution using Moment Generating Functions
 - B. Independence of Sample Mean (\bar{x}) and Sample Variance (s^2)
 - C. Expected Value of Sample Variance

- IV. Introduction to the Central Limit Theorem
 - A. The Central Limit Theorem
 - 1. Sampling Distribution of Sums
 - 2. Bernoulli Trials
 - 3. Simulations of the Central Limit Theorem
 - B. Confidence Intervals for μ

- V. Estimators and Parameters
 - A. Method of Moments
 - B. Maximum Likelihood
 - C. Sufficient Statistics

- VI. Introduction to Hypothesis Testing
 - A. Binomial Setting

- 1. Power
 - 2. Sign Test
 - B. Most Powerful Tests: Neyman-Pearson Lemma
 - C. Likelihood Ratio Tests
 - D. Chi-Square and Z-Tests
 - 1. Equivalence of Tests for Proportions
 - 2. Two Proportion Tests
- VII. The 2-Sample t Tests
- A. The Pooled t Test
 - B. The Unpooled t Test
- VIII. Linear Statistical Models
- A. The Method of Least Squares
 - B. Standardized Variables
 - C. Properties of the Least Squares Estimator for Slope

Appendix 1: Derivation of Stirling's Formula $n! \approx e^{-n} n^n \sqrt{2\pi n}$

Appendix 2: Plausibility argument for $\lim_{u \rightarrow \infty} \frac{\Gamma\left[\frac{u+1}{2}\right]}{\sqrt{up} \Gamma\left[\frac{u}{2}\right]} = \frac{1}{\sqrt{2p}}$.

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