Introduction to Notes of Regression Analysis
Introduction

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 John Cryer, Professor of Statistics at the University of Iowa, and by Jeff Witmer, Professor of Statistics at Oberlin College. These sessions were given during 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. Any errors in these notes are the responsibility of the present author.

Each of the four major topics is presented as a separate section. Each topic begins with a short introduction, followed by from 3 to 7 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 Bob Stehpenson of Iowa State University on Regression Analysis. Notes for this session were organized by Chris Olsen and Gloria Barrett.

Bob Stephenson is Professor of Statistics at Iowa State University in Ames, IA. In his position, Bob is heavily involved with teaching at all levels. He is the course coordinator for STAT 101: Principles of Statistics. As coordinator, he teaches one section of 100-110 students and supervises the instructors for the other four sections. Bob is also involved in distance education. He developed a two semester course, Applied Statistics for Industry, which is delivered via videotape delay to students at industrial sites across the country and in Mexico. Students from General Motors, 3M, Rockwell-Collins, John Deere, Frigidaire, Maytag, Gillette, and several smaller Iowa companies have taken this course over the past 5 years. Other courses Bob teaches on a regular basis include Statistical Methods for Research Workers and Statistical Design and Analysis of Experiments. This past semester, Bob was a member of a team that developed a new course on Advanced Statistical Methods for Research Workers. He and a colleague are also revamping a statistical computing course.

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 $\left(T^3\right)$ 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

I. Simple Linear Regression
   - A. Data Exploration and Inference (HIV/AIDS Example)
   - B. Linear Model
   - C. Quadratic Models
     1. Simple Quadratic Model
     2. General Quadratic Model
        - a. Multicolinearity
        - b. Coding Data
     3. Cubic Model
   - D. Exponential Model

II. Multiple Regression (Gas Consumption Example)
   - A. Building a Regression Model
   - B. Adding a Categorical Variable
   - C. Questions of Statistical Significance
   - D. Adding an Interaction Term
     1. Insulation only Model
     2. Fitting Separate Linear Models
     3. 2-Sample *t*-test

III. Logistic Regression (Challenger and Turtle Eggs Examples)
   - A. Logistic Curvilinear Model
     1. The Logit Transformation
     2. Maximum Likelihood Approach
   - B. Assessing the Fit
   - C. Challenger Disaster Revisited

IV. Appendices
   - A. A Vector Approach to Linear Regression
   - B. A Matrix Approach to Multivariate Regression
   - C. That ANOVA table at the Bottom of Regression Output

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