This course provides an introduction to statistical computing using SAS. It is intended for upper level undergraduates only and graduate students. Students entering the course are assumed to have had a prior course exposing them to the basic principles of statistical inference (e.g., ST501 or ST516 or equivalent) with additional prerequisites for the later two modules, as described below. This course cannot be taken by someone with only ST515 as prior coursework in Statistics.

The primary objective, as the title indicates, is to teach the student useful programming skills for addressing a variety of problems in statistics and probability, including carrying out monte-carlo simulations. More detail on coverage is described for each of the modules below. In addition, the course(s) also i) provides reinforcement of fundamental statistical concepts and their use in applications, ii) provides an introduction to a number of topics which may be new for many of the students (e.g., power analysis, sample size determination, bootstrap techniques, etc.)

- **ST597B. Statistical Computing I:**
  This course will provide an introduction to the use of SAS for statistical analyses. Coverage includes data entry and manipulation, and the use of “canned” procedures in SAS for descriptive statistics and basic statistical analyses, including inferences for means and proportions in one and multi-sample settings and contingency tables. It is assumed that students have had a previous course covering descriptive statistics and basic statistical methods. No prior exposure to SAS is assumed.

- **ST597C. Statistical Computing II:**
  Prerequisite: ST597B or equivalent, including some prior knowledge of probability and sampling distributions.
  This course will provide an introduction to SAS as a programming language. Topics include writing to external files, macros, arrays, do loops, if-then statements, probability and mathematical functions, random number (and random variable) generators, and more. The computing will be carried out in the context of handling interesting problems in probability, sampling distributions and statistics, including power calculations and simulations.

- **ST597D. Statistical Computing III (Prerequisite: ST597C or equivalent, plus basic matrix theory.)**
  This course will provide an introduction to programming using SAS-IML. We first introduce the basics of working with matrices in IML and then turn to various applications including (but not limited to) calculation of univariate and multivariate
statistics in IML, linear and nonlinear least squares, random vectors and linear combinations, simulations, and bootstrap resampling.

Detailed table of contents for each of the three modules are below.
Contents

1 An overview of SAS system and how to run it. 3
  1.1 What is SAS? ........................................ 3
  1.2 Where do I run SAS? .................................. 3
  1.3 Where do I read about things? ......................... 3
  1.4 How do I run SAS? .................................... 4
  1.5 The general structure of running SAS using programming statements versus “click and go”. 4

2 Basic elements in the data step. (Creating and listing a SAS data set). 5
  2.1 Reading from external text files. ....................... 5
     2.1.1 Reading space delimited data from a text file with unformatted input list. ............. 5
  2.2 Creating new variables. (expressions, if-then, concatenating character variables, etc.) .... 9
  2.3 Selecting cases and variables(SAS data option keep, drop, where and selecting cases via the if statement.) .......................................................... 12
  2.4 Turning one case into many using “output”. .......... 13
  2.5 CONCATENATING CHARACTER VARIABLES. .......... 14

3 Sorting data (proc sort) and the ”by” statement. 14

4 PERMANENT SAS DATA SETS. 14

5 Merging and combining data sets. 16

6 Describing data for a single variable. 19
  6.1 Numerical and tabular summaries. .................... 19
  6.2 Graphical Displays. ................................... 26

7 Inferences for a single random sample. 29
  7.0.1 Simple Random Sampling from a finite population. ............................................. 34
  7.1 More on estimating proportions via the analyst ......................................................... 34

8 Scatterplot for two variables. 35

9 Describing one variable across multiple groups. 37

10 Comparing two means or proportions with independent samples. 39
11 Comparing means or proportions with paired samples. 43

12 Comparing two proportions when the responses are paired on a unit or individual. 44

13 Creating SAS files with summary statistics. 46
   13.1 Attaching the grand mean to all cases. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 48

14 Examining relationships among variables. 49
   14.1 Correlation and related tests . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 49
   14.2 Two-way contingency tables. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53
Contents

1 More on Summary Statistics. 2

2 Basic use of macros. 3
   2.1 Running the same analysis on many data sets. 3
   2.2 Running the same analysis on different variables. 4
   2.3 % let. 5
   2.4 A macro within a macro. 6

3 Writing to external files via the PUT statement. 7

4 Arrays 9

5 Using SAS GRAPH to create high resolution plots. 14

6 Some programming features. 16

7 Probability Functions. 18
   7.1 Plotting the normal density. 23
   7.2 Exact inferences for a proportion based on the Binomial. 23

8 Inverse probability functions. 26

9 The noncentral t-distribution and power functions. 30

10 Determining sample size for a single mean or proportion via a confidence interval or standard error. 34
   10.1 Estimating $\mu$ with a random sample of size $n$. 34
   10.2 Sample size for estimating a proportion with a random sample. 35

11 Power and sample size for test for a single proportion. 36

12 Comparison of two means; independent samples. Power and sample size. 38

13 Generating random quantities. 42
   13.1 Generating random integers from between 1 and $M$. 42
   13.1.1 Sampling with proc surveyselect. 43
13.2 Generating Normal random variables. ......................................................... 44
13.3 Simulating samples from the exponential distribution. ............................... 46
13.4 Simulating the number of arrivals in a fixed period of time. .......................... 49
13.5 Simulating survival over time. ................................................................. 52
    13.5.1 Simulating the survival of a cohort over time. .................................... 52
    13.5.2 Simulating Time until extinction ....................................................... 53
13.6 Simulating the performance of a crude estimate of survival rate. .................. 55
# Contents

1  Introduction to IML.  
   1.1  Naming rows and columns, formats, reset  
   1.2  Referring to elements  

2  Reading from a SAS data set into matrices.  

3  Descriptive statistics etc. in a one-sample problem  
   3.1  Chi-square Goodness of Fit test.  

4  MISSING VALUES.  

5  Writing to external files in IML.  

6  Statistics for multiple groups.  

7  Multivariate observations.  

8  Creating a SAS file from a matrix.  

9  Least squares/regression via IML.  

10 Non-linear least squares and calling subroutines.  

11 User define modules/subroutines/functions  

12 Simulating sampling from a set of values.  

13 Simulating the performance of least squares estimates.  

14 Bootstrapping.  

15 Simulating the bootstrap percentile intervals.  
   15.1  Bootstrapping in Two Sample Problems  

16 An example with nonlinear regression and bootstrapping.