

STA 215: Statistical Inference --- Program Cover Document

I. Basic Course Information

A: STA 215 – Statistical Inference

B. The course will be scheduled for two lecture periods and a third meeting period to be used at the instructor's discretion such as for a computer lab, help session, etc.

One semester of calculus is a prerequisite for the course.

II. Learning Goals

A. Content Goals:

The field of statistics in its broadest terms involves the analysis of data. More specifically it may involve the planning for data collection, the actual collection of data, the subsequent analysis of the data, the drawing of inferences from these analyses and the presentation of the results.

Students in STA215 are not expected to have any prior exposure to statistics but are expected to have completed a one semester course in Calculus. Thus, the majority of the course is a systematic introduction to the various techniques employed in the statistical analysis of data and, in addition, some of the mathematical underpinnings of the techniques are introduced. Because many of the analysis techniques require extensive computations, for other than the simplest data sets, each student will become familiar with one of the software packages that are available. The student, on satisfactory completion of the course, is prepared to take further courses either in statistical theory or in applied areas, though more mathematics may be needed for some courses.

On completion of the course the student should have a foundation in the following areas:

- a. Identifying a problem that requires the collection of data and structuring the objectives
- b. Planning for the collection of the data – including issues of random sampling
- c. Be involved to a greater or lesser extent in the collection of data
- d. Performing analyses of the data using relevant techniques using a software package
- e. Draw conclusions and inferences from the analysis.
- f. Make recommendations to the client as part of a final report.

The statistical techniques covered in the course include as a minimum:

1. Basic techniques of data display (histogram, bar chart, stem-and-leaf) and data summarization (mean, median, standard deviation, IQR) which may culminate in a Box-Plot.
2. Introduction to Probability Theory as it applies to statistics
3. Introduction of Random Variables for both Discrete and Continuous Distributions
4. Sampling distributions – normal, binomial and the distribution of the sample mean
5. Techniques of Inference – Estimation involving Confidence Intervals and Tests of Hypotheses for both one and two sample situations for both means and proportions
6. Simple Linear Regression and Correlation analysis

Manual computations can sometimes be useful in learning a new concept, but extensive computations to arrive at an answer do not result in any significant learning. Therefore, once a student has a feel for a concept, he/she should be able to perform the computations using a software package such as Data Desk, SPSS, Minitab, etc. The student needs to be able to manage data files within the system as well as to perform analyses. The student needs to use a variety of

checks to ensure that the computer output is exactly what the student is expecting. It is obvious that the use of the computer allows the students to analyze larger, more realistic data sets.

Students in Stat 215 have recently been exposed to calculus in which problems generally have a unique solution with some variation on the approach possible. Statistics, on the other hand, starts with a general problem for which there is not a single correct answer. The criterion for success is whether or not the solution represents a reasonable approach to the analysis. For example, there are a variety of ways to display and summarize something as simple as the SAT scores for a group of students. In this context, mathematics (calculus and probability in this case) provides the tools to develop the techniques. Appreciation of the subtleties of the approach are a major learning objective of the course.

The principal enhancement to the course is the major increase in the utilization of computer technology. During the lecture the instructor can easily demonstrate a concept such as the adverse effect on the mean of an outlying observation as the value of the outlying observation changes. In addition, the Monte Carlo Simulation capabilities allow the student to witness the formation of the distribution of the sample mean during the lecture. Virtually every concept in the course can be illustrated using the computer.

III. Student Assessment

Given a specific technique, such as confidence intervals, the first assessment is to determine the level to which the student understands the concept, can apply the concept and interpret the results. The second assessment is to determine the level to which the student can integrate the technique into a broader analysis of data. The third level of assessment is to determine the level to which the student can think in the broadest context from what is the underlying research need in a problem through all phases of data collection and analysis to a final report and recommendation. Integral to this process is the need for the student to utilize effectively a software package that will perform the necessary analyses and also graphically display the data and results in a form that is meaningful to the reader. Also integral to the process is an assessment of the student's ability to write technical material in a well organized manner that conveys the essence of the analysis.

IV. Learning Activities

Given the nature of statistics, the learning activities can take a variety of forms. There is the classroom lecture which can be supplemented with the use of a computer in the classroom to illustrate new techniques. In the classroom the students can work together to experience a new concept or technique. The student does reading assignments and performs analyses of data (using the computer in some cases) in order to learn the material.

In addition to the student using the computer to perform statistical analyses, the student can be asked to simulate data from a statistical distribution such as the normal distribution. The student can, in this way, further his/her understanding of basic concepts like forming sampling distributions or observing the proportion of confidence intervals that do not contain the population mean.

STA 215: Statistical Inference

Course Syllabus

Instructor: Richard P. Thayer

Phone: x-2500, Email: Thayer@tcnj.edu

Office: Science Complex 213

Office Hours: 12-2 Tuesday and Friday

Text: ActivStats CD-ROM, Paul Velleman, Addison Wesley
Introductory Statistics, Neil Weiss, Addison Wesley (Optional)

Course Goals:

The overall objective of statistics is the analysis of data – data that arises in surveys in business and in the social sciences, data that arises in designed experiments in the sciences and engineering, data that arises in various business applications, etc. More broadly, the field of statistics encompasses the planning of a study, the collection of data for the study, the analysis of the data, the reporting of results and the making of recommendations. The goal of a course in statistics, then, is that the student will understand the mathematical development and application of various techniques that are useful in analyzing data. (Calculus is a prerequisite for the course, so that the instructor can use calculus to development and illustrate various techniques.) A secondary goal is that the student will become familiar with a software package useful in performing the calculations necessary with large data sets. This goal is satisfied with the Data Desk software which is integral to the ActivStats CD-ROM. A tertiary goal is that the student be able to document a study in a manner that is comprehensible to the average reader who is not schooled in statistical terminology and methodology.

The software, in addition to its capability to perform analyses, has the capability to perform Monte Carlo simulations. The goal is that the student will have a much deeper understanding of such concepts as sampling distributions, outlier impacts, the alpha level in hypothesis testing, etc. In achieving this goal, the instructor can draw samples which illustrate concepts during the lecture. In addition, the student gains insights from the presentation on the CD-ROM and the student can be assigned Projects and Homework Assignments which require the use of sampling distributions.

Course Content Overview:

The first third of the course is concerned with a variety of methods for the graphical presentation of data and for calculating various summary statistics that describe the patterns in the data.

During this period issues of Experimental Design are introduced.

The second third of the course begins the study of statistical inference – the concept of drawing conclusions about a population based on information obtained from a sample. This leads to the last third of the course which deals with particular areas of inference, that of tests of confidence intervals and tests of hypotheses. Also included in the course is a study of simple model building, called regression analysis. Throughout the course a software package will be used for analysis of data. Also students will be required to understand the broader aspects of a study including the

design of the study, the statistical analysis and the presentation of results. Because statistics is widely used in many fields, problems will be presented to illustrate a variety of applications.

Course and Performance Goals:

The goal of the course is for the student to gain insight into how useful information can be obtained from a large array of data: by displaying and summarizing the data appropriately, by quantifying and understanding relationships among the variables and by drawing inferences. Statistical concepts such as random variables and sampling distributions appear to be straightforward concepts, but they take a considerable amount of effort to internalize. It is only by working a variety of problems that the student comes to grip with these concepts and can then demonstrate his/her mastery. Similarly, it takes a lot of experience to work with a software package to produce the desired analyses.

Assessment:

The assessment of student achievement in the course is accomplished through formal examinations, unannounced quizzes in class and assigned projects. A formal in-class examination is given after the first and second thirds of the class – each being worth 25% of the grade. A final examination will be given at the end of the course and will be worth 30% of the final grade. All examinations will consist of the following:

1. Problems that require the student to select and implement a method of analysis,
2. Problems in which the student will have to write a coherent summary of an analysis,
3. Problems in which the student will have to discuss a statistical issue such choosing a sample from a population, and
4. Problems in which the student demonstrates, in the computer lab, facility with the software.

Unannounced quizzes are given at the beginning of a period to ensure that students are doing the homework. The quizzes are immediately discussed and graded for the next class meeting. At no time are there make-ups given for lateness or absence. The lowest quiz grade will be dropped. Quizzes are worth 10% of the grade.

Projects will be assigned for which the student will perform a more extensive analysis or perform a simulation. It is expected that the results will be documented in a professional manner. Credit is given for creativity, accuracy, completeness, effective use of software in analysis and graphical display, etc. The projects are worth 10% of the grade.

Computer Lab:

Classes will be given in a room equipped with a computer and projection equipment so that the instructor can demonstrate statistical content and use of software. Students will be required to use the computer lab or to use Data Desk on another computer. The instructor will conduct some office hours in the computer lab so that students have the opportunity to ask questions regarding the software.