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

C. One semester of calculus is a co-requisite 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 or be enrolled in 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. Identifying and performing analyses of the data using relevant techniques
d. Drawing conclusions and inferences from the analysis.
e. Making recommendations as part of a final report.

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

1. Basic techniques of data display and data summarization
   - Histograms and boxplots
   - Five number summary, mean, and variance
2. Introduction to Probability Theory
   - Laws of probability including rules for disjoint and independent events
   - Bayes' theorem, conditional and marginal probability
3. Introduction to Random Variables for both Discrete and Continuous Distributions
   - Rules for expectation and variance of combined distributions
4. Sampling distributions
   - Normal and binomial distributions
   - Normal approximation of the binomial
   - The t-distribution
5. Techniques of Inference for One and Two Samples
   - Hypothesis tests including t-tests and chi-squared tests
   - Confidence intervals on proportions and means
   - The definition of the p-value and effect size
   - Type I and Type II error and power
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. 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 STA215 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 is 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 that 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.

STA215: Statistical Inference

Sample Course Syllabus

Textbook: DeVeaux, Welleman, Bock: Intro Stats, 4th Ed.

Policy on Older Editions: Students may use earlier versions of the text so long as they take full responsibility to obtain assigned problems and cover any information that might be missing relative to the latest edition.

Students are responsible for all information presented in the textbook for chapters we cover whether or not we discuss the topic in class.

Instructor: Professor

Office: Phone:

Email (preferred):

Office Hours: Wed 10:00 AM – 12:00 Noon, Th 12:30 – 1:30 PM, and Additional Hours as Requested

Course Description: This course will present the basics of statistical inference, including treatment of categorical and quantitative data, concepts of random error and distributions, and limits on prediction and confidence. The focus is on the concepts needed to correctly apply statistical reasoning and on the correct application of statistical frameworks for inference.

Course Philosophy: Our world has become data full, to the point of bursting in many domains. The recovery of knowledge from data has been the focus of statistical study since the early days of the scientific revolution, demonstrated by the work of Bayes and Laplace. Thanks to computers and user-friendly programs, it has become trivially simple to apply many statistical methods to data. However, proper application of these methods requires an understanding of the mathematical assumptions underlying the methods and knowledge of the available methods. This course will provide an initial introduction to these areas.

Statistical concepts can appear amorphous when first encountered. As such, the course will include problem solving throughout to reinforce key concepts. To remove the technical barriers to learning as much as possible, calculators will be used throughout this course for homework and exams. Built-in statistical functions will be discussed, however the key issue in all cases will be the determination of the correct statistical method to apply and the setting up of the solution.

Evaluation: Evaluation will be based on psets (10%), weekly quizzes (10%), classroom participation in examples (10%), two exams during the semester (20% each), and the final exam (30%). Each pset will include a series of problems from the book with solutions in the back, which students are expected to understand *completely* but not turn in. Each problem set will also include a difficult problem to be turned in. **Students are expected to work together on the difficult problems, however all psets must be written individually without copying each other's work.** Also note that it will be statistically impossible to pass this course without doing sufficiently well on exams, so students should insure that they understand the problems. **Each pset is due at the beginning of class on the day noted in the syllabus.** Late homework turned into the Math-Stat office by 4:00 PM on the day prior to the next class meeting will receive a 25% penalty. No credit will be given for homework after this time.

Schedule (Subject to Non-Random Variation) – **Bold** indicates **PSET DUE**

Date	Topic	Chapter	Date	Topic	Chapter
27-Jan	Introduction		20-Mar	SPRING BREAK	
30-Jan	Categorical Data	2	24-Mar	Sampling Distribution and CLT	15
3-Feb	Random Variables and Probability	9, 12	**27-Mar**	Distributions of Proportions	16
6-Feb	Hypothesis Testing - Chi-Squared Tests	22	31-Mar	One Sample Z-Test	17
10-Feb	Chi-Squared Tests	22	**3-Apr**	Distributions of Means and the t-Test	18
13-Feb	NO CLASS		7-Apr	Two Sample Tests	20
17-Feb	Quantitative Data - Visualization and Basics	3	10-Apr	EXAM 2 (Chaps 1-5, 13 - 17)	
20-Feb	Distributions	3	14-Apr	Hypothesis Testing - Confidence Intervals	19
24-Feb	The Normal Distribution	5	**17-Apr**	ROC Analysis, Wilcoxon Tests	
27-Feb	Z-Scores and the Standard Normal Curve	5	21-Apr	Sampling and Surveys	10
3-Mar	Probability I	13	**24-Apr**	Experiments and Observational Studies	11
6-Mar	EXAM 1 (Chaps 1 - 5, 12, 22)		28-Apr	Two Variables and Correlation	6
10-Mar	Probability II	14	**1-May**	Regression Preview	7
13-Mar	Random Variables and Distributions		5-May	Review	
17-Mar	SPRING BREAK		8-May	Review	

Classroom Policies

In this class, the deep learning outcomes associated with TCNJ's 4th hour are accomplished by a series of rigorous educational assignments that extend beyond the typical scheduled class time. These include learning to use statistical analysis tools on calculators, completion of extended homework problems set by the professor, which will include working with fellow students on these difficult statistical problems. In addition, it is expected that many students will require additional time with tutors or the professor during the semester to develop the skills needed to apply the statistical concepts learned.

Attendance: All students are expected to attend all classes and are responsible for all information provided. A student who is absent for a test will not be permitted to make up the test unless prior arrangements with the instructor have been made. Approval for missing a test will only be permitted in exceptional circumstances. In the case of illness, a doctor's note will be required. Please view TCNJ's attendance policy at http://policies.tcnj.edu/policies/digest.php?docId=9134

Academic Honesty: Please make sure you are familiar with TCNJ's academic integrity policy. Any suspected violation of this policy will be confronted in the strict accordance with the policy: http://policies.tcnj.edu/policies/digest.php?docId=7642

Americans with Disability Act Policy:
http://policies.tcnj.edu/policies/digest.php?docId=8082

Final Exam-Evaluation-Reading Days Policy:
http://policies.tcnj.edu/policies/digest.php?docId=9136

Learning 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. 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. Identifying and performing analyses of the data using relevant techniques
d. Drawing conclusions and inferences from the analysis.
e. Making recommendations as part of a final report.

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

1. Basic techniques of data display and data summarization
    - Histograms and boxplots
    - Five number summary, mean, and variance
2. Introduction to Probability Theory
    - Laws of probability including rules for disjoint and independent events
    - Bayes' theorem, conditional and marginal probability
3. Introduction to Random Variables for both Discrete and Continuous Distributions
    - Rules for expectation and variance of combined distributions
4. Sampling distributions
    - Normal and binomial distributions
    - Normal approximation of the binomial
    - The t-distribution
5. Techniques of Inference for One and Two Samples
    - Hypothesis tests including t-tests and chi-squared tests
    - Confidence intervals on proportions and means
    - The definition of the p-value and effect size
    - Type I and Type II error and power
6. Simple Linear Regression and Correlation analysis

Students in STA215 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 is the result of 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 is a major learning objective of the course.