Course Module
Stephen F. Austin State University
MATH 1342 - 002, Introduction to Statistics
Math (Bush) 202 10:30AM-12:35PM MTWR
ZOOM: Mtg#: 956 3693 6089, PC: 419958
https://sfasu.zoom.us/j/95636936089?pwd=V1UzYVIWMnU0ekVBaDBUd0oyOUU3UT09

Instructor
Robert (Bob) Henderson
Department: Mathematics & Statistics; Office: Bush 344
E-mail: hendersork@sfasu.edu
Phone: (936) 615-7796
BA in Math & History – Trinity University, San Antonio, TX (1978)
MS in Mathematical Statistics – Southern Methodist University, Dallas, TX (1980)
PhD in Mathematical Statistics – Southern Methodist University, Dallas, TX (1982)
MBA – University of Delaware, Newark, DE (1988)
Worked in industry for 27 years: ~6 years with DuPont as internal consultant for a variety of
businesses and staff groups; then ~21 years in the semiconductor business, most with a supplier of a
key enabling material for semiconductor production; and later with Samsung working primarily
with engineers in process control efforts. The entire 27 years included many training delivery, as
well as course development activities related to basic statistics, experimental design, and process
control systems. Fall 2009 was first semester working at SFA.

Teaching Hours – 8:00AM to 10:05AM & 10:30AM – 12:35PM MTWR
Office Hours – 2:00PM-4:00PM MTW, and by appointment.

Course Goals
Statistics is a subject that has a tendency to evoke emotions for students. These emotions can run
from fear and loathing to apathy, and there are the occasional students who are excited about the
subject. However, the latter is generally not the norm. One of the goals of the course will be to
change any primary negative emotion associated with statistics to one of at least respect for the
subject, if not excitement, and, of course to nurture even greater enthusiasm for anyone already
interested in the subject.

There is substantial evidence that statistics can be challenging for many students. Traditionally, a
significant barrier to having students embrace learning about statistics has been the relatively large
number of formulas involved. These are not going to go away, and there are many of them in the
course material; however, it is NOT going to be a goal of this course for the student to memorize
formulas. Students will have access to formulas for all work and exams, just as you would in any
future employment.

My experience is that much of the challenge for many people with the subject of statistics is that the
subject often requires us to think about things differently than we have before encountering this
subject. Statistics is a subject that requires admission of ignorance, and this takes a certain level of
humility, security, and self-esteem to do. Fortunately, it does not require an admission of complete
and total ignorance, and certainly no one enrolled here at SFA would ever be so described.
However, it does require admission of ignorance about a specific question or set of questions, which
are sufficiently important that effort to reduce our ignorance related to this question (or these questions) is deemed to be of some value.

In order to reduce the level of ignorance, often an experiment is run, an observational study is conducted, or a survey of some sort is initiated, all of which generate data related to the question(s) of interest. With the presence of data, there is always that lurking discipline of statistics, which is simply the art of extracting meaningful information from data.

Unfortunately, what you will find out in this course is that there are many ways to do this, and that rarely, if ever, will one approach provide the exact answer(s) to the original question(s). This can make the study and use of statistics frustrating for some. However, if the process of collecting, analyzing, and interpreting the relevant data is done well, then knowledge will be improved and the level of ignorance will be reduced. If done poorly, then the level of ignorance may actually be increased. Consequently, a second goal of this course is to provide the student with the knowledge necessary to understand when a statistical endeavor has been done well, and how such studies, even the well executed ones, may be lacking.

This course is titled “Introduction to Statistics”, and this is what it is – an introduction. You will not leave this course as a world-class statistician. I have a PhD in statistics and over 27 years working with data and statistics in industry and do not consider myself a world-class statistician. However, the desire is for each of you to leave this class with respect for the discipline of statistics (and perhaps some with a level of excitement about the subject), as well as an understanding of its limitations. In addition, you will acquire the background to be able to work with professional statisticians as a subject matter expert, including being able to more easily converse with a professional statistician, and ask pertinent questions about the approaches and assumptions being made in their efforts to help you address your question(s) of interest. Another take-away from this course would ideally be a healthy perspective on what can and cannot be learned from a specific set of data and or a statistical report attempting to summarize that data.

Text
None. The course reference materials will consist primarily of the Case Study Manual, as well as the Course Notes, both of which will be posted on D2L as discussed through the semester.

Computer Access/Skills
It will be necessary to have access to a Microsoft Office tools: PowerPoint, Word, and Excel. The Class Notes are all PowerPoint files, the Case Study Manual and most of the assignments will be Word files, and the Utilities to do calculations, the Data Sets, and some of the assignments and Class Exercises will utilize Excel files. Almost every workplace will work with these Microsoft programs, and generally expect employees to have competency in working with them when hired. Hence, gaining some experience in using it will likely be helpful beyond this course. In addition, it may be helpful to download the university version of the JMP software (available on MySFA) for work later in the semester; however, while the Case Study Manual will reference it a few times, this is certainly not necessary for this course.
Prerequisites
Students will be expected to have some basic math skills (enough to obtain an acceptable score on the Math element in the SAT or ACT), and some facility with college-level algebra will be helpful. Any familiarity with calculus would be an additional plus, but it certainly is not required.

Course Rationale
Data surrounds us from all media: radio, TV, internet, etc. Data can be collected, summarized, and interpreted as statistics. Decisions are often based on data and statistical summarizations of data. To help us better understand and live in our world, it is helpful to know something about statistics.

Across the domains of human knowledge, information is becoming more quantitative. A basic understanding of statistics is necessary to make some sense out of all the data. The use of statistics has increased in the workplace. Market research, analysis of business trends, manufacturing, and quality assurance all make use of statistical analysis. Statistics are also used both correctly and incorrectly in matters of political and public debate to achieve desired results without deliberately falsifying the data.

The course is designed to introduce the statistical methodology that might be encountered in any of the above situations or others. It is hoped that with this knowledge, the student will have some appreciation for how to critically evaluate the actual information conveyed by some of the more standard types of statistical analyses.

Course Overview
Week 1: Introduction & Case 1A – Small Sample Proportion (Binomial Distribution)
Week 2: Case 1B – Large Sample Proportion (Normal Distribution, Confidence Intervals)
    Case 2A – Small Sample Population Mean (Descriptive Statistics, t Distribution)
Week 3: Case 2B – Large Sample Population Mean (Central Limit Theorem)
    Case 3A – Two Sample Means
Week 4: Case 4A – Correlation & Regression
Week 5: Case 5A – Two Sample Proportions
The above is a general plan, how the term actually proceeds depends on many factors, but is most influenced by students asking questions in class about the assignments.

Course Student Learning Objectives
The ideas it would be expected a MATH 1342 student to successfully address at the time of course completion include:

1) Data from real-world processes exhibits variation. The strongest form of model for this variation is a distribution. As such, distributions, including those of frequent application are of importance in statistics.

2) Because the exact distribution of a population of values is rarely known, statistical science relies on sampling in order to investigate the distribution and its features.

3) We estimate the distribution of a population, or more often, specific features of populations (parameters) with appropriate and corresponding features from a sample (statistics).
4) Knowledge gained from a sample is imperfect. It is not possible to make definitive claims about populations from taking samples. It is, however, possible to provide (non-definitive) claims about populations with **margins of error** attached.

5) Non-definitive claims about populations can be sufficient for **decision making**.

6) Claims about populations must be made in recognition that the collected sample is one of many that were possible. As such, statistics **vary from sample to sample** and therefore have distributions themselves. The most important statistic whose distribution deserves study is the average or proportion (which is merely an average in disguise) of a sample.

7) Investigating the **distribution** of a statistic leads to the ability to create a **margin of error** to be provided along with any statistical estimate (confidence interval). This margin of error is intimately tied to two things: the amount of **variation present** and the **size** of sample taken.

8) Investigating the distribution of a statistic under an assumption about a parameter leads to a **null distribution**. The null distribution has information in it that allows an experimenter to assess how common it would be to observe the value of the statistic seen in the sample (**p-value**). The **p-value** facilitates a decision regarding the validity of the assumption.

9) Samples should be taken in a way that is **representative** of the population. Data from the sample should be **summarized, described, graphed** – generally - **communicated** in efficient, reliable ways.

10) Data can be of different **types**. As such, the way we describe and utilize data depends on this type. **Descriptive statistic techniques** and **inferential statistics procedures** are intimately tied to data types (students will be exposed to continuous and categorical data, to univariate and bivariate data and to statistical methods for each type).

**Departmental Course Syllabus Link:**
http://www3.sfasu.edu/math/docs/syllabi/MATH1342Syllabus.pdf

**About Assignments**
Completed assignments will need to be e-mailed to me at hendersork@sfasu.edu . There will generally be homework and/or quiz problems assigned every day. However, usually, only a subset of the problems assigned will actually be evaluated/graded.

No credit will be given for the correct answer when no work is shown, and/or no information is supplied related to how answers were obtained. Since the grading will primarily be focused on the steps and/or approaches used to reach a final solution, neatness will count. If the steps cannot be followed, or it is unclear how a specific step is reached in a given problem solution, then points will be lost.

Of the problems assigned for homework, it is highly likely only 1 or 2 (or parts thereof) will be **chosen at random** to be evaluated for each homework result. There generally will be approximately 15-20 homework assignments during the term, and your best ~70-90% of scores will be used in the calculation of the homework grade.

There will be 7 quizzes assigned throughout the course, and again, there will often be multiple problems assigned for each quiz, but only 1 or 2 problems (or parts thereof) chosen at random may
be graded for each specific quiz result. The best 5 of the quiz scores will be used in the calculation of the quiz grade.

There will be a project of some nature assigned later in the session. It will be a group project and details will be provided at a later date.

**Grading**
Final grade will be determined based on the following proportions:

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework (including Reading HW Review Quizzes)</td>
<td>30%</td>
</tr>
<tr>
<td>Quizzes</td>
<td>30%</td>
</tr>
<tr>
<td>Project</td>
<td>10%</td>
</tr>
<tr>
<td>Exams</td>
<td></td>
</tr>
<tr>
<td>Mid-Term</td>
<td>10%</td>
</tr>
<tr>
<td>Final</td>
<td>20%</td>
</tr>
</tbody>
</table>

Homework and quizzes will be expected to be e-mailed to me by 3PM of the day they are due. Homework will be graded and returned as quickly as possible.

The mid-term will cover up through Case 2A, and the final will be comprehensive. Details on how these exams will be administered will be relayed closer to the actual event.

**Attendance**
Since homework/quizzes account for 60% of the grade, and will be assigned frequently (and the policy is to **not accept late homework**), it will serve the student to do all the assigned problems by their indicated due times. This is nothing more than is going to be expected of you at any place of employment where they expect you to show up and do the work every day.