COURSE: Stat 1312 – Statistical Literacy (3 – 3 – 0)

CATALOG DESCRIPTION: An introduction to statistical ideas and statistical reasoning, and their relevance to modern life. STAT 1312 is designed for students requiring one college-level mathematics course. This course cannot be applied toward any degree in the Department of Mathematics and Statistics.

PREREQUISITE: A grade of C or better in MATH 0300, TSI MATH score 343, TSI MATH complete, or TSI MATH exempt, and credit or enrollment in ENG 1301.

AUDIENCE: This is a freshman-level statistics course, which requires a background consisting of two years of high school mathematics or MATH 0300. The course is primarily intended for majors in liberal arts, and social and behavioral sciences. The course has more reading and writing than a traditional freshman statistics course, and has an emphasis on ideas and reasoning.

PURPOSE: This course satisfies the general education core mathematics requirement by elevating the student’s quantitative literacy to college-level by introducing statistical ideas and statistical reasoning that demonstrate the broad usefulness and importance of statistics to modern life.

LEARNING OUTCOMES: At the completion of the course, the student should be able to:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Describe the utility and limitations of data and statistical methods for solving real-world problems.</td>
</tr>
<tr>
<td>2.</td>
<td>Describe methods of data collection and explain potential pitfalls, biases, and ethical issues.</td>
</tr>
<tr>
<td>3.</td>
<td>Demonstrate facility with the terminology, notation, and numerical methods generally found in an introductory statistics course, such as graphical summaries, measures of central tendency and dispersion and use these to compare data sets.</td>
</tr>
<tr>
<td>4.</td>
<td>Demonstrate a basic understanding of probability and how it relates to statistics, particularly the ideas of randomness and statistical significance.</td>
</tr>
<tr>
<td>5.</td>
<td>Demonstrate facility with basic inferential statistical methods, including confidence intervals and hypothesis tests, and understand when a particular method is appropriate, and the ability to interpret results in the context of a stated problem.</td>
</tr>
<tr>
<td>6.</td>
<td>Demonstrate facility to use appropriate technology such as statistical software and/or calculators.</td>
</tr>
<tr>
<td>7.</td>
<td>Formulate and communicate solutions to statistical problems in clear, grammatically correct, precise English.</td>
</tr>
</tbody>
</table>


STATISTICAL SOFTWARE: Students will be required to use statistical software such as CrunchIt, StatCrunch or similar software.
GENERAL EDUCATION CORE LEARNING OUTCOMES:
(THECB Objectives: Communication, Critical Thinking, Empirical/Quantitative Reasoning)
At the completion of the course, the student should be able to:

1. Describe and communicate mathematical information verbally, numerically, graphically, and symbolically.
2. Use appropriate mathematical techniques to model situations from a variety of settings, including real-world applications in generalized mathematical forms.
3. Interpret mathematical models, such as formulas, graphs, tables, and schematics, and draw inferences from them.
4. Discern relationships and patterns in quantitative data to arrive at informed conclusions.
5. Utilize appropriate technology to enhance mathematical thinking and understanding, to solve mathematical problems, and to judge the reasonableness of the results.

GENERAL EDUCATION CORE LEARNING OUTCOMES VS. COURSE LEARNING OUTCOMES

<table>
<thead>
<tr>
<th>Core Outcome</th>
<th>Course Outcome 1</th>
<th>Course Outcome 2</th>
<th>Course Outcome 3</th>
<th>Course Outcome 4</th>
<th>Course Outcome 5</th>
<th>Course Outcome 6</th>
<th>Course Outcome 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

METHOD OF EVALUATION: Departmental policy requires that:
1. A maximum of four and a minimum of three in-class tests and a comprehensive final exam must be given. The final exam must be taken by all students.
2. All major tests should be announced at least one week, or the equivalent, in advance.
3. The final exam counts 1/3 of the course grade.
4. The final course average will be used to assign the final course grade according to the standard college formula:
   - 90-100 → “A”
   - 80-89 → “B”
   - 70-79 → “C”
   - 60-69 → “D”
   - 0-59 → “F”
5. Neither an open book nor a take-home major test may be given.
6. The instructor may assign a short paper to count the same as an in-class exam.
7. Instructors must include in the course a peer-interview activity and post-interview quiz to evaluate the activity, and count this grade as part of the course grade. This activity is designed to satisfy the oral communication learning outcome. During the activity students should explain to each other orally their solution to a statistical problem and/or concepts from a particular topic in statistics. Choice of a topic and the specifics are left to the instructor (see example in the Appendix). The expectation is that students should be able to communicate statistics using correct terminology and sound reasoning. Students should be able to express themselves accurately and follow logical steps in their solutions. Students must not only give a correct solution but they must be able to describe their thinking process to their peers.

8. Instructors are strongly encouraged to require homework assignments and in-class group activities in the course, and to count these grades equivalent to a regular test grade. A similar version of a test may not be distributed to students before a major test. Any review sheets should be comprehensive. Students should not feel that classroom notes, homework or the text may be ignored in favor of review sheets.

SUGGESTED INSTRUCTIONAL METHODS: It is helpful in this class to solve meaningful application problems and give problems in context wherever possible. Within the stated course objectives, instructors may experiment with various methods and share their results with the department. Suggestions include: have students write explanations of answers and methods where possible; have students work in groups during class when appropriate; have students do activities with computers and/or calculators where appropriate. Classroom kits of graphing calculators are available from the department. Instructors should assign a short (4-5 page) typed paper or oral presentation as part of the course requirements. The topic of such a paper may be an article that fits the student's individual interests and that cites summarized statistics. The student will dissect the article into relevant course-related components and relevant statistics.
COURSE CONTENT:

<table>
<thead>
<tr>
<th>UNITS WITH APPROXIMATE TIME</th>
<th>TEXT REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unit I – Producing Data</strong> (10 hours)</td>
<td>Chapters 1 - 6</td>
</tr>
<tr>
<td>Topics or techniques to be covered include: Where Do Data Come From; Samples, Good and Bad; What Do Samples Tell Us; Sample Surveys in the Real World; Experiments, Good and Bad; Experiments in the Real World.</td>
<td></td>
</tr>
<tr>
<td><strong>Unit II – Organizing Data</strong> (10 hours)</td>
<td>Chapters 10 - 15</td>
</tr>
<tr>
<td>Topics or techniques to be covered include: Graphs, Good and Bad; Displaying Distributions with Graphs; Describing Distributions with Numbers; Normal Distributions; Describing Relationships: Scatterplots; Describing Relationships: Regression.</td>
<td></td>
</tr>
<tr>
<td><strong>Unit III – Chance</strong> (10 hours)</td>
<td>Chapters 17, 18, 21</td>
</tr>
<tr>
<td>Topics or techniques to be covered include: Thinking About Chance; Probability Models; What is a Confidence Interval.</td>
<td></td>
</tr>
<tr>
<td><strong>Unit IV – Inference</strong> (4 hours)</td>
<td>Chapters 22 - 23</td>
</tr>
<tr>
<td>Topics or techniques to be covered include: What is a Test of Significance; Use and Abuse of Statistical Inference</td>
<td></td>
</tr>
</tbody>
</table>

RESOURCE MATERIALS: Students enrolled in STAT 1312 at UHD have access to the Center for Math & Statistics Support (C4MS²) in the Academic Support Center (N925) where they may receive additional tutoring on understanding concepts or improving their skills. The C4MS² is staffed with mathematics faculty and student assistants, and offers tutorial help, videos, calculators, and computers with web access. The Center maintains extensive hours which are published each semester by the department. Students are encouraged to visit the C4MS². No appointments are necessary.

GENERAL UNIVERSITY POLICIES: All students are subject to UH-Downtown’s Academic Honesty Policy and to all other university-wide policies and procedures as they are set forth in the UH-Downtown University Catalog and Student Handbook.

STATEMENT ON REASONABLE ACCOMMODATIONS: UH-Downtown complies with Section 504 of the Rehabilitation Act of 1973 and the Americans with Disabilities Act of 1990, pertaining to the provision of reasonable academic adjustments/auxiliary aids for students with a disability. In accordance with Section 504 and ADA guidelines, UHD strives to provide reasonable academic adjustments/auxiliary aids to students who request and require them. If you believe that you have a documented disability requiring academic adjustments/auxiliary aids, please contact the Office of Disability Services, One Main St., Suite 409-South, Houston, TX 77002. (Office) 713-226-5227 (Website) [www.uhd.edu/disability/](http://www.uhd.edu/disability/) (Email) [disabilityservices@uhd.edu](mailto:disabilityservices@uhd.edu)
APPENDIX: Example of a Peer-Interview Activity – STAT 1312

Directions: Assign students to work in pairs (this will be better with two students but could accommodate three). Assign each student a problem from the course—the problem and topic should be up to the instructor. Each student is to prepare to present their strategy and solution to the problem to their partner(s). The presentation should take no more than 20 minutes total and students can present the problem to each other outside of class time if necessary. After their presentations to each other, each student will fill out a peer-interview assessment. The entire activity—the presentation and the assessment—should count for some portion of the course grade.

Here is one example:

STAT 1312: The instructor decides to assign central tendency problems for the peer-interview activity.

Consider the following numbers.
2,  3   4,   5,   5

All students should compute the mode, median and mean.

Student 1: If the numbers are codes for the colors of T-shirts ordered from a catalog, which average is most appropriate? Explain your answer.

Student 2: If the numbers are one-way mileages for trails to different lakes, which average is most appropriate? Explain your answer.

Student 3: If the numbers are survey responses from 1 to 5, with 1=disagree strongly, 2=disagree, 3=agree, 4=agree strongly, 5=agree very strongly, which average is most appropriate? Explain your answer.

Purpose of activity: This activity illustrates the expectation that students should be able to communicate statistics using correct terminology. They should be able to discuss statistics accurately and follow logical steps to get to their solution. Students should be to be able to do more than just give the correct solution to a problem — they must also be able to communicate, in written and oral form, their knowledge of statistics and describe their thinking process to their peers.
Peer-Interview Assessment – STAT 1312

A. Rate your partner’s use of statistical terminology.
   1. Used words incorrectly
   2. Mostly used words correctly
   3. Used words correctly

B. Rate your partner’s explanation of his/her solution to the problem.
   1. I could not follow it at all
   2. I understood most of what he/she was saying
   3. The solution was clear and easy to follow

C. Rate your partner ability to answer your questions in a helpful way.
   1. He/she could not answer all of my questions
   2. He/she answered most of my questions satisfactorily
   3. He/she was able to answer all of my questions thoroughly

D. Rate your partner’s presentation of his/her solution to the problem.
   1. Inadequately
   2. Adequately
   3. Superior
APPENDIX: Examples of Extended Group Activities – STAT 1312

The sample problems listed below illustrate some of the ways the oral, visual, reading and writing components are imbedded in the statistical literacy course, STAT 1312. The emphasis in this course is on ideas, reasoning, and interpreting the results of statistical calculations. Students will be given problems sets in class and time to read the problem, analyze the situation presented, and come up with oral and written response to the discussion questions. At the end of the time allotted, one person from each group will give a short summary of the question and report on the group’s discussion. A short question and answer period in the class will follow each presentation. The grades on these assignments will be part of the overall grade in the course. During the semester, each student will be required to be the facilitator of the question and answer period, the scribe, and the oral presenter for the discussion questions.

1. **Ask Marilyn. Parade Magazine, 4 April 2004, Marilyn vos Savant.**

Marilyn issues a correction to her column from May 27, 2001, in which a reader asked: My husband takes more than 100 flights a year for his job. Although I know that the odds of an accident are the same for every flight, I still think the frequency of the flights should matter. Isn't it at least somewhat more likely that he would be involved in an accident than a person who flies only once or twice a year? I really would like you to agree with me because he maintains that this is not an actual risk for him.

**In the earlier column, Marilyn had replied as follows:**

I agree with both of you. If your husband flies 100 times a year, he runs 100 times the risk of a person who flies only once a year. On the other hand, his risk is still statistically insignificant. Flying is amazingly safe.

Marilyn now acknowledges an error in the above that was recently called to her attention by another reader. She writes:

The sentence should have been phrased this way: "If your husband flies 100 times a year, he runs the risk of a person who flies only once a year 100 times." (The words "100 times" are misplaced.)

**DISCUSSION QUESTION:**

(a) Do you understand the difference between the two calculations implied here?

(b) Dr. Arnold Barnett of MIT calculated that during the 10-year period from 1987-1996, the death risk per flight on established US domestic jet carriers was 1 in 7 million. (reference INFORMS online.) What would the resulting calculations yield?


In their article, Garcia-Berthou and Carles Alcaraz investigated 32 papers from editions of Nature published in 2001, and 12 papers from the British Medical Journal in the same year. The authors checked the calculations of the p values. They also looked at the distribution of the digits of numbers used in the tests. They found that 38% of those sampled from Nature, and 25% of those sampled in the British Medical Journal contained one or more statistical errors. The authors concluded that 4% of the errors may have caused non-significant findings to be misrepresented as being significant. As the article suggests, these errors would be more likely to be noticed by the reviewers and others if authors were
required to make the raw data available. Maxine Clarke, publishing executive editor of Nature remarked: We do not explicitly ask authors, as routine, for the raw data underlying their reported statistical results. This suggestion is now on the agenda for our next editorial meeting on editorial practices and criteria. Kamran Abbasi, deputy editor of the British Medical Journal said: We certainly do not spend our time recalculating all these numbers, and our whole review process would likely grind to a halt if we tried to do so. The BMC Medical Research Methodology journal is an "open source" journal.

DISCUSSION QUESTION:

(a) Do you think that journals should require their authors to provide the raw data for their study?

(b) The Dartmouth Committee for the Protection of Human Subjects requires that, before a study involving humans is carried out, the authors must submit a proposal which provides, among other things, an explanation of the study procedures and data collection and analysis process. Would this information be useful in evaluating the results of the study? If so, why?

3. **Boston inspects but doesn't grade restaurants. Boston Globe, 8 February 2000, E3, Bruce Mohl.**

As reported in Chance News 7.06, the article "Dining out in L.A. Comes to Crunching Numbers" discussed a numerical rating system being used by inspectors to give letter grades to Los Angeles restaurants (A for 90-100, B for 80-89, etc.). By law, the grades must be displayed in restaurant windows; they can also be viewed on a county web site. The present article reports that, when the system was adopted in 1988, only half the restaurants got A grades. Today, 75% make A's.

Boston's restaurant inspections also result in a point score, which could in principle be converted to a letter grade. However, patrons get only pass/fail information. A restaurant failing inspection is closed down until the problems are corrected, usually only a matter of days. A pass grade certifies that it is safe to eat there, and the mayor's office argues that this is all the public needs to know. But Dr. Jonathan Fielding, the former Massachusetts commissioner of public health who now holds that title in L.A. county, disagrees. He says that B or C grades can discourage business, which gives restaurants an incentive to improve to A's rather than being content merely to keep their doors open.

Steven Grover of the National Restaurant Association notes that grades are attractive as a "quick fix" to public health concerns but warns that many components of inspections are subjective. Furthermore, just reporting a final score does not tell consumers what problems actually exist. Orange County, California has found a middle ground position. Restaurants display notices indicating that they passed their last inspection, and copies of the full report must be available to patrons on request.

DISCUSSION QUESTIONS:

(a) The Globe visited 20 well-known Boston restaurants and found that the average score from their most recent inspection was 79.7. According to the article "that would have been on the border between a B and C by Los Angeles standards." Are
you convinced that this group is worse than most L.A. restaurants?

(b) When visiting a restaurant, do you think you would be inclined to request a copy of its inspection report? What kind of information would lead you to leave without eating?

4. The impact of No Child Left Behind (graphic). *New York Times, 17 August, 2004*

This data graphic appeared in the *Times*. It is intended to compare two variables at the state level: percentage of schools facing penalties under the No Child Left Behind Act, and percentage with fourth graders not meeting the basic reading standard.

**DISCUSSION QUESTIONS:**

(a) Do you think the graphs make the intended comparison clear? Can you suggest another way to present the data?

(b) What consequences might there be in aggregating the data in this way?

5. In defense of the Harvard Nurses' Health Study. *Washington Post, 8 February 2000, Z4 Letter from Walter Willett, M.D. Harvard School of Public Health*

Food surveys have wide margins of error; Researchers know that questionnaire results don't always reflect actual eating habits. The Washington Post, 1 February 2000, Z9, Lawrence Lindner

The news media frequently feature stories on the latest link found between diet and disease. The controversy reported here concerns Harvard University's Nurses' Health Study. Dr. Willett was concerned by the tone of the Post article which criticized the methodology of food surveys in general and the Harvard study in particular. The Post cited a number of reasons to doubt the quality of data produced by food surveys. One was the complexity of some of the questions asked. The article cites the following example from
the Harvard study:
How often, on average, did you eat a quarter of a cantaloupe during the past year? One to three times a month? Once a week? Two to four times a week? Once a day? Please try to average your seasonal use over the entire year. For example, if cantaloupe is eaten four times a week during the approximately three months it is in season, then the average use would be once a week.

Even when they understand the questions, people may not accurately report their eating habits. Writing in the American Journal of Clinical Nutrition, psychologist John Bludell cited a recent study finding that obese men under-reported the calorie intake by 36%. He expects that, if people are told fat intake is bad, they will similarly begin to under-report that. For their part, the news media do not do a good job distinguishing between correlation and causation. For example, a recent Washington Post headline announced "Study Links Hot Dogs, Cancer: Ingestion by children Boosts Leukemia Risk, Report Says." The article began: "Children who eat more than 12 hot dogs per month have nine times the normal risk of developing childhood leukemia..." Furthermore, since data for the study came from parents recollections of their children's eating habits, they are subject to the same concerns cited above. Nevertheless, the article does not conclude that food surveys are worthless. According to nutrition research James Fleet of the University of North Carolina, such research "generates new hypotheses to test in more controlled settings."

In his letter to the editor, Dr. Willett noted that the cantaloupe question was presented out of context in the article. He points out that researchers designing questionnaires give careful consideration to the ordering and wording of questions. Furthermore, he points out that researchers do not blindly accept survey data. Efforts are made to validate reported results with already established dietary risk factors. For example the ratio of polyunsaturated to saturated fat intake is known to be related to cardiovascular disease. The fact that this relationship has been confirmed in the Nurses' Heath study gives additional confidence in the reporting. Finally, the fact that the data have been recorded over a 20-year period strengthens conclusions about long-term effects of diet.

DISCUSSION QUESTIONS:

(a) Do you understand how the average seasonal use of once a week was computed for the cantaloupe example? Estimate your own average consumption of cantaloupe.

(b) Do you agree that the hot dog headline is too sensationalistic? How would you rewrite it to be more accurate? What about the lead sentence?
Appendix: Sample Problem Set for STAT 1312 – Statistical Literacy

These sample problems are meant to illustrate that STAT 1312 has more reading and writing than a traditional freshman statistics course and that the emphasis is on ideas and reasoning.

Producing Data

1. Below is language used by the Zogby Poll to explain the accuracy of a recent poll: “The margin of error is plus or minus 1.2 percentage points. Zogby International’s sampling and weighting procedures also have been validated through its political polling: more than 95% of the firm’s polls have come within 1% of actual election-day outcomes.” What does Zogby mean by “95% of the firm’s polls have come within 1% of actual election-day outcomes”?

2. An article in the Washington Post reported that, according to two large studies, obese people are significantly less likely to die prematurely if they undergo stomach surgery to lose weight. But people choose whether to have stomach surgery. Explain why this fact makes any conclusion about cause and effect untrustworthy. Use the language of lurking variables and confounding variables in your explanation and draw a diagram to illustrate this.

Organizing Data

1. A survey of college freshmen in 2007 asked what field they planned to study. The results were: 12.8% arts and humanities; 17.7% business; 9.2% education; 19.3% engineering, biological sciences, or physical sciences; 14.5% professional; and 11.1% social science.
   (a) What percentage of college freshmen plan to study fields other than those listed?
   (b) Make a graph that compares the percentages of college freshmen planning to study various fields.

2. The mean and standard deviation measure center and spread, but are not a complete description of a distribution. Data sets with different shapes can have the same mean and standard deviation. To demonstrate this fact, use your calculator or statistical software to calculate the mean and standard deviation for these two small data sets. Then make a stemplot of each data set and comment on the shape of each distribution.

```
Data B  6.58  5.76  7.71  8.84  8.47  7.04  5.25  5.56  7.91  6.89  12.50
```

Chance

1. Results from a summer 2010 Gallup Student Survey reveal that 37% of students in the 5th through 12th grades feel either not engaged or actively disengaged in school. The report of this sample survey of 642 students aged 10 to 18 stated that, with 95% confidence, the margin of sampling error was ± 4.94%. Explain to someone who knows no statistics what the phrase “95% confidence” means in this report. Make sure you
write your answer in context.

2. The probability that a randomly chosen driver will be involved in an accident in the next year is about 0.2. This is based on the proportion of millions of drivers who have accidents. “Accident” includes things like crumpling a fender in your own driveway, not just highway accidents.

(a) What do you think is your own probability of being in an accident in the next year? This is a personal probability.

(b) Give some reasons why your personal probability might be a more accurate prediction of your “true chance” of having an accident than the probability for a random driver.

(c) Almost everyone says that his or her personal probability is lower than the random driver probability. Why do you think this is true?

Inference

1. The national unemployment rate in a recent month was 9.0%. You think that rate may be different in your city, so you plan a sample survey that will ask the same questions as the Current Population Survey. To see if the local rate differs significantly from 9.0%, what hypotheses will you test?

2. How much education children get is strongly associated with the wealth and social status of their parents. In social science jargon, this is “socioeconomic status,” or SES. But the SES of parents has little influence on whether children who have graduated from college go on to yet more education. One study looked at whether college graduates took the graduate admissions test for business, law, and other graduate programs. The effects of the parents’ SES on taking the LSAT test for law school were “both statistically insignificant and small.”

(a) What does “statistically insignificant” mean?

(b) Why is it important that the effects were small in size as well as insignificant?