2018

STAT 380: Statistics and Applications

Yumou Qiu
University of Nebraska-Lincoln, yumouqiu@unl.edu

UNIVERSITY OF NEBRASKA LINCOLN

Stat 380: Statistics and Applications

Yumou Qiu

Department of Statistics

5/27/2018

Abstract: This portfolio prepares for the introduction level undergraduate statistic course (Stat 380) for the students with calculus background, which includes the course goals, contents as well as teaching methods for large session classes. Analysis and evaluation of student learning are also included. In-class activities and discussion helping students' engagement are introduced. The homework is assigned by Canvas, which summaries the students' solution and provides analysis for each student.
Key words: Canvas; Online quiz/homework; Statistics; Students performance assessment; Teaching portfolio.

Contents

Objectives of Peer Review Course Portfolio

The author's objectives

The STAT/MATH 380: Introduction to Probability and Statistics Concepts is the second-level introduction course in "undergraduate statistics and mathematics" sequence. The goal of this portfolio is to (1) provide in-class activities and discussions that arouse students' interest in statistics, enhance their in-class engagement and help them to understand the course context; (2) discuss teaching methods that are suitable and efficient for large session classes; and (3) discuss evaluation methods for students' performance.

Description of the Course

Course description

The target students of this course are the ones with college mathematics background who have few or no statistic courses before. This is an introduction level statistic course for mostly the mathematic, computer science and engineer undergraduates. The prerequisite course is MATH 107 "Analytic Geometry and Calculus II".

The course provides an introduction to probability concepts, i.e., random variables, probability distributions, expectation, variance, covariance, correlation; and statistical concepts, i.e., fundamental sampling distributions and data descriptions, one- and two-sample estimation and testing problems, and simple linear regression. This course lays the foundation for many 400-level courses in Engineering, Mathematics and Statistics, and is also offered for Honors credit.

This course is also accredited as Achievement Centered Education (ACE) course and satisfies ACE outcome 3: to use mathematical, computational, statistical, or formal reasoning (including reasoning based on principles of logic) to solve problems, draw inferences, and determine reasonableness. Therefore, the reinforced skill for this course is critical/statistical thinking. The student will obtain the opportunities to achieve this learning outcome through readings, quizzes, homework and in-class activities.

Course Goals

First, students will learn the core basic probability and statistics concepts necessary for more advanced courses; second, students will learn the role of probability and statistics in the modern life; third, students will understand the logic of statistical thinking and reasoning; and fourth students will be able to apply some simple statistical inference methods to solve some real problem in everyday life and the research questions in their disciplines. More specifically,

1. Students will understand essential theoretical concepts (and logic behind them) including probability definition, random variables, common discrete and continuous probability distributions, sampling distributions and data description techniques, hypothesis testing problems, confidence intervals, modeling approaches and interpretation on the results.
2. Students will be able to state a hypothesis testing (or estimating/confidence interval) problem (out of the ones considered in the course) for a given research question, perform testing (construct the corresponding confidence interval) and draw statistical inferences.
3. For a research question and a relevant data set, students will be able to fit an appropriate statistical model (including simple linear regression, one-sample or two-sample models, binomial model for proportions), conduct statistical inference, interpret the obtained results in the context of the problem, and make conclusions.

Context

The course context covers preliminary probability concepts in the first half part of the course, which includes definition of probability, counting theories, conditional probability, random variables, discrete and continuous probability distributions, and statistical moments like expectation, variance, covariance and correlation between two random variables. The second half of the course covers fundamental statistical concepts. It includes sampling distributions based on random samples of a population, one-sample and two-sample statistical estimation and inference for continuous means and categorical proportions, and simple linear regression.

Enrollment

The enrollment of the Spring 2018 class is 170. Most of the students are from the departments of math, actual science and computer science. Some of the students are from animal science and plant science departments. The students have some calculus background. The actual science students will take more advanced statistical courses Stat 462 and 463 in the following semesters. This course serves as a pre-requisite for those two courses.

Teaching Methods

Teaching methods

The course is taught by lectures and in-class discussions and activities. Incomplete slides will be posted on Canvas before the class, which show the main topics of the next lecture, and help the students to prepare for the class. Sample slides are given in Figure 1. Some key parts of the slides, such as the properties of certain probability rules and statistical methods and the solutions to some examples, are left as blank, which will be filled in during the lecture.

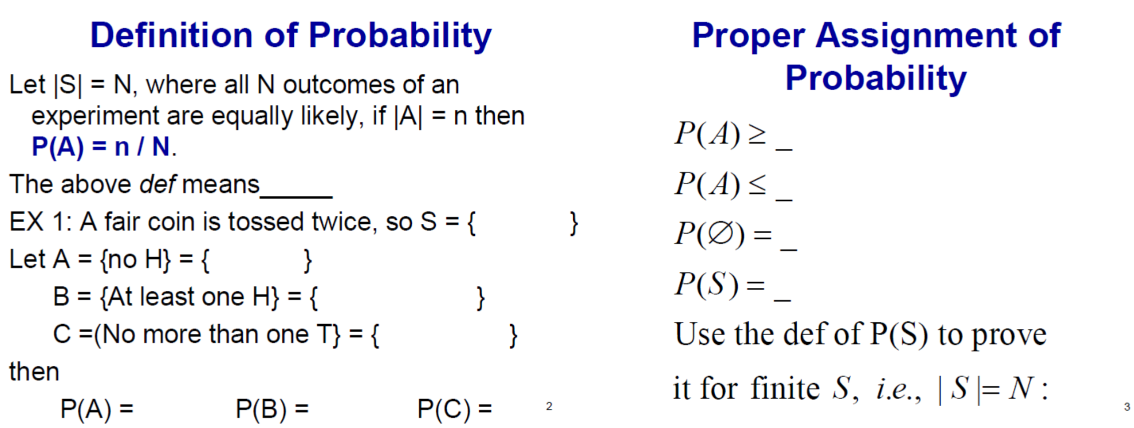


Figure 1: sample slides for probability theories.

The class will be divided into groups of 3 to 4 students, which will result in 45 to 50 groups in the large session. Those blanks in the slides are usually used as questions for students to discuss within group first during the class. And then, between group discussion will be conducted. Several hand-on activities will also be taken through the semester. See the course materials in the next subsection for details. As a statistics course, those activities usually involves random sampling in some ways. A merit of group activities and discussion is that each group may take different samples, and have different statistic results based on those samples. Although those results are different in values, they may reach the same conclusion by accounting for the randomness of the sample. This helps the students to understand the concepts of sampling distributions and statistical inference procedures like hypothesis testing and confidence intervals.

Through the semester, we have two midterm exams, and one final exam. The first exam mainly covers Chapters 1 to 4 for the basic concepts of probabilities. The second exam is for Chapters 5 to 6, which are about the concept and definition of random variables, and their associated distributions. Several specific continuous and discrete distributions are the key points for exam 2. The final exam is mainly for the rest of the chapters for statistical inference, which includes sampling distributions, hypothesis testing, confidence intervals and simple linear regression. All the exams are weighted equally. Several in-class short quizzes are given during the lecture. The quizzes assess students' understanding of the key concepts. The common mistakes of the quizzes will be emphasized in the next lecture.

Homework is assigned through Canvas. The written homework is not required for submission. Instead, a homework quiz will be posted on Canvas, which needs to be completed within 30 minutes. The students have to finish the homework first before taking the quiz. The homework quiz is usually in the form of multiple choice questions. Canvas will automatically grade the quiz, and summarize the mistakes for the whole class. It will provide the correctness percentage and the percentages for each wrong choices for every question. This makes the instructor easily track the students learning. Similar as the in-class quizzes, the common mistakes will be discussed and explained in the class, and similar questions can be assigned in the next homework to assess the improvement of certain concepts.

Course materials and rationale

In this subsection, we will introduce an in-class activity which is used several times through this course for multiple purposes. This activity is about President Lincoln's Gettysburg address. The full address and several of the questions, as well as the goals and rationales of discussing those questions are provided for illustration.

President Abraham Lincoln's Speech

The Gettysburg Address, 1863

> *Four score and seven years ago, our fathers brought forth upon this continent a new nation: conceived in liberty, and dedicated to the proposition that all men are created equal.*
>
> *Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived and so dedicated, can long endure. We are met on a great battlefield of that war.*
>
> *We have come to dedicate a portion of that field as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this.*

But, in a larger sense, we cannot dedicate, we cannot consecrate, we cannot hallow this ground. The brave men, living and dead, who struggled here have consecrated it, far above our poor power to add or detract. The world will little note, nor long remember, what we say here, but it can never forget what they did here.

It is for us the living, rather, to be dedicated here to the unfinished work which they who fought here have thus far so nobly advanced. It is rather for us to be here dedicated to the great task remaining before us, that from these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion, that we here highly resolve that these dead shall not have died in vain, that this nation, under God, shall have a new birth of freedom, and that government of the people, by the people, for the people, shall not perish from the earth.

First, Gettysburg address can be used to illustrate the ideas of population, sample and simple random sampling. The question is:

We aim to estimate the average word length of Gettysburg address. Randomly sample 10 words from this address, and record your selected words in the following table. What is the population and sample? Calculate the average of your sample. Record your data:

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Word										
# letters										

For this activity, we are considering all the words in Lincoln's Gettysburg Address as the population, and the 10 words you selected are considered a sample from this population. In this case, we know the population, since we can count the word length for each word in the address. In most studies, we do not have access to the entire population and can only study for a small sample from that population. This question about the population and sample can help students understand the fundamental goal of statistics, using the information of the sample to learn the population information. In practice, we want to study the population, but we only have access to a small part of the population.

This arouses the question of how to choose a representative sample for the population. The population average of word length is 4.295. How to unbiasedly estimate this population mean? We ask each student to choose a sample of 10 words and calculate the sample average. Then, the students are required to report their sample means on a histogram on the backboard. The students will find out the center of their histogram is significant over the true population mean, which implies most of the students over-estimate the population mean. This is due to the sample obtained by the students are a convenient and biased sample, since those long words are more “attractive”, and more frequently chosen comparing to the short and simple words. Then, simple random sampling, an efficient method to obtain a representative sample for a target population is introduced as a way to solve this problem. Through this activity and discussion, the students can have a better and deeper understanding of biased samples and how to acquire an unbiased and representative sample.

Second, we can introduce the idea of sampling distribution through this activity. We require the students to sample 10 words randomly from this speech. The question is as follows.

Randomly sample 10 words from Lincoln’s Gettysburg Address (use the sample frame in a separate sheet). Record the word you sampled and the length of those words. In this activity, we try to estimate the population mean of the word length of Gettysburg Address by using the sample you obtained. Calculate the sample mean and sample standard deviation. Is your sample mean the same as those of your partners? If not, why are they different?

The students will obtain different samples, which give different sample means. Through discussion with their partners, the students have a real example of statistic randomness. By plotting the sample mean in a histogram on the blackboard, the concept of sampling distribution, the distribution of the sample mean in this example, is introduced. This sampling distribution is bell shape, normal distribution like. By comparison with the population distribution, the distribution of the word length in the Gettysburg Address, the students can see their commons and differences. This helps them to understand the distribution of statistics, and how we use sample mean to estimate the population mean. The concept of sampling distribution is hard to understand, and it is also the key point for statistical inference in the sequence of the undergraduate statistic courses. Through the activity and discussion, the students have a hand-on experience of the randomness of statistics and a real example to illustrate the idea of statistical inference.

This activity can also be used in the learning of hypotheses testing and confidence intervals. We only explain the questions related to understanding and interpreting confidence intervals here.

Assume the population word length follows a normal distribution. Calculate the 90% confidence interval for the population mean. Does your confidence interval include the true population mean?

First, this question is a practice for the students to calculate the confidence interval of the population mean, which relates to the sampling distribution of sample means in the previous questions. Second, since we know the true mean is 4.295, the student can see whether their confidence intervals include the true mean. Due to the randomness of the sample and the choice of the confidence level, about 90% of the confidence intervals will cover the true mean. This shows that the statistical results may not be always correct. There is a possibility of making wrong conclusions, either Type I or Type II errors. This possibility is related to the confidence level in this question or the significant level in hypotheses testing problems.

This question also helps the students to understand the meaning of confidence level and how to interpret the result of the confidence interval. Since about 10% of the students may obtain confidence intervals which don't cover the true population mean, the 90% confidence level should be interpreted as "if we conduct this experience a large number of times repeatedly, about 90% of the times, we will obtain a confidence interval covering the true population mean". Of course, in real applications, we don't know the population mean, which is our target to estimate. We could also interpret the confidence intervals as "we are 90% confidence that the obtained confidence interval based on our sample will cover the true population parameter of interest".

This Gettysburg Address activity can be used to illustrate several key concepts of this course, ranging from representative data acquisition to sampling distributions and statistical inference. By re-visiting this activity from time to time, the students can see the connections between each chapter, and have a deeper understanding of those important concepts.

Illustration of changes

Compared to the previous semesters, this is an extra-large session of Stat 380. There are two major changes. For the teaching method, several hand-on activities for data collection and analysis are implemented, for example, the Gettysburg Address activity introduced in the previous sub-section. Through those activities, students have hand-on experience and a better understanding for many of the difficulty and important concepts of statistical learning. For student assessment, a new online homework quiz system is setup. The online homework provides feedback for each student. The instructors can re-direct the focus of a lecture based on the online assessment. Both overall assessment for the whole class, and the comparison and analysis of student learning over time can be conducted. Given the extra-large session of students, the latter analysis is essentially useful to study the performance of a particular student, to see his/her engagement, improvement and weakness, and provide advice for future learning.

Broader curriculum

Stat 380 is a calculus based introduction level statistic course for students who are not major in statistics. It is a required course for many undergraduate programs at University of Nebraska Lincoln, like mathematics, computer science, engineering, animal science and actual science. Its twin sister, Stat 218 “Introduction to Statistics”, is another introduction level statistic course. The main difference is that Stat 218 does not require calculus background. Stat 380 covers some fundamental probability theories that include continuous random variables, their probability density functions, and bivariate joint distributions. The math background on double integration is a pre-requisite for understanding probability theories.

We are embracing the era of big data. With the advance of modern technology, data acquisition and quantitative analysis become more and more important in many scientific and social studies, like the areas of social network, bioinformatics, agronomy and animal science. Stat 380 also introduces some elementary statistical inference procedures, which include one-sample and two-sample inference for means and proportions, and simple linear regression. For the students in related fields, the skills to conduct statistical analysis are vital to the success of their future career. Stat 380, as the first statistical course in the college study, plays an important role in arousing and culturing the interest of students in data science. If the students can see the importance and understand the rationale of statistic thinking, they are more likely to apply the statistical methods learned in the class to their fields of study.

Stat 380 is also the pre-requisite for Stat 462 (Introduction to Mathematical Statistics I: Distribution Theory) and Stat 463 (Introduction to Mathematical Statistics II: Statistical Inference). Both of the two courses are required for actual science undergraduate program at UNL. Stat 380 introduces the basic probability theories and statistical inference. Stat 462 covers a more complete set of distribution functions and general theories for probability. Stat 463 introduces the theories for statistical inference that include maximal likelihood estimation, sufficient and complete statistics, uniformly most powerful tests, and etc. Both those two courses prepare the actual science undergraduates for their actuarial credentialing and exams. Learning Stat 380 well is the key to success in the following courses.

Analysis of Student Learning

Analysis of particular students and assignments

The grades for homework, quizzes, attendance and exams are recorded on Canvas. Analysis for each student can be generated by Canvas, which provides the quantile of each assignment/quiz for every student. Figure 2 shows the grade distributions for three selected students. The first 6 assignments are for homework, the quizzes are recoded in the 7th, 8th, 9th, 15th and 16th columns, the 10th, 20th and 21st are three exams with total score 100, the rest of the recordings are attendances, and the last one is the final cumulative score, out of 100. For each assignment, the boxplot of all the scores over the whole class is plotted, which shows the grade distribution for that assignment. The score of the target student is highlighted by a point, where the green circle, yellow triangle and red square represent good, fair and poor scores, respectively.

Figure 2: Grade distributions for three selected students.

As shown in Figure 2, the first student did fairly well in all the homework, quizzes and exams, and attended all the lectures. Her score was ranked among the top of the class. The second student also did well in the class, although he had one poor quiz score, and the final exam was not as good as the previous two midterm exams. Since the final exam is mainly on statistical inference and first two are about probability theories, this indicates the student failed to understand the logic of hypotheses testing and confidence intervals very well. Note that this student had a well attitude to learning. He attended all the lectures without a missing. Additional practice and explanations should help him in understanding this part of the course. The last student got fair scores toward the lower side on all the three exams, which were not as good as the previous two students. The poor scores on most of the homework and quizzes indicate the third student struggled to manage this course through the whole semester. However, through the practicing on the assignments and attending all the lectures, his performance on the exams were improved.

The homework is submitted through Canvas, which can be graded automatically. The following figure shows a question answered incorrectly by a student. As shown in Figure 3, comments can be provided for each incorrect choice. The multiple choices are designed to reflect the common mistakes. Although we strongly encourage the students to ask questions about their homework and quiz mistakes, few students come to the office hour or contact with the instructors and TAs. Those automatic feedbacks help the students to understand which part of their solution is wrong, and figure out how to obtain the correct answer.

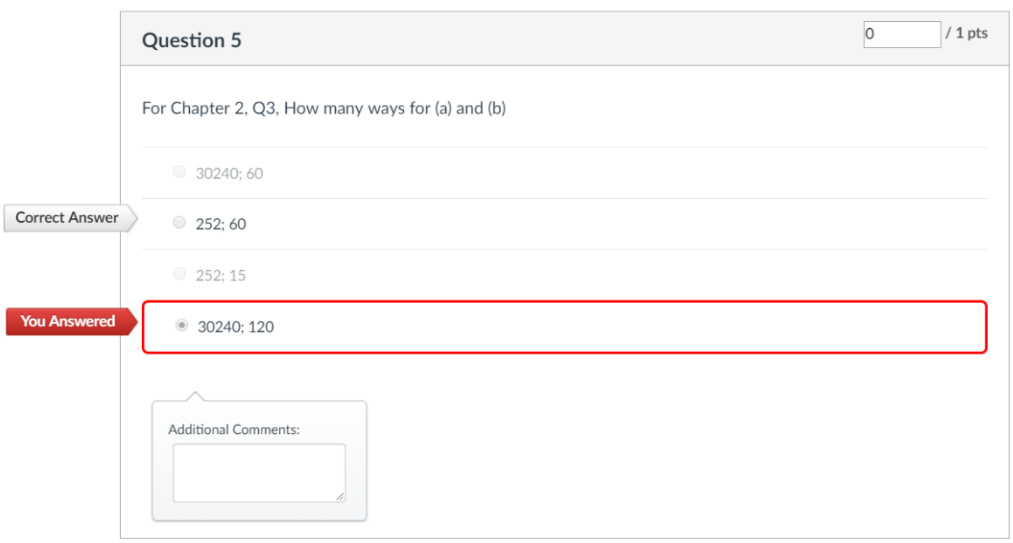


Figure 3: An incorrected answered question by a student.

The summary of each question is also provided, which helps the instructor to monitor the progress of the lectures. Figure 4 gives an example on the summary of two selected questions of an online quiz. Most of the students get right answer for the second question, however, half of the students answered incorrectly on the first question, and the wrong answer is mainly N(2, 3/5). This question is about sampling distribution of the sample average. The population variance is 9, and standard deviation 3. Many of the students mistakenly consider 3 as the variance of the population distribution. This reveals that sampling distribution should be re-visited and more practice and examples for this section should be given in the class.

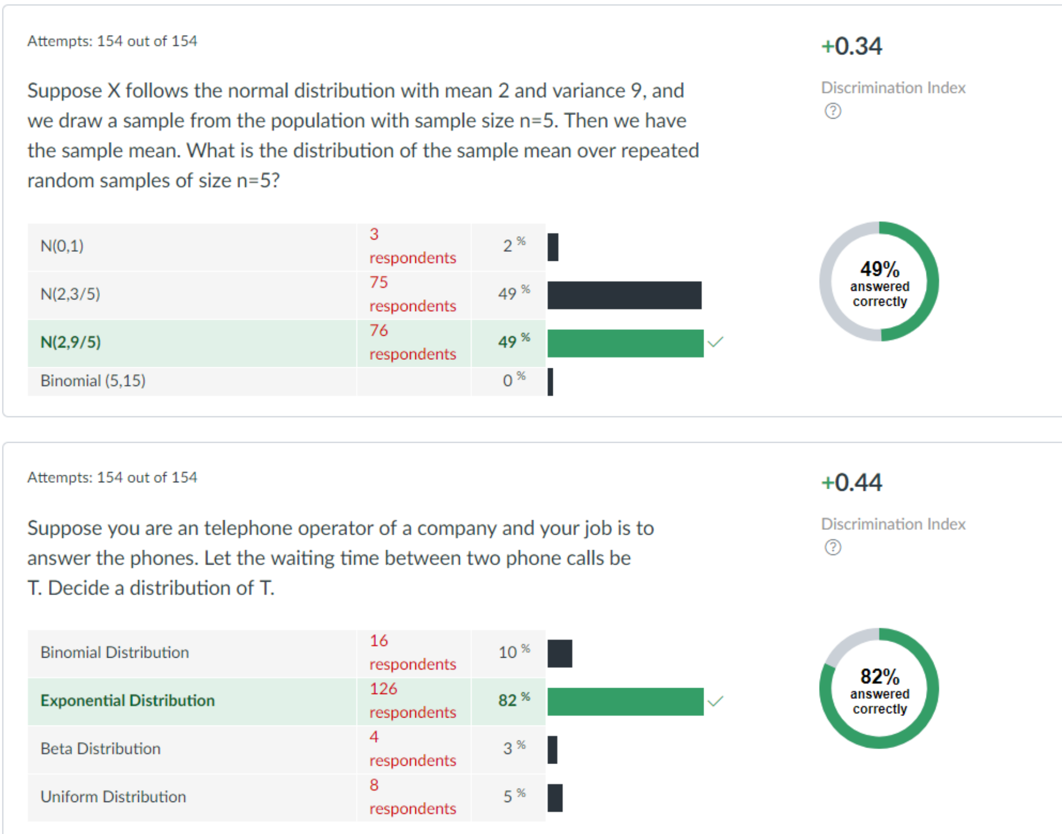


Figure 4: The summary of the answers for two selected questions in an assignment.

Analysis of grades and grade trends

Figure 5 provides the histograms of the scores of two midterm exams. As we see, the average score of the first exam was much higher than the second one. More than half of the students got a score higher than 80, and about one third of the students scored higher than 90. The first exam was on some basic probability theories, which was not hard to most of the students. However, the average of the second exam was much lower, around 70. Few students scored higher than 90 in the second exam. This exam was about probability distributions and sampling distributions, which was much more difficult. The students should not feel dis-encouraged by a lower score.

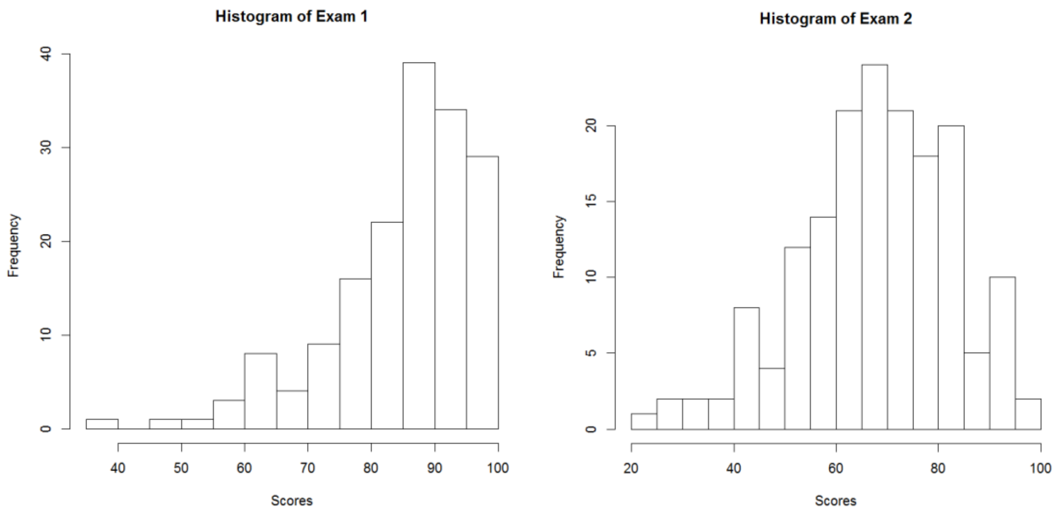


Figure 5: Histograms of the scores of the two midterm exams.

Scatter plot of the two exams is given in the following figure. As we see, for almost all the students, their exam 2 scores were lower than their exam 1. The red line is the 45 degree line. Several of the students scored near this line. Those students actually performed quite well since Exam 2 was much harder. There are two students marked in the red circle, who did significantly better in Exam 2. Their exam 1 were around 65, but exam 2 were over 80. Given the difficulty of Exam 2, those two students achieved even higher improvement than the 15 points reflected through the scores of the two exams. There are also several students in the lower right corner of Figure 6, marked in blue. Those students did much worse in Exam 2 compared to their Exam 1. Especially, one student scored nearly 100 in Exam 1, but below 30 in Exam 2. Special attention like meeting in personal should be paid on those students.

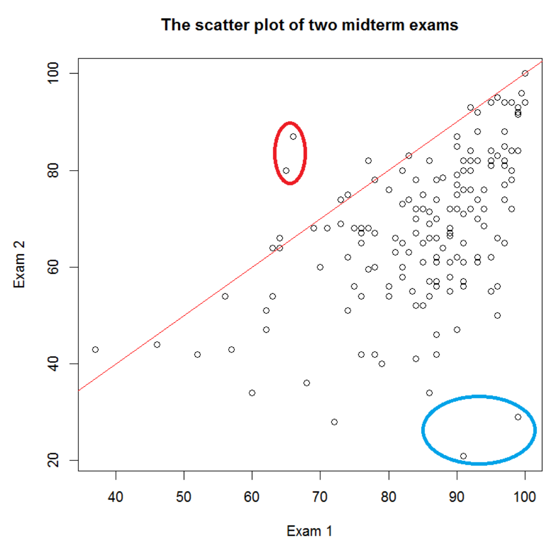


Figure 6: Scatter plot of the scores of the two midterm exams.

Planned changes

The challenging of teaching large session is keeping students involved during the lecture and obtaining real time feedback from the students. Although in-class activities and online quizzes can help student learn this course, it is hard for the instructor to know how well the students understand in each class. We plan to use Tophat in the future semesters for this course. Tophat is a teaching platform, by which the students read the lecture slides, and use their laptops or smart phones to provide answers to the in-class practice problems. Similar as what Canvas does for the online homework, it automatically summarizes the answers for each question, and provides real time feedbacks from the students. The instructor can teach the course to fit the needs of students based on those feedbacks.

Compared to Clickers which are widely used as an interactive tool between instructors and students, Tophat is linked to the grade roaster of students. The in-class practice can be graded by accuracy or served as attendance and participation scores. The students will take those practices more seriously once they realized that they are related to their final grade. I am eager to see how students feel about Tophat in the next semester.

Summary

This portfolio introduces teaching and assessment methods for Stat 380, with an especial target for the large session class. In-class activities, online quizzes and analysis for student performance and assignments are discussed. Through preparing the teaching portfolio, the author feels the effectiveness of backward design, from objectives to course activities, assessment and finally enhancing student learning. The future changes of this course will follow this process of design.

Statistics and Applications

Instructor: Yumou Qiu and (Merlin Mpoudeu)
Office: 343D Hardin Hall, East Campus
Email: yumouqiu@unl.edu and merlin.mpoudeu@huskers.unl.edu (Merlin)
Phone: 402-472-6454
Classroom: 03:30pm - 04:45pm TR at HENZ-124 City
Office Hours: By appointment

TAs: Vamsi Manthena, Yinglun Zhan, Surani Wijenayaka
Email: vamsi.manthena@gmail.com (Vamsi), zhanylun@gmail.com (Yinglun) and suraniwijenayake@gmail.com (Surani)
Office Hours: Wed. 1:00 - 3:00pm or by appointment
Location: Avery 250, City Campus

Text Book: *PROBABILITY AND STATISTICS FOR ENGINEERS AND SCIENTISTS*, 9th edition; Walpole, Myers, Myers and Ye.

Prerequisites: MATH 107 (MATH 107H): Analytic Geometry and Calculus II

Course Description: The STAT/MATH 380: Introduction to Probability and Statistics Concepts is the second-level course in "undergraduate statistics and mathematics" sequence. The course provides an introduction to probability concepts, i.e., random variables, probability distributions, expectation, variance, covariance, correlation; and statistical concepts, i.e., fundamental sampling distributions and data descriptions, one- and two-sample estimation and testing problems, and simple linear regression. This course lays the foundation for many 400-level courses in Engineering, Mathematics and Statistics, and is also offered for Honors credit.

ACE Outcome 3: The STAT/MATH 380 course is accredited as Achievement Centered Education (ACE) course and satisfies ACE outcome 3: to use mathematical, computational, statistical, or formal reasoning (including reasoning based on principles of logic) to solve problems, draw inferences, and determine reasonableness. Therefore, the reinforced skill for STAT/MATH 380 is Critical Thinking. STAT 380 will provide the student opportunities to achieve this learning outcome through readings, quizzes, homeworks and in-class activities. These assignments (together with exams) will be used by the instructor to assess achievement of the outcome. The final exam will be used by the Department of Statistics to carry out an overall assessment of STAT 380, and the course's effectiveness at assisting students in achieving Outcome 3.

Course Goals: This course will help you learn to think and reason statistically, and to construct arguments based on numerical evidence. My role as the instructor is to facilitate this type of learning by providing you with a variety of meaningful activities and opportunities to learn, as well as creating an environment conducive to learning. This will manifest in a variety of ways: group work, direct instruction, individual practice, exploration and discovery activities, writing, discussions and/or student-led instruction. Ultimately, **you are responsible for your own learning**, so please put into the class what you hope to get out of it.

- Students will understand essential theoretical concepts (and logic behind them) including common discrete and continuous probability distributions, sampling distributions and data description techniques, hypothesis testing problems, and modeling approaches.culations, and comment on the result.
- Students will be able to state a hypothesis testing (or estimating) problem (out of the ones considered in the course) for a given research question, perform testing (construct the corresponding confidence interval) and draw statistical inferences.
- For a given data set, students will be able to fit a simple linear regression model (obtain regression coefficients), provide an equation of a fitted model, interpret the observed relationship in words, and use the model to make predictions.

Course Content:

- Introduction to Statistics and Data Analysis (sample, population, observational and experimental studies, measures of location and variability, basics of data cleaning, graphical methods and data description)
- Probability
- Random Variables, Probability Distributions and Expectation
- Discrete and Continuous Probability Distributions
- Fundamental Sampling Distributions and Data Descriptions (random sampling, key statistics, sampling distributions of a sample mean and sample variance)
- One- and Two-Sample Estimation Problems (statistical inference, confidence intervals for the mean, difference between two means, proportion and the difference between two proportions: their derivations and applications)
- One- and Two-Sample Testing Problems (statistical hypotheses and their testing, tests for a single mean and the difference between two means: cases with known and unknown variances; large sample tests for a single proportion and the difference between two proportions, applications)
- Simple Linear Regression

Course Expectations: In this course, you are expected to have professional behavior. You are expected to attend all class meetings, be curious, ask questions, seek opportunities to learn, and be open and responsive to constructive feedback. You are also expected to exhibit a professional demeanor (language, attitude) toward others. Disagreement during discussions is welcome and often productive in developing a deeper understanding of the concepts being discussed. However, disagreement does not warrant yelling or disrespectful language or behavior. Unprofessional behavior will not be tolerated, and appropriate actions will be taken to prevent future occurrences.

Lectures: Lecture slides are used in all lectures. The slides are posted on Canvas. Students are expected to make notes to supplement the slides. If a student plans to miss a lecture then he/she should notify the instructor in advance.

Grading:

Homework	15%	Exam I	20%
Quizzes	15%	Exam II	20%
Participation	10%	Exam III	20%

A final average of 90% will guarantee an A-, 80% a B-, 70% a C- and 60% a D-.

Homework: Each student should print out and complete each homework assignment, and also take an on-line homework test on Canvas. Answer keys are posted on Canvas after the due date. Students are expected to compare their solutions with the answer keys and discuss their questions and concerns with the instructor. Homework tests are equally weighted. All assignments should be submitted before the deadline. There is no makeup possible for any assignments unless prior arramgements with the instructor have been made.

Participation: Most of class time will be devoted to group and individual activities, rather than lecture. Coming to class prepared to participate in these activities is vital for your success in the course. These activities may (but will not necessarily) be graded by accuracy or on a participation basis (turn in in-class group work with members' names or the instructor will call group number for answering questions). They may also require additional work outside of class. Don't write down the name for members missing the class. One ponit will be taken off if missing one group work or one in-class question call.

Quizzes: All quizzes will be in-class (unless otherwise is stated by the instructor). The instructor does reserve the right to give pop quizzes at any time. Any missed quiz is given 0 points. One lowest quiz grade is dropped. All quizzes are equally weighted.

Exams: Exams are closed book and notes. Students are allowed to use a formula sheet (in addition to distribution tables and a calculator) but no lecture notes, slides or texts are allowed. The formula sheets are to be prepared individually by each student and to be used when sample exams and fun reviews are taken. You are expected to take exams at the scheduled times. If this is impossible due to extreme circumstances (illness, death in the family, previously scheduled activities vital to academic program), please notify me. No make-up exams will be given if I am not notified prior to the examination. You will be required to obtain a note from your physician or adviser explaining the nature of the conflict.

Note: All work must be legible, and when appropriate all work must be shown to receive credit. No late work will be accepted unless other arrangements have been made before class. If you miss class, it is your responsibility to obtain the information missed. All exams and quizzes must be taken at the scheduled time. A missed exam or quiz will result in a zero unless you contact me ahead of time with adequate documentation (instructor, physician, or organization note).

Calculator: You will need a calculator for in-class activities, quizzes, and exams. You may use any calculator that contains basic statistical functions; an advanced calculator is not needed to be successful in this course. However, cell phones and computers may not be used to perform calculations on quizzes and exams.

Department Grade Appeal Policy: Students who believe their academic evaluation has been prejudiced or is capricious have recourse for appeals to, in order: their instructor; the Chair of the Statistics Department; the undergraduate academic grading appeals committee; and lastly, the college grading appeals committee.

Disability: Disabilities are encouraged to contact the instructor for a confidential discussion of their individual needs for academic accommodation. It is the policy of the University of Nebraska-Lincoln to provide flexible and individualized accommodation to students with documented disabilities that may affect their ability to fully participate in course activities or to meet course requirements. To receive accommodation services, students must be registered with the Services for Students with Disabilities (SSD) office, 132 Canfield Administration, 472-3787 voice or TTY.

Academic Integrity: You are encouraged to work together on problems and exercises, but the work you turn in must be your own (unless the assignment specifically states otherwise) Work on exams must be your own. University policy will be followed in cases of academic dishonesty:

In cases where an instructor finds that a student has committed any act of academic dishonesty, the instructor may in the exercise of his or her professional judgment impose an academic sanction as severe as giving the student a failing grade in the course. Before imposing an academic sanction the instructor shall first attempt to discuss the matter with the student. If deemed necessary by either the instructor or the student, the matter may be brought to the attention of the student's major adviser, the instructor's department chairperson or head, or the dean of the college in which the student is enrolled.

For additional details see http://stuafs.unl.edu/ja/code/three.shtml.

Disclaimer: Information contained in this syllabus was, to the best knowledge of the instructor, considered correct and complete when distributed at the beginning of the term. However, the instructor reserves the right, acting within the policies and procedures of UNL, to make changes in course content or instructional technique without notice or obligation. However, any changes will be explained to the class as a whole including reasons for the change.

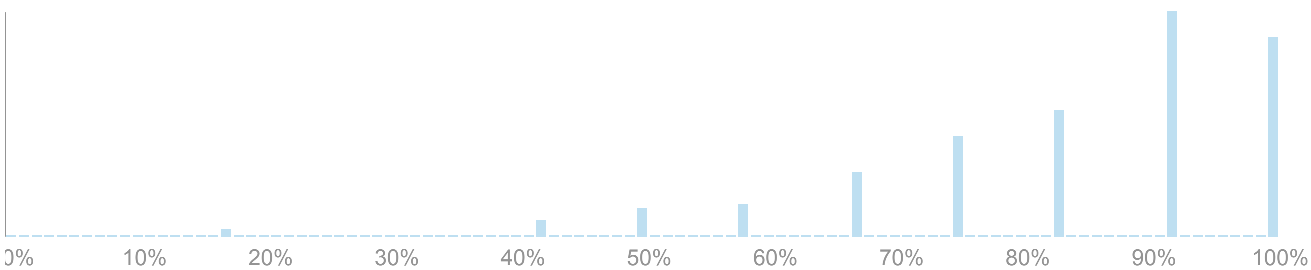
Table 1: TENTATIVE COURSE OUTLINE

Week	Day	Topic
1	Tue, Jan 9 Thur, Jan 11	Syllabus, Chapter 1 Chapter 1
2	Tue, Jan 16 Thur, Jan 18	Chapter 2 Chapter 2
3	Tue, Jan 23 Thur, Jan 25	Chapter 3 Chapter 3
4	Tue, Jan 30 Thur, Feb 1	Chapter 3, 4 Chapter 4
5	Tue, Feb 6 Thur, Feb 8	Chapter 4 Review (Ch 1-4)
6	Tue, Feb 13 Thur, Feb 15	Exam 1 Exam 1 Discussion, Chapter 5
7	Tue, Feb 20 Thur, Feb 22	Chapter 5 Chapter 6
8	Tue, Feb 27 Thur, Mar 1	Chapter 6 Chapter 6, 8
9	Tue, Mar 6 Thur, Mar 8	Chapter 8 Chapter 8
10	Tue, Mar 13 Thur, Mar 15	Review (Ch 5-6, 8) Exam 2
11	Tue, Mar 20 Thur, Mar 22	Spring Vacation
12	Tue, Mar 27 Thur, Mar 29	Exam 2 Discussion, Chapter 9 Chapter 9
13	Tue, April 3 Thur, April 5	Chapters 9, 10 Chapters 10
14	Tue, April 10 Thur, April 12	Chapter 10 Chapter 10
15	Tue, April 17 Thur, April 19	Chapter 11 Chapter 11
16	Tue, April 24 Thur, April 26	Review Review
17	Final Week	Final, 8:30 to 10:30 p.m. Monday, April 30

Quiz Summary

Section Filter ▾ | Student Analysis | Item Analysis

Average Score	High Score	Low Score	Standard Deviation	Average Time
84%	100%	17%	1.88	23:31



Question Breakdown

Attempts: 150 out of 150

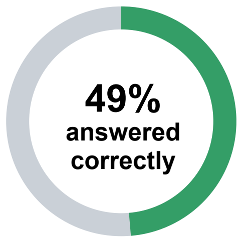
Suppose X follows the normal distribution with mean 2 and variance 9, and we draw a sample from the population with sample size n=5. Then we have the sample mean. What is the distribution of the sample mean over repeated random samples of size n=5?

+0.34

Discrimination Index

N(0,1) 3 respondents 2 %

N(2,3/5)	74 respondents	49 %	
N(2,9/5)	73 respondents	**49 %**	✓
Binomial (5,15)		0 %	



Attempts: 150 out of 150

A chimpanzee named Sarah was the subject in a study of whether chimpanzees can solve problems. Sarah was shown 30-second videos of a human actor struggling with one of several problems (for example, not able to reach bananas hanging from the ceiling). Then Sarah was shown two photographs, one that depicted a solution to the problem (like stepping onto a box) and one that did not match that scenario. Researchers watched Sarah select one of the photos, and they kept track of whether Sarah chose the correct photo depicting a solution to the problem. Sarah chose the correct photo in seven of eight scenarios that she was presented.
We want to run a test of significance to determine whether Sarah understands how to solve problems and will thus pick the photo of the correct solution more often than what would be done by random chance. Which of the following represents the null and alternative hypotheses?

a. $H_0 : \pi = 0.5$
$H_a : \pi > 0.5$

b. $H_0 : \pi = 0.5$
$H_a : \pi \neq 0.5$

c. $H_0 : \pi > 0.5$
$H_a : \pi < 0.5$

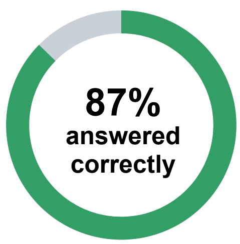
d. $H_0 : \pi = 0.5$

$H_a : \pi < 0.5$

+0.46

Discrimination Index ?

a	131 respondents	**87 %**	✓
b	5 respondents	3 %	
c	13 respondents	9 %	
d	1 respondents	1 %	



Attempts: 150 out of 150

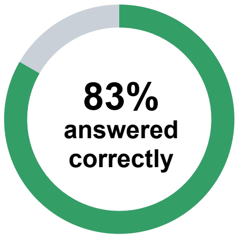
Suppose you are an telephone operator of a company and your job is to answer the phones. Let the waiting time between two phone calls be T. Decide a distribution of T.

+0.45

Discrimination Index ?

Binomial Distribution	15 respondents	10 %	
Exponential Distribution	125 respondents	**83 %**	✓
Beta Distribution	2 respondents	1 %	

Uniform Distribution	8 respondents	5 %



Attempts: 150 out of 150

A chimpanzee named Sarah was the subject in a study of whether chimpanzees can solve problems. Sarah was shown 30-second videos of a human actor struggling with one of several problems (for example, not able to reach bananas hanging from the ceiling). Then Sarah was shown two photographs, one that depicted a solution to the problem (like stepping onto a box) and one that did not match that scenario. Researchers watched Sarah select one of the photos, and they kept track of whether Sarah chose the correct photo depicting a solution to the problem. Sarah chose the correct photo in seven of eight scenarios that she was presented.

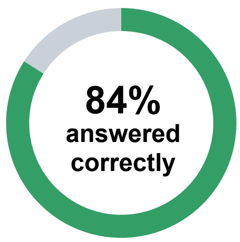
Which of the following best describes the meaning of the p-value in this situation?

+0.45

Discrimination Index ?

If Sarah doesn't understand how to solve problems and is just guessing at which picture to select, the probability she would get 7 or more correct out of 8 is 0.04.	126 respondents	**84 %** ✓
If Sarah understands how to solve problems and is not just guessing at which picture to select, the probability she would get 5 or more correct out of 8 is 0.04.	3 respondents	2 %

If Sarah understands how to solve problems and is not just guessing at which picture to select, the probability she would get 7 or more correct out of 8 is 0.04.	15 respondents	10 %
If Sarah doesn't understand how to solve problems and is randomly choosing which picture to select, the probability she would get less than 7 correct out of 8 is 0.04.	6 respondents	4 %



Attempts: 150 out of 150

A chimpanzee named Sarah was the subject in a study of whether chimpanzees can solve problems. Sarah was shown 30-second videos of a human actor struggling with one of several problems (for example, not able to reach bananas hanging from the ceiling). Then Sarah was shown two photographs, one that depicted a solution to the problem (like stepping onto a box) and one that did not match that scenario. Researchers watched Sarah select one of the photos, and they kept track of whether Sarah chose the correct photo depicting a solution to the problem. Sarah chose the correct photo in seven of eight scenarios that she was presented.
Based on your p-value, do you have strong evidence that Sarah understands how to solve problems similar to those she was presented?

+0.48

Discrimination Index ?

We have no evidence that Sarah is not simply guessing, since 7 out of 8 commonly occurs (if just guessing).	1 respondents	1 %

We have strong evidence that Sarah is not simply guessing, since 7 out of 8 rarely occurs by chance (if just guessing).	144 respondents	**96 %**
We have less evidence that Sarah is simply guessing, since 7 out of 8 always occurs by chance (if just guessing).	3 respondents	2 %
We have strong evidence that Sarah is simply guessing, since 7 out of 8 occurs every time (if just guessing).	2 respondents	1 %

Attempts: 149 out of 150

The heights of a random sample of 50 college students showed a sample mean of 174.5 centimeters and a sample standard deviation of 6.9 centimeters.

Construct a 98% confidence interval for the mean height of all college students.

+0.38

Discrimination Index ?

(172.15,176.85)	147 respondents	**98 %**
(172, 186)	2 respondents	1 %
(0,1)		0 %
(172.15, 300.65)		0 %
No Answer	1	1 %

respondents

Attempts: 149 out of 150

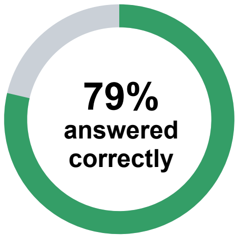
A study was conducted in which two types of engines, *A* and *B*, were compared. Gas mileage, in miles per gallon, was measured. 50 experiments were conducted using engine type *A* and 75 experiments were done with engine type *B*. The gasoline used and other conditions were held constant. The average gas mileage was 36 miles per gallon for engine *A* and 42 miles per gallon for engine *B*. Assume that the population standard deviations are 6 and 8 for engines *A* and *B*, respectively. We found a 95% confidence interval on the difference of population mean gas mileages for engines *A* and *B* is *(-8.46,-3.54).* How do you interpret this interval?

+0.61

Discrimination Index ?

We are 95% confident that the difference of population mean gas mileages for engines A and B is between -8.46 and -3.54.	118 respondents	**79 %**	✓
The probability that "the difference of population mean gas mileages for engines A and B is between -8.46 and -3.54" is 95%.	8 respondents	5 %	
We are 95% confident that the average value of population mean gas mileages for engines A and B is between -8.46 and -3.54.	7 respondents	5 %	
All the above are correct.	16 respondents	11 %	

No Answer	1 respondents	1 %



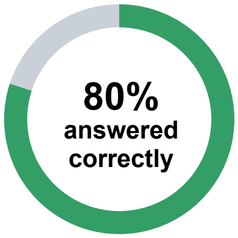
Attempts: 149 out of 150

A company with a large fleet of cars hopes to keep gasoline costs down and sets a goal of attaining a fleet average of at least 26 miles per gallon. To see if the goal is being met, they check the gasoline usage for 50 company trips chosen at random, finding a mean of 25.02 mpg and a standard deviation of 4.83 mpg. We construct a 90% confidence interval for **μ** is:(23.87, 26.15). Which of the following conclusions on the true average of mpg is correct?

+0.59

Discrimination Index ?

We are 90% confident that the true average mph is within (23.87,26.15).	120 respondents	**80 %**	✓
The probability that the true population mean is between 23.87 and 26.15 is 90%.	14 respondents	9 %	
We are 90% confident that the mean can be calculated.	5 respondents	3 %	
We are 90% confident that the sample average mph is within (23.87,26.15).	10 respondents	7 %	
No Answer	1 respondents	1 %	



Attempts: 149 out of 150

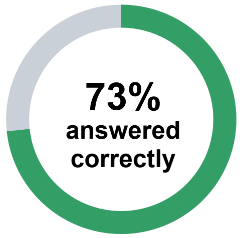
Consider the data from a population of N=10,000 people, indicating whether they do ("Yes") or don't ("No") have wireless internet service at home. The proportion of people who have wireless internet service is π, which is unknown. We call it the unknown population proportion. Now if we randomly choose n=100 people from the population, we can get the sample proportion , which is equal to (the number of people who have wireless internet in this sample)/(n=100). Suppose the observed sample proportion equals to 0.55, and we calculate the 95% confidence interval for π is: (0.45,0.65). Interpret this interval.

+0.58

Discrimination Index ?

The probability that the true population proportion is between 0.45 and 0.65 is 95%.	13 respondents	9 %
We are 95% confident that the true proportion of people having wireless internet service at home is between 45% and 65%.	110 respondents	**73 %** ✓
95% of the people who have wireless internet have been calculated by using this interval.	3 respondents	2 %
We are 95% confident that the sample proportion of people having wireless internet service at home is between 45% and 65%.	23 respondents	15 %
No Answer	1	1 %

respondents



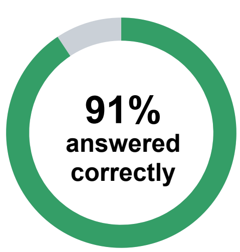
Attempts: 149 out of 150

A STAT 380 instructor claims " the average final exam score of students who passed STAT/MATH 380 is 75." To check validity of this claim you inspect a large SRS of final exam scores of students who passed the course and calculate the 90% confidence interval for the true mean score as (70,74). Thus, with 90% confidence ,do you conclude that the instructor's claim is reasonable? If yes, select True. If no, select False.

+0.56

Discrimination Index ?

True	13 respondents	9 %
False	136 respondents	**91 %** ✓
No Answer	1 respondents	1 %



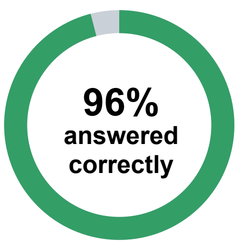
Attempts: 149 out of 150

The 95% two-sided confidence interval for the mean is (-0.1,0.2). One of your friends claims that the mean could be 0. Is your friend's claim reasonable? If yes, select True. If no, select False.

+0.37

Discrimination Index ⓘ

True	144 respondents	**96 %** ✓
False	5 respondents	3 %
No Answer	1 respondents	1 %



Attempts: 149 out of 150

An insurance company checks police records on 582 accidents selected at random and notes that 91 of them involved teenage drivers. A politician urging tighter restrictions on drivers' licenses issued to teens says, "In one of every five auto accidents, a teenager is involved." Is this statement reasonable?If yes, select True. If no, select False.

+0.24

Discrimination Index ⓘ

True	9	6 %

	respondents		
False	140 respondents	**93 %**	✓
No Answer	1 respondents	1 %	

Attempts: 1 out of 1

Suppose you are an telephone operator of a company and your job is to answer the phones. Let the waiting time between two phone calls be T. Decide a distribution of T.

+1.00

Discrimination Index ?

Binomial Distribution		0 %	
Exponential Distribution	1 respondents	**100 %**	✓
Beta Distribution		0 %	
Uniform Distribution		0 %	

Attempts: 1 out of 1