Title

An Analysis of the Discrepancy between the Students' Responses to Multiple-Choice and Open-Ended Questions

Author

Dabagh, Safaa M.

Publication Date

2016

University of California
Los Angeles

An Analysis of the Discrepancy between the Students Responses to Multiple-Choice and Open-Ended Questions on Confidence Interval and P-value: an Exploratory Study

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Statistics

by

Safaa M Dabagh

2016

Abstract of the Thesis

An Analysis of the Discrepancy between the Students Responses to Multiple-Choice and Open-Ended Questions on Confidence Interval and P-value: an Exploratory Study

by

Safaa M Dabagh

Master of Science in Statistics

University of California, Los Angeles, 2016

Professor Frederic Paik R. Schoenberg, Chair

Guidelines for Assessment and Instruction in Statistics Education (GAISE) provide a framework for revising introductory statistics courses. GAISE promotes statistical literacy and statistical thinking, where students not only understand concepts but are able to critically evaluate and make arguments based on quantitative information.

There are multiple definitions of statistical literacy and statistical thinking in the literature. In this thesis we are interested in the aspect of statistical literacy that aims for an educated consumer, one who can process everyday statistical information. Further, Statistical thinking is present when students can apply the statistical information learned to situations such as: 1) using statistics to solve real world problems, 2) critique and interpretation of statistical information reported in the mass media, and 3) interpretation and communication of statistical findings within context .

Despite this growing consensus that students must acquire higher-order think-

ing and performance skills in order to apply statistical reasoning and thinking to their research problems successfully, a gap still exists between course objectives and student outcomes. After completion of the coursework, students are more likely to forget the materials learned in course and continue to struggle with applying statistical reasoning and thinking. The type of assessment used in statistics courses provides one explanation for this gap. With the increase number of students enrolled in introductory courses, instructors rely heavily on multiple choice questions to evaluate students′undesrtanding of course materials and principles. Furthermore, in subjects such as mathematics, statistics, chemistry, biology, and physics, research has shown that about 70% of the questions are at the recall or comprehension level with very little attention paid to the questions that target application, analysis, synthesis, and evaluation. Therefore, and based on Bloom′s taxonomy, the student outcomes focus on lower level thinking skills (knowledge and comprehension), whereas the course objectives and expectations involve higher level of thinking skills (application, analysis, synthesis, and evaluation). In order to bridge the gap between objectives and outcomes, tests must provide students opportunities to employ higher order thinking.
The objectives of this thesis include:

1. Comparison of students′responses to open-ended questions on confidence interval and P-value.

2. Prediction of the students′ final scores from their scores on "upper level thinking", "application", "lower level computation and "upper level computation".

3. Pinpoint the students′misconceptions of the P-value by comparing the proportion of correct answers under two conditions including: a) deciding about the null hypothesis by comparison of the P-value Vs. level of significance or alpha, and b) deciding about the null hypothesis by examination of the confidence interval and the interpretation of the P-vlaue.

The thesis of Safaa M Dabagh is approved.

Hongquan Xu

Yingnian Wu

Maryam M Esfandiari

Frederic Paik R. Schoenberg, Committee Chair

University of California, Los Angeles

2016

To my mother; my siblings, Intissar, Abir, Mirna, Toufic, and Abdul Hay; and my wonderful children, Adam, Aya, and Abed Al Rahman...

who—among so many other things—

saw to it that I learned to be patient, strong, and persistant ...

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my sincere gratitude to my advisor Professor Mahtash Esfandiari for the continuous support of my M.S. study and research, for her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me in all the research and writing of this thesis.

I would like to further thank my committee members, Honquan Xu, Frederic Paik Schoenberg, and Yingnian Wu for their support and academic advice. Finally, My sincere gratitude to my children Adam, Aya, and Abed and my best friend Walid Chaarawi for supporting me throughout my life, for their patience and endless love.

CHAPTER 1

Introduction/Background of the Study

In the department of Statistics at UCLA, more than 3000 students are enrolled in "Introduction to Statistical Reasoning" per academic year. One of the major reasons for this high enrollment is the growing importance of statistics for different majors across campus. Another reason is that this course is a General Education (GE) requirement at UCLA, which means that all students need to complete it. Additionally, multiple sections of this course are offered through extension and taught by Ph.D. students who are trained in the teaching of Statistics.

One of the major goals of this course, which relates to this study, is training novice students in the field of statistics to engage in upper level thinking and to be able to communicate the results of statistical inference including confidence interval and P-value resulting from hypothesis testing in their own words. Upper level thinking means training the students to analyse and synthesis statistical results, and to evaluate the statistical methods used to solve real world problems.

In order to assess the students ability to engage in upper level thinking and express confidence interval and P-value in their own words, a combination of multiple-choice questions and short open-ended questions was used in Statistics 10, Introduction to Statistical Reasoning . Statistics 10 emphasizes statistical literacy and statistical thinking. Students attend lecture given by the professor three times a week for ten weeks. They meet in smaller groups of 45 students twice a week with a graduate-student Teaching Assistant (TA) once in a discussion session, and once in the computer lab. This study was conducted through analysing the students responses to questions on midterm and final examinations in two Statistics 10 courses in Fall 2012 and Spring 2013. These two courses were taught by the same instructor and teaching assistants. Additionally, homework assignments, labs, and questions asked on confidence interval and hypothesis testing were similar.

The following results were found:

1. With respect to confidence interval:

   1– In Fall 2012, The percentage of correct answers to multiple-choice questions and open-ended questions on confidence interval were 89 % and 56% respectively.

   2– In Spring 2013, the statistical software "Statcrunch" was introduced in lecture to demonstrate computation and interpretation of confidence interval within context of real data. Additionally, in weekly homeworks, students were required to solve similar problems using the software. In Spring 2013, The percentage of correct answers to multiple-choice questions and open-ended questions on confidence interval were 86% and 78% respectively.

2. With respect to the P-value,

   1– In Fall 2012, The percentage of correct answers to multiple-choice questions and open-ended questions on P-value were 81% and 23% respectively. In open-ended questions, students were asked to interpret the P-value within context. Only 23% did this correctly. 46% just compared the P-value with alpha and concluded they reject the null, and the other 31% had an inaccurate interpretation.

   2– In Spring 2013, there were no open-ended questions on the P-value. The proportion of correct answers to multiple-choice question on P-value was 81% on the midterm and 90% on the final. These percentages are similar to the Fall quarter.

3. With respect to prediction of the final score, the correlations between the overall performance on the test and the performance on the four clusters were: Upper level thinking (0.876), Application (0.729), Lower level computation (0.545), and Upper level computation (0.797). As predicted, questions from the upper-level thinking cluster is the best predictor of students′ final scores in the class.

4. With respect to students′ misconceptions on the P-value, when students were asked to decide about the null hypothesis by comparison of the P-value with the level of significance or alpha, 80% to 90% got the correct answer. The reason for this is that they memorize this fact or they have it on their cheat sheet; this is basically a recall of information or the lowest level of the cognitive domain. Consequently, getting the correct answer does not require any kind of upper level thinking. However, when they are asked to interpret the P-value within context, they will not be able to do this unless they understand the meaning of the actual risk they need to take to reject a true null and can express it within the context of the problem. In order to do this, they need to engage in analysis and comparison of type I error and the level of significance . Additionally they need to be able to write it within context. So, the results of the open-ended question on interpretation of the P-value highlight the students misconception of the interpretation of the P-value.

In chapters to follow I will : provide in chapter two an overview of the original Bloom′s taxonomy as well as the revised one; propose in chapter three the theoretical framework for the classification of the questions given on tests in Statistics 10 courses in Fall 2012 and Spring 2013; provide in chapter four the analysis of the results; and conclude in chapter five with recommendations.

CHAPTER 2

An Overview of Bloom′s Taxonomy

2.1 The Original Taxonomy

There is an extensive review of the literature connecting the two types of questions(open-ended and multiple choice questions) to the Bloom′s Taxonomy of the Cognitive Domain. In the 1950s, Benjamin Bloom and his colleagues developed a classification hierarchy for types of knowledge, cognitive processes, and skills. Bloom′s Taxonomy has had a profound effect on education and educators, as it provides a mental model for thinking about the relative difficulty of different objectives that students are expected to master and provides guidance for how teachers should approach and assess various objectives. The framework of Bloom′s Taxonomy of learning goals consisted of six major categories: Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation. Lower levels of learning focus on recall of information, middle levels require application of knowledge, and higher thinking levels involve the deconstruction and construction of concepts.

2.1.1 Knowledge

Knowledge represents the lowest level of learning outcomes in the cognitive domain. It involves recalling memorized information. The learning objectives at this level are knowing common terms, facts, methods and procedures, and basic concepts and principles.

2.1.2 Comprehension

Comprehension represents the lowest level of understanding. It involves the ability to grasp the meaning of materials, translating materials from one form to another, interpreting, and estimating future trends. The learning objectives at this level are understanding facts and principles, interpreting verbal material, charts and graphs, translating verbal material to mathematical formulae, estimating the future consequences implied in data, and finally justifying methods and procedures.

2.1.3 Application

Application refers to the ability to use learned materials in new and concrete situations. The learning outcomes in application require a higher level of understanding than those under comprehension. The learning objectives at this level are applying concepts and laws to new and practical situations, solve mathematical problems, construct graphs and charts, and demonstrate the correct usage of a method or procedure.

2.1.4 Analysis

Analysis represents the ability to break down material into its component parts. Identifying parts, analysis of relationships between parts, recognition of the organizational principles involved. Learning outcomes here represent a higher intellectual level than comprehension and application because they require an understanding of both the content and the structural form of the material. Learning objectives at this level: recognize unstated assumptions, recognizes logical fallacies in reasoning, distinguish between facts and inferences, evaluate the relevancy of data, analyze the organizational structure of a work (art, music, writing).

2.1.5 Synthesis

Synthesis represents the ability to put parts together to form a new whole. This may involve the production of a unique communication (theme or speech), a plan of operations (research proposal), or a set of abstract relations (scheme for classifying information). Learning outcomes in this area stress creative behaviors, with major emphasis on the formulation of new patterns or structure. Learning objectives at this level: write a well organized paper, give a well organized speech, write a creative short story (or poem or music), propose a plan for an experiment, integrate learning from different areas into a plan for solving a problem, formulate a new scheme for classifying objects (or events, or ideas).

2.1.6 Evaluation

Evaluation represents the ability to judge the value of material (statement, novel, poem, research report) for a given purpose. The judgments are to be based on definite criteria, which may be internal (organization) or external (relevance to the purpose). The student may determine the criteria or be given them. Learning outcomes in this area are highest in the cognitive hierarchy because they contain elements of all the other categories, plus conscious value judgments based on clearly defined criteria. Learning objectives at this level: judge the logical consistency of written material, judge the adequacy with which conclusions are supported by data, judge the value of a work (art, music, writing) by the use of internal criteria, judge the value of a work (art, music, writing) by use of external standards of excellence.

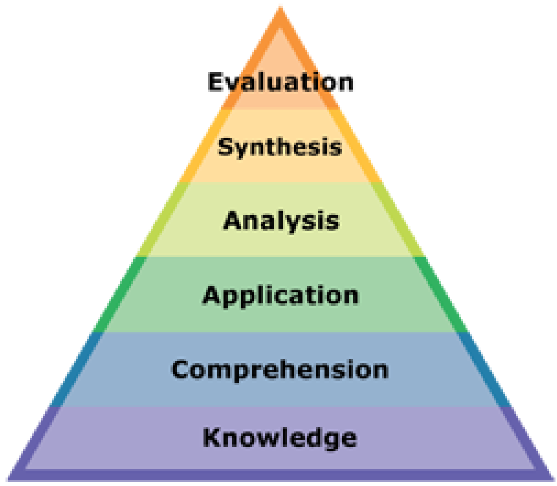


Figure 2.1: Original Bloom′s Taxonomy.

2.2 The Revised Bloom′s Taxonomy

In the 1990s, one of Bloom′s students, Lorin Anderson, revised the original taxonomy. In the amended version of Blooms Taxonomy, the names of the major cognitive process categories were changed to indicate action because thinking implies active engagements. Instead of listing knowledge as a part of the taxonomy, the category is divided into different types of knowledge: factual, conceptual, procedural, and metacognitive. This newer taxonomy also moves the evaluation stage down a level and the highest element becomes "creating."

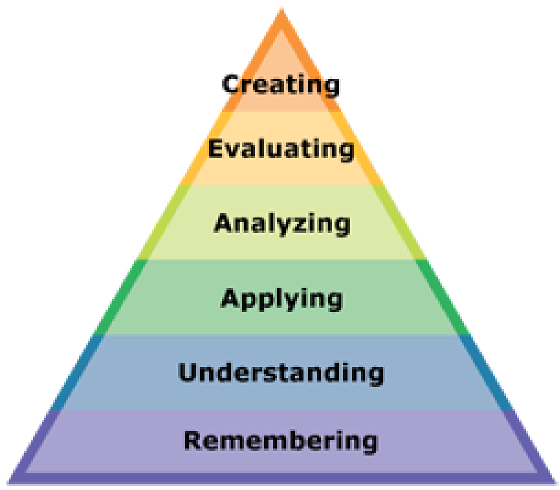


Figure 2.2: The Revised Taxonomy.

CHAPTER 3

The Theoretical Framework Proposed for the Classification of the Questions

The data were collected during the fall of 2012 from a final exam of statistics 10 class (N=176). There were fifteen questions on the exam. Those questions consisted of 10 multiple-choice questions (1 through 10) and five open-ended questions (11 through 15.) Further, and based on Bloom′s Taxonomy, we categorize the final exam questions into four categories:

1. Recall and lower level computation (recall, knowledge), Questions 2 and 6 (Appendix 6.1.2): Recall questions requiring students to remember definitions and to distinguish between them, for example, "the difference between a parameter or a statistic", "inferential versus descriptive statistics", "definition of bias", "requirements for the application of Central Limit Theorem", etc. On the other hand, lower level computation questions require students to make simple computations such as calculation of the mean, standard deviation Z score, etc. All they need to do is to copy the formula off of their cheat sheet, plug some numbers in, and make the necessary computations.

2. Upper level computation (explain results, comprehension), Questions 1, 5,13a, 15a, 15b (Appendix 6.1.2) :These types of questions require students to apply concepts and principles to new situations. In question 12 for example, students are given four scenarios and asked to determine the appropriate method to analyze the relationship between two variables.

3. Application, questions 10,11, and 12(Appendix 6.1.2): This category aligns with the application level defined by Bloom′s Taxonomy; it requires students to use learned material in new and concrete situations.

4. Upper level thinking (analysis, synthesis, evaluation), Questions 4, 7, 8, 9, 13b, 13c, 14a, 14b, and 15c (Appendix 6.1.2): This category combines the three levels: analysis, synthesis, and evaluation from Bloom′s Taxonomy. These types of questions require students to recognize unstated assumptions, recognize logical fallacies in reasoning, distinguish between facts and inferences, evaluate the relevancy of data, and analyze the organizational structure of a work. By synthesizing, students are required to put parts together to form a new whole. In addition, students are required to judge the value of material for a given purpose. In question fourteen, for example, students are required to interpret the p-value and the confidence interval within context.

In specific, questions 7 and 14b were designed to test students knowledge/ understanding of the concept of confidence interval, they fall into the fourth category: the upper level thinking category. Questions 9 and 14a were designed to test students′knowledge/ understanding of the concept of p-value.They aslo fall into the fourth category: the upper level thinking category.

CHAPTER 4

Analysis of Results

4.1 Analysis of students′responses to multiple-choice and open-ended questions on confidence interval and p-value.

4.1.1 Results from Fall 2012

Students were allowed to use 8.5 by 11 cheat-sheet on both sides. When Grading open-ended questions on confidence interval, we divided students′ answers into three categories:

1. Correct: An example of a correct answer:"We are 95%confident that the difference in proportions of democratic females and democratic males is between .13 to .27."

2. Partially correct: An example of a partially correct answer:"The proportion of democratic women is between 0.13 to 0.27 higher than the proportion of democratic men."

3. wrong: An example of a wrong answer:"Given the sample, we are 95% confident that the true p-value of the population falls between 0.13 and 0.27 , and if repeated, 95% of the time we would calculate this confidence interval given similar samples that are representative of the population."

When grading open-ended questions on P-value, we divided students′answers into four categories:

1. Correct answer (7 points): The student provides the correct interpretation of the P-value. An example of student′s answer: "Assuming H_0 is true, the $P-value$ is the probability that if the experiment were repeated, you would get a test statistic as extreme as, or more extreme than the one you actually got. Since p-value is smaller and close to 0, the outcome is surprising and therefore discredits the null hypothesis."

2. Partially correct (5 points): The student provides an answer that reflects her vague understanding of the P-value, and her explanation is incomplete. An example that falls into this category: "the p-value of 0.02 shows that the results of the sample are so extreme that they can't be by chance, because the p-value is smaller then he can consider it statistically significant. Reject H_0."

3. Wrong interpretation (2 points): The student rejects H_0 without providing any interpretation of the P-value. They conclude that since P-value is less than α, they reject H_0. They try to interpret the results using their recall knowledge and this is more of comprehension meaning, since they are using their memorized knowledge to conclude something about the null like telling there is a relationship between gender and political affiliation or concluding that the proportion of democrats are higher among women. Hence, they are somewhat operating between the first and the second level of Bloom′s Taxonomy; specially given that they have the guidelines for when they can reject or fail to reject the null as well as sample interpretations of when they should reject null on their cheat sheets. An example:"the p-value is less than 0.05 so the null is rejected. There is evidence for a statistical relationship between gender and political affiliation." We decided to include this category

because the percentage of students with this answer is high 46%.

4. Wrong answer (0 points): the student provides meaningless interpretation of the P-value; for example:" the p-value is the true probability value at the proportion of democratic women being equal to the proportion of democratic men."

The percentage of correct answers to the multiple-choice questions on confidence interval (89%) is 32 % higher than the percentage of correct answers to open-ended questions(56%).
The percentage of correct answers to the multiple-choice questions on P-value (81%) is 58 % higher than the percentage of correct answers to open-ended questions(23%).

Table 4.1: Preliminary Results on Confidence Interval for Fall 2012

	Multiple-Choice on Confidence interval	Open-Ended on Confidence interval		
	correct	Correct	Partially Correct	Wrong
Fall 2012	89%	56%	14%	30%

Table 4.2: Preliminary Results on P-Value for Fall 2012

	Multiple-Choice	Open-Ended			
	correct	Correct	Partially Correct	Wrong	Wrong Interpretation
Fall 2012	81%	23%	17%	14%	46%

It is interesting that this gap is a lot lower than the gap on P value (32% vs.58%). Of course, this is expected because the correct interpretation of P value requires a good understanding of type I error or nominal risk you are willing take to reject a true null vs. the actual risk of rejecting a true null. Thus, the understanding of the P value requires seeing the connection between a number of abstract concepts. Generally speaking, students in both lower division and upper division classes seem to have a harder time grasping the P value compared to confidence interval. Another reason could be that it is easier to interpret confidence interval by using a real context to which the students can relate. Maybe in introductory statistics classes we need to put more emphasis on the teaching and interpretation of confidence interval.

Another reason to take into consideration is that confidence interval generally provides the consumers of statistics with a lot more information than the P-value. That is why majority of journals in the applied fields such as medicine, sociology, and psychology report confidence intervals compared to P-value and they even address rejection of the null based on the confidence interval.

4.1.2 Results from Spring 2013

In Spring 2013, the statistical software Statcrunch was introduced in lecture to demonstrate computation and interpretation of confidence interval within context of real data, and students continued to use their own cheat sheet. Additionally, in weekly homeworks, students were required to solve similar problems using the software. The percentage of correct answers to multiple-choice questions and open-ended questions on confidence interval were 86% and 78% respectively.

4.1.2.1 Confidence Interval Results for Spring 2013

Table 4.3: Preliminary Results on Confidence Interval for Spring 2013

Spring 2013	Multiple-Choice	Open-Ended		
	%correct	%Correct	%Partially Correct	%Wrong
midterm	89%	52%	23%	25%
Final	86%	78%	15%	7%

4.1.2.2 P-Value Results from Spring 2013

In Spring 2013, students were not given open-ended questions on the topic of P-value.

Table 4.4: Preliminary Results on P-value for Spring 2013

Spring 2013	Multiple-Choice	Open-Ended		
	%correct	%Correct	%Partially Correct	%Wrong
midterm	81%	NA	NA	NA
Final	90%	NA	NA	NA

4.1.3 Results for all Terms

4.1.3.1 Results on Confidence Interval for all Terms

In this section, we compare the results on confidence interval for all terms: Fall 2012, Spring 2013, and Summer 2013. In Fall 2012, students were allowed to use their own cheat sheet, and Statcrunch was not used. In Spring 2013, students were introduced to StatsCrunch but still used their own cheat sheet. In Summer 2013, which is a 6 weeks instruction, students used Statscrunch, and a formula sheet was given to them during exams. However, during the summer session, no open-ended was given.

NA means no open-ended question was given on the Confidence Interval.

The following observations were made:

- As predicted, When Statcrunch was used as an integral part of the lecture and homework, the gap between the correct answer to multiple-choice and open-ended questions on confidence interval decreased from 8% to 33%. Without Statcrunch the percentage of correct answers to multiple choice

Table 4.5: Preliminary Results on Confidence Interval for all Terms

	Multiple-Choice	Open-Ended		
	%correct	%Correct	%Partially Correct	%Wrong
Fall 2012 Cheat sheet was allowed Statcrunch was not used	89%	56%	14%	30%
Spring 2013 Cheat sheet was allowed Statcrunch was used	86%	78%	15%	7%
Summer 2013 Cheat sheet was not allowed Statcrunch was used	75%	NA	NA	NA [1]

and open-ended questions on confidence interval were 89% and 56% respectively. After using statcruch these percentages changed to 86% and 78% respectively.

- When Statcrunch was used the percentage of the correct answers on open-ended questions increased almost 20%.
- When the cheat sheet was taken away the percentage of correct answers to multiple-choice questions decreases by about 11% to 15%.

4.1.3.2 Results on P-Value for all Terms

NA means no open-ended question was given on the P-value.
By no interpretation, it is implied that the instead of interpreting the meaning of the P-value within context, the students used the information on the cheat sheet or their memorized information and stated we reject the null since P is less than 0.05.

The followings observations were made:

- When the cheat sheet was taken away the percentage of correct answers to

Table 4.6: Preliminary Results on P-Value for all Terms

	Multiple-Choice	Open-Ended			
	correct	Correct	Partially Correct	Wrong	Wrong Interpretation
Fall 2012 Cheat sheet was allowed Statcrunch was not used	81%	23%	17%	14%	46%
Spring 2013 Cheat sheet was allowed Statcrunch was used	81%	NA	NA	NA	NA
Summer 2013 Cheat sheet was not allowed Statcrunch was used	59%	NA	NA	NA	NA [2]

the interpretation of the P-value on the multiple-choice question decreased between 20% to 24%. The percentage of correct answers to multiple-choice questions on confidence interval and P value in the fall and spring quarters are almost similar and range from 81 % to 89%.The percentage of correct answers to open-ended questions on P value is a lot lower than the confidence interval (56% vs.23%).

- When we compare the two tables, it is clear that the students find interpretation of the P-value more difficult than the confidence interval. That may be the reason that most applied journals in medicine and social science report their results in terms of the confidence interval because it is easier to understand and gives the reader more information. This could imply that we should be placing more emphasis on the teaching of confidence interval than the P-value in introductory statistics classes.

4.2 Fall 2012 Analysis of Relationship between Score on Final Exam and Questions on Upper Level Thinking, Recall, Lower Level and Upper Level Computation

4.2.1 Correlation between the Final Score and the Four Suggested Cores

Table 4.7: Correlation between the Final Score and the Four Suggested Cores

	finalscore	upperlevel	application	computelower	computeupper
finalscore	1	.876**	.729**	.545**	.797**
upperlevel	.876**	1	.550**	.339**	.506**
application	.729**	.550**	1	.313**	.495**
computelower	.545**	.339**	.313**	1	.309**
computeupper	.797**	.506**	.495**	.309**	1

The correlation between the overall performance on the test and the performance on the four clusters: Upper level thinking (0.876), Application (0.729), Lower level computation (0.545), and Upper level computation (0.797). As predicted, questions from the upper-level thinking cluster is the best predictor of students final scores in the class.

4.2.2 Correlation between Questions Seven, Nine, and Fourteen from Fall 2012 Final Exam

The following questions were given to students in Fall 2012 final exam (Figure 4.1,4.2,4.3). Question seven and nine are part of multiple choice questions; question 14 is an open ended one:

1. the correlation between questions 7 and 14b is 0.386(which is low) [both questions test the students understanding of confidence interval]. 2. the correlation between questions 9 and 14a is 0.076 (which is very low) [both questions test

Table 4.8: Correlation between Questions 7,9, and 14

		Q14*-a	Q14*-b	Q7*	Q9*
Q14*a	Pearson Correlation	1	.292**	.251**	0.076
	Sig. (2-tailed)		0	0.001	0.336
	N	164	164	164	164
Q14*-b	Pearson Correlation	.292**	1	.386**	.264**
	Sig. (2-tailed)	0		0	0.001
	N	164	164	164	164
Q7*	Pearson Correlation	.251**	.386**	1	0.129
	Sig. (2-tailed)	0.001	0		0.099
	N	164	164	164	164
Q9*	Pearson Correlation	0.076	.264**	0.129	1
	Sig. (2-tailed)	0.336	0.001	0.099	
	N	164	164	164	164

Question Seven.

A researcher wants to find out whether, in rural areas, there is any relationship between gender and attitude toward pursuing a science-related career. For two random samples of 200 female and 200 male high-school seniors in a rural area, 70 of the males and 30 of the females indicated that they aspired to pursue a science-related career.

Given that the 95% confidence interval for the difference in proportions of males and females who aspire to a science-related career is from 0.315 to 0.485, what is your conclusion regarding the hypothesis test posed by this researcher?

(a) The p-value is less than 0.05. We reject the null, and conclude that in rural areas, the proportion of male seniors who aspire to pursue a science-related career is between 0.315 to 0.485 higher than female seniors.

(b) The p-value is less than 0.05. We reject the null, and conclude that in rural areas, the proportion of female seniors who aspire to pursue a science-related career is between 0.31.5 to 0.485 higher than male seniors.

(c) The p-value is greater than 0.05. We fail to reject the null, and conclude that in rural areas, the proportion of male and female seniors who aspire to pursue a science-related career is not significantly different.

(d) The p-value is less than 0.05. We reject the null, and conclude that, in rural areas, the proportion of male seniors who aspire to pursue a science-related career is different than the percentage of female seniors.

Figure 4.1: 2012 Final Exam Question 7

Question Nine.

A researcher wants to find out if the mean birth weight is different for working moms than for stay-at-home moms.

Descriptives	Working Moms	Stay-at-home Moms
Mean birth weight	7.15	7.42
SD	1.23	1.29

Given the above data for a random sample of 100 working moms and 100 stay-at-home moms, what is best the answer to the researchers question? Let the significance level be $\alpha = 0.05$.

(a) The p-value is greater than 0.05. We fail to reject the null, and conclude that there is no significant difference between the mean birth weight for working moms and stay-at-home moms.

(b) The p-value is greater than 0.05. We reject the null, and conclude that there is no significant difference between the mean birth weight for working moms and stay-at-home moms.

(c) The p-value is less than 0.05. We reject the null, and conclude that the mean birth weight is higher for stay-at-home moms than working moms.

(d) The p-value is less than 0.05. We reject the null, and conclude that the mean birth weight is different for working moms and stay-at-home moms.

Figure 4.2: 2012 Final Exam Question 9

Question Fourteen.

The following question was given on a Stats 10 exam:

A researcher wants to find out if there is a relationship between gender and political affiliation (Democrat vs. Republican). He tests the null hypothesis that the proportion of Democratic women is equal to the proportion of Democratic men ($P_1 - P_2 = 0$, where P_1 represents the proportion of Democratic women). For his sample, he finds the proportion of Democratic females and males to be 0.70 and 0.50, respectively. When he tests the null, he finds the p-value to be 0.02 and the 95% confidence interval for $P_1 - P_2$ to be 0.13 to 0.27. Assume all the necessary conditions are met.

(a) Interpret the p-value. NOT MORE THAN THREE LINES.

(b) Interpret the 95% confidence interval. NOT MORE THAN THREE LINES.

Figure 4.3: 2012 Final Exam Question 14

the students understanding of the p-value] 3. Studying the correlation between questions 7, 9, 14a, and 14b and the final score, we found that, from lowest to highest, Question 9 (MC) has the lowest correlation, question 14a (OE) comes next- observe that both questions test the students understanding of the p-value. Question 7 (MC) comes next, and finally, question 14b has the highest correlation with final score.

4.2.3 Correlation between Questions Seven, Nine, fourteen, and the Final Score

Table 4.9: Correlation between Questions 7,9, and 14 and final score

		Q14*-a	Q14*-b	Q7*	Q9*	finalscore
Q14*-a	Pearson Correlation	1	.292**	.251**	0.076	.457**
	Sig. (2-tailed)		0	0.001	0.336	0
	N	164	164	164	164	162
Q14*-b	Pearson Correlation	.292**	1	.386**	.264**	.562**
	Sig. (2-tailed)	0		0	0.001	0
	N	164	164	164	164	162
Q7*	Pearson Correlation	.251**	.386**	1	0.129	.526**
	Sig. (2-tailed)	0.001	0		0.099	0
	N	164	164	164	164	162
Q9*	Pearson Correlation	0.076	.264**	0.129	1	.421**
	Sig. (2-tailed)	0.336	0.001	0.099		0
	N	164	164	164	164	162
finalscore	Pearson Correlation	.457**	.562**	.526**	.421**	1
	Sig. (2-tailed)	0	0	0	0	
	N	162	162	162	162	162

1- From lowest to highest, Question nine (MC) has the lowest correlation(r=0.421), question fourteen a (OE) comes next (r=0.457)- observe that both questions test the students understanding of the p-value. Question seven (MC) comes next (r=0.526), and finally, question fourteen b has the highest correlation with final score (r=0.562).

4.2.4 Correlation between Multiple Choice, Open-ended Questions, and the Final Score

Studying the relationship between total score on OE, on MC, and the final score (after rescaling); we find that the total on open ended questions best predicts the total score (r=0.936). MC questions are also good predictors of final score.

Table 4.10: Correlations bteween Open Ended, Multiple Choice Questions, and Final Score

		openendedp	multipchoicep	finalscore
openendedp	Pearson Correlation	1	.649**	.936**
	Sig. (2-tailed)		.000	.000
	N	162	162	162
multipchoicep	Pearson Correlation	.649**	1	.875**
	Sig. (2-tailed)	.000		.000
	N	162	164	162
finalscore	Pearson Correlation	.936**	.875**	1
	Sig. (2-tailed)	.000	.000	
	N	162	162	162

4.2.5 Correlation between Multiple Choice (MC) Questions from Fall 2012 Final Exam

Next, we consider the correlation between the final score and the MC questions on P-value and Confidence intervals for Fall 2012 Final exam. Below are questions four and ten:

Question Four.

For a population of 1000 students in introductory statistics,

200 read the book
200 read the lecture notes
200 read their own notes
50 read the book and lecture notes
50 read the book and their own notes
50 read the lecture notes and their own notes
20 read all three

If you pick a student at random, what is the probability that she reads at least two of the above?

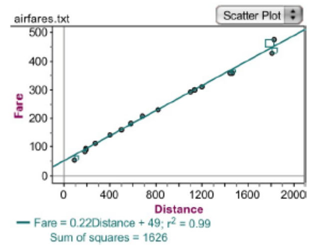
(a) 0.11
(b) 0.15
(c) 0.02
(d) 0.17

Figure 4.4: 2012 Final Exam Question 4

Question seven on confidence interval (r=0.526), question four on concept of probability (r=0.473), question ten on regression (0.453), and question nine on p-value (r=0.421) have the highest correlation with the final score. Those questions fall in the upper level thinking category: analysis, evaluation, and synthesis based on Bloom′s Taxonomy.

Question Ten.

Given the following plot, what is the best interpretation of the given information?



(a) We are predicting airfare from distance. If we increase distance by one mile, on the average, airfare increases by 49 cents. 99% of the variance in distance is explained by airfare, and 1626 is $\sum(x-\bar{x})^2$.

(b) We are predicting distance from airfare. If we increase airfare by one dollar, on the average, distance will increase 0.22 miles. 99% of the variance in distance is explained by airfare, and 1626 is $\sum e^2$.

(c) We are predicting distance from airfare. If we increase airfare by one dollar, on the average, distance will increase 49 miles. 99% of the variance in distance is explained by airfare, and1626 is $\sum(y-\bar{y})^2$.

(d) We are predicting airfare from distance. If we increase distance by one mile, on the average, airfare increases by 22 cents. 99% of the variance in airfare is explained by distance, and 1626 is $\sum e^2$.

Figure 4.5: 2012 Final Exam Question 10

Table 4.11: Correlation between MC Questions and Final Score

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9*	Q10	finalsco
Q1	1	0.017	.168*	.183*	-0.035	-0.038	.221**	0.046	0.046	0.345	.381*
Q2	0.017	1	0.057	0.064	-0.02	0.044	0.057	0.008	0.086	0.04	0.355
Q3	.168*	0.057	1	.224**	0.104	0.02	0.064	.178*	0.129	0.172	0.36
Q4	.183*	0.064	.224**	1	0.098	0.077	0.145	0.139	0.201	0.146	.473**
Q5	-0.035	-0.02	0.104	0.098	1	0.065	0.026	0.031	0.051	0.154	0.286
Q6	-0.038	0.044	0.02	0.077	0.065	1	0.075	0.151	0.064	0.12	0.275
Q7*	.221**	0.057	0.064	0.145	0.026	0.075	1	.263**	0.129	0.312	0.526
Q8	0.046	0.008	.178*	0.139	0.031	0.151	.263**	1	0.095	-0.019	.321*
Q9*	0.046	0.086	0.129	.201**	0.051	0.064	0.129	0.095	1	0.131	0.421
Q10	.345**	0.04	.172*	0.146	.154*	0.12	.312**	-0.019	0.131	1	.453*
finalscor	.381**	.355**	.360**	.473**	.286**	.275*	.526**	.321**	0.421	0.453	1**

CHAPTER 5

Recommendations

Since the number of students who want to enroll in Statistics 10, "Introduction to Statistical Reasoning,"is increasing exponentially, we have to:

1. Devise teaching methods that allow the students to develop a deep understanding of abstract concepts like the P-value and confidence interval and be able to communicate the findings of statistical inference verbally and orally.

2. Devise assessment strategies that test students′true understanding of these concepts.

Based on the finding of this study, it is recommended to:

1. Use statistical softwares such as Statcrunch, SPSS, etc... to demonstrate calculation and interpretation of P-value and confidence interval during lectures.

2. Have students use software to solve and interpret problems on P-value and confidence interval as part of their homeworks.

3. Provide students with outputs resulting from statistical software in the exams and have them interpret it.

4. Include short open-ended questions on midterms and finals, which test upper level thinking and are easy to grade.

   The following question was given on a statistics exam.

At a university with an enrollment of 30,000 students, we want to estimate the proportion of students who work part-time. In a random sample of 225 students, 150 students work part-time. We find the 95% confidence interval to be from 61% to 73%. Interpret the confidence interval.

John commented on those results as follows:

We are 95% confident that the percentage of students who work part-time in this sample is between 61% to 73%. If we select one hundred random samples, each of size 225, then about 95% of the corresponding confidence intervals would contain the true proportion of part-time workers in the population (represented by the letter P). Would you give John hire John?

5. If Clicker is used in lecture,

   (a) avoid writing clicker questions that only require students to recall information.

   (b) Write clicker questions at the upper-level thinking and have students discuss them as pairs; then have them solve the problem, write the results, and help them choose the correct response.

   (c) Based on the feedback you get from the clicker, you can diagnose, students′ misconception. Go through every wrong answer and have students tell you why that answer is wrong.

In phase II of the study, similar data was collected on students who were enrolled in a six-week course in Statistics 10 during summer 2013. The book, instructor, TAs, assignments, and labs were the same as Spring 2013. Statcrunch was used in lecture, homework, and final. The main difference was taking away the cheat-sheet, in order to see whether the percentage of correct answers to abstract concepts would change. No open-ended questions were given during summer. Results indicated that the proportion of correct answers to confidence interval and P-value were 75% and 59% respectively. It is interesting to notice that after the cheat sheet was taken away the percentage of correct answers dropped 11% to 15% on confidence interval and 20% to 24% on P-value. Hence, it′s recommended to provide students with a sheet that include all the necessary formulas during midterms and finals.

CHAPTER 6

Appendix

6.1 Fall 2012

6.1.1 Fall 2012 Midterm Two

6.1.2 Fall 2012 Final Exam

6.1.3 Spring 2013 Midterm Two Exam

6.1.4 Spring 2013 Final Exam

6.1.5 Summer 2013 Final Exam

Statistics 10 - Midterm 2
Professor Esfandiari
Fall 2012

Last Name: ____________________ **First Name:** ____________________

Student ID: ____________________

Name of TA: ____________________ **Section #:** ____________________

	Question #	Possible Points	Actual Points
SCORING	One to Six	36 points	________
	Seven	10 points	________
	Eight	10 points	________
	Nine	10 points	________
	Ten	34 points	________
TOTAL		100 points	________

- **You can have a cheat sheet 8.5 by 11 written on both sides.**
- **No access to phone, email, or computer.**
- **You need to have your ID to take the test.**
- **You need a calculator that takes a square root.**

Good Luck!

1

Question One.

Given the following scenarios, what do you expect the qqplots to look like?

> **<u>Scenario One:</u>**
> Histogram of IQ scores for a random sample of two thousand seventeen year-old students in the State of California.
>
> **<u>Scenario Two:</u>**
> Histogram of annual income in a country in which 70% earn below the mean.
>
> **<u>Scenario Three:</u>**
> Histogram of scores on an easy final exam in which more than 80% score above the mean.

If we draw the qqplots for the above scenarios, which of the following is true?

(a) For all three qqplots, the points will be on the line.
(b) For two out of three qqplots, the points will be on the line.
(c) For one out of three qqplots, the points will be on the line.
(d) We do not have enough information to decide what the qqplots will look like.

Question Two.

Professor Rodriguez teaches a small graduate class of ten students. Of these ten students, six major in science, and the rest major in other fields. If you pick five students randomly from this group, what is the probability that at least four of them will be science majors?

(a) 0.05184

(b) 0.25920

(c) 0.07776

(d) 0.33696

Question Three.

According to the exit polls of 2012, 71% of the Hispanic population voted for President Obama, with a margin of error of 2%. How would you interpret this?

(a) We are 95% confident that between 69% to 73% of the Hispanic sample voted for Obama in 2012 election.

(b) We are 95% confident that between 69% to 73% of the Hispanic population voted for Obama in 2012 election.

(c) We are 95% confident that 71% of the Hispanic population voted for Obama in 2012 election.

Question Four.

Given the following three scenarios, how would you evaluate them in terms of accuracy and precision? Note that $P = 0.40$ (proportion in the population).

Scenario A: Three samples with $\widehat{p_1} = 0.60$, $\widehat{p_2} = 0.66$, $\widehat{p_3} = 0.67$.
Scenario B: Three samples with $\widehat{p_1} = 0.20$, $\widehat{p_2} = 0.60$, $\widehat{p_3} = 0.80$.
Scenario C: Three samples with $\widehat{p_1} = 0.36$, $\widehat{p_2} = 0.37$, $\widehat{p_3} = 0.38$.

The sample sizes in all three scenarios are equal to 50.

(a) Scenario A has precision not accuracy, Scenario B has neither accuracy nor precision, and Scenario C has both accuracy and precision.

(b) Scenario A has accuracy not precision, Scenario B has both precision and accuracy, and Scenario C has neither accuracy nor precision.

(c) Scenario A has both accuracy and precision, Scenario B has precision not accuracy, and Scenario C has neither accuracy nor precision.

Question Five.

In the overall population, IQ scores follow a normal distribution with mean equal to 100 and SD = 16. We define

- Gifted as IQ > 120
- And mentally challenged as IQ < 65

Event A: Randomly picking three gifted students from the population.
Event B: Randomly picking one gifted and one mentally challenged student from the population.

Given Event A and Event B above, which of the following is true?

(a) Events A and B are both mutually exclusive.
(b) Event A is independent and event B is mutually exclusive.
(c) Event A is mutually exclusive and event B is independent.
(d) Events A and B are both independent.

Question Six.

Assume that undergraduates at a state school with 60,000 students are ambivalent toward the issue of prohibiting smoking on campus ($P = 0.50$). We have the computer take one hundred simple random samples of size 200 from this population, compute the proportion of the sample who are in favor of prohibiting smoking on campus, and for each sample construct a confidence interval. We find that ninety-five of these intervals include 0.50. What does this imply?

(a) The approximate probability that a randomly selected interval includes 0.50 is 0.95.

(b) The approximate probability that a randomly selected interval includes the sample proportion of those favoring prohibiting smoking on campus is 0.95.

(c) The approximate probability that a randomly selected interval does not include the sample proportion is 0.05.

(d) The approximate probability that a randomly selected interval includes 0.50 is 0.05.

Question Seven.

Given the following table, if you pick an individual at random, what is the probability that he will have depression or insomnia?

Psychological Disorder	Has depression	Does not have depression	Total
Has insomnia	70	30	100
Does not have insomnia	80	120	200
Total	150	150	300

Question Eight.

A researcher wants to determine the sample size needed for estimating the proportion of the population who will vote for a candidate who plans to run for governor. Given that they want to set the confidence level at 90% and the margin of error at 2%, what is the estimated value of the sample size? Show your work.

Question Nine.

The following question was given on a statistics exam:

> At a university with an enrollment of 30,000 students, we want to estimate the proportion of students who work part-time. In a random sample of 225 students, 150 students work part-time. We find the 95% confidence interval to be from 61% to 73%. Interpret the confidence interval.

John commented on those results as follows:

> "We are 95% confident that the percentage of students who work part-time in this sample is between 61% to 73%. If we select one hundred random samples, each of size 225, then about 95% of the corresponding confidence intervals would contain the true proportion of part-time workers in the population (represented by the letter P)."

If you were the professor, would you give John...

Full Credit ________ Partial Credit ________ No Credit ________

Explain why? NOT MORE THAN FOUR LINES

____

____

____

____

9

Question Ten.

In a simple random sample of 1500 young Americans 1200 did some kind of exercise.

(a) Calculate the 99% confidence interval. Show all of your work. A final answer with no work gets NO CREDIT.

(b) How can we make this confidence interval narrower?
No calculations are necessary. Use the relevant formulas to explain your answer.

(c) Suppose that in the past only 70% of young Americans did some kind of exercise. Does the confidence interval that you computed in part (a) support or refute the claim that the percentage of young Americans who do some kind of exercise has increased?

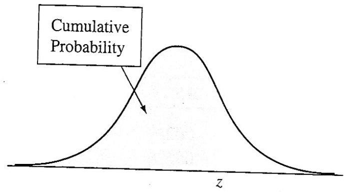
Support ________ Refute ________

Explain. NOT MORE THAN THREE LINES

____

____

____

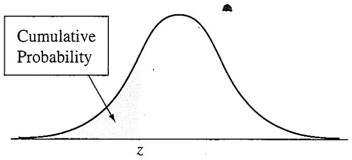


Cumulative probability for z is the area under the standard normal curve to the left of z

Standard Normal Cumulative Probabilities (*continued*)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

z	.00
3.5	.999767
4.0	.9999683
4.5	.9999966
5.0	.999999713



Cumulative Probability for z is The Area Under the Standard Normal Curve to The Left of z

Table 2: Standard Normal Cumulative Probabilities

z	.00
−5.0	.000000287
−4.5	.00000340
−4.0	.0000317
−3.5	.000233

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Statistics 10 - FINAL EXAM
Professor Esfandiari
Fall 2012

Last Name: ______________________ **First Name:** ______________________

Student ID: ______________________

Name of TA: ______________________ **Section #:** ______________________

	Question #	Possible Points	Actual Points
SCORING	1 - 10	40 points	________
	11	7 points	________
	12	10 points	________
	13	10 points	________
	14	18 points	________
	15	15 points	________
TOTAL		100 points	________

- **No access to phone, email, or computer.**
- **You need to have your ID to take the test.**
- **You need a calculator that takes a square root.**

Good Luck!

Question One.

We have a data set with 200 observations. The data has complete symmetry with four outliers on the right end of the distribution (four standard deviations above the mean) and four outliers on the left end of the distribution (four standard deviations below the mean). If we remove the eight outliers, how will the variance change?

(a) The variance will increase.

(b) The variance will decrease.

(c) The variance will not change.

(d) We need more information to answer this question.

Question Two.

Given the following information on the annual salary of a company with over 500 employees, what is the best way to find the salary of an individual who earns less than 90% of the employees.

Mean of annual income = median of annual income = \$80K
Curve has symmetry and a single mode. Data follows the normal model.
SD of annual income = \$20 K

(a) Use the Z-score corresponding to the 10^{th} percentile ($Z = -1.28$) and use the formula $Z = (X - Mean)/SD$ to solve for X.

(b) Create a table of frequencies and look up the annual salary that corresponds to the cumulative percent of 90.

(c) Use the Z-score corresponding to the 90^{th} percentile ($Z = +1.28$) and use the formula $Z = (X - Mean)/SD$ and solve for X.

(d) Create a table of frequencies and look up the annual salary that corresponds to the cumulative percent of 10.

Question Three.

A researcher believes that wearing a magnetized bracelet lowers arthritis pain. 200 arthritis sufferers are randomly assigned to wear a magnetized bracelet (population one) or a placebo bracelet (population two). After six weeks, 50 of the individuals who wore the magnetized bracelet and 40 of those who wore the placebo bracelet indicated that they had relief from arthritis pain. What would be the best way to state the null and the alternative hypotheses?

(a) $H_0 : \mu_1 - \mu_2 = 0,\ \ H_a : \mu_1 > \mu_2$

(b) $H_0 : P_1 - P_2 = 0,\ \ H_a : P_1 > P_2$

(c) $H_0 : \widehat{p_1} - \widehat{p_2} = 0,\ \ \ H_a : \widehat{p_1} > \widehat{p_2}$

(d) $H_0 : \bar{X}_1 - \bar{X}_2 = 0,\ \ ,\ H_a : \bar{X}_1 > \bar{X}_2$

Question Four.

For a population of 1000 students in introductory statistics,

200 read the book
200 read the lecture notes
200 read their own notes
50 read the book and lecture notes
50 read the book and their own notes
50 read the lecture notes and their own notes
20 read all three

If you pick a student at random, what is the probability that she reads at least two of the above?

(a) 0.11

(b) 0.15

(c) 0.02

(d) 0.17

Question Five.

A researcher has computed the coefficient of correlation between the height (x) and weight (y) of 500 newborns and he has also drawn the scatterplot (r stands for coefficient of correlation and Z stands for standard scores). Which of the following is true?

(a) $r_{x,y}$ will be larger than $r_{z(x),z(y)}$, and due to different scales, the scatterplot of Y vs. X will look different from the scatterplot of Z_x vs. Z_y.

(b) $r_{x,y}$ will be equal to $r_{z(x),z(y)}$, and the scatterplot of Y vs. X will look different from the scatterplot of Z_x vs. Z_y.

(c) $r_{x,y}$ will be equal to $r_{z(x),z(y)}$, and the scatterplot of Y vs. X will look exactly the same as the scatterplot of Z_x vs. Z_y.

(d) $r_{x,y}$ is not comparable to $r_{z(x),z(y)}$, and the scatterplot of Y vs. X is not comparable to the scatterplot of Z_x vs. Z_y.

Question Six.

Suppose you have surveyed a random sample of 100 men and 100 women and asked them how they feel about same-sex marriage. 40 of the women and 30 of the men indicate that same-sex marriage should be allowed. If you pick an individual at random, and given that they endorse same-sex marriage, what is the probability that the person is a male?

(a) 0.43

(b) 0.30

(c) 0.30

(d) 0.35

Question Seven.

A researcher wants to find out whether, in rural areas, there is any relationship between gender and attitude toward pursuing a science-related career. For two random samples of 200 female and 200 male high-school seniors in a rural area, 70 of the males and 30 of the females indicated that they aspired to pursue a science-related career.

Given that the 95% confidence interval for the difference in proportions of males and females who aspire to a science-related career is from 0.315 to 0.485, what is your conclusion regarding the hypothesis test posed by this researcher?

(a) The p-value is less than 0.05. We reject the null, and conclude that in rural areas, the proportion of male seniors who aspire to pursue a science-related career is between 0.315 to 0.485 higher than female seniors.

(b) The p-value is less than 0.05. We reject the null, and conclude that in rural areas, the proportion of female seniors who aspire to pursue a science-related career is between 0.31.5 to 0.485 higher than male seniors.

(c) The p-value is greater than 0.05. We fail to reject the null, and conclude that in rural areas, the proportion of male and female seniors who aspire to pursue a science-related career is not significantly different.

(d) The p-value is less than 0.05. We reject the null, and conclude that, in rural areas, the proportion of male seniors who aspire to pursue a science-related career is different than the percentage of female seniors.

Question Eight.

A study was conducted to find out whether the proportion of high school seniors who plan to attend a four-year college is higher for males than females. For a random sample of 10,000 males and 10,000 females, 7500 of the males and 7300 of the females indicated that they wanted to attend a four-year college. A two-sample test of proportion was conducted to test the null hypothesis and the Z-value was found to be 3.33. What is the best answer?

(a) The findings of this study have both statistical and practical significance.

(b) The findings of this study have neither practical nor statistical significance.

(c) The findings of this study have statistical but not practical significance.

(d) The findings of this study have practical but not statistical significance.

Question Nine.

A researcher wants to find out if the mean birth weight is different for working moms than for stay-at-home moms.

Descriptives	**Working Moms**	**Stay-at-home Moms**
Mean birth weight	7.15	7.42
SD	1.23	1.29

Given the above data for a random sample of 100 working moms and 100 stay-at-home moms, what is best the answer to the researchers question? Let the significance level be $\alpha = 0.05$.

(a) The p-value is greater than 0.05. We fail to reject the null, and conclude that there is no significant difference between the mean birth weight for working moms and stay-at-home moms.

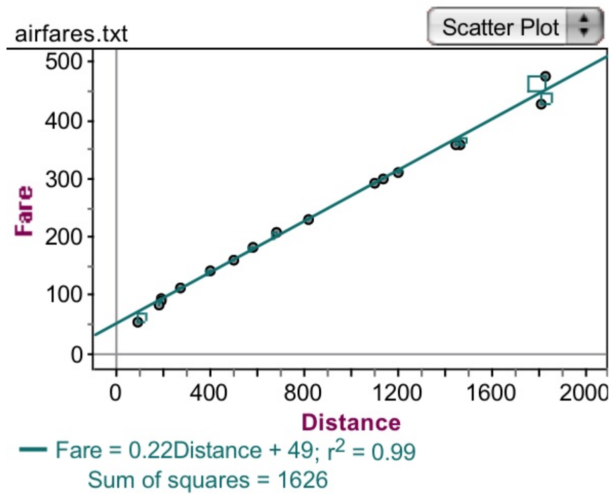
(b) The p-value is greater than 0.05. We reject the null, and conclude that there is no significant difference between the mean birth weight for working moms and stay-at-home moms.

(c) The p-value is less than 0.05. We reject the null, and conclude that the mean birth weight is higher for stay-at-home moms than working moms.

(d) The p-value is less than 0.05. We reject the null, and conclude that the mean birth weight is different for working moms and stay-at-home moms.

Question Ten.

Given the following plot, what is the best interpretation of the given information?



(a) We are predicting airfare from distance. If we increase distance by one mile, on the average, airfare increases by 49 cents. 99% of the variance in distance is explained by airfare, and 1626 is $\sum (x - \bar{x})^2$.

(b) We are predicting distance from airfare. If we increase airfare by one dollar, on the average, distance will increase 0.22 miles. 99% of the variance in distance is explained by airfare, and 1626 is $\sum e^2$.

(c) We are predicting distance from airfare. If we increase airfare by one dollar, on the average, distance will increase 49 miles. 99% of the variance in distance is explained by airfare, and1626 is $\sum (y - \bar{y})^2$.

(d) We are predicting airfare from distance. If we increase distance by one mile, on the average, airfare increases by 22 cents. 99% of the variance in airfare is explained by distance, and 1626 is $\sum e^2$.

Question Eleven.

At a major state university with 70,000 undergraduates, they want to estimate the proportion of students who endorse online teaching. Agency A suggests to choose a random sample of 2000 students and estimate a confidence interval. They are going to charge \$5000 for this. Agency B suggests to choose a sample of 5000 students by polling 50 classes of 100 or more that meet between 8:00 and 11:00 AM. Agency B charges \$2000 to do this.

If you were in charge of this project, which agency would you hire?

Agency A ________ **Agency B** ________

Explain why? NOT MORE THAN TWO LINES

Question Twelve.

You are given the following scenarios:

Scenario One: Are the parents of high school seniors with college degrees more likely to help their children with filling out college applications?

Scenario Two: Is mean annual income higher for people with graduate education compared to those with a four-year college degree?

Scenario Three: Can middle school students' standardized scores in math be predated from their standardized scores in reading?

Scenario Four: Is the average weight of the newborns whose mothers diet during pregnancy lower than seven pounds?

Which of the following statistical methods would you use for analyzing each scenarios data? Write the letter of the method next to the relevant scenario on the given line below.

(a) Correlation
(b) Linear regression
(c) One-sample test of proportion
(d) Two-sample test of proportion
(e)One-sample test of the mean
(f) Two-sample test of the mean

Statistical method recommended for **Scenario One**: ________________

Statistical method recommended for **Scenario Two**: ________________

Statistical method recommended for **Scenario Three**: ________________

Statistical method recommended for **Scenario Four**: ________________

Question Thirteen.

A study was designed to examine the effectiveness of a new medication on lowering insomnia. 200 volunteers in the 50-70 year age-range who suffered from insomnia volunteered to participate in the study. They were randomly assigned to two groups of one hundred each to either take the new medication or the placebo. Results indicted that after six weeks, 62 of the patients who took the new mediation and 33 of the patients who took the placebo showed improvement in lowering insomnia.

(a) Test the null hypothesis by the calculation of the P-value.
Let $\alpha = 0.01$. (Conduct a one-tailed test). Show all of your work.

(b) Based on the results of this study, do you recommend the new medication?

(c) If you computed the 95% CI for $P_1 - P_2$, would you expect it to include zero or not? EXPLAIN WHY. No computations are necessary.

Question Fourteen.

The following question was given on a Stats 10 exam:

A researcher wants to find out if there is a relationship between gender and political affiliation (Democrat vs. Republican). He tests the null hypothesis that the proportion of Democratic women is equal to the proportion of Democratic men ($P_1 - P_2 = 0$, where P_1 represents the proportion of Democratic women). For his sample, he finds the proportion of Democratic females and males to be 0.70 and 0.50, respectively. When he tests the null, he finds the p-value to be 0.02 and the 95% confidence interval for $P_1 - P_2$ to be 0.13 to 0.27. Assume all the necessary conditions are met.

(a) Interpret the p-value. NOT MORE THAN THREE LINES.

(b) Interpret the 95% confidence interval. NOT MORE THAN THREE LINES.

Question Fifteen.

Based on the 2011 records, the mean MCAT scores of Medical Programs in the United States follows the normal distribution with mean equal to 31.1 and standard deviation equal to 1.1. For a random sample of 20 students who got admitted to some medical school X, the mean MCAT score was equal to 31.5. The MCAT measurements were independent of each other.

(a) Determine whether the mean MCAT of the students who are admitted to medical school X is significantly higher than the population mean of 31.1. Use a significance level of $\alpha = 0.05$.

(b) Now suppose the sample consisted of 100 students who got admitted to school X and repeat the test.

(c) How do the results in part (a) and (b) compare?

Same ________ **Different** ________

Explain why? NOT MORE THAN TWO LINES

13

Statistics 10 – Midterm Two
Esfandiari, Dabagh, Johnson
Spring 2013

Name: ______________________ **UCLA ID:** ______________________

Section/TA: ______________________

Questions	**Possible points**	**Actual points**
1-12	36	----------
13	9	-----------
Total points	45	-----------
Percentage		-----------

Rules of the exam

- You can have a cheat sheet 8.5 by 11 written on both sides
- You need to have your ID to take the test
- You can have no access to I-phone, cell phone or computer
- You need to have a calculator to take the square root
- You cannot use the calculator on your phone
- You will have 50 minutes and we need to leave the room at 12:50 since there is a lecture right after and the next professor needs to set up

Question one. Given the following contingency table, what is the probability of randomly picking an individual who is male or does not do weight lifting?

	Male	Female	Total
Does weight lifting	140	60	200
Does not do weight lifting	60	40	100
Total	200	100	300

a) $\frac{200}{300} + \frac{100}{300}$

b) $\frac{200}{300} + \frac{100}{300} - \frac{60}{300}$

c) $\frac{60}{300}$

d) $\frac{60}{200}$

Question two: In a certain country, about 10 women in 1000 have breast cancer. Given that a woman does not have cancer, the chance that she tests positive for it is 0.05. What is the probability that a randomly chosen woman who is tested for breast cancer is **both free of breast cancer AND tests positive for it (a false positive)?**

a) 0.04950
b) 0.00100
c) 0.00500
d) 0.00095

Question three. In a certain state, 70% of the people who take the bar exam pass it. What is the approximate probability that at least 75% of the 200 randomly sampled people taking the bar exam (that is 150 put of 200) will pass?

a) 0.1236
b) 0.0618
c) 0.1540
d) 0.3080

Question four. We want to find out if more than 20% of the UCLA male undergraduates buy their casual clothing online. For a random sample of 300 UCLA male undergraduates, 56 indicate that they buy their casual clothing online. Donna analyzes the data and reports the following findings:

P-value = 0.002
CI = (14% to 23%)

What is the best answer?

a) She meets the requirements of CLT, and does a good job of reporting both the confidence interval and the P value.

b) She meets the requirements of CLT, but the results that she has reported are contradictory.

c) We cannot be sure that she meets the requirements of CLT because we do not know if the distribution of $\hat{p}$ follows normality.

d) She does not meet the requirements of CLT. She should not be using the Z formula to calculate the confidence interval and P-value.

Question five. John works in the administrative office of a major university with 40,000 students. At the end of a school year (June 2012), he draws the scatterplot of freshman GPA and high school GPA for the 5,000 freshmen that were admitted in September 2011. Terri distributes a survey in five courses required for all freshmen (N = 500) and asks them if they know what they plan to major in.

a) Both John and Terri are doing inferential statistics.

b) Both John and Terri are doing descriptive statistics.

c) John is doing inferential statistics and Terri is doing descriptive statistics.

d) John is doing descriptive statistics and Terri is doing inferential statistics.

Question six. A researcher wonders whether the proportion of undergraduates who like taking online classes is more than 30%. For a random sample of 200 students, 70 responded "yes." Given that the level of significance is set equal to 0.01, what is your decision regarding the null?

a) We reject the null hypothesis and conclude that the proportion of undergraduates who like taking online classes is higher than 30%.

b) We fail to reject the null and conclude that the proportion of undergraduates who like taking online classes is not significantly different from 30%.

c) We cannot answer this question because we are not sure if the distribution of sample proportions follows the normal model or not.

d) We reject the alternative hypothesis and conclude that the proportion of undergraduates who like taking online classes is not significantly different from 30%.

Question seven. Given the following output, what is your conclusion regarding the null and how do you interpret the P-value?

One sample Proportion with summary

Options

Hypothesis test results:

p : proportion of successes for population

$H_0 : p = 0.167$

$H_A : p > 0.167$

Proportion	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
p	17	100	0.17	0.037297588	0.08043416	0.4679

a) We fail to reject the null hypothesis that the proportion in the population is equal 16.7%, and conclude the probability of getting a proportion as large as 17% or higher is 0.4679.

b) We fail to reject the alternative hypothesis that the proportion in the population is larger than 16.7%, and conclude that the probability of getting a proportion as large as 17% or higher is 0.4679.

c) If the proportion in the sample is 17%, we fail to reject the null hypothesis, and conclude that the probability of getting a proportion as large as 16.7% or higher is 0.4679.

d) If the proportion in the population is 16.7%, we reject the alternative hypothesis, and conclude that the probability of getting a proportion as large as 16.7% or higher is 0.4679.

Question eight: At a nuclear research institute, they want to find out what extent of their employees are happy with their new pension plan. They have a total of 10,000 employees, 40% of whom are females. Of the 4,000 female employees, 40% are in research, and 60% are administrative. Of the 6,000 male employees, 80% are in research, and 20% are administrative. What would be the best strategy for picking a representative sample of 500 from the 10,000 employees?

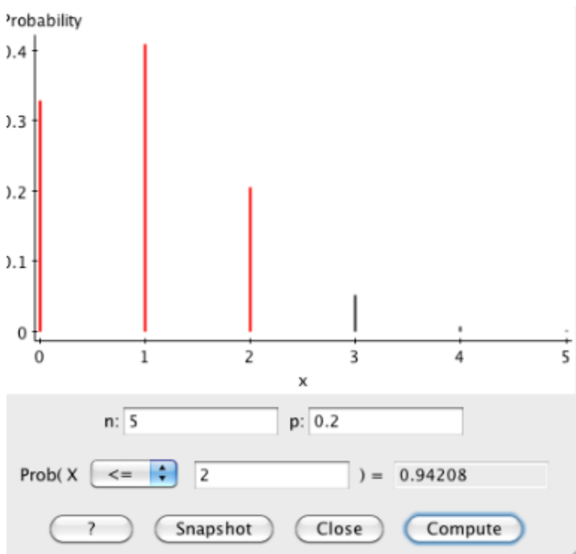
a) Cluster sampling

b) Simple random sampling

c) Stratified random sampling

d) Systematic sampling

Question nine: Given the following screenshot, what is the best answer?



a) It shows the probability of randomly guessing the correct answer to at least two out of five questions on a multiple-choice test with five options.

b) It shows the probability of randomly guessing the correct answer to at most two out of five questions on a multiple-choice test with five options.

c) It shows the probability of getting the correct answer to at least two out of five questions on a True-False test.

d) It shows the probability of getting three, four,or five correct answers on a multiple-choice exam with five questions.

6

Question ten: Given the following printout, how do you interpret the confidence interval?

Two sample Proportion with summary

Options

95% confidence interval results:

p_1 : proportion of successes for population 1

p_2 : proportion of successes for population 2

$p_1 - p_2$: difference in proportions

Difference	Count1	Total1	Count2	Total2	Sample Diff.	Std. Err.	L. Limit	U. Limit
$p_1 - p_2$	120	200	80	200	0.2	0.048989795	0.10398176	0.29601824

a) We are 95% confident that the proportion in population one is between 10.3% to 29.6% higher than population two.

b) We are 95% confident that the proportion in sample one is between 10.3% to 29.6% is higher than sample two.

c) We are 95% confident that the proportion in sample one is 60% and the proportion in sample two is 40%; give or take the margin of error.

d) We are 95% confident that the proportion in population one is 60% and the proportion in population two is 40%; give or take the margin of error.

Question eleven. Assume that the proportion of the population who endorse higher taxes for the wealthy is 60% in the overall population. Given the following three scenarios, what is the best answer?

Scenario A: Terri takes three samples and finds the $\hat{p}$'s to be 0.30, 0.32, and 0.28
Scenario B: Jane takes three samples and finds the $\hat{p}$'s to be 0.58, 0.60, and 0.62
Scenario C: Sarah takes three samples and finds the $\hat{p}$'s to be 0.50, 0.65, and 0.70

a) A has consistency but not accuracy, B has both consistency and accuracy, C has accuracy but not consistency.

b) A has consistency and accuracy, B has accuracy but not consistency, C has consistency but not accuracy.

c) A has accuracy but not consistency, B has consistency but not accuracy, C has neither accuracy nor consistency.

d) A has neither accuracy nor consistency, B has accuracy but not consistency, C has accuracy but not consistency.

Question twelve: A researcher believes the proportion of men who endorse support for nuclear energy is higher than the proportion of women. What is the best way to state the null and the alternative hypotheses?

a) $H_0: P_{men} = P_{women}$

$H_a: P_{men} > P_{women}$

b) $H_0: \widehat{p^{\wedge}}_{men} = \hat{p}_{\wedge women}$

$H_a: \widehat{p^{\wedge}}_{men} > \hat{p}_{\wedge women}$

c) $H_0: P_{men} = P_{women} = 0.50$

$H_a: P_{men} > \mathit{0.50}$

d) $H_0: \hat{p}_{men} = \hat{p}_{women} = 0.50$

$H_a: \hat{p}_{men} > \mathit{0.50}$

8

Question thirteen. Olga attends Adams High School with 5,000 students. She wants to estimate the proportion of tenth grade students who are interested in taking AP classes. She asks tenth grade students who are enrolled in two math, two science, and two social science classes (N = 225) whether they are interested in taking AP classes. She finds $\hat{p}$ to be to be 0.30 and she calculates the 95% confidence interval as follows:

$$0.30 \pm 1.96\sqrt{\frac{0.50 * 0.50}{225}} = (0.23, 0.37)$$

She then interprets the 95% confidence interval as follows:

The probability that tenth grade students in Adams High School are interested in taking AP classes is between 23% to 37%.

If you were looking for somebody to help you on a project, would you hire Olga?

a) Yes

b) No

c) With reservation

Explain why. List your reasons as (1), (2), (3), etc and explain each reasoning in NO MORE THAN TWO LINES.

Form A

Statistics 10/ Final Exam
Spring 2013: Esfandiari, Dabagh, Johnson

Last Name_______________________First Name _____________________________

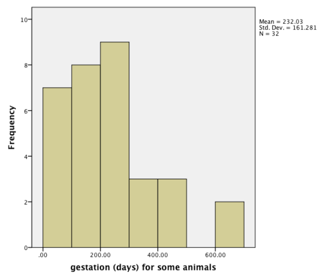
Student ID ___________________________

Email___________________________

Questions	Possible points	Actual points
1-29	87	_______________
30	13	_______________
Total points	100	_______________

No cell phones, I-phones, I-pads, or computers allowed
You can have a cheat sheet 8.5 by 11 written on both sides
You need to have a calculator that takes square root
You need to have your ID to take the test

Question 1. In the following histogram you are given the gestation period for a number of animals. Given this histogram, what are the best measures of center and spread?



a) The best measures of center and spread are mean standard deviation.
b) The best measures of center and spread are mean and IQR
c) The best measures of center and spread are median and standard deviation.
d) The best measures of center and spread are median and IQR.

Question 2. You are given standardized math scores for 200 students with IQ equal to 115 and higher and 200 students with IQ equal to 85 and lower. IQ scores follow the normal model with mean equal to 100 and standard deviation equal to 15. What would be the best way to display this data?

a) Side-by-side boxplot
b) Scatterplot
c) Contingency table
d) Side-by-side barchart

Question 3. You are given the histogram of weight for 100 football players and 100 male undergraduates in the 18-22 year age range. What is the best measure of center for this histogram?

a) The best measure of center is the mean weight of football players plus the mean weight of the undergraduates divided by two.
b) The best measure of center is the median weight of football players plus the median weight of the undergraduates divided by two.
c) The best measure of center is the median weight of football players plus the mean weight of undergraduates divided by two.
d) The nature of the data does not allow computation of a single measure of center for this histogram.

Question 4. A medical researcher wants to examine the effect of a new medication on lowering cholesterol. 300 people with cholesterol issues volunteered to participate in the study. He first divides the volunteers into two groups (below median weight and above median weight). Then he randomly assigns half of the participants in each of the two weight groups to the drug and the other half to the placebo group. After a six-month period, he measures the drop in the cholesterol level of all the participants. What is the best answer?

a) This is an experimental study. He used randomization and blocking for weight as a potential confounder. He can draw causal conclusions.
b) This is an observational study. The participants are volunteers. He cannot draw causal conclusions.
c) This is an experimental study. Blocking by weight would have been enough to warrant causal conclusions. He did not need to randomize the participants.
d) This is an observational study. He should have first selected a random sample of people with high cholesterol and then randomly assigned them to the drug and the placebo group.

Question 5. We have a data set with 30 observations. The mean is equal to 100 and standard deviation is equal to 15. The data set has four observations all of which are equal to 160. If we add four observations all of which are equal to 40, how will the standard deviation change? Three potential answers are given below.

Answer one: The standard deviation will increase.
Answer two: The standard deviation will not change because we had four outliers on the right, and we added four outliers on the left. The positive and negative deviations from the mean cancel each other out.
Answer three: We need to have the actual data set or at least the histogram in order to answer this question.

If you were the professor, how would you grade the above answers?

a) I would assign full credit to answer one and no credit to answers two and three.
b) I would assign full credit to answer two and no credit to answers one and three.
c) I would assign full credit to answer three and no credit to answers one and two.

Question 6. Given the following information on the annual salary of a company with over 5000 employees, what is the best way to find the salary of an individual who earns less than 90% of the employees.

- Mean of annual income = median of annual income = 80K
- Histogram of annual salary has symmetry and a single mode
- SD of annual income = 20 K
- 90% of the employees earn between 60K to 100K.

a) Use the Z score corresponding to the 10^{th} percentile (-1.28) and use the formula X = (X – Mean)/SD and solve for X.

b) Create a table of frequencies and look up the annual salary that corresponds to the 90^{th} percentile.

c) Use the Z score corresponding to the 90^{th} percentile (+1.28) and use the formula X = (X – Mean)/SD and solve for X.

d) Create a table of frequencies and look up the annual salary that corresponds to the 10th percentile.

Question 7. A researcher believes that average IQ of students who are enrolled in AP classes is higher than the general population (μ = 100; i.e. average IQ for the whole population is 100).
The following output was reported based on random sample of 200 students enrolled in AP classes. What is the best conclusion?

Hypothesis test results:
μ : population mean
$H_0 : \mu = 100$
$H_A : \mu > 100$

Mean	Sample Mean	Std. Err.	DF	T-Stat	P-value
μ	108	1.0606601	199	7.5424724	<0.0001

a) Reject the null hypothesis and conclude that on the average IQ of students enrolled in AP classes is higher than the general population.
b) Reject the null hypothesis and conclude that the students in this sample have higher IQ than the general population.
c) Reject the null hypothesis and conclude that we do not expect the confidence interval for the mean IQ of students enrolled in AP classes to include zero.
d) Reject the null hypothesis and conclude that since P = 0.0001, we can conclude that high IQ leads to taking AP classes.

Question 8. Given the following what is the best answer?

Event one. Being over the age of 65 vs. being under the age of 16.
Even two. Getting two "6's" in a row in tossing a fair die.

a) Events one and two are both independent
b) Event one is independent and event two is mutually exclusive.
c) Events one and two are both mutually exclusive.
d) Event one is mutually exclusive and event two is independent.

Question 9. A researcher believes that wearing a magnetized bracelet lowers arthritis pain. 200 arthritis sufferers are randomly assigned to wear a magnetized bracelet or a placebo bracelet (100 in each group). After six weeks, 50 of the individuals who wore the magnetized bracelet and 40 of those who wore the placebo bracelet indicated that they had relief from arthritis pain. What would be the best way to state the null and the alterative hypotheses?

a)

$$\mu_{Experimentl} - \mu_{Control} = 0$$

$$\mu_{Experimental} - \mu_{Control} < 0$$

b)

$$\overline{X}(Experimental) - \overline{X}(Control) = 0$$

$$\overline{X}(Experimental) - \overline{X}(Control) > 0$$

c)

$$P_{Experimental} - P_{Control} = 0$$

$$P_{Experimental} - P_{Control} > 0$$

d)

$$\hat{p}_{Experimental} - \hat{p}_{Control} = 0.10$$

$$\hat{p}_{Experimental} - \hat{p}_{Control} > 0.10$$

Question 10. For a population of 1000 students in introductory statistics,

- 200 read the book
- 200 read the lecture notes
- 200 read their own notes
- 50 read the book and lecture notes
- 50 read the book and their own notes
- 50 read the lecture notes and their own notes
- 20 read all three

If you pick a student at random, what is the probability that she reads at least two of the above?

a) 0.11
b) 0.15
c) 0.02
d) 0.17

Question 11. Jerry is provided with the following information on midterm (X) and final (Y) scores for 200 students enrolled in Stat 10. Jerry computes the coefficient of correlation between X and Y as well as the coefficient of correlation between Z_X *and* Z_Y. What is the best answer?

Midterm (X)	Final score (Y)
$\overline{X}$	$\overline{Y}$
S_X	S_y
Z_X	Z_Y

a) The coefficient of correlation between X and Y will be larger than the coefficient of correlation between Z_X *and* Z_Y.

b) The coefficient of correlation between X and Y will be smaller than the coefficient of correlation between Z_X *and* Z_Y.

c) The coefficient of correlation between X and Y will be equal to the coefficient of correlation between Z_X *and* Z_Y.

d) This question cannot be answered because we have not been provided qqplot to make sure that the assumption of normality is tenable.

Question 12. Suppose you have surveyed a random sample of 100 men and 100 women and asked them how they feel about same sex marriage. 40 of the women and 30 of the men indicate that same sex marriage should be allowed. If you pick an individual at random, and given that they endorse same sex marriage, what is the probability that he is a male?

a) 0.43
b) 0.15
c) 0.20
d) 0.35

Question 13. A researcher wonders whether in rural areas the proportion of male seniors who want to pursue a science-related career is higher than female seniors. For two random samples of 200 female and 200 male seniors, 70 of the males and 30 of the females indicated that they aspired to pursue a science-related career. He finds the 95% confidence interval for the difference in proportions to be from 0.315 to 0.485. If we set alpha equal to 0.05, what would be the best answer?

a) The p-value should be less than 0.05. We reject the null hypothesis. We conclude that in rural areas the proportion of male seniors who aspire to pursue a science-related career is between 0.315 to 0.485 higher than female seniors.

b) The p-value should be less than 0.05. We reject the null hypothesis. We conclude that in rural areas the proportion of female and male seniors who aspire to pursue a science-related career is 0.31.5 and 48.5 respectively.

c) The p-value is greater than 0.05. We fail to reject the null hypothesis. We conclude that in rural areas the proportion of male and female seniors who aspire to pursue a science-related career is not significantly different.

d) The p-value is less than 0.05. We do not reject the alternative hypothesis. We conclude that in rural areas the proportion of male seniors who aspire to pursue a science-related career is not different from female seniors.

Question 14. A researcher wants to find out if the mean birth weight of newborns is different for working moms than for stay-home moms.

Given the following data for a random sample of 100 working moms and 100 stay-home moms, what is the answer to the question? Let the significance level be equal to 0.05.

Descriptives	Working moms	Stay home moms
Mean of birth weight	7.15	7.42
SD	1.23	1.29

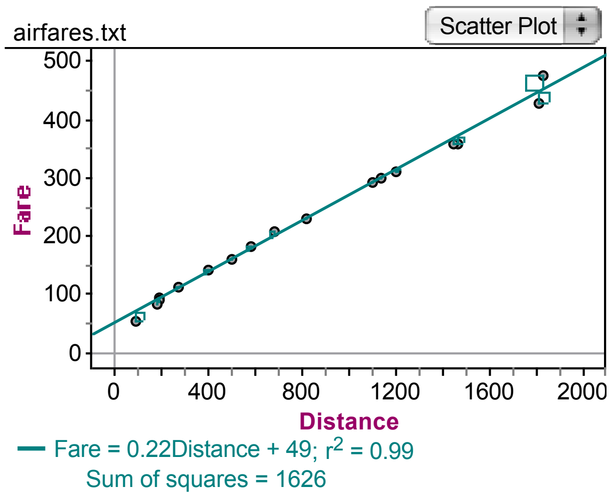
a) The p-value is greater than 0.05. We fail to reject the null hypothesis. We expect the 95% confidence interval for the difference of the two means not to include zero.

b) The p-value is greater than 0.05. We fail to reject the null hypothesis. We conclude that the mean birth weight of newborns is not different for working moms and stay-home moms.

c) The p-value is less than 0.05. We reject the null hypothesis and we expect the 95% confidence interval for the difference of the two means not to include zero.

d) The p-value is less than 0.05. We reject the null hypothesis and we conclude that the average weight of newborns is higher for stay home moms.

Question 15. Given the following plot, what is the best interpretation of the given information?



a) We are predicting fare from distance. If we increase distance by one mile, on the average, fare increases by 49 cents. 99% of the variance in distance is explained by fare, and 1626 is $\sum(x - xbar)^2$

b) We are predicting distance from fare. If we increase fare by one dollar, on the average distance will increase 0.22 miles. 99% of the variance in distance is explained by fair, and 1626 is $\sum e^2$

c) We are predicting distance from fare. If we increase fare by one dollar, on the average distance will increase 49 miles. 99% of the variance in distance is explained by fair, and1626 is $\sum(y - ybar)^2$

d) We are predicting fare from distance. If we increase distance by one mile, on the average, fare increases by 22 cents. 99% of the variance in fare is explained by distance, and 1626 is $\sum e^2$.

Question 16. Given the following two outputs, what is the best conclusion that you can draw about output one?

Output one

H0: $\mu_1 - \mu_2 = 0$

$\overline{X}(1) = 101$

$\overline{X}(2) = 102$

$N_1 = N_2 = 2000$

df = 3998, t = 2.108, P = 0.0351

Output two

H0: $\mu_1 - \mu_2 = 0$

$\overline{X}(1) = 101$

$\overline{X}(2) = 102$

$N_1 = N_2 = 200$

df = 398, t = 0.667, P = 0.505

a) Output one is an example of statistical but not practical significance.
b) Output one is an example of both statistical and practical significance.
c) Output one is an example of practical but not statistical significance.
d) Output one is an example of neither practical nor statistical significance.

Questions 17- 21: Which of the following statistical methods (a through e) will you use to analyze the data and answer research questions asked in scenarios one to five? YOU CANNOT TRANSFORM THE VARIABLES. FOR EXAMPLE YOU CANNOT TAKE NUMERICAL VARIABLES AND MAKE THEM CATEGORICAL. YOU NEED TO USE THEM THE WAY THEY ARE.

Question 17
Scenario one. Are the parents of ninth grade students with college degrees more likely to help them with math homework?

a) One-sample test of proportion
b) Two-sample test of proportion
c) One-sample test of the mean
d) Two-sample test of the mean
e) Paired sample test of the mean

Question 18
Scenario two. Is the mean annual income for lawyers who graduate from the top 20 schools higher than those who graduate from other law schools?

a) One-sample test of proportion
b) Two-sample test of proportion
c) One-sample test of the mean
d) Two-sample test of the mean
e) Paired sample test of the mean

Question 19
Scenario three. Two hundred at risk students were assigned to participate in a workshop designed to improve self-concept. We want to know if the students' self-concept improved after participation in this workshop. Self concept was measured before and after participation in the workshop by administration of a survey (range of scores 20-60).

a) One-sample test of proportion
b) Two-sample test of proportion
c) One-sample test of the mean
d) Two-sample test of the mean
e) Paired sample test of the mean

Question 20
Scenario four. Is the average GREV (Verbal score on the Graduate Record Exam) of the students who are admitted to the top ten graduate schools in the field of psychology at least 750?

a) One-sample test of proportion
b) Two-sample test of proportion
c) One-sample test of the mean
d) Two-sample test of the mean
e) Paired sample test of the mean

Question 21
Scenario five. Is the proportion of adults who endorse same sex marriage at least 0.50?

a) One-sample test of proportion
b) Two-sample test of proportion
c) One-sample test of the mean
d) Two-sample test of the mean
e) Paired sample test of the mean

Question 22. How would you interpret the following printout?

Options

95% confidence interval results:

μ_1 : mean of population 1

μ_2 : mean of population 2

$\mu_1 - \mu_2$: mean difference

(with pooled variances)

Difference	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
$\mu_1 - \mu_2$	5	1.0663049	108	2.8863988	7.113601

a) We reject the null hypothesis that mean in population one is equal to the mean in population two. We are 95% confident that, on the average, the mean in population one is 2.88 to 7.11 points higher than the mean in population two.

b) We reject the null hypothesis that mean in population one is equal to the mean in population two. We are 95% confident that the mean in the populations one and two are 2.88 and 7.11 respectively.

c) We reject the null hypothesis that mean in population one is equal to the mean in population two. We conclude that there is a 95% chance for that the mean difference between population one and population two to be between 2.88 to 7.11.

d) We reject the null hypothesis that mean in population one is equal to the mean in population two. We are 95% confident that if we construct 100 confidence intervals, in 95 cases the confidence interval will be between 2.88 to 7.11.

Question 23. Given the following two outputs, what is the best conclusion?

Printout one

Options

Hypothesis test results:

μ_1 : mean of population 1

μ_2 : mean of population 2

$\mu_1 - \mu_2$: mean difference

$H_0 : \mu_1 - \mu_2 = 0$

$H_A : \mu_1 - \mu_2 \neq 0$

(with pooled variances)

Difference	Sample Mean	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	1	1.2909944	58	0.7745967	0.4417

Printout two

Options

Hypothesis test results:

μ_1 : mean of population 1

μ_2 : mean of population 2

$\mu_1 - \mu_2$: mean difference

$H_0 : \mu_1 - \mu_2 = 0$

$H_A : \mu_1 - \mu_2 \neq 0$

(with pooled variances)

Difference	Sample Mean	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	1	0.31622776	998	3.1622777	0.0016

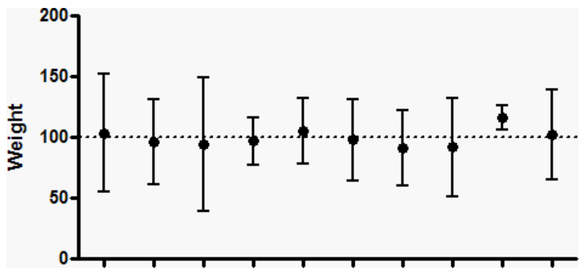
a) Due to the larger sample size, degrees of freedom increases and we can use the Z test instead of the t-test. That is why printout two displays both statistical and practical significance.

b) Due to the larger sample size, the standard error decreases and P value decreases. That is why in printout two the small difference between the mean of the two samples looks statistically significant.

c) Due to the larger sample size, degrees of freedom increases and we can use the Z test instead of the t-test. That is why printout two displays neither practical nor statistically significance.

d) Neither the degrees of freedom nor the sample size affects the sample mean or the level of significance. That is why printout two displays the importance of using a larger sample.

Question 24. In the following you are given ten confidence intervals based on random samples of N = 5. If we increase the sample size to N = 50, how will the ten confidence intervals change?



a) The confidence intervals will become narrower and more precise.
b) The behavior of confidence intervals is not predictable because each sample is different.
c) The confidence intervals will become wider and more precise.
d) The change in confidence interval depends on the value of sample mean.

Question 25. The distribution of income in a country is right skewed with mean equal to 30K and standard deviation equal to 20K. If you choose, 100 random samples of size five (N = 5), and draw the distribution of sample means, how will it look like?

a) The distribution of sample means will be normal.
b) The distribution of sample means will be symmetrical but not normal.
c) We need to know the size of the population in order to answer this question.
d) The distribution of sample means will be right skewed.

Question 26 . At a polling agency they want to estimate the proportion of residents who would vote for a candidate running for senate. Given that they are interested in 90% confidence and margin of error of 2%, how large of a sample do you recommend?

a) Around 1500
b) Around 2400
c) Around 750
d) Around 1200

Question 27. At a school district they want to estimate the proportion of ninth grade students who felt comfortable transferring from middle school to high school. This district has four high schools and each high school has five ninth grade classes. The researcher randomly chooses one out of the five ninth grade classes in each of the four high schools. Then he asks everybody in the four classes chosen to indicate how comfortable they felt with transferring to high school.

The researcher used

a) Systematic sampling
b) Stratified sampling
c) Cluster sampling
d) Simple random sampling

Question 28. Given the following output, and given that there are no outliers in the data, what is the best answer? (CLT conditions are met)

95% confidence interval results:
$\mu_1 - \mu_2$: mean of the paired difference between midterm two and midterm one

Difference	Sample Diff.	Std. Err.	DF	L. Limit	U. Limit
midterm two – midterm one	2.4666667	0.8157186	14	0.7171243	4.216209

a) We reject the null hypothesis resulting from the t-test. We are 95% confident that on the average, the scores on midterm two are between 0.72 to 4.22 points higher than midterm one.

b) We fail to reject the null hypothesis resulting from the Z-test. We conclude that on the average there was no statistically significant difference between the scores on midterm one and midterm two.

c) We fail to reject the null hypothesis resulting from the t-test. We expect the P value or the probability of rejecting the true null to be more than 0.05.

d) We reject the null hypothesis resulting from the t-test. There is a 95% chance for the scores to improve an average of 2.47 points from midterm one to midterm two.

Question 29. Based on the data from the Insurance Information Institute, as age increases from twenty years to forty years, the death rate per 100,000 drivers decreases from 30 to 5. It remains constant between the ages of 40 to 60. As age increases from 60 to 90, the death rate per 100,000 drivers increases from 5 to 25. Jerry who is working for the Insurance Information Institute analyzed this data and reported the following:

The coefficient of correlation between age and death rate per 100,000 drivers is equal to zero. We can conclude that age is not the cause of increasing death rate in automobile accidents.

If you were looking for somebody to work on your data, which of the following will you do?

a) I would hire Jerry. He did a good job.
b) I would hire Jerry with reservation. He used the right statistical method. But, his interpretation is incorrect.
c) I would hire Jerry with reservation. He used the right statistical method. He forgot to draw the scatterplot.
d) I would not hire Jerry. His work is not acceptable.

Question 30

95% confidence interval results:

p_1 : proportion of successes for population 1

p_2 : proportion of successes for population 2

$p_1 - p_2$: difference in proportions

Difference	Count1	Total1	Count2	Total2	Sample Diff.	Std. Err.	L. Limit	U. Limit
$p_1 - p_2$	120	200	60	200	0.3	0.047434166	0.20703074	0.39296925

a) Show how the standard error was calculated? (four points)

b) Show how the confidence interval was calculated. (four points)

c) Interpret the confidence interval within context. (five points)

Statistics 10/ Final Exam
Summer 2013: Esfandiari, Dabagh, Johnson
Form A

Last Name______________________ First Name ____________________________

Student ID _____________________________________

Email____

Questions	Possible points	Actual points
1-20	100	______________
Total points	100	______________

No cell phones, I-phones, I-pads, or computers allowed
You need to have a calculator that takes square root
You need to have your ID to take the test

Question one. Research has shown that the proportion of teenagers who text and drive is higher than the proportion of adults who text and drive. What would be the best way to state the null and the alternative hypotheses?

a)
H0: P adults – P teenagers = 0
Ha: P adults > P teenagers

b)
H0: P^ adults – P^ teenagers = 0
Ha: P^ adults > P^ teenagers

c)
H0: $\overline{X}$ adults – $\overline{X}$ teenagers = 0
Ha: $\overline{X}$ adults > $\overline{X}$ teenagers

d)
H0: μ teenagers – μof adults= 0
Ha: μof teenagers –μof adults> 0

Question two. SAT quantitative scores follow the normal model with mean equal to 500 and standard deviation equal to 100. Jose scores 625 on this exam, what is the best conclusion?

a) Jose scored higher than 62% of the population.
b) Jose scored higher than 89% of the population.
c) Jose scored higher than 11% of the population
d) Jose scored lower than 89% of the population.

Question three. Professor Stewart teaches a class of statistics with forty students. The students have been divided into eight groups of five (five students per group). Group one consists of two math majors, two statistics majors, and one science major. If you pick three students at random from group one, what is the probability that at most one (means one or less) of them will be a statistics major?

a) 0.352
b) 0.648
c) 0.432
d) 0.144

Question four.
Question 10. For a population of 1000 students in introductory statistics,

- 200 read the book
- 200 read the lecture notes
- 200 read their own notes
- 50 read the book and lecture notes
- 50 read the book and their own notes
- 50 read the lecture notes and their own notes
- 20 read all three

If you pick a student at random, what is the probability that she reads at least two of the above?

a) 0.15
b) 0.02
c) 0.11
d) 0.17

Question five. Which of the following statistical methods will you use to answer this question: "The percentage of the American Public that endorses having health insurance for everybody is at least 80%."

a) Correlation
b) Regression
c) One sample test of proportion
d) Two-sample test of proportion

Question six. Which of the following statistical methods will you use to answer this question: "Are the parents of high school seniors with a graduate degree more likely to help their children with filling out their college applications than parents without a graduate degree?"

a) Correlation
b) Regression
c) One sample test of proportion
d) Two-sample test of proportion

Question seven. On a statistics exam, the students are asked to test the null hypothesis that a dice is fair. They are required to do this by the calculation of the P value and the 95% confidence interval and check their results. They are asked to set alpha = 0.05

Jonah calculates the P value to be 0.12. He also calculates the confidence interval to be from 0.14 to 0.17. He uses Z = +/-1.96

How would you grade Jonah's paper?

a) Jonah should get full credit because the conclusions that he draws about the null hypothesis via the P value and confidence interval are consistent.
b) Jonah should be given no credit because his results are contradictory and lead to different conclusions about the null hypothesis.
c) Jonah should be given full credit full credit because based on both the P value and the confidence interval he rejects the null hypothesis.
d) Jonah should be given partial credit because based on the P value he fails to reject the null hypothesis. But, his confidence interval does not include 0.50.

Question eight. In a large school district with 50,000 employees, the annual salary of teachers follows the normal model with mean equal to 60K and standard deviation equal to 10K. What is the annual salary of a teacher who earns more than 25% of the teachers employed in this district?

a) His salary should be around 45K
b) His salary should be around 80K
c) His salary should be around 67K
d) His salary should be around 53K

Question nine. Suppose that you choose a random sample of UCLA students (50 males and 50 females) and ask them whether they enjoy classical music (Yes, No). Which of the following methods will you use to show whether gender and enjoyment of classical music are independent of each other?

a) If probability of (smoke ∪ enjoyment) = P(smoke) * P(enjoyment), then smoking and enjoyment of classical music are independent events.
b) If the probability of (smoke │enjoyment) = P(enjoyment), then smoking and enjoyment of classical music are independent.
c) If the probability of (smoke │enjoyment) = P(smoke), then smoking and enjoyment are independent events.
d) If probability of (smoke ∩ enjoyment) = P(smoke) + P(enjoyment), then smoking and enjoyment of classical music are independent events.

Question ten. A fair dice is tossed 100 times (A dice has six sides). What is the probability of getting at least fifteen "6" 's?

a) This probability is about 33%
b) This probability is about 67%.
c) This probability is about 15%.
d) This probability is about 85%.

Question eleven. You are watching the news and you see the following:
The proportion of the American public who endorses spending more money on global warming is…

45% +/- 3%

How was the 3% calculated? (* means multiplied by)?

a) Proportion in the population +/- 1.96 * $\sqrt{P*(1-P)/n}$

b) Proportion in the sample +/- 1.96 * $\sqrt{n*P*(1-P)}$

c) Proportion in the population +/- 1.96* $\sqrt{n*p\hat{}*(1-p\hat{})}$

d) Proportion in the sample +/-1.96* $\sqrt{(p\hat{}*(1-p\hat{}))/n}$

Question twelve. You want to estimate the proportion of UCLA students that have some kind of a scholarship. For a 99% confidence interval with a margin of error of 3 percent, estimate the required sample size.

a) The sample size is around 1067
b) The sample size is around 1834
c) The sample size is around 756
d) The sample size is around 1500

Question thirteen. In a polling they wonder if the American public is pro allocation of more funding to stem cell research. The following printout resulted from a random sample of 250 participants. Given the following output, what is the best conclusion?

95% confidence interval results:
p : proportion of successes for population
Method: Standard-Wald

Proportion	Count	Total	Sample Prop.	Std. Err.	L. Limit	U. Limit
p	150	250	0.6	0.030983867	0.5392727	0.66072726

a) There is strong evidence that American people support allocation of more funding to stem cell research.
b) There is strong evidence that American public are indifferent toward allocation of more funding to stem cell research.
c) There is strong evidence that American public do not support allocation of more funding to stem cell research.
d) We are 95% confident that 54% to 66% of the sample tested support allocation of more funding to stem cell research.

Question fourteen. Which one of the following scenarios represents independent events and which one represents mutually exclusive events?

Scenario one: Randomly choosing two people from the population. The first person is older than sixty-five and the second person is younger than sixteen.

Scenario two: Getting two "6's" in a row in tossing of a fair die.

a) Events represented in both scenarios are independent
b) Event represented in scenario one is independent and event represented in scenario two is mutually exclusive.
c) Events represented in both scenarios are mutually exclusive.
d) Event represented in scenario one is mutually exclusive and event represented in scenario two is independent.

Question fifteen. Suppose you have surveyed a random sample of 100 men and 100 women and asked them how they feel about same sex marriage. 40 of the women and 30 of the men indicate that same sex marriage should be allowed. If you pick an individual at random, and given that they endorse same sex marriage, what is the probability that he is a male? Complete the following table prior to answering the question asked.

	male	female	
endorse	30		
doesn't endorse		60	
	100	100	200

a) Around 0.15
b) Around 0.20
c) Around 0.35
d) Around 0.43

Question sixteen. John takes a random sample of 200 UCLA students and asks them if they plan to do an internship sometime in their field. 125 respond "Yes". He uses this information and finds the 95% confidence interval. Jerry asks his classmates enrolled in Economics 101 whether they endorse a tax increase for the wealthy. He calculate the P^ and creates a pie chart.

a) Both John and Jerry are doing inferential statistics.
b) Both John and Jerry are doing descriptive statistics.
c) John is doing descriptive statistics and Jerry is doing inferential statistics.
d) John is doing inferential statistics and Jerry is doing descriptive statistics.

Question seventeen. At a school district, they want to estimate the proportion of ninth grade students who felt comfortable transferring from middle school to high school. This district has four high schools and each high school has five ninth grade classes. The researcher randomly chooses one out of the five ninth grade classes in each of the four high schools. Then he asks everybody in the four classes chosen to indicate how comfortable they felt with transferring to high school. The researcher used

a) Systematic sampling
b) Stratified sampling
c) Cluster sampling
d) Simple random sampling

Question eighteen. Given the following output, what is the null hypothesis being tested and what conclusion do you draw about testing the null hypothesis?

95% confidence interval results:
p_1 : proportion of successes for population 1
p_2 : proportion of successes for population 2
$p_1 - p_2$: difference in proportions

Difference	Count1	Total1	Count2	Total2	Sample Diff.	Std. Err.	L. Limit	U. Limit
$p_1 - p_2$	150	200	100	200	0.25	0.046770718	0.15833108	0.34166893

a) We are testing the null hypothesis that the proportion in the two samples are equal. We reject the null hypothesis.
b) We are testing the null hypothesis that the proportion in the two samples are equal. We fail to reject the null hypothesis.
c) We are testing the null hypothesis that the proportion in the two populations are equal. We fail to reject the null hypothesis.
d) We are testing the null hypothesis that the proportion in the two populations are equal. We reject the null hypothesis.

Question nineteen. We are wondering if a coin is fair or not.

H0: P = 0.50
Ha: P # 0.50
We get 47 heads in 100 tosses of this coin. What is the P value?

a) The P value is around 0.55
b) The P value is around 0.27
c) The P value is around 0.47
d) The P value is around 0.60

Question twenty. Let us assume that the proportion in the population is equal to 0.40. Given the following three scenarios, what is the best answer?

Scenario A. For three samples of N = 200, p^ was found to be 0.60, 0.61, and 0.62 respectively.
Scenario B. For three samples of N = 200, the p^ was found to be 0.35, 0.40, and 0.45 respectively.
Scenario C. For three samples of N = 200, the p^ was found to be 0.40, 0.41, and 0.42 respectively.

a) Scenario A has accuracy but not precision, scenario B has both accuracy and precision, and scenario C has precision but not accuracy.
b) Scenario A has precision but not accuracy, scenario B has accuracy but not precision, and scenario C has both precision and accuracy.
c) Scenario A has both precision and accuracy, scenario B neither accuracy nor precision, and scenario C has accuracy but not precision.

References

[1] American Statistical Association (2005) GAISE college report. Retrieved October 16, 2007.

[2] Anderson, L.W., Krathwohl, D. R. , Airasian P.W. , Cruikshank K.A., Mayer R. E., et al. (2001) A taxonomy for learning, teaching, and assessing: A revision of Bloom s taxonomy of educational objectives, abridged edition. White Plains, NY: Longman.

[3] Bloom, B., Hastings, J., & Maduas, M. (1971). Handbook of Formative and Summative Evaluation of Student Learning. Mc Graw Hill.

[4] Bloom, B., Englehart, M., Furst, E., Hill, W., & Krathwohl, D. (1956) : Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York, Toronto: Longmans, Green.

[5] Bush HM, Daddysman J, Charnigo R (2014) Improving Outcomes with Bloom′ s Taxonomy: From Statistics Education to Research Partnerships. J Biomet Biostat 5: e130.

[6] Chance, B. L. (2002). Components of Statistical Thinking and Implications for Instruction and Assessment. Journal of Statistics Education, 10(3).

[7] Cobb, G.W. (1992). Teaching statistics: more data, less lecturing in Heeding the Call for Change, Lynn Steen, editor, MAA Notes, Number 22, 3-43.

[8] Esfandiari, M., & Phares, S. (2004). Testing for Upper Level Thinking in Lower Division and Upper Division Statistics Classes and the Role of Technology. UCLA Department of Statistics Lecture Series.

[9] Esfandiari, M., Barr, C., & Sugano, A.P. (2007). Using Technology to Enhance Teaching, Learning, and Assessment in Statistics Education. University of California, Riverside. Invited Paper.

[10] Esfandiari, M., & Nguyen, H. (2008). Development of an Automated Essay Grading Software for Statistics (AEGSS): A Prototype. UCLA Department of Statistics Lecture Series.

[11] Esfandiari, M., Nguyen, H., Yaglovskaya, Y., Gould, R. (2010). "Enhancing statistical literacy through short open-ended questions that involve context, data, and upper level thinking," Proceedings of 8th International Conference on Teaching Statistics.

[12] Forehand, M. (2005) Blooms taxonomy: Original and Revised. In M. Orey (Ed.), Emerging perspectives on learning, teaching, and technology.

[13] Franklin, C., Garfield, J. (2006) Guidelines for Statistics Education Endorsed by ASA Board of Directors. Amstat News (Education)

[14] Garfield, J., Aliaga, M., Cobb, G., Cuff, C., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, R., Utts, J., Velleman, P., & Witmer, J., (2005). GAISE College Report, American Statistical Association. Online: http://www.amstat.org/education/GAISE/ GAISECollege.htm.

[15] Gould, R., Davis, G., Patel, R., Esfndiari, M., (2010). "Enhancing conceptual understanding with data driven labs," Proceedings of 8th International Conference on Teaching Statistics.

[16] Krathwohl ,D. R. (2002) A revision of Bloom ′ s taxonomy: An overview. Theory into practice 41: 212-218.

[17] Moore, D. S. (1998). Should mathematicians teach statistics? The College Mathematica Journal, 19, 3-35.

[18] Rumsey D. J. (2002) Statistical literacy as a goal for introductory statistics courses. Journal of Statistics Education 10: 6-13.

[19] Singer, J. S., & Willet, J. (1990). Improving the teaching of applied statistics: Putting the data back into data analysis. The American Statistician, 44(3), 223-230.

[20] Wilson, W. J. (1992) American Statistical Association. The American Statistician 46: 295-298.

[21] Wittrock, M. C. (1974). A generative model of mathematics learning. Journal for Research in Mathematics Education, 5(4), 181-197.