

Classroom Voting Questions: Elementary Statistics

Chapter 1: Looking at Data - Distributions

Section 1.1: Displaying Distributions with Graphs

1. A data set consists of fifty three-digit numbers ranging from 180 to 510. The best choice for stems in a stem-and-leaf display would be to use _____.
 - (a) 1 digit stems (1, 2, ..., 5)
 - (b) 2 digit stems (18, 19, ..., 51)
 - (c) 3 digit stems (180, 181, ..., 510)
2. Which of the following statements is most completely true in comparing an appropriately drawn histogram to a stem-and-leaf display of the same data?
 - (a) Both convey the same information about the shape of the distribution.
 - (b) Both convey the same information about gaps in the distribution.
 - (c) Both convey the same information about outliers.
 - (d) Both convey the same amount of information generally.
 - (e) Two from (A)-(D) are correct.
 - (f) Three from (A)-(D) are correct.
 - (g) All from (A)-(D) are correct.
3. Suppose you take a random sample of 10 juniors who bought the same model laptop when they were freshmen. You test these 10 laptops to determine how long their batteries last before needing to be recharged, and you obtain the following data (in hours): 1.2, 1.3, 3.8, 3.9, 3.9, 4.0, 4.1, 4.1, 4.2, 4.3. What should be done with the values 1.2 and 1.3? Which of the following is the best course of action?
 - (a) Delete them from the data set since they are outliers.
 - (b) Keep them in the data set even though they are outliers.
 - (c) Determine why these values were so much lower than the rest, then delete them.
 - (d) Determine why these values were so much lower than the rest, then keep them in the data set, provided they weren't due to data entry errors.

Section 1.2: Describing Distributions with Numbers

4. In a certain university there are three types of professors. Their salaries are approximately normally distributed within each of the following types:

- Assistant Professors make a median salary of \$50K, with a minimum of \$40K and a maximum of \$60K.
- Associate Professors make a median salary of \$65K per year, with minimum of \$57K and a maximum of \$80K.
- Full Professors make a median salary of \$90K per year, with a minimum of \$70K and a maximum of \$110K.

There are 1600 total Professors at this University, with the following distribution: 50% of all Professors are Assistants, 30% are Associates, and 20% are Fulls.

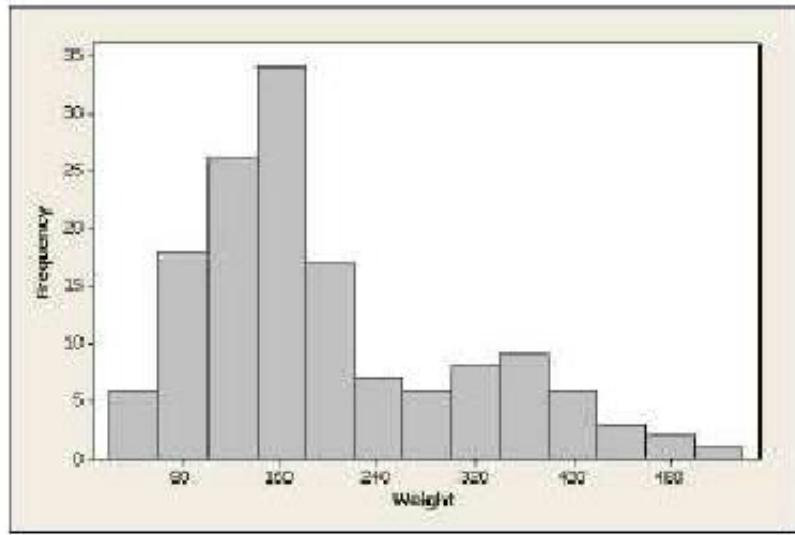
What can we say about the average salary at this university?

- (a) mean < median
- (b) mean = median
- (c) mean > median
- (d) insufficient information

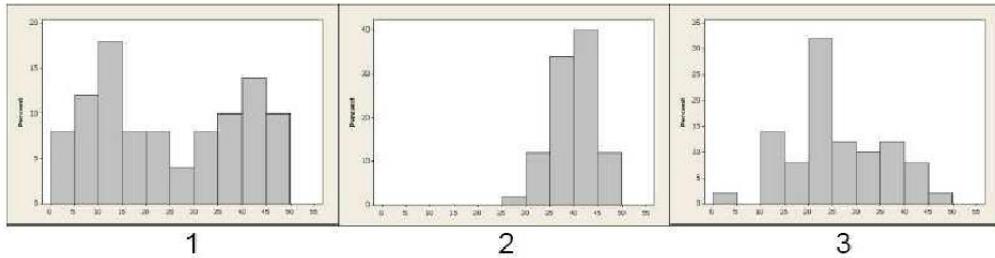
5. Many individuals, after the loss of a job, receive temporary pay unemployment compensation until they are re-employed. Consider the distribution of time to re-employment as obtained in an employment survey. One broadcast reporting on the survey said that the average time until re-employment was 4.5 weeks. A second broadcast reported that the average was 9.9 weeks. One of your colleagues wanted a better understanding of the situation and learned (through a Google search) that one report was referring to the mean and the other to the median and also that the standard deviation was about 14 weeks. Knowing that you are a statistically-savvy person, your colleague asked you which is most likely the mean and which is the median?

- (a) 4.5 is the mean and 9.9 is the median.
- (b) 4.5 is the median and 9.9 is the mean.
- (c) Neither (A) nor (B) is possible given the SD of the data.
- (d) I am not a statistically-savvy person, so how should I know?

6. For the data set displayed in the following histogram, which would be larger?



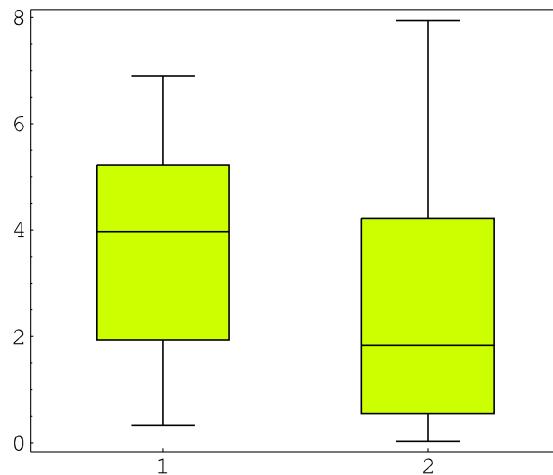
- (a) mean
 (b) median
 (c) Can't tell from the given histogram.
7. Why is the term $(n - 1)$ used in the denominator of the formula for sample variance?
- (a) There are $(n - 1)$ observations.
 (b) There are $(n - 1)$ uncorrelated pieces of information.
 (c) The $(n - 1)$ term gives the correct answer.
 (d) There are $(n - 1)$ samples from the population.
 (e) There are $(n - 1)$ degrees of freedom.
8. Which of the three histograms shown summarizes the data set with the smallest standard deviation?



9. Suppose your statistics instructor tells you that you scored 70 on an exam and that the class mean was 74. You should hope that the standard deviation of exam scores was _____.

- (a) Small
- (b) Large

10. Below are boxplots for two data sets.



TRUE or FALSE: There is a greater proportion of values outside the box for the set on the right than for the set on the left.

- (a) True
- (b) False

11. The five-number summary for all student scores on an exam is 29, 42, 70, 75, 79. Suppose 200 students took the test. How many students had scores between 42 and 70?

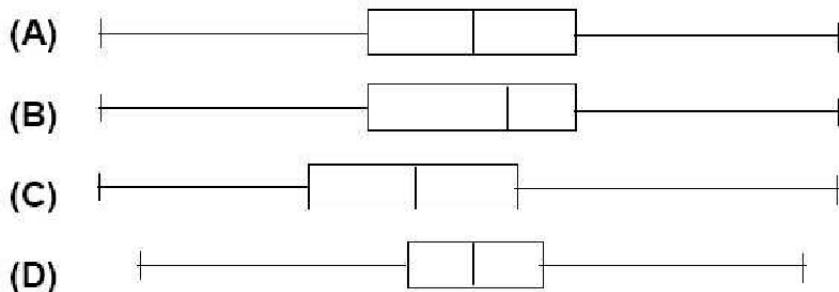
- (a) 25
- (b) 28
- (c) 50
- (d) 100

12. The five-number summary for all student scores on an exam is 40, 60, 70, 75, 79. Suppose 500 students took the test. What should you conclude about the distribution of scores?

- (a) Skewed to the left.
- (b) Skewed to the right.
- (c) Not skewed.
- (d) Not enough information given to determine skew.

Section 1.3: Density Curves and Normal Distributions

13. If a large sample were drawn from a normal distribution and accurately represented the population, which of the following is most likely to be a box plot of that sample?



(E) Two from (A)-(D) are correct.

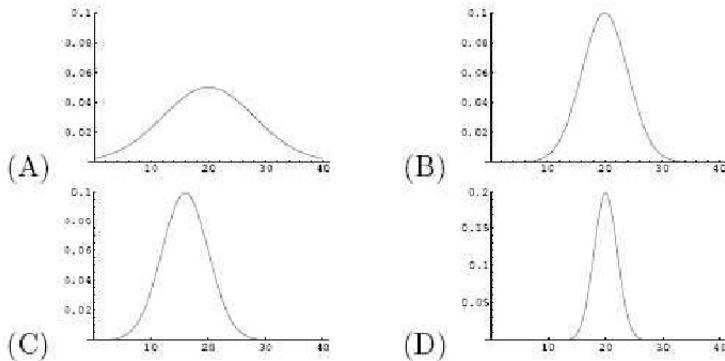
(F) Three from (A)-(D) are correct.

(G) All from (A)-(D) are correct.

14. Consider the continuous random variable X = the weight in pounds of a randomly selected newborn baby born in the United States last year. Suppose that X can be modeled with a normal distribution with mean $\mu = 7.57$ and standard deviation $\sigma = 1.06$. If the standard deviation were $\sigma = 1.26$ instead, how would that change the graph of the pdf of X ?

- (a) The graph would be narrower and have a greater maximum value.
- (b) The graph would be narrower and have a lesser maximum value.
- (c) The graph would be narrower and have the same maximum value.
- (d) The graph would be wider and have a greater maximum value.
- (e) The graph would be wider and have a lesser maximum value.

- (f) The graph would be wider and have the same maximum value.
15. Consider the continuous random variable X = the weight in pounds of a randomly newborn baby born in the United States during 2006. Suppose that X can be modeled with a normal distribution with mean $\mu = 7.57$ and standard deviation $\sigma = 1.06$. If the mean were $\mu = 7.27$ instead, how would that change the graph of the pdf of X ?
- The graph would be shifted to the left.
 - The graph would be shifted to the right.
 - The graph would become more negatively skewed.
 - The graph would become more positively skewed.
 - The graph would have a greater maximum value.
 - The graph would have a lesser maximum value.
16. If X is a normal random variable with mean $\mu = 20$ and standard deviation $\sigma = 4$, which of the following could be the graph of the pdf of X ?



17. Yogurt is sold in cartons labeled as containing 6 oz, but the actual contents vary slightly from container to container. Suppose that the content distribution is approximately normal in shape with a mean of 6 oz and a standard deviation of 0.05 oz. What can be said about the percentage of cartons that have actual contents less than 5.95 oz?
- The percentage is approximately 68%
 - The percentage is approximately 34%
 - The percentage is approximately 32%
 - The percentage is approximately 16%

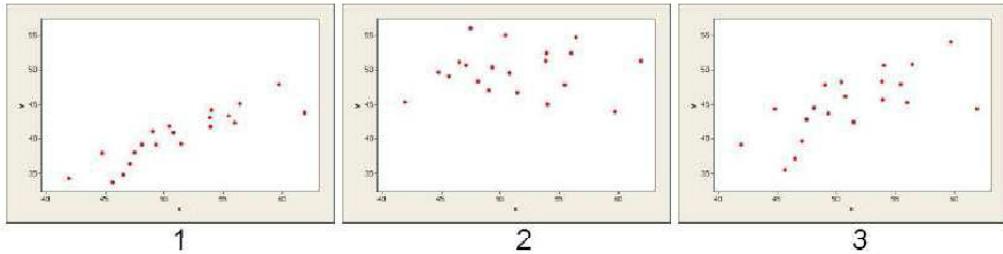
18. The University of Oklahoma has changed its admission standards to require an ACT-score of 26. We know the ACT is normally distributed with a mean of 21 and an SD of 5. If we sample 100 students who took the ACT at random, how many would be expected to qualify for admission to OU?
- (a) 5
 - (b) 16
 - (c) 34
 - (d) 84
 - (e) none of the above
19. A colleague has collected 1000 old VW vans for resale. The colleague, an old stats professor, will only sell a van to those who can answer the following question: The -2 SD sales price for one of these vans is set at \$550; and $+2$ SD sales price is set at \$1100. He tells you the distribution of sales prices is approximately normal. What is the expected number of vans for sale between \$550 and \$1100?
- (a) 500
 - (b) 680
 - (c) 750
 - (d) 888
 - (e) 950
20. The heights of women are normally distributed with a mean of 65 inches and an SD of 2.5 inches. The heights of men are also normal with a mean of 70 inches. What percent of women are taller than a man of average height?
- (a) 0.15%
 - (b) 2.5%
 - (c) 5%
 - (d) 16%
 - (e) insufficient information
21. Many psychological disorders (e.g. Depression, ADHD) are based on the application of the 2 SD rule assuming a normal distribution of reported symptoms. This means that anyone who reports a symptom count that is greater than the 2 SD point in a normal population can be considered to be “abnormal” or “disordered”.
Given this definition of “disorder”, what is expected prevalence rate of these disorders based on the 2 SD rule?

- (a) 0.15%
 - (b) 2.5%
 - (c) 5%
 - (d) 16%
 - (e) 95%
22. The ACT has a mean of 21 and an SD of 5. The SAT has a mean of 1000 and a SD of 200. Joe Bob Keith took the ACT and he needs a score of 1300 on the SAT to get into UNC-Chapel Hill and a score of 1400 on the SAT to get into Duke. UNC and Duke both told Joe Bob Keith that they will convert the ACT to the SAT using a z-score (or standard-score) transformation. Joe Bob Keith has decided to go to the school with the highest standards that will accept him. If he doesn't qualify for either Duke or UNC, then it's Faber College for Joe Bob Keith. As it turns out, Joe Bob Keith got a 30 on the ACT, but he cannot figure out what that means for his choice of college. Help Joe Bob Keith out. Where is he going to school?
- (a) UNC
 - (b) Duke
 - (c) Faber
23. Let Z be a standard normal random variable. Which of the following probabilities is the smallest?
- (a) $P(-2 < Z < -1)$
 - (b) $P(0 < Z < 2)$
 - (c) $P(Z < 1)$
 - (d) $P(Z > 2)$
24. Let Z be a standard normal random variable. Which of the following probabilities is the smallest?
- (a) $P(0 \leq Z \leq 2.07)$
 - (b) $P(-0.64 \leq Z \leq -0.11)$
 - (c) $P(Z > -1.06)$
 - (d) $P(Z < -0.88)$

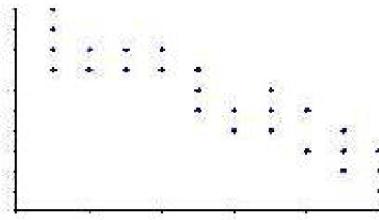
Chapter 2: Looking at Data - Relationships

Section 2.2: Correlation

25. The scatterplots below display three bivariate data sets. The correlation coefficients for these data sets are 0.03, 0.68, and 0.89. Which scatter plot corresponds to the data set with $r = 0.03$?



- (a) Plot 1
(b) Plot 2
(c) Plot 3
26. Joe Bob found a strong correlation in an empirical study showing that individuals' physical ability decreased significantly with age. Which numerical result below best describes this situation?
- (a) -1.2
(b) -1.0
(c) -0.8
(d) $+0.8$
(e) $+1.0$
(f) $+1.2$
27. Which correlation best describes the scatterplot?



- (a) -0.7
(b) -0.3
(c) 0
(d) $+0.3$
(e) $+0.7$
28. If you believed strongly in the idea that the more hours per week full-time students work in a job, the lower their GPA would be, then which correlation would you realistically expect to find?
(a) -0.97
(b) -0.72
(c) -0.20
(d) $+0.20$
(e) $+0.72$
(f) $+0.97$
29. A researcher found that $r = +.92$ between the high temperature of the day and the number of ice cream cones sold at Cone Island. This result tells us that
(a) high temperatures cause people to buy ice cream.
(b) buying ice cream causes the temperature to go up.
(c) some extraneous variable causes both high temperatures and high ice cream sales.
(d) temperature and ice cream sales have a strong positive linear relationship.

Section 2.3: Least-Squares Regression

30. A store manager conducted an experiment in which he systematically varied the width of a display for toothpaste from 3 ft. to 6 ft. and recorded the corresponding number of tubes of toothpaste sold per day. The data was used to fit a regression line, which was

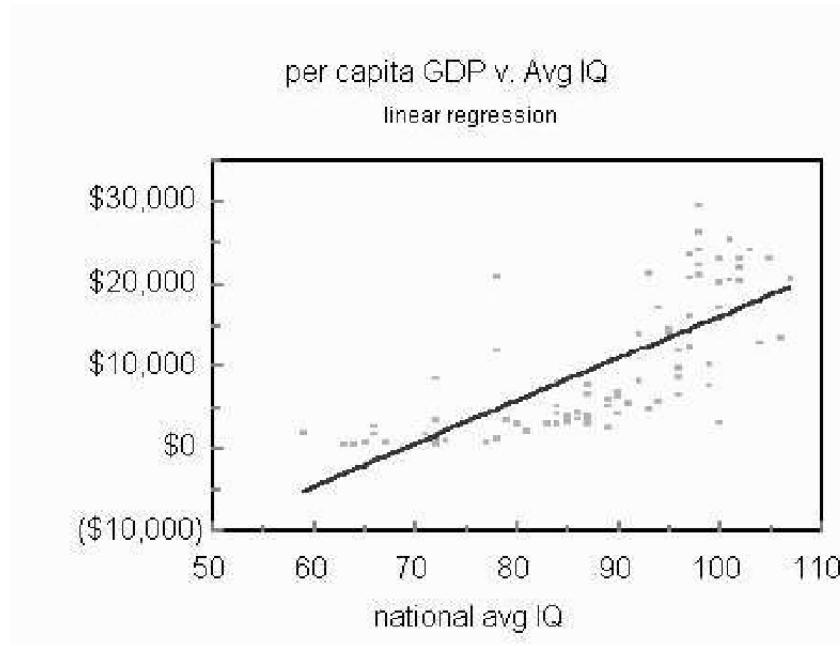
$$\text{tubes sold per day} = 20 + 10(\text{display width})$$

What is the predicted number of tubes sold per day for a display width of 12 feet?

- (a) 120
- (b) 140
- (c) It would be unwise to use the regression line to make a prediction for a display width of 12 ft.

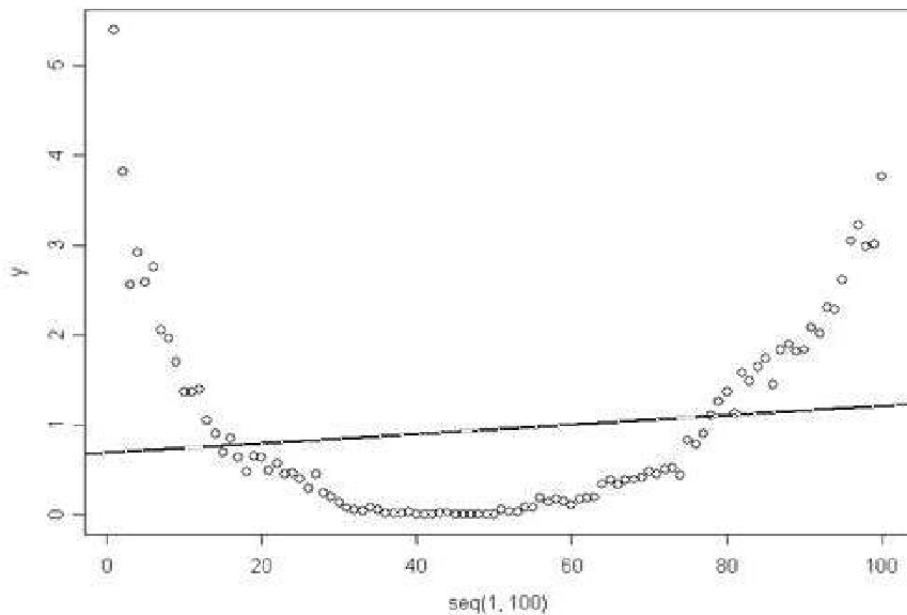
Section 2.4: Cautions about Correlation and Regression

31. Gas mileage and weight were recorded for each automobile in a sample of 20 compact cars. There was a strong negative correlation, with $r = -.87$. Based on the value of r , it is reasonable to conclude that increasing the weight of a compact car causes a decrease in gas mileage.
- (a) True
 - (b) False
32. Which of the following characteristics in a residual plot are indicative of potential problems?
- (a) A strong pattern in the residual plot
 - (b) Isolated points in the residual plot
 - (c) A lack of any strong pattern in the residual plot
 - (d) Both (a) and (b) above are indicative of potential problems
 - (e) (a), (b), and (c) above are all indicative of potential problems
33. Which phrase best describes the scatterplot?



- (a) strong $+r$
 - (b) strong $-r$
 - (c) weak $+r$
 - (d) weak $-r$
 - (e) influential outliers
 - (f) non-linearity
 - (g) Two from (A)-(F) are true.
 - (h) Three from (A)-(F) are true.
34. Why is it important to look for outliers in data prior to applying regression?
- (a) Outliers always affect the magnitude of the regression slope.
 - (b) Outliers are always bad data.
 - (c) Outliers should always be eliminated from the data set.
 - (d) Outliers should always be considered because of their potential influence.
 - (e) We shouldn't look for outliers, because all the data must be analyzed.
35. Which of the following factors is *NOT important* to consider when interpreting a correlation coefficient?
- (a) restriction of range

- (b) problems associated with aggregated data
 - (c) outliers
 - (d) lurking variables
 - (e) unit of measurement
36. What is the greatest concern about the regression below?



- (a) It has a small slope.
- (b) It has a high R^2 .
- (c) The investigator should not be using a linear regression on these data.
- (d) The residuals are too large.
- (e) The regression line does not pass through the origin.

Chapter 3: Producting Data

Section 3.1: Design of Experiments

37. "Graduating is good for your health," according to a headline in the Boston Globe (3 April 1998). The article noted "According to the Center for Disease Control, college

graduates feel better emotionally and physically than do high school dropouts.” Do you think the headline is justified based on this statement?

- (a) Yes, as long as the data was from random samples of college graduates and high school dropouts.
 - (b) Yes, because this must have been an observational study. As long as it was a well-designed study, the headline is justified.
 - (c) No, because the headline implies a cause and effect relationship, which is not justified based on an observational study.
 - (d) No, because this study must have been an experiment and we can't draw cause and effect conclusions from an experiment.
38. In a study of perceived importance of money, 100 attorneys were selected at random from those in private practice and 100 attorneys were selected at random from those employed by government agencies as district attorneys. The attorneys in each group were asked to respond to a set of questions designed to assess level of stress in the workplace. This study is _____.
- (a) an observational study
 - (b) an experiment
39. When is it unreasonable to reach a cause-and-effect conclusion based on data from a statistical study?
- (a) Any time the study is based on a random sample from a population of interest.
 - (b) When the study is observational.
 - (c) When the study is a well-designed experiment that uses random assignment to experimental conditions (treatments).
 - (d) It is always reasonable to reach a cause-and-effect conclusion based on data from a statistical study.

Section 3.2: Sampling Design

40. Researchers believe that one possible cause of Very Low Birth Weight (VLBW) infants is the presence of undiagnosed infections in the mother. To assess this possibility, they collected data on all pregnant women presenting themselves for prenatal care at large urban hospitals. What is the *appropriate population* for this study?
- (a) All infants.
 - (b) All infants born as VLBW infant.

- (c) All infants born in large urban centers.
 - (d) All pregnant women.
 - (e) All pregnant women living in large urban centers.
41. A Gallup survey was taken recently regarding peoples current preference for Democratic nominee for President for which there are 11 candidates. The survey also collected gender information, in order to capture male female differences in preference. For this poll, what is the *primary variable* of interest and *how many values* does it take?
- (a) gender; 2
 - (b) gender; more than 2
 - (c) candidate preference; 2
 - (d) candidate preference; more than 2
 - (e) political party; 2
 - (f) political party; more than 2
42. Increasing sample size
- (a) has no effect on bias.
 - (b) increases bias.
 - (c) decreases bias.
43. If you were trying to obtain a random sample of a population of interest for a political poll for a local mayoral race, which of the following approaches would be best to obtain the random sample?
- (a) Randomly assign a number to local companies and, using random-number generation, go to those companies selected and conduct interviews.
 - (b) Randomly select a busy street corner in your city and conduct on-site interviews.
 - (c) Assign a number to people in the local phone book and, using random-number generation, call those randomly selected.
 - (d) Randomly select a couple of television stations from your local cable company using random number generation and ask people through advertising to call a polling line.
 - (e) Randomly dial phone numbers within the selected area and interview those who answer the phone.

44. In order to estimate the proportion of students at a small liberal arts college who watch reality TV for more than 4 hours per week, a random sample of students at the school is selected and each is interviewed about his or her reality TV viewing habits. The students conducting the survey are worried that people that watch reality TV might be embarrassed to admit it and that they may not respond to the survey with honest answers. What type of bias are the students conducting the survey worried about?

- (a) They shouldn't worry - there is no obvious source of bias.
- (b) Voluntary bias
- (c) Nonresponse bias
- (d) Response bias

Section 4.2: Probability Models

45. Consider a standard 52-card deck, with four suits (hearts(red), diamonds(red), spades(black), clubs(black)), 13 cards per suit (2-10, J, Q, K, A). Define an event space on the standard deck such that it consists of 52 simple outcomes, one for each card in the deck.

Which of the following is a true statement?

- (a) Black is not an event.
- (b) Black is an event with 1 simple outcome.
- (c) Black is an event with 26 simple outcomes.
- (d) Black is an event with 52 simple outcomes.
- (e) None of the above is true.

Section 4.3: Random Variables

46. Draw the following dart board: A dart board is constructed from three concentric circles with radii 1 inch, 2 inches, and 3 inches, respectively. If a dart lands in the innermost circle, the player receives 4 points. If the dart lands between the innermost circle and the middle circle, the player receives 2 points. If the dart lands between the middle circle and the outermost circle, the player receives 1 point. Assume that the probability of a dart landing in any particular region is proportional to the area of that region.

Define the random variable X to be the sum of the player's score on two successive throws. Then X is what type of random variable?

- (a) discrete

- (b) continuous
47. Draw the following dart board: A dart board is constructed from three concentric circles with radii 1 inch, 2 inches, and 3 inches, respectively. If a dart lands in the innermost circle, the player receives 4 points. If the dart lands between the innermost circle and the middle circle, the player receives 2 points. If the dart lands between the middle circle and the outermost circle, the player receives 1 point. Assume that the probability of a dart landing in any particular region is proportional to the area of that region.
- Suppose that a player's score on a single dart throw is defined to be the distance between the dart and the center of the board. Define the random variable X to be the sum of the player's score on two successive throws. Then X is what type of random variable?
- (a) discrete
(b) continuous
48. A radioactive mass emits particles at an average rate of 15 particles per minute. Define the random variable X to be the number of particles emitted in a 10-minute time frame. Then X is what type of random variable?
- (a) discrete
(b) continuous
49. A radioactive mass emits particles at an average rate of 15 particles per minute. A particle is emitted at noon today. Define the random variable X to be the time elapsed between noon and the next emission. Then X is what type of random variable?
- (a) discrete
(b) continuous
50. Consider the continuous random variable X = the weight in pounds of a randomly selected newborn baby born in the United States during 2006. Let f be the probability density function for X . It is probably safe to say that $P(X < 0) = 0$ and $P(X < 20) = 1$. Which of the following is *not* a justifiable conclusion about f given this information?
- (a) No portion of the graph of f can lie below the x -axis.
(b) The area under the entire graph of f equals 1.
(c) The area under the graph of f between $x = 0$ and $x = 20$ is 1.
(d) The nonzero portion of the graph of f lies entirely between $x = 0$ and $x = 19$.

Section 4.4: Means and Variances of Random Variables

51. Suppose that a random variable X has only two values, 0 and 1. If $P(X = 0) = 0.5$ then what can we say about $E(X)$?
 - (a) $E(X) = 0$
 - (b) $E(X) = 0.5$
 - (c) $E(X) = 1$
 - (d) Either (A) or (C) is possible.
 - (e) Both (A) and (C).
 - (f) insufficient information
52. Suppose that a random variable X has only two values, 0 and 1. If $P(X = 0) = 0.5$ then what can we say about $\text{Var}(X)$?
 - (a) $\text{Var}(X) = -0.25$
 - (b) $\text{Var}(X) = 0$
 - (c) $\text{Var}(X) = 0.25$
 - (d) $\text{Var}(X) = 0.5$
 - (e) $\text{Var}(X) = 1$
 - (f) insufficient information
53. Suppose that a random variable X has only two values, 3 and 4. If $P(X = 3) = 0.5$ then what can we say about $E(X)$?
 - (a) $E(X) = 0.5$
 - (b) $E(X) = 1$
 - (c) $E(X) = 3$
 - (d) $E(X) = 3.5$
 - (e) $E(X) = 4$
54. Suppose that a random variable X has only two values, 3 and 4. If $P(X = 3) = 0.5$ then what can we say about $\text{Var}(X)$?
 - (a) $\text{Var}(X) = 0.25$
 - (b) $\text{Var}(X) = 0.5$

- (c) $\text{Var}(X) = 0.75$
 (d) $\text{Var}(X) = 1.0$
 (e) $\text{Var}(X) = 3.25$
 (f) $\text{Var}(X) = 3.5$
55. Suppose your instructor asks you a multiple-choice question with three answer choices in class. You are to submit your answer and also rate the confidence (low, medium, or high) with which you believe in that answer. You will be scored based on the following chart.
- | Confidence | Correct Answer | Incorrect Answer |
|------------|----------------|------------------|
| Low | 3 | 2 |
| Medium | 4 | 1 |
| High | 5 | 0 |
- If have no idea what the answer to the question is and you have to guess randomly among the three available answer choices, what confidence level should you choose in order to maximize your points?
- (a) Low
 (b) Medium
 (c) High
 (d) It doesn't matter.
56. Suppose your instructor asks you a multiple-choice question with *two* answer choices in class. You are to submit your answer and also rate the confidence (low, medium, or high) with which you believe in that answer. You will be scored based on the following chart.

Confidence	Correct Answer	Incorrect Answer
Low	3	2
Medium	4	1
High	5	0

- If have no idea what the answer to the question is and you have to guess randomly among the two available answer choices, what confidence level should you choose in order to maximize your points?
- (a) Low
 (b) Medium
 (c) High
 (d) It doesn't matter.

Chapter 5: Sampling Distributions

Section 5.1: Sampling Distributions for Counts and Proportions

57. Consider the following experiment. On a Friday night, a highway patrol officer sets up a roadblock and stops 100 drivers. A given driver is considered a success if he or she is wearing a seat belt; the driver is considered a failure otherwise. Can we consider this experiment a binomial experiment?
- (a) Yes
(b) No
58. Consider the following experiment. A particular car club has 100 members, 70 of which regularly wear their seat belts and 30 of which do not. Ten of these members are selected at random without replacement as they leave a car show. A given driver is considered a success if he or she is wearing a seat belt. The driver is considered a failure otherwise. Can we consider this experiment a binomial experiment?
- (a) Yes
(b) No
59. In 1938, Duke University researchers Pratt and Woodruff conducted an experiment looking for evidence of ESP (extrasensory perception). In the experiment, students were presented with five standard ESP symbols (square, wavy lines, circle, star, cross). The experimenter shuffled a deck of ESP cards, each of which had one of the five symbols on it. The experimenter drew a card from this deck, looked at it, and concentrated on the symbol on the card. The student would then guess the symbol, perhaps by reading the experimenter's mind. This experiment was repeated with 32 students for a total of 60,000 trials. The students were correct 12,489 times.
If the students were selecting one of the five symbols as random, the probability of success would be $p = 0.2$ and we would expect the students to be correct 12,000 times out of 60,000. Should we write off the observed excess of 489 as nothing more than random variation?
- (a) Yes
(b) No
60. For which of the following probabilities would the binomial distribution be appropriate?

- (a) The probability of a randomly selected professional basketball player making half of his free throws throughout a regular 82-game NBA season.
 - (b) The probability that a randomly selected student from a randomly selected high-school within the greater New York City metropolitan area will be accepted to an elite University.
 - (c) The probability that a randomly selected engineering student from a specific University will take at least 3 attempts to pass the licensure exam.
 - (d) Two of the above are appropriate for the binomial distribution.
 - (e) All of the above are appropriate for the binomial distribution.
 - (f) None of the above is appropriate for the binomial distribution.
61. Suppose a family is randomly selected from among all families with 3 children. What is the probability that the family has exactly one boy? You may assume that $\Pr(\text{boy}) = \Pr(\text{girl})$ for each birth.
- (a) $1/8$
 - (b) $1/6$
 - (c) $1/3$
 - (d) $3/8$
 - (e) $1/2$
 - (f) $5/6$
 - (g) $7/8$
62. Suppose a family is randomly selected from among all families with 4 children. What is the probability that the family has exactly two boys? You may assume that $\Pr(\text{boy}) = \Pr(\text{girl})$ for each birth.
- (a) $1/24$
 - (b) $1/16$
 - (c) $1/6$
 - (d) $3/8$
 - (e) $1/2$
63. To measure the success of the latest treatment for iPod-related deafness among young adults, researchers measured the sound sensitivity of 100 young adults by having them stand 20 feet away from a speaker playing “Slim Whitman Favorite Hits.” It was found that 35% of the sample could not repeat any song lyrics from the CD. What is the mean of this distribution?

- (a) (20)(.35)
 - (b) (20)(.65)
 - (c) (20)(.35)(.65)
 - (d) (.35)(.65)
 - (e) (100)(.35)
 - (f) (100)(.65)
 - (g) (100)(.35)(.65)
 - (h) insufficient information
64. To measure the success of the latest treatment for iPod-related deafness among young adults, researchers measured the sound sensitivity of 100 young adults by having them stand 20 feet away from a speaker playing "Slim Whitman Favorite Hits." It was found that 35% of the sample could not repeat any song lyrics from the CD. What is the variance of this distribution?
- (a) (20)(.35)
 - (b) (20)(.65)
 - (c) (20)(.35)(.65)
 - (d) (.35)(.65)
 - (e) (100)(.35)
 - (f) (100)(.65)
 - (g) (100)(.35)(.65)
 - (h) insufficient information

Chapter 6: Introduction to Inference

Section 6.1: Estimating with Confidence

65. The fundamental concept underlying statistical inference is that
- (a) through the use of sample data we are able to draw conclusions about a sample from which the data were drawn.
 - (b) through the examination of sample data we can derive appropriate conclusions about a population from which the data were drawn.
 - (c) when generalizing results to a sample we must make sure that the correct statistical procedure has been applied.

- (d) Two of the above are true.
(e) All of the above are true.
66. A 95% confidence interval is an interval calculated from
- (a) sample data that will capture the true population parameter for at least 95% of all samples randomly drawn from the same population.
 - (b) population data that will capture the true population parameter for at least 95% of all samples randomly drawn from the same population.
 - (c) sample data that will capture the true sample statistic for at least 95% of all samples randomly drawn from the same population.
 - (d) population data that will capture the true sample statistic for at least 95% of all samples randomly drawn from the same population.
67. A 95% confidence interval has been constructed around a sample mean of 28. The interval is (21, 35). Which of the following statement(s) is true?
- (a) The margin of error in the interval is 7.
 - (b) 95 out of 100 confidence intervals constructed around sample means will contain the true population mean.
 - (c) The interval (21,35) contains the true population mean.
 - (d) Both (a) and (b) are true.
 - (e) (a), (b), and (c) are true.
68. A 95% confidence interval for the mean of a population is given as (6.85, 7.61). Is it correct to say that there is a 95% chance that μ is between 6.85 and 7.61?
- (a) Yes
 - (b) No
69. Is it correct to say the following? If the process of selecting a sample of size 30 and then computing the corresponding 95% confidence interval is repeated 100 times, 95 of the resulting intervals will include μ .
- (a) Yes
 - (b) No
70. A 95% confidence intervals for birthweights is found to be (6.85, 7.61). Is it correct to say that 95% of all birth weights will be between 6.85 and 7.61 pounds?

- (a) Yes
 - (b) No
71. Suppose that a random sample of size 60 resulted in a 90% confidence interval for the proportion of students who carry more than 2 credit cards of (0.52, 0.76). Which of the following is a correct interpretation of the 90% confidence level?
- (a) 90% of the time the population proportion will be between 0.52 and 0.76
 - (b) The method used to construct the interval will produce an interval that includes the value of the population proportion about 90% of the time in repeated sampling.
 - (c) If 100 different random samples of size 60 from this population were each used to construct a confidence 90% confidence interval, 90 of them will contain the value population proportion.
 - (d) The probability that the population proportion is between 0.52 and 0.76 is 0.90.
72. Suppose that for babies born in the United States, birth weight is normally distributed about some unknown mean μ with standard deviation $\sigma = 1.06$ pounds. What is the minimum sample size necessary to ensure that the resulting 99% confidence interval has a width of at most 0.5?
- (a) 70
 - (b) 119
 - (c) 120
 - (d) 140

Section 6.2: Tests of Significance

73. Child and Protective Services, a branch of the Department of Health and Human Services is investigating the monthly average number of children in foster care over the last several years. They are interested in seeing if the average is dropping from 235 children per month in 2001.

The null hypothesis for this problem would be:

- (a) $H_0 : \mu < 235$
- (b) $H_0 = 235$
- (c) $H_0 : p = 235$
- (d) $H_0 : \mu = 235$

- (e) None of the above
74. Which of the following pairs gives a legitimate null and alternative hypothesis for carrying out a hypothesis test?
- (a) $H_0 : \pi = 0.4$ $H_a : \pi > 0.6$
 - (b) $H_0 : \mu < 80$ $H_a : \mu > 80$
 - (c) $H_0 : \mu = 80$ $H_a : \mu < 80$
 - (d) $H_0 : x = 25$ $H_a : x > 25$
75. A grocery store manager is interested in determining if the proportion of customers who pay by credit card at his store is greater than the reported national figure of 0.10. What hypotheses should the store manager test?
- (a) $H_0 : \mu > 0.1$ $H_a : \mu = 0.1$
 - (b) $H_0 : \mu = 0.1$ $H_a : \mu > 0.1$
 - (c) $H_0 : \pi = 0.1$ $H_a : \pi > 0.1$
 - (d) $H_0 : \pi > 0.1$ $H_a : \pi = 0.1$
 - (e) $H_0 : \pi > 0.1$ $H_a : \pi < 0.1$
76. A drug company believes their newest drug for controlling cardiac arrhythmias is more effective and has less side effects than the current drug being used in the market. They submit their new drug to the FDA for a clinical trial to assess the efficacy of their drug in comparison to the current drug. What is the most appropriate null hypothesis for this clinical trial?
- (a) H_0 : Efficacy of new drug = Efficacy of old drug
 - (b) H_0 : Efficacy of new drug > Efficacy of old drug
 - (c) H_0 : Efficacy of new drug < Efficacy of old drug
 - (d) H_0 : Efficacy of new drug \neq Efficacy of old drug
 - (e) None of the above
77. If the conclusion in a hypothesis test is to fail to reject H_0 , we can conclude that there is strong evidence that the null hypothesis is true.
- (a) True
 - (b) False

Section 6.3: Use and Abuse of Tests

78. Robert is asked to conduct a clinical trial on the comparative efficacy of Aleve versus Tylenol for relieving the pain associated with muscle strains. He creates a carefully controlled study and collects the relevant data. To be most informative in his presentation of the results, Robert should report
- (a) whether a statistically significant difference was found between the two drug effects.
 - (b) a P -value for the test of no drug effect.
 - (c) the mean difference and the variability associated with each drug's effect.
 - (d) a confidence interval constructed around the observed difference between the two drugs.
79. A P -value represents
- (a) the probability, given the null hypothesis is true, that the results could have been obtained purely on the basis of chance alone.
 - (b) the probability, given the alternative hypothesis is true, that the results could have been obtained purely on the basis of chance alone.
 - (c) the probability that the results could have been obtained purely on the basis of chance alone.
 - (d) Two of the above are proper representations of a P -value.
 - (e) None of the above is a proper representation of a P -value.
80. Two studies investigating the effect of motivation upon job performance found different results. With the exception of the sample size the studies were identical. The first study used a sample size of 500 and found statistically significant results, whereas the second study used a sample size of 100 and could not reject the null hypothesis. Which of the following is true?
- (a) The first study showed a larger effect than the second.
 - (b) The first study was less biased than the second study for estimating the effect size because of the larger sample size.
 - (c) The first study results are less likely to be due to chance than the second study results.
 - (d) Two of the above are true.
 - (e) All of the above are true.

Section 6.4: Power and Inference as a Decision

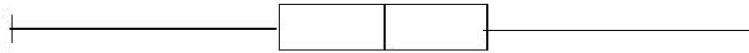
81. The manager of a university computing help line is trying to decide whether to hire additional staff. She has decided to hire if there is evidence that the average time callers to the help line must wait on hold before receiving assistance is greater than 5 minutes. She decides to collect data in order to test $H_0 : \mu = 5$ versus $H_a : \mu > 5$ where μ is the mean time on hold. From the callers perspective, which type of error would be more serious?
 - (a) Type I error
 - (b) Type II error
 - (c) Both types of error would be considered equally serious
82. Suppose that the P -value in a hypothesis test is 0.08. If the significance level for the test is $\alpha = 0.05$, which of the following is the appropriate decision?
 - (a) Fail to reject H_0
 - (b) Reject H_0
 - (c) There is not enough information given to know whether or not H_0 should be rejected.
83. In order to investigate a claim that the average time required for the county fire department to respond to a reported fire is greater than 15 minutes, county staff determined the response times for 40 randomly selected fire reports. The data was used to test $H_0 : \mu = 15$ versus $H_a : \mu > 15$ and the computed P -value was 0.12. If a 0.05 level of significance is used, what conclusions can be drawn?
 - (a) There is convincing evidence that the mean response time is 15 minutes (or less).
 - (b) There is convincing evidence that the mean response time is greater than 15 minutes.
 - (c) There is not convincing evidence that the mean response time is greater than 15 minutes.
84. Carol reports a statistically significant result ($P < 0.02$) in one of her journal articles. The editor suggests that because of the small sample size of the study ($n = 20$), the result cannot be trusted and she needs to collect more data before the article can be published. He is concerned that the study has too little power. How would you respond to the editor?
 - (a) The study has enough power to detect the effect since the significant result was obtained.

- (b) Because the sample size so small, increasing the sample size to 200 should ensure sufficient power to detect a small effect.
- (c) Setting the $\alpha = 0.01$ would be an alternative to collecting more data.
- (d) Because the P -value is so close to $\alpha = 0.05$, the effect size is likely to be small and hence more information is needed.

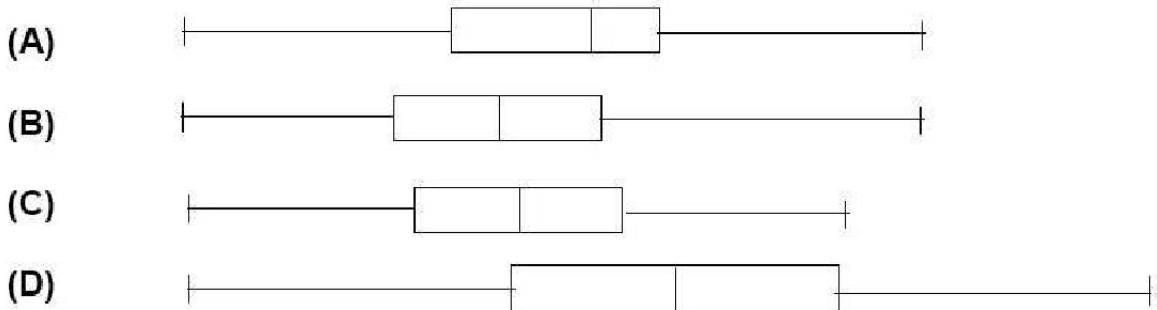
Chapter 7: Inference for Distributions

Section 7.1: Inference for the Mean of a Population

- 85. If you are testing two groups of individuals to see if they differ in regards to their working memory capacity, your alternative hypothesis would be that the two groups
 - (a) differ significantly in terms of working memory capacity.
 - (b) differ in terms of working memory capacity.
 - (c) differ, but not significantly, in terms of working memory capacity.
 - (d) do not differ in terms of working memory capacity.
 - (e) do not differ significantly in terms of working memory capacity.
- 86. This box plot is for a sample that accurately represents a normal distribution:



Which of the following box plots is for a sample that represents a Students t -distribution with the same standard deviation and sample size as the normal distribution above?



(E) Two from (A)-(D) are correct.

(F) Three from (A)-(D) are correct.

(G) All from (A)-(D) are correct.

87. A random sample of 25 observations, with a mean of 44.4 and a sample standard deviation of 3.5, is drawn from a population that is approximately normally distributed. If one sets up a hypothesis test with population mean equal to 43 against an alternative that the population mean is not 43, using $\alpha = 0.01$, what is the value of the test statistic?

- (a) 2.000
- (b) 2.576
- (c) 2.797
- (d) 2.857
- (e) 10.000

88. A climate researcher sets up an experiment that the mean global temperature is $\mu = 60^\circ$ F, looking for an indication of global warming in a climate model projection. For the year 2050, the series of 10 models predict an average temperature of 65° F. A standard one-tailed t -test is run on the data. Then the power of the test

- (a) increases as μ decreases.
- (b) remains constant as μ changes.
- (c) increases as μ increases.
- (d) decreases as μ increases.

Chapter 8: Inference for Proportions

Section 8.1: Inference for a Single Proportion

89. To estimate the proportion of students at a university who watch reality TV shows, a random sample of 50 students was selected and resulted in a sample proportion of .3. A 95% confidence interval for the proportion that watches reality TV would be _____ a 90% confidence interval.
- (a) narrower than
 - (b) the same width as
 - (c) wider than
90. Suppose we wish to estimate the percentage of students who smoke marijuana at each of several liberal arts colleges. Two such colleges are StonyCreek (enrollment 5,000) and Whimsy (enrollment 13,000). The Dean of each college decides to take a random sample of 10% of the entire student population. The margin of error for a simple random sample of 10% of the population of students at each school will be
- (a) smaller for Whimsy than for StonyCreek.
 - (b) smaller for StonyCreek than for Whimsy.
 - (c) the same for each school.
 - (d) insufficient information
91. Suppose we wish to estimate the percentage of people who speed while driving in a college town. We choose to sample the populations of Austin, TX (University of Texas) and Norman, OK (University of Oklahoma). We know that both cities have populations over 100,000 and that Austin is approximately 5 times bigger (in population) than Norman. We also expect the rates of speeding to be about the same in each city. Suppose we were to take a random sample of 1000 drivers from each city. The margin of error for a simple random sample of the population of drivers from each city will be
- (a) smaller for the Austin sample than the Norman sample.
 - (b) smaller for the Norman sample than the Austin sample.
 - (c) the same for both samples.
 - (d) not possible to determine without more precise information about the population sizes.

92. A parachute manufacturer is concerned that the failure rate of 0.1% advertised by his company may in fact be higher. What is the null hypothesis for the test he would run to address his worries.
- (a) $H_0 : \mu = 0.001$
 - (b) $H_0 : p > 0.001$
 - (c) $H_0 : \mu < 0.001$
 - (d) $H_0 : p = 0.001$
93. A parachute manufacturer is concerned that the failure rate of 0.1% advertised by his company may in fact be higher. A hypothesis test was run and the result was a P -value of 0.03333. The most likely conclusion the manufacturer might make is:
- (a) My parachutes are safer than I claim.
 - (b) My parachutes are not as safe as I claim them to be.
 - (c) I can make no assumption of safety based on a statistical test.
 - (d) The probability of a parachute failure is 0.03333.
 - (e) Both (b) and (d) are true.
94. To explain the meaning of a P -value of 0.033, you could say:
- (a) There is approximately a 96.7% chance of obtaining my sample results.
 - (b) Assuming the null hypothesis is accurate, my sample results should occur only 3.3% of the time.
 - (c) Cant say anything for sure without knowing the sample results.
 - (d) There is approximately a 3.3% chance of obtaining my sample results.
95. Suppose we have the results of a Gallup survey (simple random sampling) which asks participants for their opinions regarding their attitudes toward technology. Based on 1500 interviews, the Gallup report makes confidence statements about its conclusions. If 64% of those interviewed favored modern technology, we can be 95% confident that the percent of those who favored modern technology is
- (a) 95% of 64%, or 60.8%
 - (b) $95\% \pm 3\%$
 - (c) 64%
 - (d) $64\% \pm 3\%$

96. A confidence interval for a proportion is constructed using a sample proportion of 0.5. If the sample proportion was 0.9 instead of 0.5, what would happen to the width of the resulting confidence interval?
- (a) The new CI would be narrower.
 - (b) The new CI would have the same width.
 - (c) The new CI would be wider.
97. A sample needs to be taken to answer the question Have you ever shoplifted? Assuming a random sample can be found, how many people would need to be polled to insure a margin of error of no more than 3% with 90% confidence?
- (a) 1068
 - (b) 752
 - (c) 23
 - (d) None of the above
98. The margin of error is computed for a poll with a sample size of 50. Approximately what sample size would you need if you wanted to cut the margin of error in half?
- (a) 25
 - (b) 100
 - (c) 200
 - (d) 400
99. Which of the following does *not* result in a larger margin of error?
- (a) Increasing the confidence level
 - (b) Decreasing the sample size
 - (c) Having a larger population size

Section 8.2: Comparing Two Proportions

100. A two proportion z interval was constructed for the difference in the two population proportions, p_1 and p_2 . The resulting 99% confidence interval was $(-0.004, 0.12)$. A conclusion that could be drawn is:
- (a) There is no significant difference between p_1 and p_2 .

- (b) There is a significant difference between p_1 and p_2 .
 - (c) The range of possible differences between the two proportions could be from a 0.4% difference with p_2 being larger up to a 12% difference with p_1 being larger.
 - (d) Both (a) and (c) are correct.
 - (e) Both (b) and (c) are correct.
101. Two methods are used to predict the shear strength for steel plate girders. Each method is applied to nine specific girders and the ratio of predicted load to observed load is calculated for each method and each girder. What kind of t-test should we use to compare these data?
- (a) Independent *t*-test
 - (b) Paired *t*-test
102. Two catalysts are being analyzed to determine how they affect the mean yield of a chemical process. Catalyst 1 is used in the process eight times and the yield in percent is measured each time. Then catalyst 2 is used in the process eight times and the yield is measured each time. What kind of t-test should be used to compare these data?
- (a) Independent *t*-test
 - (b) Paired *t*-test
103. Six river locations are selected and the zinc concentration is determined for both surface water and bottom water at each location. What kind of t-test should be used to compare these data?
- (a) Independent *t*-test
 - (b) Paired *t*-test

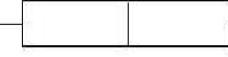
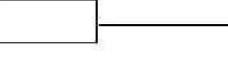
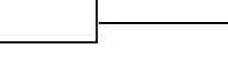
Chapter 9: Analysis of Two-Way Tables

Section 9.1: Inference for Two-Way Tables

104. In a 2×2 table of the frequency of sexual intercourse by age, we observe a chi-square (χ^2) statistic of 2.5. What should be the conclusion?
- (a) There is observed evidence that sex and age are associated.
 - (b) There is little observed evidence of anything but a chance association.
 - (c) It is not possible to obtain an observed chi-square statistic this large.
 - (d) It would be unlikely to obtain an observed chi-square statistic this large.
 - (e) No conclusion is appropriate without sample size information.

Section 9.3: Goodness of Fit

105. If a large sample were drawn from a chi-square (χ^2) distribution (with degrees of freedom ≤ 10) and accurately represented the population, which of the following is most likely to be a box plot of that sample?

- (A)  A box plot with a median at the first quartile (Q1), whiskers extending from the minimum to the maximum.
- (B)  A box plot with a median at the third quartile (Q3), whiskers extending from the minimum to the maximum.
- (C)  A box plot with a median at the first quartile (Q1), whiskers extending from the minimum to the maximum.
- (D)  A box plot with a median at the third quartile (Q3), whiskers extending from the minimum to the maximum.
- (E) Two from (A)-(D) are correct.
- (F) Three from (A)-(D) are correct.
- (G) All from (A)-(D) are correct.

Chapter 10: Inference for Regression

Secton 10.1: Simple Linear Regression

106. What is the most common rationale for significance testing of simple linear regression?
- to test if the intercept is significantly large
 - to test is the slope of the regression line is positive
 - to test if the slope of the regression line is negative
 - to test if the slope is different from zero
 - to appease an editor or reviewer when publishing the results
107. Which of the following does *not* result in more accurate estimates for β_1 ?
- An increase in the sample size

- (b) An increase in the coefficient of determination
- (c) An increase in the variance of the observed x -values
- (d) An increase in the variance of the observed y -values