

- The heights of American men aged 18 to 24 are approximately normally distributed with mean 68 inches and standard deviation 2.5 inches. Half of all young men are shorter than
 - 65.5 inches
 - 68 inches
 - 70.5 inches
 - can't tell, median height is not given
- The heights of American men aged 18 to 24 are approximately normally distributed with mean 68 inches and standard deviation 2.5 inches. Only about 5% of young men have heights outside the range
 - 65.5 inches to 70.5 inches
 - 63 inches to 73 inches
 - 60.5 inches to 75.5 inches
 - 58 inches to 78 inches
- The scores on a statistics exam are strongly skewed to the left. So it is best to describe the distribution by reporting
 - the five-number summary
 - the mean and standard deviation
 - the mean, median, and mode
 - the correlation and its square
- The grade point averages (GPA) of 7 randomly chosen students from your statistics class are
3.14 2.37 2.94 3.60 1.70 4.00 1.85
The median GPA for these students is
 - 2.8
 - 2.37
 - 3.6
 - 2.94
 - none of the above
- In a linear regression analysis, which of the following may be a negative number?
 - s_e
 - correlation
 - residual
 - both B and C
- A standardized test designed to measure math anxiety has a mean of 100 and a standard deviation of 10 in the population of first year college students. Which of the following observations would you suspect is an outlier?
 - 150
 - 100
 - 90
 - all of the above
 - none of the above
- Scores on a standard IQ test are normally distributed with mean 100 and standard deviation 15. What percentage of the population have IQ's less than 90?
 - 15%
 - 74.86%
 - 66.67%
 - 25.14%
- A study found correlation $r = 0.61$ between the sex of a worker and his or her income. You conclude that
 - women earn more than men on the average
 - women earn less than men on the average
 - men typically earn 71 cents for every \$1 that women earn for doing the same job.
 - an arithmetic mistake was made; this is not a possible value of r
 - this is nonsense because r makes no sense here

9. A copy machine dealer has data on the number x of copy machines at each of 89 customer locations and the number y of service calls in a month at each location. Summary calculations give $\bar{x} = 8.4$, $S_x = 2.1$, $\bar{y} = 14.2$, $S_y = 3.8$, and $r = .86$. What is the slope of the least squares regression line of the number of service calls on number of copiers?
- A) 0.86 B) 1.56 C) 0.48 D) none of these
10. A copy machine dealer has data on the number x of copy machines at each of 89 customer locations and the number y of service calls in a month at each location. Summary calculations give $\bar{x} = 8.4$, $S_x = 2.1$, $\bar{y} = 14.2$, $S_y = 3.8$, and $r = .86$. About what percent of the variation in number of service calls is explained by the linear relation between number of service calls and number of machines?
- A) 86% B) 93% C) 74% D) none of these
11. In a statistics course a linear regression equation was computed to predict the final exam score from the score on the first test. The equation was $\hat{y} = 10 + 0.9x$ where y is the final exam score and x is the score on the first exam. Eve scored 95 on the first test and 98 on the final exam. What is the predicted value of her score on the final exam?
- A) 95 B) 85.5 C) 90 D) 95.5 E) none of the above
12. In a statistics course a linear regression equation was computed to predict the final exam score from the score on the first test. The equation was $\hat{y} = 10 + 0.9x$ where y is the final exam score and x is the score on the first exam. Eve scored 95 on the first test and 98 on the final exam. What is the value of her residual?
- A) 98 B) 2.5 C) -2.5 D) 0 E) none of the above
13. You notice that your car seems to run better when you use Brand A of gasoline than when you use Brand B. Can you conclude that Brand A is better than Brand B for your car?
- A) Yes. This is a simple random sample D) No. The two brands are the same
 B) No. The evidence is anecdotal E) None of the above
 C) Yes. This is a comparative experiment
14. A new headache remedy is given to a group of 25 patients who suffer severe headaches. Of these, 20 report that the remedy is very helpful in treating their headaches. From this information you conclude
- A) the remedy is effective for the treatment of headaches
 B) nothing, because the sample size is too small
 C) nothing, because there is not a control group for comparison
 D) the new treatment is better than aspirin
 E) none of the above
15. A student organization wants to assess the attitudes of students toward a proposed change in the hours that the library is open. They randomly select 50 freshmen, 50 sophomores, 50 juniors, and 50 seniors. The situation described is
- A) a stratified sample D) anecdotal evidence
 B) a simple random sample E) a block sample
 C) a comparative experiment

16. In a study of the effects of acid rain, a random sample of 100 trees from a particular forest is examined. Forty percent of these show some signs of damage. Which of the following statements are correct?
- A) 40% is a parameter
 - B) 40% is a statistic
 - C) 40% of all trees in this forest show signs of damage
 - D) more than 40% of the trees in this forest show some signs of damage
 - E) less than 40% of the trees in this forest show some signs of damage
17. In a study of acid rain effects, a random sample of 100 trees from a forest is examined. Forty percent of these show some signs of damage. Which of the following statements are correct?
- A) the sampling distribution of the proportion of damaged trees is approximately normal
 - B) if we took another random sample of trees, we would find that 40% of these would show some signs of damage
 - C) if a sample of 1000 trees was examined, the variability of the sample proportion would be larger than for a sample of 100 trees
 - D) this is a comparative experiment
 - E) none of the above
18. A simple linear regression was used to predict the score y on a final exam from the score x on the first exam. The slope of the least squares regression line is .75. The standard error of the slope is .11 and then sample size is 200. To test the null hypotheses that the slope is zero versus the one-sided alternative that the slope is positive, we use the statistic $t =$
- A) 6.82 B) 0.05 C) 0.15 D) 0.95 E) 0.75
19. University X conducts a sample survey of families of undergraduate students to obtain their opinions about health insurance for students. A questionnaire is sent to a simple random sample of 100 families. One question asks for a response to the statement "University X should provide full medical insurance for students even if tuition must be raised to pay for it." The possible responses are (1) Strongly agree, (2) Agree, (3) Neutral, (4) Disagree, and (5) Strongly disagree. The population in this sample survey is
- A) all University X students
 - B) all families of University X undergraduate students
 - C) the 100 families sent the questionnaire
 - D) the five possible answers
20. A randomly selected student is asked to respond yes, no, or maybe to the question: "Do you intend to vote in the next presidential election?" The sample space is {yes, no, maybe}. Which of the following represent a legitimate assignment of probabilities for this sample space?
- A) 0.4, 0.4, 0.2 B) 0.4, 0.6, 0.4 C) 0.3, 0.3, 0.3 D) 0.5, 0.3, -0.2 E) none of these

21. In a population of students, the number of calculators owned is a random variable X with $P(X=0) = 0.2$, $P(X=1) = 0.6$, $P(X=2) = 0.2$. The mean of this probability distribution is
- A) 0 B) 2 C) 1 D) 0.5
E) the answer cannot be computed from the information given
22. In a population of students, the number of calculators owned is a random variable X with $P(X=0) = .2$, $P(X=1) = .6$, $P(X=2) = .2$. The variance for this probability distribution is
- A) 1 B) 0.63 C) 0.5 D) 0.4
E) the answer cannot be computed from the information given
23. You play tennis regularly with a friend and from past experience you believe that the outcome of each match is independent. For any given match you have a probability of .6 of winning. The probability that you win the next two matches is
- A) 0.36 B) 0.6 C) 0.4 D) 0.16 E) 1.2
24. The number of calories in a one-ounce serving of a certain breakfast cereal is a random variable with mean 110. The number of calories in a full cup of whole milk is a random variable with mean 140. For breakfast you eat one ounce of the cereal with 1/2 cup of whole milk. Let Z be the random variable that represents the total number of calories in this breakfast. The mean of Z is
- A) 110 B) 140 C) 180 D) 250 E) 195
25. An athlete suspected of having used steroids is given two tests that operate independently of each other. Test A has probability 0.9 of being positive if steroids have been used. Test B has probability 0.8 of being positive if steroids have been used. What is the probability that *neither* test is positive if steroids have been used?
- A) 0.02 B) 0.72 C) 0.30 D) 0.28 E) none of the above
26. An instant lottery game gives you probability 0.02 of winning on any one play. Plays are independent of each other. If you play 5 times, what is the probability that you win at least once?
- A) 0.0961 B) 0.0922 C) 0.9039 D) 0.1 E) none of the above
27. A sample survey chooses a sample of households and measures their annual income and their savings. Some events of interest are:
- A = the household chosen has income of at least \$50,000
C = the household chosen has at least \$10,000 in savings
- The complement A^c of the event A is the event that a household...
- A) has income of at least \$50,000
B) has income less than \$50,000
C) has income at least \$50,000 and savings at least \$10,000
D) is not selected by the sample survey

28. A sample survey chooses a sample of households and measures their annual income and their savings. Some events of interest are:
- A = the household chosen has income at least \$50,000
C = the household chosen has at least \$10,000 in savings
- Based on the sample survey, we estimate that $P(A) = 0.2$ and $P(C) = 0.3$. The probability that a household has income at least \$50,000 *or* savings at least \$10,000
- A) is 0.5
B) is 0.06
C) can't be found from this information because A and C are probably not disjoint
D) can't be found from this information because A and C are probably not independent
29. In a large population of college students, 20% of the students have experienced feelings of math anxiety. If you take a random sample of 10 students from this population, the probability that exactly 2 students have experienced math anxiety is
- A) 0.3020 B) 0.2634 C) 0.2013 D) 0.5 E) 1.0
30. In a large population of college students, 20% of the students have experienced feelings of math anxiety. Suppose you take a random sample of 10 students from this population. The standard deviation of the sample proportion of students who have experienced math anxiety is
- A) 0.0160 B) 0.1265 C) 0.2530 D) 1.0 E) 0.2070
31. If a population has a standard deviation σ , then the standard deviation of the mean of 100 randomly selected items from this population is
- A) σ B) 100σ C) $\sigma/10$ D) $\sigma/100$ E) 0.1
32. The Central Limit Theorem states that
- A) the sample mean is unbiased
B) the sample mean is approximately normal
C) the binomial distribution is skewed
D) the sample standard deviation is approximately normal
E) none of the above
33. You want to compute a 95% confidence interval for a population mean. Assume that the population standard deviation is known to be 10 and the sample size is 51. The critical value to be used in this calculation is
- A) 1.645 B) 2.009 C) 1.960 D) 1.676 E) 0.8352
34. You want to estimate the mean SAT score for a population of students with a 90% confidence interval. Assume that the population standard deviation is $\sigma = 100$. If you want the margin of error to be approximately 10, you will need a sample size of
- A) 16 B) 271 C) 38 D) 1476 E) none of the above

35. You have measured the systolic blood pressure of a random sample of 25 employees of a company located near you. A 95% confidence interval for the mean systolic blood pressure for the employees of this company is (122, 138). Which of the following statements gives a valid interpretation of this interval?
- A) 95% of the sample of employees has a systolic blood pressure between 122 and 138
 - B) 95% of the population of employees has a systolic blood pressure between 122 and 138
 - C) if the procedure were repeated many times, 95% of the resulting confidence intervals would contain the population mean systolic blood pressure
 - D) the probability that the population mean blood pressure is between 122 and 138 is .95
 - E) if the procedure were repeated many times, 95% of the sample means would be between 122 and 138
36. A significance test was performed to test the null hypothesis $H_0: \mu = 2$ versus the alternative $H_a: \mu \neq 2$. The data for the test was collected from a random sample of 11 subjects. The test statistic is $t = 1.40$. The P-value for this test is approximately
- A) 0.16
 - B) 0.08
 - C) 0.1918
 - D) 0.0959
 - E) 0.1891
37. A significance test gives a P-value of .04. From this we can...
- A) reject H_0 with $\alpha = 0.01$.
 - B) reject H_0 with $\alpha = 0.05$.
 - C) say that the probability that H_0 is false is 0.04.
 - D) say that the probability that H_0 is true is .04.
 - E) none of the above
38. Based on a random sample of 3rd grade students, a 95% confidence interval for the mean reading achievement score is (44.2, 54.2). The margin of error for this interval is
- A) 0.95
 - B) 5
 - C) 10
 - D) 54.2
 - E) 0.475
39. Based on a random sample of 3rd grade students, a 95% confidence interval for the mean reading achievement score is (44.2, 54.2). The point estimate from which this interval was calculated is
- A) 44.2
 - B) 54.2
 - C) 0.95
 - D) 49.2
 - E) 1.96
40. Based on a random sample of 3rd grade students, a 95% confidence interval for the mean reading achievement score for a population of third grade students is (44.2, 54.2). Suppose you compute a 99% confidence interval. Which of the following statements is correct?
- A) the two intervals have the same width
 - B) the 99% interval is wider
 - C) the 95% interval is wider
 - D) you cannot determine which interval is wider unless you know n and s
41. You are reading an article in your field that reports several statistical analyses. The article says that the P-value for a significance test is 0.045. Is this result significant at the 1% significance level?
- A) yes
 - B) no

Free Response Questions – Please work these problems out on separate paper.

51. R. B. D’Agnostino (1973) wrote a delightful article about the weight of a 40-pound box of bananas. The problem facing the banana shipper is the rule that a 40-pound box must at least 40 pounds upon arrival. Suppose a shipper knows from past experience that, when boxes are packed to have 40 pounds, the standard deviation is four ounces and that each box loses eight ounces in transit.
- a) What mean weight should the shipper establish so that only one-fourth of 1% of the boxes will arrive with less than 40 pounds?
- b) Suppose the shipper adopts the weight you established and ships out 5 million boxes. How many will arrive with less than 40 pounds of bananas?
52. A researcher asked 173 people with different educational accomplishments to respond to the statement “Our planet is being observed by intelligent life forms from outer space.” Responses were given on a 7-point scale from “Strong agreement” to “Strong disagreement.” The number of people responding in each educational category follows. Do the data in the table below provide statistical evidence of a relationship between level of education and agreement with the statement?

Level of Agreement	College Educated	High School Dropouts
strong agreement	2	4
mild agreement	7	10
slight agreement	15	24
neutral	19	21
slight disagreement	23	18
mild disagreement	12	9
strong disagreement	6	3

53. One school of thought about the nature of creativity is that it is related to humor. The data that follows is for ten college sophomores. For each score there is a score on a test of creativity and the number of puns produced in eight minutes while looking at a list of “wise sayings.” Draw a scatterplot, find the correlation coefficient, and the coefficient of determination. Write the regression equation and plot the line on your scatterplot. Write an explanation of what your analysis shows (5 points)

Subject	1	2	3	4	5	6	7	8	9	10
Creativity Test Score	60	57	52	46	41	38	32	29	25	19
Number of Puns	28	32	24	16	21	14	18	11	9	12

54. As a result of his research before 1900, E. L. Thorndike concluded that animals were incapable of learning by imitation. In 1901, however, L. L. Hobhouse reported the results of experiments with cats, dogs, otters, elephants, monkeys, and chimpanzees indicating that these animals could learn by imitation. Suppose the following study was carried out. One group of hungry cats is shown food being obtained from under a vase. Another group is not. Shortly afterward, time (in seconds) required to upset the vase and find the food is recorded for each animal. Results are given below. Do these results provide statistical evidence in support of either of the two theorists? Explain.

Shown	18	15	15	12	11	9
Not Shown	25	22	21	19	16	15

55. “The self confidence of that group of recruits is negatively correlated with their success in the obstacle course.” Tell what this statement means.
56. Suppose you wanted to know whether the weight of vegetarians was less than that of the general population in the United States. Suppose also that you were fortunate enough to have the weights of a representative sample of 49 male vegetarians who were college age. Now, it is a fact that the average weight of 18 to 24 year old male Americans (5’ 10” tall) is 166 pounds (Statistical Abstract of the United States: 1987). It is also a fact that weight is not normally distributed but is positively skewed. Can you use the techniques we’ve learned in class to determine the probability that the mean weight of college-age male vegetarians came from a population with mean weight of 166? Write your conclusion and your reasoning. (Don’t perform the test. HINT: Check out the Central Limit Theorem.)