

Name \_\_\_\_\_

1. 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

- A) 65.5 inches    B) 68 inches    C) 70.5 inches    D) can't tell, because the median height is not given

2. 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

- A) 65.5 inches to 70.5 inches    C) 60.5 inches to 75.5 inches  
B) 63 inches to 73 inches    D) 58 inches to 78 inches

3. The scores on a statistics exam are strongly skewed to the left. So it is best to describe the distribution by reporting

- A) the five-number summary    C) the mean, median, and mode  
B) the mean and standard deviation    D) the correlation and its square

4. 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 mean GPA for these students is

- A) 2.8    B) 2.94    C) 3.6    D) none of the above

5. 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

If  $\sum (x - \bar{x})^2 = 4.51$ , then the standard deviation is

- A) .75    B) .87    C) .64    D) .80    E) none of the above

6. 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?

- A) 150    B) 100    C) 90    D) all of the above    E) none of the above

7. 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?

- A) 15%    B) 74.86%    C) 66.67%    D) 25.14%

8. A study found correlation  $r = .61$  between the sex of a worker and his or her income. You conclude that

- A) women earn more than men on the average
- B) women earn less than men on the average
- C) an arithmetic mistake was made; this is not a possible value of  $r$
- D) 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      E) can't tell from the information given

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      E) can't tell from the information given

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. 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 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 freshman, 50 sophomores, 50 juniors, and 50 seniors. The situation described is

- A) a block sample
- B) a simple random sample
- C) a comparative experiment
- D) anecdotal evidence
- E) none of the above

16. In a study of the effects of acid rain, a random sample of 100 trees from a particular forest are 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 the effects of acid rain, a random sample of 100 trees from a particular forest are 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. You must choose a simple random sample of five from a lot of 64 vials of a pharmaceutical product for assay. You label the vials beginning with 00 and use the table of random digits beginning with line 136 to choose your sample. Your sample contains the vials labeled

- A) 84 21 44 75 37
- B) 08 44 77 28 75
- C) 08 42 14 47 53
- D) 84 53 66 55 12
- E) none of these

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) .4, .4, .2    B) .4, .6, .4    C) .3, .3, .3    D) .5, .3, -.2    E) none of the above

21. In a population of students, the number of calculators owned is a random variable  $X$  with  $P(X=0) = .2$ ,  $P(X=1) = .6$ , and  $P(X=2) = .2$ . The mean of this probability distribution is

A) 0    B) 2    C) 1    D) .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  $P(X=0) = .2$ ,  $P(X=1) = .6$ , and  $P(X=2) = .2$ . The variance for this probability distribution is

A) 1    B) .63    C) .5    D) .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) .36    B) .6    C) .4    D) .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 the *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) .3020      B) .2634      C) .2013      D) .5      E) 1

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) .0160      B) .1265      C) .2530      D) 1      E) .2070

31. If a population has a standard deviation  $\sigma$ , then the standard deviation of the mean of 100 randomly selected items from this population is

A)  $\sigma$       B)  $100\sigma$       C)  $\sigma/10$       D)  $\sigma/100$       E) .1

32. The central limit theorem states that

A) the sample mean is unbiased

D) the sample standard deviation is approximately normal

B) the sample mean is approximately normal E) none of the above

C) the binomial distribution is skewed

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 50. The value of  $z^*$  to be used in this calculation is

A) 1.645      B) 2.009      C) 1.960      D) .8289      E) .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 have a systolic blood pressure between 122 and 138.
- B) 95% of the population of employees have 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 test statistic is  $z = 1.40$ . The P-value for this test is approximately

- A) .16      B) .08      C) .003      D) .92      E) .70

37. A significance test gives a P-value of .04. From this we can

- A) reject  $H_0$  with  $\alpha = .01$
- B) reject  $H_0$  with  $\alpha = .05$
- C) say that the probability that  $H_0$  is false is .04.
- D) say that the probability that  $H_0$  is true is .04.
- E) none of the above

38. A 95% confidence interval for the mean reading achievement score for a population of third grade students is (44.2, 54.2). The width of this interval is

- A) 95%    B) 10    C) 5    D) 54.2    E) the answer cannot be determined from the information given

39. A 95% confidence interval for the mean reading achievement score for a population of third grade students is (44.2, 54.2). The sample mean is

- A) 44.2    B) 54.2    C) .95    D) 49.2    E) the answer cannot be determined from the information given

40. 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 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

42. You want to compute a 90% confidence interval for the mean of a population with unknown population standard deviation. The sample size is 30. The value of  $t^*$  you would use for this interval is

- A) 1.96      B) 1.645      C) 1.699      D) .90      E) 1.311

43. To use the two-sample  $t$  procedure to perform a significance test on the difference between two means, we assume

- A) the population standard deviations are known
- B) the samples from each population are independent
- C) the distributions are exactly normal in each population
- D) the sample sizes are large
- E) all of the above

44. In an effort to determine what percent of Americans live on farms, a polling organization selects a random sample of 500 Americans. Of this sample, 26 live on farms. Suppose you computed a 95% confidence interval (Do not compute the interval.) Would the margin of error of your confidence interval increase or decrease if you increased the sample size to 1500 persons?

- A) increase     B) decrease

45. In an opinion poll, 25% of 200 people sampled said that they were strongly opposed to having a state lottery. The standard error of the sample proportion is approximately

- A) .03            B) .25            C) .00094       D) 6.12        E) .06

46. You want to design a study to estimate the proportion of students on your campus who agree with the statement, "The student government is an effective organization for expressing the needs of students to the administration." You will use a 95% confidence interval and you would like the margin of error to be .05 or less. The minimum sample size required is approximately

- A) 22            B) 1795          C) 385           D) 271          E) none of the above

47. A two-way table of counts is analyzed to examine the hypothesis that the row and column classifications are independent. There are 3 rows and 4 columns. The degrees of freedom for the  $X^2$  statistic are

- A) 12            B) 6            C)  $n-1$             D) the minimum of  $n_1-1$  and  $n_2-1$        E) none of the above

48. 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 the sample size is 200. A 90% confidence interval for the true slope is

- A) .64 to .86     B) .53 to .97     C) .57 to .93     D)  $-.64$  to .86     E) none of the above

49. 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 the 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) .05            C) .15            D) .95            E) .75

50. Which of the following may be a negative number?

- A) range            B) variance            C) deviation score            D) both B and C

Short Answer Questions - Show all work NEATLY on your own paper. Use graph paper where necessary.

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 weigh 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 on the average a box loses eight ounces in transit. (10 points)

a) What mean weight should the shipper establish so that only one-fourth of 1% (.0025) 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. Analyze the data, and comment on the relationship between level of education and agreement with the statement. (5 points)

| 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 follow are 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. You are quite familiar with gauges that measure. (As examples, think of a fever thermometer, a watch, and an automobile odometer.) How close to the truth is the number you get from such a gauge? Consider odometers, which measure automobile mileage. Suppose 20 cars traveled exactly 10 miles (as determined by surveying) and that the following summary data were obtained. State the null hypothesis, choose an alternative hypothesis, perform  $t$  test, and write a conclusion. (10 points)

$$\Sigma x = 202.60$$

$$\Sigma x^2 = 2053.10$$

Note: the user friendly formula for standard deviation is:  $s = \sqrt{\frac{1}{n-1} \left[ \sum x_i^2 - \frac{1}{n} \left( \sum x_i \right)^2 \right]}$

(Doesn't look very friendly, does it?)

55. 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 support either of the two theorists?

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

56. "The self confidence of that group of recruits is negatively correlated with their success in the obstacle course." Tell what this statement means.

57. 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 a mean weight of 166? Write your conclusion and your reasoning. (Don't perform the test. HINT: Check out the Central Limit Theorem.)

58. Draw a random sample with  $N=6$  from the following scores. Write down each step in your procedure.

21    31    17    13    02    09    57    26    72    140