Exam 1 Statistics 216 Spring 2005

Name _____ Section _____

In each of the following True/False questions (1-10) circle the best answer. Each True/False question is worth 2 points.

1. **TRUE FALSE**
   Linear transformations do not change the overall shape of a distribution.

2. **TRUE FALSE**
   The square of the correlation is the fraction of the variation in values of y that is explained by the least-squares regression of y on x.

3. **TRUE FALSE**
   The standard deviation and the Interquartile Range are both measures of spread.

4. **TRUE FALSE**
   Categorical variables can be normally distributed.

5. **TRUE FALSE**
   The units of measurement of correlation are the same as the units of measurement of the explanatory variable.

6. **TRUE FALSE**
   If the correlation between height in inches and weight in pounds is 0.76 then the correlation between height in centimeters and weight in kilograms will be 1/0.76.

7. **TRUE FALSE**
   If 2 different lists of numbers have exactly the same average of 50 and standard deviation of 10, then the percentage of entries between 40 and 60 must be exactly the same for each list.

8. **TRUE FALSE**
   Randomization creates treatment groups that are similar (except for chance variation) before treatments are applied.

9. **TRUE FALSE**
   The zip code at an address is an example of a quantitative variable.

10. **TRUE FALSE**
    An explanatory variable explains or causes changes in a response variable.

In each of the following multiple choice questions (11-27) choose the **single best** answer. Each multiple choice question is worth 3 points.

11. Which of the following is resistant to outliers.

    A) Correlation

    B) Mean

    C) Median

    D) Standard deviation.

12. As part of the coursework, a class collects data on streams each year. The table shows data on the substrate of the stream.

Substrate	Number of Streams
Limestone	77
Mixed	26
Shale	69

    Which of the following is an appropriate graphical summary for this data?

    A) Histogram

    B) Bar graph

    C) Scatterplot

    D) Stemplot

13. At a party there are 30 students over age 21 and 20 students under age 21. You choose at random 3 of those over 21 and separately choose at random 2 of those under 21 to interview about attitudes about alcohol. This is an example of a

    A) simple random sample.

    B) voluntary response sample.

    C) block sample.

    D) stratified random sample.

14. Sixty-three college men were asked what they thought was their ideal weight. A **five-number summary** of the responses (in pounds) is

123	155	175	190	225

    What is the Interquartile Range?

    A) 102 lbs

    B) 20 lbs

    C) 35 lbs

    D) 175 lbs

15. A scatterplot shows a strong positive linear association between 2 quantitative variables. Which of the following is TRUE about the correlation?

    A) $r > 0$
    B) $r = 0$
    C) $r < 0$
    D) We cannot tell anything about r from the information given.

16. Researchers studying acid rain measured the acidity of precipitation in an isolated wilderness area in Colorado for 150 consecutive weeks. The acidity of a solution is measured by pH, with lower pH values indicating that the solution is more acidic. The researchers reported that the least-squares regression line

$$pH = 5.43 - 0.053weeks$$

    with $r^2 = 0.49$. What is the correlation between pH and weeks?

    A) 0.49
    B) 0.7
    C) -0.7
    D) -0.053

17. Which of the following is NOT a basic principle of experimental design?

    A) stratification
    B) control
    C) randomization
    D) replication

18. A survey question asked of a simple random sample of unmarried men was, "What is the most important feature you consider when deciding whether to date somebody?" The results were found to depend upon whether the interviewer was male or female. This is an example of

    A) bias due to nonresponse.
    B) response bias.
    C) bias due to undercoverage.
    D) bias due to voluntary response.

19. An investigator has a computer file showing family incomes for 1000 subjects in a certain study. These range from \$5,800 a year to \$ 98,600 a year. By accident, the highest income in the file gets changed to \$986,000. Which of the following is affected by this error. Which of the following is a true statement about the effect on the standard deviation?

    A) The standard deviation increases.
    B) The standard deviation stays the same.
    C) The standard deviation decreases.
    D) There is not enough information given to answer this question.

20. A county government wants to determine whether all taxpayers support increasing local taxes in order to provide more public funding to schools. They randomly select 500 schoolchildren from a list of all children enrolled in local schools and then survey the parents of these children about possible tax increases. What is the population of interest?

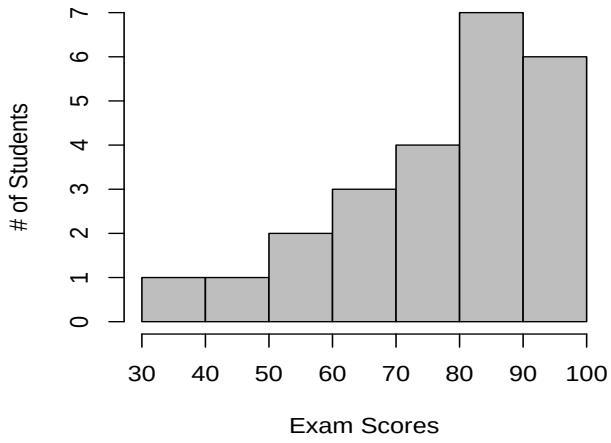
    A) The 500 parents of the children who were selected from the list.

    B) All taxpayers with school aged children.

    C) Children whose names appear on the list.

    D) All taxpayers in the county.

21. Below is a histogram of scores on an exam in a Statistics class. Which of the following best describes this distribution?



    A) Skewed to the right.

    B) Skewed to the left.

    C) Symmetric.

    D) Symmetric with outliers to the left.

22. Data collected on y = annual income in dollars and x = age in years of men aged 20 to 40 yielded a least-squares regression equation of

$$\hat{y} = 15000 + 600x.$$

    The equation is used to predict the annual income of a 60 year old man. This is an example of

    A) a randomization error.

    B) extrapolation.

    C) sampling variability.

    D) skewness.

23. Correlations were computed for the following pairs of variables. Which one is invalid?

    A) Social security number and phone number.

    B) Age and income.

    C) Height and weight

    D) Score on exam 1 and score on exam 2.

24. An observation from a normal distribution has a standardized value of 0. This observation is the

    A) correlation.

    B) standard deviation.

    C) mean.

    D) Interquartile Range.

25. Twenty students agreed to participate in a study on colds. Ten were randomly assigned to receive vitamin C, and the remaining 10 received a tablet that looked and tasted like vitamin C but in fact contained only sugar and flavoring. Neither the students nor the researchers who actually carried out the study and analyzed the data knew which treatment group the students were in. The students were followed for 2 months to see who came down with a cold and who did not. Below are some terms. Which one does not apply to this study?

    A) double-blind

    B) matching

    C) randomized experiment

    D) placebo

26. In May of 2001, the U.S. Bureau of Labor Statistics issued a news release on unemployment rates in 322 metropolitan areas that said, in part:

    In April, 223 metropolitan areas recorded unemployment rates below the U.S. average of 4.2 percent while 99 areas registered higher rates.

    The distribution of unemployment rates is probably

    A) symmetric.

    B) skewed right.

    C) skewed left.

    D) skewed symmetrically.

27. Eleanor scores 680 on the mathematics part of the SAT examination. The distribution of SAT scores in a reference population is normal with mean 500 and standard deviation 100. Gerald takes the ACT mathematics test and scores 27. ACT scores are normally distributed with mean 18 and standard deviation 6. Assuming both tests measure the same kind of ability, who has the higher score?

    A) Eleanor

    B) Gerald

    C) Neither. Both Eleanor and Gerald performed the same.

    D) There is no way to tell.

Show your work problems

28. In 2000, the scores of men on the math part of the SAT approximately followed a normal distribution with mean 533 and standard deviation 110.

    A) What proportion scored over 500? (4pts)

    B) Find the 99th percentile of the test scores. (4pts)

29. A study on college students found that the men had an average weight of 66 kilograms (kg) and a standard deviation of 9 kg. The women had an average weight of 55 kg and a standard deviation of 9 kg.

    A) Find the average and standard deviation of the weight of the **men** in pounds (1 kg = 2.2 pounds). (4pts)

B) If you took the men and women together, would the standard deviation of their weight be less than 9 kg, about 9 kg, or greater than 9 kg? Why? (3pts)

30. Suppose that x = height in inches and y = weight in pounds of men aged 18 to 29 years old and that the equation of the least-squares line is

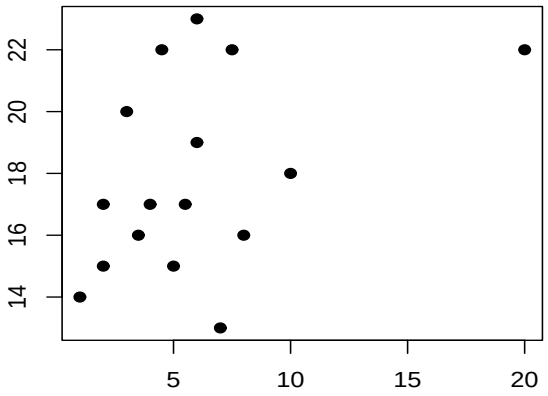
$$\hat{y} = -250 + 6x.$$

A) Interpret the value of 6 in terms of the problem. (4pts)

B) Predict the weight of a 23 year old male who is 72 inches tall. (3pts)

C) A 23 year old male who is 72 inches tall is found to weigh 178 pounds. What is the residual? (3pts)

31. The scatterplot below shows a cluster of points and one outlier.



A) If that point were removed from the data would the correlation become stronger or weaker? (2pts)

B) If that point were removed from the data would the slope of the regression line increase or decrease? (2pts)