STATISTICS 8
CHAPTERS 1 TO 6, SAMPLE MULTIPLE CHOICE QUESTIONS

Circle the best answer.

This scenario applies to Questions 1 and 2: A study was done to compare the lung capacity of coal miners to the lung capacity of farm workers. The researcher studied 200 workers of each type. Other factors that might affect lung capacity are smoking habits and exercise habits. The smoking habits of the two worker types are similar, but the coal miners generally exercise less than the farm workers.

1. Which of the following is the explanatory variable in this study?
   a. Exercise
   b. Lung capacity
   c. Smoking or not
   d. Occupation

2. Which of the following is a confounding variable in this study?
   a. Exercise
   b. Lung capacity
   c. Smoking or not
   d. Occupation

This scenario applies to Questions 3 to 5: A randomized experiment was done by randomly assigning each participant either to walk for half an hour three times a week or to sit quietly reading a book for half an hour three times a week. At the end of a year the change in participants' blood pressure over the year was measured, and the change was compared for the two groups.

3. This is a randomized experiment rather than an observational study because:
   a. Blood pressure was measured at the beginning and end of the study.
   b. The two groups were compared at the end of the study.
   c. The participants were randomly assigned to either walk or read, rather than choosing their own activity.
   d. A random sample of participants was used.

4. The two treatments in this study were:
   a. Walking for half an hour three times a week and reading a book for half an hour three times a week.
   b. Having blood pressure measured at the beginning of the study and having blood pressure measured at the end of the study.
   c. Walking or reading a book for half an hour three times a week and having blood pressure measured.
   d. Walking or reading a book for half an hour three times a week and doing nothing.

Scenario for Questions 3 to 5, continued

5. If a statistically significant difference in blood pressure change at the end of a year for the two activities was found, then:
   a. It cannot be concluded that the difference in activity *caused* a difference in the change in blood pressure because in the course of a year there are lots of possible confounding variables.
   b. Whether or not the difference was caused by the difference in activity depends on what else the participants did during the year.
   c. It cannot be concluded that the difference in activity *caused* a difference in the change in blood pressure because it might be the opposite, that people with high blood pressure were more likely to read a book than to walk.
   d. It can be concluded that the difference in activity *caused* a difference in the change in blood pressure because of the way the study was done.

6. What is one of the distinctions between a population parameter and a sample statistic?
   a. A population parameter is only based on conceptual measurements, but a sample statistic is based on a combination of real and conceptual measurements.
   b. A sample statistic changes each time you try to measure it, but a population parameter remains fixed.
   c. A population parameter changes each time you try to measure it, but a sample statistic remains fixed across samples.
   d. The true value of a sample statistic can never be known but the true value of a population parameter can be known.

7. A magazine printed a survey in its monthly issue and asked readers to fill it out and send it in. Over 1000 readers did so. This type of sample is called
   a. a cluster sample.
   b. a self-selected sample.
   c. a stratified sample.
   d. a simple random sample.

8. Which of the following would be most likely to produce selection bias in a survey?
   a. Using questions with biased wording.
   b. Only receiving responses from half of the people in the sample.
   c. Conducting interviews by telephone instead of in person.
   d. Using a random sample of students at a university to estimate the proportion of people who think the legal drinking age should be lowered.

9. Which one of the following variables is not categorical?
   a. Age of a person.
   b. Gender of a person: male or female.
   c. Choice on a test item: true or false.
   d. Marital status of a person (single, married, divorced, other)

10. A polling agency conducted a survey of 100 doctors on the question "Are you willing to treat women patients with the recently approved pill RU-486"? The conservative margin of error associated with the 95% confidence interval for the percent who say 'yes' is
    a. 50% b. 10% c. 5% d. 2%

11. Which one of these statistics is unaffected by outliers?
    a. Mean
    b. Interquartile range
    c. Standard deviation
    d. Range

12. A list of 5 pulse rates is: 70, 64, 80, 74, 92. What is the median for this list?
    a. 74 b. 76 c. 77 d. 80

13. Which of the following would indicate that a dataset is <u>not</u> bell-shaped?
    a. The range is equal to 5 standard deviations.
    b. The range is larger than the interquartile range.
    c. The mean is much smaller than the median.
    d. There are no outliers.

14. A scatter plot of number of teachers and number of people with college degrees for cities in California reveals a positive association. The most likely explanation for this positive association is:
    a. Teachers encourage people to get college degrees, so an increase in the number of teachers is causing an increase in the number of people with college degrees.
    b. Larger cities tend to have both more teachers and more people with college degrees, so the association is explained by a third variable, the size of the city.
    c. Teaching is a common profession for people with college degrees, so an increase in the number of people with college degrees causes an increase in the number of teachers.
    d. Cities with higher incomes tend to have more teachers and more people going to college, so income is a confounding variable, making causation between number of teachers and number of people with college degrees difficult to prove.

15. The value of a correlation is reported by a researcher to be $r = -0.5$. Which of the following statements is correct?
    a. The x-variable explains 25% of the variability in the *y*-variable.
    b. The x-variable explains –25% of the variability in the *y*-variable.
    c. The x-variable explains 50% of the variability in the *y*-variable.
    d. The x-variable explains –50% of the variability in the *y*-variable.

16. What is the effect of an outlier on the value of a correlation coefficient?
    a. An outlier will always decrease a correlation coefficient.
    b. An outlier will always increase a correlation coefficient.
    c. An outlier might either decrease or increase a correlation coefficient, depending on where it is in relation to the other points.
    d. An outlier will have no effect on a correlation coefficient.

17. One use of a regression line is
    a. to determine if any x-values are outliers.
    b. to determine if any y-values are outliers.
    c. to determine if a change in x causes a change in y.
    d. to estimate the change in y for a one-unit change in x.

18. Past data has shown that the regression line relating the final exam score and the midterm exam score for students who take statistics from a certain professor is:

$$\textit{final exam} = 50 + 0.5 \times \textit{midterm}$$

    One interpretation of the slope is
    a. a student who scored 0 on the midterm would be predicted to score 50 on the final exam.
    b. a student who scored 0 on the final exam would be predicted to score 50 on the midterm exam.
    c. a student who scored 10 points higher than another student on the midterm would be predicted to score 5 points higher than the other student on the final exam.
    d. students only receive half as much credit (.5) for a correct answer on the final exam compared to a correct answer on the midterm exam.

Questions 19 to 21: A survey asked people how often they exceed speed limits. The data are then categorized into the following contingency table of counts showing the relationship between age group and response.

	Exceed Limit if Possible?		
Age	**Always**	**Not Always**	**Total**
Under 30	100	100	200
Over 30	40	160	200
Total	140	260	400

19. Among people with age *over 30*, what's the "risk" of always exceeding the speed limit?
    a. 0.20
    b. 0.40
    c. 0.33
    d. 0.50

20. Among people with age *under 30* what are the odds that they always exceed the speed limit?
    a. 1 to 2
    b. 2 to 1
    c. 1 to 1
    d. 50%

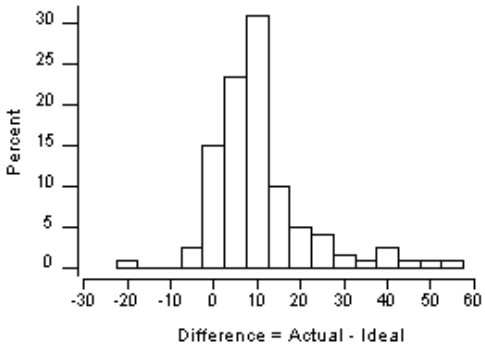
21. What is the relative risk of always exceeding the speed limit for people under 30 compared to people over 30?
    a. 2.5
    b. 0.4
    c. 0.5
    d. 30%

Questions 22 and 23: A newspaper article reported that "Children who routinely compete in vigorous after-school sports on smoggy days are three times more likely to get asthma than their non-athletic peers." (*Sacramento Bee*, Feb 1, 2002, p. A1)

22. Of the following, which is the most important additional information that would be useful before making a decision about participation in school sports?
    a. Where was the study conducted?
    b. How many students in the study participated in after-school sports?
    c. What is the baseline risk for getting asthma?
    d. Who funded the study?

23. The newspaper also reported that "The number of children in the study who contracted asthma was relatively small, 265 of 3,535." Which of the following is represented by 265/3535 = .075?
    a. The overall risk of getting asthma for the children in this study.
    b. The baseline risk of getting asthma for the "non-athletic peers" in the study.
    c. The risk of getting asthma for children in the study who participated in sports.
    d. The relative risk of getting asthma for children who routinely participate in vigorous after-school sports on smoggy days and their non-athletic peers.

Questions 24 to 26: The following histogram shows the distribution of the difference between the actual and "ideal" weights for 119 female students. Notice that percent is given on the vertical axis. Ideal weights are responses to the question "What is your ideal weight"? The difference = actual –ideal. (*Source*: **idealwtwomen** dataset on CD.)



24. What is the approximate shape of the distribution?
    a. Nearly symmetric.
    b. Skewed to the left.
    c. Skewed to the right.
    d. Bimodal (has more than one peak).

25. The median of the distribution is approximately
    a. –10 pounds.
    b. 10 pounds.
    c. 30 pounds.
    d. 50 pounds.

Scenario for Questions 24 to 26, continued

26. Most of the women in this sample felt that their actual weight was
    a. about the same as their ideal weight.
    b. less than their ideal weight.
    c. greater than their ideal weight.
    d. no more than 2 pounds different from their ideal weight.

27. A chi-square test of the relationship between personal perception of emotional health and marital status led to rejection of the null hypothesis, indicating that there is a relationship between these two variables. One conclusion that can be drawn is:
    a. Marriage leads to better emotional health.
    b. Better emotional health leads to marriage.
    c. The more emotionally healthy someone is, the more likely they are to be married.
    d. There are likely to be confounding variables related to both emotional health and marital status.

28. A chi-square test involves a set of counts called "expected counts." What are the expected counts?
    a. Hypothetical counts that would occur of the alternative hypothesis were true.
    b. Hypothetical counts that would occur if the null hypothesis were true.
    c. The actual counts that did occur in the observed data.
    d. The long-run counts that would be expected if the observed counts are representative.

29. Pick the choice that best completes the following sentence. If a relationship between two variables is called statistically significant, it means the investigators think the variables are
    a. related in the population represented by the sample.
    b. not related in the population represented by the sample.
    c. related in the sample due to chance alone.
    d. very important.

30. Simpson's Paradox occurs when
    a. No baseline risk is given, so it is not know whether or not a high relative risk has practical importance.
    b. A confounding variable rather than the explanatory variable is responsible for a change in the response variable.
    c. The direction of the relationship between two variables changes when the categories of a confounding variable are taken into account.
    d. The results of a test are statistically significant but are really due to chance.