1. In a statistics class with 136 students, the professor records how much money each student has in his or her possession during the first class of the semester. The histogram below is of the data collected.

The histogram
A) is skewed right.
B) has an outlier.
C) is asymmetric.
D) has a mean less than 30.
E) all of the above.

2. A set of data has a median that is much larger than the mean. Which of the following statements is most consistent with this information?
A) A stemplot of the data is assymetrical.
B) A stemplot of the data is skewed left.
C) A stemplot of the data is skewed right.
D) The data set must be so large that it would be better to draw a histogram than a stemplot.
E) A stemplot of the data is symmetric.

3. In a statistics class with 136 students, the professor records how much money each student has in his or her possession during the first class of the semester. The histogram below is of the data collected.
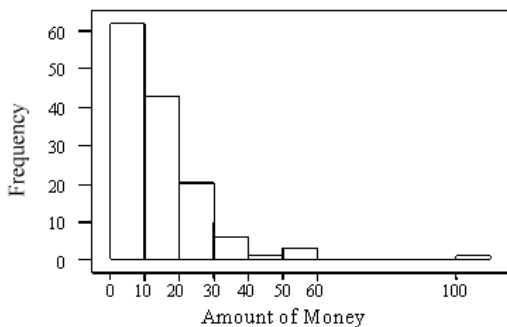
From the histogram, which of the following is true?
A) The mean is much larger than the median.
B) The mean is much smaller than the median.
C) The mean must be at least $10 below the median.
D) It is impossible to compare the mean and median for these data.
E) The mean and median are approximately equal.

4. The Insurance Institute for Highway Safety publishes data on the total damage suffered by compact automobiles in a series of controlled, low-speed collisions. A sample of the data in dollars, with brand names removed, is

      1000   600   800   1000

The interquartile range of the data is
A) 300.   B) 200.   C) 400.   D) 450.   E) none of the above.

5. A sample was taken of the salaries of 20 employees of a large company. The following are the salaries (in thousands of dollars) for this year. For convenience, the data are ordered.

| 28	31	34	35	37	41	42	42	42	47
49	51	52	52	60	61	67	72	75	77

Find the upper and lower bounds to determine if any outliers exist in this data.

Chapter 2

6. If bricks are fired at a temperature above 1125°F, they will crack and must be discarded. If the bricks are placed randomly throughout the kiln, the proportion of bricks that crack during the firing process is closest to
A) 0.62%.   B) 2.28%.   C) 6.2%.   D) 47.72%.   E) 49.38%.

7. A company produces boxes of soap powder labeled "Giant Size 32 Ounces." The actual weight of soap powder in a box has a normal distribution with a mean of 33 ounces and a standard deviation of 0.7 ounces. What proportion of boxes are underweight (i.e., weigh less than 32 ounces)?
A) 0.0764.   B) 0.2420.   C) 0.4236.   D) 0.7580.   E) 0.9236.

Chapter 3 / 4

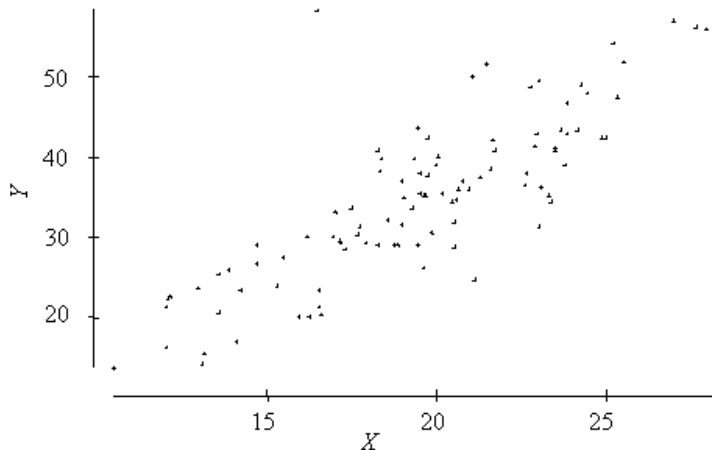8. Consider the scatterplot below.
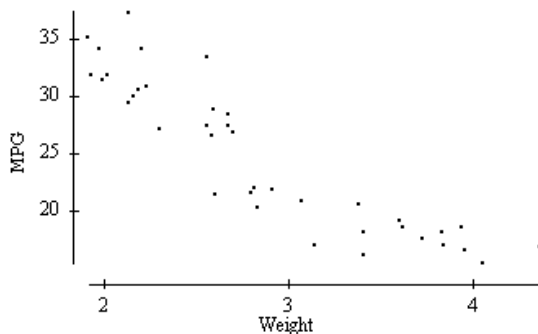
According to the scatterplot, which of the following is a plausible value for the correlation coefficient between weight and MPG?
A) −1.0.   B) −0.9.   C) −0.5.   D) 0.2.   E) 0.7.

9.
Consider the scatterplot below.

The correlation between X and Y is approximately
A) 0.999.   B) 0.8.   C) 0.5.   D) 0.   E) −0.7.

10.
In a statistics course, a linear regression equation was computed to predict a student's final exam score from his/her score on the first test. The equation of the least-squares regression line was

$$y = 10 + 0.9x$$

where y represents the final exam score and x is the score on the first exam. Suppose Joe scores a 90 on the first exam. What would be the predicted value of his score on the final exam?
A) 91.
B) 90.
C) 89.
D) 81.
E) It cannot be determined from the information given. We also need to know the correlation coefficient.

11.

A researcher wishes to study how the average weight Y (in kilograms) of children changes during the first year of life. He plots these averages versus the children's age X (in months) and decides to fit a least-squares regression line to the data with X as the explanatory variable and Y as the response variable. He computes the following quantities.

r = correlation between X and Y = 0.9

$\overline{X}$ = mean of the values of X = 6.5

$\overline{Y}$ = mean of the values of Y = 6.6

s_X = standard deviation of the values of X = 3.6

s_Y = standard deviation of the values of Y = 1.2

The slope of the least-squares line is
A) 0.30.   B) 0.88.   C) 1.01.   D) 2.7.   E) 3.0.

The scatterplot below plots, for each of the 50 states, the percent of 18-year-olds in the state Y in 1990 that graduated from high school versus the state's infant mortality rate (deaths per 1,000 births) X in 1990
.
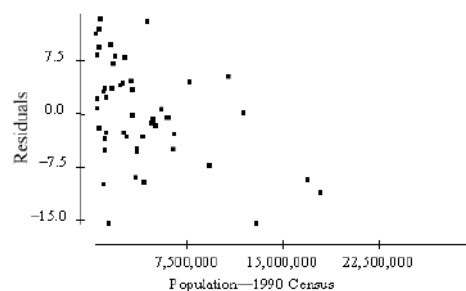
12. Referring to the information above, the least-squares regression line was fitted to the data in the scatterplot and the residuals were computed. A plot of the residuals versus the 1990 population in the state is given above.

    This plot suggests
    A)   that states with larger populations have lower infant mortality rates due to superior hospital facilities.
    B)   that high infant mortality rates imply low nutrition and thus higher dropout rates later in life, but only for states with small populations.
    C)   that population may be a lurking variable in understanding the association between infant mortality rate and percent graduating from high school.
    D)   that high infant mortality rates imply low nutrition and thus higher dropout rates later in life, but only for states with large populations.
    E)   none of the above.

Chapter 5

A study of human development showed two types of movies to groups of children. Crackers were available in a bowl, and the investigators compared the number of crackers eaten by children watching the different kinds of movies. One kind of movie was shown at 8 a.m. (right after the children had breakfast) and another at 11 a.m. (right before the children had lunch). It was found that during the movie shown at 11 a.m., more crackers were eaten than during the movie shown at 8 a.m.. The investigators concluded that the different types of movies had an effect on appetite.

13. The results cannot be trusted because
    A)   the study was not double-blind. Neither the investigators nor the children should have been aware of which movie was being shown.
    B)   children are usually too sleepy early in the morning to watch movies.
    C)   the investigators should have used several bowls, with crackers randomly placed in each.
    D)   the time the movie was shown is a confounding variable.
    E)   the investigators were biased. They knew beforehand what they hoped the study would show.

14. The response variable in this experiment is
    A)   the number of crackers eaten.   D)   the bowls.
    B)   the different kinds of movies.   E)   the time the movie was shown.

C) the children in the study.

Chapter 6

15. I toss a thumbtack 60 times and it lands point up on 35 of the tosses. The approximate probability of landing point up is
A) 35.   B) 0.35.   C) 0.58.   D) 0.65.   E) 58.

16. Suppose we have a *loaded* die that gives the outcomes 1–6 according to the following probability distribution:

X	1	2	3	4	5	6
P(X)	0.1	0.2	0.3	0.2	0.1	0.1

Note that for this die all outcomes are *not* equally likely, as they would be if the die were fair. If this die is rolled 6000 times, the number of times we get a 2 or a 3 should be about
A) 1000.   B) 2000.   C) 3000.   D) 4000.
E) The answer cannot be determined since the probabilities are only approximate.

17. I select two cards from a deck of 52 cards and observe the color of each (26 cards in the deck are red and 26 are black). Which of the following is an appropriate sample space S for the possible outcomes?
A) $S = \{red, black\}$.
B) $S = \{(red, red), (red, black), (black, red), (black, black)\}$, where, for example, (red, red) stands for the event "the first card is red and the second card is red."
C) $S = \{0, 1, 2\}$.
D) All of the above.
E) The results will vary since the cards are drawn without replacement.

18. Event A occurs with probability 0.3. If event A and B are disjoint, then
A) $P(B) \leq 0.3$.   B) $P(B) \geq 0.3$.   C) $P(B) \leq 0.7$.   D) $P(B) \geq 0.7$.
E) $P(B) = 0.21$.

19. Event A occurs with probability 0.2. Event B occurs with probability 0.8. If A and B are disjoint (mutually exclusive), then
A) $P(A$ and $B) = 0.16$.
B) $P(A$ and $B) = 0.84$.
C) $P(A$ and $B) = 1$.
D) $P(A$ or $B) = 0.16$.
E) $P(A$ or $B) = 1$.

20. Suppose there are three cards in a deck, one marked with a 1, one marked with a 2, and one marked with a 5. You draw *two* cards at random and without replacement from the deck of three cards. The sample space $S = \{(1, 2), (1, 5), (2, 5)\}$ consists of these three equally likely outcomes. Let X be the sum of the numbers on the two cards drawn. Which of the following is the correct set of probabilities for X?

A)
X	1	2	5
$P(X)$	1/3	1/3	1/3

B)
X	3	6	7
$P(X)$	1/3	1/3	1/3

C)
X	3	6	7
$P(X)$	3/16	6/16	7/16

D)
X	3	6	7
$P(X)$	1/4	1/2	1/4

E)
X	1	2	5
$P(X)$	1/4	1/2	1/4

Use the following to answer questions 32-33:

An event A will occur with probability 0.5. An event B will occur with probability 0.6. The probability that both A and B will occur is 0.1.

21. Referring to the information above, the conditional probability of A given B
    A) is 0.3.
    B) is 0.2.
    C) is 1/6.
    D) is 0.1.
    E) cannot be determined from the information given.

22. Referring to the information above, we may conclude
    A) that events A and B are independent.   D) that events A and B are disjoint.
    B) that events A and B are complements.   E) none of the above.
    C) that either A or B always occurs.

23. The probability of a randomly selected adult having a rare disease for which a diagnostic test has been developed is 0.001. The diagnostic test is not perfect. The probability the test will be positive (indicating that the person has the disease) is 0.99 for a person with the disease and 0.02 for a person without the disease. The proportion of adults for which the test would be positive is
A) 0.00002.   B) 0.00099.   C) 0.01998.   D) 0.02097.   E) 0.02100.

Chapter 7

In a particular game, a fair die is tossed. If the number of spots showing is either 4 or 5, you win $1; if number of spots showing is 6, you win $4; and if the number of spots showing is 1, 2, or 3, you win nothing. Let X be the amount that you win on a single play of the game.

24. Referring to the information above, the expected value of X is
A) $0.   B) $1.   C) $1.33.   D) $2.50.   E) $4.

25. Referring to the information above, the variance of X is
A) 1.   B) 1.414.   C) 3/2.   D) 2.   E) 13/6.

26. Suppose X is a random variable with mean μ_X and standard deviation σ_X. Suppose Y is a random variable with mean μ_Y and standard deviation σ_Y. The mean of $X + Y$ is
A) $\mu_X + \mu_Y$.
B) $(\mu_X/\sigma_X) + (\mu_Y/\sigma_Y)$.
C) $\mu_X + \mu_Y$, but only if X and Y are independent.
D) $(\mu_X + \mu_Y)/(\sigma_X + \sigma_Y)$.
E) $(\mu_X/\sigma_X) + (\mu_Y/\sigma_Y)$, but only if X and Y are independent.

27. Suppose X is a random variable with mean μ_X and standard deviation σ_X. Suppose Y is a random variable with mean μ_Y and standard deviation σ_Y. The variance of $X + Y$ is
A) $\sigma_X + \sigma_Y$.
B) $(\sigma_X)^2 + (\sigma_Y)^2$.
C) $\sigma_X + \sigma_Y$, but only if X and Y are independent.
D) $(\sigma_X)^2 + (\sigma_Y)^2$, but only if X and Y are independent.
E) $(\sigma_X + \sigma_Y)^2$.

28. Suppose there are three balls in a box. On one of the balls is the number 1, on another is the number 2, and on the third is the number 3. You select two balls at random and without replacement from the box and note the two numbers observed. The sample space S consists of the three equally likely outcomes {(1, 2), (1, 3), (2, 3)}. X, the sum of the numbers on the two balls selected, has the following probability distribution:

X	3	4	5
Probability	1/3	1/3	1/3

The probability that X is at least 4 is
A) 0.   B) 1/3.   C) 9/20.   D) 2/3.   E) 1.

29. A fourth-grade teacher gives homework every night in both mathematics and language arts. The time to complete the mathematics homework has a mean of 10 minutes and a standard deviation of 3 minutes. The time to complete the language arts assignment has a mean of 12 minutes and a standard deviation of 4 minutes. Assuming the times to complete homework assignments in math and language arts are independent, the standard deviation of the time required to complete the entire homework assignment is
    A) 16 minutes.   B) 5 minutes.   C) 4 minutes.   D) 3 minutes.   E) $16/9$ minutes.

Chapter 8

30. A small class has 10 students. Five of the students are male and five are female. I write the name of each student on a 3-by-5 card. The cards are shuffled thoroughly and I choose one at random, observe the name of the student, and replace it in the set. The cards are thoroughly reshuffled and I again choose a card at random, observe the name, and replace it in the set. This is done a total of four times. Let X be the number of cards observed in these four trials with a name corresponding to a male student. The random variable X has which of the following probability distributions?
    A) The normal distribution with mean 2 and variance 1.
    B) The binomial distribution with parameters $n = 4$ and $p = 0.5$.
    C) The binomial distribution with parameters $n = 4$ and $p = 0.1$.
    D) The uniform distribution on 0, 1, 2, 3, and 4.
    E) None of the above.

31. For which of the following counts would a binomial probability model be reasonable?
    A) The number of traffic tickets written by each police officer in a large city during one month.
    B) The number of hearts in a hand of five cards dealt from a standard deck of 52 cards that has been thoroughly shuffled.
    C) The number of 7's in a randomly selected set of five random digits from a table of random digits.
    D) The number of phone calls received in a one-hour period.
    E) All of the above.

33. A set of 10 cards consists of five red cards and five black cards. The cards are shuffled thoroughly and I choose one at random, observe its color, and replace it in the set. The cards are thoroughly reshuffled and I again choose a card at random, observe its color, and replace it in the set. This is done a total of four times. Let X be the number of red cards observed in these four trials. The mean of X is
   A) 4.   B) 2.   C) 1.   D) 0.5.   E) 0.1.

34. A college basketball player makes 80% of his free throws. At the end of a game, his team is losing by two points. He is fouled attempting a three-point shot and is awarded three free throws. Assuming free throw attempts are independent, what is the probability that he makes at least two of the free throws?
   A) 0.896.   B) 0.80.   C) 0.64.   D) 0.512.   E) 0.384.

FRQ example A
Use the following to answer questions 1 through 3:

A survey asks a random sample of 1500 adults in Ohio if they support an increase in the state sales tax from 5% to 6%, with the additional revenue going to education. Let X denote the number in the sample that say they support the increase. Suppose that 40% of *all* adults in Ohio support the increase.

1. The mean of X is

2. The standard deviation of X is

3. The probability that X is more than 750 is

FRQ Example B
Each state conducts an annual study of seat belt use by drivers following guidelines set by the federal government. Seat belt use is observed at randomly chosen road locations at random times during daylight hours. The locations are not an SRS of all locations in

35. ~~Which of the following random variables is geometric?~~
   A) ~~The number of phone calls received in a one-hour period.~~
   B) ~~The number of cards I need to deal from a deck of 52 cards that has been thoroughly shuffled so that at least one of the cards is a heart.~~
   C) ~~The number of digits I will read beginning at a randomly selected starting point in a table of random digits until I find a 7.~~
   D) ~~The length of time a traffic light must be green so that 5 cars can make a left turn.~~
   E) ~~All of the above.~~

the state but rather a sample using the state's counties. In Hawaii, the counties are the islands that make up the state's territory. The survey was conducted on the 4 most populated islands: Oahu, Maui, Hawaii,

and Kauai. The sample sizes on the islands are proportional to the amount of road traffic. Suppose there are 476 road locations in Kauai and we need to randomly select 22 of them to be in the sample.

(a) Use the portion of Table B below to choose the first 3 road locations for the seat belt survey sample. Explain you method clearly.

```
81486   69487   60513   09297   00412   71238   27649   39950

59636   88804   04634   71197   19352   73089   84898   45785
```

(b) What is the name for this kind of sampling design?

FRQ Example C

Researchers are interested in the relationship between cigarette smoking and lung cancer. Suppose an adult male is randomly selected from a particular population. Assume that the following table shows some probabilities involving the compound event that the individual does or does not smoke and the person is or is not diagnosed with cancer:

Event	Probability
smokes and gets cancer	0.08
smokes and does not get cancer	0.17
does not smoke and gets cancer	0.04
does not smoke and does not get cancer	0.71

Suppose further that the probability that the randomly selected individual is a smoker is 0.25.

(a) Find the probability that the individual gets cancer, given that he is a smoker.
(b) Find the probability that the individual does not get cancer, given that he is a smoker.
(c) Find the probability that the individual gets cancer, given that he does not smoke.
(d) Find the probability that the individual does not get cancer, given that he does not smoke.