1. The student government at a high school wants to conduct a survey of student opinion. It wants to begin with a simple random sample of 60 students. Which of the following survey methods will produce a simple random sample?
   A. Survey the first 60 students to arrive at school in the morning.
   B. Survey every 10th student entering the school library until 60 students are surveyed.
   C. Use random numbers to choose 15 each of first-year, second-year, third-year, and fourth-year students.
   D. Number the cafeteria seats. Use a table of random numbers to choose seats and interview the students until 60 have been interviewed.
   E. Number the students in the official roster. Use a table of random numbers to choose 60 students from this roster for the survey.

2. A large elementary school has 15 classrooms, with 24 children in each classroom. A sample of 30 children is chosen by the following procedure. Each of the 15 teachers selects 2 children from his or her classroom to be in the sample by numbering the children from 1 to 24, then using the random digit table to select two different random numbers between 01 and 24. The 2 children with those numbers are in the sample. Did this procedure give a simple random sample of 30 children from the elementary school?
   A. No because we the teachers were not selected randomly.
   B. No, because not all possible groups of 30 children had the same chance of being chosen.
   C. No, because not all children had the same chance of being chosen.
   D. Yes, because each child had the same chance of being chosen.
   E. Yes, because the numbers were assigned randomly to the children.

3. Which of the following can be used to show a cause-and-effect relationship between two variables?
   A. Census
   B. Controlled experiment
   C. Observational study
   D. Sample survey
   E. Cross-sectional survey

4. Which of the following are important in the design of experiments?
   I. Control of confounding variables.
   II. Randomization in assigning subjects to different treatments.
   III. Replication of the experiments using sufficient numbers of subjects.
   A. I and II
   B. I and III
   C. II and III
   D. I, II, and III
   E. None of the above gives the complete set of true responses.

5. The primary reason for using blocking when designing an experiment is to reduce
   A. The sensitivity of the experiment.
   B. Variation.
   C. The need for randomization.
   D. Bias
   E. Confounding.
6. A study is made to determine whether studying Latin helps students achieve higher scores on the verbal section of the SAT exam. In comparing records of 200 students, half of whom have taken at least 1 year of Latin, it is noted that the average SAT verbal score is higher for those 100 students who have taken Latin than for those who have not. Based on this study, guidance counselors, begin to recommend Latin for students who want to do well on the SAT exam. Which of the following are true statements?
   I. While this study indicates a relation, it does not prove causation.
   II. There could well be a confounding variable responsible for the seeming relationship.
   III. Self-selection here makes drawing the counselor’s conclusion difficult.
   A. I and II  B. I and III  C. II and III  D. I, II and III  E. None of the above gives the complete set of true responses.

7. In a 1927 – 1932 Western Electric Company study on the effect of lighting on worker productivity, productivity increased with each increase in lighting but then also increased with every decrease in lighting. If it is assumed that the workers knew a study was in progress, this is an example of
   A. The effect of a treatment unit.
   B. the placebo effect
   C. the control group effect
   D. sampling error
   E. voluntary response bias.

8. Some researchers believe that too much iron in the blood can raise the level of cholesterol. The iron level in the blood can be lowered by making periodic blood donations. A study is performed by randomly selecting half of a group of volunteers to give periodic blood donations while the rest do not. Is this an experiment or an observational study?
   A. An experiment with control group and blinding.
   B. An experiment with blocking.
   C. An observational study with comparison and randomization.
   D. An observational study with little if any bias.
   E. None of the above.

9. Twenty men and twenty women with high blood pressure were subjects in an experiment to determine the effectiveness of a new drug in lowering blood pressure. Ten of the 20 men and 10 of the 20 women were chosen at random to receive the new drug. The remaining 10 men and 10 women received a placebo. The change in blood pressure was measured for each subject. The design of this experiment is
   A. Completely randomized with one factor, drug.
   B. Completely randomized with one factor, gender.
   C. Randomized block, block by drug and gender.
   D. Randomized block, block by drug.
   E. Randomized block, block by gender.
10. To check the effect of cold temperatures on the elasticity of two brands of rubber bands, one box of Brand A and one box of Brand B rubber bands are tested. Ten bands from the Brand A box are placed in a freezer for two hours and ten bands from the Brand B box are kept at room temperature. The amount of stretch before breakage is measured on each rubber band, and the mean for the cold bands is compared to the mean for the others. Is this a good experimental design?
A. No, because the means are not proper statistics for comparison.
B. No, because more than two brands should be used.
C. No, because more temperatures should be used.
D. No, because temperature is confounded with brand.
E. Yes.

11. Suppose you wish to compare the average class size of mathematics classes to the average class size of English classes in your high school. Which is the most appropriate technique for gathering the needed data?
A. Census    B. Sample survey    C. Experiment    D. Observational study
E. None of these methods is appropriate.

12. Which of the following are true statements?
I. Based on careful use of control groups, experiments can often indicate cause-and-effect relationships.
II. While observational studies may suggest relationships, great care must be taken in concluding that there is cause-and-effect because of the lack of control over lurking variables.
III. A complete census is the only way to establish a cause-and-effect relationship absolutely.
A. I and II    B. I and III    C. II and III    D. I, II, and III
E. None of the above gives the complete set of responses.

13. Ann Landers, who writes a daily advice column appearing in newspapers across the country, once asked her readers, “If you had it to do again, would you have children?” Of the more than 10,000 readers who responded, 70% said no. What does this show?
A. The survey is meaningless because of voluntary response bias.
B. No meaningful conclusion is possible without knowing something more about the characteristics of her readers.
C. The survey would have been more meaningful if she had picked a random sample of 10,000 readers who responded.
D. The survey would have been more meaningful if she had used a control group.
E. This is a legitimate sample, randomly drawn from her readers and of sufficient size to allow the conclusion that most of her readers who are parents would have second thoughts about having children.

14. Each of the 29 NBA teams has 12 players. A sample of 58 players is to be chosen as follows. Each team will be asked to place 12 cards with their players’ names into a hat and randomly draw out two names. The two names from each team will be combined to make up the sample. Will this method result in a simple random sample of 348 players?
A. Yes, because each player has the same chance of being selected.
B. Yes, because each team is equally represented.
C. Yes, because this is an example of stratified sampling, which is a special case of simple random sampling.
D. No, because the teams are not chosen randomly.
E. No, because not each group of 58 players has the same chance of being selected.
15. A researcher planning a survey of heads of households in a particular state has census lists for each of the 23 counties in that state. The procedure will be to obtain a random sample of heads of households from each of the counties rather than grouping all the census lists together and obtaining a sample from the entire group. Which of the following is a true statement about the resulting stratified sample?

I. It is not a simple random sample.
II. It is easier and less costly than a simple random sample.
III. It gives comparative information that a simple random sample wouldn’t give.

A. I only. B. I and II. C. I and III. D. I, II and III. E. None of the above gives a complete set of true responses.

16. In one study on the effect of niacin on cholesterol level, 100 subjects who acknowledged being long-time niacin takers had their cholesterol levels compared with those of 100 people who had never taken niacin. In a second study, 50 subjects were randomly chosen to receive niacin and 50 were chosen to receive a placebo.

A. The first study was a controlled experiment, while the second was an observational study.
B. The first study was an observational study, while the second was a controlled experiment.
C. Both studies were controlled experiments.
D. Both studies were observational studies.
E. Each study was part controlled experiment and part observational study.

17. In one study subjects were randomly given either 500 or 1000mg of vitamin C daily, and the number of colds they came down with during a winter season was noted. In a second study people responded to a questionnaire asking about the average number of hours of sleep per night and the number of colds they came down with during a winter season.

A. The first study was an experiment without a control group, while the second was an observational study.
B. The first study was an observational study, while the second was a controlled experiment.
C. Both studies were controlled experiments.
D. Both studies were observational studies.
E. None of the above is a correct statement.

18. Following are the SAT math scores for an AP Statistics class of 20 students: 664, 658, 610, 670, 640, 643, 675, 650, 676, 575, 660, 661, 520, 667, 668, 635, 671, 673, 645, and 650. The distribution of scores is

A. Symmetric B. Skewed to the left C. Skewed to the right D. Uniform E. Bell-shaped

19. A random sample of size 10 was taken from a population. The sample has a variance of zero. Which of the following statements must be true?

I. The population also has a variance of zero.
II. The sample mean is equal to the sample median.
III. The ten data points in the sample are equal in numerical value.

A. I only B. II only C. III only D. I and II E. II and III
20. A company wanted to determine the health care costs of its employees. A sample of 25 employees were interviewed and their medical expenses for the previous year were determined. Later the company discovered that the highest medical expense in the sample was mistakenly recorded as 10 times the actual amount. However, after correcting the error, the corrected amount was still greater than or equal to any other medical expense in the sample. Which of the following sample statistics must have remained the same after the correction was made?

A. Mean  B. Median  C. Mode  D. Range  E. Variance

21. Suppose the average score on a national test is 500 with a standard deviation of 100. If each score is increased by 25, what are the new mean and standard deviation?

A. 500, 100  B. 500, 125  C. 525, 100  D. 525, 105  E. 525, 125

22. Suppose the average score on a national test is 500 with a standard deviation of 100. If each score is increased by 25%, what are the new mean and standard deviation?

A. 500, 100  B. 525, 100  C. 625, 100  D. 625, 105  E. 625, 125

23. Which of the following are true statements?

I. If the sample has variance of 0, the variance of the population is also 0.

II. If the population has variance 0, the variance of the sample is also 0.

III. If the sample has variance of 0, the sample mean and the sample median are equal.

A. I and II  B. I and III  C. II and III  D. I, II, and III  E. None of the above gives the complete set of true responses.

24. A professor teaches two statistics classes. The morning class has 25 students and their average on the first test was 82. The evening class has 15 students and their average on the same test was 74. What is the average on this test if the professor combines the scores for both classes?

A. 76  B. 78  C. 79  D. 80  E. The average cannot be calculated since individual scores of each student are not available.

25. If the range of a data set is 144, which of the following are reasonable values for the standard deviation?

A. 1.40, 1.42, 1.44  B. 12, 14, 16  C. 25, 30, 35  D. 48, 72, 96  E. All of the above

26. Mathematically speaking, casinos and life insurance companies make a profit because of

A. Their understanding of sampling error and sources of bias.

B. Their use of well-designed, well-conducted surveys and experiments.

C. Their use of simulation and probability distributions.

D. The central limit theorem

E. The law of large numbers.
27. Suppose that for any given year, the probabilities that the stock market declines, that women’s hemlines are lower, and that both events occur are 0.40, 0.35 and 0.30 respectively. Are the two events independent?
   A. Yes because (0.4)(0.35) ≠ 0.3
   B. No because (0.4)(0.35) ≠ 0.3
   C. Yes because (0.4) > (0.35) > (0.3)
   D. No because 0.5(0.3 + 0.4) = 0.35
   E. There is insufficient information to answer this question.

28. Suppose that among 6000 students at a high school, 1500 are taking honors courses, 1800 prefer watching basketball to watching football. If taking honors courses and preferring basketball are independent, how many students are both taking honors course and prefer basketball to football?
   A. 300   B. 330   C. 450   D. 825   E. There is insufficient information to answer.

29. The Russian Health Ministry announced that one-quarter of the country’s hospitals has no sewage system and one-seventh had no running water. What is the probability that a Russian hospital will have at least one of these problems
   i. If the two problems are independent?
   ii. If the hospitals with a running water problem are a subset of those with a sewage problem?
   A. 11/28, 1/4   B. 11/28, 1/7   C. 9/28, 1/4   D. 9/28, 1/7   E. 5/14, 1/4

30. An author was asked the following question, “Suppose a person was having two surgeries performed at the same time. If the chances of success for surgery A are 85%, and the chances of success for surgery B are 90%, what are the chances that both would fail?” What do you think of the author’s solution: (0.15)(0.10) = 0.015 or 1.5%?
   A. The solution is mathematically correct but not explained very well.
   B. The solution is both mathematically correct and intuitively obvious.
   C. The use of complementary events is incorrect.
   D. The use of the general addition formula is incorrect.
   E. There is an assumption of independent events, which is most likely wrong.

31. For college-bound high school seniors in 2014, the nationwide mean Sat verbal score was 505 with a standard deviation of about 110, and the mean Sat math score was 508 with a standard deviation of about 110. Students who do well on the verbal portion of the SAT tend to do well on the mathematics portion. If the two scores for each student are added, the mean of the combined score is 1,013. What is the standard deviation of the combined verbal and math scores?
   A. 77.78   B. 110   C. 155.56   D. 220
   E. The standard deviation cannot be computed because the events are dependent.
32. Suppose X and Y are random variables with $\mu_X = 10, \sigma_X = 3, \mu_Y = 15, \text{ and } \sigma_Y = 4$. Given that X and Y are independent, what are the mean and standard deviation of the random variable $X + Y$?

A. $\mu_{X+Y} = 25, \sigma_{X+Y} = 3.5$  
B. $\mu_{X+Y} = 25, \sigma_{X+Y} = 5$  
C. $\mu_{X+Y} = 25, \sigma_{X+Y} = 7$  
D. $\mu_{X+Y} = 12.5, \sigma_{X+Y} = 7$  
E. There is insufficient information to answer this question.

33. Suppose X and Y are random variables with $E[X] = 500, Var(X) = 50, E[Y] = 400, \text{ and } Var(Y) = 30$. Given that X and Y are independent, what are the mean and standard deviation of the random variable $X - Y$?

A. $E[X - Y] = 100, Var(X - Y) = 20$  
B. $E[X - Y] = 100, Var(X - Y) = 80$  
C. $E[X - Y] = 900, Var(X - Y) = 20$  
D. $E[X - Y] = 100, Var(X - Y) = 80$  
E. There is insufficient information to answer this question.

34. Suppose the average height of policemen is 71 inches with a standard deviation of 4 inches, while the average for policewomen is 66 inches with a standard deviation of 3 inches. If a committee looks at all the ways of pairing up one male with one female officer, what will be the mean and standard deviation for the difference in heights for the set of possible partners?

A. Mean of 5 inches with a standard deviation of 1 inch.  
B. Mean of 5 inches with a standard deviation of 3.5 inches.  
C. Mean of 5 inches with a standard deviation of 5 inches.  
D. Mean of 68.5 inches with a standard deviation of 1 inch.  
E. Mean of 68.5 inches with a standard deviation of 3.5 inches.

35. Joe and Matthew plan to visit a bookstore. Based on their previous visits to this bookstore, the probability distributions of the number of books they will buy are given below.

<table>
<thead>
<tr>
<th>Number of books Joe will buy</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>Number of books Matthew will buy</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.50</td>
<td>0.25</td>
<td>0.25</td>
<td>Probability</td>
<td>0.25</td>
<td>0.50</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Assuming that Joe and Matthew make their decisions independently, what is the probability that they will purchase no books on this visit to the bookstore?

A. 0.0625  
B. 0.1250  
C. 0.1875  
D. 0.2500  
E. 0.7500

36. Given the following probability distributions:

<table>
<thead>
<tr>
<th>X</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(X=x)</td>
<td>1/6</td>
<td>2/3</td>
<td>?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Y</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(Y = y)</td>
<td>?</td>
<td>1/4</td>
<td>1/4</td>
<td>?</td>
</tr>
</tbody>
</table>

The tables above show part of the probability distribution for random variables X and Y. If X and Y are independent and the joint probability $P(X = 3, Y = 4) = 1/16$, then $P(Y = 1)$ =

A. 1/8  
B. 1/6  
C. 1/4  
D. 3/8  
E. 1/2
37. Following are parts of the probability distributions for the random variables X and Y.

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th></th>
<th></th>
<th>Y</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P(X=x)</td>
<td>?</td>
<td>0.4</td>
<td>?</td>
<td>P(Y = y)</td>
<td>0.3</td>
<td>?</td>
</tr>
</tbody>
</table>

If X and Y are independent and the joint probability \( P(X = 1, Y = 1) = 0.06 \) and \( P(X = 2, Y = 2) = 0.1 \), what is \( P(X = 3, Y = 3) \)?

A. 0.16  
B. 0.18  
C. 0.25  
D. 0.40  
E. 0.425

38. Following are parts of the probability distributions for the random variables X and Y.

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th></th>
<th></th>
<th></th>
<th>Y</th>
<th></th>
<th></th>
</tr>
</thead>
</table>

If X and Y are independent and the joint probability \( P(X = 1, Y = 1) = 0.14 \) and \( P(X = 1, Y = 2) = 0.06 \), what is \( P(Y = 2) \)?

A. 0.08  
B. 0.20  
C. 0.30  
D. 0.70  
E. It cannot be determined from the given information.

39. Following are parts of the probability distributions for the random variables X and Y.

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th></th>
<th></th>
<th>Y</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P(X=x)</td>
<td>0.30</td>
<td>?</td>
<td>0.50</td>
<td>P(Y = y)</td>
<td>0.30</td>
<td>?</td>
</tr>
</tbody>
</table>

If \( P(X = 2, Y = 2) = 0.10 \), is it still possible that X and Y are independent?

A. Yes, but only if \( P(X = 1, Y = 1) = 0.09 \).
B. Yes, but only if \( P(X = 3, Y = 3) = 0.05 \).
C. Yes, but only if both of the above hold.
D. No.
E. There is insufficient information to answer this question.

40. Suppose that in a certain part of the world, in any 50 year period the probability of a major plague is 0.39, the probability of a major famine is 0.52, and the probability of both a plague and a famine is 0.15. What is the probability of a famine given that there is a plague?

A. 0.240  
B. 0.288  
C. 0.370  
D. 0.385  
E. 0.760

41. Suppose that 2% of a clinic’s patients are known to have cancer. A blood test is developed that is positive in 98% of patients with cancer but also positive in 3% of patients who do not have cancer. If a person who is chosen at random from the clinic’s patients is given the test and it comes out positive, what is the probability that the person actually has cancer?

A. 0.02  
B. 0.40  
C. 0.50  
D. 0.60  
E. 0.98
42. A computer technician notes that 40% of computers fail because of the hard drive, 25% because of the monitor, 20% because of the disk drive, and 15% because of the microprocessor. If the problem is not in the monitor, what is the probability that it is in the hard drive?

A. 0.150  B. 0.400  C. 0.417  D. 0.533  E. 0.650

43. A magazine has 1,620,000 subscribers, of whom 640,000 are women and 980,000 are men. Thirty percent of the women read the advertisements in the magazine and 50% of the men read the advertisements in the magazine. A random sample of 100 subscribers is selected. What is the expected number of subscribers in the sample who read the advertisements?

A. 30  B. 40  C. 42  D. 50  E. 80

44. Suppose that 60% of students who take the AP Statistics exam score 4 or 5, 25% score 3 and the rest score 1 or 2. Suppose further that 95% of those scoring 4 or 5 receive college credit, 50% of those scoring 3 receive such credit and 4% of those scoring a 1 or 2 receive credit. If a student who is chosen at random from among those taking the exam receives college credit, what is the probability that she received a 3 on the exam?

A. 0.125  B. 0.178  C. 0.701  D. 0.813  E. 0.822

45. There are two games involving flipping a coin. In the first game you win a prize if you can throw between 40% and 60% heads. In the second game you win if you can throw more than 75% heads. For each game would you rather flip the coin 50 times or 500 times?

A. 50 times for each game.  B. 500 times for each game.  C. 50 times for the first game and 500 for the second.  D. 500 times for the first game and 50 for the second.  E. The outcomes of the games do not depend on the number of flips.

46. Suppose you toss a coin ten times and it comes up heads every time. Which of the following is a true statement?

A. By the Law of Large Numbers, the next toss is more likely to be tails than another heads.  B. By the properties of conditional probability, the next toss is more likely to be heads given that ten tosses in a row have been heads.  C. Coins actually do have memories, and thus what comes up on the next toss is influenced by past tosses.  D. The Law of Large Numbers tells how many tosses will be necessary before the percentages of heads and tails are again in balance.  E. The probability that the next toss will again be heads is 0.50.

47. Circuit boards are assembled by selecting 4 computer chips at random from a large batch of chips. In this batch of chips, 90% of the chips are acceptable. Let X denote the number of acceptable chips out of a sample of 4 chips from this batch. What is the least probable value of X?

A. 0  B. 1  C. 2  D. 3  E. 4
48. The heights of adult women are approximately normally distributed with a mean of 65 inches and a standard deviation of 2 inches. If Rachel is at the 99th percentile in height for adult women, then her height, in inches, is closest to
A. 60 B. 62 C. 68 D. 70 E. 74

49. Which of the following are true statements?
I. The area under a normal curve is always equal to 1, no matter what the mean and standard deviation are.
II. The smaller the standard deviation of a normal curve, the higher and narrower the graph.
III. Normal curves with different means are centered around different numbers.
A. I and II B. I and III C. II and III D. I, II and III E. None of the above gives the complete set of true responses.

50. Which of the following are true statements?
I. The area under the standard normal curve between 0 and 2 is twice the area between 0 and 1.
II. The area under the standard normal curve between 0 and 2 is half the area between -2 and 2.
III. For the standard normal curve, the interquartile range is approximately 3.
A. I and II B. I and III C. II and III D. I, II and III E. None of the above gives the complete set of true responses.

51. The distribution of weights of loaves of bread from a certain bakery follows approximately a normal distribution. Based on a very large sample, it was found that 10 percent of the loaves weighed less than 15.34 ounces and 20 percent of the loaves weighed more than 16.31 ounces. What are the mean and standard deviation of the distribution of the weights of loaves of bread?
A. \( \mu = 15.82, \sigma = 0.48 \) B. \( \mu = 15.82, \sigma = 0.69 \) C. \( \mu = 15.87, \sigma = 0.50 \)
D. \( \mu = 15.93, \sigma = 0.46 \) E. \( \mu = 16.00, \sigma = 0.50 \)

52. Gina’s doctor told her that the standardized score for her systolic blood pressure, as compared to the blood pressure of other women her age, is 1.50. Which of the following is the best interpretation of this standardized score?
A. Gina’s systolic blood pressure is 150.
B. Gina’s systolic blood pressure is 1.50 standard deviations above the average systolic blood pressure of women her age.
C. Gina’s systolic blood pressure is 1.50 above the average systolic blood pressure of women her age.
D. Gina’s systolic blood pressure is 1.50 times the average systolic blood pressure of women her age.
E. Only 1.5% of women Gina’s age have a higher systolic blood pressure than she does.
53. At a college the scores on the chemistry final exam are approximately normally distributed, with a mean of 75 and a standard deviation of 12. The scores on the calculus final are also approximately normally distributed, with a mean of 80 and a standard deviation of 8. A student scored 81 on the chemistry final and 84 on the calculus final. Relative to the students in each representative class, in which subject did this student do better?
A. The student did better in chemistry.
B. The student did better in calculus.
C. The student did equally well in each course.
D. There is no basis for comparison, since the subjects are different from each other and are in different departments.
E. There is not enough information for comparison, because the number of students in each class is not known.

54. Which of the following would you expect to be true about the correlation between distances and tolls on the Pennsylvania Turnpike?
A. Strong and positive
B. Weak and positive
C. Strong and negative
D. Weak and negative
E. Zero

55. If every woman married a man who was exactly 2 inches taller than she, what would the correlation between the heights of married men and women be?
A. Somewhat negative
B. 0
C. Somewhat positive
D. Nearly 1
E. 1

56. Which of the following statements about the correlation coefficient $r$ are true?
I. It is not affected by changes in the measurement units of the variables.
II. It is not affected by which variable is called $x$ and which is called $y$.
III. It is not affected by extreme values.
A. I and II
B. I and III
C. II and III
D. I, II and III
E. None of the above gives a complete set of true responses.

57. Suppose the correlation between two variables is $r = 0.23$. What will the new correlation be if 0.14 is added to all values of the $x$-variable, every value of the $y$-variable is doubled, and the two variables are interchanged?
A. 0.23
B. 0.37
C. 0.74
D. -0.23
E. -0.74

58. Consider the set of points $\{(2, 5), (3, 7), (4, 9), (5, 12), (10, n)\}$. What should $n$ be so that the correlation between the $x$ and $y$ values is 1?
A. 21
B. 24
C. 25
D. A value different from any of the above.
E. No value for $n$ can make $r = 1$. 
59. Which of the following statements about correlation $r$ are true?

I. The correlation coefficient and the slope of the regression line have the same sign.
II. A correlation of -0.35 and a correlation of +0.35 show the same degree of clustering around the regression line.
III. A correlation of 0.75 indicated a relationship that is 3 times as linear as one for which the correlation is only 0.25.

A. I and II  
B. I and III  
C. II and III  
D. I, II and III  
E. None of the above gives a complete set of true responses.

60. Suppose that the scatterplot of logX and logY shows a strong positive correlation close to 1. Which of the following is true?

I. The variables X and Y also have a correlation close to 1.
II. A scatterplot of the variables X and Y shows a strong nonlinear pattern.
III. The residual plot of the variables X and Y shows a random pattern.

A. I only  
B. II only  
C. III only  
D. I and II  
E. I, II and III

61. There is a linear relationship between the number of chirps made by the striped ground cricket and the air temperature. A least squares fit of some data collected by a biologist gives the model \( \hat{y} = 25.2 + 3.3x \) where \( 9 < x < 25 \). Where \( x \) is the number of chirps per minute and \( \hat{y} \) is the estimated temperature in degrees Fahrenheit. What is the estimated increase in temperature that corresponds to an increase of 5 chirps per minute?

A. 3.3°F  
B. 16.5°F  
C. 25.2°F  
D. 28.5°F  
E. 41.7°F

62. Suppose the regression line for a set of data, \( y = 3x + b \), passes through the point (2, 5). If \( \bar{x} \) and \( \bar{y} \) are the sample means of the x and y values respectively, then \( \bar{y} = \)

A. \( \bar{x} \)  
B. \( \bar{x} - 2 \)  
C. \( \bar{x} + 5 \)  
D. \( 3\bar{x} \)  
E. \( 3\bar{x} - 1 \)

63. Data are obtained for a group of college freshmen examining their SAT scores (math plus verbal) from their senior year of high school and their GPAs during their first year of college. The resulting regression equation is \( \hat{y} = 0.00161x + 1.35 \) with \( r = 0.632 \). What percentage of variation in GPAs can be explained by looking at SAT scores?

A. 0.161%  
B. 16.1%  
C. 39.9%  
D. 63.2%  
E. This value cannot be computed from the information given.

64. The data below gives the heart disease rates per 100,000 people in the United States for certain years.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Death rate</td>
<td>307.6</td>
<td>286.2</td>
<td>253.6</td>
<td>217.8</td>
<td>202.0</td>
</tr>
</tbody>
</table>

Which of the following is a correct interpretation of the slope of the best-fitting straight line for the above data?

A. The heart disease rate per 100,000 people has been dropping about 3,627 per year.
B. The baseline heart disease rate is 7,386.87
C. The regression line explain 96.28% of the variation in heart disease death rates over the years.
D. The regression line explains 98.12% of the variation in heart disease rates over the years.
E. Heart disease will be cured in the year 2036.
65. Based on the regression line from problem #64, what is the predicted death rate for the year 1983?
A. 145.8 per 100,000 people  
B. 192.5 per 100,000 people  
C. 196.8 per 100,000 people  
D. 198.5 per 100,000 people  
E. None of the above

66. Which of the following statements about influential points are true?
I. Influential scores have large residuals.
II. Removal of an influential score sharply affects the regression line.
III. An x-value that is an outlier in the x-variable is more indicative that a point is influential than a y-value that is an outlier in the y-variable.
A. I and II  
B. I and III  
C. II and III  
D. I, II and III  
E. None of the above gives a complete set of true responses.

67. With regard to regression, which of the following statements about outliers are true?
I. Outliers have large residuals.
II. A point may not be an outlier even though its x-value is an outlier in the x-variable and its y-value may be an outlier in the y-variable.
III. Removal of an outlier sharply affects the regression line.
A. I and II  
B. I and III  
C. II and III  
D. I, II and III  
E. None of the above gives a complete set of true responses.

68. Which of the following statements about residuals are true?
I. The mean of the residuals is always zero.
II. The regression line for a residual plot is a horizontal line.
III. A definite pattern in the residual plot is an indication that a nonlinear model will show a better fit to the data than the straight regression line.
A. I and II  
B. I and III  
C. II and III  
D. I, II and III  
E. None of the above gives a complete set of true responses.

69. For which of the following distributions is the mean greater than the median?

(A)  
(B)  
(C)  
(D)  
(E)
70. The XYZ Office Supplies Company sells calculators in bulk at wholesale prices as well as individually at retail prices. Next year’s sales will depend on market conditions, but executives use probability to find estimates of sales for the coming year. The following tables are estimates for next year’s sales.

**WHOLESALE SALES**

<table>
<thead>
<tr>
<th>Number sold</th>
<th>2,000</th>
<th>5,000</th>
<th>10,000</th>
<th>20,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.1</td>
<td>0.3</td>
<td>0.4</td>
<td>0.2</td>
</tr>
</tbody>
</table>

**RETAIL SALES**

<table>
<thead>
<tr>
<th>Number sold</th>
<th>600</th>
<th>1,000</th>
<th>1,500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.4</td>
<td>0.5</td>
<td>0.1</td>
</tr>
</tbody>
</table>

What profit does XYZ expect to make for the next year if the profit from this each calculator sold is $20 at wholesale and $30 at retail?

A. $10,590   B. $220,700   C. $264,750   D. $833,100   E. $1,002,500

71. A large medical experiment involving 22,000 male physicians, attempted to determine whether aspirin could help prevent heart attacks. In this study, one group of about 11,000 physicians took an aspirin a day, while a control group took a placebo. After several years, it was determined that the physicians in the group that took aspirin has significantly fewer heart attacks than the physicians in the control group. Which of the following explains why it would NOT be appropriate to say that everyone should take aspirin every other day?

I. The study included only physicians, and different results may occur in individuals in other occupations.
II. The study included only males and there may be different results for females.
III. Although taking aspirin may be helpful in preventing heart attacks, it may be harmful to some other aspects of health.

A. I only   B. II only   C. III only   D. II and III only   E. I, II and III

72. When travelers chance airlines during connecting flights, each airline receives a portion of the fare. Several years ago, the major airlines used a sample trial period to determine what percentage of certain fares each should collect. Using these statistical results to determine fare splits, the airlines now claim huge savings over previous clerical costs. Which of the following is true?

I. The airlines ran an experiment using a trial period for the control group.
II. The airlines ran an observational study using calculations from a trial period as a sample.
III. The airlines feel that any monetary error in fare splitting resulting from using a statistical sample is smaller than the previous clerical costs necessary to calculate exact fare splits.

A. I only   B. II only   C. III only   D. I and III   E. II and III
73. The histogram below displays a set of measurements. Which of the following boxplots below the histogram displays the same set of measurements?

![Boxplot Image]

74. Consider the following back-to-back stemplot:

![Stemplot Image]

Which of the following are true statements?

I. The distributions have the same mean.
II. The distributions have the same range.
III. The distributions have the same variance.

A. II only  
B. I and II  
C. I and III  
D. II and III  
E. I, II and III
Every Thursday, the video store has “roll the dice” day. A customer may choose to roll two fair dice and rent a second movie for an amount, in cents, equal to the numbers on the top of the dice, with the larger number first. For example, if the customer rolls a two and a four, a second movie may be rented for $0.42. Let $X$ represent the amount paid for a second movie on “roll the dice” day. The expected value of $X$ is $0.47 and the standard deviation of $X$ is $0.15.

75. If a customer rolls the dice and rents a second movie every Thursday for 20 consecutive weeks, what is the total amount that the customer would expect to pay for these second movies?

A. $0.45  
B. $0.47  
C. $0.67  
D. $3.00  
E. $9.40

76. If a customer rolls the dice and rents a second movie every Thursday for 30 consecutive weeks, what is the approximate probability that the total amount paid for these second movies will exceed $15.00?

A. 0  
B. 0.09  
C. 0.14  
D. 0.86  
E. 0.91

77. Which of the following is the best estimate of the standard deviation of the distribution shown in the figure below?

A. 5  
B. 10  
C. 30  
D. 50  
E. 60

78. Consider the following scatterplot of midterm and final exam scores for a class of 15 students.

Which of the following are true statements?

I. The same number of students scored 100 on the midterm exam as scored 100 on the final exam.
II. Students who scored higher on the midterm exam tended to score higher on the final exam.
III. The scatterplot shows a moderate negative correlation between midterm and final exam scores.

A. I and II  
B. I and III  
C. II and III  
D. I, II and III  
E. None of the above gives the complete set of true responses.
79. Which of the following are true statements about blocking?

I. Blocking is to experimental design as stratification is to sampling design.
II. By controlling certain variables, blocking can make conclusions more specific.
III. The paired comparison design is a special case of blocking.

A. I and II  
B. I and III  
C. II and III  
D. I, II and III  
E. None of the above gives a complete set of true responses.

PART II: Free Response

Directions: Answer the following questions in a complete and concise fashion. Be sure to show all work.

80. A study is to be designed to examine the life expectancies of tall people versus those of short people. Which is more appropriate, an observational study or an experiment and why?

81. Researchers often mark wildlife in in order to identify particular individuals across time and space. A study of butterfly migration is designed to determine which location on the butterflies’ wings is best for marking. The six possible locations are those shown as A through F in the figure below. The butterfly in the figure is a monarch.

Because marks in certain locations may be more likely to attract predators or cause problems than marks in other locations, the goal is to determine whether the six marking locations result in equivalent chances of successful migration. To test this, researchers plan to mark 3,600 butterflies and release them, then count how many arrive displaying each marking location at the end of the migratory path.

A. Briefly describe a method you could use to assign the marking locations if you wanted to ensure that exactly 600 butterflies were marked in each location.

B. Briefly describe a method you could use to assign the marking location if you wanted the location to be independent from one butterfly to the next, and wanted each location assigned with probability 1/6 each time.

82. There is a pressure point on the wrist that some doctors believe can be used to help control the nausea experienced following certain medical procedures. The idea is to place a band containing a small marble firmly on a patient’s wrist so that the marble is located directly over the pressure point. Describe how an experiment might be run on 50 postoperative patients.
83. A chemical fertilizer company wishes to test whether using their product results in superior vegetables. After dividing a large field into small plots, how might the experiment proceed?

84. The famous Pepsi-Coke challenge had subjects compare the taste of samples of each drink. How could such a paired comparison be set up?

85. A new type of fish food has become available for salmon raised on fish farms. Your task is to design an experiment to compare the weight gain of salmon raised over a six-month period on the new and the old types of food. The salmon you will use for this experiment have already been randomly placed in eight large tanks in a room that has a considerable temperature gradient. Specifically, tanks on the north side of the room tend to be much colder than those on the south side. The arrangement of tanks is shown on the diagram below.

![Diagram of tanks and temperature gradient]

Describe a design for this experiment that takes account of the temperature gradient.

86. Suppose a new drug is developed that appears in laboratory settings to completely prevent people who test positive for HIV from ever developing full blown AIDS. Putting all ethical considerations aside, design an experiment to test the drug. What ethical considerations might arise during the testing that would force an early end to the experiment?

87. Suppose that in your statistics class of 30 students, 2 students discover they have the same birthday (month and date). You decide to test whether this is an unusual occurrence by using a random number table to simulate picking 30 randomly selected birth dates. Start at line 110 of the table from your textbook.
   A. Clearly explain how to perform this test and then do so, showing your procedure step by step and your result.
   B. Do you feel justified in making some conclusion based on the result of your test? Explain.
   C. Can you make any definite conclusion about the probability of finding two students with the same birth month? Explain.
88. Following are parallel boxplots showing the daily price fluctuations of a certain common stock over the course of 5 years. What trends do the boxplots show?

89. A plot of the number of defective items produced during 20 consecutive days at a factory is shown below.

A. Draw a histogram that shows the frequencies of the number of defective items.

B. Give one fact that is obvious from the histogram but is not obvious from the scatterplot.

C. Give one fact that is obvious from the scatterplot but is not obvious from the histogram.

90. The graph below shows the cumulative proportions plotted against age distribution for the year 1900 with the age distribution for the year 2000.

A. Approximate the median age for each distribution.

B. Approximate the interquartile range for each distribution.

C. Using the results from parts A and B, write a sentence or two comparing the age distributions for the years 1900 and 2000.
91. The monthly rainfall in a small South American city is given below:

<table>
<thead>
<tr>
<th>Month</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>Aug</th>
<th>Sept</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rain (in)</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>37</td>
<td>31</td>
<td>16</td>
<td>28</td>
<td>24</td>
<td>9</td>
<td>4</td>
</tr>
</tbody>
</table>

What are the measures of variability?

92. Suppose the probability that a construction company will be awarded a certain contract is 0.25, the probability that it will be awarded a second contract is 0.21, and the probability that it will get both contracts is 0.13. What is the probability that the company will win at least one of the two contracts?

93. Concessionaires know that attendance at a football stadium will be 60,000 on a clear day, 45,000 if there is light snow, and 15,000 if there is heavy snow. Furthermore, the probability of clear skies, light snow, or heavy snow on any particular day is 1/2, 1/3, and 1/6 respectively. What average attendance should be expected for the season?

94. In a lottery, 10,000 tickets are sold at $1 each with a prize of $7,500 for one winner. What is the average result for each bettor?

95. A manager must choose among three options. Option A has a 10% chance of resulting in a $250,000 gain but otherwise will result in a $10,000 loss. Option B has a 50% chance of gaining $40,000 and a 50% chance of losing $2,000. Finally, option C has a 5% chance of gaining $800,000 but otherwise will result in a loss of $20,000. Which option should the manager choose?

96. Die A has four 9’s and two 0’s on its faces. Die B has four 3’s and two 11’s on its faces. When either of these dice is rolled, each face has an equal chance of landing on top. Two players are going to play a game. The first player selects a die and rolls it. The second player rolls the remaining die. The winner is the player whose die has the higher number on top.

A. Suppose you are the first player and you want to win the game. Which die would you select? Justify your answer.

B. Suppose the player using die A receives 45 tokens each time he or she wins the game. How many tokens must the player using die B receive each time he or she wins in order for this to be a fair game? Explain how you found your answer. (A fair game is one in which the player using die A and the player using die B both end up with the same number of tokens in the long run.)
97. The probability that a student will receive a state grant is $1/3$, while the probability that she will be awarded a federal grant is $1/2$. If whether or not she receives one grant is not influenced by whether or not she receives the other, what is the probability of her receiving both grants?

98. The following are the probability distributions for random variables $X$ and $Y$.

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th></th>
<th>Y</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>$P(X=x)$</td>
<td>0.3</td>
<td>0.7</td>
<td>0.6</td>
<td>0.4</td>
</tr>
</tbody>
</table>

A. If $X$ and $Y$ are independent random variables, what is their joint probability table?

B. Give a possible joint probability table for which $X$ and $Y$ are not independent.

99. If 90% of the households in a certain region have answering machines and 50% have both answering machines and call waiting, what is the probability that a household chosen at random and found to have an answering machine also has call waiting?

100. A laboratory test for detection of a certain disease gives a positive result 5 percent of the time for people who do not have the disease. The test gives a negative result 0.3 percent of the time for people who have the disease. Large scale studies have shown that the disease occurs in 2% of the population.

A. What is the probability that a person selected at random would test positive for this disease? Show your work.

B. What is the probability that a person selected at random who tests positive for the disease does not have the disease? Show your work.

101. Assume there is no overlap between the 56% of the population who wear glasses and 4% who wear contacts. If 55% of those who wear glasses are women and 63% of those who wear contacts are women, what is the probability that the next person you encounter on the street will be

A. A woman with glasses?

B. A woman with contacts?

C. A man with contacts?

D. A person not wearing glasses or contacts? (Explain all)
102. Joe DiMaggio has a career batting average of 0.325. What is the probability that he would get at least one hit in five official times at bat?

103. The life expectancy of a particular brand of light bulb is normally distributed with a mean of 1500 hours and a standard deviation of 75 hours.
   A. What is the probability that a light bulb will last less than 1410 hours?
   B. What is the probability that a light bulb will last between 1563 and 1648 hours?
   C. What is the probability that a light bulb will last between 1416 and 1677 hours?

104. A packing machine is set to fill a cardboard box with a mean average of 16.1 ounces of cereal. Suppose the amounts per box form a normal distribution with a standard deviation equal to 0.04 ounce.
   A. What percentage of the boxes will end up with at least 1 pound of cereal?
   B. Ten percent of the boxes will contain more than what number of ounces?
   C. Eighty percent of the boxes will contain more than what number of ounces?
   D. The middle 90% of the boxes will be between what two weights?

105. A company is considering implementing one of two quality control plans for monitoring the weights of automobile batteries that it manufactures. If the manufacturing process is working properly, the battery weights are approximately normally distributed with a specified mean and standard deviation.

   Quality control plan A calls for rejecting a battery as defective if its weight falls more than two standard deviations below a specified mean.

   Quality control plan B calls for rejecting a battery as defective if its weight falls more than 1.5 interquartile ranges below the lower quartile of the specified population.

Assuming the manufacturing process is under control.
   A. What proportion of batteries will be rejected by plan A?
   B. What is the probability that at least one of two randomly selected batteries will be rejected by plan A?
   C. What proportion of batteries will be rejected by plan B?
106. Suppose the probability that someone will make a major mistake on an income tax return is 0.23. One day, an IRS agent plans to audit as many returns as necessary until she finds one with a major mistake. Use simulation to estimate the probability that a major mistake will be found before the 5th return.

107. An outlier can have a striking effect on the correlation coefficient \( r \). For example, comment of the following 3 scatterplots:

108. An insurance company conducts a survey of 15 of its life insurance agents. The average number of minutes spent with each customer and the number of policies sold in a week are noted for each agent. Let \( X \) and \( Y \) represent the average number of minutes and number of sales, respectively, we have

\[
\begin{array}{c|cccccccccccccccc}
\end{array}
\]

A. Find the equation of the LSRL.

B. Interpret the slope in the context of the problem.

C. Interpret the y-intercept in the context of the problem.

D. State the correlation coefficient.

E. State and interpret the coefficient of determination in the context of the problem.

109. The equation of the LSRL for the points on the scatterplot above is \( \hat{y} = 1.3 + 0.73x \). What is the residual for the point (4,7)?
The dentists in a dental clinic would like to determine if there is a difference between the number of new cavities in people who eat an apple a day and people who eat less than one apple a week. They are going to conduct a study with 50 people in each group. Fifty clinic patients who report that they routinely eat an apple a day and 50 clinic patients who report they eat less than one apple a week will be identified. The dentists will examine their patients and their records to determine the number of new cavities the patients have had over the past two years. They will then compare the number of new cavities in the two groups.

A. Why is this an observational study and not an experiment?

B. Explain the concept of confounding in the context of this study, include an example of a possible confounding variable.

C. If the mean number of new cavities for those who ate an apple a day was statistically significantly smaller than the mean number of new cavities for those who ate less than one apple a week, could one conclude that the lower number of new cavities can be attributed to eating an apple a day? Explain.