1. Which of the following variables is continuous?

(A) Gender
(B) Outcomes on a roulette wheel
(C) Marital status
(D) Temperature

2. This graph represents the score distributions of four different classes, P, Q, R, and S. Scores varied from 0 to 40, with a passing score of 20. Which class had the greatest percentage of passing grades?

(A) P
(B) Q
(C) R
(D) S
3. In a probability distribution such as the one shown, which of the following must be true?

(A) The total area under the curve is 1.00.
(B) The greatest height of the curve is 1.00.
(C) The width of the distribution is 1.00.
(D) The height times the width of the curve is 1.00.

4. The data is most likely to come from a distribution that looks like

(A)  

(B)  

(C)  

(D)  

<table>
<thead>
<tr>
<th>Stem</th>
<th>Leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>11346679</td>
</tr>
<tr>
<td>3</td>
<td>12234</td>
</tr>
<tr>
<td>4</td>
<td>123</td>
</tr>
<tr>
<td>5</td>
<td>01</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
</tr>
</tbody>
</table>
5.

The median of the data shown is

(A) 20
(B) 30
(C) 32
(D) 45

6. What is the correct calculation formula for the sample variance shown?

\[ s^2 = \frac{\sum x^2 - \bar{x}^2}{n - 1} \]

(A) \( s^2 = \frac{n(\sum x)^2 - (\sum x)^2}{n(n-1)} \)

(B) \( s^2 = \frac{(\sum x)^2 - n(\sum x)^2}{n(n-1)} \)

(C) \( s^2 = \frac{(\sum x)^2 - (\sum x)^2}{n - 1} \)

(D) \( s^2 = \frac{(\sum x)^2 - (\sum x)^2}{n} \)
7. As a sample, five high school students were asked to record the number of hours they spent studying during a one-week period. The ranked data is: 0, 1, 2, 3, 3. Which of the following statements about the data is correct?

(A) The mean is 1.8.
(B) The range is 2.
(C) The midrange is 3.
(D) The mode is 2.

8. If a person invests $1,000 at 10% and $3,000 at 30%, the average return of the total investment is

(A) 15%
(B) 20%
(C) 25%
(D) 30%

9. If the mean of a set of 50 numbers is 20, the mean of a set of 30 numbers is 40, and the mean of a set of 20 numbers is 70, what is the mean of the 100 numbers?

(A) \( \frac{33}{3} \)
(B) 36
(C) 40
(D) 45

10. When performing a regression analysis and calculating a least squares regression line, \( \hat{y} = a + bx \), the statistic that measures the strength of the relationship between \( x \) and \( y \) is shown in which option?

(A) \( a \)
(B) \( b \)
(C) \( \frac{a+b}{2} \)
(D) \( r \)

11. Of the following sets of numbers, which has the greatest standard deviation?

(A) \{1,2,3,4,5\}
(B) \{1,3,5\}
(C) \{1,5\}
(D) \{41,42,43,44,45\}
12. Suppose \( x \) and \( y \) represent the number of hours studied for a test and the score on the test for a sample of 30 students, respectively. If the method of least squares gives \( a = 24 \) and \( b = 4 \) in \( y = bx + a \), what test score would be predicted for someone who studied for 3 hours?

(A) 24
(B) 36
(C) 76
(D) 99

13. Which of the following statements about the square of the correlation coefficient is true?

(A) It gives the percentage of the fluctuation in \( y \) accounted for by \( x \).
(B) It gives the value of the slope of the least squares regression line.
(C) It gives the value of the \( y \) intercept of the least squares regression line.
(D) It relates to the reason why \( y \) varies with \( x \).

14. In a sample of 2,000 drivers, 860 had their seat belts fastened. If one driver is chosen at random, the probability that the driver chosen was not using a seat belt is

(A) 0.14
(B) 0.43
(C) 0.57
(D) 0.86

15. Assume \( A \) and \( B \) are mutually exclusive events. If \( P(A) = 0.2 \) and \( P(B) = 0.6 \), then \( P(A \text{ or } B) \) equals

(A) 0.2.
(B) 0.4.
(C) 0.6.
(D) 0.8.

16. How many different 4-digit sequences can be formed using the digits 0, 1, 2, 3, 4, and 5 if repetition of digits is allowed?

(A) 64
(B) 256
(C) 1,296
(D) 4,096
17. A given intelligence test for 8-year-olds is normally distributed with a mean of \( \mu = 100 \) and a standard deviation of 15. An 8-year-old child is considered gifted if he or she scores above 130. In this situation, the percentage of 8-year-olds considered to be gifted is closest to

(A) 1.0%
(B) 2.0%
(C) 2.5%
(D) 5.0%

18. Carla took four midterm exams. Her raw scores, the class means, and the class SD’s are shown. If the scores on each exam are approximately normally distributed, on which exam did Carla do the best relative to the rest of the class?

(A) English
(B) History
(C) Mathematics
(D) Biology

<table>
<thead>
<tr>
<th>MIDTERM EXAM DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Score</td>
</tr>
<tr>
<td>English</td>
</tr>
<tr>
<td>History</td>
</tr>
<tr>
<td>Mathematics</td>
</tr>
<tr>
<td>Biology</td>
</tr>
</tbody>
</table>

19. Which of the following scatter diagrams best illustrates the data above?

(A)

(B)

(C)

(D)
20. A sample of 100 students at a university showed that 43 had completed one or more remedial college courses. A sample of 200 students at a junior college showed that 90 had completed one or more remedial college courses. Which of the following statements about the null hypothesis that there is no difference in proportion of students who complete remedial courses at a university or a junior college is true? (Assume \( \alpha = 0.05 \).)

(A) There is overwhelming evidence to accept the hypothesis.
(B) There is overwhelming evidence to reject the hypothesis.
(C) There is not enough evidence to accept the hypothesis.
(D) There is not enough evidence to reject the hypothesis.

21. When two sets of scores have symmetric distributions and the same means, which of the following must be true?

(A) Both sets have the same standard deviation.
(B) Both sets have the same median.
(C) Both sets have the same range.
(D) The sums of the scores of the two sets are the same.

22. A committee of 3 people can be selected from 12 people in how many different ways?

(A) 220
(B) 416
(C) 1,320
(D) 1,728

23. What percentage of the area under the standard normal curve is between \( z = -1 \) and \( z = 2 \)?

(A) 95%
(B) 81.5%
(C) 68%
(D) 47.5%

24. Let \( P(A) = \frac{1}{3} \), \( P(B) = \frac{1}{2} \), and \( P(A \cup B) = \frac{2}{5} \). Which of the following statements about \( A \) and \( B \) is true?

(A) \( A \) and \( B \) are mutually exclusive
(B) \( A \) and \( B \) are independent
(C) \( P(A|B) = P(B|A) \)
(D) \( P(A \cap B) = \frac{13}{30} \)
25. The aspect of a normal distribution curve that represents probability is the

(A) area
(B) ordinate
(C) z-score
(D) standard deviation

26. Which of the above plots shows the strongest linear correlation?

(A) Plot A
(B) Plot B
(C) Plot C
(D) Plot D

27. If a die is rolled twice, what is the probability that each roll comes up 3 or greater?

(A) \( \frac{1}{9} \)
(B) \( \frac{1}{4} \)
(C) \( \frac{4}{9} \)
(D) \( \frac{9}{16} \)

28. The standard deviation of the sample data 6, 8, 9, 9, 8 is shown in

(A) \( \sqrt{1.5} \)
(B) 1.5
(C) \((1.5)^2\)
(D) 3
29. To ensure that there will be a 95% probability that the sample proportion is within 3% of the population proportion, the sample size \( n \) must be no less than

(A) 1,200  
(B) 1,100  
(C) 1,000  
(D) 900

30. Joe has three extra concert tickets to share with his eight friends. The number of ways he can distribute the tickets is

(A) 0.3!  
(B) 24  
(C) 56  
(D) 5!

31. A box contains 3 balls numbered 1, 1, and 7, respectively. A ball is drawn at random, its number is noted, and it is put back in the box. This process is performed 100 times. Which of the following is expected to be closest to the sum of the 100 numbers noted?

(A) 200  
(B) 300  
(C) 400  
(D) 500

32. A 100-question multiple-choice test has four alternative responses for each question. If a person picks one of the responses for each question at random, what is the person’s expected score?

(A) 4  
(B) 8  
(C) 20  
(D) 25

33. Test scores of 1,000 students were found to have a mean and a standard deviation of 500 and 100, respectively. If the distribution of scores is bell-shaped, the number of scores that fall in the interval 400 to 600 is approximately

(A) 680  
(B) 600  
(C) 500  
(D) 400
34. Which of the following distributions would have the smallest standard deviation?

(A)  
(B)  
(C)  
(D)  

35. A study of 150 managers at large manufacturing plant A showed that 50 had a salary exceeding $32,000. A study at large manufacturing plant B showed that 95 of 200 sampled managers had salaries exceeding $32,000. Which option shows correct estimated proportion of managers at plants A and B combined earning more than $32,000?

(A) \( \frac{50+95}{150+200} \)
(B) \( \frac{50(95)}{(150)(200)} \)
(C) \( \frac{145}{64,000} \)
(D) \( \frac{145}{350} \)

36. In testing hypothesis \( H_0 \) versus \( H_a \), it’s determined that a \( p \)-value is 0.032 at the 0.05 level of significance. In this situation one should

(A) reject \( H_0 \) at 3.2%
(B) not reject \( H_0 \) at 3.2%
(C) reject \( H_0 \) at 5%
(D) not reject \( H_0 \) at 5%
37. Len and Kim roll a die 60 times. If a 2, 3, or 5 comes up, Kim pays Len that many dollars, but if a 1, 4, or 6 comes up, Len pays Kim that many dollars. Len can expect to

(A) win $60
(B) win $10
(C) lose $10
(D) lose $60

38. A college student visits two coffee houses on campus: coffee house A and coffee house B. She goes to coffee house A 60% of the time and coffee house B 40% of the time. Regardless of the coffee house, she buys a café mocha on 50% of her visits. The next time she visits a coffee house, what is the probability that she goes to coffee house A and orders a café mocha?

(A) 0.3
(B) 0.4
(C) 0.5
(D) 0.6

39. A manufacturer has 1,000 light bulbs of which 90% are non-defective. If the manufacturer selects 5 at random, with replacement, what is the probability that all of the light bulbs selected are non-defective?

(A) \[ \frac{9}{1000} \times \frac{899}{999} \times \frac{898}{997} \times \frac{897}{996} \]
(B) \( \left( \frac{1}{10} \right)^5 \)
(C) \( \left( \frac{1}{2} \right)^5 \)
(D) \( \left( \frac{9}{10} \right)^5 \)
40. Which diagram best describes a correlation coefficient \( r = -0.8 \)?

(A)

(B)

(C)

(D)

41. A student is testing the claim that the mean time needed for a group of 49 third graders to complete a task is 40 minutes against the alternative that the mean time is longer than 40 minutes. The group had a mean time of 44 minutes with a standard deviation of 14 minutes. The smallest \( \alpha \) at which the null hypothesis could be rejected is

(A) 0.01
(B) 0.025
(C) 0.05
(D) 0.10
42. When taking a large sample from a population and applying the Central Limit Theorem, which of the following statements about the distribution of the population being sampled is true?

(A) It must be $t$.
(B) It must be $x^2$.
(C) It must be normal.
(D) It can be anything.

43. \{5, 9, 15, 21, 21, 49\}

Adding the value 20 to the set above would

(A) increase the value of its median
(B) increase the value of its mean
(C) decrease the value of its mode
(D) decrease the value of its range

44. In which of the following problems can the binomial formula be applied?

(A) A die is rolled 10 times. What is the probability of getting exactly two 1s?
(B) A die is rolled until a 6 appears. What is the probability of getting a 2 before a 6 appears?
(C) Four draws are randomly made from a deck of cards, without replacement. What is the probability of drawing two hearts?
(D) A box contains 5 red marbles and 4 green ones. Three draws are made without replacement. What is the probability of getting 3 red marbles?

45. If a die is to be rolled 6,000 times, which of the following is the best answer to the question: "How many 2s will be rolled"?

(A) The number of 2s is exactly 1,000.
(B) The number of 2s is very likely exactly 1,000, but there is a small probability that it will not be equal to 1,000.
(C) The number of 2s is likely to be different from 1,000, but the difference is most probably small compared with 6,000.
(D) It is quite likely that the number of 2s is greater than or equal to 1,080.
46. If all of the 16-ounce cans of chicken soup manufactured by a company have a mean of 16 ounces and a standard deviation of 0.5 ounces and the weights are normally distributed, what percentage of the cans can be expected to weigh between 15.5 ounces and 16 ounces?

(A) 5%
(B) 16%
(C) 32%
(D) 34%

47. \( H_0 : \mu = 10, \ H_a : \mu \neq 10 \)

When testing the hypothesis above, which of the following is necessary for using the \( t \) statistic when the \( n \) is small?

(A) The population from which the sample is drawn must be normal.
(B) The population from which the sample is drawn must be \( t \) distributed.
(C) The population standard deviation must be known.
(D) The sample mean must be close to 10.

48.

![Histogram](image)

Which of the following classes would contain the median for the histogram above?

(A) I
(B) II
(C) III
(D) IV
49. The weight limit on an elevator is 1,600 pounds. The elevator contains 10 people with mean weight of 150 pounds and median weight of 175 pounds. Which of the following statements about the load on the elevator is true?

(A) The load is under 1,600 lbs.
(B) The load is over 1,600 lbs.
(C) The load is exactly 1,600 lbs.
(D) The load cannot be determined from the information given.

50. There are 2,000 voters in a town. In an experiment of randomly selecting a voter to be interviewed, the event $A$ consists of being in favor of buying park land. The event $B$ consists of having lived in the town less than 10 years. The table above gives the numbers of voters in various categories. The probability $(A \cap B')$ is

(A) $\frac{1}{20}$
(B) $\frac{1}{10}$
(C) $\frac{7}{20}$
(D) $\frac{1}{2}$

51. A student wants to estimate the average income of families in a particular community with a 90% confidence interval. If the standard deviation is $900 and the sample size chosen is 81, the sample mean should be accurate within approximately

(A) $100$
(B) $165$
(C) $196$
(D) $900$
52. A biologist suspected that males aged 20 to 24 have a higher mean systolic blood pressure than females in the same age group. Independent random samples of males and females were selected. The data are as follows. (Assume \( \mu_A \) is mean for males and \( \mu_B \) is mean for females.)

\[
\begin{align*}
(A) & & \mu X_A - \mu X_B = \mu_A - \mu_B \quad \text{and} \quad \sigma X_A - \sigma X_B = \sqrt{\frac{13.9^2}{31} + \frac{12.1^2}{41}} \\
(B) & & \mu X_A - \mu X_B = \mu_A - \mu_B \quad \text{and} \quad \sigma X_A - \sigma X_B = \frac{13.9^2}{31} + \frac{12.1^2}{41} \\
(C) & & \mu X_A - \mu X_B = \mu_A - \mu_B \quad \text{and} \quad \sigma X_A - \sigma X_B = \frac{13.9}{\sqrt{31}} + \frac{12.1}{\sqrt{41}} \\
(D) & & \mu X_A - \mu X_B = \mu_A - \mu_B \quad \text{and} \quad \sigma X_A - \sigma X_B = \frac{13.9}{\sqrt{31}} \sqrt{12.1} \\
\end{align*}
\]

53. Which of the following statements is true for the scatter diagrams above?

(A) The correlation coefficient of diagram P is least.
(B) The correlation coefficient of diagram Q is greater than that of diagram P.
(C) The correlation coefficient of diagram Q is greatest.
(D) All three diagrams have the same correlation coefficient.

54. If \( P(A) = .6 \), \( P(A') \) equals

(A) 0.50
(B) 0.40
(C) 0.05
(D) 0.04
55. Assume a 90% confidence interval in replacement time for television sets in the random sample shown. What is the margin of error in estimating the population mean? (The critical value \( z_\alpha = \frac{0.645}{2} \) at 90% is 0.645.)

(A) 0.05  
(B) 0.1  
(C) 1.1  
(D) 1.5

56. According to the U.S. Bureau of Labor Statistics, the mean weekly income of workers is $487 with a standard deviation of $64. If 64 workers are selected at random, the probability that the mean weekly income of those workers is less than $495 is

(A) 0.67  
(B) 0.80  
(C) 0.84  
(D) 0.95

57. Which of these hypotheses formulations is correct?

(A) \( H_0 : \mu = 20 \)  
\( H_a : \mu \neq 20 \)
(B) \( H_0 : \mu > 20 \)  
\( H_a : \mu = 20 \)
(C) \( H_0 : \mu < 20 \)  
\( H_a : \mu \geq 20 \)
(D) \( H_0 : \mu \leq 20 \)  
\( H_a : \mu < 20 \)

58. If \( y = 3.5x - 10.2 \) is a \( y \)-on-\( x \) regression equation, and if \( x \) equals 6.0, then the predicted value of \( y \) is

(A) 31.2  
(B) 21.0  
(C) 13.7  
(D) 10.8
59. The probability that person A will vote in the next election is 0.40. The probability that person B will vote in the next election is 0.35. The probability that both will vote is 0.20. The probability that at least one of them will vote is

(A) 0.55
(B) 0.60
(C) 0.75
(D) 0.95

60. | x  | p(x) |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>.2</td>
</tr>
<tr>
<td>0</td>
<td>.5</td>
</tr>
<tr>
<td>1</td>
<td>.3</td>
</tr>
</tbody>
</table>

Given the probability distribution above, the expected value of \(x^2\) is

(A) 0.0
(B) 0.1
(C) 0.5
(D) 0.9

61. If \(s_x\) and \(s_y\) represent the SDs of the independent variable \(x\) and the dependent variable \(y\), respectively, and if \(r\) represents the correlation coefficient between \(x\) and \(y\), then the slope of the \(y\)-on-\(x\) regression equation is

(A) \(\frac{s_y}{s_x}\)
(B) \(\frac{s_y}{s_x} (r)\)
(C) \(\frac{s_x}{s_y} (r)\)
(D) \(\left(\frac{s_y}{s_x}\right) \left(\frac{1}{r}\right)\)
62. The probability of rejecting H₀ for a true null hypothesis is

(A) \( 1 - \beta \)
(B) \( \beta \)
(C) \( 1 - \alpha \)
(D) \( \alpha \)

63. If a correlation is perfect, the regression line of \( x \)-on-\( y \) is

(A) parallel to the \( y \)-on-\( x \) regression line
(B) intersecting the \( y \)-on-\( x \) regression line
(C) coincident with the \( y \)-on-\( x \) regression line
(D) perpendicular to the \( y \)-on-\( x \) regression line

64. For the equation of the line, \( y = a + bx \), \( b \) represents the

(A) \( y \)-intercept of the line
(B) \( x \)-intercept of the line
(C) slope of the line
(D) value of \( y \) when \( x = 0 \)

65. Which of the following statements must be true about a frequency distribution whose histogram is symmetric about the median?

(A) The median is approximately equal to the mean.
(B) The median is approximately equal to the mode.
(C) The distribution is approximately bell-shaped.
(D) The distribution is two-tailed.

66. A game of chance has a probability of 0.30 of paying off $400 and a probability of 0.70 of paying off $200. There’s a charge for playing the game. What would be a fair price (break-even point) for playing the game?

(A) $225
(B) $260
(C) $310
(D) $350
67. If the value of the linear correlation coefficient \( r = -0.83 \), the \( r \)-value is

(A) weak and direct
(B) strong and direct
(C) weak and inverse
(D) strong and inverse

68. A coin is tossed 1,000 times. The probability that it lands on heads at least once is

(A) \( 0.001^{1000} \)
(B) \( 0.5^{1000} \)
(C) \( 1 - 0.5^{1000} \)
(D) \( 1 - 0.001^{1000} \)

69. Given the following information:

<table>
<thead>
<tr>
<th>( x )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(x) )</td>
<td>0.2</td>
<td>0.3</td>
<td>0.1</td>
<td>0.15</td>
<td>0.25</td>
</tr>
</tbody>
</table>

\[ \mu_x = 2.95 \quad \sigma_x = \sqrt{2.247} \]

Which of the following options is the mean of \( \bar{x}(\mu_x) \) and the standard deviation of \( x(\sigma_x) \)?

(Note: Take a random sample of size 3 from the probability distribution shown.)

(A) \( \mu_x = 2.95, \sigma_x = \sqrt{\frac{2.247}{3}} \)
(B) \( \mu_x = 2.95, \sigma_x = \frac{2.247}{3} \)
(C) \( \mu_x = 2.95, \sigma_x = \sqrt{2.247} \)
(D) \( \mu_x = \frac{2.95}{3}, \sigma_x = \sqrt{\frac{2.247}{3}} \)
70. For the observed values 3, 8, 8, 10, 12, and 18, the number of degrees of freedom associated with their mean is

(A) 5
(B) 6
(C) 7
(D) 8

71. Given the following information:

\[ \hat{y} = \frac{1}{3} \bar{x} + b_0 \]

\( \bar{x} \) equals the sample mean of \( x \) values.
The least squares regression line for the data set passes through the point (3,3).

If \( \bar{y} \) equals the sample mean of \( y \) values, then \( \bar{y} \) equals

(A) \( \frac{1}{3} \bar{x} \)
(B) \( \frac{1}{3} \bar{x} + 1 \)
(C) \( \frac{1}{3} \bar{x} + 2 \)
(D) \( \bar{x} + 1 \)

72. 3, 5, 7, 8, 9

The midrange for the sample data given is

(A) 5
(B) 6
(C) 7
(D) 9
73. Suppose 30% of families own a computer. If a random sample of 10 families is selected, the probability that between 2 and 4 of the sample families, inclusive, own a computer is

(A) \((45)(.3^2)(.7^8) + (120)(.3^3)(.7^7) + (210)(.3^4)(.7^6)\)

(B) \(\binom{10}{2}(.3^2)(.7^8) + \binom{7}{3}(.3^3)(.7^7)\)

(C) \(P(-1 < Z < 1)\)

(D) \(P(-.5 < Z < .5)\)

74. As the chance of making a Type I error is increased, which of the following statements is true about the chance of making a Type II error?

(A) It is increased.

(B) It is decreased.

(C) It remains unchanged.

(D) It cannot be determined from the information given.

75. A box contains a large number of tickets with numbers on them. The mean of these numbers is 50 and their SD is 20. If \(n\) draws are made at random with replacement, then the standard error for the mean of these draws is

(A) \(20n\)

(B) \(20\sqrt{n}\)

(C) \(\frac{20}{n}\)

(D) \(\frac{20}{\sqrt{n}}\)

76. Using the normal approximation to the binomial to approximate the probability that \(x\) is at least 10, the area under the normal curve should be calculated from

(A) 9.5 to \(\infty\)

(B) 10.0 to \(\infty\)

(C) 10.5 to \(\infty\)

(D) 11.0 to \(\infty\)
77. In a random sample of 25, if the SD of the distribution is $k$, what is the approximate value of the standard error of the mean?

(A) $0.01k$
(B) $0.1k$
(C) $0.2k$
(D) $0.5k$

78. If the standard error of the mean is 5.0 and the mean equals 50.0, then the probability is approximately 0.95 that the population mean is between

(A) 20 and 80
(B) 30 and 70
(C) 40 and 60
(D) 45 and 55

79. Which of the following is most likely to be the equation of the regression line for the data above?

(A) $\hat{y} = -2 + 1.5x$
(B) $\hat{y} = \frac{1}{2} + 0.6x$
(C) $\hat{y} = 2 - 2x$
(D) $\hat{y} = 1 + 7x$

80. What are the 95% confidence limits if a sample of 1,000 has a mean of 100 and a standard error of 0.5?

(A) 98.0 to 102.0
(B) 98.5 to 101.5
(C) 99.0 to 101.0
(D) 99.5 to 100.5
81. Values of $x^2$ (chi-square) are always

(A) between $-1$ and $+1$
(B) between $-3$ and $+3$
(C) negative
(D) positive

82. A student is testing the claim that the standard deviation in the weight of 24-ounce cans of tomatoes is at most 0.3 ounces against the alternative that it is more than 0.3 ounces. A sample of 28 cans had a mean of 24.3 ounces with a standard deviation of 0.32 ounces. The proper test statistic would be calculated by which of the following options?

(A) Option A: $\frac{27\.32}{(.3)}$
(B) Option B: $\frac{27\.32^2}{(.3)^2}$
(C) Option C: $\frac{28\.3}{(.32)}$
(D) Option D: $\frac{28\.3^2}{(.32)^2}$

83. To use the normal approximation to the binomial, which of the following is true?

(A) $np < 5$ and $n(1-p) < 5$
(B) $np < 5$ and $n(1-p) \geq 5$
(C) $np \geq 5$ and $n(1-p) < 5$
(D) $np \geq 5$ and $n(1-p) \geq 5$

84. Which of the following is true of $x^2$ curves?

(A) They are symmetric.
(B) They have a long left-hand tail.
(C) They have a long right-hand tail.
(D) They have long tails both to the left and to the right.
85. If the number of terms in the sum for calculating $x^2$ in the goodness of fit test is $k$, then the number of degrees of freedom is

(A) $k - 1$
(B) $k$
(C) $k + 1$
(D) $k^2$

86. If $o$ represents observed frequencies and $e$ represents expected frequencies, then the calculated value of $x^2$ equals the sum of

(A) $\frac{(o - e)^2}{e}$
(B) $\frac{(o^2 - e^2)}{e}$
(C) $\frac{(o - e)}{e^2}$
(D) $\frac{o - e^2}{e}$

87. The null hypothesis concerning the difference between two population means has been rejected at the 1% level. Which of the following statements can be validly made about the difference existing between the two population means?

(A) There is no difference.
(B) There is a difference.
(C) The likelihood of no difference is very large.
(D) The likelihood of no difference is very small.
88. A die was tossed. Given that an even number occurred, the probability that it was a 4 is

(A) \( \frac{1}{4} \)

(B) \( \frac{1}{3} \)

(C) \( \frac{1}{2} \)

(D) \( \frac{3}{4} \)

89. According to a soccer coach, the mean height of all female high school soccer players is 60.0 inches. A random sample of 25 such players produced a mean height of 61.2 inches with a standard deviation of 2.0 inches. Assuming that the heights of all female soccer players are normally distributed, do a test at the 1% significance level to see if their mean height is greater than 60.0 inches.

Which of the following are the correct null and alternative hypotheses for the situation above?

(A) \( H_0 : \mu = 61.2 \) and \( H_a : \mu > 61.2 \)

(B) \( H_0 : \mu = 60.0 \) and \( H_a : \mu < 60.0 \)

(C) \( H_0 : \mu \leq 60.0 \) and \( H_a : \mu > 60.0 \)

(D) \( H_0 : \mu < 60.0 \) and \( H_a : \mu \geq 60.0 \)

90. If the critical value for \( t \) is 2.492 and the test statistic for \( t \) is 3.00, the correct conclusion is to

(A) reject \( H_o \) at 5%

(B) do not reject \( H_o \) at 5%

(C) reject \( H_o \) at 1%

(D) do not reject \( H_o \) at 1%

91. For the goodness of fit test, which of the following statements is true?

(A) If the calculated value of \( x^2 \) is "small," the null hypothesis is accepted.

(B) If the calculated value of \( x^2 \) is "large," the null hypothesis is accepted.

(C) If the calculated value of \( x^2 \) lies between two given finite values, the null hypothesis is accepted, otherwise, it is rejected.

(D) The total area under the \( x^2 \) curves depends on the number of degrees of freedom.
92. Which of the following equations would provide the least squares regression line for the data above?

(A) \( \hat{y} = -x \)
(B) \( \hat{y} = 0 \)
(C) \( \hat{y} = x \)
(D) \( \hat{y} = 1 + x \)

93. A box contains 10 red marbles and 15 green marbles. A marble is drawn randomly, its color is noted, and then it is returned to the box. The process is performed 100 times. Approximately how many red marbles might one expect to draw?

(A) 60
(B) 50
(C) 40
(D) 30

94. According to Chebyshev’s theorem, the probability that a random variable will take on values within 5 standard deviations of the mean is

(A) at most \( \frac{4}{5} \)
(B) at least \( \frac{4}{5} \)
(C) at most \( \frac{24}{25} \)
(D) at least \( \frac{24}{25} \)
95. If there is a 5% chance of making a Type I error, the

(A) confidence interval was calculated at the 5% confidence rate
(B) confidence interval was calculated at the 90% confidence rate
(C) confidence interval was calculated at the 95% confidence rate
(D) none of the above

96. Which of the following is true if the regression line equation for a set of ordered points is $\hat{y} = -2 + 1.5x$?

(A) Low values of $x$ are associated with low values of $y$.
(B) The $y$-intercept is $-2$.
(C) Given $x = 8$, $\hat{y} = 10$.
(D) All of the above

97. A pile of 24 cards contains two different kinds of cards: some with a picture of a star and some with a picture of a moon. A card is drawn randomly, its picture is noted, and then it is returned to the deck. If the process is performed 80 times and it is expected that 30 of the cards drawn will be stars, how many of the cards in the pile must be stars?

(A) 8
(B) 9
(C) 12
(D) 15

98. Which of the following could NOT be the value of a chi-square?

(A) $-1$
(B) 0
(C) 0.652
(D) 11
1. **The correct answer is D.** A continuous variable is one for which, within the limits the variable ranges, any value is possible. For example, the variable "time to run a race" is continuous since it could take 5 minutes, 5.25 minutes, etc. to finish the race. The variable "number of students in a class" is not a continuous variable since it isn't possible to have 10.5 students in a class. Continuous variables can have an infinite number of different values between two given points. The only choice that is a continuous variable is temperature, so choice D is correct.

2. **The correct answer is D.** The area under any part of the curve is the proportion of the population in the designated region. For example, if 20% of a population has an income level between $150,000 and $250,000, 20% of the area of the histogram will be between $150,000 and $250,000. From these graphs, you can see that the passing grade is 20. The class with the highest percentage of passing grades will be the one with the greatest area between 20 and 40. This is shown in graph S, so choice D is correct.

3. **The correct answer is A.** A probability distribution illustrates the proportion or percentage of the population in each designated region. The total area of the curve represents 100% of the population or 1, so choice A is correct.

4. **The correct answer is B.** A stem-and-leaf plot resembles a histogram on its side. Rotating it 90 degrees gives an approximation of what the curve representing the distribution looks like. The peak in data corresponds with the 20s cohort, followed by the 30s and 40s. The graph that best represents this is graph B, so choice B is correct.

5. **The correct answer is C.** The median of a set of data is the middle value in a set of numerically arranged data. If you count the leaves, you see that there are 23 values total in this data set, so the middle value is the 12th value. The 12th leaf is a 2 that corresponds to a stem of 3, so the median is 32.

6. **The correct answer is A.** The raw score method of calculating the sample variance of a set of data uses the formula:

   \[ S^2 = \frac{\sum x^2 - (\sum x)^2}{n} \]

   To get rid of the fraction in the numerator of this formula, multiply the formula by n/n, which results in:

   \[ S^2 = \frac{(n)\sum x^2 - (\sum x)^2}{n(n-1)} \]

   so choice A is correct.

7. **The correct answer is A.** The mean of a set of data is the sum of the terms divided by the number of terms. In this case, \( \frac{0 + 1 + 2 + 3 + 3}{5} = \frac{9}{5} = 1.8 \), so choice A is correct. The range of a set of data is the
maximum value minus the minimum value. In this case, $3 - 0 = 3$, so choice B is incorrect. The midrange of a set of data is the sum of the maximum and minimum values divided by 2. In this case, $(0 + 3) / 2 = 1.5$, so choice C is incorrect. The mode of a set of data is the term that appears most often. In this case, 3, so choice D is incorrect.

8. The correct answer is C. The average return of the total investment is equal to the total return divided by the total investment. An investment of $1,000 at 10% yields a return of $100. An investment of $3,000 at 30% yields a return of $900. The average return is

$$\frac{100 + 900}{1,000 + 3,000} = \frac{1,000}{4,000} = 0.25, \text{ or } 25\%,$$

so choice C is correct.

9. The correct answer is B. The mean of the 100 numbers is the sum of all 100 numbers divided by 100. Since the mean times the number of terms gives you the sum of the terms, you can find the sum of each subset of numbers by multiplying each mean by the number of terms it represents: $20(50) + 40(30) + 70(20) = 1,000 + 1,200 + 1,400 = 3,600$. Divide by 100 to get the mean of the 100 numbers: $\frac{3,600}{100} = 36$, so choice B is correct.

10. The correct answer is D. The correlation coefficient, $r$, quantifies the relationship between $x$ and $y$. It defines the degree to which data points adhere to an imaginary trend line passing through the points. When $y$ increases as $x$ increases, the correlation is positive. When $y$ decreases as $x$ increases, the correlation is negative.

11. The correct answer is C. The sample standard deviation is found as the square root of the sample variance, or $S = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$. Calculate the standard deviation for each set of data:

- Set A: $\{1,2,3,4,5\}$; SD = 1.58114
- Set B: $\{1,3,5\}$; SD = 2
- Set C: $\{1,5\}$; SD = 2.82843
- Set D: $\{41,42,43,44,45\}$; SD = 1.58114

Since choice C is greatest, it's correct.

12. The correct answer is B. To find the predicted score of someone who studied for 3 hours, plug in $x = 3$, $a = 24$, and $b = 4$: score $= bx + a = 4(3) + 24 = 12 + 24 = 36$, so choice B is correct.

13. The correct answer is A. The square of the correlation coefficient, $r^2$, is a useful value in linear regression. This value represents the fraction of the variation in one variable that may be explained by the other variable. Thus, if a correlation of 0.5 is observed between two variables (for example, height and
weight), then a linear regression model attempting to explain either variable in terms of the other variable will account for 25% of the variability in the data, so choice A is correct.

14. The correct answer is C. Since 860 drivers had their seat belts fastened, $2,000 - 860 = 1,140$ did not. The probability that a driver chosen at random was not using a seat belt is $\frac{1,140}{2,000} = 0.57$.

15. The correct answer is D. Events A and B are said to be mutually exclusive if it's impossible for them to occur simultaneously. In such cases, the expression for the union of these two events reduces to the following, since the probability of the intersection of these events is defined as zero: $P(A \text{ or } B) = P(A) + P(B)$. In this case, $0.2 + 0.6 = 0.8$.

16. The correct answer is C. This is a permutation with repetition. When you have $n$ things to choose from, you have $n$ choices each time. When choosing $r$ of them, the number of permutations is $n^r$. In this case, there are a total of 6 digits to choose from, so $n = 6$. You're asked to find the number of 4-digit sequences, so $r = 4$. There are a total of $6^4 = (6)(6)(6)(6) = 1,296$ different sequences possible.

17. The correct answer is C. The mean score is 100 and the standard deviation is 15. Since an 8-year-old is considered gifted with a score above 130, scores that are more than two standard deviations greater than the mean are considered gifted. This is true since gifted scores are those that are greater than $100 + 2(15) = 130$. This corresponds to the regions on the graph more than two standard deviations above the mean.

The percentage of 8-year-olds that fall in this category is $1.7% + 0.5% + 0.1% = 2.3\%$, which is closest to 2.5%, so choice C is correct.

18. The correct answer is C. On her English exam, Carla scored $75 - 75 = 0$ points above the class mean. This is equal to $\frac{0}{5} = 0$ SD above the mean. On her history exam, Carla scored $82 - 60 = 22$ points above the class mean. This is equal to $\frac{22}{15} = \frac{7}{5}$ SD above the mean. On her mathematics exam, Carla scored $68 - 52 =$
16 points above the class mean. This is equal to \( \frac{16}{8} = 2 \) SD above the mean. On her biology exam, Carla scored \( 90 - 80 = 10 \) points above the class mean. This is equal to \( \frac{10}{6} = \frac{5}{3} \) SD above the class mean. Since she scored 2 SD above the class mean on her mathematics exam, this was Carla's best performance relative to the rest of the class.

19. **The correct answer is B.** According to the data, the greatest \( x \) value is 3. This \( x \) value of 3 corresponds to a negative \( y \) value, -2. The only scatter diagram on which the greatest \( x \) value has a negative \( y \) value is Diagram B, so choice B is correct.

20. **The correct answer is D.** The proportion of students at a university that completed remedial courses is \( \frac{43}{100} = 0.43 \). The proportion of students as a junior college that completed remedial courses is \( \frac{90}{200} = 0.45 \). The null hypothesis states that there is no difference in proportion of students who complete remedial courses at a university or a junior college, or \( u_1 - u_2 = 0 \). In this case \( 0.45 - 0.43 = 0.02 \). Since 0.02 < 0.05, there isn't enough evidence to reject the hypothesis.

21. **The correct answer is B.** If the distribution is symmetric, then the mean equals the median and there is no skew, so choice B is correct.

22. **The correct answer is A.** This is a combinations question, asking for the combination of 12 objects taken 3 at a time. The formula for this is

\[
\binom{n}{r} = \frac{n!}{r!(n-r)!},
\]

where \( n \) is the number of objects and \( r \) is the number taken at a time:

\[
_{12}\text{C}_3 = \frac{12!}{3!(12-3)!} = \frac{(12)(11)(10)}{(3)(2)(1)} = (2)(11)(10) = 222
\]

23. **The correct answer is B.** The \( z \)-score always reflects the number of standard deviations above or below the mean of a particular score. The question is asking for the percentage of the normal curve between -1 and 2 on the graph below:
15.0% + 19.1% + 19.1% + 15.0% + 9.2% + 4.4% = 81.8%, which is closest to 81.5%, so choice B is correct.

24. The correct answer is D. \( P(A \cup B) = P(A) + P(B) - P(A \cap B) \). Since \( \frac{1}{3} = \frac{1}{2} - \frac{13}{20} = \frac{2}{5} \), choice D is correct. If A and B were mutually exclusive, \( P(A) + P(B) = P(A \cup B) \). Since \( \frac{1}{3} + \frac{1}{2} \neq \frac{2}{5} \), A and B aren’t mutually exclusive and choice A is incorrect. Since \( P(A | B) = \frac{P(A \cap B)}{P(B)} \) and \( P(B | A) = \frac{P(B \cap A)}{P(A)} \), choice C is incorrect.

25. The correct answer is A. The area between the normal curve and the horizontal axis is 1, so the area of any interval under it depicts the probability that a value selected at random will fall in that interval. The ordinate is the height of the curve, which reaches its maximum at \( z = 0 \). The \( z \)-score reflects the number of standard deviations above or below the mean of a particular score. The standard deviation is a measure of how spread out a distribution is, and on the curve, it depicts how far above or below the mean a value falls.

26. The correct answer is B. Graph B shows the strongest linear correlation, since as \( x \) increases, \( y \) tends to decrease linearly. Graph D depicts a nonlinear correlation, since as \( x \) increases, \( y \) decreases at first and then changes direction and increases. Graphs A and C depict no correlation as the two variables seem to have no impact on each other.

27. The correct choice is C. Rolling a 3 or greater means rolling a 3, 4, 5, or 6. This is 4 of the possible 6 outcomes on the roll of a die, so the probability of doing so is \( \frac{4}{6} = \frac{2}{3} \). This is a case of independent probability, since the result of the previous roll has no bearing on subsequent rolls. The probability of rolling a 3 or greater on two rolls is \( \frac{2}{3} \times \frac{2}{3} = \frac{4}{9} \), so choice C is correct.

28. The correct answer is A. The standard deviation is the square root of the variance. The sample has five values, whose mean is \( \frac{6+8+9+9+8}{5} = \frac{40}{5} = 8 \). To calculate the sample variance, first compute the difference of each data point from the mean and square the result:
\[ \begin{align*}
6 - 8 &= -2^2 = 4 \\
8 - 8 &= 0^2 = 0 \\
9 - 8 &= 1^2 = 1 \\
9 - 8 &= 1^2 = 1 \\
8 - 8 &= 0^2 = 0 \\
\end{align*} \]

Now divide the sum of these values by \( n - 1 \):

\[
\frac{4 + 0 + 1 + 1 + 0}{5 - 1} = \frac{6}{4} = 1.5
\]

The standard deviation is the square root of the variance or \( \sqrt{1.5} \), so choice A is correct.

**29. The correct answer is B.** In planning a study, we want to collect enough data to estimate population proportion \( p \) with adequate precision. Use the formula

\[
\frac{z^2 \sigma pq}{d^2}
\]

For 95% confidence \( z = 1.96 \).

Let \( d \) represent the margin of error. When we have no reasonable guess for \( p \), we use \( p = 0.50 \) to provide more than enough data. Since \( p = 0.50 \),

\[
q = 1 - p \\
= 1 - 0.50 \\
= 0.50
\]

To ensure that the sample proportion is within 3% of the population proportion, the sample size \( n \) must be no less than:

\[
n = \frac{(1.96^2)(0.50)(0.50)}{0.03^2}
\]

\[
= \frac{0.9604}{0.0009}
\]

\[
= 1,067.1111.
\]
Round up to ensure adequate precision, and this is closest to 1,100, so choice B is correct.

30. The correct answer is C. This is a combinations question, asking for the combination of 8 objects taken 3 at a time. The formula for this is

\[
\binom{n}{r} = \frac{n!}{r!(n-r)!}
\]

where \(n\) is the number of objects and \(r\) is the number taken at a time:

\[
\binom{8}{3} = \frac{8!}{3!(8-3)!} = \frac{8!}{3!5!} = \frac{8 \times 7 \times 6}{3 \times 2 \times 1} = 56
\]

31. The correct answer is B. Since there are a total of 3 balls in the box, the probability of selecting each one is \(\frac{1}{3}\). Of the 100 draws, each ball can be expected to be drawn \(\frac{1}{3} \times 100 = 33\) times. The sum of the 100 draws is likely to be:

\[
33(1) + 33(1) + 33(7) = 66 + 231 = 297
\]

Since this is closest to 300, choice B is correct.

32. The correct answer is D. Each question on the test has four choices, one of which is correct. With random selection, a person has a \(\frac{1}{4}\) chance of choosing the right answer. Since there are 100 questions in total on the test, a person could expect to answer \(100 \times \frac{1}{4} = 25\) correctly.

33. The correct answer is A. The mean score is 500 and the SD is 100, so scores in the interval 400 to 600 are those within plus or minus 1 SD. By the empirical rule (or 68-95-99.7 rule) for bell-shaped or normal distributions, 68% of the curve lies within one standard deviation of the mean. Since there are 1,000 students, this represents approximately \((0.68)(1,000) = 680\).
34. **The correct answer is B.** Spread addresses the extent to which data vary around the center of the distribution. When all values in the data set are the same, there's no spread and standard deviation \( s = 0 \). In all other instances, the standard deviation is a positive value. When the standard deviation is large, there's a lot of spread in the data. When the standard deviation is small there's not a lot of spread in the data. The distribution with the smallest standard deviation will show the least amount of spread, with the data most closely clustered around the center of the distribution. This is shown in distribution B, so choice B is correct.

35. **The correct answer is D.** To find the proportion of managers at plants A and B combined earning more than $32,000, divide the sum of the employees earning more than $32,000 at each plant by the sum of all the employees at each plant:

\[
\frac{50 + 95}{150 + 200} = \frac{145}{350},
\]

so choice D is correct.

36. **The correct answer is C.** In hypothesis testing, the probability value is compared with the chosen significance level. If the probability is less than or equal to the significance level, then the null hypothesis is rejected. If the probability is greater than the significance level, then the null hypothesis is not rejected. Since 0.032 < 0.05, the null hypothesis is rejected at the chosen significance level of 0.05 or 5%, so choice C is correct.

37. **The correct answer is C.** Since there are six values on a die, each number has a \( \frac{1}{6} \) chance of coming up on a roll. Since there are a total of 60 rolls, the expectation is that each number will be rolled \( \frac{1}{6} \times (60) = 10 \) times. Consider the financial implications for Len with each number:

2: Len wins \((10)(\$2) = \$20\)
3: Len wins \((10)(\$3) = \$30\)
5: Len wins \((10)(\$5) = \$50\)
1: Len loses \((10)(\$1) = -\$10\)
4: Len loses \((10)(\$4) = -\$40\)
6: Len loses \((10)(\$6) = -\$60\)

Len wins a total of \$100, but loses a total of \$110, for a net loss of \$10, so choice C is correct.

38. **The correct answer is A.** There is a 0.6 chance that the student will visit coffee house A and a 0.5 chance that, once there, she will order a café mocha. The probability that she will go to coffee house A and order a café mocha is \( (0.6)(0.5) = 0.3 \), so choice A is correct.

39. **The correct answer is D.** Since 90% of the 1,000 bulbs are non-defective, the probability of selecting a non-defective bulb at random is \( \frac{90}{100} = \frac{9}{10} \). You're asked to find the probability of selecting 5 non-defective bulbs at random, with replacement. With replacement means that each time there are a total of 1,000 bulbs,
900 of which are non-defective, so each of the five times the probability is \( \frac{9}{10} \). That makes the probability of selecting 5 bulbs at random that are non-defective:

\[
\frac{9}{10} \times \frac{9}{10} \times \frac{9}{10} \times \frac{9}{10} = \left( \frac{9}{10} \right)^5.
\]

40. **The correct answer is A.** In a graph with a negative linear correlation, as \( x \) increases, \( y \) decreases linearly. The diagram that best displays this is diagram A. In diagram B, there appears to be little or no linear correlation. In diagram C, there is a nonlinear correlation, since as \( x \) increases, \( y \) decreases, but then changes direction and increases. In diagram D, there is a positive linear correlation, since as \( x \) increases, \( y \) increases linearly.

41. **The correct answer is B.** In testing the null hypothesis that the population mean is equal to a specified value \( \mu_0 \), one uses the statistic:

\[
t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}
\]

where \( s \) is the sample standard deviation and \( n \) is the sample size. In this case, the population mean is 40, the sample mean is 44, \( s = 14 \) and \( n = 49 \):

\[
t = \frac{44 - 40}{14} \times \frac{14}{\sqrt{49}}
= \frac{4}{14} \times \frac{14}{7}
= \frac{4}{2}
= 2
\]

If you look up \( t = 2 \) in a \( t \)-table, you find that it's associated with a value of 0.025, so choice B is correct.

42. **The correct answer is D.** The Central Limit Theorem states that given a distribution with a mean \( \mu \) and variance \( \sigma^2 \), the sampling distribution of the mean approaches a normal distribution with a mean (\( \mu \)) and a variance \( \frac{\sigma^2}{N} \) as \( N \), the sample size increases. No matter what the shape of the original distribution, the sampling distribution of the mean approaches a normal distribution, so choice D is correct.
43. **The correct answer is A.** The original set has 6 values, with a median of \( \frac{15 + 21}{2} = 18 \), a mean of \( \frac{120}{6} = 20 \), a mode of 21, and a range of \( 49 - 5 = 44 \). Adding a seventh value, specifically 20, to the set will have the following effect: median = 20; mean = \( \frac{140}{7} = 20 \); mode = 21; range \( 49 - 5 = 44 \). Since only the median will change, increasing from 18 to 20, choice A is correct.

44. **The correct answer is A.** Rolling a die 10 times to determine the probability of getting exactly two 1s lends itself to the binomial formula. This is because it has \( n \) repeated trials (10), each of which has only two possible outcomes, a success (1) or a failure (not 1). The probability of success is the same on every trial, \( \frac{1}{6} \). The trials are independent since the previous roll has no effect on the subsequent rolls.

45. **The correct answer is C.** The theoretical probability of an event is determined by noting all the possible outcomes and determining the likely outcome. The formula to use is number of favorable outcomes divided by total number of outcomes. As the number of trials keeps increasing, the experimental probability tends toward the theoretical probability. The theoretical probability of rolling a 2 is \( \frac{1}{6} \) on each roll. If the die is rolled 6,000 times you can expect a 2 to be rolled \( \frac{1}{6} \times 6,000 = 1,000 \) times. While this may not be the exact number of times a 2 would actually come up in an experiment, the difference would not be significant, so choice C is correct.

46. **The correct answer is D.** The can of soup has a mean weight of 16 ounces and an SD of 0.5 ounces. By the empirical rule (or 68-95-99.7 rule) for bell-shaped or normal distributions, 68% of the curve lies within one standard deviation of the mean. That means that 68% of the cans are between \( 16 - 0.5 = 15.5 \) ounces and \( 16 + 0.5 = 16.5 \) ounces. The cans that are between 15.5 and 16 ounces represent half of that, or \( \frac{68\%}{2} = 34\% \).

47. **The correct answer is A.** A \( t \)-test is any statistical hypothesis test in which the test statistic follows a Student's \( t \) distribution if the null hypothesis is supported. It's most commonly applied when the test statistic would follow a normal distribution if the value of a scaling term in the test statistic were known, so choice A is correct.

48. **The correct answer is B.** In this frequency histogram, you can see that there are 10 values in I, 25 in II, 15 in III, 5 in IV, and 5 in V. This is a total of 60 values. The median is the middle value in a set, and in this case, the mean of the 30th and 31st since there is an even number of terms. Since 10 values fall in I, the 11th through 35th fall in II. The 30th and 31st terms fall in II, so choice B is correct.

49. **The correct answer is A.** There are 10 people in the elevator with a mean weight of 150 pounds. The mean times the number of terms is equal to the sum of the terms. The weight of the people in the elevator is \( (10)(150) = 1,500 \), so choice A is correct.
50. The correct answer is D. The question asks for the intersection of $A$ and $B'$. If you look at the table, you can see that the number of voters who fall box where $A$ meets $B'$ is 1,000. Since there is a total of 2,000 voters, this represents a probability of \[ \frac{1,000}{2,000} = \frac{1}{2}. \]

51. The correct answer is B. The critical $z$-score for a 90% confidence level is 1.645. Multiply it by the standard error of the mean to solve this question:

\[
\text{Standard Error of the Mean} = \frac{SD}{\sqrt{n}} \\
= \frac{900}{\sqrt{81}} \\
= \frac{900}{9} \\
= 100
\]

Multiply this by 1.645 to get $164.50$, which is closest to $165$, so choice B is correct.

52. The correct answer is A. The two mean hypothesis test for large samples is

\[
\sqrt{\frac{s_A^2}{n_A} + \frac{s_b^2}{n_b}}
\]

In this case,

\[
\sqrt{\frac{13.9^2}{31} + \frac{12.1^2}{41}}
\]

so choice A is correct.

53. The correct answer is D. The correlation between two variables reflects the degree to which the variables are related. The correlation coefficient always takes a value between $-1$ and 1, with 1 or $-1$ indicating perfect correlation (all points would lie along a straight line in this case).

A positive correlation indicates a positive or direct association between the variables (increasing values in one variable correspond to increasing values in the other variable), while a negative or inverse correlation indicates a negative association between the variables (increasing values in one variable correspond to decreasing values in the other variable).

The points on each of the three scatter diagrams have the same relative relationship to each other, so they have the same correlation coefficient. If you plug the $x$ and $y$ values for each scatter diagram into the equation
\[ r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}. \]

you will find that all have a correlation coefficient of 0.62017.

54. The correct answer is B. By convention, the notation \( P(A) \) represents the probability of event A occurring and the notation \( P(A') \) represents the probability of event A not occurring. If the probability of an event occurring is \( P \), the probability of the event not occurring is \( 1 - P \). If \( P(A) = 0.6 \), then \( P(A') = 1 - 0.6 = 0.40 \).

55. The correct answer is C. The critical \( z \)-score for a 90% confidence level is 1.645. Multiply it by the standard error of the mean to solve this question:

\[
\text{standard error of the mean} = \frac{SD}{\sqrt{n}} = \frac{2}{\sqrt{9}} = \frac{2}{3} = 0.6667
\]

Multiply this by 1.645 to get 1.0967, which rounds up to 1.1, so choice C is correct.

56. The correct answer is C. Calculate the test statistic for this situation:

\[
\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{495 - 487}{64/\sqrt{64}} = \frac{8}{64} = \frac{8}{8} = 1
\]

Looking up \( z = 1 \) on the \( z \)-table you find that the probability is 0.8413, which is closest to 0.84, so choice C is correct.

57. The correct answer is C. The null hypothesis is denoted by \( H_0 \) and involves stating the belief that the mean of the population is \( \leq \), \( = \), or \( \geq \) a specific value. The alternative hypothesis is denoted by \( H_a \) and represents the opposite of the null hypothesis and holds true if the null hypothesis is found to be false. The alternative hypothesis always states the mean of the population is \( < \), \( \neq \), or \( > \) a specific value. Formula D follows these guidelines, so choice D is correct.

58. The correct answer is D. Plug \( x = 6.0 \) into the given equation to solve for the predicted value of \( y \):
3.5 10.2
y = 3.5x – 10.2
y = 3.5(6.0) – 10.2
y = 21 – 10.2
y = 10.8

59. The correct answer is A. The probability that at least one of them will vote is the sum of the probabilities that either will vote, minus the probability that both will vote: (0.4 + 0.35) – 0.2 = 0.75 – 0.2 = 0.55.

60. The correct answer is C. The chart gives three values for x, as well as the probability of each. There is a 0.2 probability that the value of x is –1, so a 0.2 probability that \( x^2 = (-1)^2 = 1 \). There is a 0.5 probability that the value of x is 0, so a 0.5 probability that \( x^2 = 0^2 = 0 \). There is a 0.3 probability that the value of x is 1, so a 0.3 probability that \( x^2 = 1^2 = 1 \). The expected value of \( x^2 \) is 0.2(1) + 0.5(0) + 0.3(1) = 0.2 + 0 + 0.3 = 0.5.

61. The correct answer is B. The equation of a regression line is \( \hat{y} = bx + a \), where b represents the slope of the line. The formula for \( b \) is \( b = r \frac{s_y}{s_x} \), where \( r \) is the correlation between \( x \) and \( y \), \( s_y \) is the standard deviation of \( y \), and \( s_x \) is the standard deviation of \( x \), so choice B is correct.

62. The correct answer is D. In a statistical hypothesis test, there are two types of incorrect conclusions that can be made. The hypothesis can be inappropriately rejected, which is a Type I error or a false positive. The level of significance, \( \alpha \), is the probability of making a Type I error, so choice D is correct.

63. The correct answer is C. When there is a reasonable amount of scatter, we can draw two different regression lines depending upon which variable we consider to be the most accurate. The first is a line of regression of \( y \) on \( x \), which can be used to estimate \( y \) given \( x \). The other is a line of regression of \( x \) on \( y \) used to estimate \( x \) given \( y \). If there is a perfect correlation between the data (in other words, if all the points lie on a straight line), then the two regression lines will be the same, so choice C is correct.

64. The correct answer is C. A linear regression line has an equation of the form \( \hat{y} = bx + a \), where \( x \) is the independent variable and \( y \) is the dependent variable. The slope of the line is \( b \), and \( a \) is the \( y \)-intercept (the value of \( y \) when \( x = 0 \)), so choice C is correct.

65. The correct answer is A. The key to this question is not assuming that just because a curve is symmetric that it's normal. By definition, a symmetric curve is symmetric about its center. For a curve to be symmetric, its mean and median must be at its center, so choice A is true. All of the other statements apply to normal curves, but not necessarily symmetric curves. A symmetric curve could be bimodal or even trimodal, making the other statements false.

66. The correct answer is B. In this game, there's a 30% chance that the payoff will be $400 and a 70% chance that the payoff will be $200. A fair price for playing the game would be 0.3($400) + 0.7($200) = $120 + $140 = $260.
67. The correct answer is D. The correlation coefficient always takes a value between –1 and 1, with 1 or –1 indicating perfect correlation (all points would lie along a straight line in this case). A positive correlation indicates a positive or direct association between the variables (increasing values in one variable correspond to increasing values in the other variable), while a negative or inverse correlation indicates a negative association between the variables (increasing values in one variable corresponds to decreasing values in the other variable). A correlation value close to 0 indicates that there's no association between the variables. The farther away the value of the correlation coefficient is from 0, the stronger it is. If the correlation coefficient is –0.83, it's negative and much closer to –1 than to 0, so it's strong and inverse.

68. The correct answer is C. The chance of getting no heads at all is the same as the probability of getting tails every time. Since the probability of getting a tail is \( \frac{1}{2} = 0.5 \), the probability of getting all tails is \( 0.5^{1000} \). The probability of getting at least one head (which is one head, or two heads, ... or all heads) is everything else, or \( 1 - 0.5^{1000} \), so choice C is correct.

69. The correct answer is A. To find the mean: \( 0.2 \times 1 + 0.3 \times 2 + 0.1 \times 3 + 0.15 \times 4 + 0.25 \times 5 = 2.95 \). Using a random sample of size 3,

\[
SEM = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{2.247}}{\sqrt{3}} = \frac{\sqrt[3]{2.247}}{3},
\]

so choice A is correct.

70. The correct answer is B. The number of degrees of freedom is the number of values in the final calculation of a statistic that are free to vary. There are said to be \( n - 1 \) degrees of freedom, so in the case of 7 observed variables, there are \( 7 - 1 = 6 \) degrees of freedom.

71. The correct answer is C. Plugging point \((3,3)\) into the given equation gives you

\[
\begin{align*}
3 &= \frac{1}{3} \times 3 + b_o \\
3 &= 1 + b_o \\
2 &= b_o,
\end{align*}
\]

so choice C is correct.
72. The correct answer is B. The midrange of a set is the mean of its minimum and maximum values. The midrange for this sample data is \(\frac{3 + 9}{2} = \frac{12}{2} = 6\), so choice B is correct.

73. The correct answer is A. The binomial distribution will calculate the probability of \(r\) successes in \(n\) trials where \(p\) is the probability of success for each trial and \(q\) is the probability of failure for each trial. In this case \(p = 0.3, q = 0.7, n = 10\), and \(r\) will be 2, 3, and 4:

\[
P[r, n] = \frac{n!}{(n-r)!}r!(p^r)(q^{n-r})
\]

\[
P[2, 10] = \frac{10!}{8!2!}(0.3^2)(0.7^8) = 45(0.3^2)(0.7^8)
\]

\[
P[3, 10] = \frac{10!}{7!3!}(0.3^3)(0.7^7) = 120(0.3^3)(0.7^7)
\]

\[
P[4, 10] = \frac{10!}{6!4!}(0.3^4)(0.7^6) = 210(0.3^4)(0.7^6)
\]

\[
P[2, 10] + P[3, 10] + P[4, 10] = 45(0.3^2)(0.7^8) + (0.3^3)(0.7^7) + 210(0.3^4)(0.7^6)
\]

74. The correct answer is B. In a statistical hypothesis test, there are two types of incorrect conclusions that can be made. The hypothesis can be inappropriately rejected, which is a Type I error or a false positive. Alternately the hypothesis can be inappropriately retained, which is a Type II error or false negative. Statistical tests involve a trade-off between the acceptable level of false positives and the acceptable level of false negatives. A threshold value can be varied to make the test more restrictive or more sensitive. The more restrictive test increases the risk of rejecting true positives and the more sensitive test increases the risk of accepting false positives. As the chance of making a Type I error is increased, the chance of making a Type II error is decreased.

75. The correct answer is D. A way to quantify how much any given sample mean varies from the population mean is by determining the standard deviation of the sampling distribution mean. This standard deviation is called the standard error of the mean or SEM. For a sample of \(n\) items, SEM = \(\frac{\sigma}{\sqrt{n}}\). In this case, that is \(\frac{20}{\sqrt{n}}\).
76. The correct answer is A. Because the normal distribution is continuous and the binomial is discrete, you must use the continuity correction factor. This involves compensating by subtracting 0.5 from the low interval endpoint (and if there had been a high interval endpoint, adding 0.5 to it). In this case, that means calculating the area under the normal curve starting at \(10 - 0.5 = 9.5\).

77. The correct answer is C. For a sample of \(n\) items, \(SEM = \frac{\sigma}{\sqrt{n}}\). In this case that is

\[
\frac{k}{\sqrt{25}} = \frac{k}{5} = \frac{1}{5k} = 0.2k.
\]

78. The correct answer is C. A \((1-\alpha)100\%\) confidence interval for \(\mu\) is found as: the mean \(\pm (z_{1-\alpha/2})(SEM)\). A 95\% confidence interval means a value for \(\alpha\) of 0.05, which means a \(z_{1-0.05/2} = z_{0.975} = 1.96\).

\[
50 \pm (1.96)(5) = 50 \pm 9.8 \text{ or } 40.2 - 59.8. \text{ The closest range to this is } 40-60, \text{ so choice C is correct.}
\]

79. The correct answer is B. A linear regression line has an equation of the form \(\hat{y} = bx + a\). The slope of the line is \(b\), and \(a\) is the \(y\)-intercept (the value of \(y\) when \(x = 0\)). The scatterplot shows data for which as \(x\) increases, \(y\) increases. That means that the regression line equation must have a positive slope. From the scatterplot, you can see that the regression line will hit the \(y\)-axis somewhere between \(x = 0\) and \(x = 1\). Since the \(y\)-intercept in choice B is 0.6, this is the correct equation.

80. The correct answer is C. A \((1-\alpha)100\%\) confidence interval for \(\mu\) is found as: the mean \(\pm (z_{1-\alpha/2})(SEM)\). A 95\% confidence interval means a value for \(\alpha\) of 0.05, which means a \(z_{1-0.05/2} = z_{0.975} = 1.96\).

\[
100 \pm (1.96)(0.5) = 100 \pm 0.98 \text{ or } 99.02 - 100.98. \text{ The closest range to this is } 99.0 \text{ to } 101.0, \text{ so choice C is correct.}
\]

81. The correct answer is D. When no relationship exists between the variables, chi-square equals zero. The stronger the relationship, the greater the value of chi-square. The chi-square is always positive and it provides no information about the direction of the relationship, so choice D is correct.

82. The correct answer is B. Given a random sample of size \(n\) from a normal population with a standard deviation equal to \(\sigma\), and a sample standard deviation equal to \(s\), we can define a statistic, called chi-square, using the following equation: \(x = \frac{(n-1)s^2}{\sigma^2}\). In this case, \(n = 28\), the sample standard deviation equals 0.32, and the population standard deviation is 0.3: \(x = \frac{(27)(0.32^2)}{0.3^2}\), so choice B is correct.

83. The correct answer is D. The binomial equation can be used to calculate the probability of \(r\) successes in \(n\) trials with \(p\) = the probability of a success for each trial and \(q\) = the probability of a failure. If \(np \geq 5\) and \(nq \geq 5\)
≥ 5, the normal distribution can be used to approximate the binomial. Since there are only two possible outcomes in a binomial trial, the probability of a failure is equivalent to 1 − p, so choice D is correct.

84. The correct answer is C. The chi-square curve is non-symmetrical and skewed to the right. The skew is less with more degrees of freedom. Since chi square curves skew right, they have long right tails, so choice C is correct.

85. The correct answer is A. The number of degrees of freedom is the number of values in the final calculation of a statistic that are free to vary. There are said to be n − 1 degrees of freedom, so in this case of k terms, there are k − 1 degrees of freedom.

86. The correct answer is A. The chi-square value is the sum of the squares of the differences between each observed value and the expected value, divided by the expected value. If o represents observed frequencies and e represents expected frequencies, the calculated value of chi-squared is equal to the sum of \( \frac{(o-e)^2}{e} \).

87. The correct answer is D. The null hypothesis is a claim of "no difference." Alpha (α) is a probability threshold for a decision. If \( P \leq \alpha \), we will reject the null hypothesis, meaning we reject the claim of no difference, so there is a small likelihood that there is no difference. If the null hypothesis concerning the difference between two population means has been rejected, it means that the likelihood of no difference between the two population means is very small, so choice D is correct.

88. The correct answer is B. On a die, three of the values are even and only one of them is a four, so the probability that a die will land on a 4 is \( \frac{1}{3} \).

89. The correct answer is C. The null hypothesis H₀ is a claim of no difference. The opposing hypothesis is the alternative hypothesis Hₐ. The alternative hypothesis is a claim of a difference in the population, and is the hypothesis the researcher often hopes to bolster. The question talks about a test at the 1% significance level to see if the soccer players' mean height is greater than 60.0 inches. In this case, the null hypothesis would be the mean is less than or equal to 60.0 inches and the alternative would be that the height is greater than 60.0, so choice C is correct.

90. The correct answer is C. A critical value is used in significance testing. It's the value that a test statistic must exceed for the null hypothesis to be rejected. Since the test statistic 3.00 exceeds the critical value 2.492, the null hypothesis should be rejected. Looking at the t-table, we see that the t-value is 2.492 at 24 degrees of freedom at the 1% confidence level.

91. The correct answer is A. When the data indicate that no relationship exists between the variables, the values of observed and expected frequencies must be identical. Likewise, the greater the relationship, the greater the difference between the observed and expected frequencies. When no relationship exists between the variables, chi-square equals zero. The greater the relationship, the greater the value of chi-square. Since a low chi-square value would indicate little to no difference between the expected and observed frequencies, choice A is correct.
92. The correct answer is B. The scatterplot shows 24 points with no correlation. That means there is no slope, and no y-intercept. The only equation that represents this situation is choice B.

93. The correct answer is C. There are 10 red and 15 green marbles for a total of 25 marbles. Since 10 are red, there is a $\frac{10}{25} = \frac{2}{5} = 40\%$ probability that a marble chosen at random will be red. Since the chosen marble is returned to the box each time, there is always a total of 25, 10 of which are always red. In 100 draws, 40% or 40 will be red, so choice C is correct.

94. The correct answer is D. According to Chebyshev's theorem, for any positive $k$, the proportion of the data that lies within $k$ standard deviations of the mean is at least $1 - \frac{1}{k^2}$. The probability that a random variable will take on values within 5 standard deviations of the mean is at least

$$1 - \left(\frac{1}{5^2}\right) = 1 - \frac{1}{25} = \frac{24}{25}.$$

95. The correct answer is C. A confidence level is the probability that the interval estimate will include the population parameter. The complement to the confidence level is the significance level, $\alpha$. It represents the probability that any given confidence interval will NOT include the true population mean. It also represents the probability of making a Type I error. So if $\alpha$ equals 5%, the confidence rate must be 95%, so choice C is correct.

96. The correct answer is D. A linear regression line has an equation of the form $\hat{y} = a + bx$. The slope of the line is $b$, and $a$ is the y-intercept (the value of $y$ when $x = 0$). Since $b$ equals positive 1.5, the line has a positive slope. That means that there is a direct relationship between $x$ and $y$; high values of $x$ are associated with high values of $y$, just as low values of $x$ are associated with low values of $y$. So choice A is true. The y-intercept is the $y$-value when $x = 0$ and the line crosses the $y$-axis. Since $b = -2$, choice B is true. The linear equation will predict a value for $y$ given a value of $x$. Plugging in $x = 8$, you find that the predicted value for $y$ is 10, and choice C is true. Since choices A, B, and C are all true, choice D is the correct answer.

97. The correct answer is B. Since the cards are replaced after each draw, on each draw there is the same probability of getting a star card. If it is expected that in 80 draws there will be 30 stars, that means it is expected that $\frac{30}{80}$, or $\frac{3}{8}$, of the draws will be stars. Since there are a total of 24 cards in the pile, 9 of them must be stars:

$$\frac{3}{8} \times 24 = 9,$$

and choice B is correct.

98. The correct choice is A. The chi-square is always positive. The stronger a relationship, the greater the
value of chi-square. When no relationship exists between variables, chi-square equals zero.