1. After calculating the sample size needed to estimate a population proportion to within 0.05, you have been told that the maximum allowable error must be reduced by half to just 0.025. If the original calculation led to a sample size of 1,000, the sample size will now have to be:
   a. 1,000  b. 2,000  c. 4,000  d. 8,000  e. 500

2. Coors Field in Denver Colorado is the home ballpark of the Colorado Rockies, a major league baseball team. The elevation of Coors Field is approximately 1 mile above sea level, much higher than any other ballpark in major league baseball. Because of its elevation, Coors Field has always been regarded as a home run friendly ballpark; a batted ball travels approximately 10% farther at Coors Field than at a ballpark at sea level. The Rockies were aware of this when Coors Field was built in 1995 so the field was designed to make it more difficult to hit a home run, but home runs still were hit at Coors Field at a much higher rate than at any other ballpark.

   In an attempt to compensate for the extra distance that batted balls travel, in 2002 the Rockies began humidifying the baseballs. They keep them in a room where the temperature is 70 degrees and the humidity is 50%, conditions that are similar to those at ballparks at sea level. The idea is that the humidity will make the baseballs slightly larger and softer, so they would not fly as far.

   Has humidifying the baseballs helped? Here is the data: between 2002 and 2008 the number of home runs hit in all games at Coors Field (Rockies' home games) by all teams is 1,380; the number of home runs hit in all Rockies' away games is 1,178. What is the approximate P-value of the hypothesis test, \( H_0 : \pi = 0.5 \), \( H_A : \pi > 0.5 \), where \( \pi \) is the proportion of all home runs hit in Rockies' home and away games that are hit at Coors Field?
   a. 0.05  b. 3.94  c. 0  d. .539  e. 1

3. Dewey, Lie, and Howe is a national company that sells cell phones. Company management is interested in estimating \( \pi \), the proportion of customer sales transactions in which the customer is convinced to purchase more expensive, but useless, cell phone upgrades. A random sample of customer sales transactions results in a 95% confidence interval of .73 ± .06 for \( \pi \). What choice below correctly describes this confidence interval?
   a. On 95% of days, between .67 and .79 of customer sales transactions result in purchases of more expensive but useless upgrades.
   b. 95% of the company's salespersons can convince customers to purchase more expensive but useless upgrades in between .67 and .79 of their transactions.
   c. The probability is .95 that the true proportion \( \pi \) of sales transactions that result in purchases of more expensive but useless upgrades is between .67 and .79.
   d. 95% of all random samples of customer sales transactions will show that .73 of customer sales transactions result in purchases of more expensive but useless upgrades.
   e. We are 95% confident that the interval (.67, .79) contains the true, but unknown, proportion \( \pi \) of all customer sales transactions that result in purchases of more expensive but useless upgrades.
4. A P-value is the probability:
   a. that the null hypothesis \( H_0 \) is true
   b. that we have made the wrong conclusion when doing the hypothesis test
   c. calculated under the assumption that the null hypothesis \( H_0 \) is true, that the test statistic would
      assume a value as or more extreme than the value stated in the null hypothesis
   d. calculated under the assumption that the null hypothesis \( H_0 \) is true, that the test statistic would
      assume a value as or more extreme than the observed value of the test statistic.
   e. that the null hypothesis \( H_0 \) is in fact false.

5. An online book seller is concerned about the timeliness of the delivery of their books. The VP of
   Operations and Marketing recently stated that she wanted the percentage of book orders delivered on
time to be at least 90%, and she wants to know if the company has succeeded. Let \( \pi \) be the proportion
   of book orders that are delivered on time. Choose the correct null and alternative hypotheses below.
   a. \( H_0 : \pi = .90, H_a : \pi < .90 \)
   b. \( H_0 : \pi = .90, H_a : \pi \neq .90 \)
   c. \( H_0 : \pi = .90, H_a : \pi > .90 \)
   d. \( H_0 : \pi > .90, H_a : \pi = .90 \)
   e. \( H_0 : \pi = .90, H_a : \pi < .90 \)

Use the following information for question 6-9
A consumer research organization poll asked consumers if they trusted "eco friendly" labels on
cleaning products. Out of 1000 adults surveyed, 498 respondents "yes." We would like to test if the
proportion \( \pi \) of consumers that trust these labels is more than .50. The null hypothesis is
\( H_0 : \pi = .50 \).

6. The correct alternative hypothesis \( H_A \) is
   a. \( H_A : \pi = .50 \)
   b. \( H_A : \pi \geq .50 \)
   c. \( H_A : \pi \geq .50 \)
   d. \( H_A : \pi > .50 \)
   e. \( H_A : \pi > .50 \)

7. The value of the test statistic \( z \) is
   a. \(-0.126\)
   b. \(-2.45\)
   c. \(0.498\)
   d. \(-1.89\)
   e. \(0.139\)

8. If the value of the test statistic above is denoted as \( z_0 \), then the correct calculation of the \( P-value \) is
   a. \( P(z < z_0) \)
   b. \( P(z > z_0) \)
   c. \(2P(z > |z_0|) \)
   d. \(1 - P(z > z_0) \)
   e. \(1 - P(z < z_0) \)

9. Suppose the \( P-value \) is .550. Then the correct conclusion is
   a. Since the \( P-value \) is greater than .05, do not reject the null hypothesis. There is insufficient
evidence to conclude that the proportion of consumers that trust “eco-friendly” labels differs from
\( .50 \)
   b. Since the \( P-value \) is greater than .05, reject the null hypothesis. There is sufficient evidence to
conclude that the proportion of consumers that trust “eco-friendly” labels is greater than .50
   c. We cannot conclude anything from the hypothesis test because there is a 0.550 chance that we have
made an error.
   d. Since there is only a 0.550 chance that \( H_0 : \pi = .50 \) is true, we have to reject \( H_0 \).
   e. Since 0.550 is the probability the alternative hypothesis \( H_A \) is true, we should reject \( H_0 \).

10. A copy machine dealer has data on the number \( x \) of copy machines at each of 89 customer locations
    and the number \( y \) of service calls in a month at each location. Summary calculations give \( \bar{x} = 8.4, \)
    \( s_x = 2.1, \bar{y} = 14.2, s_y = 3.8, \) and \( r = .86 \). What is the slope of the least squares regression line of
    number of service calls on number of copiers?
    a. \(0.86\)
    b. \(1.56\)
    c. \(0.48\)
    d. none of these
    e. can't tell from information given

11. In the setting of the previous problem, about what percent of the variation in number of service calls is
    explained by the linear relation between number of service calls and number of machines?
    a. 86%
    b. 93%
    c. 74%
    d. none of these
    e. can't tell from information given
12. Outdoor temperature influences natural gas consumption for the purpose of heating a house. The usual measure of the need for heating is heating degree days. The number of heating degree days for a particular day is the number of degrees the average temperature for that day is below 65°F, where the average temperature for a day is the mean of the high and low temperatures for that day. An average temperature of 20°F, for example, corresponds to 45 heating degree days. A homeowner interested in switching to solar heating panels collects the following data on her natural gas use for the months October through June, where $x$ is heating degree days per day for the month and $y$ is gas consumption per day in hundreds of cubic feet.

<table>
<thead>
<tr>
<th>Month</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>June</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>15.6</td>
<td>26.8</td>
<td>37.8</td>
<td>36.4</td>
<td>35.5</td>
<td>18.6</td>
<td>15.3</td>
<td>7.9</td>
<td>0</td>
</tr>
<tr>
<td>$y$</td>
<td>5.2</td>
<td>6.1</td>
<td>8.7</td>
<td>8.5</td>
<td>8.8</td>
<td>4.9</td>
<td>4.5</td>
<td>2.5</td>
<td>1.1</td>
</tr>
</tbody>
</table>

If $\sum (x_i - \bar{x})(y_i - \bar{y}) = 291.31$, $s_x = 13.42$, $s_y = 2.74$, calculate the correlation coefficient $r$ and interpret its value; draw a scatterplot of the data.

13. A sample of 44 cordless telephones had a sample mean working range of 142 feet and a sample standard deviation $s = 12$ feet. A 98% confidence interval for the mean working range $\mu$ is
   a. (139.02, 144.98)   b. (138.45, 145.55)   c. (137.79, 146.21)   d. (137.34, 146.66)   e. (137.63, 146.37)

14. If the level of confidence is changed from 90% to 95% and the sample size remains the same, then
   a. this means that the interval is less likely to include the population value
   b. this means that the interval is more likely to include the population value
   c. this means that the interval will necessarily be narrower
   d. this means that the probability of error has increased
   e. the 95% confidence interval will be 5% longer than the 90% confidence interval

15. A study reports that a 95% confidence interval for the mean SAT math score $\mu$ of Illinois high school seniors is 452 to 470. The correct statistical interpretation of this interval is that
   a. 95% of Illinois high school seniors have SAT math scores between 452 and 470
   b. there is a 95% chance that the true mean SAT math score $\mu$ of Illinois high school seniors is in the interval (452, 470).
   c. the probability is .95 that an individual Illinois high school senior will score between 452 and 470 on the SAT math test
   d. if we take many samples of Illinois high school senior SAT math scores and compute the sample mean $\bar{x}$ for each sample, 95% of the sample means will be in the interval (452, 470).
   e. we are 95% confident that the interval (452, 470) contains the true mean SAT math score $\mu$ of Illinois high school seniors.

16. Twenty-five weather stations measure rainfall at random locations in a state. In 2015 they recorded a mean rainfall of $\bar{x} = 12$ inches with a standard deviation $s = 1.2$ inches. Construct a 95% confidence interval for the mean rainfall $\mu$ in this state in 2015.
   a. (11.505, 12.495)   b. (11.530, 12.470)   c. (11.906, 12.094)   d. (11.571, 12.429)

Questions 17 and 18 refer to the following information:

A biologist has collected bivariate data where the independent ($x$) variable is the number of cricket chips per minute and the dependent ($y$) variable is the temperature in °F. The data can be summarized as follows:

$n = 7$, $\bar{x} = 17.429$, $s_x = 1.988$, $\bar{y} = 80.571$, $s_y = 8.696$, $r = .976$

17. The slope of the least squares line is approximately
   a. 6.13   b. 4.27   c. .234   d. .163   e. .976

18. What proportion of the variation in temperature ($y$) is explained by the linear relationship between the
number of cricket chirps ($x$) and temperature?

a. .024  b. .976  c. $\frac{1.988}{80.00}$  d. $(.976)^2$  e. $\frac{17.429}{80.00}$

19. In developing a confidence interval for a population mean $\mu$ from sample data, the confidence interval was 52.84 to 59.84. The population standard deviation was assumed to be 6.50, and a sample of 100 observations was used. The mean $\bar{x}$ of the sample was

a. 56.34  b. 62.96  c. 13.24  d. 66.15  e. 65.00

20. You are interested in purchasing a new car. One of the many points you wish to consider is the resale value of the car after 5 years. Since you are particularly interested in a certain foreign sedan, you decide to estimate the resale value of this car with a 99% confidence interval. You manage to obtain data on 17 recently resold 5-year-old foreign sedans of the same model. These 17 cars were resold at an average price of $12,610 with a standard deviation of $700. What is the 99% confidence interval for the true mean resale value of a 5-year-old car of this model?

a. $12,610 \pm 2.9208$  b. $12,610 \pm 2.8982$  c. $12,610 \pm 2.58$  d. $12,610 \pm 2.9208$  e. $12,610 \pm 2.9208$

Use the following information to answer questions 21 and 22.

A philosophy professor has found a correlation of 0.80 between the number of hours students study for his exams and their exam score. During the time he collected the data, students studied an average of 10 hours with a standard deviation of 2.5 hours, and scored an average of 82 points with a standard deviation of 7.5 points.

21. Calculate the slope of the least squares regression line where the x-variable is the number of hours students study for the professor's exams and the y-variable is their exam score.

a. 0.267  b. 0.80  c. 3.0  d. 2.40  e. 0.33

22. A particular student usually studies 5 hours for an exam. If this student studies an additional 2 hours for an exam, on average by how many points will the student's exam score change?

a. increase by 0.80  b. increase by 6.0  c. decrease by 6.0  d. increase by 4.8  e. decrease by 4.8

23. A lakeside restaurant uses a least squares line to predict the number of meals they will serve in a day (the y-variable) from the daily temperature (the x-variable). The correlation between the daily temperature and the number of meals they serve is $r = 0.40$. On a day when the temperature is 2 standard deviations above the mean, the number of meals they should plan on serving is ______ the mean?

a. equal to  b. 0.16 SD above  c. 0.4 SD above  d. 0.8 SD above  e. 2.0 SD above

24. An economist studying energy trends used data for the years 1995 to 2005 to construct a least squares line where the independent variable ($x$) is the price (in US dollars) of a barrel of crude oil and the dependent variable ($y$) is the price (in cents) of a gallon of gasoline. The least squares prediction line is $\hat{y} = 33.42 + 2.93x$. Choose the correct interpretation of the slope of the least squares line.

a. If the price of a barrel of crude oil increases by $2.93$, the price of a gallon of gasoline increases by 1¢.
b. If the price of a barrel of crude oil increases by $2.93$, the price of a gallon of gasoline increases by 33.42¢.c. If the price of a barrel of crude oil increases by $1$, the price of a gallon of gasoline increases by 2.93¢.
d. If the price of a barrel of crude oil increases by $1$, the price of a gallon of gasoline increases by 33.42¢.

25. A student wonders if people of similar heights tend to date each other. She measures herself and five other women in her dormitory. Then she measures the height of the next man each woman dates. The data are shown in the table below (heights are in inches).
The summary data is as follows:
\[ \bar{x} = 66, \ s_x = 2.10, \ \bar{y} = 70.17, \ s_y = 2.40, \ r = .913 \]
If women height is the \( x \)-variable and men height is the \( y \)-variable, the slope of the least squares line is:
\[ a. \ 0.9565 \quad b. \ 1.043 \quad c. \ -0.9565 \quad d. \ 0.913 \quad e. \ 1.167 \]

26. Medical research indicate that people with more education tend to live longer. Let \( x \) be a person’s years of education and \( y \) the person’s lifespan. The slope of the least squares line that predicts lifespan from years of education is 0.8 and the correlation between the variables \( x \) and \( y \) is 0.48. If a medical researcher wants to reverse the roles of the variables and predict years of education from lifespan, what is the slope of the line that predicts years of education from lifespan?
\[ a. \ 0.288 \quad b. \ \frac{1}{0.8} \quad c. \ -0.8 \quad d. \ -\frac{1}{0.8} \quad e. \ \frac{0.8}{0.48} \]

27. Private colleges and universities rely on money contributed by individuals and corporations for their operating expenses. Much of this money is invested in a fund called an endowment, and the college spends only the interest earned by the fund. A recent survey of eight private colleges in the United States resulted in the following summary statistics for their endowments (in millions of dollars): \( \bar{x} = 180.975 \) and \( s = 143.042 \). Calculate a 99% confidence interval for the mean endowment of all private colleges in the United States.
\[ a. \ 180.975 \pm 176.980 \quad b. \ 180.975 \pm 189.199 \quad c. \ 180.975 \pm 130.478 \quad d. \ 180.975 \pm 169.693 \]

28. The exercise habits of Russian male and female adults was compared. A random sample of Russian adults showed that 28% of 200 males exercise regularly compared to only 18% of 200 females who exercise regularly. Which formula calculates the 95% confidence interval for the difference in the proportions of Russian male and female adults who exercise regularly?
\[ a. \ (.28 - .18) \pm 1.96 \sqrt{\frac{(.28)(.72)}{200}} \quad b. \ (.28 - .18) \pm 1.96 \sqrt{\frac{(.28)(.72)}{200}} + \frac{(.23)(.77)}{200} \]
\[ c. \ (.28 - .18) \pm 1.96 \sqrt{\frac{(.23)(.77)}{400}} \quad d. \ (.28 - .18) \pm 1.96 \sqrt{\frac{(.28)(.72)}{200}} + \frac{(.18)(.82)}{200} \]
\[ e. \ (.28 - .18) \pm 1.96 \sqrt{\frac{(.23)(.77)}{400}} + \frac{(.28)(.72)}{200} \]

29. Is it a good idea to listen to music when studying for a big test? In a study conducted by some Statistics students, 62 people were randomly assigned to listen either to rap music, music by Mozart, or no music while attempting to memorize objects pictured on a page. They were then asked to list all the objects they could remember. The data are summarized below.

<table>
<thead>
<tr>
<th>Rap</th>
<th>Mozart</th>
<th>No Music</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_1 = 29 )</td>
<td>( n_2 = 20 )</td>
<td>( n_3 = 13 )</td>
</tr>
<tr>
<td>( \bar{x}_1 = 10.72 )</td>
<td>( \bar{x}_2 = 10.00 )</td>
<td>( 12.77 )</td>
</tr>
<tr>
<td>( s_1 = 3.99 )</td>
<td>( s_2 = 3.19 )</td>
<td>( 4.73 )</td>
</tr>
</tbody>
</table>

Calculate a 99% confidence interval for the difference \( \mu_1 - \mu_2 \) in mean memory scores for students who study while listening to rap music and students who study while listening to Mozart. (\( \mu_1 \) is the mean memory score for students who listen to rap music). Use \( \min(n_1 - 1, n_2 - 1) \) to estimate the degrees of freedom.
\[ a. \ (-1.933, 3.373) \quad b. \ (-2.222, 3.662) \quad c. \ (-0.839, 2.279) \quad d. \ (-2.044, 3.484) \]

Use the following to answer questions 30 and 31:
A major airline is concerned that the mean waiting time \( \mu \) for customers at their ticket counter may be exceeding their target mean of 190 seconds. To test \( H_0 : \mu = 190 \) vs. \( H_A : \mu > 190 \), the company selects a random sample of \( n = 50 \) customers and times them from when the customer first arrives at the checkout line until he or she is at the counter being served by the ticket agent. The mean time for this sample is \( \bar{x} = 202 \) seconds and the standard deviation is \( s = 38 \) seconds.

30. The value of the test statistic is
   a. \( z = 2.268 \)   b. \( t = 2.233 \)   c. \( z = .3158 \)   d. \( t = -2.268 \)   e. \( t = -2.233 \)

31. Lower and upper bounds on the \( P - value \) for this hypothesis test are:
   a. \( (0.02, 0.05) \)   b. \( (2.0096, 2.4049) \)   c. \( (0.01, 0.025) \)   d. \( (-2.4049, -2.0096) \)   e. \( (-0.025, -0.01) \)

32. If the hypothesis test \( H_0 : \mu = 10 \) vs \( H_A : \mu \neq 10 \) has test statistic value \( t = 2.4786 \) and a sample size \( n = 27 \), what is the \( P - value \)?
   a. .025   b. .05   c. .01   d. .02   e. .005

**ANSWERS**

1. c 2. c. \( \hat{p} = \frac{1380}{1389 + 1767} = .539; \) \( SD(\hat{p}) = \sqrt{\frac{(0.5)(0.5)}{2588}} = 0.0099; \) test statistic \( z = \frac{0.539 - 0.5}{0.0099} = 3.94; \)
   \( P - value = P(z > 3.94) = 0 \) (from \( z \) table) or \( P - value = P(z > 3.94) = 4.07E-05 \) (from Excel).
   Since the \( P - value \) is less than 0.05, we reject \( H_0 : p = 0.5 \) and conclude \( p > 0.5 \), that is, Coors Field is still a home run friendly ballpark even though the baseballs are humidified.

3. e 4. d 5. e 6. e 7. a 8. b 9. a

10. b. since \( b = \hat{r}\left(\frac{n}{2}\right) = .86\left(\frac{18}{2}\right) = 1.56. \) 11. c since \( r^2 = (.86)^2 = .74 \)

12. \( r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = .990. \) There is a strong positive linear relationship between heating degree days and gas consumption.

**SCATTERPLOT: Heat. Deg. Days (x), Gas Consump.(y)**

13. e. \( 142 \pm 2.4163 \frac{12}{\sqrt{44}} \); 14. b; 15. e; 16. a. \( 12 \pm 2.0639 \frac{1.2}{\sqrt{20}} \)

17. b. slope \( b_1 = .976 \frac{2.696}{1388} = 4.27; \)

18. d; 19. a the sample mean is the midpoint of the interval; 20. a

21. d, \( b_1 = r_{XY} = 0.80\frac{2.5}{2} = 2.4. \)

22. d, the slope 2.4 means that for each additional hour spent studying, the exam score will increase by 2.4 points, so if a student studies an additional 2 hours, the exam score will increase by \( 2 \times 2.4 = 4.8 \) points;
23. d, since the slope is $b_1 = r \frac{s_y}{s_x}$, for each increase in $x$ of one standard deviation $s_x$, $y$ will change by one $y$ standard deviation $s_y$ times $r$; so if $x$ is 2 standard deviations above $\mu$, $y$ will change by $2r = 2(0.4) = 0.8$ standard deviations, that is, $y$ will increase by 0.8 $y$ standard deviations.

24. c; 25. b, $b_1 = r \frac{s_y}{s_x} = 0.913 \cdot \frac{0.40}{0.31} = 1.043$; 26. a, $b_1 = r \frac{s_y}{s_x} \Rightarrow .8 = .48 \frac{s_y}{s_x} \Rightarrow \frac{s_y}{s_x} = \frac{.8}{.48}$; since the roles of the variables are reversed the new slope $b_1^* = .48 \frac{s_x}{s_y} = .48 \left( \frac{s_x}{s_y} \right)^{-1} = .48(0.5999) = .288$.

27. a, $180.975 \pm 3.4995 \cdot 14.30.42 \sqrt{18}$; 28. d; 29. b, $0.72 \pm 2.8609 \sqrt{\frac{3.99^2}{29} + \frac{3.19^2}{29}}$;

30. b, $t = \frac{202-190}{\sqrt{26}} = 2.233$;

31. c, for 49 degrees of freedom the test statistic 2.233 is between the table values 2.0096 and 2.4049. The area to the right of 2.4049 is .01 and the area to the right of 2.0096 is .025.

32. d, $P-value = 2P(t > 2.4786) = .02$ from the $t$-table with 26 df.