Provide an appropriate response.

✓ 1) A student surveyed a simple random sample of students at her college. Is this sample likely to be representative of all students at her college? Of all adults in the United States? Explain.

A) Yes
B) No
C) In order to represent all adults the sample would have to be taken from all adults.

Form a conclusion about statistical significance. Do not make any formal calculations. Either use the results provided or make subjective judgments about the results.

✓ 2) Last year, the average math SAT score for students at one school was 473. The headmaster introduced new teaching methods hoping to improve scores. This year, the mean math SAT score for a sample of students was 481. Is there statistically significant evidence that the new teaching method is effective? If the teaching method had no effect, there would be roughly a 3 in 10 chance of seeing such an increase. Does the result have statistical significance? Why or why not? Does the result have practical significance?

No - too close to old value. (Not substantially higher.)
I + 10 is not significant. The results would have to be less than or equal to 5-70 to be significant.

Determine whether the given value is from a discrete or continuous data set.

✓ 3) The total number of phone calls a sales representative makes in a month is 425.

A) Continuous
B) Discrete

Find the original data from the stemplot.

✓ 4)

<table>
<thead>
<tr>
<th>Stem</th>
<th>Leaves</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6 8</td>
</tr>
<tr>
<td>3</td>
<td>1 8</td>
</tr>
<tr>
<td>4</td>
<td>6 6</td>
</tr>
</tbody>
</table>

A) 26, 21, 28, 31, 41, 46  
B) 26, 28, 31, 46, 46  
C) 26, 28, 31, 38, 46, 46  
D) 21, 26, 21, 38, 48, 46
Determine whether the given description corresponds to an observational study or an experiment.

✓ 5) A clinic gives a drug to a group of ten patients and a placebo to another group of ten patients to find out if the drug has an effect on the patients' illness.

A) Observational study      B) Experiment

✓ 6) The frequency distribution below summarizes the home sale prices in the city of Summerhill for the month of June. Find the class boundaries for class 80.0-110.9.

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency</th>
<th>cumulative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>80.0 - 110.9</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>111.0 - 141.9</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>142.0 - 172.9</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>173.0 - 203.9</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>204.0 - 234.9</td>
<td>3</td>
<td>27</td>
</tr>
<tr>
<td>235.0 - 265.9</td>
<td>1</td>
<td>28</td>
</tr>
</tbody>
</table>

Use the pie chart to solve the problem.

✓ 7) A survey of the 9854 vehicles on the campus of State University yielded the following pie chart.

Motorcycles 9%
Convertibles 16%
Hatchbacks 36%
Vans 7%
Sedans 3%
Pickups 29%

Find the number of Pickup vehicles.

\[ n = \frac{29}{100} \times 9854 \]

Round your result to the nearest whole number.

\[ n = 2858 \]

✓ 8) Listed below are the amounts of time (in months) that the employees of a restaurant have been working at the restaurant. Find the median.

12 4 6 8.5 12 16 17 32 53 85 99 123 140 167

Find the median for the given sample data.

\[ \text{Median} = 84.5 \]

Find the mean of the data.

\[ \bar{x} = 55.3 \]

Find the midrange for the given sample data.

\[ \text{Midrange} = 81.5 \]

Find the variance for the given data.

\[ s^2 = 199.1 \]

Use the range rule of thumb to estimate the standard deviation. Round results to the nearest tenth.

✓ 9) A distribution of data has a maximum value of 64, a median value of 55, and a minimum of 46.

\[ s = 4.5 \]
Solve the problem.

10) A student earned grades of C, A, B, and A. Those courses had these corresponding numbers of credit hours: 4, 5, 2, and 5. The grading system assigns quality points to letter grades as follows: A = 4, B = 3, C = 2, D = 1, and F = 0. Compute the grade point average (GPA) and round the result to two decimal places.

\[ \frac{4 + 3 + 2 + 4}{16} = \frac{13}{16} \]

3.31 \leq \text{GPA} \leq 3.37

Use the empirical rule to solve the problem.

11) The systolic blood pressure of 18-year-old women is normally distributed with a mean of 120 mmHg and a standard deviation of 12 mmHg. What percentage of 18-year-old women have a systolic blood pressure between 96 mmHg and 144 mmHg?

\[ \frac{96 - 120}{12} = -2 \]
\[ \frac{144 - 120}{12} = 2 \]

\[ P(-2 \leq z \leq 2) = 0.95\% \]

Provide an appropriate response.

12) Suppose that all the values in a data set are converted to z-scores. Which of the statements below is true?

A: The mean of the z-scores will be zero, and the standard deviation of the z-scores will be the same as the standard deviation of the original data values.
B: The mean and standard deviation of the z-scores will be the same as the mean and standard deviation of the original data values.

C: The mean of the z-scores will be 0, and the standard deviation of the z-scores will be 1.
D: The mean and the standard deviation of the z-scores will both be zero.

Construct a boxplot for the given data. Include values of the 5-number summary in all boxplots.

13) The weights (in pounds) of 30 newborn babies are listed below. Construct a boxplot for the data set.

5.5 5.8 5.9 6.1 6.1 6.3 6.4 6.5 6.6 6.7 6.7 6.9 7.0 7.0 7.1 7.2 7.2 7.4 7.5 7.7 7.7 7.8 8.0 8.1 8.1 8.3 8.7

A) [Boxplot image]

B) [Boxplot image]
Find the number of standard deviations from the mean. Round your answer to two decimal places.

\[ Z = \frac{X - \mu}{\sigma} = \frac{11 - 5}{1.3} = \frac{6}{1.3} \approx 4.62 \]

Find the z-score corresponding to the given value and use the z-score to determine whether the value is unusual. Consider a score to be unusual if its z-score is less than -2.00 or greater than 2.00. Round the z-score to the nearest tenth if necessary.

\[ Z = \frac{X - \mu}{\sigma} = \frac{85 - 68}{12} = \frac{17}{12} \approx 1.417 \]

Determine which score corresponds to the higher relative position.

\[ Z = \frac{X - \mu}{\sigma} = \frac{53.8 - 37}{3} = \frac{16.8}{3} = 5.6 \]

Find the percentile for the data value.

Data set: 14, 12, 10, 18, 22, 18, 12, 18, 12, 18, 12, 18, 12, 18, 12, 18

\[ P = \frac{14}{15} \approx 0.933 \text{ or } 93.3\% \]
Find the indicated probability.

✓ 1) On a multiple choice test, each question has 4 possible answers. If you make a random guess on the first question, what is the probability that you are correct?

\[ P = \frac{1}{4} = 0.25 \]

Answer the question, considering an event to be "unusual" if its probability is less than or equal to 0.05.

✓ 2) Assume that one student in your class of 22 students is randomly selected to win a prize. Would it be "unusual" for you to win?

(A) Yes \( \frac{1}{22} = 0.045 < 0.05 \)
(B) No

Find the indicated complement.

✓ 3) Based on meteorological records, the probability that it will snow in a certain town on March 15th is 0.107. Find the probability that it will not snow on March 15th in that town.

\[ P(\text{not snow}) = 1 - P(\text{snow}) = 1 - 0.107 = 0.893 \]

Find the indicated probability.

✓ 4) A study of consumer smoking habits includes 155 people in the 18–22 age bracket (57 of whom smoke), 136 people in the 23–30 age bracket (37 of whom smoke), and 87 people in the 31–40 age bracket (24 of whom smoke). If one person is randomly selected from this sample, find the probability of getting someone who is age 18–22 or does not smoke.

\[ P(\text{18–22 or does not smoke}) = \frac{57 + 98}{155 + 136 + 87} = \frac{155}{378} = 0.3986 \]

Find the indicated probability. Express your answer as a simplified fraction unless otherwise noted.

✓ 5) The table below shows the soft drinks preferences of people in three age groups.

<table>
<thead>
<tr>
<th></th>
<th>cola</th>
<th>root beer</th>
<th>lemon-lime</th>
</tr>
</thead>
<tbody>
<tr>
<td>under 21 years of age</td>
<td>40</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>between 21 and 40</td>
<td>35</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>over 40 years of age</td>
<td>20</td>
<td>30</td>
<td>35</td>
</tr>
</tbody>
</table>

If one of the 255 subjects is randomly selected, find the probability that the person drinks root beer given that they are over 40.

\[ P(\text{Root beer} | \text{over 40}) = \frac{P(\text{Root beer} \cap \text{over 40})}{P(\text{over 40})} = \frac{30}{85} = 0.3529 = \frac{6}{17} \]

\[ P = \frac{353}{17} \]
Find the indicated probability. Round to the nearest thousandth.

6) A sample of 4 different calculators is randomly selected from a group containing 12 that are defective and 37 that have no defects. What is the probability that at least one of the calculators is defective?

\[
P(1^+) = 1 - P(0) = 1 - \left( \frac{\binom{37}{4}}{\binom{49}{4}} \right) = 1 - \left( \frac{\frac{37!}{4!(37-4)!}}{\frac{49!}{4!(49-4)!}} \right) = 1 - \left( \frac{47 \cdot 46}{49 \cdot 47} \right) \approx 0.317
\]

\[
P = 0.688
\]

Solve the problem.

7) A firm uses trend projection and seasonal factors to simulate sales for a given time period. It assigns "0" if sales fall, "1" if sales are steady, "2" if sales rise moderately, and "3" if sales rise a lot. The simulator generates the following output:

\[
011001110210102112321032102101
\]

Estimate the probability that sales will rise at least moderately.

\[
P = 0.323
\]

8) 8 basketball players are to be selected to play in a special game. The players will be selected from a list of 27 players. If the players are selected randomly, what is the probability that the 8 tallest players will be selected?

\[
P = \frac{\binom{8}{8}}{\binom{27}{8}} = \frac{1}{\binom{27}{8}} = \frac{1}{\frac{27!}{8!(27-8)!}} = \frac{1}{\frac{27 \cdot 26 \cdot 25 \cdot 24 \cdot 23 \cdot 22 \cdot 21 \cdot 20}{8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}} = \frac{1}{2,220,075}
\]

\[
P = 4.504 \times 10^{-7}
\]

Provide an appropriate response. Round to the nearest hundredth.

9) The random variable \( x \) is the number of houses sold by a realtor in a single month at the Sondson's Real Estate Office. Its probability distribution is as follows. Find the standard deviation for the probability distribution.

<table>
<thead>
<tr>
<th>Houses Sold (x)</th>
<th>Probability P(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.24</td>
</tr>
<tr>
<td>1</td>
<td>0.01</td>
</tr>
<tr>
<td>2</td>
<td>0.12</td>
</tr>
<tr>
<td>3</td>
<td>0.16</td>
</tr>
<tr>
<td>4</td>
<td>0.01</td>
</tr>
<tr>
<td>5</td>
<td>0.14</td>
</tr>
<tr>
<td>6</td>
<td>0.11</td>
</tr>
<tr>
<td>7</td>
<td>0.21</td>
</tr>
</tbody>
</table>

\[
\mu = 3.6
\]

\[
\sigma = 2.6
\]

\[
\sigma = 2.6222
\]
Provide an appropriate response.

10) Suppose you buy 1 ticket for $1 out of a lottery of 1,000 tickets where the prize for the one winning ticket is to be $500. What is your expected value?

\[
\begin{array}{c|c|c}
X & P(x) & XP(x) \\
1 & .999 & .999 \\
500 & .001 & .499 \\
\hline \\
& & -1.499 \\
\end{array}
\]

\[\text{Expected Value} = -1.50 \text{ or } 50 \text{ }.
\]

Assume that a procedure yields a binomial distribution with a trial repeated \(n\) times. Use the binomial probability formula to find the probability of \(x\) successes given the probability \(p\) of success on a single trial. Round to three decimal places.

11) \(n = 5, x = 2, p = 0.70\)

\[
\binom{n}{x} p^x q^{n-x} = \binom{5}{2} (0.70)^2 (0.30)^3
\]

Over table \(P = 0.749\)

\[p = 0.132\]

Find the mean, \(\mu\), for the binomial distribution which has the stated values of \(n\) and \(p\). Round answer to the nearest tenth.

12) \(n = 1742; p = 0.57\)

\[\mu = np = (1742)(0.57) = 992.94\]

\[\sigma = \sqrt{npq} = \sqrt{(1742)(0.57)(0.43)}\]

\[\sigma = 20.7\]

Minimum usual value \(
\mu - 2\sigma = 951.6
\)

\[992.94 - 2(20.7) = 951.6\]

Maximum usual value \(
\mu + 2\sigma = 1034.3
\)

\[992.94 + 2(20.7) = 1034.3\]

(17 each or almost 17 each rounds up to 18)

Use the Poisson Distribution to find the indicated probability.

13) The town of Fastville has been experiencing a mean of 64.1 car accidents per year. Find the probability that on a given day the number of car accidents in Fastville is 1. (Assume 365 days in a year.)

\[P(1) = \text{Poisson}(\mu, x) = \text{Poisson}(64.1, 1)\]

\[= 0.147\]

Over 1.1473

Over 1.14733
Provide an appropriate response.

1) If selecting samples of size \( n > 30 \) from a population with a known mean and standard deviation, what requirement, if any, must be satisfied in order to assume that the distribution of the sample means is a normal distribution?

(A) None; the distribution of sample means will be approximately normal.
(B) The population must have a standard deviation of 0.
(C) The mean must be equal to the standard deviation.
(D) The population must have a normal distribution.

2) Shaded area is 0.8599.

\[ Z = \frac{1.08}{1.080} \]

Solve the problem.

3) For a standard normal distribution, find the percentage of data that are between 3 standard deviations below the mean and 1 standard deviation above the mean.

\[ P (= -3, 1) \]

Find the indicated probability.

4) The incomes of trainees at a local mill are normally distributed with a mean of $1100 and a standard deviation of $150. What percentage of trainees earn less than $900 a month?

\[ P = \text{NormalCDF} (L_1, R, \bar{x}, s) \]
\[ (0, 900, 1100, 150) \]

Use the binomial distribution.

5) A coin is tossed 20 times. A person, who claims to have extrasensory perception, is asked to predict the outcome of each flip in advance. She predicts correctly on 16 tosses. What is the probability of being correct 16 or more times by guessing? Does this probability seem to verify her claim?

\[ P(16^-) = 1 - P(15^-) \]
\[ = 1 - \text{binomialCDF} (20, .5, 16) \]
\[ = .0059084 \]

\[ P = .00591 \]
6) A bank's loan officer rates applicants for credit. The ratings are normally distributed with a mean of 200 and a standard deviation of 50. If 40 different applicants are randomly selected, find the probability that their mean is above 215.

\[ P(z > z^*) = \frac{\text{Normalcdf}(215, 1000, 200, 50 \div \sqrt{40})}{0.02889} = 0.02889 \]

Ch 6

7) Of 238 employees selected randomly from one company, 10.08% of them commute by carpooling. Construct a 90% confidence interval for the true percentage of all employees of the company who carpool.

\[ \hat{p} = \frac{x}{n} = \frac{23.94 \div 238}{0.06874, 0.13232} \]

\[ x = \hat{p} = 23.94 \div 238 = 0.06874, 0.13232 \]

\[ \hat{p} = 0.06874, 0.13232 \]

Solve the problem.

8) The following confidence interval is obtained for a population proportion, p: (0.528, 0.554). Use these confidence interval limits to find the margin of error, E.

\[ \hat{p} - E < p < \hat{p} + E \]

\[ 0.528 < p < 0.554 \]

\[ E = 0.554 - 0.528 = 0.026 \]

9) How many weeks of data must be randomly sampled to estimate the mean weekly sales of a new line of athletic footwear? We want 99% confidence that the sample mean is within $500 of the population mean, and the population standard deviation is known to be $1500.

\[ n = \left[ \frac{z_{0.005}}{E} \right]^2 = \left[ \frac{2.576(1500)}{500} \right]^2 \]

\[ = 59.6756 \]
Use the given degree of confidence and sample data to construct a confidence interval for the population mean \( \mu \). Assume that the population has a normal distribution.

10) A sociologist develops a test to measure attitudes towards public transportation, and 27 randomly selected subjects are given the test. Their mean score is 76.2 and their standard deviation is 21.4. Construct the 95% confidence interval for the mean score of all such subjects.

\[
\bar{x} = 76.2, \quad s = 21.4
\]

\[
\mu - \bar{x} \leq \mu \leq \mu + \bar{x}
\]

\[
\mu \in [76.2 - 8.5, 76.2 + 8.5]
\]

67.7, 84.7

8 Points

10) Use the given degree of confidence and sample data to find a confidence interval for the population standard deviation \( \sigma \). Assume that the population has a normal distribution. Round the confidence interval limits to the same number of decimal places as the sample standard deviation.

11) The mean replacement time for a random sample of 20 washing machines is 11.6 years and the standard deviation is 2.3 years. Construct a 99% confidence interval for the standard deviation, \( \sigma \), of the replacement times of all washing machines of this type.

\[
\bar{x} = 11.6
\]

\[
s = 2.3
\]

99% Conf.

\[
\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}
\]

\[
\frac{(19)(2.3)^2}{38.582} \leq \sigma^2 \leq \frac{(19)(2.3)^2}{6.644}
\]

1.6 \leq \sigma \leq 3.8

20 Points
Assume that the data has a normal distribution and the number of observations is greater than fifty. Find the critical z value used to test a null hypothesis.

1) \( \alpha = 0.05 \) for a left-tailed test.

\[ z = -1.645 \]

Use the given information to find the P-value. Also, use a 0.05 significance level and state the conclusion about the null hypothesis (reject the null hypothesis or fail to reject the null hypothesis).

2) With \( H_1: \mu < 25 \), the test statistic is \( z = -1.68 \).

\[ P = 0.046 
\]

3) The principal of a school claims that the percentage of students at his school that come from single-parent homes is 20%. Identify the type I or type II error for the test.

A. Fail to reject the claim that the percentage of students that come from single-parent homes is equal to 20% when that percentage is actually different from 20%.

B. Reject the claim that the percentage of students that come from single-parent homes is equal to 20% when that percentage is actually 20%.

C. Fail to reject the claim that the percentage of students that come from single-parent homes is equal to 20% when that percentage is actually 20%.

D. Reject the claim that the percentage of students that come from single-parent homes is equal to 20% when that percentage is actually less than 20%.

4) Determine whether the hypothesis test involves a sampling distribution of means that is a normal distribution, Student t distribution, or neither.

A. Neither

B. Normal

C. Student t
Identify the null hypothesis, alternative hypothesis, test statistic, P-value, conclusion about the null hypothesis, and final conclusion that addresses the original claim.

5) A medical school claims that more than 28% of its students plan to go into general practice. It is found that among a random sample of 130 of the school’s students, 32% of them plan to go into general practice. Find the P-value for a test of the school’s claim.

\[ \begin{align*}
H_0 : P &= 0.28 \\
H_1 : P &> 0.28 \text{ (original)} \\
\text{TS} : Z &= 1.09
\end{align*} \]

\[ \begin{align*}
\text{2 CV} : Z &= 1.645 \\
\text{3 P} : P &= 0.13700 > 0.05 = \alpha \\
N &= 130 \\
\hat{p} &= \frac{X}{n} = 0.32 \\
x &= np = (130)(0.32) = 41.6 \approx 42
\end{align*} \]

\[ \begin{align*}
\text{Treat } H_0 : \text{ Fail to reject } H_0
\end{align*} \]

Conclusion: There is not sufficient sample evidence to support the claim that more than 28% of its students plan to go into general practice.

Assume that a simple random sample has been selected from a normally distributed population. Find the test statistic, P-value, critical value(s), and state the final conclusion.

6) Test the claim that the mean lifetime of car engines of a particular type is greater than 220,000 miles. Sample data are summarized as \( n = 23, \bar{x} = 226,450 \) miles, and \( s = 11,500 \) miles. Use a significance level of \( \alpha = 0.01 \).

\[ \begin{align*}
H_0 : \mu &= 220,000 \\
H_1 : \mu > 220,000 \text{ (original)} \\
\text{TS} : t &= 2.690 \\
\text{2 CV} : t_{0.01, 22} &= 2.508 \\
\text{3 P} : P = 0.00669 < 0.01 = \alpha
\end{align*} \]

Treat \( H_0 : \text{ Reject } H_0 

Conclusion: Core 5. The sample data support the claim that the mean lifetime of car engines of a particular type is greater than 220,000.
Use the traditional method to test the given hypothesis. Assume that the population is normally distributed and that the sample has been randomly selected.

7) When 14 bolts are tested for hardness, their indexes have a standard deviation of 41.7. Test the claim that the standard deviation of the hardness indexes for all such bolts is greater than 30.0. Use a 0.025 level of significance.

\[ H_0 : \sigma = 30 \]
\[ H_1 : \sigma > 30 \text{(Revised)} \]

(1) \[ \chi^2 = (n-1) \frac{s^2}{\sigma^2} = \frac{(13)(41.7)^2}{30^2} = 25.117 \]

n = 14

(2) TS: \[ \chi^2 = 25.117 \]

(3) CV: \[ \chi_{0.025}^2 = 24.736 \]

(4) P: \[ P = 0.02228 \]

(1) Treat \( H_0 \): reject \( H_0 \) \[ \chi^2_{CDF} (L, R, df) \]

(2) Conclusion: \( \text{Conc} \neq 3 \)

The sample data support the claim that the standard deviation of the hardness indexes for all such bolts is greater than 30.0.
Determine whether the samples are independent or dependent.

\[ \checkmark 1) \text{ The effectiveness of a headache medicine is tested by measuring the intensity of a headache in patients before and after drug treatment. The data consist of before and after intensities for each patient.} \]

[A) Dependent samples | B) Independent samples

State what the given confidence interval suggests about the two population means.

\[ \checkmark 2) \text{ A paint manufacturer made a modification to a paint to speed up its drying time. Independent simple random samples of 11 cans of type A (the original paint) and 9 cans of type B (the modified paint) were selected and applied to similar surfaces. The drying times, in hours, were recorded. The summary statistics are as follows.} \]

<table>
<thead>
<tr>
<th>Type A</th>
<th>Type B</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 = 75.9 \text{ hrs} )</td>
<td>( x_2 = 65.6 \text{ hrs} )</td>
</tr>
<tr>
<td>( s_1 = 4.5 \text{ hrs} )</td>
<td>( s_2 = 5.1 \text{ hrs} )</td>
</tr>
<tr>
<td>( n_1 = 11 )</td>
<td>( n_2 = 9 )</td>
</tr>
</tbody>
</table>

The following 98% confidence interval was obtained for \( \mu_1 - \mu_2 \), the difference between the mean drying time for paint cans of type A and the mean drying time for paint cans of type B: \( 4.69 < \mu_1 - \mu_2 < 13.91 \text{ hrs} \)

What does the confidence interval suggest about the population means?

A) The confidence interval includes only positive values which suggests that the mean drying time for paint type A is smaller than the mean drying time for paint type B. \[ \text{NO} \]

B) The confidence interval includes 0 which suggests that the two population means might be equal. There doesn't appear to be a significant difference between the mean drying time for paint type A and the mean drying time for paint type B. The modification does not seem to be effective in reducing drying times. \[ \text{NO} \]

C) The confidence interval includes only positive values which suggests that the two population means might be equal. There doesn't appear to be a significant difference between the mean drying time for paint type A and the mean drying time for paint type B. The modification does not seem to be effective in reducing drying times. \[ \text{NO} \]

D) The confidence interval includes only positive values which suggests that the mean drying time for paint type A is greater than the mean drying time for paint type B. The modification seems to be effective in reducing drying times. \[ \text{YES} \]
Find the critical z value(s) for the given hypothesis test.

3) The table shows the number of smokers in a random sample of 500 adults aged 20–24 and the number of smokers in a random sample of 450 adults aged 25–29. Do the data provide sufficient evidence that the proportion of smokers in the 20–24 age group is different from the proportion of smokers in the 25–29 age group? Assume that you plan to use a significance level of \( \alpha = 0.10 \) to test the claim that \( p_1 \neq p_2 \).

<table>
<thead>
<tr>
<th>Age 20–24</th>
<th>Age 25–29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number in sample</td>
<td>500</td>
</tr>
<tr>
<td>Number of smokers</td>
<td>110</td>
</tr>
</tbody>
</table>

\[
\hat{p}_1 = \frac{110}{500} \quad \hat{p}_2 = \frac{63}{450}
\]

(1) \( H_0 : p_1 = p_2 \)  \( \alpha = 0.10 \)

(2) \( Z = 3.19 \)

(3) \( 0.0014 \leq 0.10 \)

(4) Treat \( H_0 \): \( \text{Reject } H_0 \)

(5) Conclusion: \#3 The sample data supports the claim that \( p_1 \neq p_2 \)

4) A researcher wishes to determine whether the blood pressure of vegetarians is, on average, lower than the blood pressure of nonvegetarians. Independent simple random samples of 85 vegetarians and 75 nonvegetarians yielded the following sample statistics for systolic blood pressure:

<table>
<thead>
<tr>
<th>Vegetarians</th>
<th>Nonvegetarians</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_1 = 85 )</td>
<td>( n_2 = 75 )</td>
</tr>
<tr>
<td>( x_1 = 124.1 ) mmHg</td>
<td>( x_2 = 138.7 ) mmHg</td>
</tr>
<tr>
<td>( s_1 = 38.7 ) mmHg</td>
<td>( s_2 = 39.2 ) mmHg</td>
</tr>
</tbody>
</table>

(1) \( H_0 : \mu_1 = \mu_2 \)

(2) \( Z = -2.365 \) \( df = 155 \)

(3) \( CV : t = -2.345 \) \( df = 100 \)

(4) \( 0.00963 \leq 0.01 \)

(5) Treat \( H_0 \): \( \text{Reject } H_0 \)

(6) Conclusion: \#3 The sample data supports the claim that the mean systolic blood pressure for vegetarians is lower than the mean systolic blood pressure for nonvegetarians.
Statistics 50  Test4B  Chapters 9  Fall 2011  Page 3.

Construct a confidence interval for \( \mu_d \), the mean of the differences \( d \) for the population of paired data. Assume that the population of paired differences is normally distributed.

\[ t = \frac{\overline{d} - \mu_d}{S_d / \sqrt{n}} \]

5) The table below shows the weights of 9 subjects before and after following a particular diet for two months.

<table>
<thead>
<tr>
<th>Subject</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>168</td>
<td>180</td>
<td>157</td>
<td>132</td>
<td>202</td>
<td>124</td>
<td>190</td>
<td>210</td>
<td>171</td>
</tr>
<tr>
<td>After</td>
<td>162</td>
<td>178</td>
<td>145</td>
<td>125</td>
<td>171</td>
<td>126</td>
<td>180</td>
<td>195</td>
<td>163</td>
</tr>
</tbody>
</table>

Construct a 99% confidence interval for the mean difference of the "before" minus "after" weights.

\[ -6.91 < \mu_d < 20.4 \]

Assume that you want to test the claim that the paired sample data come from a population for which the mean difference is \( \mu_d = 0 \). Compute the value of the test statistic.

\[ H_0 : \mu_d = 0 \]
\[ H_1 : \mu_d \neq 0 \]

\[ t = 3.156 \]
\[ P = 0.0134 < \alpha = 0.05 \]

\( H_0 \) : Fail to reject \( H_0 \)

2 Conclusion: #2. There is not sufficient evidence to warrant rejection of the claim that the paired sample data come from a population for which the mean difference \( \mu_d = 0 \).

Test the indicated claim about the variances or standard deviations of two populations. Assume that both samples are independent simple random samples from populations having normal distributions.

\[ F = \frac{S_X^2}{S_Y^2} \]

\( \alpha = 0.05 \)

6) Two types of flares are tested and their burning times are recorded. The summary statistics are given below. Use a 0.05 significance level to test the claim that the burning times for Brand X flares have the same variance as the burning times for Brand Y flares.

<table>
<thead>
<tr>
<th>Brand X</th>
<th>Brand Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 35 )</td>
<td>( n = 40 )</td>
</tr>
<tr>
<td>( \overline{x} = 19.4 \text{ min} )</td>
<td>( \overline{x} = 15.1 \text{ min} )</td>
</tr>
<tr>
<td>( s = 1.4 \text{ min} )</td>
<td>( s = 0.8 \text{ min} )</td>
</tr>
</tbody>
</table>

\[ H_0 : \sigma_X^2 = \sigma_Y^2 \]
\[ H_1 : \sigma_X^2 \neq \sigma_Y^2 \]

\[ F = 3.0625 \]

\( F_{30,0.025} = 1.9429 \)

\( P = 0.1298 \leq 0.025 \)

\( P = 0.00091298 \)

\( P = 2 \times 10^{-4} \)

\( P = 2 \times \text{CDF}(L, R, df_X, df_Y) \)

(3.0625, 100, 34.89)

by hand.

\( d_f = 30 \)

\( d_f = 100 \)

\( d_f = 34.89 \)

Treat \( H_0 \): reject \( H_0 \)

Conclusion: #1. There is sufficient evidence to warrant rejection of the claim that the burning times for Brand X flares have the same variance as the burning times for Brand Y flares.
Describe the error in the stated conclusion.

1) Given: Each school in a state reports the average SAT score of its students. There is a significant linear correlation between the average SAT score of a school and the average annual income in the district in which the school is located.

Conclusion: There is a significant linear correlation between individual SAT scores and family income. Error results with data based on averages. Averages suppress individual variation, and may inflate the correlation coefficient. A significant correlation between average SAT score of a school and average score at a school does not imply a significant correlation between individual SAT score and family income.

Given the linear correlation coefficient \( r \) and the sample size \( n \), determine the critical values of \( r \) and use your findings to state whether or not the given \( r \) represents a significant linear correlation. Use a significance level of 0.05.

2) \( r = 0.898, n = 10 \)

Critical values: \( r = \pm 1.632 \)

4 Points each part.

Find the value of the linear correlation coefficient \( r \).

3) The paired data below consist of the costs of advertising (in thousands of dollars) and the number of products sold (in thousands):

<table>
<thead>
<tr>
<th>Cost</th>
<th>9 2 3 4 2 5 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>85 52 55 68 67 86 83</td>
</tr>
</tbody>
</table>

1. \( H_0 : \rho = 0 \)  
2. \( H_1 : \rho \neq 0 \)  
3. \( r = \frac{808}{30.65} \)  
4. \( (\bar{x}, \bar{y}) = (48.6, 70.1) \)

The standard error of estimate = 9.16

5. CV: \( r = 7.54 \)  
6. The equation of the regression line is \( \hat{y} = \) \( \bar{y} = 52.5 + 3.79x \)

1. Treat \( H_0 \): reject \( H_0 \)

3. Conclusion: The correlation is statistically significant.
Use the given data to find the equation of the regression line. Round the final values to three significant digits, if necessary.

\[
\begin{array}{c|cccc}
4) & x & 6 & 8 & 20 \\ 
\hline
y & 2 & 4 & 13 & 20 \\
\end{array}
\]

\[
\begin{align*}
\text{4) } i &= 999 \\
\text{5) } i^2 &= 999 \\
\text{6) } (\bar{x}, \bar{y}) &= (15.5, 9.75) \\
\text{7) } \hat{y} &= -2.70 + 8.03x
\end{align*}
\]

The equation of the regression line is 
\[ \hat{y} = -2.70 + 8.03x \]

Use the given information to find the coefficient of determination.

5) A regression equation is obtained for a collection of paired data. It is found that the explained variation is 84, and the unexplained variation is 44.6.

\[
\frac{84}{44.6} = \frac{84}{128.6} = 0.653
\]

Use the computer display to answer the question.

6) A collection of paired data consists of the number of years that students have studied Spanish and their scores on a Spanish language proficiency test. A computer program was used to obtain the least squares linear regression line and the computer output is shown below. Along with the paired sample data, the program was also given an \(x\) value of 2 (years of study) to be used for predicting test score.

The regression equation is
\[ \hat{y} = 31.55 + 10.90x \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>StdDev</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>31.55</td>
<td>6.360</td>
<td>4.96</td>
<td>0.000</td>
</tr>
<tr>
<td>Years</td>
<td>10.90</td>
<td>1.744</td>
<td>6.25</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[ S = 5.651 \quad \text{R-Sq} = 83.0\% \quad \text{R-Sq (Adj)} = 82.7\% \]

Predicted values

\[ \hat{y} = 31.55 + 10.90(2) = 80.6 \]

If a person studies 4.5 years, what is the single value that is the best predicted test score? Assume that there is a significant linear correlation between years of study and test score.

\[ \hat{y} = 31.55 + 10.90(4.5) = 80.6 \]
Perform the indicated goodness-of-fit test.

7) In studying the responses to a multiple-choice test question, the following sample data were obtained. At the 0.05 significance level, test the claim that the responses occur with the same frequency.

<table>
<thead>
<tr>
<th>Response</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>12</td>
<td>15</td>
<td>16</td>
<td>18</td>
<td>19</td>
</tr>
</tbody>
</table>

1. \( H_0: \frac{P_A}{P_B} = \frac{P_C}{P_D} = \frac{P_E}{P_D} \) or original responses occur with same frequency.

2. \( H_1: 2 \) or more are different from others.

3. \( \chi^2 = 1.475 \)

4. \( \chi^2 \approx 1.15 \approx 9.485 \)

5. \( x^2 = \begin{bmatrix} 12 & 15 & 16 & 18 & 19 \\ 10 & 10 & 10 & 10 & 10 \end{bmatrix} \)

6. \( P = .75 \approx .74 \approx .05 = \alpha \)

1. Treat \( H_0: \) Fail to reject \( H_0 \)

2. Conclusion: \( \chi^2 \approx 1.15 \). There is not sufficient evidence to warrant rejection of the claim that responses occur with the same frequency.

Solve the problem.

8) Use a 0.01 significance level to test the claim that the proportion of men who plan to vote in the next election is the same as the proportion of women who plan to vote. 300 men and 300 women were randomly selected and asked whether they planned to vote in the next election. The results are shown below.

<table>
<thead>
<tr>
<th>Plan to vote</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>170</td>
<td>185</td>
</tr>
<tr>
<td>Do not plan to vote</td>
<td>130</td>
<td>115</td>
</tr>
</tbody>
</table>

1. \( H_0: \) Gender and Voting independent (original) \( P_M = P_W \)

2. \( H_1: \) Gender and Voting pattern dependent \( P_M \neq P_W \)

3. \( \chi^2 = 1.55 \)

4. \( \chi^2_{.01, .04} = 6.635 \)

5. \( P = .212 > .01 = \alpha \)

1. Treat \( H_0: \) Fail to reject \( H_0 \)

2. Conclusion: \( \chi^2 \approx 1.55 \). There is not sufficient evidence to warrant rejection of the claim that the proportion of men who plan to vote in the next election is the same as the proportion of women who plan to vote.
Given below are the analysis of variance results from a Minitab display. Assume that you want to use a 0.05 significance level in testing the null hypothesis that the different samples come from populations with the same mean.

\[ F = 1.6 \]
\[ P = 0.264 \]
\[ F^2_{0.05} = 4.0662 \]

Test the claim that the samples come from populations with the same mean. Assume that the populations are normally distributed with the same variance.

\( F = \frac{MS_A}{MS_E} = \frac{6 \cdot 2.6263}{23.716} = 6.6443 \)
\[ P = \text{Fcdf}(6.6443, 200, 2, 15) = 0.0085 > 0.05 \]

(1) \( H_0 : \mu_A = \mu_B = \mu_C \) (original)
(2) \( H_1 : \) At least one different
(3) \( \text{TS} : F = 6.6443 \)
(4) \( \text{CV} : F_{0.05}^2 = 3.6823 \)
(5) \( P : P = 0.0085 < 0.05 = \alpha \)

Treat \( H_0 : \) Reject \( H_0 \)
Conclusion: There is sufficient evidence to warrant rejection of the claim that the populations have the same mean.
3) At the 0.025 significance level, test the claim that the brands have the same mean if the following sample results have been obtained.

<table>
<thead>
<tr>
<th>Brand A</th>
<th>Brand B</th>
<th>Brand C</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>15</td>
<td>21</td>
</tr>
<tr>
<td>25</td>
<td>15</td>
<td>22</td>
</tr>
<tr>
<td>21</td>
<td>14</td>
<td>20</td>
</tr>
<tr>
<td>23</td>
<td>23</td>
<td>19</td>
</tr>
<tr>
<td>22</td>
<td>22</td>
<td>18</td>
</tr>
<tr>
<td>20</td>
<td>28</td>
<td>28</td>
</tr>
</tbody>
</table>

- **H₀**: \( M_A = M_B = M_C \)
- **CV**: \( \bar{X}_{15, 025} = 4.7650 \)
- **H₁**: At least two different \( \mu_i \)
- **P**: \( P = .92963 \)
- **TS**: \( F = .07333 \)

Compare \( P \) to \( \alpha \)

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares (SS)</th>
<th>Degrees of Freedom</th>
<th>Mean Square (MS)</th>
<th>F Test Statistic</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td>2.6492</td>
<td>2</td>
<td>1.4246</td>
<td>.07883</td>
<td>9.2963</td>
</tr>
<tr>
<td>Error</td>
<td>291.42857</td>
<td>15</td>
<td>19.42857</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>294.27</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( \bar{X}_A = \frac{6}{2} = 3 \)
\( \bar{X}_B = \frac{21}{2} = 10.5 \)
\( \bar{X}_C = \frac{20}{2} = 10 \)

\( \bar{X} = \frac{371}{18} = 20.61 \)

\( F = \frac{\sum (\bar{X}_i - \bar{X})^2}{k-1} \)

\( F = \frac{\sum (\bar{X}_i - \bar{X})^2}{k-1} \)

\( = \frac{6(21-20.61)^2 + 7(20.7143-20)^2 + 5(20-20.61)^2}{18} \)

\( = \frac{6(21-20.61)^2 + 7(20.7143-20)^2 + 5(20-20.61)^2}{18} \)

\( = \frac{371}{18} = 20.61 \)

- **Treat H₀**: Fail to reject H₀
- **Conclusion**: Conclusion: There is not sufficient evidence to warrant rejection of the claim that the populations have the same mean.
Use the Minitab display to test the indicated claim.

4) A manager records the production output of three employees who each work on three different machines for three different days. The sample results are given below and the Minitab results follow.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employee</td>
<td>I 23, 27, 29</td>
<td>30, 27, 25</td>
<td>18, 20, 22</td>
</tr>
<tr>
<td></td>
<td>II 25, 26, 24</td>
<td>24, 29, 26</td>
<td>19, 16, 14</td>
</tr>
<tr>
<td></td>
<td>III 28, 25, 26</td>
<td>25, 27, 23</td>
<td>15, 11, 17</td>
</tr>
</tbody>
</table>

**ANALYSIS OF VARIANCE ITEMS**

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>MACHINE</td>
<td>2</td>
<td>34.66</td>
<td>17.33</td>
<td>3.1857</td>
</tr>
<tr>
<td>EMPLOYEE</td>
<td>2</td>
<td>504.67</td>
<td>-</td>
<td>252.385</td>
</tr>
<tr>
<td>INTERACTION</td>
<td>4</td>
<td>26.68</td>
<td>6.67</td>
<td>1.2261</td>
</tr>
<tr>
<td>ERROR</td>
<td>18</td>
<td>98.00</td>
<td>-</td>
<td>5.44</td>
</tr>
<tr>
<td>TOTAL</td>
<td>26</td>
<td>664.01</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

5) Using a 0.05 significance level, test the claim that the interaction between employee and machine has no effect on the number of items produced.

\[
\begin{align*}
\text{H}_0 & : \quad \text{NO interaction effect (original)} \\
\text{H}_1 & : \quad \text{There is an interaction effect} \\
\text{TS} & : \quad F = 1.23 \\
\text{CV} & : \quad F_{18,0.05} = 2.9277 \\
\text{P} & : \quad \frac{F_{18,0.05}}{(1.23, 200, 2, 18)} = 33.31 \geq 0.05 = \alpha \\
\text{Treat H}_0 : \quad \text{ Fail to reject H}_0 \\
\text{Conclusion:} & \quad \text{There is not sufficient evidence to warrant rejection of the claim that the interaction between employee and machine has no effect on the number of items produced.}
\end{align*}
\]
6. A placement test is required for students desiring to take a finite mathematics course at a university. The instructor of the course studies the relationship between students' placement test score and final course score. A random sample of eight students yields the following data.

<table>
<thead>
<tr>
<th>Placement Score</th>
<th>Final Course Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>63</td>
</tr>
<tr>
<td>9</td>
<td>41</td>
</tr>
<tr>
<td>5</td>
<td>54</td>
</tr>
<tr>
<td>3</td>
<td>32</td>
</tr>
<tr>
<td>7</td>
<td>93</td>
</tr>
<tr>
<td>7</td>
<td>70</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
</tr>
<tr>
<td>6</td>
<td>61</td>
</tr>
</tbody>
</table>

\[ r_s = \frac{6 \cdot 1 \cdot 0 - 8 \cdot 5 \cdot 25}{\sqrt{(6 \cdot 1 \cdot 0 - 8) \cdot (6 \cdot 5 \cdot 25)}} = -1.179 \]

Compute the rank correlation coefficient, \( r_s \), of the data and test the claim of correlation between placement score and final course score. Use a significance level of 0.05.

1. \( H_0: \rho_s = 0 \)
2. \( H_1: \rho_s \neq 0 \)
3. \( T_S: r_s = -1.179 \)
4. \( C_V: r_s = \frac{1.96}{\sqrt{1(9)}} \)

Conclusion: The rank correlation is significant.

Use the rank correlation coefficient to test for a correlation between the two variables.

14. Given that the rank correlation coefficient, \( r_s \), for 97 pairs of data is -0.526, test the claim of correlation between the two variables. Use a significance level of 0.05.

1. \( T_S: r_s = -0.526 \)
2. \( C_V: r_s = \pm 1.96 \)

Conclusion: The correlation is significant.