Use common sense to determine whether the given event is impossible; possible, but very unlikely; or possible and likely.

1) When Dave picked two marbles from a bag containing one red, one blue, and one yellow marble, he got two marbles of the same color. (Assume that he didn't replace the first marble before picking the second).
   A) Possible and likely
   B) Impossible
   C) Possible, but very unlikely

2) In a random sample of 160 women, 78% favored stricter gun control laws. In a random sample of 220 men, 61% favored stricter gun control laws. Is there statistically significant evidence that a larger proportion of women than men favor stricter gun control laws?
   Yes, 78% is significantly larger than 61% for samples this large. If they were equal, it is unlikely sample would be this different.

Determine which of the four levels of measurement (nominal, ordinal, interval, ratio) is most appropriate.

3) Student's grades, A, B, or C, on a test.
   A) Ratio
   B) Ordinal
   C) Nominal
   D) Interval

Use critical thinking to address the key issue.

4) A company accused of downsizing workers defended itself with the following statement: "Yes, we were forced to lay off 20% of our workforce last year, but this year we increased our workforce by 20%, and we therefore now have the same number of employees as before the layoff." What is the flaw in this argument?
   If we start with 100 laying off 20% would take us to 80. Then if we add 20% we will be up to 96.

Provide an appropriate response.

5) A bus company claims that in the past year it has reduced the number of late departures of buses by 100%. What is wrong with this statement?
   This implies we now have no late departures which is not likely.
6) The frequency distribution below summarizes employee years of service for Alpha Corporation. Find the class boundaries for class 26-30.

<table>
<thead>
<tr>
<th>Years of Service</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-5</td>
<td>5</td>
</tr>
<tr>
<td>6-10</td>
<td>20</td>
</tr>
<tr>
<td>11-15</td>
<td>25</td>
</tr>
<tr>
<td>16-20</td>
<td>10</td>
</tr>
<tr>
<td>21-25</td>
<td>5</td>
</tr>
<tr>
<td>26-30</td>
<td>3</td>
</tr>
</tbody>
</table>

\[
\bar{x} = \frac{17.9}{68} = 0.264\]  \[s = 6.0\]  \[s^2 = 36.2\]

Solve the problem.

7) The histogram below shows the distribution of the assets (in millions of dollars) of 71 companies. Does the distribution appear to be normal?

Provide an appropriate response.

8) Suppose that a state introduces a state income tax which will be at a flat rate of 3%. The state legislature wishes to estimate how much money they will receive in taxes, and to do this they need to know the average income of residents of the state. Which information would be most useful, the mean income, the median income, or the mode of the incomes? Why?

Mean - It represents all scores.

9) The ages (in years) of the eight passengers on a bus are listed below.

Find the mean: 23.8  \[\frac{42 + 3}{2} = 22.5\]  \[\frac{45}{2} = 22.5\]

Find the median: 24  \[\sqrt{200.6} = 14.2\]

Find the mode(s): None
Statistics 50  Test 1  Chapters 1-3  Spring 2012  Page 3.

10) A student earned grades of 78, 78, 91, and 91 on her four regular tests. She earned a grade of 78 on the final exam and 87 on her class projects. Her combined homework grade was 87. The four regular tests count for 40% of the final grade, the final exam counts for 30%, the project counts for 10%, and homework counts for 20%. What is her weighted mean grade? Round to one decimal place.

83.3

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

11) The following is a set of data showing the water temperature in a heated tub at different time intervals.

114.0  113.2  113.4  115.9  114.2  113.0  114.1

\[ \frac{116.3 - 113.0}{4} = \frac{3.3}{4} = 0.825 \]

Use the empirical rule to solve the problem.

12) The amount of Jen's monthly phone bill is normally distributed with a mean of $79 and a standard deviation of $8. What percentage of her phone bills are between $55 and $103?

\[ z = \frac{X - \mu}{\sigma} = \frac{79 - 55}{8} = 3 \]

Solve the problem.

13) In chemistry, the Kelvin scale is often used to measure temperatures. On the Kelvin scale, zero degrees is absolute zero. Temperatures on the Kelvin scale are related to temperatures on the Celsius scale as follows: \( K = C + 273 \). Temperatures on the Fahrenheit scale are related to temperatures on the Celsius scale as follows:

\[ F = \frac{9}{5} C + 32 \]

A set of temperatures is given in Celsius, Kelvin, and Fahrenheit. How will the standard deviations of the three sets of data compare?

Celsius and Kelvin are the same.
Fahrenheit will be \( \frac{9}{5} \) as large as the other two.

Find the number of standard deviations from the mean. Round your answer to two decimal places.

14) The number of hours per day a college student spends on homework has a mean of 4 hours and a standard deviation of 1.25 hours. Yesterday she spent 2 hours on homework. How many standard deviations from the mean is that?

\[ z = \frac{X - \mu}{\sigma} = \frac{2 - 4}{1.25} = \frac{-2}{1.25} = -1.6 \]
Statistics 50 Test 1 Chapters 1-3 Spring 2012 Page 4.

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.

15) A time for the 100 meter sprint of 20.4 seconds at a school where the mean time for the 100 meter sprint is 17.5 seconds and the standard deviation is 2.1 seconds.

\[ z = \frac{x - \overline{x}}{s} = \frac{20.4 - 17.5}{2.1} = 1.3660 \]

Determine which score corresponds to the higher relative position.

16) Which score has a higher relative position, a score of 278.4 on a test for which \(\overline{x} = 240\) and \(s = 24\), or a score of 60 on a test for which \(\overline{x} = 60\) and \(s = 6\)?

A) A score of 60
B) A score of 278.4
C) Both scores have the same relative position.

Calculate and display both z-scores

Find the percentile for the data value.

17) Data set: 53 41 42 66 70 74 31; data value: 53

\[ z = \frac{x - \overline{x}}{s} = \frac{53 - 61}{2.5} = \frac{53 - 60}{6} = -\frac{1}{2} \]

Find the indicated measure.

18) Use the given sample data to find \(P_{0.25}\)

\[ P_{0.25} = 42.57 \]

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

19) The weights (in ounces) of 27 tomatoes are shown below. Construct a boxplot for the data set.

\[ 2.0 2.1 2.2 2.2 2.4 2.4 2.5 2.5 2.5 2.6 2.6 2.7 2.7 2.7 2.8 2.8 2.8 2.9 2.9 3.0 3.0 3.1 3.1 3.2 3.2 3.4 \]

A)\(\)

B)\(\)

C)\(\)

D)\(\)

Provide an appropriate response.

20) Human body temperatures have a mean of 98.2°F and a standard deviation of 0.62°F. Sally's temperature can be described by \(z = -1.3\). What is her temperature? Round your answer to the nearest hundredth.

A) 97.39°F B) 99.01°F C) 96.11°F D) 96.90°F

\[ z = \frac{x - \overline{x}}{s} \]

\[ x = z \cdot s + \overline{x} = (-1.3)(0.62) + 98.2 \]
Find the indicated probability. Round to the nearest thousandth.

1) Fly Best Airlines boasts that this year among 316 flights from Los Angeles to New York City 310 arrived on time. What is the probability that the next flight on Fly Best from Los Angeles to New York City will be on time?

\[ P = \frac{310}{316} = 0.961 \]

2) Of the 45 people who answered "yes" to a question, 13 were male. Of the 83 people that answered "no" to the question, 7 were male. If one person is selected at random from the group, what is the probability that the person answered "yes" or was male?

\[ P = \frac{13}{45} + \frac{7}{83} = 0.406 \]

3) A bag contains 6 red marbles, 3 blue marbles, and 1 green marble. Find \( P(\text{not blue}) \).

\[ P = \frac{10}{10} = 0.7 \]

4) A IRS auditor randomly selects 3 tax returns from 59 returns of which 11 contain errors. What is the probability that she selects none of those containing errors? Round to four decimal places.

\[ P = \frac{\binom{11}{0} \cdot \binom{48}{3}}{\binom{59}{3}} = 0.532037 \]

5) In a batch of 8,000 clock radios 9% are defective. A sample of 8 clock radios is randomly selected without replacement from the 8,000 and tested. The entire batch will be rejected if at least one of those tested is defective. What is the probability that the entire batch will be rejected?

\[ U = 8000, \quad 9\% \text{ defective}, \quad 9190 \text{ good} = 0.91 \]

\[ \text{Accepted if all good,} \quad 8900 \text{ (91)} = 4702 \text{ defective,} \quad 4702 \text{ (5.52) = 0.529747} \]

\[ \text{Reject} \quad P(1+) = 1 - P(0) \]

\[ = 1 - 0.4702 \]

\[ = 0.529747 \]

6) The following table contains data from a study of two airlines which fly to Small Town, USA.

<table>
<thead>
<tr>
<th></th>
<th>Number of flights which were on time</th>
<th>Number of flights which were late</th>
</tr>
</thead>
<tbody>
<tr>
<td>Podunk Airlines</td>
<td>33</td>
<td>6</td>
</tr>
<tr>
<td>Upstate Airlines</td>
<td>43</td>
<td>5</td>
</tr>
</tbody>
</table>

\[ P(U|L) = \frac{p(U\cap L)}{P(L)} = \frac{5}{11} \]

If one of the 87 flights is randomly selected, find the probability that the flight selected is an Upstate Airlines flight given that it was late.

\[ P = \frac{5}{11} = 0.454545 = 0.4545 \text{ or } 0.4545 \]
Solve the problem.

7) How many ways can an IRS auditor select 6 of 11 tax returns for an audit?

\[ \binom{11}{6} n = \frac{462}{1} \]

Evaluate the expression.

8) \[ 8 \cdot 4 \cdot 2 = 70 \]

n = 70

Answer the question.

9) Suppose that computer literacy among people ages 40 and older is being studied and that the accompanying tables describes the probability distribution for four randomly selected people, where x is the number that are computer literate. Is it unusual to find four computer literates among four randomly selected people?

\[ \mu = 1.7 \]

\[ \sigma = 1.1 \]

\[ \sigma^2 = 1.3 \]

A) Yes

B) No

\[ \frac{\mu}{\sigma} = 1.6 \]

\[ \frac{\sigma}{\mu} = 0.65 \]

\[ \frac{\sigma^2}{\mu} = 0.76 \]

Assume that a researcher randomly selects 14 newborn babies and counts the number of girls selected, x. The probabilities corresponding to the 14 possible values of x are summarized in the given table. Answer the question using the table.

Probabilities of Girls

<table>
<thead>
<tr>
<th>x(girls)</th>
<th>P(x)</th>
<th>x(girls)</th>
<th>P(x)</th>
<th>x(girls)</th>
<th>P(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.000</td>
<td>5</td>
<td>0.122</td>
<td>10</td>
<td>0.061</td>
</tr>
<tr>
<td>1</td>
<td>0.001</td>
<td>6</td>
<td>0.183</td>
<td>11</td>
<td>0.022</td>
</tr>
<tr>
<td>2</td>
<td>0.006</td>
<td>7</td>
<td>0.209</td>
<td>12</td>
<td>0.006</td>
</tr>
<tr>
<td>3</td>
<td>0.022</td>
<td>8</td>
<td>0.183</td>
<td>13</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>0.061</td>
<td>9</td>
<td>0.122</td>
<td>14</td>
<td>0.000</td>
</tr>
</tbody>
</table>

P = 0.183

10) Find the probability of selecting exactly 8 girls.

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 = 11, x = 5, p = 0.5

\[ P(\text{Binomial}) \left( n, p, x \right) \]

\[ P(\text{Binomial}) \left( 11, 0.5, 5 \right) = \frac{22 \leq x \leq 9}{0.226} \]

P = 0.226
Find the mean, $\mu$, for the binomial distribution which has the stated values of $n$ and $p$. Round answer to the nearest tenth. Find the standard deviation, $\sigma$.

$$\mu = np = 29 \cdot 0.2 = 5.8$$

$$\sigma = \sqrt{npq} = \sqrt{29 \cdot 0.2 \cdot 0.8} = 2.2$$

Use the given values of $n$ and $p$ to find the minimum usual value $\mu - 2\sigma$ and the maximum usual value $\mu + 2\sigma$. Round your answer to the nearest hundredth unless otherwise noted.

Minimum: $1.4$  
Maximum: $10.2$

$$\mu - 2\sigma = 5.8 - 2(2.2) = 5.8 - 4.3 = 1.5$$

$$\mu + 2\sigma = 5.8 + 2(2.2) = 10.1$$

Determine if the outcome is unusual. Consider as unusual any result that differs from the mean by more than 2 standard deviations. That is, unusual values are either less than $\mu - 2\sigma$ or greater than $\mu + 2\sigma$.

13) According to AccuData Media Research, 36% of televisions within the Chicago city limits are tuned to "Eyewitness News" at 5:00 pm on Sunday nights. At 5:00 pm on a given Sunday, 2500 such televisions are randomly selected and checked to determine what is being watched. Would it be unusual to find that 950 of the 2500 televisions are tuned to "Eyewitness News"?

A) Yes
B) No

14) The number of calls received by a car towing service averages 14.4 per day (per 24-hour period). After finding the mean number of calls per hour, find the probability that in a randomly selected hour the number of calls is 2.

$$P(X) = \frac{e^{-\mu} \cdot \mu^x}{x!} = \frac{e^{-14.4} \cdot 14.4^2}{2!} = \frac{14.4^2 \cdot e^{-14.4}}{2} = 0.98786$$

$$P = 0.98786$$
Find the indicated probability.

6) The volumes of soda in quart soda bottles are normally distributed with a mean of 32.3 oz and a standard deviation of 1.2 oz. What is the probability that the volume of soda in a randomly selected bottle will be less than 32 oz?

\[ P(X < 32) = \Phi\left(\frac{32 - 32.3}{1.2}\right) = \Phi(-0.25) = 0.40129 \]

7) Suppose that replacement times for washing machines are normally distributed with a mean of 9.3 years and a standard deviation of 1.1 years. Find the probability that 70 randomly selected washing machines will have a mean replacement time less than 9.1 years.

\[ x = 9.3, \sigma = 1.1, n = 70 \]

\[ P(X < 9.1) = \Phi\left(\frac{9.1 - 9.3}{1.1}\right) = \Phi(-0.18) = 0.42868 \]

Use the binomial distribution.

8) 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 11 tosses. What is the probability of being correct 11 or more times by guessing? Does this probability seem to verify her claim?

\[ P(X \geq 11) = 1 - P(X < 11) = 1 - \sum_{k=0}^{10} \binom{20}{k} (0.5)^k (0.5)^{20-k} = 0.04109 \]

Yes

Solve the problem.

9) The following confidence interval is obtained for a population proportion, \( p \): (0.639, 0.663). Use these confidence interval limits to find the point estimate, \( \hat{p} \).

\[ \hat{p} = \frac{0.639 + 0.663}{2} = 0.651 \]

Assume that a sample is used to estimate a population proportion \( p \). Find the margin of error \( E \) that corresponds to the given statistics and confidence level. Round the margin of error to four decimal places.

10) 95% confidence; \( n = 2448, \hat{x} = 1763 \)

\[ E = z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96 \sqrt{\frac{0.67(0.33)}{2448}} = 0.0178 \]

Use the given data to find the minimum sample size required to estimate the population proportion.

11) Margin of error: 0.01; confidence level: 95%; from a prior study, \( \hat{p} \) is estimated by the decimal equivalent of 67%.

\[ n = \left( \frac{z_{\alpha/2}}{E} \right)^2 \hat{p}(1-\hat{p}) = \left( \frac{1.96}{0.01} \right)^2 \times 0.67 \times 0.33 = 8494 \]

Use the given degree of confidence and sample data to construct a confidence interval for the population proportion \( p \).

12) Of 84 adults selected randomly from one town, 62 have health insurance. Find a 90% confidence interval for the true proportion of all adults in the town who have health insurance.

\[ n = 84, x = 62 \]

Confidence interval: 0.659 < \( \hat{p} \) < 0.717
Find the indicated critical \( z \) value.

13) Find the critical value \( z_{\alpha/2} \) that corresponds to a 99% confidence level.

\[
\begin{align*}
\alpha/2 &= 0.005 \Rightarrow \text{norm}(0.995) & \Rightarrow z = z_{0.005} = 2.575 \\
\end{align*}
\]

Do one of the following, as appropriate: (a) Find the critical value \( z_{\alpha/2} \), (b) find the critical value \( t_{\alpha/2} \), (c) state that neither the normal nor the \( t \) distribution applies.

\[
\begin{align*}
\checkmark \quad 91\%: n = 45; \sigma \text{ is known}; \text{population appears to be very skewed.}
A) z_{\alpha/2} = 1.75 & \quad B) t_{\alpha/2} = 1.645 & \quad C) z_{\alpha/2} = 1.70 & \quad D) t_{\alpha/2} = 1.34
\end{align*}
\]

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.

15) Thirty randomly selected students took the calculus final. If the sample mean was 89 and the standard deviation was 6.2, construct a 99% confidence interval for the mean score of all students.

\[
\begin{align*}
\bar{x} - E < \mu < \bar{x} + E \\
E = t_{\alpha/2} \frac{s}{\sqrt{n}} \\
\therefore \quad Z_{1.96} \quad [85.86, 92.14] \quad [4]
\end{align*}
\]

Find the appropriate minimum sample size.

16) To be able to say with 95% confidence level that the standard deviation of a data set is within 10% of the population's standard deviation, the number of observations within the data set must be greater than or equal to what quantity?

\[
\begin{align*}
n = 192
\end{align*}
\]

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 one more decimal place than is used for the original set of data.

17) The daily intakes of milk (in ounces) for ten randomly selected people were:

\[
\begin{align*}
19.9 & \quad 20.2 & \quad 18.7 & \quad 31.5 & \quad 13.8 \\
12.6 & \quad 15.2 & \quad 31.4 & \quad 13.4 & \quad 24.5
\end{align*}
\]

\[
\begin{align*}
N = 10 \quad df = 9
\end{align*}
\]

Find a 99% confidence interval for the population standard deviation \( \sigma \).

\[
\begin{align*}
\sigma = 7.02263 \ldots \quad s^2 = 49.3173 \\
\left( \frac{(n-1)s^2}{\chi^2_{l}} \right)^{1/2} < \sigma < \left( \frac{(n-1)s^2}{\chi^2_{u}} \right)^{1/2} \\
\left( \frac{(9)(49.3173)}{23.569} \right)^{1/2} < \sigma < \left( \frac{(9)(49.3173)}{17.35} \right)^{1/2} \\
4.34 < \sigma < 15.99 \quad \therefore \quad 4.3 \leq \sigma \leq 15.99
\end{align*}
\]
Solve the problem.

1) Write the claim that is suggested by the given statement, then write a conclusion about the claim. Do not use symbolic expressions or formal procedures; use common sense.

Of a group of 1000 people suffering from arthritis, 500 receive acupuncture treatment and 500 receive a placebo. Among those in the placebo group, 24% noticed an improvement, while of those receiving acupuncture, 44% noticed an improvement.

The claim is that the proportion who noticed an improvement in the treatment group is greater than the proportion in the placebo group, i.e., acupuncture is more effective than the placebo. It is unlikely this is chance occurrence.

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.

2) \( z = \frac{1.34}{0.09} \)

Formulate the indicated conclusion in nontechnical terms. Be sure to address the original claim.

3) The manufacturer of a refrigerator system for beer kegs produces refrigerators that are supposed to maintain a true mean temperature, \( \mu \), of 45°F, ideal for a certain type of German pilsner. The owner of the brewery does not agree with the refrigerator manufacturer, and claims he can prove that the true mean temperature is incorrect. Assuming that a hypothesis test of the claim has been conducted and that the conclusion is to reject the null hypothesis, state the conclusion in nontechnical terms.

A) There is not sufficient evidence to support the claim that the mean temperature is equal to 45°F.

B) There is sufficient evidence to support the claim that the mean temperature is different from 45°F.

C) There is sufficient evidence to support the claim that the mean temperature is equal to 45°F.

D) There is not sufficient evidence to support the claim that the mean temperature is different from 45°F.

The sample data support the claim.

Find the critical value or values of \( \chi^2 \) based on the given information.

4) \( H_1: \sigma > 3.5 \)
\( n = 25 \)
\( \alpha = 0.05 \)

\( \chi^2 = \chi^2_{24,0.05} = 36.415 \)

Find the P-value for the indicated hypothesis test.

5) A manufacturer claims that fewer than 6% of its fax machines are defective. In a random sample of 97 such fax machines, 5% are defective. Find the P-value for a test of the manufacturer's claim.

\[ H_0: \quad p = 0.06 \]
\[ H_1: \quad p < 0.06 \text{ (original)} \]

\[ Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.05 - 0.06}{\sqrt{\frac{(0.06)(0.94)}{97}}} \]

\[ = -0.41 \text{ by hand} \]

Conclusion: There is not sufficient sample evidence to conclude that fewer than 6% of its fax machines are defective.

6) Test the claim that the mean age of the prison population in one city is less than 26 years.

Sample data are summarized as \( n = 25 \), \( \bar{x} = 24.4 \) years, and \( s = 9.2 \) years. Use a significance level of \( \alpha = 0.05 \).

\[ H_0: \quad \mu = 26 \]
\[ H_1: \quad \mu < 26 \text{ (original)} \]

\[ t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{24.4 - 26}{9.2/\sqrt{25}} \]

\[ = -1.711 \]

\[ P = .19658 \]

Conclusion: There is not sufficient sample evidence to support the claim that the mean age of the prison population in one city is less than 26 years.
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) Systolic blood pressure levels for men have a standard deviation of 19.7 mm Hg. A random sample of 51 women resulted in blood pressure levels with a standard deviation of 23.2 mm Hg. Use a 0.05 significance level to test the claim that blood pressure levels for women have the same variation as those for men.

\[ H_0 : \sigma_w = \sigma_m \quad (\text{ovl 1g}) \]
\[ H_1 : \sigma_w \neq \sigma_m \]

\[ \text{TS: } X^2 = 69.345 \]
\[ \text{CV: } X^2_L = 32.357 \quad X^2_R = 71.420 \]

\[ P = 0.07275 \]

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

2. Conclusion: There is not sufficient evidence to warrant rejection of the claim that blood pressure levels for women have the same variation as those for men.

\[ P = 2 \times X^2 \text{cdf} (L, \text{Ridf} \quad (69.345, 500, 50) = 0.07275 \quad 0.07275 \]

\[ = 0.07275 \quad 3 \text{ pts} \quad 0.0363 \quad 2 \text{ pts} \]
Find the number of successes \( x \) suggested by the given statement.

1) A computer manufacturer randomly selects 2280 of its computers for quality assurance and finds that 2.72% of these computers are found to be defective.

\[
\hat{p} = \frac{x}{n} = \frac{62}{620} = 0.16
\]

Use the traditional method to test the given hypothesis. Assume that the samples are independent and that they have been randomly selected.

2) A researcher finds that of 1000 people who said that they attend a religious service at least once a week, 31 stopped to help a person with car trouble. Of 1200 people interviewed who had not attended a religious service at least once a month, 20 stopped to help a person with car trouble. At the 0.05 significance level, test the claim that the two proportions are equal.

\[
H_0 : \hat{p}_1 = \hat{p}_2 \quad \text{(null)}
\]

\[
H_1 : \hat{p}_1 \neq \hat{p}_2 \quad \text{(alternative)}
\]

TS : \[ z = 2.22 \]

\[ z_{0.05/2} = \pm 1.96 \]

\[ p = 0.0261 \leq 0.05 = \alpha \]

1) Treat \( H_0 \) : Reject \( H_0 \)

2) **Conclusion:** There is sufficient evidence to warrant rejection of the claim that the two proportions are equal.
Perform the indicated hypothesis test. Assume that the two samples are independent simple random samples selected from normally distributed populations. Also assume that the population standard deviations are equal ($\sigma_1 = \sigma_2$), so that the standard error of the difference between means is obtained by pooling the sample variances.

3) A researcher was interested in comparing the response times of two different cab companies. Companies A and B were each called at 50 randomly selected times. The calls to company A were made independently of the calls to company B. The response times were recorded and the summary statistics were as follows:

<table>
<thead>
<tr>
<th></th>
<th>Company A</th>
<th>Company B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean response time</td>
<td>7.6 mins</td>
<td>6.9 mins</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1.4 mins</td>
<td>1.7 mins</td>
</tr>
</tbody>
</table>

Use a 0.02 significance level to test the claim that the mean response time for company A differs from the mean response time for company B. Use the $P$-value method of hypothesis testing.

$T$-test

$\bar{x}_A = 2.248$, $s_A = 0.50$, $n_A = 50$

$\bar{x}_B = 2.364$, $s_B = 0.64$, $n_B = 50$

$H_0 : \mu_A = \mu_B$

$H_1 : \mu_A \neq \mu_B$ (original)

$P = 0.2693 > 0.02 = \alpha$

1) Treat $H_0$ to reject $H_0$

2) Conclusion: There is not sufficient sample evidence to support the claim that the mean response time for company A differs from the mean response time for company B.

4) When testing for a difference between the means of a treatment group and a placebo group, the computer display below is obtained. Using a 0.01 significance level, is there sufficient evidence to support the claim that the treatment group (variable 1) comes from a population with a mean that is greater than the mean for the placebo population?

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>171.6292</td>
<td>168.7718</td>
</tr>
<tr>
<td>Known Variance</td>
<td>47.51672</td>
<td>41.08293</td>
</tr>
<tr>
<td>Observations</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Hypothesized Mean Difference</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>2.154</td>
<td></td>
</tr>
<tr>
<td>$P(T&gt;t)$ one-tail</td>
<td>0.0316</td>
<td></td>
</tr>
<tr>
<td>$T$ Critical one-tail</td>
<td>1.644853</td>
<td></td>
</tr>
<tr>
<td>$P(T=t)$ two-tail</td>
<td>0.0316</td>
<td></td>
</tr>
<tr>
<td>$t$ Critical two-tail</td>
<td>1.995961</td>
<td></td>
</tr>
</tbody>
</table>

1) Treat $H_0$ to reject $H_0$

2) Conclusion: There is not sufficient sample evidence to support the claim that the treatment group (variable 1) comes from a population with a mean that is greater than the mean for the placebo population.
Statistics 50 Test 4B Chapters 9 Spring 2012 Page 3.

Use the traditional method of hypothesis testing to test the given claim about the means of two populations. Assume that two dependent samples have been randomly selected from normally distributed populations.

5) A coach uses a new technique in training middle distance runners. The times for 8 different athletes to run 800 meters before and after this training are shown below.

<table>
<thead>
<tr>
<th>Athlete</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>Time before training (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>114.3</td>
<td>110.1</td>
<td>111.8</td>
<td>109</td>
<td>113.7</td>
<td>108</td>
<td>110.5</td>
<td>119.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>114.9</td>
<td>108.8</td>
<td>109.4</td>
<td>109.8</td>
<td>111.9</td>
<td>108.1</td>
<td>106.9</td>
<td>115.8</td>
<td></td>
</tr>
</tbody>
</table>

Using a 0.05 level of significance, test the claim that the training helps to improve the athletes' times for the 800 meters.

(2) TS: $t = 2.227$

(1) CV: $t_{0.05} = 1.895$

(2) P: $P = 0.0306 < 0.05 = \alpha$

H$_0$: $\mu_d = 0$

H$_1$: $\mu_d > 0$ (Original)

(1) Treat H$_0$: Reject H$_0$

(2) Conclusion: Conclude #3 The sample data support the claim that the training helps to improve the athletes' times for the 800 meters.

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.

(2) \[
\frac{2.12}{2.66} \leq \mu_d \leq \frac{2.606}{2.441},
\]

(6) When 25 randomly selected customers enter any one of several waiting lines, their waiting times have a standard deviation of 5.59 minutes. When 16 randomly selected customers enter a single main waiting line, their waiting times have a standard deviation of 2.15 minutes. Use a 0.05 significance level to test the claim that there is more variation in the waiting times when several lines are used.

(2) TS: $F = 6.76$

(2) CV: $F = 2.2878 F_{15,0.05}$

(2) P: $P = 1.9164 \times 10^{-4}$ $\alpha$, $0.0019164 < 0.05 = \alpha$

(1) Treat H$_0$: Reject H$_0$

(2) Conclusion: Conclude #3 The sample data support the claim that there is more variation in the waiting times when several lines are used.
Provide an appropriate response.

1) A regression equation is obtained for a set of data. After examining a scatter diagram, the researcher notices a data point that is potentially an influential point. How could she confirm that this data point is indeed an influential point?

Run the data with and without the potential influential point. See if the regression equation is quite different with the potential influential point. If it is, the point is an influential point.

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

2) $r = 0.434$, $n = 35$

A) Critical values: $r = \pm 0.335$

B) significant linear correlation

Find the value of the linear correlation coefficient $r$.

3) The paired data below consist of the temperatures on randomly chosen days and the amount a certain kind of plant grew (in millimeters):

<table>
<thead>
<tr>
<th>Temp</th>
<th>62</th>
<th>76</th>
<th>50</th>
<th>71</th>
<th>46</th>
<th>51</th>
<th>44</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth</td>
<td>36</td>
<td>39</td>
<td>50</td>
<td>13</td>
<td>33</td>
<td>33</td>
<td>17</td>
</tr>
</tbody>
</table>

$r = \frac{44}{7} = 6.286$

$r^2 = 0.200$

the standard error of estimate $= \frac{14.4}{7} = 2.057$

$\overline{X}, \overline{Y} = (56.4, 28.4)$

The equation of the regression line

$\hat{y} = -3.17 + 0.60 x$

Describe the error in the stated conclusion.

4) Given: There is a significant linear correlation between the number of homicides in a town and the number of movie theaters in a town.

Conclusion: Building more movie theaters will cause the homicide rate to rise.

Correlation does not imply causality.
Use the given data to find the best predicted value of the response variable.

5) Four pairs of data yield \( r = 0.942 \) and the regression equation \( \hat{y} = 3x \). Also, \( \bar{y} = 12.75 \).

What is the best predicted value of \( y \) for \( x = 4.8 \)?

\[
\begin{align*}
TS: \quad & r = 0.942 \\
\bar{y} - y &= b \cdot x - a \\
\bar{y} &= 4 \cdot 4.8 - .05 = .95 \quad \text{Use } \bar{y} = 12.75
\end{align*}
\]

Use the given information to find the coefficient of determination.

6) Find the coefficient of determination, given that the value of the linear correlation coefficient, \( r \), is -0.346.

\[
\begin{align*}
v^2 &= (0.346)^2 = 0.120
\end{align*}
\]

Use the given data to find the equation of the regression line. Round the final values to three significant digits, if necessary.

7) \[
\begin{array}{c|c|c|c|c|c|c|c|c|c}
x & 0 & 3 & 4 & 5 \\
y & 8 & 2 & 6 & 9 \\
\end{array}
\]

\( r = \) \quad \frac{3.79}{\text{The standard error of estimate}} = \quad \frac{3.79}{\text{The equation of the regression line}} = \quad \frac{6.25 + 0}{-0.95 + 0} \quad \text{H}_0: \quad \rho = 0 \\
\text{TS: } t = 0 \quad \text{CV: } t = 2.433 \quad \text{P: } 1 = 0.05 \quad \text{Treat } \text{H}_0:
\]

Conclusion: The correlation is not significant.

Is the data point, P, an outlier, an influential point, both, or neither?

8) The regression equation for a set of paired data is \( \hat{y} = 1.18 + 1.90x \). The values of \( x \) run from 3 to 20. A new data point, P(30, 60), is added to the set.

A) Influential point \( \quad \text{Outlier} \quad \text{B) Outlier} \quad \text{C) Both} \quad \text{D) Neither}

Use the computer display to answer the question.

9) 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

\[
\text{Score} = 31.55 + 10.90 \text{ Years.} \quad S = 5.651 \quad \text{R-Sq} = 83.0\% \quad \text{explained}
\]

What percentage of the total variation in test scores is unexplained by the linear relationship between years of study and test scores?

\[
p = \frac{17}{17} = 17\% \]
Perform the indicated goodness-of-fit test.

10) You roll a die 48 times with the following results.

<table>
<thead>
<tr>
<th>Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>3</td>
<td>12</td>
<td>15</td>
<td>14</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

Use a significance level of 0.05 to test the claim that the die is fair.

\[
\chi^2 = \frac{(O - E)^2}{E} = 8
\]

\[
\text{TS: } \chi^2 = 25.000 \quad x = 24.9
\]

\[
\text{CV: } \chi^2_{.05} = 11.071
\]

\[
P: \quad 1.3933 \leq 0.0013933
\]

1. Treat \( H_0: \) Reject \( H_0 \)
2. Conclusion: \( H_0 \) is not rejected. There is sufficient evidence to warrant rejection of the claim that the die is fair.

Use a \( \chi^2 \) test to test the claim that in the given contingency table, the row variable and the column variable are independent.

11) Responses to a survey question are broken down according to gender and the sample results are given below. At the 0.05 significance level, test the claim that response and gender are independent.

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
<th>Undecided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>25</td>
<td>35</td>
<td>15</td>
</tr>
<tr>
<td>Female</td>
<td>20</td>
<td>30</td>
<td>10</td>
</tr>
</tbody>
</table>

\[
H_0: \text{ Response and gender are independent (original)}
\]

1. \( H_1: \) Response and gender are dependent.
2. TS: \( \chi^2 = 5.79 \)
3. CV: \( \chi^2_{.05} = 5.991 \)
4. P: \( .07 < .05 = \alpha \)

1. Treat \( H_0: \) Fail to reject \( H_0 \)
2. Conclusion: \( H_0 \) is not rejected. There is not sufficient evidence to warrant rejection of the claim that response and gender are independent.
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.

1) Source | DF | SS | MS | F  | P  |
---|---|---|---|---|---|
Factor | 3 | 13.500 | 4.500 | 5.17 | 0.011 |
Error  | 16 | 13.925 | 0.870 | | |
Total  | 19 | 27.425 | | | |

Identify the value of the test statistic. \( F = 5.17 \)
Identify the p-value. \( P = 0.011 \)
Find the critical value. \( F_{0.05}^{3,16} = 3.2389 \)

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

2) Given the sample data below, test the claim that the populations have the same mean.

<table>
<thead>
<tr>
<th>Brand A</th>
<th>Brand B</th>
<th>Brand C</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 16</td>
<td>n = 16</td>
<td>n = 16</td>
</tr>
<tr>
<td>( \bar{x} = 3.48 )</td>
<td>( \bar{x} = 1.86 )</td>
<td>( \bar{x} = 2.84 )</td>
</tr>
<tr>
<td>s = 0.61</td>
<td>s = 0.45</td>
<td>s = 0.53</td>
</tr>
</tbody>
</table>

\( s_x = 0.6189 \)

(i) \( H_0: \mu_1 = \mu_2 = \mu_3 \) (or 0)

(ii) \( H_1: \) At least one different from others

TS: \( 37.3527 = F \)

\( F = \frac{\sum x^2 - \frac{\sum x^2}{n}}{\frac{\sum x^2}{n}} = \frac{16 (66575)}{(1.611^2 + (45)^2 + (53)^2)} \)

CV: \( F_{0.05}^{45,05} = 3.2317 \)

P: \( P = 2.7524 \times 10^{-10} = 37.3527 \times 10^{-5} \)

(i) Treat \( H_0: \) Reject \( H_0 \)

(ii) Conclusion: \# 1. There is sufficient evidence to warrant rejecting the claim that the proportions 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>Brands</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>15</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>14</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>23</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>22</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>28</td>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>

\[ H_0 : \mu_A = \mu_B = \mu_C \]
\[ H_1 : \text{At least one differs} \]
\[ \alpha = 0.025 \]
\[ P = \] Compare P to \( \alpha \)

\[ F_{16, 1025} = 46.867 \]

\[ F = 1.869 \]

\[ \text{Source of Variation} \]

<table>
<thead>
<tr>
<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>10.4749</td>
<td>5.2375</td>
<td>1.1869</td>
<td>0.3128</td>
</tr>
<tr>
<td>Error</td>
<td>448.2619</td>
<td>28.0164</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>458.7328</td>
<td>18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \bar{x}_A = \frac{6}{3} = 2 \]
\[ \bar{x}_B = \frac{7}{3} = 2.333 \]
\[ \bar{x}_C = \frac{6}{3} = 2 \]

\[ \text{Equation Development} \]

\[ F = \frac{\sum_i (\bar{x}_i - \bar{x})^2}{k-1} \]
\[ = \frac{6(19.5 - 20.5)^2 + 7(20.7 - 20.5)^2 + 6(20.3 - 20.5)^2}{3-1} \]
\[ = \frac{5(32.3) + 6(37.2) + 5(12.7)}{19-3} \]

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

2) Conclusion: There is not sufficient evidence to warrant rejection of the claim that the brands 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>Employee</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>31, 34, 32</td>
<td>29, 23, 22</td>
<td>21, 20, 24</td>
</tr>
<tr>
<td>II</td>
<td>19, 26, 22</td>
<td>35, 33, 30</td>
<td>25, 19, 23</td>
</tr>
<tr>
<td>III</td>
<td>21, 18, 26</td>
<td>20, 23, 24</td>
<td>36, 37, 31</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>1.19</td>
<td>0.595</td>
<td>0.669</td>
</tr>
<tr>
<td>EMPLOYEE</td>
<td>2</td>
<td>5.86</td>
<td>2.93</td>
<td>3.296</td>
</tr>
<tr>
<td>INTERACTION</td>
<td>4</td>
<td>710.81</td>
<td>177.702</td>
<td>19.9890</td>
</tr>
<tr>
<td>ERROR</td>
<td>18</td>
<td>160.02</td>
<td>8.89</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>26</td>
<td>877.88</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.

\[
\begin{align*}
H_0: & \text{ there is no interactive effect (original)} \\
H_1: & \text{ there is an interactive effect} \\
TS: & F = 19.9890 \\
CV: & F_{18, .05} = 2.9277 \\
P: & P = F_{cdf} \left( L_{1R}, df_{u}, df_{d} \right) = 1.9937 \times 10^{-6} \\
\end{align*}
\]

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

Conclusion: There is sufficient evidence to warrant rejection of the claim that the interaction between employee and machine has no effect on the number of items produced.
6) Use the sample data below to find the rank correlation coefficient and test the claim of correlation between math and verbal scores. Use a significance level of 0.05.

<table>
<thead>
<tr>
<th>Mathematics</th>
<th>2 7 1 6 3 4 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>347 440 327 456 427 349 377 398</td>
</tr>
<tr>
<td>Verbal</td>
<td>285 378 243 371 340 271 294 322</td>
</tr>
<tr>
<td></td>
<td>3 8 1 7 6 2 4 5</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\alpha & = 0.05 \\
\beta & = 0.05 \\
\end{align*}
\]

\[
\alpha^2 = 0.01 \\
\beta^2 = 0.01
\]

\[
\begin{align*}
R_s &= 1 - \frac{6 \sum d^2}{n(n^2 - 1)} \\
&= 1 - \frac{6 \times 4}{8(8^2 - 1)} \\
&= 1 - \frac{24}{8(63)} = 1 - \frac{1}{2.1} \\
&= 0.952 \approx 0.952
\end{align*}
\]

\[H_0: \rho_s = 0\]

\[H_1: \rho_s \neq 0\]

TS: \[R_s = 0.952\]

CV: \[R_s = 0.738\]

1. Treat \(H_0: \) reject \(H_0\)
2. Conclusion: The rank correlation is significant.

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

7) Given that the rank correlation coefficient, \(r_s\), for 20 pairs of data is 0.528, test the claim of correlation between the two variables. Use a significance level of 0.05.

\[TS: \ R_s = 0.528\]

\[CV: \ R_s = \pm 0.447\]

1. Treat \(H_0: \) reject \(H_0\)
2. Conclusion: The rank correlation is significant.