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

4)  1) Survey responses of "good, better, best".
    A) Ratio  B) Nominal  C) Interval  D) Ordinal

Use critical thinking to address the key issue.

4) A researcher published this survey result: "74% of people would be willing to spend 10 percent more for energy from a non-polluting source". The survey question was announced on a national radio show and 1,200 listeners responded by calling in. What is wrong with this survey?
   **Self Selecting Survey. Not Necessarily Representative. We Can't Generalize to the Population.**
   **Voluntary = Potential for Bias**

Determine whether the given description corresponds to an observational study or an experiment.

4) A quality control specialist compares the output from a machine with a new lubricant to the output of machines with the old lubricant.
   A) Experiment  B) Observational study

Use the pie chart to solve the problem.

4) The pie chart shows the percent of the total population of 78,100 of Springfield living in the given types of housing. Round your result to the nearest whole number.

- Single family: 39%  
- Apts: 32%  
- Duplex: 4%  
- Townhouse: 6%  
- Condo: 19%

Find the number of people who live in Apts:

\[ n = \frac{78100 \times 32}{100} = 24992 \]

Use the empirical rule to solve the problem.

5) At one college, GPA's are normally distributed with a mean of 2.6 and a standard deviation of 0.4. What percentage of students at the college have a GPA between 2.2 and 3?

\[ P = 68.7\% \]
6) A television manufacturer sold three times as many televisions in 2005 as it did in 1995. To illustrate this fact, the manufacturer draws a graph as shown below. The television on the right is three times as tall and three times as wide as the television on the left. Why is this graph misleading? What visual impression is created by the graph?

Find the mean for the given sample data. Unless indicated otherwise, round your answer to one more decimal place than is present in the original data values.

\[
\text{Find the mean for the given sample data.} \quad \frac{52 + 68 + 64 + 50 + 64 + 58 + 60 + 50}{8} = 57.3
\]

Find the midrange for the given sample data.

\[
\text{Find the midrange for the given sample data.} \quad \frac{50 + 68}{2} = 59
\]

Find the mode(s) for the given sample data.

\[
\text{Find the mode(s) for the given sample data.} \quad 50
\]

Find the median for the given sample data.

\[
\text{Find the median for the given sample data.} \quad 58
\]

Find the range for the given data.

\[
\text{Find the range for the given data.} \quad 18
\]

Find the variance for the given data.

\[
\text{Find the variance for the given data.} \quad 50
\]

Find the standard deviation for the given sample data.

\[
\text{Find the standard deviation for the given sample data.} \quad 7.1
\]

Solve the problem.

7) Last year, nine employees of an electronics company retired. Their ages at retirement are listed below. Find the mean retirement age.

\[
\text{Find the mean for the given sample data.} \quad \frac{52 + 68 + 64 + 50 + 64 + 58 + 60 + 50}{8} = 57.3
\]

Find the midrange for the given sample data.

\[
\text{Find the midrange for the given sample data.} \quad \frac{50 + 68}{2} = 59
\]

Find the mode(s) for the given sample data.

\[
\text{Find the mode(s) for the given sample data.} \quad 50
\]

Find the median for the given sample data.

\[
\text{Find the median for the given sample data.} \quad 58
\]

Find the range for the given data.

\[
\text{Find the range for the given data.} \quad 18
\]

Find the variance for the given data.

\[
\text{Find the variance for the given data.} \quad 50
\]

Find the standard deviation for the given sample data.

\[
\text{Find the standard deviation for the given sample data.} \quad 7.1
\]
Find the mean of the data summarized in the given frequency distribution.

The test scores of 40 students are summarized in the frequency distribution below. Find the mean score.

<table>
<thead>
<tr>
<th>Score</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>50-59</td>
<td>8</td>
</tr>
<tr>
<td>60-69</td>
<td>7</td>
</tr>
<tr>
<td>70-79</td>
<td>10</td>
</tr>
<tr>
<td>80-89</td>
<td>6</td>
</tr>
<tr>
<td>90-99</td>
<td>9</td>
</tr>
</tbody>
</table>

Find the mean for the given sample data. **74.8**

Find the variance for the given data. **217.6**

Find the standard deviation for the given sample data. **14.4**

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

The weekly salaries (in dollars) of 24 randomly selected employees of a company are shown below. Construct a boxplot for the data set.

310 320 450 460 470 500 320 340
580 600 650 700 710 840 370 900
1000 1200 1250 1300 1400 1720 2500 3700

A)

B)

C)

D)
Find the indicated measure.

11) Use the given sample data to find Q₃.
   \[
   \begin{align*}
   &49 \ 52 \ 52 \ 74 \ 67 \ 35 \ 55 \ 55 \ 55 \ 61 \ 74 \\
   &\frac{55 + 67}{2} = 61
   \end{align*}
   \]

Find the percentile for the data value.

12) Data set: 12 9 6 4 6 12 6 4 12 2 12 15 9 6 12;
    data value: 6
    \[
    \frac{7}{10} = 0.7 \quad P = 35.05112
    \]

Solve the problem.

13) The data set below consists of the scores of 15 students on a quiz. For this data set, which measure of variation do you think is more appropriate, the range or the standard deviation? Explain your thinking:
   90 90 91 91 89
   90 89 91 91 90
   68 90 91 90 91

Determine which score corresponds to the higher relative position.

14) Which score has the highest relative position: a score of 51.5 on a test for which \( \bar{x} = 47 \) and \( s = 9 \), a score of 5.9 on a test for which \( \bar{x} = 4.2 \) and \( s = 1.2 \) or a score of 460.8 on a test for which \( \bar{x} = 444 \) and \( s = 42 \)?
   A) A score of 51.5
   B) A score of 460.8
   C) A score of 5.9

\[
\begin{align*}
Z &= \frac{X - \bar{X}}{S} \\
Z &= \frac{51.5 - 47}{9} = 0.5555555555555556 = 0.56
\end{align*}
\]

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 weight of 99 pounds among a population having a mean weight of 162 pounds and a standard deviation of 24.3 pounds.

\[
\begin{align*}
Z &= \frac{X - \bar{X}}{S} \\
Z &= \frac{99 - 162}{24.3} = -2.59
\end{align*}
\]
Statistics 50 SPRING 2011
Test2 Chapters 4-5 Time 7:00 8:20 11:00 Seat Letter Number

Directions: Circle the correct choice for each response set. If required, show calculations in the blank spaces near the problems.

Find the indicated probability.

✓ 1) A bag contains 4 red marbles, 3 blue marbles, and 7 green marbles. If a marble is randomly selected from the bag, what is the probability that it is blue?

\[ P = \frac{3}{14} \]

Answer the question.

✓ 2) Find the odds against correctly guessing the answer to a multiple choice question with 7 possible answers.

A) 7 : 1 B) 7 : 6 C) 6 : 7 D) 6 : 1

✓ 3) 100 employees of a company are asked how they get to work and whether they work full time or part time. The figure below shows the results. If one of the 100 employees is randomly selected, find the probability that the person drives alone or cycles to work.

Provide a written description of the complement of the given event.

✓ 5) Of ten adults, at least one of them has high blood pressure.

A) All of the adults have high blood pressure.
B) None of the adults have high blood pressure
C) Nine of the adults have high blood pressure.
D) At most one of the adults has high blood pressure.

Find the indicated probability. Round to the nearest thousandth.

✓ 6) A study conducted at a certain college shows that 52% of the school's graduates find a job in their chosen field within a year after graduation. Find the probability that among 8 randomly selected graduates, at least one finds a job in his or her chosen field within a year of graduating.

A) 0.125 B) 0.995 C) 0.520 D) 0.997
Statistics 50  Test 2  Chapters 4-5  Spring 2011  Page 2.

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

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

<table>
<thead>
<tr>
<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>
</tr>
<tr>
<td>between 21 and 40</td>
<td>35</td>
<td>20</td>
</tr>
<tr>
<td>over 40 years of age</td>
<td>20</td>
<td>30</td>
</tr>
</tbody>
</table>

\[
P(40 \oplus 1 | R \oplus B) = \frac{30}{75} = \frac{2}{5}
\]

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

P = \frac{2}{5}

8) 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.

\[
01120011002101021201203102101
\]

Estimate the probability that sales will remain steady.

A) 0.125  B) 0.419  C) 0.412  (D) 0.355

9) A state lottery involves the random selection of six different numbers between 1 and 27. If you select one six number combination, what is the probability that it will be the winning combination?

\[
\frac{1}{27C_6} = \frac{1}{29,360,100} = \frac{3.3782 \times 10^{-6}} = P = 0.00000338
\]

Find the mean of the given probability distribution.

10) The probabilities that a batch of 4 computers will contain 0, 1, 2, 3, and 4 defective computers are 0.4979, 0.3793, 0.1084, 0.0138, and 0.0007, respectively. Round answer to the nearest hundredth.

\[
\mu = \frac{0.4979 + 0.3793 + 0.1084 + 0.0138 + 0.0007}{5} = 0.173  \quad \sigma = \frac{1.73}{1.73}
\]

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(\text{girls}) )</th>
<th>( P(x) )</th>
<th>( x(\text{girls}) )</th>
<th>( P(x) )</th>
<th>( x(\text{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>

11) Find the probability of selecting exactly 5 girls.

\[
P = \frac{122}{122}
\]
Statistics 50  Test2  Chapters 4-5  Spring 2011  Page 3.

Determine whether the given procedure results in a binomial distribution. If not, state the reason why.

12) Choosing 5 people (without replacement) from a group of 34 people, of which 15 are women, keeping track of the number of men chosen.
   \[A\) Not binomial: the trials are not independent.\]
   \[B\) Not binomial: there are more than two outcomes for each trial.\]
   \[C\) Procedure results in a binomial distribution.\]
   \[D\) Not binomial: there are too many trials.\]

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.

\[13) n = 4, x = 3, p = \frac{1}{6} \]

\[
\binom{n}{x} \left( p \right)^x \left( 1 - p \right)^{n-x} = 0.154
\]

\[
p = 0.154
\]

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

\[14) n = 2142; p = 0.63 \]

\[
\mu = np = 1349.46
\]

\[
(A) \mu = 1349.5 \hspace{1cm} (B) \mu = 1353.2 \hspace{1cm} (C) \mu = 1341.0 \hspace{1cm} (D) \mu = 1344.2
\]

\[
\mu + 2\sigma = 1349.41 \text{ or } 1349.8 \hspace{1cm} \mu - 2\sigma = 1344.7 \text{ or } 1349.2
\]

\[
\sigma = 22.3
\]

\[
1349.5 \pm 2(22.3)
\]

\[
\sqrt{n} = \sqrt{2142} = 223.45
\]

Use the Poisson Distribution to find the indicated probability.

15) Sunita's job is to provide technical support to computer users. Suppose the arrival of calls can be modeled by a Poisson distribution with a mean of 4.7 calls per hour. What's the probability that in the next 10 minutes there will be 2 calls?

\[
\mu = \frac{4.7}{6} \hspace{1cm} p = \frac{1}{140}
\]

\[
\text{Poisson}\left( \mu, x \right) = \left( \frac{4.7}{6} \right)^2 = 0.14017
\]
Find the area of the shaded region. The graph depicts the standard normal distribution with mean 0 and standard deviation 1.

\[ \text{Normalcdf}(-2.26, 1.13) = 0.5 \]

Find the indicated z score. The graph depicts the standard normal distribution with mean 0 and standard deviation 1.

\[ \text{invnorm}(0.4013) = -0.25 \]

Solve the problem.

3) For a standard normal distribution, find the percentage of data that are more than 1 standard deviation away from the mean.

\[ 1 - 68.26\% = 31.74\% \quad \text{(A)} \]

Solve the problem. Round to the nearest tenth unless indicated otherwise.

4) Scores on an English test are normally distributed with a mean of 33.8 and a standard deviation of 8.5. Find the score that separates the top 59% from the bottom 41%.

\[ X = 31.9 \]

5) In one region, the September energy consumption levels for single-family homes are found to be normally distributed with a mean of 1050 kWh and a standard deviation of 218 kWh. If 50 different homes are randomly selected, find the probability that their mean energy consumption level for September is greater than 1075 kWh.

\[ P = \text{normalcdf}(1075, 2000, 1050, 218/\sqrt{50}) \]

\[ P = 0.2087 \]
Use the binomial distribution.

6) 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 14 tosses. What is the probability of being correct 14 or more times by guessing? Does this probability seem to verify her claim?

\[ P = \frac{0.5}{6} \leq 0.6 \]

\[ P(14^+) = 1 - P(14^-) \\
= 1 - \text{binomcdf}(n, p, x) \]

\[ (20, 0.5, 14) \]

\[ \approx 0.05765 < 0.05 \]

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

7) Of 375 randomly selected medical students, 30 said that they planned to work in a rural community. Find a 95% confidence interval for the true proportion of all medical students who plan to work in a rural community.

\[ \hat{p} = \frac{30}{375} \approx 0.08 \]

\[ \hat{p} = 0.082 \pm 1.96 \times \sqrt{0.08 \times 0.92 / 375} \]

\[ 0.075 < \hat{p} < 0.107 \]

Solve the problem.

8) The following confidence interval is obtained for a population proportion, \( p \): 0.883 < \( p \) < 0.911. Use these confidence interval limits to find the margin of error, \( E \).

\[ \hat{p} = 0.883 \]

\[ l = 0.883 - 0.023 \]

\[ 0.860 \]

\[ u = 0.883 + 0.023 \]

\[ 0.906 \]

\[ E = 0.014 \]

Use the confidence level and sample data to find a confidence interval for estimating the population \( \mu \). Round your answer to the same number of decimal places as the sample mean.

9) Test scores: \( n = 105, \bar{x} = 70.5, \sigma = 6.8 \); 99% confidence

\[ (68.791, 72.209) \]

\[ (68.8, 72.2) \]

\[ 68.8 < \mu < 72.2 \]
Assume that a sample is used to estimate a population mean $\mu$. Use the given confidence level and sample data to find the margin of error. Assume that the sample is a simple random sample and the population has a normal distribution. Round your answer to one more decimal place than the sample standard deviation.

10) 95% confidence; $n = 21; \bar{x} = 0.53; s = 0.53$

$$E = \frac{t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}}{1 - \frac{1}{n}} = \frac{(2.086)(0.53)}{\sqrt{21}} = 0.241$$

$$\bar{x} - E = 0.241$$

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) To find the standard deviation of the diameter of wooden dowels, the manufacturer measures 19 randomly selected dowels and finds the standard deviation of the sample to be $s = 0.16$. Find the 95% confidence interval for the population standard deviation $\sigma$.

$$\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}} < \sigma < \frac{(n-1)s^2}{\chi^2_{1 - \frac{\alpha}{2}}}$$

$$\frac{(18)(0.16)^2}{31.526} < \sigma < \frac{(18)(0.16)^2}{8.231}$$

$$0.1208 < \sigma < 0.2366$$

$$0.12 < \sigma < 0.24$$

Provide an appropriate response.

12) A paper published the results of a poll. It stated that, based on a sample of 1000 married men, 51% of married men say that they would marry the same woman again. The margin of error was given as $\pm 3$ percentage points and the confidence level was given as 95%. What does it mean that the margin of error was $\pm 3$ percentage points?

We are 95% certain $0.48 < p < 0.54$

We are 95% confident the true percentage is within 3% of 51%.
Statistics 50 SPRING 2011
Test4A Chapters 8  Time 7:00 8:20 11:00  Seat  Letter  Number

Solve the problem.

✓ 1) What do you conclude about the claim below? Do not use formal procedures or exact calculations. Use only the rare event rule and make a subjective estimate to determine whether the event is likely.

Claim: An employee of a company is equally likely to take a sick day on any day of the week. Last year, the total number of sick days taken by all the employees of the company was 143. Of these, 52 were Mondays, 14 were Tuesdays, 17 were Wednesdays, 17 were Thursdays, and 43 were Fridays.

\[ H_0: p_M = p_T = p_W = p_{Th} = p_F \]

\[ \frac{52}{143}, \frac{14}{143}, \frac{17}{143}, \frac{17}{143}, \frac{43}{143} \]

Probabilities on Monday = Friday are approximately 3 times Tuesday, Wednesday, and Thursday. It is quite unlikely this is a chance happening, the claim of the employee is not correct.

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 \( H_1: p = 3/4 \), the test statistic is \( z = 0.77999997 \).

(A) 0.43540001; reject the null hypothesis
(B) 0.43540001; fail to reject the null hypothesis
(C) 0.2177; reject the null hypothesis
(D) 0.2177; reject the null hypothesis

\[ P = 2 \text{ NormalCDF}( z = 1.2) \]

\[ = 0.77999997 \]

Assume that a hypothesis test of the given claim will be conducted. Identify the type I or type II error for the test.

✓ 3) A medical researcher claims that 20% of children suffer from a certain disorder. Identify the type I error for the test.

(A) Fail to reject the claim that the percentage of children who suffer from the disorder is equal to 20% when that percentage is actually 20%.

(B) Reject the claim that the percentage of children who suffer from the disorder is different from 20% when that percentage really is different from 20%.

(C) Fail to reject the claim that the percentage of children who suffer from the disorder is equal to 20% when that percentage is actually different from 20%.

(D) Reject the claim that the percentage of children who suffer from the disorder is equal to 20% when that percentage is actually 20%.
Determine whether the hypothesis test involves a sampling distribution of means that is a normal distribution, Student t distribution, or neither.

4) Claim: \( \mu = 959 \). Sample data: \( n = 25, x = 951, s = 25 \). The sample data, for this simple random sample, appear to come from a normally distributed population with \( \sigma = 28 \).

A) Neither    B) Normal    C) Student t

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.

\[ n > 30 \quad S \quad \alpha = .05 \quad \text{for a 2 tailed test} \]

\[ Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \]

Find the P-value for the indicated hypothesis test.

6) In a sample of 47 adults selected randomly from one town, it is found that 9 of them have been exposed to a particular strain of the flu. Find the P-value for a test of the claim that the proportion of all adults in the town that have been exposed to this strain of the flu is 8%.

\[ H_0 : P = .08 \quad \text{(original)} \]
\[ H_1 : P \neq .08 \]

1. Prop Z test.

\[ n = 47 \quad \bar{x} = 9 \]

\( \alpha = .05 \quad 2t \)

\[ Z = 2.82 \]

\[ \pm 1.96 \]

\[ P = .004 \quad \text{If } .05 = \alpha \]

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

2. Conclusion: There is sufficient evidence to warrant rejection of the claim that the proportion of all adults in that town that have been exposed to this strain of the flu is 8%.
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.

1. Test the claim that for the population of female college students, the mean weight is given by $\mu = 132$ lb. Sample data are summarized as $n = 20$, $\bar{x} = 137$ lb, and $s = 14.2$ lb. Use a significance level of $\alpha = 0.1$.

   **Hypothesis:**
   - $H_0 : \mu = 132$ (Original)
   - $H_1 : \mu \neq 132$

   **Test Statistic (TS):**
   $$ t = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{137 - 132}{14.2 / \sqrt{20}} $$

   **Critical Value (CV):**
   $$ t = \pm 4.75 $$

   **P-Value (P):**
   $$ P = 0.0358 $$

1. **Treat $H_0$: Fail to reject $H_0$**

2. **Conclusion:** There is not sufficient evidence to reject the claim that the mean weight is $\mu = 132$ lb.

---

3. **Conclusion: Test of a computer component.**

   In tests of a computer component, it is found that the mean time between failures is 520 hours. A modification is made which is supposed to increase the time between failures. Tests on a random sample of 10 modified components resulted in the following times (in hours) between failures:
   
   518 548 561 523 536 499 538 557 528 563
   
   At the 0.05 significance level, test the claim that for the modified components, the mean time between failures is greater than 520 hours. Use the P-value method of testing hypotheses.

   **Hypothesis:**
   - $H_0 : \mu = 520$
   - $H_1 : \mu > 520$ (Original)

   **Test Statistic (TS):**
   $$ t = 2.612 $$

   **Critical Value (CV):**
   $$ t = 1.833 $$

   **P-Value (P):**
   $$ P = 0.0409 < 0.05 = \alpha $$

1. **Treat $H_0$: Reject $H_0$**

2. **Conclusion:** The sample data supports the claim that the mean time between failures is greater than 520 hours.
Provide an appropriate response.

1) A set of data consists of the number of years that applicants for foreign service jobs have studied German and the grades that they received on a proficiency test. The following regression equation is obtained: \( \hat{y} = 31.6 + 10.9x \), where \( x \) represents the number of years of study and \( y \) represents the grade on the test. Identify the predictor and response variables.

Predictor: \( x \) years of study
Response: \( y \) grade

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) \( r = -0.844 \), \( n = 7 \)

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

3) \( r = -0.844 \)

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

Conclusion: The correlation is significant.

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

3) The paired data below consist of the test scores of 6 randomly selected students and the number of hours they studied for the test.

<table>
<thead>
<tr>
<th>Hours</th>
<th>5 10 4 6 10 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>64 85 69 86 59 87</td>
</tr>
</tbody>
</table>

\( r = 0.876 \)

4) The standard error of estimate = 14.1

5) \((\bar{x}, \bar{y}) = (7, 75.2)\)

The equation of the regression line is \( \hat{y} = 72.1 + 43.8x \)

Describe the error in the stated conclusion.

4) Given: There is no significant linear correlation between scores on a math test and scores on a verbal test. Conclusion: There is no relationship between scores on the math test and scores on the verbal test.

Just because there is no linear relationship does not imply there is no relation ship.

Use the given data to find the best predicted value of the response variable.

5) Eight pairs of data yield \( r = 0.706 \) and the regression equation \( \hat{y} = 55.8 + 2.79x \). Also, \( \bar{y} = 71.125 \). What is the best predicted value of \( y \) for \( x = 5.7 \)?

\[ \overline{y} = 71.125 \]

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

6) \[ \begin{array}{c|cccc} x & 0 & 3 & 4 & 5 \\ \hline y & 12 & 8 & 6 & 9 \end{array} \]

\[ r = 0.706 \]

\[ \bar{y} = 7.0 \]

\[ \bar{x} = 3.7 \]

\[ \text{the standard error of estimate} = 3.20 \]

\[ (\bar{x}, \bar{y}) = (3.7, 7.0) \]

The equation of the regression line is \( \hat{y} = 5.36 + 3.43x \).

Use the given information to find the coefficient of determination.

7) A regression equation is obtained for a collection of paired data. It is found that the total variation is 29.045, the explained variation is 15.212, and the unexplained variation is 13.833. Find the coefficient of determination.

\[ r^2 = \frac{15.212}{29.045} = 0.523739 \]

\[ r^2 = 0.524 \]

Find the standard error of estimate for the given paired data.

8) The equation of the regression line for the paired data below is \( \hat{y} = 3x \). Find the standard error of estimate.

\[ \begin{array}{c|cccc} x & 2 & 4 & 5 & 6 \\ \hline y & 7 & 11 & 13 & 20 \end{array} \]

A) 5.00  B) 4.1892  C) 6.2750  D) 2.2361

Provide an appropriate response.

9) The following residual plot is obtained after a regression equation is determined for a set of data. Does the residual plot suggest that the regression equation is a bad model? Why or why not?

\[ \text{No, no noticeable pattern. Enough} \]

\[ \text{Doesn't become thinner to thinner implies consistent standard deviation} \]
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

\[
\text{Score} = 31.55 + 10.90 \text{Years.}
\]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>StDev</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>

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

Predicted values

<table>
<thead>
<tr>
<th>Fit</th>
<th>StDev</th>
<th>95.0% CI</th>
<th>95.0% PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>53.3</td>
<td>3.168</td>
<td>(42.72, 63.98)</td>
<td>(31.61, 75.09)</td>
</tr>
</tbody>
</table>

For a person who studies for 2 years, obtain the 95% prediction interval and write a statement interpreting the interval.

A) (31.61, 75.09): We can be 95% confident that the test score of an individual who studies 2 years will lie in the interval (31.61, 75.09)

B) (42.72, 63.98): We can be 95% confident that the test score of an individual who studies 2 years will lie in the interval (42.72, 63.98)

C) (31.61, 75.09): We can be 95% confident that the mean test score of all individuals who study 2 years will lie in the interval (31.61, 75.09)

D) (42.72, 63.98): We can be 95% confident that the mean test score of all individuals who study 2 years will lie in the interval (42.72, 63.98)
Perform the indicated goodness-of-fit test.

11) A company manager wishes to test a union leader's claim that absences occur on the different week days with the same frequencies. Test this claim at the 0.05 level of significance if the following sample data have been compiled.

<table>
<thead>
<tr>
<th>Day</th>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thur</th>
<th>Fri</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absences</td>
<td>37</td>
<td>15</td>
<td>12</td>
<td>23</td>
<td>43</td>
</tr>
</tbody>
</table>

\[
H_0: \mu_M = \mu_T = \mu_W = \mu_Th = \mu_F \quad \text{or all the same} \\
H_1: \quad \text{At least one is different from other or not all same.}
\]

\[
\chi^2 = 28.308
\]

\[
\chi^2 = 9.488
\]

\[
P = 0.0804 \quad E^{-5} \leq 0.05 = \alpha
\]

P < \alpha

\[0\]

1. Treat \( H_0 \): \textbf{Reject} \( H_0 \)

2. Conclusion: \textbf{There is sufficient evidence to warrant rejection of the claim that absences occur on different week days with the same frequency.}

Solve the problem.

12) Use the sample data below to test whether car color affects the likelihood of being in an accident. Use a significance level of 0.01.

<table>
<thead>
<tr>
<th>Car has been in accident</th>
<th>Red</th>
<th>Blue</th>
<th>White</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car has not been in accident</td>
<td>28</td>
<td>33</td>
<td>36</td>
</tr>
<tr>
<td>Car has been in accident</td>
<td>23</td>
<td>22</td>
<td>30</td>
</tr>
</tbody>
</table>

\[
H_0: \quad \text{Vehicle color and accidents independent}
\]

\[
H_1: \quad \text{Vehicle color and accidents dependent}
\]

\[
\chi^2 = 4.29
\]

\[
\chi^2 = 9.210
\]

\[
P = 0.00706 > 0.01 = \alpha
\]

1. Treat \( H_0 \): \textbf{Fail to reject} \( H_0 \)

2. Conclusion: \textbf{There is not sufficient evidence to warrant rejection of the claim that color does not affect the likelihood of an accident.}
1) What can you conclude about the equality of the population means?

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
<td>3</td>
<td>13.500</td>
<td>4.500</td>
<td>5.17</td>
<td>0.011</td>
</tr>
<tr>
<td>Error</td>
<td>16</td>
<td>13.925</td>
<td>0.870</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>27.425</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ F = \frac{MS_{\text{Factor}}}{MS_{\text{Error}}} = \frac{5.17}{0.870} = 5.97 \]

A) Reject the null hypothesis since the p-value is greater than the significance level.
B) Accept the null hypothesis since the p-value is greater than the significance level.
C) Accept the null hypothesis since the p-value is less than the significance level.
D) Reject the null hypothesis since the p-value is less than the significance level.

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. Use a significance level of 0.05.

<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} = 2.84 )</td>
<td>( \bar{x} = 2.09 )</td>
<td>( \bar{x} = 1.86 )</td>
</tr>
<tr>
<td>( s = 0.53 )</td>
<td>( s = 0.37 )</td>
<td>( s = 0.45 )</td>
</tr>
</tbody>
</table>

\[ F = \frac{\sum x^2}{n} - \frac{\sum (\bar{x})^2}{n} = \frac{16 (2.6263)}{20} = 2.328 \]

\[ F = \frac{s_{\text{pooled}}^2}{s_{\text{pooled}}^2} = \frac{16 (0.26263)}{2.0676} = 2.328 \]

1. \( H_0: M_A = M_B = M_C \) or \( M_1 = M_2 = M_3 \) (original)
2. \( H_1: \) At least one different.
3. TS: \( F = 2.328 \)
4. CV: \( F^{2}_{45,0.05} = 3.2317 \)
5. P: \( 5.1454 \leq \) 0.05

Treat \( H_0: \) reject \( H_0 \) \( P = F_{df}(1, R, df_1, df_2) \)

Conclusion: There is sufficient evidence to warrant rejection of the claim that the populations have the same mean.
8) 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>20</td>
<td>21</td>
</tr>
<tr>
<td>25</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td>21</td>
<td>22</td>
<td>20</td>
</tr>
<tr>
<td>23</td>
<td>23</td>
<td>19</td>
</tr>
<tr>
<td>22</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \overline{x} = \left( \frac{6(21) + 4(20.5) + 5(20)}{15} \right) = \frac{306}{15} = 20.5 \]

\[ H_0: \mu_A = \mu_B = \mu_C \]

\[ F = \frac{MS_A}{MS_E} = \frac{1.35}{0.75} = 1.8 \]

\[ \overline{x} = 20.5 \]

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>2.75</td>
<td>2</td>
<td>1.35</td>
<td>1.8</td>
<td>.8342</td>
</tr>
<tr>
<td>Error</td>
<td>9.75</td>
<td>12</td>
<td>0.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>12.5</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ F = \frac{S_A}{k-1} \frac{(x_i - \overline{x})^2}{(n_i-1)S^2} \]

\[ = \frac{6(21-20.5)^2 + 4(20.5-20.5)^2 + 5(20-20.5)^2}{15} \]

\[ = \frac{5(11.4) + 3(0) + 4(0.09)}{15} \]

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

2. Conclusion: \( H_0 \) There is not sufficient evidence to warrant rejection of the claim that the brands have the same mean.
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>16, 18, 19</td>
<td>15, 17, 20</td>
<td>14, 18, 16</td>
</tr>
<tr>
<td>Machine</td>
<td>II</td>
<td>20, 27, 29</td>
<td>25, 28, 27</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>15, 18, 17</td>
<td>16, 16, 19</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>588.74</td>
<td>294.37</td>
<td>53.7172</td>
</tr>
<tr>
<td>EMPLOYEE</td>
<td>2</td>
<td>2.08</td>
<td>1.04</td>
<td>18.98</td>
</tr>
<tr>
<td>INTERACTION</td>
<td>4</td>
<td>15.48</td>
<td>3.87</td>
<td>7.062</td>
</tr>
<tr>
<td>ERROR</td>
<td>18</td>
<td>98.64</td>
<td>5.48</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>26</td>
<td>704.94</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5) Assume that the number of items produced is not affected by an interaction between employee and machine. Using a 0.05 significance level, test the claim that the machine has no effect on the number of items produced.

- $H_0$: Machine has no effect (original)
- $H_1$: Machine has an effect

1. $TS: F = 53.7172$
2. $CV: F_{18,05} = 3.5548$
3. $P: P = 2.56 \times 10^{-8}$
4. Reject $H_0$; the machine has an effect on the number of items produced.

**Conclusion**: There is sufficient evidence to warrant rejection of the claim that the machine has no effect on the number of items produced.
Use the rank correlation coefficient to test for a correlation between the two variables.

6) Given that the rank correlation coefficient, \( r_s \), for 35 pairs of data is 0.321, test the claim of correlation between the two variables. Use a significance level of 0.01.

\[
\begin{align*}
H_0 : & \quad \rho_s = 0 \\
H_1 : & \quad \rho_s \neq 0 \\
TS : & \quad r_s = 0.321 \\
CV : & \quad r_{35,01} = 0.442 \\
\end{align*}
\]

1. Treat \( H_0 \): Fail to reject \( H_0 \)
2. Conclusion: The rank correlation is not significant.

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

7) A standard aptitude test is given to several randomly selected programmers, and the scores are given below for the mathematics and verbal portions of the test.

<table>
<thead>
<tr>
<th>Mathematics</th>
<th>440</th>
<th>327</th>
<th>347</th>
<th>427</th>
<th>349</th>
<th>377</th>
<th>425</th>
<th>398</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal</td>
<td>285</td>
<td>243</td>
<td>371</td>
<td>271</td>
<td>322</td>
<td>385</td>
<td>259</td>
<td>248</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
H_0 : \quad & \rho_s = 0 \\
H_1 : \quad & \rho_s \neq 0 \\
TS : \quad & r_s = 0.867 \\
CV : \quad & r_{9,05} = 0.700 \\
\end{align*}
\]

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