Chapter 17

Data Analysis: Hypothesis Testing Related to Differences
### Focus of this Chapter

- Hypothesis Testing Related to Differences
- Means
- Proportions

### Relationship to Previous Chapters

- Research Questions and Hypothesis (Chapter 2)
- Data Analysis Strategy (Chapter 15)
- General Procedure of Hypothesis Testing (Chapter 16)

### Relationship to Marketing Research Process

1. Problem Definition
2. Approach to Problem
3. Research Design
4. Field Work
5. Data Preparation and Analysis
6. Report Preparation and Presentation
Figure 17.2  Hypothesis Testing Related to Differences: An Overview
Hypothesis Testing Related to Differences
(Fig 17.3, 17.4 & 17.5)

- t-Distribution
- Testing Hypothesis

**t-Tests**
(Table 17.1, 17.2, 17.3, 17.4, 17.5)

- One Sample
  - Independent
  - Paired
- Two Samples

Testing Hypothesis for More Than Two Samples
(Fig 17.6) (Table 17.6, 17.7 & 17.8)

Application to Contemporary Issues
(Figs 17.7 & 17.8)

International | Social Media | Ethics
Hypothesis Testing Related to Differences

- **Parametric tests** assume that the variables of interest are measured on at least an interval scale.

- These tests can be further classified based on whether one or two or more samples are involved.

- The samples are **independent** if they are drawn randomly from different populations. For the purpose of analysis, data pertaining to different groups of respondents, e.g., males and females, are generally treated as independent samples.

- The samples are **paired** when the data for the two samples relate to the same group of respondents.
Figure 17.3
Hypothesis Tests Related to Differences

Tests of Differences

- One Sample
  - Means
  - Proportions
- Two Independent Samples
  - Means
  - Proportions
- Paired Samples
  - Means
  - Proportions
- More Than Two Samples
  - Means
  - Proportions
The $t$ Distribution

- The $t$ statistic assumes that the variable is normally distributed and the mean is known (or assumed to be known) and the population variance is estimated from the sample.

- Assume that the random variable $X$ is normally distributed, with mean $\mu$ and unknown population variance $\sigma^2$, which is estimated by the sample variance $s^2$.

- Then, $t = (\bar{X} - \mu)/s_{\bar{X}}$ is $t$ distributed with $n - 1$ degrees of freedom.

- The $t$ distribution is similar to the normal distribution in appearance. Both distributions are bell-shaped and symmetric. As the number of degrees of freedom increases, the $t$ distribution approaches the normal distribution.
Hypothesis Testing Using the t Statistic

1. Formulate the null \((H_0)\) and the alternative \((H_1)\) hypotheses.
2. Select the appropriate formula for the \(t\) statistic.
3. Select a significance level, \(\alpha\), for testing \(H_0\). Typically, the 0.05 level is selected.
4. Take one or two samples and compute the mean and standard deviation for each sample.
5. Calculate the \(t\) statistic assuming \(H_0\) is true.
6. Calculate the degrees of freedom and estimate the probability of getting a more extreme value of the statistic from Table 4 (Alternatively, calculate the critical value of the \(t\) statistic).
Hypothesis Testing Using the t Statistic (Cont.)

7. If the probability computed in step 5 is smaller than the significance level selected in step 2, reject \( H_0 \). If the probability is larger, do not reject \( H_0 \). (Alternatively, if the value of the calculated \( t \) statistic in step 4 is larger than the critical value determined in step 5, reject \( H_0 \). If the calculated value is smaller than the critical value, do not reject \( H_0 \).) Failure to reject \( H_0 \) does not necessarily imply that \( H_0 \) is true. It only means that the true state is not significantly different than that assumed by \( H_0 \).

8. Express the conclusion reached by the \( t \) test in terms of the marketing research problem.
Formulate $H_0$ and $H_1$

Select Appropriate t-Test

Choose Level of Significance, $\alpha$

Collect Data and Calculate Test Statistic

a) Determine Probability Associated with Test Statistic ($T_{CAL}$)

b) Determine Critical Value of Test Statistic ($T_{CR}$)

a) Compare with Level of Significance, $\alpha$

b) Determine if $T_{CAL}$ falls into (Non) Rejection Region

Reject or Do Not Reject $H_0$

Draw Marketing Research Conclusion
One Sample t Test

\[ H_0 = \mu \leq 7.0 \]
\[ H_1 = \mu > 7.0 \]
\[ t = \frac{\bar{X} - \mu}{s_{\bar{X}}} \]
\[ s_{\bar{X}} = \frac{s}{\sqrt{n}} \]
\[ s_{\bar{X}} = \frac{1.6}{\sqrt{20}} = \frac{1.6}{4.472} = 0.358 \]
\[ t = \frac{7.9 - 7.0}{0.358} = \frac{0.9}{0.358} = 2.514 \]
The degrees of freedom for the $t$ statistic to test the hypothesis about one mean are $n - 1$. In this case, $n - 1 = 20 - 1$, or 19. From Table 4 in the Statistical Appendix, the probability of getting a more extreme value than 2.514 is less than 0.05 (Alternatively, the critical $t$ value for 19 degrees of freedom and a significance level of 0.05 is 1.7291, which is less than the calculated value). Hence, the null hypothesis is rejected.
One Sample $z$ Test

Note that if the population standard deviation was assumed to be known as 1.5, a $z$ test would be appropriate.

\[ z = (\bar{X} - \mu) / \sigma_{\bar{X}} \]

where

\[ \sigma_{\bar{X}} = \frac{1.5}{\sqrt{20}} = \frac{1.5}{4.472} = 0.335 \]

and

\[ z = \frac{7.9 - 7.0}{0.335} = \frac{0.9}{0.335} = 2.687 \]
One Sample $z$ Test

- From Table 2 in the Statistical Appendix, the probability of getting a more extreme value of $z$ than 2.687 is less than 0.05. (Alternatively, the critical $z$ value for a one-tailed test and a significance level of 0.05 is 1.645, which is less than the calculated value.) Therefore, the null hypothesis is rejected, reaching the same conclusion arrived at earlier by the $t$ test.

- The procedure for testing a null hypothesis with respect to a proportion was illustrated in Chapter 16 when we introduced hypothesis testing.
<table>
<thead>
<tr>
<th>No.</th>
<th>Sample</th>
<th>Preference for Disney Before</th>
<th>Preference for Disney After</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>1.00</td>
<td>7.00</td>
<td>9.00</td>
</tr>
<tr>
<td>2.</td>
<td>1.00</td>
<td>6.00</td>
<td>8.00</td>
</tr>
<tr>
<td>3.</td>
<td>1.00</td>
<td>5.00</td>
<td>8.00</td>
</tr>
<tr>
<td>4.</td>
<td>1.00</td>
<td>6.00</td>
<td>9.00</td>
</tr>
<tr>
<td>5.</td>
<td>1.00</td>
<td>4.00</td>
<td>7.00</td>
</tr>
<tr>
<td>6.</td>
<td>1.00</td>
<td>6.00</td>
<td>8.00</td>
</tr>
<tr>
<td>7.</td>
<td>1.00</td>
<td>5.00</td>
<td>7.00</td>
</tr>
<tr>
<td>8.</td>
<td>1.00</td>
<td>4.00</td>
<td>7.00</td>
</tr>
<tr>
<td>9.</td>
<td>1.00</td>
<td>7.00</td>
<td>9.00</td>
</tr>
<tr>
<td>10.</td>
<td>1.00</td>
<td>5.00</td>
<td>7.00</td>
</tr>
<tr>
<td>11.</td>
<td>2.00</td>
<td>3.00</td>
<td>7.00</td>
</tr>
<tr>
<td>12.</td>
<td>2.00</td>
<td>4.00</td>
<td>8.00</td>
</tr>
<tr>
<td>13.</td>
<td>2.00</td>
<td>4.00</td>
<td>7.00</td>
</tr>
<tr>
<td>14.</td>
<td>2.00</td>
<td>3.00</td>
<td>6.00</td>
</tr>
<tr>
<td>15.</td>
<td>2.00</td>
<td>6.00</td>
<td>8.00</td>
</tr>
<tr>
<td>16.</td>
<td>2.00</td>
<td>5.00</td>
<td>8.00</td>
</tr>
<tr>
<td>17.</td>
<td>2.00</td>
<td>4.00</td>
<td>9.00</td>
</tr>
<tr>
<td>18.</td>
<td>2.00</td>
<td>3.00</td>
<td>6.00</td>
</tr>
<tr>
<td>19.</td>
<td>2.00</td>
<td>3.00</td>
<td>7.00</td>
</tr>
<tr>
<td>20.</td>
<td>2.00</td>
<td>5.00</td>
<td>9.00</td>
</tr>
</tbody>
</table>
Table 17.2 One Sample t-test

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of Cases</th>
<th>Mean</th>
<th>SD</th>
<th>SE of Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAR00002</td>
<td>10</td>
<td>5.500</td>
<td>1.080</td>
<td>.342</td>
</tr>
</tbody>
</table>

Test Value = 5

<table>
<thead>
<tr>
<th>Mean Difference</th>
<th>95% CI</th>
<th>t-value</th>
<th>df</th>
<th>2-Tail Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>.50</td>
<td>-.273</td>
<td>1.273</td>
<td>1.46</td>
<td>9</td>
</tr>
</tbody>
</table>
Two Independent Samples: Means

- In this case the hypotheses take the following form.

  \[ H_0: \mu_1 = \mu_2 \]
  \[ H_1: \mu_1 \neq \mu_2 \]

- If both populations are found to have the same variance, a pooled variance estimate is computed from the two sample variances as follows:

  \[ s^2 = \frac{\sum_{i=1}^{n_1} (X_{i1} - \bar{X}_1)^2 + \sum_{i=1}^{n_2} (X_{i2} - \bar{X}_2)^2}{n_1 + n_2 - 2} \]

  or

  \[ s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \]
Two Independent Samples: Means (Cont.)

The standard deviation of the test statistic can be estimated as:

\[ s_{(\bar{x}_1-\bar{x}_2)} = \sqrt{s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \]

The appropriate value of \( t \) can be calculated as:

\[ t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{(\bar{x}_1-\bar{x}_2)}} \]

The degrees of freedom in this case are \( (n_1 + n_2 - 2) \)
Two Independent Samples: F Test

An *F test* of sample variance may be performed if it is not known whether the two populations have equal variance. In this case, the hypotheses are:

\[
H_0 : \sigma_1^2 = \sigma_2^2 \\
H_1 : \sigma_1^2 \neq \sigma_2^2
\]
Two Independent Samples: F Statistic

$$F_{(n_1-1),(n_2-1)} = \frac{s_1^2}{s_2^2}$$

Where

- $n_1$ = size of sample 1
- $n_2$ = size of sample 2
- $n_1 - 1$ = degrees of freedom for sample 1, the larger sample variance
- $n_2 - 1$ = degrees of freedom for sample 2, the smaller sample variance
- $s_1^2$ = sample variance for sample 1
- $s_2^2$ = sample variance for sample 2

Using the data of Table 17.1, we determine whether teenagers have a different preference than adults. A two-independent-samples $t$ test was conducted. The results are presented in Table 17.3.
### Table 17.3

**t-tests for Independent Samples**

<table>
<thead>
<tr>
<th>Variable Sample</th>
<th>Number of Cases</th>
<th>Mean</th>
<th>SD</th>
<th>SE of Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teenagers</td>
<td>10</td>
<td>5.5000</td>
<td>1.080</td>
<td>.342</td>
</tr>
<tr>
<td>Adults</td>
<td>10</td>
<td>4.0000</td>
<td>1.054</td>
<td>.333</td>
</tr>
</tbody>
</table>

Mean Difference = 1.5000

Levene's Test for Equality of Variances: F = .150  P = .703

### t-test for Equality of Means

<table>
<thead>
<tr>
<th>Variances</th>
<th>t-value</th>
<th>df</th>
<th>2-Tail Sig</th>
<th>SE of Diff</th>
<th>95% CI for Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equal</td>
<td>3.14</td>
<td>18</td>
<td>.006</td>
<td>.477</td>
<td>(.497, 2.503)</td>
</tr>
<tr>
<td>Unequal</td>
<td>3.14</td>
<td>17.99</td>
<td>.006</td>
<td>.477</td>
<td>(.497, 2.503)</td>
</tr>
</tbody>
</table>

Copyright © 2009 Pearson Education, Inc.
Figure 17.5
Calculating the Critical Value of the Test Statistic: \( \text{TS}_{CR} \) for Two-Tailed and One-Tailed Tests

(a) Two-tailed test

(b) One-tailed test
Two Independent Samples: Proportions

The case involving proportions for two independent samples is illustrated using the data of Table 17.4, which gives the number of users and nonusers of jeans in the U.S. and Hong Kong. The null and alternative hypotheses are:

\[ H_0: \pi_1 = \pi_2 \]

\[ H_1: \pi_1 \neq \pi_2 \]

A Z test is used as in testing the proportion for one sample. However, in this case the test statistic is given by:

\[ Z = \frac{P_1 - P_2}{S_{P_1-P_2}} \]
Two Independent Samples: Proportions (Cont.)

In the test statistic, the numerator is the difference between the proportions in the two samples, \( P_1 \) and \( P_2 \). The denominator is the standard error of the difference in the two proportions and is given by

\[
S_{P_1 - P_2} = \sqrt{P(1 - P)\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}
\]

where

\[
P = \frac{n_1P_1 + n_2P_2}{n_1 + n_2}
\]
Two Independent Samples: Proportions (Cont.)

A significance level of $\alpha = 0.05$ is selected. Given the data of Table 17.4, the test statistic can be calculated as:

\[
P_1 - P_2 = 0.8 - 0.6 = 0.2
\]

\[
P = \frac{200 \times 0.8 + 200 \times 0.6}{200 + 200} = 0.7
\]

\[
SP_1 - P_2 = \sqrt{(0.7)(0.3)\left[\frac{1}{200} + \frac{1}{200}\right]} = 0.04583
\]

\[
Z = \frac{0.2}{0.04583} = 4.36
\]
Given a two-tail test, the area to the right of the critical value is 0.025. Hence, the critical value of the test statistic is 1.96. Since the calculated value exceeds the critical value, the null hypothesis is rejected. Thus, the proportion of users (0.8 for the U.S. and 0.6 for HK) is significantly different for the two samples.
Table 17.4
Comparing Jeans Users for USA and Hong Kong

Usage of Jeans

<table>
<thead>
<tr>
<th></th>
<th>Users</th>
<th>Non-Users</th>
<th>Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>160</td>
<td>40</td>
<td>200</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>120</td>
<td>80</td>
<td>200</td>
</tr>
<tr>
<td>Column Totals</td>
<td>280</td>
<td>120</td>
<td></td>
</tr>
</tbody>
</table>

Copyright © 2011 Pearson Education, Inc. Chapter 17 - 27
Paired Samples

The difference in these cases is examined by a paired samples t test. To compute $t$ for paired samples, the paired difference variable, denoted by $D$, is formed and its mean and variance calculated. Then the $t$ statistic is computed. The degrees of freedom are $n - 1$, where $n$ is the number of pairs. The relevant formulas are:

$$H_0: \mu = 0$$
$$H_1: \mu \neq 0$$

$$t_{n-1} = \frac{\bar{D} - \mu_D}{S_D / \sqrt{n}}$$

continued…
For the data in Table 17.1, a paired $t$ test could be used to determine if the teenagers differed in their preference before and after visiting Disney Park. The resulting output is shown in Table 17.5.
### Table 17.5
**t-Tests for Paired Samples**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of pairs</th>
<th>Corr</th>
<th>2-tail Sig</th>
<th>Mean</th>
<th>SD</th>
<th>SE of Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preference Before</td>
<td>10</td>
<td>.881</td>
<td>.001</td>
<td>5.5000</td>
<td>1.080</td>
<td>0.342</td>
</tr>
<tr>
<td>Preference After</td>
<td></td>
<td></td>
<td></td>
<td>7.9000</td>
<td>.876</td>
<td>0.277</td>
</tr>
</tbody>
</table>

**Paired Differences**

<table>
<thead>
<tr>
<th>Mean</th>
<th>SD</th>
<th>SE of Mean</th>
<th>t-value</th>
<th>df</th>
<th>2-tail Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.4000</td>
<td>0.516</td>
<td>0.163</td>
<td>-14.70</td>
<td>9</td>
<td>.000</td>
</tr>
</tbody>
</table>

95% CI (-2.769, -2.031)
Analysis of Variance

- **Analysis of variance (ANOVA)** is used as a test of means for two or more populations. The null hypothesis, typically, is that all means are equal.

- Analysis of variance must have a dependent variable that is metric (measured using an interval or ratio scale).

- There must also be one or more independent variables that are all categorical (nonmetric). Categorical independent variables are also called **factors**.
One-Way Analysis of Variance

- A particular combination of factor levels, or categories, is called a **treatment**.

- **One-way analysis of variance** involves only one categorical variable, or a single factor. In one-way analysis of variance, a treatment is the same as a factor level.
One-Way Analysis of Variance (Cont.)

Marketing researchers are often interested in examining the differences in the mean values of the dependent variable for several categories of a single independent variable or factor. For example:

- Do the various segments differ in terms of their volume of product consumption?
- Do the brand evaluations of groups exposed to different commercials vary?
- What is the effect of consumers' familiarity with the store (measured as high, medium, and low) on preference for the store?
Statistics Associated with One-Way Analysis of Variance

- **eta\(^2\) (\(\eta^2\))**. The strength of the effects of \(X\) (independent variable or factor) on \(Y\) (dependent variable) is measured by \(\text{eta}^2 (\eta^2)\). The value of \(\eta^2\) varies between 0 and 1.

- **\(F\) statistic**. The null hypothesis that the category means are equal in the population is tested by an **\(F\) statistic** based on the ratio of mean square related to \(X\) and mean square related to error.

- **Mean square**. This is the sum of squares divided by the appropriate degrees of freedom.
Statistics Associated with One-Way Analysis of Variance (Cont.)

- $SS_{between}$. Also denoted as $SS_x$, this is the variation in $Y$ related to the variation in the means of the categories of $X$. This represents variation between the categories of $X$, or the portion of the sum of squares in $Y$ related to $X$.

- $SS_{within}$. Also referred to as $SS_{error}$, this is the variation in $Y$ due to the variation within each of the categories of $X$. This variation is not accounted for by $X$.

- $SS_y$. This is the total variation in $Y$. 

Copyright © 2011 Pearson Education, Inc.   Chapter 17 - 35
Figure 17.6 Conducting One-Way ANOVA

1. Identify the Dependent and Independent Variables
2. Decompose the Total Variation
3. Measure the Effects
4. Test the Significance
5. Interpret the Results
Conducting One-Way Analysis of Variance
Decompose the Total Variation

The total variation in $Y$, denoted by $SS_y$, can be decomposed into two components:

$$SS_y = SS_{between} + SS_{within}$$

where the subscripts $between$ and $within$ refer to the categories of $X$. $SS_{between}$ is the variation in $Y$ related to the variation in the means of the categories of $X$. For this reason, $SS_{between}$ is also denoted as $SS_x$. $SS_{within}$ is the variation in $Y$ related to the variation within each category of $X$. $SS_{within}$ is not accounted for by $X$. Therefore it is referred to as $SS_{error}$. 
Decompose the Total Variation

\[ SS_y = SS_x + SS_{\text{error}} \]

Where

\[ SS_y = \sum_{i=1}^{N} (Y_i - \bar{Y})^2 \]

\[ SS_x = \sum_{j=1}^{c} n (\bar{Y}_j - \bar{Y})^2 \]

\[ SS_{\text{error}} = \sum_{j=1}^{c} \sum_{i=1}^{n} (Y_{ij} - \bar{Y}_j)^2 \]

\( Y_i \) = individual observation
\( \bar{Y}_j \) = mean for category \( j \)
\( \bar{Y} \) = mean over the whole sample, or grand mean
\( Y_{ij} \) = \( i \)th observation in the \( j \)th category
In analysis of variance, we estimate two measures of variation: within groups (\(SS_{\text{within}}\)) and between groups (\(SS_{\text{between}}\)). Thus, by comparing the \(Y\) variance estimates based on between-group and within-group variation, we can test the null hypothesis.

The strength of the effects of \(X\) on \(Y\) are measured as follows:

\[\eta^2 = SS_x / SS_y = (SS_y - SS_{\text{error}}) / SS_y\]

The value of \(\eta^2\) varies between 0 and 1.
Test Significance

In one-way analysis of variance, the interest lies in testing the null hypothesis that the category means are equal in the population.

$$H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_c$$

Under the null hypothesis, $SS_x$ and $SS_{error}$ come from the same source of variation. In other words, the estimate of the population variance of $Y$,

$$S_y^2 = \frac{SS_x}{(c - 1)}$$
= Mean square due to $X$
= $MS_x$

or

$$S_y^2 = \frac{SS_{error}}{(N - c)}$$
= Mean square due to error
= $MS_{error}$
Test Significance

The null hypothesis may be tested by the $F$ statistic based on the ratio between these two estimates:

$$F = \frac{SS_x/(c - 1)}{SS_{error}/(N - c)} = \frac{MS_x}{MS_{error}}$$

This statistic follows the $F$ distribution, with $(c - 1)$ and $(N - c)$ degrees of freedom (df).
Interpret the Results

- If the null hypothesis of equal category means is not rejected, then the independent variable does not have a significant effect on the dependent variable.

- On the other hand, if the null hypothesis is rejected, then the effect of the independent variable is significant.

- A comparison of the category mean values will indicate the nature of the effect of the independent variable.
We illustrate the concepts discussed in this chapter using the data presented in Table 17.6.

The supermarket is attempting to determine the effect of in-store advertising \((X)\) on sales \((Y)\).

The null hypothesis is that the category means are equal:

\[ H_0: \mu_1 = \mu_2 = \mu_3. \]
Table 17.6
Effect of In-Store Promotion On Sales

<table>
<thead>
<tr>
<th>Store No.</th>
<th>Level of In-store Promotion</th>
<th>Normalized Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>5</td>
</tr>
</tbody>
</table>
Calculation of Means

- Category means $j$:

\[
\begin{array}{ccc}
45/5 & 25/5 & 20/5 \\
= 9 & = 5 & = 4 \\
\end{array}
\]

- Grand mean: $= (45 + 25 + 20)/15 = 6$
Sums of Squares

\[SS_y = (10 - 6)^2 + (9 - 6)^2 + (10 - 6)^2 + (8 - 6)^2 + (8 - 6)^2 + (6 - 6)^2 + (4 - 6)^2 + (7 - 6)^2 + (3 - 6)^2 + (5 - 6)^2 + (5 - 6)^2 + (6 - 6)^2 + (5 - 6)^2 + (2 - 6)^2 + (2 - 6)^2 \]

\[= 16 + 9 + 16 + 4 + 4 + 0 + 4 + 1 + 9 + 1 + 1 + 0 + 1 + 16 + 16 \]

\[= 98 \]
Sums of Squares (Cont.)

\[ SS_x = 5(9 - 6)^2 + 5(5 - 6)^2 + 5(4 - 6)^2 \]

\[ = 45 + 5 + 20 \]

\[ = 70 \]
Sums of Squares (Cont.)

\[ SS_{error} = (10 - 9)^2 + (9 - 9)^2 + (10 - 9)^2 + (8 - 9)^2 + (8 - 9)^2 + (6 - 5)^2 + (4 - 5)^2 + (7 - 5)^2 + (3 - 5)^2 + (5 - 5)^2 + (5 - 4)^2 + (6 - 4)^2 + (5 - 4)^2 + (2 - 4)^2 + (2 - 4)^2 \]

\[ = 1 + 0 + 1 + 1 + 1 + 1 + 4 + 4 + 0 + 1 + 4 + 1 + 4 + 4 \]

\[ = 28 \]
Sum of Squares (Cont.)

It can be verified that  \( SS_y = SS_x + SS_{error} \)

as follows:

\[ 98 = 78 + 20 \]
Measurement of Effects

The strength of the effects on \( X \) and \( Y \) are measured as follows:

\[
\eta^2 = \frac{SS_x}{SS_y}
\]

\[
= 70/98
\]

\[
= 0.714
\]
Testing the Null Hypothesis

\[ F = \frac{SS_X / (c - 1)}{SS_{error} / (N - c)} = \frac{MS_X}{MS_{error}} \]

\[ F = \frac{70 / (3-1)}{28 / (15-3)} \]

\[ = 15.0 \]
Testing the Null Hypothesis

- From Table 5 in the Appendix of Statistical Tables, we see that for 2 and 12 degrees of freedom and $\alpha = 0.05$, the critical value of $F$ is 3.89. Since the calculated value of $F$ is greater than the critical value, we reject the null hypothesis.

- This one-way ANOVA was also conducted using statistical software and the output is shown in Table 17.7.
<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>DF</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig of F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Effects</td>
<td>70.00</td>
<td>2</td>
<td>35.00</td>
<td>15.00</td>
<td>.001</td>
</tr>
<tr>
<td>In-store promotion</td>
<td>70.00</td>
<td>2</td>
<td>35.00</td>
<td>15.00</td>
<td>.001</td>
</tr>
<tr>
<td>Explained</td>
<td>70.00</td>
<td>2</td>
<td>35.00</td>
<td>15.00</td>
<td>.001</td>
</tr>
<tr>
<td>Residual</td>
<td>28.00</td>
<td>12</td>
<td>2.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>98.00</td>
<td>14</td>
<td>7.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Table 17.8 A Summary of Hypothesis Testing

<table>
<thead>
<tr>
<th>Sample</th>
<th>Test/Commets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>One Sample</strong></td>
<td></td>
</tr>
<tr>
<td>Means</td>
<td>( t ) test, if variance is unknown</td>
</tr>
<tr>
<td></td>
<td>( z ) test, if variance is known</td>
</tr>
<tr>
<td>Proportions</td>
<td>( z ) test</td>
</tr>
<tr>
<td><strong>Two Independent Samples</strong></td>
<td></td>
</tr>
<tr>
<td>Means</td>
<td>Two-group ( t ) test</td>
</tr>
<tr>
<td></td>
<td>( F ) test for equality of variances</td>
</tr>
<tr>
<td>Proportions</td>
<td>( z ) test</td>
</tr>
<tr>
<td></td>
<td>Chi-square test</td>
</tr>
<tr>
<td><strong>Paired Samples</strong></td>
<td></td>
</tr>
<tr>
<td>Means</td>
<td>Paired ( t ) test</td>
</tr>
<tr>
<td>Proportions</td>
<td>Chi-square test</td>
</tr>
<tr>
<td><strong>More Than Two Samples</strong></td>
<td></td>
</tr>
<tr>
<td>Means</td>
<td>One-way analysis of variance</td>
</tr>
<tr>
<td>Proportions</td>
<td>Chi-square test</td>
</tr>
</tbody>
</table>
Figure 17.7 A Concept Map for Conducting t-Tests

Hypothesis Testing
  formulate
  H₀ and H₁

Appropriate t-Test
  select
  consider
  choose

Level of Significance, α
  collect data

Calculate the Appropriate t Statistic
  determine
  determine

Probability Associated with t Statistic (TS_Cal.)
  compare with

Critical Value of t Statistic (TS_CR)
  compare with

Level of Significance, α
  results in

Calculated Value of t Statistic (TS_Cal.)
  results in

Reject or Do Not Reject H₀
  reject if
  reject if

Marketing Research Conclusion

Appropriate Tables to determine probability or the value of TS_CR

<table>
<thead>
<tr>
<th>Probability of TS_Cal. &lt; α</th>
</tr>
</thead>
<tbody>
<tr>
<td>[</td>
</tr>
</tbody>
</table>

Recommend status quo

Recommend action implied by H₁

Copyright © 2011 Pearson Education, Inc. Chapter 17 - 55
Figure 17.8 A Concept Map for One-Way ANOVA

Identify the Dependent and Independent Variables

- The dependent variable is metric (interval or ratio scaled)
- The independent variable must be categorical (nonmetric)

The Total Variation

- SS_y = SS_x + SS_{error}
- \( \eta^2 = \frac{SS_x}{SS_y} = \frac{SS_y - SS_{error}}{SS_y} \)

The Effects

- measured as
- test

The Significance

- null hypothesis is
- interpret
- use the F statistic
- \( F = \frac{SS_x/(c-1)}{SS_{error}/(N-c)} = \frac{MS_x}{MS_{error}} \)

The Results

Copyright © 2011 Pearson Education, Inc.  Chapter 17 - 56
SPSS Windows: t-Tests

The major program for conducting parametric tests in SPSS is COMPARE MEANS. This program can be used to conduct t tests on one sample or independent or paired samples. To select these procedures using SPSS for Windows click:

Analyze > Compare Means > Means …
Analyze > Compare Means > One-Sample T Test …
Analyze > Compare Means > Independent-Samples T Test …
Analyze > Compare Means > Paired-Samples T Test …

We illustrate the detailed steps using the data of Table 17.1.
SPSS Detailed Steps: One-Sample t-Test

1. Select DATA from the SPSS menu bar.
2. Click SELECT CASES.
3. Check IF CONDITION IS SATISFIED. Then click the IF button.
4. Move SAMPLE <sample> in to the SELECT CASES: IF box. Click "=" and then click "1."
5. Click CONTINUE.
6. Check FILTER OUT UNSELECTED CASES.
7. Click OK.

8. Select ANALYZE from the SPSS menu bar.

9. Click COMPARE MEANS and then ONE SAMPLE T TEST.

10. Move Preference Before Visiting <pref1> into the TEST VARIABLE(S) box.

11. Type "5" in the TEST VALUE box.

12. Click OK.
First, make sure all cases will be included in the analysis.

- Menu Bar
- DATA
- Select CASES.
- Select ALL CASES.
- Click OK.

Perform the analysis as follows:

1. Select ANALYZE from the SPSS menu bar.
2. Click COMPARE MEANS and then INDEPENDENT SAMPLES T TEST.
SPSS Detailed Steps: Two-Independent-Samples t-Test (Cont.)

3. Move "Preference Before Visiting <pref1>" into the TEST VARIABLE(S) box.

4. Move "Sample <sample>" to GROUPING VARIABLE box.

5. Click DEFINE GROUPS.

6. Type "1" in GROUP 1 box and "2" in GROUP 2 box.

7. Click CONTINUE.

8. Click OK.
SPSS Detailed Steps: Paired-Samples t-Test

1. Select DATA from the SPSS menu bar.
2. Click SELECT CASES.
3. Check IF CONDITION IS SATISFIED. Then click the IF button.
4. Move "SAMPLE<sample>" into the SELECT CASES: IF box. Click "=" and then click "1."
5. Click CONTINUE.
6. Check "FILTER OUT UNSELECTED CASES."
7. Click OK.

8. Select ANALYZE from the SPSS menu bar.

9. Click COMPARE MEANS and then PAIRED SAMPLES T TEST.

10. Select "Preference Before Visiting <pref1>" and then select "Preference After Visiting <pref2>." Move these variables into the PAIRED VARIABLE(S) box.

11. Click OK.
One-way ANOVA can be efficiently performed using the program COMPARE MEANS and then One-way ANOVA. To select this procedure using SPSS for Windows, click:

**Analyze > Compare Means > One-Way ANOVA …**

We illustrate the detailed steps using the data of Table 17.6
SPSS Detailed Steps: One-Way ANOVA

1. Select ANALYZE from the SPSS menu bar.
2. Click COMPARE MEANS and then ONE-WAY ANOVA.
3. Move "Normalized Sales <sales>" into the DEPENDENT LIST box.
4. Move "In-Store Promotion <promotio>" to the FACTOR box.
5. Click OPTIONS.
6. Click DESCRIPTIVE.
7. Click CONTINUE.
8. Click OK.
Excel

The available parametric tests in EXCEL and other spreadsheets include the t-TEST: paired sample for means; t-TEST: two independent samples assuming equal variances; t-TEST: two independent samples assuming unequal variances, z-TEST: two samples for means, and F-TEST two samples for variances.

One-way ANOVA can be accessed under the DATA > DATA ANALYSIS > ANOVA: SINGLE FACTOR.

We illustrate the detailed steps using the data of Table 17.1
Excel Detailed Steps: One-Sample t-Test

1. Add a column with the name "Dummy" beside PREF2 and fill the first 10 rows with "5."
2. Select DATA tab.
3. In the ANALYSIS tab, select DATA ANALYSIS.
4. DATA ANALYSIS Window pops up.
5. Select t-TEST: PAIRED TWO SAMPLE FOR MEANS from DATA ANALYSIS Window.
6. Click OK.
7. t-TEST: PAIRED TWO SAMPLE FOR MEANS pop-up window appears on screen.

8. t-TEST: PAIRED TWO SAMPLE FOR MEANS window has two portions:
   a. INPUT
   b. OUTPUT OPTIONS

9. Input portion asks for five inputs:
   a. Click in the VARIABLE 1 RANGE box. Select (highlight) the first 10 rows of data under PREF1. $C$2:$C$11 should appear on VARIABLE 1 RANGE.
Excel Detailed Steps: One-Sample t-Test (Cont.)

b. Click in the VARIABLE 2 RANGE box. Select (highlight) the first 10 rows of data under DUMMY. $E$2:$E$11 should appear on VARIABLE 2 RANGE.

c. Leave HYPOTHESIZED MEAN DIFFERENCE and LABELS as blank.

d. Default value of ALPHA 0.05 is seen on the ALPHA Box. Let the ALPHA value be like that.

10. In the OUTPUT OPTIONS pop-up window, select NEW WORKBOOK Options.

11. Click OK.
Excel Detailed Steps: Two-Independent Samples t-Test

1. Select DATA tab.
2. In the ANALYSIS group, select DATA ANALYSIS.
3. DATA ANALYSIS Window pops up.
4. Select t-TEST: TWO-SAMPLE ASSUMING EQUAL VARIANCES from DATA ANALYSIS Window.
5. Click OK.
Excel Detailed Steps: Two-Independent Samples t-Test (Contd.)

6. t-TEST: TWO-SAMPLE ASSUMING EQUAL VARIANCES pop-up window appears on screen.

7. t-TEST: TWO-SAMPLE ASSUMING EQUAL VARIANCES window has two portions:
   a. INPUT
   b. OUTPUT OPTIONS

8. INPUT portion asks for two inputs:
   a. Click in the VARIABLE 1 RANGE box. Select (highlight) the first 10 rows of data under PREF1. $C$2:$C$11 should appear on VARIABLE 1 RANGE.
Excel Detailed Steps: Two-Independent-Samples t-Test (Cont.)

b. Click in the VARIABLE 2 RANGE box. Select (highlight) the last 10 rows of data under PREF1. $C$12:$C$21 should appear on VARIABLE 2 RANGE.

c. Leave HYPOTHESIZED MEAN DIFFERENCE and LABELS blank.

d. Default value of ALPHA 0.05 is seen on the ALPHA Box. Let the ALPHA value be like that.

9. In the OUTPUT OPTIONS pop-up window, select NEW WORKBOOK Options.

10. Click OK.
Excel Detailed Steps: Paired-Samples t-Test

1. Select DATA tab.
2. In the ANALYSIS tab, select DATA ANALYSIS.
3. DATA ANALYSIS Window pops up.
4. Select t-TEST: PAIRED TWO SAMPLE FOR MEANS from DATA ANALYSIS Window.
5. Click OK.
6. t-TEST: PAIRED TWO SAMPLE FOR MEANS pop-up window appears on screen.
Excel Detailed Steps: Paired-Samples t-Test (Cont.)

7. t-TEST: PAIRED TWO SAMPLE FOR MEANS window has two portions:
   a. INPUT
   b. OUTPUT OPTIONS

8. INPUT portion asks for two inputs:
   a. Click in the VARIABLE 1 RANGE box. Select (highlight) the first 10 rows of data under PREF1. $C$2:$C$11 should appear on VARIABLE 1 RANGE.
Excel Detailed Steps: Paired-Samples t-Test (Cont.)

b. Click in the VARIABLE 2 RANGE box. Select (highlight) the first 10 rows of data under PREF2. $D$2:$D$11 should appear on VARIABLE 2 RANGE.

c. Leave HYPOTHESES MEAN DIFFERENCE and LABELS as blank.

d. Default value of ALPHA 0.05 is seen on the ALPHA Box. Let the ALPHA value be like that.

9. In the OUTPUT OPTIONS pop-up window, select NEW WORKBOOK Options.

10. Click OK.
Excel Detailed Steps: One-Way ANOVA

1. Select DATA tab.
2. In the ANALYSIS tab, select DATA ANALYSIS.
3. DATA ANALYSIS Window pops up.
4. Select ANOVA: SINGLE FACTOR from DATA ANALYSIS Window.
5. Click OK.
6. ANOVA: SINGLE FACTOR pop-up window appears on screen.
7. ANOVA: SINGLE FACTOR window has two portions:
   a. INPUT
   b. OUTPUT OPTIONS
Excel Detailed Steps:
One-Way ANOVA (Cont.)

8. INPUT portion asks for two inputs:
   a. Click in the INPUT RANGE box. Select (highlight) all the data under the columns High, Medium, Low at the same time. $B$2:$D$6 should appear in the INPUT RANGE.
   b. Click in COLUMNS beside GROUPED BY.
   c. LABELS in the First row should not be checked.
   d. Leave ALPHA at the default value of 0.05.
Excel Detailed Steps:
One-Way ANOVA (Cont.)

9. In the OUTPUT OPTIONS pop-up window, select NEW WORKBOOK Options.

10. Click OK.
Exhibit 17.1
Other Computer Programs for t-tests

MINITAB

Parametric test available in MINITAB under the STAT > BASIC STATISTIC function are z-TEST mean, T-TEST of the mean, and two-sample T-TEST.

SAS

In SAS, the program T TEST can be used to conduct t-tests on independent as well as paired samples.
MINITAB

Analysis of Variance can be accessed from the Stat > ANOVA function. This function performs one way ANOVA and can also handle more complex designs. In order to compute the mean and standard deviation, the cross-tab function must be used. To obtain $F$ and $p$ values, use the balanced ANOVA.

SAS

The main program for performing analysis of variance is ANOVA. This program can handle data from a wide variety of experimental designs. One-Way ANOVA can be efficiently performed using ONE-WAY ANOVA within the ANOVA task.
Acronym: *T*-test

The major characteristics of *t*-tests can be summarized by the acronym *T TEST*:

- **T** distribution is similar to the normal distribution
- **T** est of difference: Means or proportions
- **E** stimate of variance from the sample
- **S** ingle sample
- **T** wo samples: independent or paired