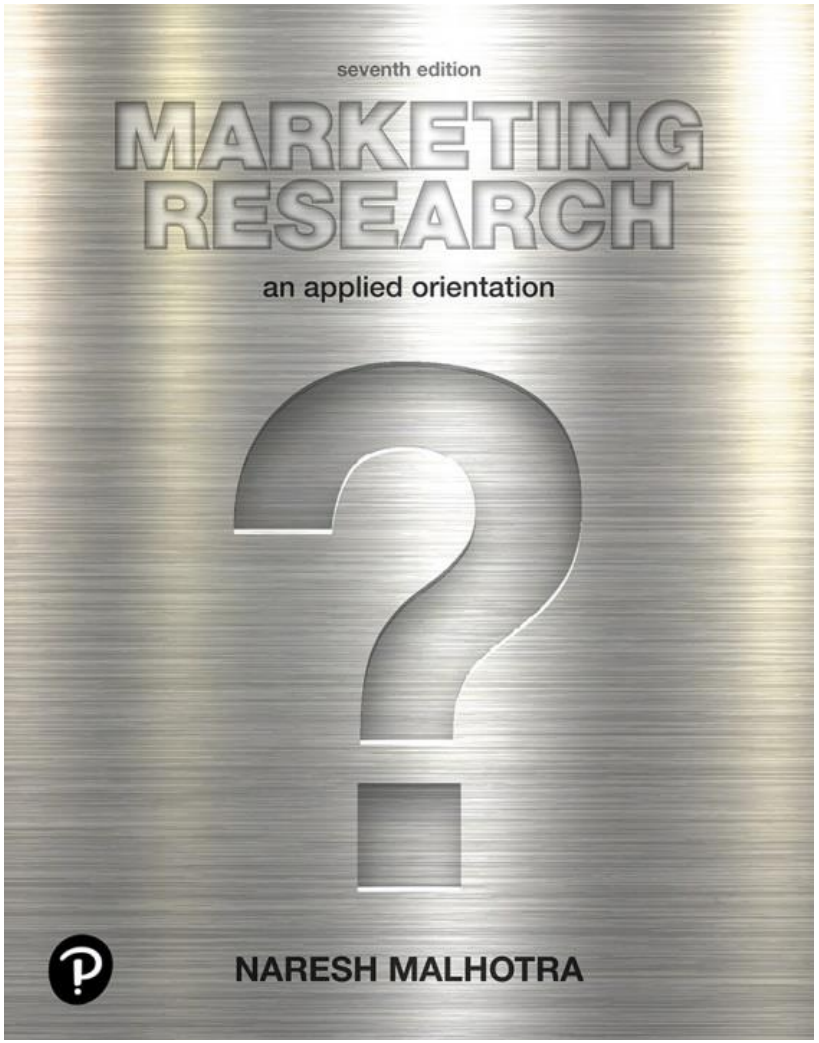
Marketing Research: An Applied Orientation

Seventh Edition



Chapter 15

Frequency Distribution, Cross-Tabulation, and Hypothesis Testing

Pearson

Restaurant Preference

Table 14.1 Restaurant Preference

Id	Preference	Quality	Quantity	Value	Service	Income
1	2	2	3	1	3	6
2	6	5	6	5	7	2
3	4	4	3	4	5	3
4	1	2	1	1	2	5
5	7	6	6	5	4	1
6	5	4	4	5	4	3
7	2	2	3	2	3	5
8	3	3	4	2	3	4
9	7	6	7	6	5	2
10	2	3	2	2	2	5
11	2	3	2	1	3	6
12	6	6	6	6	7	2
13	4	4	3	3	4	3
14	1	1	3	1	2	4
15	7	7	5	5	4	2
16	5	5	4	5	5	3
17	2	3	1	2	3	4
18	4	4	3	3	3	3
19	7	5	5	7	5	5
20	3	2	2	3	3	3

Internet Usage Data (1 of 2)

Table 15.1 Internet Usage Data

Respondent Number	Sex	Familiarity	Internet Usage	Attitude Toward Internet	Attitude Toward Technology	Usage of Internet: Shopping	Usage of Internet: Banking
1	1	7	14	7	6	1	1
2	2	2	2	3	3	2	2
3	2	3	3	4	3	1	2
4	2	3	3	7	5	1	2
5	1	7	13	7	7	1	1
6	2	4	6	5	4	1	2
7	2	2	2	4	5	2	2
8	2	3	6	5	4	2	2
9	2	3	6	6	4	1	2
10	1	9	15	7	6	1	2
11	2	4	3	4	3	2	2
12	2	5	4	6	4	2	2
13	1	6	9	6	5	2	1
14	1	6	8	3	2	2	2

Internet Usage Data (2 of 2)

[**Table 15.1** Continued]

Respondent Number	Sex	Familiarity	Internet Usage	Attitude Toward Internet	Attitude Toward Technology	Usage of Internet: Shopping	Usage of Internet: Banking
15	1	6	5	5	4	1	2
16	2	4	3	4	3	2	2
17	1	6	9	5	3	1	1
18	1	4	4	5	4	1	2
19	1	7	14	6	6	1	1
20	2	6	6	6	4	2	2
21	1	6	9	4	2	2	2
22	1	5	5	5	4	2	1
23	2	3	2	4	2	2	2
24	1	7	15	6	6	1	1
25	2	6	6	5	3	1	2
26	1	6	13	6	6	1	1
27	2	5	4	5	5	1	1
28	2	4	2	3	2	2	2
29	1	4	4	5	3	1	2
30	1	3	3	7	5	1	2

Frequency Distribution

- In a **frequency distribution**, one variable is considered at a time.
- A frequency distribution for a variable produces a table of frequency counts, percentages, and cumulative percentages for all the values associated with that variable.

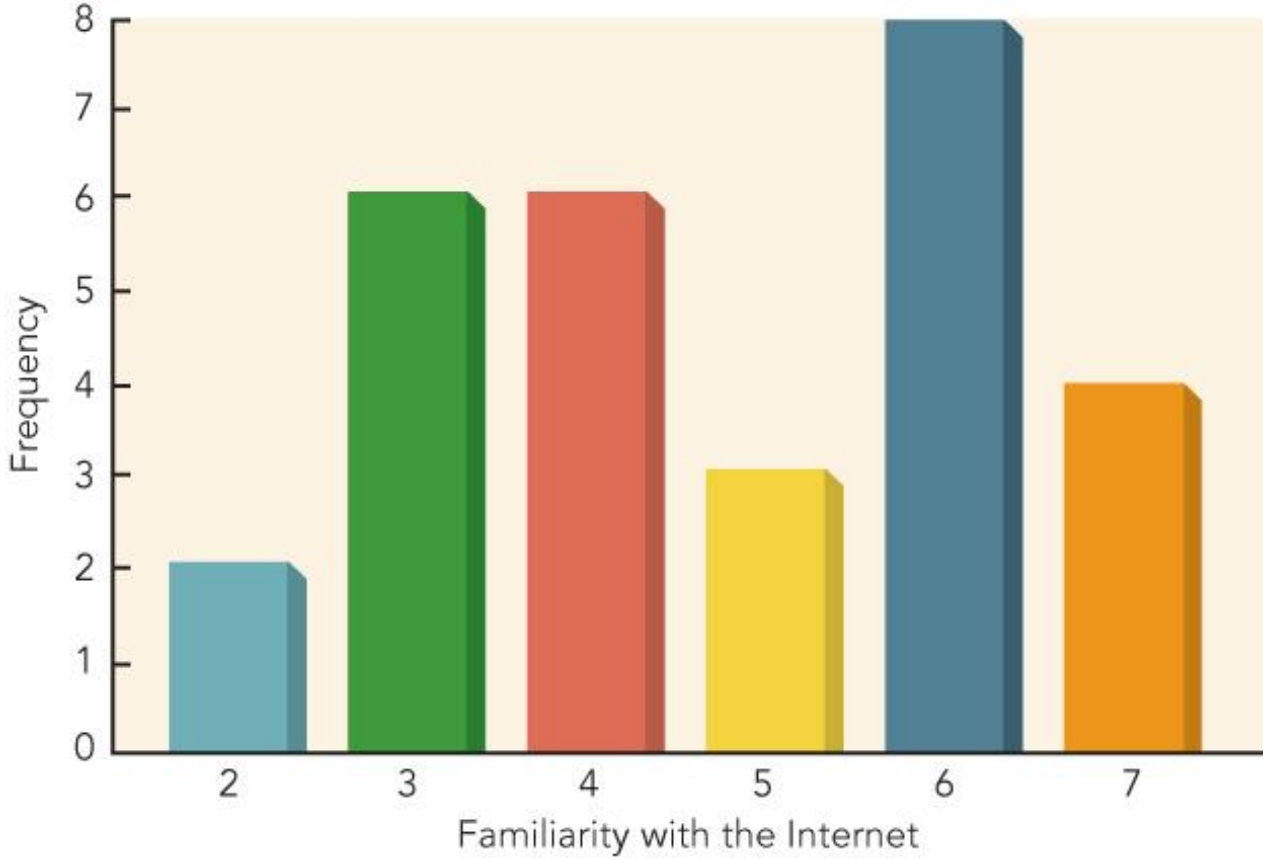
Frequency of Familiarity with the Internet

Table 15.2 Frequency Distribution of Familiarity with the Internet

Value Label	Value	Frequency (*n*)	Percentage	Valid Percentage	Cumulative Percentage
Very unfamiliar	1	0	0.0	0.0	0.0
	2	2	6.7	6.9	6.9
	3	6	20.0	20.7	27.6
	4	6	20.0	20.7	48.3
	5	3	10.0	10.3	58.6
	6	8	26.7	27.6	86.2
Very familiar	7	4	13.3	13.8	100.0
Missing	9	1	3.3		
	TOTAL	30	100.0	100.0	

Frequency Histogram

Figure 15.1 Frequency Histogram



Statistics Associated with Frequency Distribution: Measures of Location (1 of 2)

- The **mean**, or average value, is the most commonly used measure of central tendency. The mean, $\overline{X}$, is given by

$$\overline{X} = \sum_{i=1}^{n} X_i / n$$

Where,

X_i = Observed values of the variable X

n = Number of observations (sample size)

- The **mode** is the value that occurs most frequently. It represents the highest peak of the distribution. The mode is a good measure of location when the variable is inherently categorical or has otherwise been grouped into categories.

Statistics Associated with Frequency Distribution: Measures of Location (2 of 2)

- The **median** of a sample is the middle value when the data are arranged in ascending or descending order. If the number of data points is even, the median is usually estimated as the midpoint between the two middle values – by adding the two middle values and dividing their sum by 2. The median is the 50th percentile.

Statistics Associated with Frequency Distribution: Measures of Variability (1 of 2)

- The **range** measures the spread of the data. It is simply the difference between the largest and smallest values in the sample.

$$\text{Range} = X_{\textbf{largest}} - X_{\textbf{smallest}}$$

- The **interquartile range** is the difference between the 75th and 25th percentile. For a set of data points arranged in order of magnitude, the p^{th} percentile is the value that has p% of the data points below it and (100 − p)% above it.

Statistics Associated with Frequency Distribution: Measures of Variability (2 of 2)

- The **variance** is the mean squared deviation from the mean. The variance can never be negative.
- The **standard deviation** is the square root of the variance.

$$s_x = \sqrt{\sum_{i=1}^{n} \frac{\left(X_i - \overline{X}\right)^2}{n-1}}$$

- The **coefficient of variation** is the ratio of the standard deviation to the mean expressed as a percentage, and is a unitless measure of relative variability.

$$CV = s_x / \overline{X}$$

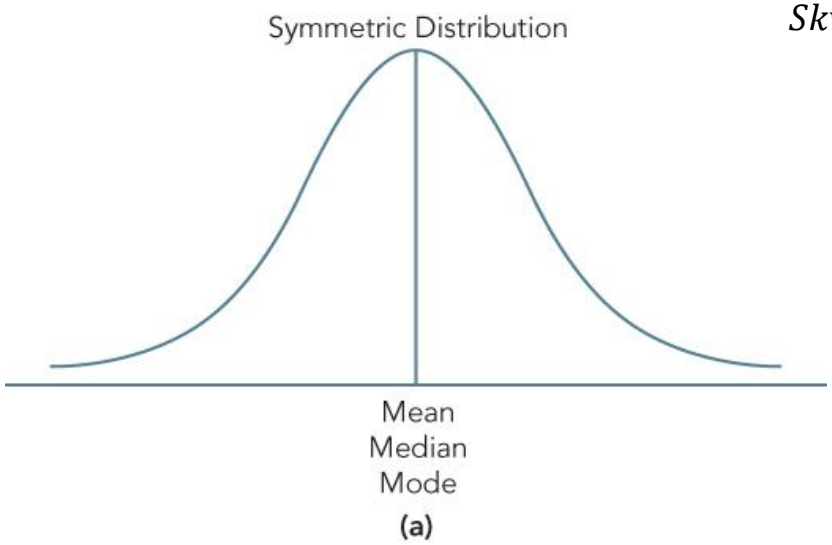
Statistics Associated with Frequency Distribution: Measures of Shape

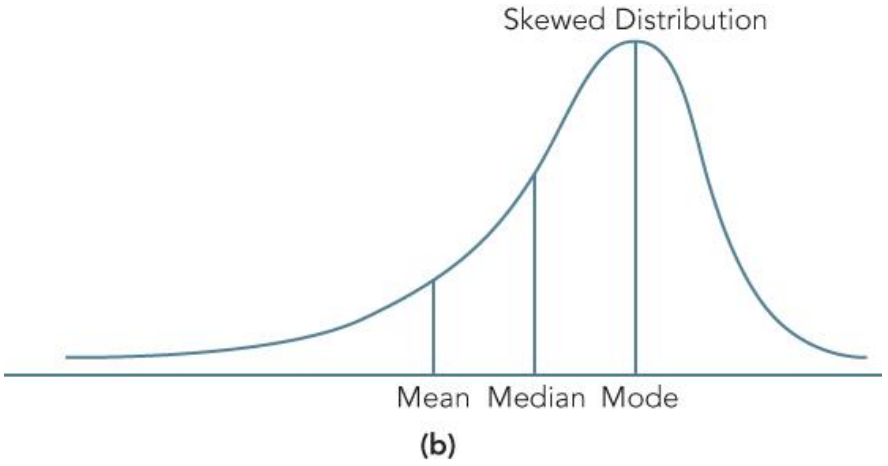
- **Skewness.** The tendency of the deviations from the mean to be larger in one direction than in the other. It can be thought of as the tendency for one tail of the distribution to be heavier than the other.
- **Kurtosis** is a measure of the relative peakedness or flatness of the curve defined by the frequency distribution. The kurtosis of a normal distribution is zero. If the kurtosis is positive, then the distribution is more peaked than a normal distribution. A negative value means that the distribution is flatter than a normal distribution.

Skewness of a Distribution

Figure 15.2 Skewness of a Distribution

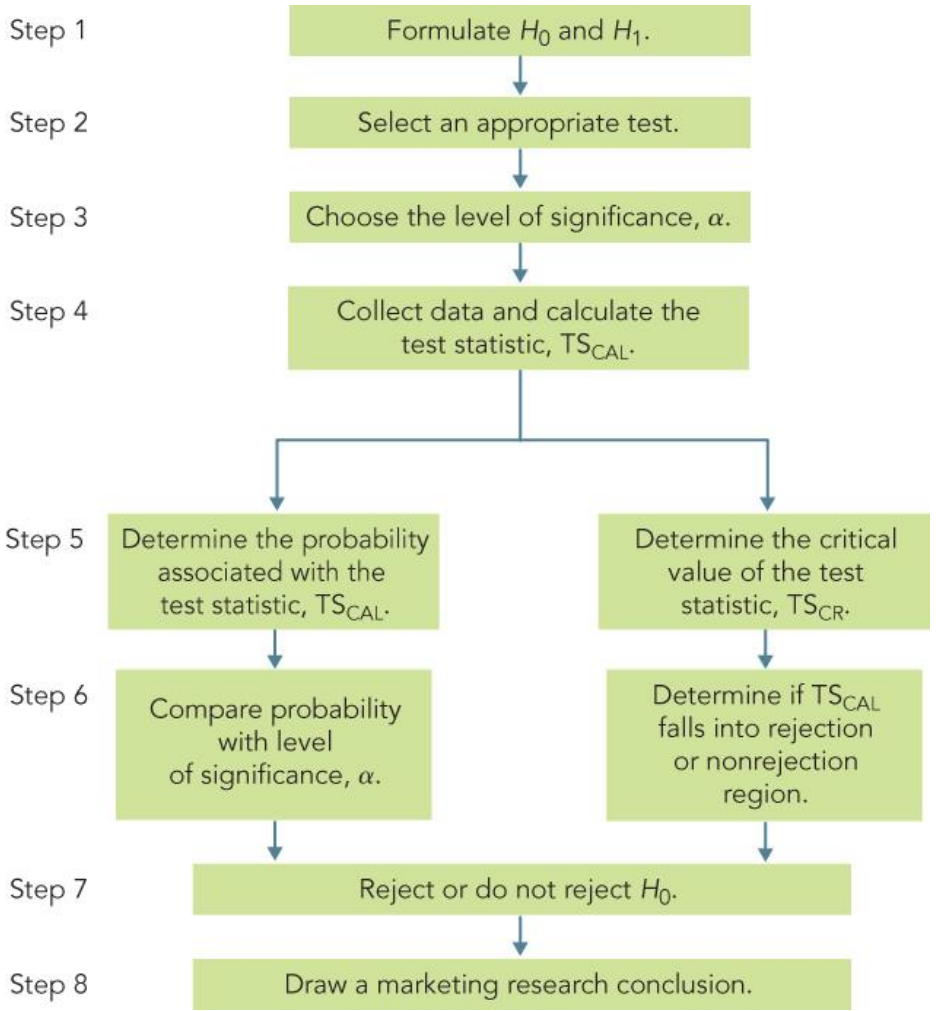
$$Skw = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left[\sqrt{\frac{1}{n-1}(x_i - \bar{x})^2}\right]^3}, Kurt = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left[\sqrt{\frac{1}{n-1}(x_i - \bar{x})^2}\right]^4}$$





Steps Involved in Hypothesis Testing

Figure 15.3 A General Procedure for Hypothesis Testing



A General Procedure for Hypothesis Testing
Step 1: Formulate the Hypothesis (1 of 3)

- A **null hypothesis** is a statement of the status quo, one of no difference, no action, or no effect. If the null hypothesis is not rejected, no changes will be made.
- An **alternative hypothesis** is one in which some difference or effect is expected. Accepting the alternative hypothesis will lead to changes in opinions or actions.
- The null hypothesis refers to a specified value of the population parameter (e.g., μ, σ, π), not a sample statistic (e.g., $\bar{X}$).

A General Procedure for Hypothesis Testing
Step 1: Formulate the Hypothesis (2 of 3)

- A null hypothesis may be rejected, but it can never be accepted based on a single test. In classical hypothesis testing, there is no way to determine whether the null hypothesis is true.
- In marketing research, the null hypothesis is formulated in such a way that its rejection leads to the acceptance of the desired conclusion. The alternative hypothesis represents the conclusion for which evidence is sought.

$$H_0: \pi \leq 0.40$$
$$H_1: \pi > 0.40$$

A General Procedure for Hypothesis Testing
Step 1: Formulate the Hypothesis (3 of 3)

- The test of the null hypothesis is a **one-tailed test**, because the alternative hypothesis is expressed directionally. If that is not the case, then a **two-tailed test** would be required, and the hypotheses would be expressed as:

$$H_0: \pi = 0.40$$
$$H_1: \pi \neq 0.40$$

A General Procedure for Hypothesis Testing Step 2: Select an Appropriate Test

- The **test statistic** measures how close the sample has come to the null hypothesis.
- The test statistic often follows a well-known distribution, such as the normal, *t*, or chi-square distribution.
- In our example, the ***z* statistic**, which follows the standard normal distribution, would be appropriate.

$$z = \frac{p - \pi}{\sigma_p}$$

where

$$\sigma_p = \sqrt{\frac{\pi\ (1 - \pi)}{n}}$$

A General Procedure for Hypothesis Testing
Step 3: Choose a Level of Significance (1 of 2)

Type I Error

- **Type I error** occurs when the sample results lead to the rejection of the null hypothesis when it is in fact true.
- The probability of type I error (α) is also called the **level of significance**.

Type II Error

- **Type II error** occurs when, based on the sample results, the null hypothesis is not rejected when it is in fact false.
- The probability of type II error is denoted by β.
- Unlike α, which is specified by the researcher, the magnitude of β depends on the actual value of the population parameter (proportion).

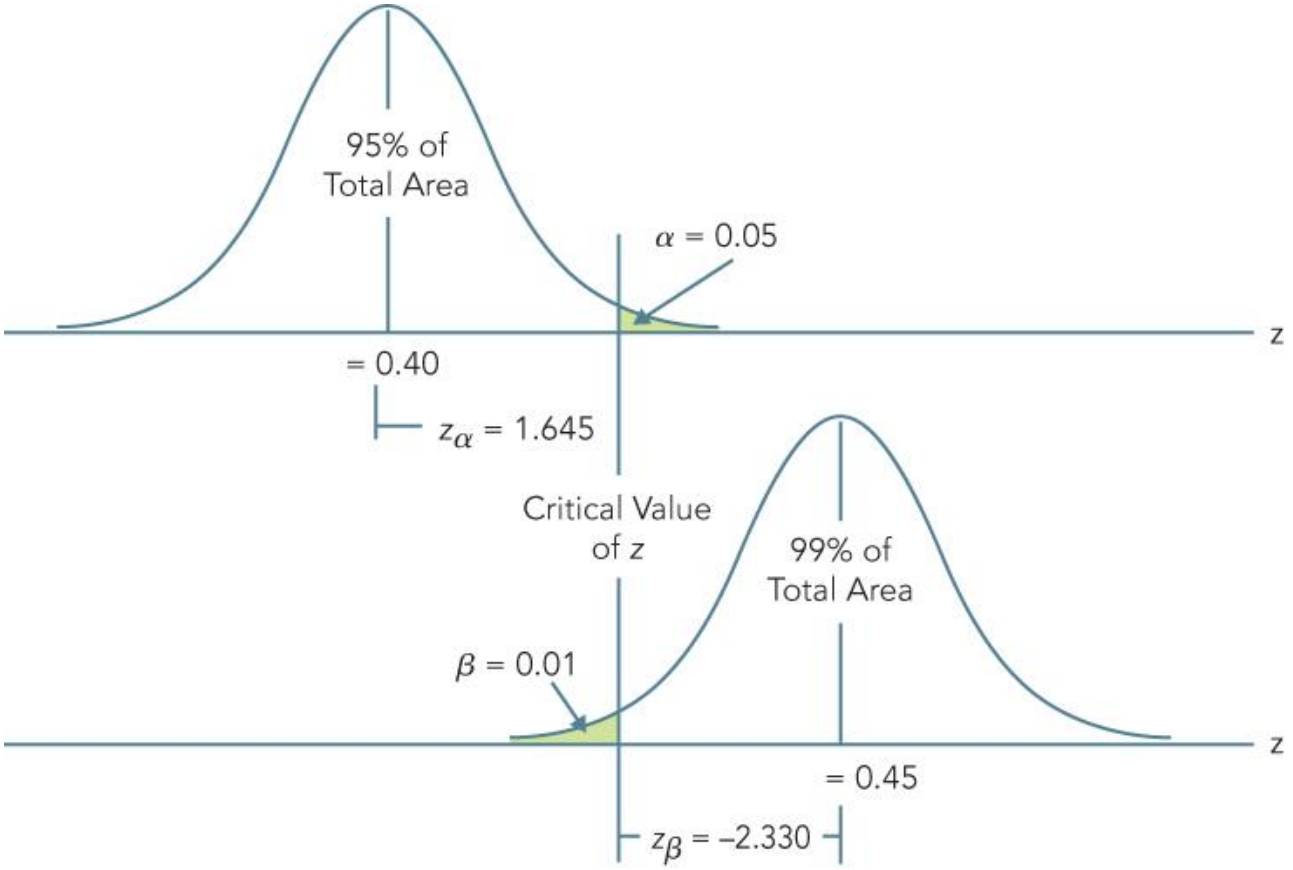
A General Procedure for Hypothesis Testing
Step 3: Choose a Level of Significance (2 of 2)

Power of a Test

- The **power of a test** is the probability (1 − β) of rejecting the null hypothesis when it is false and should be rejected.
- Although β is unknown, it is related to α. An extremely low value of α (e.g.,α = 0.001) will result in intolerably high β errors.
- So it is necessary to balance the two types of errors.

Probabilities of Type I & Type II Error

Figure 15.4 Type I Error (α) and Type II Error (β)



A General Procedure for Hypothesis Testing
Step 4: Collect Data and Calculate Test Statistic
(1 of 2)

- The required data are collected and the value of the test statistic computed.
- In our example, the value of the sample proportion is $p = 17/30 = 0.567$.
- The value of σ_p can be determined as follows:

$$\sigma_p = \sqrt{\frac{\pi\ (1-\pi)}{n}}$$
$$= \sqrt{\frac{(0.40)(0.6)}{30}}$$
$$= 0.089$$

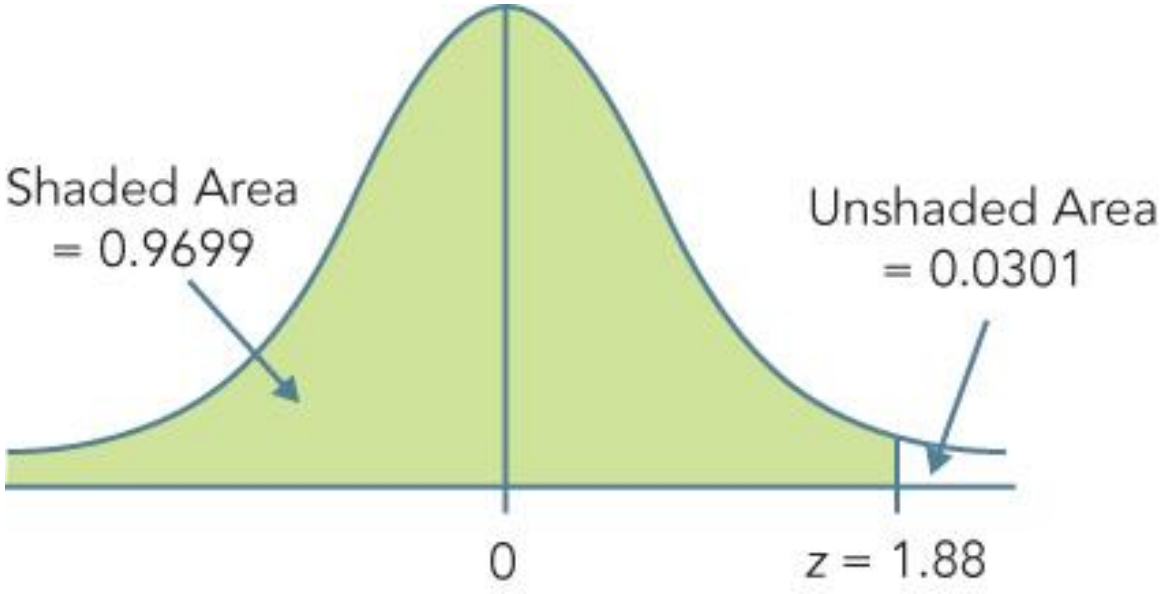
A General Procedure for Hypothesis Testing
Step 4: Collect Data and Calculate Test Statistic
(2 of 2)

The test statistic z can be calculated as follows:

$$z = \frac{\hat{p} - \pi}{\sigma_p}$$
$$= \frac{0.567 - 0.40}{0.089}$$
$$= 1.88$$

Probability of z with a One-Tailed Test

Figure 15.5 Probability of *z* with a One-Tailed Test



A General Procedure for Hypothesis Testing
Step 5: Determine the Probability (Critical Value)

- Using standard normal tables (Table 2 of the Statistical Appendix), the probability of obtaining a *z* value of 1.88 can be calculated (see Figure 15.5).
- The shaded area between - ∞ and 1.88 is 0.9699. Therefore, the area to the right of $z = 1.88$ is 1.0000 − 0.9699 = 0.0301 (*p-value*).
- Alternatively, the critical value of *z*, which will give an area to the right side of the critical value of 0.05, is between 1.64 and 1.65 and equals 1.645.
- Note, in determining the critical value of the test statistic, the area to the right of the critical value is either α or α/2. It is α for a one-tail test and α/2 for a two-tail test.

A General Procedure for Hypothesis Testing Steps 6 & 7: Compare the Probability (Critical Value) & Making the Decision (1 of 2)

- If the probability associated with the calculated or observed value of the test statistic (TSCAL) is less than the level of significance (α), the null hypothesis is rejected.
- The probability associated with the calculated or observed value of the test statistic is 0.0301. This is the probability of getting a *proportion (p)* value of 0.567 when π = 0.40. This is less than the level of significance of 0.05. Hence, the null hypothesis is rejected.
- Alternatively, if the absolute calculated value of the test statistic ($|TS_{CAL}|$) is greater than the absolute critical value of the test statistic ($|TS_{CR}|$), the null hypothesis is rejected.

A General Procedure for Hypothesis Testing Steps 6 & 7: Compare the Probability (Critical Value) & Making the Decision (2 of 2)

- The calculated value of the test statistic $z = 1.88$ lies in the rejection region, beyond the value of 1.645. Again, the same conclusion to reject the null hypothesis is reached.
- Note that the two ways of testing the null hypothesis are equivalent but mathematically opposite in the direction of comparison.
- If the probability of TS_{CAL} < significance level (α) then reject H_0 but if $|TS_{CAL}| > |TS_{CR}|$ then reject H_0.

A General Procedure for Hypothesis Testing Step 8: Marketing Research Conclusion

- The conclusion reached by hypothesis testing must be expressed in terms of the marketing research problem.

- In our example, we conclude that there is evidence that the proportion of Internet users who shop via the Internet is significantly greater than 0.40. Hence, the recommendation to the department store would be to introduce the new Internet shopping service.