



Source: © Getty Images

18

Frequency distribution, cross-tabulation and hypothesis testing

Objectives

After reading this chapter, you should be able to:

- 1 describe the significance of preliminary data analysis and the insights that can be obtained from such an analysis;
- 2 discuss data analysis associated with frequencies, including measures of location, measures of variability and measures of shape;
- 3 explain data analysis associated with cross-tabulations and the associated statistics: chi-square, phi coefficient, contingency coefficient, Cramer's V and lambda coefficient;
- 4 describe data analysis associated with parametric hypothesis testing for one sample, two independent samples and paired samples;
- 5 understand data analysis associated with non-parametric hypothesis testing for one sample, two independent samples and paired samples.

“

Frequency distribution, cross-tabulation and hypothesis testing are the fundamental building blocks of quantitative data analysis. They provide insights into the data, guide subsequent analyses and aid the interpretation of results.

”

STAGE 1
Problem definition

STAGE 2
Research approach developed

STAGE 3
Research design developed

STAGE 4
Fieldwork or data collection

STAGE 5
Data preparation and analysis

STAGE 6
Report preparation and presentation

Overview

Once the data have been prepared for analysis (Chapter 17), the researcher should conduct basic analyses. This chapter describes basic data analyses, including frequency distribution, cross-tabulation and hypothesis testing. First, we describe the frequency distribution and explain how it provides both an indication of the number of out-of-range, missing or extreme values as well as insights into the central tendency, variability and shape of the underlying distribution. Next, we introduce hypothesis testing by describing the general procedure. Hypothesis testing procedures are classified as tests of associations or tests of differences. We consider the use of cross-tabulation for understanding the associations between variables taken two or three at a time. Although the nature of the association can be observed from tables, statistics are available for examining the significance and strength of the association. Finally, we present tests for examining hypotheses related to differences based on one or two samples.

Many marketing research projects do not go beyond basic data analysis. These findings are often displayed using tables and graphs, as discussed further in Chapter 25. Although the findings of basic analysis are valuable in their own right, they also provide guidance for conducting multivariate analysis. The insights gained from the basic analysis are also invaluable in interpreting the results obtained from more sophisticated statistical techniques. The following examples provide a 'flavour' of basic data analysis techniques. We illustrate the use of cross-tabulation, chi-square analysis and hypothesis testing.

Focus on

Sports Marketing Surveys

Basic data analyses

In the Formula One Racetrack Project, basic data analysis formed the foundation for conducting subsequent multivariate analysis. Data analysis began by obtaining a frequency distribution and descriptive statistics for each variable or question asked in the survey. In addition to identifying possible problems with the data, this information provided a good idea of the data and insights into how specific variables should be treated in subsequent analyses. For example, should the interval scaled variable 'number of Grand Prix races watched on television' be treated as categorical? If so, how many categories should there be? (See the following example.) Several two- and three-variable cross-tabulations were also conducted to identify associations in the data. The effects of variables with two categories on the metric-dependent variables of interest were examined by means of t tests and other hypothesis testing procedures.

Focus on

Sports Marketing surveys

Formula One viewing habits

Formula One fans watched an average of 6.2 races during the 2003 season, a slight increase on 2002 (5.8). This was seen as good news after the impact of Michael Schumacher's runaway victory in 2002. The following table focuses upon the viewing habits of male and female respondents. The first question tackled was 'Out of the 16 Formula One Grands Prix held in the 2003 season, how many have you watched on television?' Sports Marketing Surveys used the question to create three different viewing groups:



- **Occasional** = watched 4 or less Grands Prix
- **Regular** = watched between 5 and 9 Grands Prix
- **Avid** = watched 10 or more Grand Prix.

The responses to this question were broken down into male and female respondents. Cross-tabulation and chi-square analysis provided the following:

Formula One races watched on TV	Respondent gender (%)	
	Male	Female
Occasional	59	74
Regular	6	7
Avid	35	19
	$\chi^2 = 6.51$	$p \leq 0.05$

These results indicate that there is little difference in male and female viewers in terms of regular viewing of Formula One races. However, there is a much higher proportion of men that are avid viewers.

Example Catalogues are risky business¹

Twelve product categories were examined to compare shopping by catalogue with store shopping. The hypothesis that there is no significant difference in the overall amount of risk perceived when buying products by catalogue compared with buying the same products in a retail store was rejected. The hypothesis was tested by computing 12 paired observation *t* tests, one for each product. Mean scores for overall perceived risk for some of the products in both buying situations are presented in the following table, with higher scores indicating greater risk.

Product	Overall perceived risk	
	Catalogue	Store
Shoes	58.60	50.80*
Pocket calculator	49.62	42.00*
Hi-fi	48.89	41.98*
Portable television	48.53	40.91*
Digital camera	48.13	39.52*
Athletic socks	35.22	30.22*
Perfume	34.85	29.79*
CDs	32.65	28.74*

* Significant at 0.01 level

As can be seen, a significantly ($p < 0.01$) higher overall amount of perceived risk was attached to products purchased by catalogue as compared with those purchased from a retail store.

The first Sports Marketing Surveys example illustrates the role of basic data analysis used in conjunction with multivariate procedures, whereas the other two examples show how such analysis can be useful in its own right. The cross-tabulation and chi-square analysis in the Formula One Racetrack example and the paired t tests in the catalogue shopping example enabled us to draw specific conclusions from the data. These and other concepts discussed in this chapter are illustrated in the context of explaining Internet usage for personal (non-professional) reasons. Table 18.1 contains data for 30 respondents giving the gender (1 = male, 2 = female), familiarity with the Internet (1 = very unfamiliar, 7 = very familiar), Internet usage in hours per week, attitude towards the Internet and towards technology, both measured on a seven-point scale (1 = very unfavourable, 7 = very favourable), and whether the respondents have done shopping or banking on the Internet (1 = yes, 2 = no). As a first step in the analysis, it is useful to examine the frequency distributions of the relevant variables.

Table 18.1 Internet usage data

<i>Respondent No.</i>	<i>Gender</i>	<i>Familiarity</i>	<i>Internet usage</i>	<i>Attitude towards Internet</i>	<i>Attitude towards technology</i>	<i>Usage of Internet shopping</i>	<i>Usage of Internet banking</i>
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
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



Marketing researchers often need to answer questions about a single variable. For example:

- How many users of the brand may be characterised as brand loyal?
- What percentage of the market consists of heavy users, medium users, light users and non-users?
- How many customers are very familiar with a new product offering? How many are familiar, somewhat familiar, or unfamiliar with the brand? What is the mean familiarity rating? Is there much variance in the extent to which customers are familiar with the new product?
- What is the income distribution of brand users? Is this distribution skewed towards low-income brackets?

Frequency distribution

A mathematical distribution whose objective is to obtain a count of the number of responses associated with different values of one variable and to express these counts in percentage terms.

The answers to these kinds of questions can be determined by examining **frequency distributions**. In a frequency distribution, one variable is considered at a time.

The objective is to obtain a count of the number of responses associated with different values of the variable. The relative occurrence, or frequency, of different values of the variable is expressed in percentages. A frequency distribution for a variable produces a table of frequency counts, percentages and cumulative percentages for all the values associated with that variable.

Table 18.2 gives the frequency distribution of familiarity with the Internet. In the table, the first column contains the labels assigned to the different categories of the variable and the second column indicates the codes assigned to each value. Note that a code of 9 has been assigned to missing values. The third column gives the number of respondents ticking each value. For example, three respondents ticked value 5, indicating that they were somewhat familiar with the Internet. The fourth column displays the percentage of respondents ticking each value.

The fifth column shows percentages calculated by excluding the cases with missing values. If there are no missing values, the fourth and fifth columns are identical. The last column represents cumulative percentages after adjusting for missing values. As can be seen, of the 30 respondents who participated in the survey, 10% entered a figure of '5'. If the one respondent with a missing value is excluded, this changes to 10.3%. The cumulative percentage corresponding to the value of 5 is 58.6. In other words, 58.6% of the respondents with valid responses indicated a familiarity value of 5 or less.

Table 18.2 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	

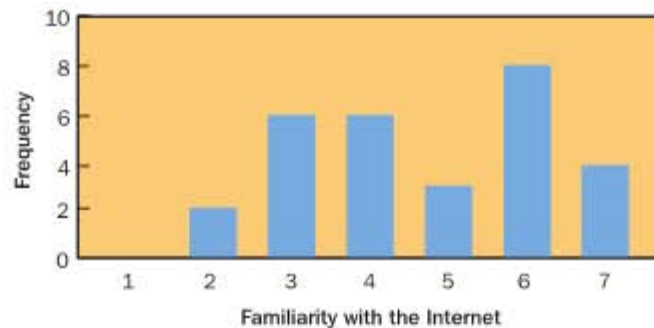


Figure 18.1
Frequency histogram

A frequency distribution helps determine the extent of item non-response (1 respondent out of 30 in Table 18.1). It also indicates the extent of illegitimate responses. Values of 0 and 8 would be illegitimate responses, or errors. The cases with these values could be identified and corrective action could be taken. The presence of outliers or cases with extreme values can also be detected. For example, in the case of a frequency distribution of household size, a few isolated families with household sizes of nine or more might be considered outliers. A frequency distribution also indicates the shape of the empirical distribution of the variable. The frequency data may be used to construct a histogram, or a vertical bar chart in which the values of the variable are portrayed along the x axis and the absolute or relative frequencies of the values are placed along the y axis.

Figure 18.1 is a histogram of the frequency data in Table 18.1. From the histogram, one could examine whether the observed distribution is consistent with an expected or assumed distribution.

Example

Basic analysis yields Olympic results²

For the 1996 Olympic Games in Atlanta, more than 2 million unique visitors came to the games and more than 11 million tickets were sold. Researchers at the University of

Motivational factors that influenced the decision to attend the Olympic Games

Motivational factor	Frequency	Percentage
Once-in-a-lifetime opportunity	95	29.7
Availability of housing	36	11.2
Availability of tickets	27	8.4
Distance away from home	24	7.5
Business/employment	17	5.3
Availability of money – overall expenses	17	5.3
Availability of time	12	3.8
Personal relationship with participant or official	8	2.5
Other motivational factor	8	2.5
Visit Atlanta	4	1.3
Security	3	0.9
Did not respond	69	21.6
Total	320	100.0



Colorado decided to discover what motivated international and domestic travellers to come to the Olympic Games in Atlanta. A survey was developed and administered to visitors via personal interviews during a nine-day period surrounding the completion of the games. Three hundred and twenty questionnaires were completed correctly and were used in the data analysis.

The results showed that the top three factors that motivated people to attend the games were: 'once-in-a-lifetime opportunity', 'availability of housing' and 'availability of tickets'. The results of this study helped planners for the 2000 Olympic Games in Sydney find what specific characteristics the city needed to improve. For instance, Sydney put funds into projects that added hotel rooms to the city. Sydney also constructed state-of-the-art transportation (a new elevated rail system) and unique venues (Olympic Park) so that visitors truly felt they were getting a once-in-a-lifetime experience. As this survey evolved over the years the findings have a cumulative effect in supporting future host cities.

Note that the numbers and percentages in the preceding example indicate the extent of the various motivational factors that attract individuals to the Olympic Games. Because numbers are involved, a frequency distribution can be used to calculate descriptive or cumulative statistics.

Statistics associated with frequency distribution



As illustrated in the previous section, a frequency distribution is a convenient way of looking at different values of a variable. A frequency table is easy to read and provides basic information, but sometimes this information may be too detailed and the researcher must summarise it by the use of descriptive statistics.³ The most commonly used statistics associated with frequencies are measures of location (mean, mode and median), measures of variability (range, interquartile range, variance, standard deviation and coefficient of variation) and measures of shape (skewness and kurtosis).⁴

Measures of location

The **measures of location** that we discuss are measures of central tendency because they tend to describe the centre of the distribution. If the entire sample is changed by adding a fixed constant to each observation, then the mean, mode and median change by the same fixed amount.

Measure of location

A statistic that describes a location within a dataset. Measures of central tendency describe the centre of the distribution.

Mean

The average; that value obtained by summing all elements in a set and dividing by the number of elements.

Mean. The **mean**, or average value, is the most commonly used measure of central tendency. The measure is used to estimate the mean when the data have been collected using an interval or ratio scale. The data should display some central tendency, with most of the responses distributed around the mean.

The mean, \bar{X} , is given by

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

where

X_i = observed values of the variable X
 n = number of observations (sample size).

Generally, the mean is a robust measure and does not change markedly as data values are added or deleted. For the frequencies given in Table 18.1, the mean value is calculated as follows:

$$\begin{aligned}\bar{X} &= (2 \times 2) + (6 \times 3) + (6 \times 4) + (3 \times 5) + (8 \times 6) + (4 \times 7)/29 \\ &= (4 + 18 + 24 + 15 + 48 + 28)/29 \\ &= 137/29 \\ &= 4.724\end{aligned}$$

Mode

A measure of central tendency given as the value that occurs with the most frequency in a sample distribution.

Median

A measure of central tendency given as the value above which half of the values fall and below which half of the values fall.

Mode. 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. The mode in Table 18.2 is 6.

Median. 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. The median is an appropriate measure of central tendency for ordinal data. In Table 18.2, the middle value is 5, so the median is 5.

As can be seen from Table 18.1, the three measures of central tendency for this distribution are different (mean = 4.724, mode = 6, median = 5). This is not surprising, since each measure defines central tendency in a different way. So which measure should be used? If the variable is measured on a nominal scale, the mode should be used. If the variable is measured on an ordinal scale, the median is appropriate. If the variable is measured on an interval or ratio scale, the mode is a poor measure of central tendency. This can be seen from Table 18.2. Although the modal value of 6 has the highest frequency, it represents only 27.6% of the sample. In general, for interval or ratio data, the median is a better measure of central tendency, although it too ignores available information about the variable. The actual values of the variable above and below the median are ignored. The mean is the most appropriate measure of central tendency for interval or ratio data. The mean makes use of all the information available since all of the values are used in computing it. However, it is sensitive to extremely small or extremely large values (outliers). When there are outliers in the data, the mean is not a good measure of central tendency, and it is useful to consider both the mean and the median.

Measures of variability

The **measures of variability**, which are calculated on interval or ratio data, include the range, interquartile range, variance or standard deviation and coefficient of variation.

Measure of variability

A statistic that indicates the distribution's dispersion.

Range

The difference between the smallest and largest values of a distribution.

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

$$\text{range} = X_{\text{largest}} - X_{\text{smallest}}$$

As such, the range is directly affected by outliers. If all the values in the data are multiplied by a constant, the range is multiplied by the same constant. The range in Table 18.2 is $7 - 2 = 5$.

Interquartile range

The range of a distribution encompassing the middle 50% of the observations.

Interquartile range. The **interquartile range** is the difference between the 75th and 25th percentiles. 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. If all the data points are multiplied by a constant, the interquartile range is multiplied by the same constant. The interquartile range in Table 18.2 is $6 - 3 = 3$.

Variance

The mean squared deviation of all the values of the mean.

Variance. The difference between the mean and an observed value is called the deviation from the mean. The **variance** is the mean squared deviation from the mean. The variance can never be negative. When the data points are clustered around the mean, the variance

is small. When the data points are scattered, the variance is large. If all the data values are multiplied by a constant, the variance is multiplied by the square of the constant.

Standard deviation

The square root of the variance.

Standard deviation. The **standard deviation** is the square root of the variance. Thus, the standard deviation is expressed in the same units as the data, rather than in squared units. The standard deviation of a sample, s is calculated as

$$s = \sqrt{\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}}$$

We divide by $n-1$ instead of n because the sample is drawn from a population and we are trying to determine how much the responses vary from the mean of the entire population. The population mean is unknown, however; therefore, the sample mean is used instead. The use of the sample mean makes the sample seem less variable than it really is. By dividing by $n-1$ instead of by n , we compensate for the smaller variability observed in the sample. For the data given in Table 18.1, the variance is calculated as follows:

$$\begin{aligned} S^2 &= [2 \times (2 - 4.724)^2 + 6 \times (3 - 4.724)^2 + 6 \times (4 - 4.724)^2 + 3 \times (5 - 4.724)^2 \\ &\quad + 8 \times (6 - 4.724)^2 + 4 \times (7 - 4.724)^2] / 28 \\ &= [14.840 + 17.833 + 3.145 + 0.229 + 13.025 + 20.721] / 28 \\ &= 69.793 / 28 \\ &= 2.493 \end{aligned}$$

The standard deviation, therefore, is calculated as

$$\begin{aligned} S &= \sqrt{2.493} \\ &= 1.579 \end{aligned}$$

Coefficient of variation

A useful expression in sampling theory for the standard deviation as a percentage of the mean.

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

$$CV = \frac{S}{\bar{X}}$$

The coefficient of variation is meaningful only if the variable is measured on a ratio scale. It remains unchanged if all the data values are multiplied by a constant. Because familiarity with the Internet is not measured on a ratio scale, it is not meaningful to calculate the coefficient of variation for the data in Table 18.2.

Measures of shape



In addition to measures of variability, measures of shape are also useful in understanding the nature of the distribution. The shape of a distribution is assessed by examining skewness and kurtosis.

Skewness. Distributions can be either symmetric or skewed. In a symmetric distribution, the values on either side of the centre of the distribution are the same, and the mean, mode and median are equal. The positive and corresponding negative deviations from the mean are also equal. In a skewed distribution, the positive and negative deviations from the mean are unequal. **Skewness** is 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

Skewness

A characteristic of a distribution that assesses its symmetry about the mean.

of the distribution to be heavier than the other (see Figure 18.2). The skewness value for the data of Table 18.1 is -0.094 , indicating a slight negative skew.

Kurtosis

A measure of the relative peakedness of the curve defined by the frequency distribution.

Kurtosis. 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. The value of this statistic for Table 18.2 is -1.261 , indicating that the distribution is flatter than a normal distribution.

A general procedure for hypothesis testing

Basic analysis invariably involves some hypothesis testing. Examples of hypotheses generated in marketing research abound:

- A cinema is being patronised by more than 10% of the households in a city.
- The heavy and light users of a brand differ in terms of psychographic characteristics.
- One hotel has a more 'luxurious' image than its close competitor.
- Familiarity with a restaurant results in greater preference for that restaurant.

Chapter 15 covered the concepts of the sampling distribution, standard error of the mean or the proportion, and the confidence interval.⁵ All these concepts are relevant to hypothesis testing and should be reviewed. We now describe a general procedure for hypothesis testing that can be applied to test hypotheses about a wide range of parameters.

The following steps are involved in hypothesis testing (Figure 18.3).

- 1 Formulate the null hypothesis H_0 and the alternative hypothesis H_1 .
- 2 Select an appropriate statistical technique and the corresponding test statistic.
- 3 Choose the level of significance, α .
- 4 Determine the sample size and collect the data. Calculate the value of the test statistic.
- 5 Determine the probability associated with the test statistic under the null hypothesis, using the sampling distribution of the test statistic. Alternatively, determine the critical values associated with the test statistic that divide the rejection and non-rejection region.
- 6 Compare the probability associated with the test statistic with the level of significance specified. Alternatively, determine whether the test statistic has fallen into the rejection or the non-rejection region.
- 7 Make the statistical decision to reject or not reject the null hypothesis.
- 8 Express the statistical decision in terms of the marketing research problem.

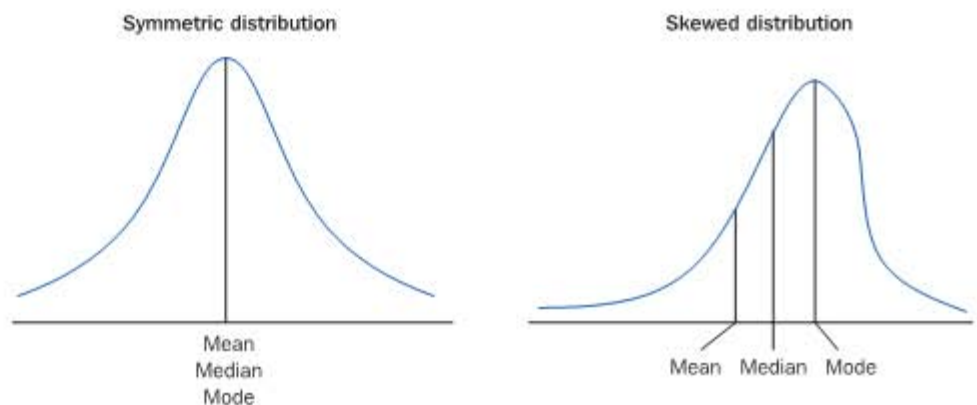


Figure 18.2
Skewness of a distribution

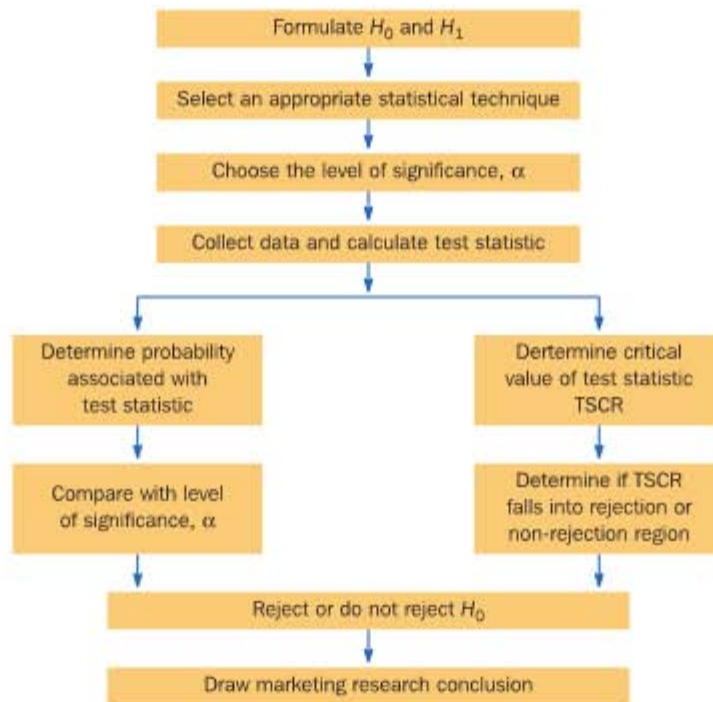


Figure 18.3
A general procedure for hypothesis testing

Step 1: Formulate the hypothesis

Null hypothesis

A statement in which no difference or effect is expected. If the null hypothesis is not rejected, no changes will be made.

Alternative hypothesis

A statement that some difference or effect is expected. Accepting the alternative hypothesis will lead to changes in opinions or actions.

The first step is to formulate the null and alternative hypotheses. A **null hypothesis** is a statement of the status quo, one of no difference 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. Thus, the alternative hypothesis is the opposite of the null hypothesis.

The null hypothesis is always the hypothesis that is tested. 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 null hypothesis may be rejected, but it can never be accepted based on a single test. A statistical test can have one of two outcomes: that the null hypothesis is rejected and the alternative hypothesis accepted, or that the null hypothesis is not rejected based on the evidence. It would be incorrect, however, to conclude that since the null hypothesis is not rejected, it can be accepted as valid. 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. For example, a Formula One sponsor is considering introducing an Internet shopping service. Given the investment in systems and personnel to make this plan work, it will only be introduced if more than 40% of Internet users shop via the Internet. The appropriate way to formulate the hypotheses is

$$H_0: \pi \leq 0.40$$

$$H_1: \pi > 0.40$$

If the null hypothesis H_0 is rejected, then the alternative hypothesis H_1 will be accepted and the new Internet shopping service introduced. On the other hand, if H_0 is not rejected, then the new Internet shopping service should not be introduced unless additional evidence is obtained. The test of the null hypothesis is a **one-tailed test** because the alternative hypothesis is expressed directionally: the proportion of customers who express a preference is greater than 0.40.

One-tailed test

A test of the null hypothesis where the alternative hypothesis is expressed directionally.

Two-tailed test

A test of the null hypothesis where the alternative hypothesis is not expressed directionally.

On the other hand, suppose that the researcher wanted to determine whether the proportion of Internet users who shop via the Internet is different than 40%. Then a **two-tailed test** would be required, and the hypotheses would be expressed as

$$H_0: \pi = 0.40$$

$$H_1: \pi \neq 0.40$$

In commercial marketing research, the one-tailed test is used more often than a two-tailed test. Typically, there is some preferred direction for the conclusion for which evidence is sought. For example, the higher the profits, sales and product quality, the better. The one-tailed test is more powerful than the two-tailed test. The power of a statistical test is discussed further in step 3.

Test statistic

A measure of how close the sample has come to the null hypothesis. It often follows a well-known distribution, such as the normal, t , or chi-square distribution.

Step 2: Select an appropriate statistical technique

To test the null hypothesis, it is necessary to select an appropriate statistical technique. The researcher should take into consideration how the **test statistic** is computed and the sampling distribution that the sample statistic (e.g. the mean) follows. 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. Guidelines for selecting an appropriate test or statistical technique are discussed later in this chapter. In our example, the z statistic, which follows the standard normal distribution, would be appropriate. This statistic would be computed as follows:

$$z = \frac{p - \pi}{\sigma_p} \text{ where } \sigma_p = \sqrt{\frac{\pi(1 - \pi)}{n}}$$

Step 3: Choose the level of significance

Whenever we draw inferences about a population, there is a risk that an incorrect conclusion will be reached. Two types of error can occur.

Type I error

An error that occurs when the sample results lead to the rejection of a null hypothesis that is in fact true. Also called alpha error (α)

Level of significance

The probability of making a Type I error.

Type II error

An error that occurs when the sample results lead to acceptance of a null hypothesis that is in fact false. Also called beta error (β).

Type I error occurs when the sample results lead to the rejection of the null hypothesis when it is in fact true. In our example, a Type I error would occur if we concluded, based on sample data, that the proportion of customers preferring the new service plan was greater than 0.40, when in fact it was less than or equal to 0.40. The probability of Type I error (α) is also called the **level of significance**. The Type I error is controlled by establishing the tolerable level of risk of rejecting a true null hypothesis. The selection of a particular risk level should depend on the cost of making a Type I error.

Type II error occurs when, based on the sample results, the null hypothesis is not rejected when it is in fact false. In our example, the Type II error would occur if we concluded, based on sample data, that the proportion of customers preferring the new service plan was less than or equal to 0.40 when in fact it was greater than 0.40. 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). The probability of Type I error (α) and the probability of Type II error (β) are shown in Figure 18.4.

The complement ($1 - \beta$) of the probability of a Type II error is called the **power of a statistical test**. The power of a test is the probability ($1 - \beta$) 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. 0.001) will result in intolerably high β errors. So it is necessary to balance the two types of errors. As a compromise, α is often set at 0.05; sometimes it is 0.01; other values of α are rare. The level of along with the sample size will determine the level of for a particular research design. The risk of both α and β can be controlled by increasing the sample size. For a given level of α , increasing the sample size will decrease β , thereby increasing the power of the test.

Power of a statistical test

The probability of rejecting the null hypothesis when it is in fact false and should be rejected.

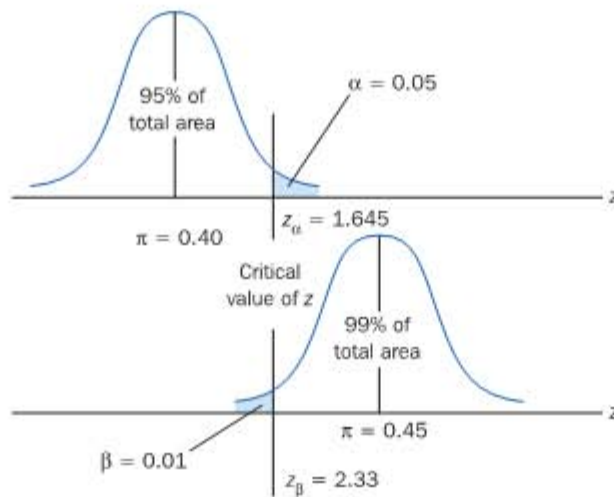


Figure 18.4
Type I error (α) and
Type II error (β)

Step 4: Collect the data and calculate the test statistic

Sample size is determined after taking into account the desired α and β errors and other qualitative considerations, such as budget constraints. Then the required data are collected and the value of the test statistic is computed. In our example, 30 users were surveyed and 17 indicated that they used the Internet for shopping. Thus the value of the sample proportion is $p = 17/30 = 0.567$. The value of σ_p can be determined as follows:

$$\begin{aligned} \sigma_p &= \sqrt{\pi(1-\pi)/n} \\ &= \sqrt{0.4 \times 0.6/30} \\ &= 0.089 \end{aligned}$$

The test statistic z can be calculated as follows:

$$\begin{aligned} z &= \frac{p - \pi}{\sigma_p} \\ &= \frac{0.567 - 0.40}{0.089} \\ &= 1.88 \end{aligned}$$

Step 5: Determine the probability or the critical value

Using standard normal tables (Table 2 of the Appendix of statistical tables), the probability of obtaining a z value of 1.88 can be calculated (see Figure 18.5). The shaded area between $-\infty$ and 1.88 is 0.96. Therefore, the area to the right of $z = 1.88$ is $1.0000 - 0.9699 = 0.03$.

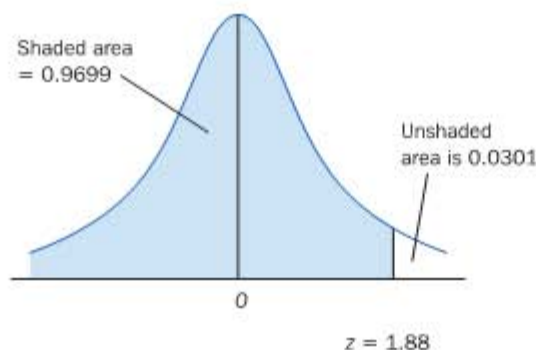


Figure 18.5
Probability of z with a
one-tailed test

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 that, in determining the critical value of the test statistic, the area to the right of the critical value is either α or $\alpha/2$. It is α for a one-tailed test and $\alpha/2$ for a two-tailed test.

Steps 6 and 7: Compare the probability or critical values and make the decision

The probability associated with the calculated or observed value of the test statistic is 0.0301. This is the probability of getting a p value of 0.567 when $\pi = 0.40$. This is less than the level of significance of 0.05. Hence, the null hypothesis is rejected. Alternatively, 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 associated with the calculated or observed value of the test statistic (TS_{CAL}) is less than the level of significance (α), the null hypothesis is rejected. If the calculated value of the test statistic is greater than the critical value of the test statistic (TS_{CR}), however, the null hypothesis is rejected. The reason for this sign shift is that the larger the value of TS_{CAL} , the smaller the probability of obtaining a more extreme value of the test statistic under the null hypothesis. This sign shift can be easily seen:

if probability of $TS_{CAL} < \text{significance level } (\alpha)$, then reject H_0

but

if $TS_{CAL} > TS_{CR}$, then reject H_0

Step 8: Draw the 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 would be to introduce the new Internet shopping service.

As can be seen from Figure 18.6, hypothesis testing can be related to either an examination of associations or an examination of differences. In tests of associations the null hypothesis is that there is no association between the variables (H_0 : ... is *not* related to ...). In tests of differences the null hypothesis is that there is no difference (H_0 : ... is *not* different than ...). Tests of differences could relate to distributions, means, proportions, or medians or rankings. First, we discuss hypotheses related to associations in the context of cross-tabulations.

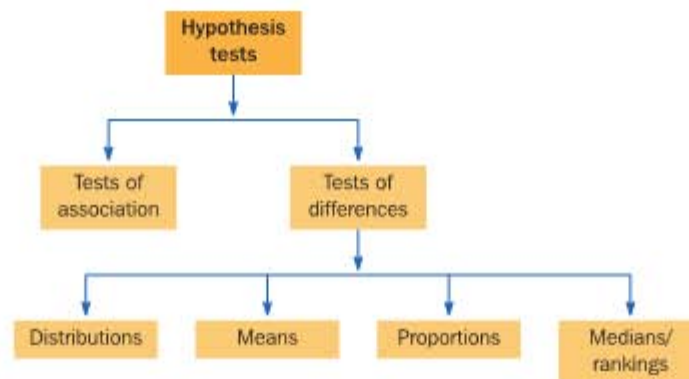


Figure 18.6
A broad classification of hypothesis testing procedures

Cross-tabulations



Although answers to questions related to a single variable are interesting, they often raise additional questions about how to link that variable to other variables. To introduce the frequency distribution, we posed several representative marketing research questions. For each of these, a researcher might pose additional questions to relate these variables to other variables. For example:

- How many brand-loyal users are males?
- Is product use (measured in terms of heavy users, medium users, light users and non-users) related to interest in outdoor leisure activities (high, medium and low)?
- Is familiarity with a new product related to age and income levels?
- Is product ownership related to income (high, medium and low)?

Cross-tabulation

A statistical technique that describes two or more variables simultaneously and results in tables that reflect the joint distribution of two or more variables that have a limited number of categories or distinct values.

The answers to such questions can be determined by examining **cross-tabulations**. A frequency distribution describes one variable at a time, but a cross-tabulation describes two or more variables simultaneously. Cross-tabulation results in tables that reflect the joint distribution of two or more variables with a limited number of categories or distinct values. The categories of one variable are cross-classified with the categories of one or more other variables. Thus, the frequency distribution of one variable is subdivided according to the values or categories of the other variables.

Suppose we are interested in determining whether Internet usage is related to gender. For the purpose of cross-tabulation, respondents are classified as 'light' or 'heavy' users. Those reporting five hours or less usage were classified as 'light' users, and the remaining were 'heavy' users. The cross-tabulation is shown in Table 18.3. A cross-tabulation includes a cell for every combination of the categories of the two variables. The number in each cell shows how many respondents gave that combination of responses. In Table 18.3, 10 respondents were females who reported light Internet usage. The marginal totals in this table indicate that of the 30 respondents with valid responses on both variables, 15 reported light usage and 15 were heavy users. In terms of gender, 15 respondents were females and 15 were males. Note that this information could have been obtained from a separate frequency distribution for each variable. In general, the margins of a cross-tabulation show the same information as the frequency tables for each of the variables.

Contingency table

A cross-tabulation table. It contains a cell for every combination of categories of the two variables.

Cross-tabulation tables are also called **contingency tables**. The data are considered to be qualitative or categorical data, because each variable is assumed to have only a nominal scale.⁶ Cross-tabulation is widely used in commercial marketing research because (1) cross-tabulation analysis and results can be easily interpreted and understood by managers who are not statistically oriented; (2) the clarity of interpretation provides a stronger link between research results and managerial action; (3) a series of cross-tabulations may provide greater insights into a complex phenomenon than a single multivariate analysis; (4) cross-tabulation may alleviate the problem of sparse cells, which could be serious in discrete multivariate analysis; and (5) cross-tabulation analysis is simple to conduct and appealing to both qualitative and quantitative researchers.⁷ We will discuss cross-tabulation for two and three variables.

Table 18.3 Gender and Internet usage

Internet usage	Gender		Row total
	Male	Female	
Light (1)	5	10	15
Heavy (2)	10	5	15
Column total	15	15	

Two variables

Cross-tabulation with two variables is also known as bivariate cross-tabulation. Consider again the cross-classification of Internet usage with gender given in Table 18.3. Is usage related to gender? It appears to be from Table 18.3. We see that disproportionately more of the respondents who are male are heavy Internet users as compared with females. Computation of percentages can provide more insight.

Because two variables have been cross-classified, percentages could be computed either column-wise, based on column totals (Table 18.4), or row-wise, based on row totals (Table 18.5). Which table is more useful?

Table 18.4 Gender and Internet usage – column totals

Internet usage	Gender	
	Male	Female
Light (1)	33.3%	66.7%
Heavy (2)	66.7%	33.3%
Column total	100%	100%

Table 18.5 Gender and Internet usage – row totals

Internet usage	Gender		Row total
	Male	Female	
Light (1)	33.3%	66.7%	100%
Heavy (2)	66.7%	33.3%	100%

The answer depends on which variable will be considered as the independent variable and which as the dependent variable.⁸ The general rule is to compute the percentages in the direction of the independent variable, across the dependent variable. In our analysis, gender may be considered as the independent variable and Internet usage as the dependent variable, and the correct way of calculating percentages is shown in Table 18.4. Note that whereas 66.7% of the males are heavy users, only 33.3% of females fall into this category. This seems to indicate that males are more likely to be heavy users of the Internet as compared with females.

Note that computing percentages in the direction of the dependent variable across the independent variable, as shown in Table 18.5, is not meaningful in this case. Table 18.5 implies that heavy Internet usage causes people to be males. This latter finding is implausible. It is possible, however, that the association between Internet usage and gender is mediated by a third variable, such as age or income. This kind of possibility points to the need to examine the effect of a third variable.

Three variables

Often the introduction of a third variable clarifies the initial association (or lack of it) observed between two variables. As shown in Figure 18.7, the introduction of a third variable can result in four possibilities:

- 1 It can refine the association observed between the two original variables.
- 2 It can indicate no association between the two original variables, although an association was initially observed. In other words, the third variable indicates that the initial association between the two variables was spurious.

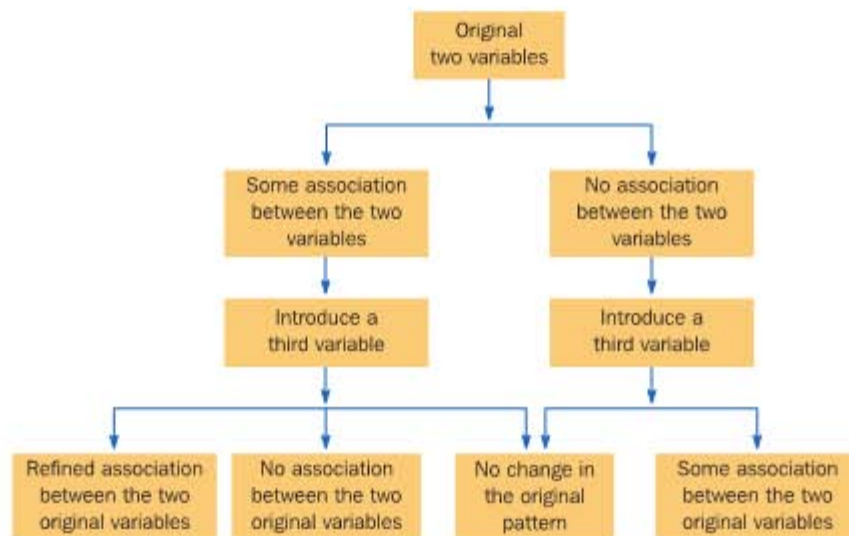


Figure 18.7
The introduction of a third variable in cross-tabulation

- 3 It can reveal some association between the two original variables, although no association was initially observed. In this case, the third variable reveals a suppressed association between the first two variables.
- 4 It can indicate no change in the initial association.⁹

These cases are explained with examples based on a sample of 1,000 respondents. Although these examples are contrived to illustrate specific cases, such cases are not uncommon in commercial marketing research.

Refine an initial relationship. An examination of the relationship between the purchase of ‘designer’ clothing and marital status resulted in the data reported in Table 18.6. The respondents were classified into either high or low categories based on their purchase of ‘designer’ clothing. Marital status was also measured in terms of two categories: currently married or unmarried. As can be seen from Table 18.6, 52% of unmarried respondents fell in the high-purchase category as opposed to 31% of the married respondents. Before concluding that unmarried respondents purchase more ‘designer’ clothing than those who are married, a third variable, the buyer’s gender, was introduced into the analysis.

Table 18.6 Purchase of ‘designer’ clothing by marital status

Purchase of ‘designer’ clothing	Marital status	
	Married	Unmarried
High	31%	52%
Low	69%	48%
Column	100%	100%
Number of respondents	700	300

The buyer’s gender was selected as the third variable based on past research. The relationship between purchase of ‘designer’ clothing and marital status was re-examined in light of the third variable, as shown in Table 18.7. In the case of females, 60% of the unmarried respondents fall in the high-purchase category compared with 25% of those who are married. On the other hand, the percentages are much closer for males, with 40%

Table 18.7 Purchase of 'designer' clothing by marital status and gender

Purchase of 'designer' clothing	Gender			
	Male marital status		Female marital status	
	Married	Unmarried	Married	Unmarried
High	35%	40%	25%	60%
Low	65%	60%	75%	40%
Column	100%	100%	100%	100%
Number of respondents	400	120	300	180

of the unmarried respondents and 35% of the married respondents falling in the high-purchase category. Hence, the introduction of gender (third variable) has refined the relationship between marital status and purchase of 'designer' clothing (original variables). Unmarried respondents are more likely to fall into the high-purchase category than married ones, and this effect is much more pronounced for females than for males.

Initial relationship was spurious. A researcher working for an advertising agency promoting a car brand costing more than €70,000 was attempting to explain the ownership of expensive cars (see Table 18.8). The table shows that 32% of those with university degrees own an expensive (more than €70,000) car compared with 21% of those without university degrees. The researcher was tempted to conclude that education influenced ownership of expensive cars. Realising that income may also be a factor, the researcher decided to re-examine the relationship between education and ownership of expensive cars in the light of income level. This resulted in Table 18.9. Note that the percentages of those with and without university degrees who own expensive cars are the same for each income group. When the data for the high-income and low-income groups are examined separately, the association between education and ownership of expensive cars disappears, indicating that the initial relationship observed between these two variables was spurious.



Source: © Alamy

Table 18.8 Ownership of expensive cars by education level

Own expensive car	Education	
	Degree	No degree
Yes	32%	21%
No	68%	79%
Column	100%	100%
Number of respondents	250	750

Table 18.9 Ownership of expensive cars by education and income levels

Own expensive car	Income			
	Low-income education		High-income education	
	Degree	No degree	Degree	No degree
Yes	20%	20%	40%	40%
No	80%	80%	60%	60%
Column totals	100%	100%	100%	100%
Number of respondents	100	700	150	50

Reveal suppressed association. A researcher suspected that desire to travel abroad may be influenced by age. A cross-tabulation of the two variables produced the results in Table 18.10, indicating no association. When gender was introduced as the third variable, Table 18.11 was obtained. Among men, 60% of those under 45 indicated a desire to travel abroad compared with 40% of those 45 or older. The pattern was reversed for women, where 35% of those under 45 indicated a desire to travel abroad as opposed to 65% of those 45 or older. Since the association between desire to travel abroad and age runs in the opposite direction for males and females, the relationship between these two variables is masked when the data are aggregated across gender as in Table 18.10. But when the effect of gender is controlled, as in Table 18.11, the suppressed association between preference and age is revealed for the separate categories of males and females.

Table 18.10 Desire to travel abroad by age

<i>Desire to travel abroad</i>	<i>Age</i>	
	<i>Under 45</i>	<i>45 or older</i>
Yes	50%	50%
No	50%	50%
Column totals	100%	100%
Number of respondents	500	500

Table 18.11 Desire to travel abroad by age and gender

<i>Desire to travel abroad</i>	<i>Gender</i>			
	<i>Male age</i>		<i>Female age</i>	
	<i>Under 45</i>	<i>45 or older</i>	<i>Under 45</i>	<i>45 or older</i>
Yes	60%	40%	35%	65%
No	40%	60%	65%	35%
Column totals	100%	100%	100%	100%
Number of respondents	300	300	200	200

No change in initial relationship. In some cases, the introduction of the third variable does not change the initial relationship observed, regardless of whether the original variables were associated. This suggests that the third variable does not influence the relationship between the first two. Consider the cross-tabulation of family size and the tendency to eat in fast-food restaurants frequently, as shown in Table 18.12. The respondents' families were classified into small- and large-size categories based on a median split of the distribution, with 500 respondents in each category. No association is observed. The

Table 18.12 Eating frequently in fast-food restaurants by family size

<i>Eat frequently in fast-food restaurants</i>	<i>Family size</i>	
	<i>Small</i>	<i>Large</i>
Yes	65%	65%
No	35%	35%
Column totals	100%	100%
Number of respondents	500	500

respondents were further classified into high- or low-income groups based on a median split. When income was introduced as a third variable in the analysis, Table 18.13 was obtained. Again, no association was observed.

Table 18.13 Eating frequently in fast-food restaurants by family size and income

Eat frequently in fast-food restaurants	Income			
	Low-income family size		High-income family size	
	Small	Large	Small	Large
Yes	65%	65%	65%	65%
No	35%	35%	35%	35%
Column total	100%	100%	100%	100%
Number of respondents	250	250	250	250

General comments on cross-tabulation

More than three variables can be cross-tabulated; the interpretation is quite complex. Also, because the number of cells increases multiplicatively, maintaining an adequate number of respondents or cases in each cell can be problematic. As a general rule, there should be at least five expected observations in each cell for the computed statistics to be reliable. Thus, cross-tabulation is an inefficient way of examining relationships when there are more than a few variables. Note that cross-tabulation examines association between variables, not causation. To examine causation, the causal research design framework should be adopted (see Chapter 11).

Statistics associated with cross-tabulation



We now discuss the statistics commonly used for assessing the statistical significance and strength of association of cross-tabulated variables. The statistical significance of the observed association is commonly measured by the chi-square statistic. The strength of association, or degree of association, is important from a practical or substantive perspective. Generally, the strength of association is of interest only if the association is statistically significant. The strength of the association can be measured by the phi correlation coefficient, the contingency coefficient, Cramer's *V* and the lambda coefficient. These statistics are described in detail.

Chi-square statistic

The statistic used to test the statistical significance of the observed association in a cross-tabulation. It assists us in determining whether a systematic association exists between the two variables.

Chi-square

The **chi-square statistic** (χ^2) is used to test the statistical significance of the observed association in a cross-tabulation. It assists us in determining whether a systematic association exists between the two variables. The null hypothesis, H_0 , is that there is no association between the variables. The test is conducted by computing the cell frequencies that would be expected if no association were present between the variables, given the existing row and column totals. These expected cell frequencies, denoted f_e , are then compared with the actual observed frequencies, f_o , found in the cross-tabulation to calculate the chi-square statistic. The greater the discrepancies between the expected and observed frequencies, the larger the value of the statistic. Assume that a cross-tabulation has r rows and c columns and a random sample of n observations. Then the expected frequency for each cell can be calculated by using a simple formula:

$$f_e = \frac{n_r n_c}{n}$$

where n_r = total number in the row
 n_c = total number in the column
 n = total sample size

For the data in Table 18.3, the expected frequencies for the cells, going from left to right and from top to bottom, are

$$15 \times 15/30 = 7.50 \quad 15 \times 15/30 = 7.50 \quad 15 \times 15/30 = 7.50 \quad 15 \times 15/30 = 7.50$$

Then the value of χ^2 is calculated as follows:

$$\chi^2 = \sum_{\text{all cells}} \frac{(f_o - f_e)^2}{f_e}$$

For the data in Table 18.2, the value of χ^2 is calculated as

$$\begin{aligned} \chi^2 &= (5 - 7.5)^2/7.5 + (10 - 7.5)^2/7.5 + (10 - 7.5)^2/7.5 + (5 - 7.5)^2/7.5 \\ &= 0.833 + 0.833 + 0.833 + 0.833 \\ &= 3.333 \end{aligned}$$

To determine whether a systematic association exists, the probability of obtaining a value of chi-square as large as or larger than the one calculated from the cross-tabulation is estimated. An important characteristic of the chi-square statistic is the number of degrees of freedom (df) associated with it. In general, the number of degrees of freedom is equal to the number of observations less the number of constraints needed to calculate a statistical term. In the case of a chi-square statistic associated with a cross-tabulation, the number of degrees of freedom is equal to the product of number of rows (r) less one and the number of columns (c) less one. That is, $df = (r - 1) \times (c - 1)$.¹⁰ The null hypothesis (H_0) of no association between the two variables will be rejected only when the calculated value of the test statistic is greater than the critical value of the chi-square distribution with the appropriate degrees of freedom, as shown in Figure 18.8.

Chi-square distribution

A skewed distribution whose shape depends solely on the number of degrees of freedom. As the number of degrees of freedom increases, the chi-square distribution becomes more symmetrical.

The **chi-square distribution** is a skewed distribution whose shape depends solely on the number of degrees of freedom.¹¹ As the number of degrees of freedom increases, the chi-square distribution becomes more symmetrical. Table 3 in the Appendix of statistical tables contains upper tail areas of the chi-square distribution for different degrees of freedom. In this table, the value at the top of each column indicates the area in the upper portion (the right side, as shown in Figure 18.8) of the chi-square distribution. To illustrate, for 1 degree of freedom, the value for an upper tail area of 0.05 is 3.841. This indicates that for 1 degree of freedom the probability of exceeding a chi-square value of 3.841 is 0.05. In other words, at the 0.05 level of significance with 1 degree of freedom, the critical value of the chi-square statistic is 3.841.

For the cross-tabulation given in Table 18.3, there are $(2 - 1) \times (2 - 1) = 1$ degree of freedom. The calculated chi-square statistic had a value of 3.333. Because this is less than the critical value of 3.841, the null hypothesis of no association cannot be rejected, indicating that the association is not statistically significant at the 0.05 level.

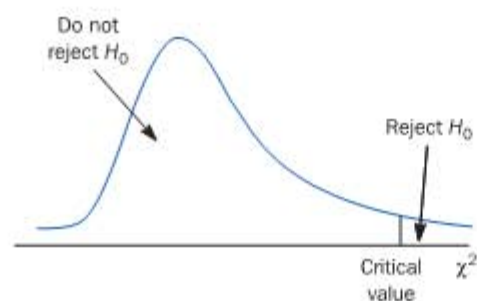


Figure 18.8
 Chi-square test of association

The chi-square statistic can also be used in goodness-of-fit tests to determine whether certain models fit the observed data. These tests are conducted by calculating the significance of sample deviations from assumed theoretical (expected) distributions and can be performed on cross-tabulations as well as on frequencies (one-way tabulations). The calculation of the chi-square statistic and the determination of its significance are the same as illustrated above.

The chi-square statistic should be estimated only on counts of data. When the data are in percentage form, they should first be converted to absolute counts or numbers. In addition, an underlying assumption of the chi-square test is that the observations are drawn independently. As a general rule, chi-square analysis should not be conducted when the expected or theoretical frequency in any of the cells is less than five. If the number of observations in any cell is less than 10, or if the table has two rows and two columns (a 2×2 table), a correction factor should be applied.¹² With the correction factor, the value is 2.133, which is not significant at the 0.05 level. In the case of a 2×2 table, the chi-square is related to the phi coefficient.

Phi coefficient

The **phi coefficient** (ϕ) is used as a measure of the strength of association in the special case of a table with two rows and two columns (a 2×2 table). The phi coefficient is proportional to the square root of the chi-square statistic. For a sample of size n , this statistic is calculated as

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

It takes the value of 0 when there is no association, which would be indicated by a chi-square value of 0 as well. When the variables are perfectly associated, phi assumes the value of 1 and all the observations fall just on the main or minor diagonal. (In some computer programs, phi assumes a value of -1 rather than $+1$ when there is perfect negative association.) In our case, because the association was not significant at the 0.05 level, we would not normally compute the phi value. However, for the purpose of illustration, we show how the values of phi and other measures of the strength of association would be computed. The value of phi is

$$\phi = \sqrt{3.333/30} = 0.333$$

Thus, the association is not very strong. In the more general case involving a table of any size, the strength of association can be assessed by using the contingency coefficient.

Contingency coefficient

Although the phi coefficient is specific to a 2×2 table, the **contingency coefficient** (C) can be used to assess the strength of association in a table of any size. This index is also related to chi-square, as follows:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

The contingency coefficient varies between 0 and 1. The value of 0 occurs in the case of no association (i.e. the variables are statistically independent), but the maximum value of 1 is never achieved. Rather, the maximum value of the contingency coefficient depends on the size of the table (number of rows and number of columns). For this reason, it should be used only to compare tables of the same size. The value of the contingency coefficient for Table 18.3 is

$$C = \sqrt{3.333/(3.333 + 30)} = 0.316$$

Phi coefficient (ϕ)

A measure of the strength of association in the special case of a table with two rows and two columns (a 2×2 table).

Contingency coefficient

A measure of the strength of association in a table of any size.

This value of C indicates that the association is not very strong. Another statistic that can be calculated for any table is Cramer's V .

Cramer's V

Cramer's V

A measure of the strength of association used in tables larger than 2×2 .

Cramer's V is a modified version of the phi correlation coefficient, ϕ , and is used in tables larger than 2×2 . When phi is calculated for a table larger than 2×2 , it has no upper limit. Cramer's V is obtained by adjusting phi for either the number of rows or the number of columns in the table based on which of the two is smaller. The adjustment is such that V will range from 0 to 1. A large value of V merely indicates a high degree of association. It does not indicate how the variables are associated. For a table with r rows and c columns, the relationship between Cramer's V and the phi correlation coefficient is expressed as

$$V = \sqrt{\frac{\phi^2}{\min(r-1), (c-1)}}$$

or

$$V = \sqrt{\frac{\chi^2/n}{\min(r-1), (c-1)}}$$

The value of Cramer's V for Table 18.3 is

$$V = \sqrt{(3.333/30)/1} = 0.333$$

Thus, the association is not very strong. As can be seen, in this case $V = \phi$. This is always the case for a 2×2 table. Another statistic commonly estimated is the lambda coefficient.

Lambda coefficient

The lambda coefficient assumes that the variables are measured on a nominal scale. **Asymmetric lambda** measures the percentage improvement in predicting the value of the dependent variable, given the value of the independent variable. The lambda coefficient also varies between 0 and 1. A value of 0 means no improvement in prediction. A value of 1 indicates that the prediction can be made without error. This happens when each independent variable category is associated with a single category of the dependent variable.

Asymmetric lambda is computed for each of the variables (treating it as the dependent variable). The two asymmetric lambdas are likely to be different, since the marginal distributions are not usually the same. A **symmetric lambda**, a kind of average of the two asymmetric values, is also computed. The symmetric lambda does not make an assumption about which variable is dependent. It measures the overall improvement when prediction is done in both directions.¹³ The value of asymmetric lambda in Table 18.3, with usage as the dependent variable, is 0.333. This indicates that knowledge of gender increases our predictive ability by the proportion of 0.333, i.e. a 33% improvement. The symmetric lambda is also 0.333.

Other statistics

Note that in the calculation of the chi-square statistic the variables are treated as being measured only on a nominal scale. Other statistics such as **tau b**, **tau c** and **gamma** are available to measure association between two ordinal-level variables. All these statistics use information about the ordering of categories of variables by considering every possible pair of cases in the table. Each pair is examined to determine whether its relative ordering on the first variable is the same as its relative ordering on the second variable (concordant), the ordering is reversed (discordant), or the pair is tied. The manner in

Asymmetric lambda

A measure of the percentage improvement in predicting the value of the dependent variable given the value of the independent variable in contingency table analysis. Lambda also varies between 0 and 1.

Symmetric lambda

The symmetric lambda does not make an assumption about which variable is dependent. It measures the overall improvement when prediction is done in both directions.

tau b

A test statistic that measures the association between two ordinal-level variables. It makes an adjustment for ties and is the most appropriate when the table of variables is square.

tau c

A test statistic that measures the association between two ordinal-level variables. It makes an adjustment for ties and is most appropriate when the table of variables is not square but a rectangle.

Gamma

A test statistic that measures the association between two ordinal-level variables. It does not make an adjustment for ties.

which the ties are treated is the basic difference between these statistics. Both tau b and tau c adjust for ties. Tau b is the most appropriate with square tables in which the number of rows and the number of columns are equal. Its value varies between +1 and -1. For a rectangular table in which the number of rows is different from the number of columns, tau c should be used. Gamma does not make an adjustment for either ties or table size. Gamma also varies between +1 and -1 and generally has a higher numerical value than tau b or tau c . For the data in Table 18.3, as gender is a nominal variable, it is not appropriate to calculate ordinal statistics. All these statistics can be estimated by using the appropriate computer software that incorporates cross-tabulation. Other statistics for measuring the strength of association, namely product moment correlation and non-metric correlation, are discussed in Chapter 20.

Cross-tabulation in practice

While conducting cross-tabulation analysis in practice, it is useful to proceed through the following steps:

- 1 Test the null hypothesis that there is no association between the variables using the chi-square statistic. If you fail to reject the null hypothesis, then there is no relationship.
- 2 If H_0 is rejected, then determine the strength of the association using an appropriate statistic (phi coefficient, contingency coefficient, Cramer's V , lambda coefficient, or other statistics).
- 3 If H_0 is rejected, interpret the pattern of the relationship by computing the percentages in the direction of the independent variable, across the dependent variable.
- 4 If the variables are treated as ordinal rather than nominal, use tau b , tau c or gamma as the test statistic. If H_0 is rejected, then determine the strength of the association using the magnitude, and the direction of the relationship using the sign of the test statistic.

Hypothesis testing related to differences



The previous section considered hypothesis testing related to associations. We now focus on hypothesis testing related to differences. A classification of hypothesis testing procedures for examining differences is presented in Figure 18.9. Note that this figure is consistent with the classification of univariate techniques presented in Figure 17.4. The major difference is that Figure 17.4 also accommodates more than two samples and thus deals with techniques such as one-way ANOVA and K-W ANOVA (Chapter 17), whereas Figure 18.9 is limited to no more than two samples. Also, one-sample techniques such as frequencies, which do not involve statistical testing, are not covered in Figure 18.9.

Hypothesis testing procedures can be broadly classified as parametric or non-parametric, based on the measurement scale of the variables involved. **Parametric tests** assume that the variables of interest are measured on at least an interval scale. **Non-parametric tests** assume that the variables are measured on a nominal or ordinal scale. These tests can be further classified based on whether one or two or more samples are involved. As explained in Chapter 17, the number of samples is determined based on how the data are treated for the purpose of analysis, not based on how the data were collected. 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. On the other hand, the samples are paired when the data for the two samples relate to the same group of respondents.

The most popular parametric test is the t test conducted for examining hypotheses about means. The t test could be conducted on the mean of one sample or two samples of observations. In the case of two samples, the samples could be independent or paired.

Parametric tests

Hypothesis testing procedures that assume that the variables of interest are measured on at least an interval scale.

Non-parametric tests

Hypothesis testing procedures that assume that the variables are measured on a nominal or ordinal scale.

Non-parametric tests based on observations drawn from one sample include the chi-square test, the Kolmogorov–Smirnov test, the runs test and the binomial test. In the case of two independent samples, the chi-square test, the Mann–Whitney U test, the median test and the Kolmogorov–Smirnov two-sample test are used for examining hypotheses about location. These tests are non-parametric counterparts of the two-group t test. For paired samples, non-parametric tests include the Wilcoxon matched-pairs signed-ranks test and the sign test. These tests are the counterparts of the paired t test. Parametric as well as non-parametric tests are also available for evaluating hypotheses relating to more than two samples. These tests are considered in later chapters.

Parametric tests

t test

A univariate hypothesis test using the t distribution, which is used when the standard deviation is unknown and the sample size is small.

t statistic

A statistic that assumes that the variable has a symmetric bell-shaped distribution, that the mean is known (or assumed to be known), and that the population variance is estimated from the sample.

t distribution

A symmetrical bell-shaped distribution that is useful for sample testing ($n < 30$). It is similar to the normal distribution in appearance.

Parametric tests provide inferences for making statements about the means of parent populations. A t test is commonly used for this purpose. This test is based on Student's t statistic. 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 μ and unknown population variance σ^2 , which is estimated by the sample variance s^2 . Recall that the standard deviation of the sample mean, \bar{X} , is estimated as $s_{\bar{X}} = s/\sqrt{n}$. 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. Compared with the normal distribution, however, the t distribution has more area in the tails and less in the centre. This is because the population variance σ^2 is unknown and is estimated by the sample variance s^2 . Given the uncertainty in the value of s^2 , the observed values of t are more variable than those of z . Thus, we must go out a larger number of standard deviations from zero to encompass a certain percentage of values from the t distribution than is the case with the normal distribution. Yet, as the number of degrees of freedom increases, the t distribution approaches

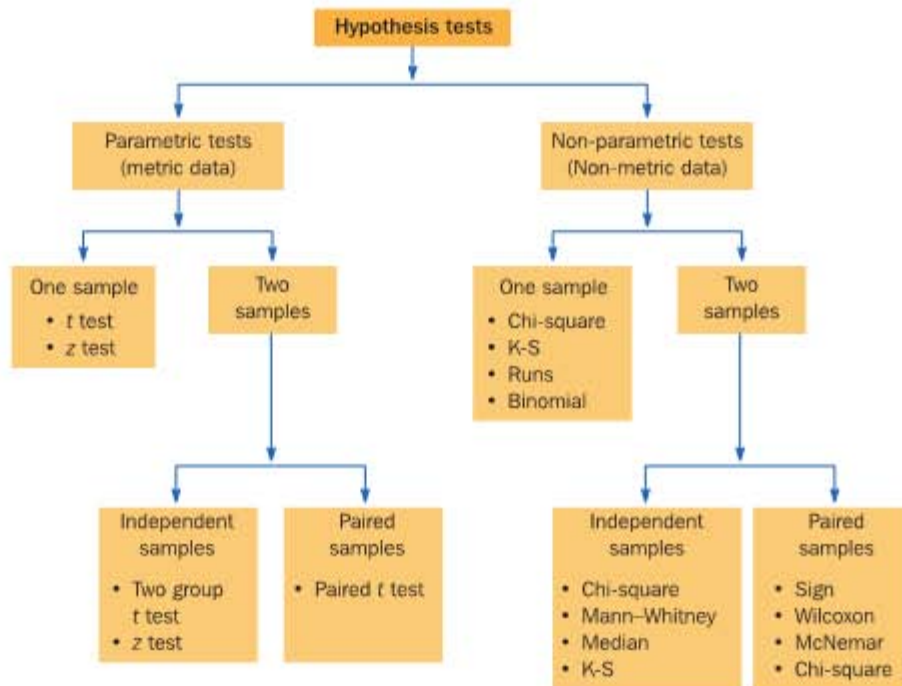


Figure 18.9 Hypothesis testing procedures

the normal distribution. In fact, for large samples of 120 or more, the t distribution and the normal distribution are virtually indistinguishable. Table 4 in the Appendix of statistical tables shows selected percentiles of the t distribution. Although normality is assumed, the t test is quite robust to departures from normality.

The procedure for hypothesis testing, for the special case when the t statistic is used, is as follows:

- 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, α , for testing H_0 . Typically, the 0.05 level is selected.¹⁴
- 4 Take one or two samples and compute the mean and standard deviation for each sample.
- 5 Calculate the t statistic assuming that 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 in the Appendix. (Alternatively, calculate the critical value of the t statistic.)
- 7 If the probability computed in step 6 is smaller than the significance level selected in step 3, reject H_0 . If the probability is larger, do not reject H_0 . (Alternatively, if the value of the calculated t statistic in step 5 is larger than the critical value determined in step 6, 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 from that assumed by H_0 .¹⁵
- 8 Express the conclusion reached by the t test in terms of the marketing research problem.

We illustrate the general procedure for conducting t tests in the following sections, beginning with the one-sample case.

One sample

In marketing research, the researcher is often interested in making statements about a single variable against a known or given standard. Examples of such statements are that the market share for a new product will exceed 15%, at least 65% of customers will like a new package design, and 80% of retailers will prefer a new pricing policy. These statements can be translated to null hypotheses that can be tested using a one-sample test, such as the t test or the z test. In the case of a t test for a single mean, the researcher is interested in testing whether the population mean conforms to a given hypothesis (H_0). For the data in Table 18.2, suppose we wanted to test the hypothesis that the mean familiarity rating exceeds 4.0, the neutral value on a seven-point scale. A significance level of $\alpha = 0.05$ is selected. The hypothesis may be formulated as

$$H_0: \mu \leq 4.0$$

$$H_1: \mu > 4.0$$

$$t = (\bar{X} - \mu) / s_{\bar{x}}$$

$$s_{\bar{x}} = s / \sqrt{n}$$

$$s_{\bar{x}} = 1.579 / \sqrt{29} = 1.579 / 5.385 = 0.293$$

$$t = (4.724 - 4.0) / 0.293 = 0.724 / 0.293 = 2.471$$

The degrees of freedom for the t statistic to test the hypothesis about one mean are $n - 1$. In this case, $n - 1 = 29 - 1$, or 28. From Table 4 in the Appendix, the probability of getting a more extreme value than 2.471 is less than 0.05. (Alternatively, the critical t value for 28 degrees of freedom and a significance level of 0.05 is 1.7011, which is less than the calculated value.) Hence, the null hypothesis is rejected. The familiarity level does not exceed 4.0.

z test

A univariate hypothesis test using the standard normal distribution.

Note that if the population standard deviation was assumed to be known as 1.5, rather than estimated from the sample, a **z test** would be appropriate. In this case, the value of the *z* statistic would be

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}}$$

where

$$\sigma_{\bar{x}} = 1.5/\sqrt{29} = 1.5/5.385 = 0.279$$

and

$$z = (4.724 - 4.0)/0.279 = 0.724/0.279 = 2.595$$

From Table 2 in the Appendix of statistical tables, the probability of getting a more extreme value of *z* than 2.595 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 earlier in this chapter when we introduced hypothesis testing.

Two independent samples

Several hypotheses in marketing relate to parameters from two different populations; for example, the users and non-users of a brand differ in terms of their perceptions of the brand, high-income consumers spend more on leisure activities than low-income consumers, or the proportion of brand-loyal users in segment I is more than the proportion in segment II. Samples drawn randomly from different populations are termed **independent samples**. As in the case for one sample, the hypotheses could relate to means or proportions.

Independent samples

The samples are independent if they are drawn randomly from different populations.

Means. In the case of means for two independent samples, the hypotheses take the following form:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

The two populations are sampled and the means and variances are computed based on samples of sizes n_1 and n_2 . If both populations are found to have the same variance, a pooled variance estimate is computed from the two sample variances as follows:

$$s^2 = \frac{\sum_{i=1}^{n_1} (X_{i1} - \bar{X}_1)^2 + \sum_{i=1}^{n_2} (X_{i2} - \bar{X}_2)^2}{n_1 + n_2 - 2} \quad s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

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 $(n_1 + n_2 - 2)$.

If the two populations have unequal variances, an exact t cannot be computed for the difference in sample means. Instead, an approximation to t is computed. The number of degrees of freedom in this case is usually not an integer, but a reasonably accurate probability can be obtained by rounding to the nearest integer.¹⁶

F test

A statistical test of the equality of the variances of two populations.

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

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

The **F statistic** is computed from the sample variances as follows:

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

where n_1 = size of sample 1

n_2 = size of sample 2

$n_1 - 1$ = degrees of freedom for sample 1

$n_2 - 1$ = degrees of freedom for sample 2

s_1^2 = sample variance for sample 1

s_2^2 = sample variance for sample 2.

F distribution

A frequency distribution that depends upon two sets of degrees of freedom: the degrees of freedom in the numerator and the degrees of freedom in the denominator.

As can be seen, the critical value of the **F distribution** depends on two sets of degrees of freedom: those in the numerator and those in the denominator. The critical values of F for various degrees of freedom for the numerator and denominator are given in Table 5 of the Appendix of statistical tables. If the probability of F is greater than the significance level α , H_0 is not rejected and t based on the pooled variance estimate can be used. On the other hand, if the probability of F is less than or equal to α , H_0 is rejected and t based on a separate variance estimate is used.

Using the data in Table 18.1, suppose we wanted to determine whether Internet usage was different for males as compared with females. A two-independent-samples t test can be conducted. The results of this test are presented in Table 18.14. Note that the F test of sample variances has a probability that is less than 0.05. Accordingly H_0 is rejected and the t test based on the 'equal variances not assumed' should be used. The t value is -4.492

Table 18.14 Two independent samples t test

Summary statistics					
	Number of cases		Mean	Standard error mean	
Male	15		9.333	1.137	
Female	15		3.867	0.435	
F test for equality of variances					
F value			Two-tail probability		
15.507			0.000		
t test					
Equal variances assumed			Equal variances not assumed		
t value	Degrees of freedom	Two-tail probability	t value	Degrees of freedom	Two-tail probability
-4.492	28	0.000	-4.492	18.014	0.000

and with 18.014 degrees of freedom, this gives a probability of 0.000, which is less than the significance level of 0.05. Therefore, the null hypothesis of equal means is rejected. Because the mean usage for males (gender = 1) is 9.333 and that for females (gender = 2) is 3.867, males use the Internet to a significantly greater extent than females. We also show the t test assuming equal variances because most software packages automatically conduct the t test both ways.

For another application of the t test, consider the following example.

Example Shops seek to suit elderly to a 't'¹⁷

A study based on a sample of 789 respondents who were 65 or older attempted to determine the effect of lack of mobility on shop patronage. A major research question related to the differences in the physical requirements of dependent and self-reliant elderly persons. That is, did the two groups require different things to get to the shop or after they arrived at the shop? A more detailed analysis of the physical requirements conducted by the t tests of two independent samples (shown in the table) indicated that dependent elderly persons are more likely to look for shops that offer home delivery and phone orders and for shops to which they have accessible transportation. They are also more likely to look for a variety of shops located close together.

Differences in physical requirements between dependent and self-reliant elderly

Physical requirement items	Mean*		
	Self-reliant	Dependent	t test probability
Delivery to home	1.787	2.000	0.023
Phone in order	2.030	2.335	0.003
Transportation to store	2.188	3.098	0.000
Convenient parking	4.001	4.095	0.305
Location close to home	3.177	3.325	0.137
Variety of shops close together	3.456	3.681	0.023

* Measured on a five-point scale from not important (1) to very important (5).

In this example, we tested the difference between means. A similar test is available for testing the difference between proportions for two independent samples.

Proportions. A case involving proportions for two independent samples is also illustrated using the data from Table 18.1, which gives the number of males and females who use the Internet for shopping. Is the proportion of respondents using the Internet for shopping the same for males and females? 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. In this case, however, the test statistic is given by

$$z = \frac{P_1 - P_2}{\sqrt{\hat{p}_1 - \hat{p}_2}}$$

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

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

A significance level of $\alpha = 0.05$ is selected. Given the data in Table 18.1, the test statistic can be calculated as

$$\begin{aligned} P_1 - P_2 &= (11/15) - (6/15) \\ &= 0.733 - 0.400 = 0.333 \\ P &= (15 \times 0.733 + 15 \times 0.4) / (15 + 15) = 0.567 \\ s_{p_1 - p_2} &= \sqrt{0.567 \times 0.433 (1/15 + 1/15)} = 0.181 \\ z &= 0.333 / 0.181 = 1.84 \end{aligned}$$

Given a two-tail test, the area to the right of the critical value is $\alpha/2$, or 0.025. Hence, the critical value of the test statistic is 1.96. Because the calculated value is less than the critical value, the null hypothesis cannot be rejected. Thus, the proportion of users (0.733) for males and (0.400) for females is not significantly different for the two samples. Note that although the difference is substantial, it is not statistically significant due to the small sample sizes (15 in each group).

Paired samples

In many marketing research applications, the observations for the two groups are not selected from independent samples. Rather, the observations relate to **paired samples** in that the two sets of observations relate to the same respondents. A sample of respondents may rate competing brands, may indicate the relative importance of two attributes of a product, or may evaluate a brand at two different times. The differences in these cases are 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 formulae are

$$H_0: \mu_D = 0$$

$$H_1: \mu_D \neq 0$$

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

where

$$\bar{D} = \frac{\sum_{i=1}^n D_i}{n}$$

$$S_{\bar{D}} = \sqrt{\frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1}}$$

$$S_{\bar{D}} = S_D / \sqrt{n}$$

Paired samples

The samples are paired when the data for the two samples relate to the same group of respondents.

Paired samples *t* test

A test for differences in the means of paired samples.

In the Internet usage example (Table 18.1), a paired t test could be used to determine if the respondents differed in their attitude towards the Internet and attitude towards technology. The resulting output is shown in Table 18.15. The mean attitude towards the Internet is 5.167 and that towards technology is 4.10. The mean difference between the variables is 1.067, with a standard deviation of 0.828 and a standard error of 0.1511. This results in a t value of $(1.067/0.1511) = 7.06$, with $30 - 1 = 29$ degrees of freedom and a probability of less than 0.001. Therefore the respondents have a more favourable attitude towards the Internet as compared with technology in general. Another application is provided in the context of determining the relative effectiveness of 15-second versus 30-second TV commercials.

Table 18.15 Paired samples t test

Variable		Number of cases		Mean		Standard deviation		Standard error	
Internet attitude		30		5.167		1.234		0.225	
Technology attitude		30		4.100		1.398		0.255	
Difference mean	Standard deviation	Standard error	Correlation	Two-tail probability	t value	Degrees of freedom	Two-tail probability		
1.067	0.828	0.1511	0.809	0.000	7.059	29	0.000		

Example

Seconds count¹⁸

A survey of 83 media directors of the largest Canadian advertising agencies was conducted to determine the relative effectiveness of 15-second versus 30-second commercial advertisements. By use of a five-point rating scale (1 being excellent and 5 being poor), 15- and 30-second commercials were rated by each respondent for brand awareness, main idea recall, persuasion, and ability to tell an emotional story. The table indicates that 30-second commercials were rated more favourably on all the dimensions. Paired t tests indicated that these differences were significant, and the 15-second commercials were evaluated as less effective.

Mean rating 15- and 30-second commercials on four communication variables

Brand awareness		Main idea recall		Persuasion		Ability to tell an emotional story	
15	30	15	30	15	30	15	30
2.5	1.9	2.7	2.0	3.7	2.1	4.3	1.9

The difference in proportions for paired samples can be tested by using the McNemar test or the chi-square test, as explained in the following section on non-parametric tests.

Non-parametric tests



Non-parametric tests are used when the variables are non-metric. Like parametric tests, non-parametric tests are available for testing variables from one sample two independent samples or two related samples.

One sample

Sometimes the researcher wants to test whether the observations for a particular variable could reasonably have come from a particular distribution, such as the normal, uniform or Poisson distribution. Knowledge of the distribution is necessary for finding probabilities corresponding to known values of the variable, or variable values corresponding to known probabilities (see the Appendix at the end of Chapter 15). The **Kolmogorov–Smirnov (K–S) one-sample test** is one such goodness-of-fit test. The K–S test compares the cumulative distribution function for a variable with a specified distribution. A_i denotes the cumulative relative frequency for each category of the theoretical (assumed) distribution, and O_i denotes the comparable value of the sample frequency. The K–S test is based on the maximum value of the absolute difference between A_i and O_i . The test statistic is

$$K = \max |A_i - O_i|$$

The decision to reject the null hypothesis is based on the value of K . The larger K is, the more confidence we have that H_0 is false. Note that this is a one-tailed test, since the value of K is always positive, and we reject H_0 for large values of K . For $\alpha = 0.05$, the critical value of K for large samples (over 35) is given by $1.36/\sqrt{n}$.¹⁹ Alternatively, K can be transformed into a normally distributed z statistic and its associated probability determined.

Table 18.16 K–S one-sample test for normality for Internet usage

Test distribution, normal				
Mean	6.600			
Standard deviation	4.296			
Cases	30			
Most extreme differences				
Absolute	Positive	Negative	K–S z	Two-tailed p
0.222	0.222	–0.142	1.217	0.103

In the context of the Internet usage example, suppose we wanted to test whether the distribution of Internet usage was normal. A K–S one-sample test is conducted, yielding the data shown in Table 18.16. The largest absolute difference between the observed and normal distribution was $K = 0.222$. Although our sample size is only 30 (less than 35), we can use the approximate formula and the critical value of K is $1.36/\sqrt{30} = 0.248$. Because the calculated value of K is smaller than the critical value, the null hypothesis cannot be rejected. Alternatively, Table 18.16 indicates that the probability of observing a K value of 0.222, as determined by the normalized z statistics, is 0.103. Because this is more than the significance level of 0.05, the null hypothesis cannot be rejected, leading to the same conclusion. Hence the distribution of Internet usage does not deviate significantly from the normal distribution.

As mentioned earlier, the chi-square test can also be performed on a single variable from one sample. In this context, the chi-square serves as a goodness-of-fit test. It tests whether a significant difference exists between the observed number of cases in each category and the expected number.

Other one-sample non-parametric tests include the **runs test** and the **binomial test**. The runs test is a test of randomness for the dichotomous variables. This test is conducted by determining whether the order or sequence in which observations are obtained is

Kolmogorov–Smirnov (K–S) one-sample test

A one-sample non-parametric goodness-of-fit test that compares the cumulative distribution function for a variable with a specified distribution.

Runs test

A test of randomness for a dichotomous variable.

Binominal test

A goodness-of-fit statistical test for dichotomous variables. It tests the goodness of fit of the observed number of observations in each category to the number expected under a specified binominal distribution.

random. The binomial test is also a goodness-of-fit test for dichotomous variables. It tests the goodness of fit of the observed number of observations in each category to the number expected under a specified binomial distribution. For more information on these tests, refer to the standard statistical literature.²⁰

Two independent samples

When the difference in the location of two populations is to be compared based on observations from two independent samples and the variable is measured on an ordinal scale, the **Mann–Whitney *U* test** can be used.²¹ This test corresponds to the two-independent-samples *t* test, for interval scale variables, when the variances of the two populations are assumed equal.

Mann–Whitney *U* test

A statistical test for a variable measured on an ordinal scale, comparing the differences in the location of two populations based on observations from two independent samples.

In the Mann–Whitney *U* test, the two samples are combined and the cases are ranked in order of increasing size. The test statistic, *U*, is computed as the number of times a score from sample or group 1 precedes a score from group 2. If the samples are from the same population, the distribution of scores from the two groups in the rank list should be random. An extreme value of *U* would indicate a non-random pattern pointing to the inequality of the two groups. For samples of less than 30, the exact significance level for *U* is computed. For larger samples, *U* is transformed into a normally distributed *z* statistic. This *z* can be corrected for ties within ranks.

We examine again the difference in the Internet usage of males and females. This time, though, the Mann–Whitney test is used. The results are given in Table 18.17. Again, a significant difference is found between the two groups, corroborating the results of the two-independent-samples *t* test reported earlier. Because the ranks are assigned from the smallest observation to the largest, the higher mean rank of males (20.93) indicates that they use the Internet to a greater extent than females (mean rank = 10.07).

Table 18.17 Mann–Whitney *U*–Wilcoxon rank sum *W* test: Internet usage by gender

Gender	Mean rank	Cases	
Male	20.93	15	
Female	10.07	15	
Total		30	
<i>U</i>	<i>W</i>	<i>z</i>	Corrected for ties, two-tailed <i>p</i>
31.000	151.000	–3.406	0.001

Note: *U* = Mann–Whitney test statistics, *W* = Wilcoxon *W* statistic, *z* = *U* transformed into normally distributed *z* statistics.

Two-sample median test

Non-parametric test statistic that determines whether two groups are drawn from populations with the same median. This is not as powerful as the Mann–Whitney *U* test.

Kolmogorov–Smirnov (K–S) two-sample test

Non-parametric test statistic that determines whether two distributions are the same. It takes into account any differences in the two distributions, including median, dispersion and skewness.

Researchers often wish to test for a significant difference in proportions obtained from two independent samples. In this case, as an alternative to the parametric *z* test considered earlier, one could also use the cross-tabulation procedure to conduct a chi-square test.²² In this case, we will have a 2×2 table. One variable will be used to denote the sample and will assume a value of 1 for sample 1 and a value of 2 for sample 2. The other variable will be the binary variable of interest.

Two other independent-samples non-parametric tests are the median test and K–S test. The **two-sample median test** determines whether the two groups are drawn from populations with the same median. It is not as powerful as the Mann–Whitney *U* test because it merely uses the location of each observation relative to the median, and not the rank, of each observation. The **K–S two-sample test** examines whether the two distributions are the same. It takes into account any differences between the two distributions, including the median, dispersion and skewness, as illustrated by the following example.

Example**Directors change direction²³**

How do marketing research directors and users in Fortune 500 manufacturing firms perceive the role of marketing research in initiating changes in marketing strategy formulation? It was found that the marketing research directors were more strongly in favour of initiating changes in strategy and less in favour of holding back than were users of marketing research. Using the K-S test, these differences of role definition were statistically significant at the 0.05 level, as shown below

The role of marketing research in strategy formulation

Sample	n	Responses (%)				
		Absolutely must	Preferably should	May or may not	Preferably should not	Absolutely must not
D	77	7	26	43	19	5
U	68	2	15	32	35	16

K-S significance = 0.05.
D = directors, U = users.

In this example, the marketing research directors and users comprised two independent samples. The samples, however, are not always independent. In the case of paired samples, a different set of tests should be used.

Paired samples

An important non-parametric test for examining differences in the location of two populations based on paired observations is the **Wilcoxon matched-pairs signed-ranks test**. This test analyses the differences between the paired observations, taking into account the magnitude of the differences. It computes the differences between the pairs of variables and ranks the absolute differences. The next step is to sum the positive and negative ranks. The test statistic, z , is computed from the positive and negative rank sums. Under the null hypothesis of no difference, z is a standard normal variate with mean 0 and variance 1 for large samples. This test corresponds to the paired t test considered earlier.²⁴

The example considered for the paired t test, whether the respondents differed in terms of attitude towards the Internet and attitude towards technology, is considered again. Suppose we assume that both these variables are measured on ordinal rather than interval scales. Accordingly, we use the Wilcoxon test. The results are shown in Table 18.18. Again, a significant difference is found in the variables, and the results are in accordance with the conclusion reached by the paired t test. There are 23 negative differences (attitude towards technology is less favourable than attitude towards the Internet). The mean rank of these negative differences is 12.72. On the other hand, there is only one positive difference (attitude towards technology is more favourable than attitude towards the Internet). The mean rank of the difference is 7.50. There are six ties, or observations with the same value for both variables. These numbers indicate that the attitude towards the Internet is more favourable than that towards technology. Furthermore, the probability associated with the z statistic is less than 0.05, indicating that the difference is indeed significant.

Wilcoxon matched-pairs signed-ranks test

A non-parametric test that analyses the differences between the paired observations, taking into account the magnitude of the differences.

Table 18.18 Wilcoxon matched-pairs signed-ranks test

Internet with technology		
(Technology–Internet)	Cases	Mean rank
– Ranks	23	12.72
+ Ranks	1	7.50
Ties	6	
Total	30	
$z = -4.207$		Two-tailed $p = 0.000$

Sign test

A non-parametric test for examining differences in the location of two populations, based on paired populations, that compares only the signs of the differences between pairs of variables without taking into account the magnitude of the differences.

Another paired sample non-parametric test is the **sign test**.²⁵ This test is not as powerful as the Wilcoxon matched-pairs signed-ranks test because it only compares the signs of the differences between pairs of variables without taking into account the magnitude of the differences. In the special case of a binary variable where the researcher wishes to test differences in proportions, the McNemar test can be used. Alternatively, the chi-square test can also be used for binary variables.

The various parametric and non-parametric tests are summarised in Table 18.19. The tests in Table 18.19 can be easily related to those in Figure 18.9. Table 18.19 classifies the tests in more detail, as parametric tests (based on metric data) are classified separately for

Table 18.19 A summary of hypothesis testing

Sample	Application	Level of scaling	Test/comments
One sample			
One sample	Distributions	Non-metric	K–S and chi-square for goodness of fit Runs test for randomness Binomial test for goodness of fit for dichotomous variables
One sample	Means	Metric	t test, if variance is unknown z test, if variance is known
One sample	Proportions	Metric	z test
Two independent samples			
Two independent samples	Distributions	Non-metric	K–S two-sample test for examining equivalence of two distributions
Two independent samples	Means	Metric	Two-group t test F test for equality of variances
Two independent samples	Proportions	Metric Non-metric	z test Chi-square test
Two independent samples	Rankings/medians	Non-metric	Mann–Whitney U test more powerful than median test
Paired samples			
Paired samples	Means	Metric	Paired t test
Paired samples	Proportions	Non-metric	McNemar test for binary variables Chi-square test
Paired samples	Rankings/medians	Non-metric	Wilcoxon matched-pairs ranked-signs test more powerful than sign test

means and proportions. Likewise, non-parametric tests (based on non-metric data) are classified separately for distributions and rankings/medians. The final example illustrates the use of hypothesis testing in international branding strategy.

Example International brand equity – the name of the game²⁶

In an era of global marketing, how can marketers develop their brands overseas in diverse social, cultural and historical environments? A study showed that, in general, a firm's international brand structure includes firm-based characteristics, product market characteristics, and market dynamics. More specifically, according to Bob Kroll, the former President of Del Monte International, uniform packaging may be an asset to marketing internationally, but catering to the culinary taste preferences of individual countries is more important. One survey on international product marketing makes this clear. Respondents included 100 brand and product managers and marketing executives from some of the world's largest food, pharmaceutical and personal product companies. For overseas markets, 39% said it was not a good idea to use uniform packaging and 38% said it was a good idea. Those who said it was not a good idea mentioned, however, the desirability of maintaining as much brand equity and package consistency from market to market. But they also believed it was necessary to tailor the package to fit the linguistic and regulatory needs of different markets.

Based on this finding, a suitable research question can be: 'Do consumers in different countries prefer to buy global name brands with different packaging customised to suit their local needs?' Based on this research question, one can frame a hypothesis that, other things being constant, standardised branding with customised packaging for a well-established brand name will result in a greater market share. The hypotheses may be formulated as follows:

H_0 : Standardised branding with customised packaging for a well-established brand name will not lead to greater market share in an international market.

H_1 : Other factors remaining equal, standardised branding with customised packaging for a well-established brand name will lead to greater market share in the international market.

To test the null hypothesis, a well-established brand such as Colgate toothpaste, which has followed a mixed strategy, can be selected. The market share in countries with standardised branding and standardised packaging can be compared with market share in countries with standardised branding and customised packaging, after controlling for the effect of other factors. A two-independent-samples t test can be used.



Internet and computer applications

SPSS (www.spss.com)

The main program in SPSS is FREQUENCIES. It produces a table of frequency counts, percentages and cumulative percentages for the values of each variable. It gives all of the associated statistics except for the coefficient of variation. If the data are interval scaled and only the summary statistics are desired, the DESCRIPTIVES procedure can be used. All of the statistics computed by DESCRIPTIVES are available in FREQUENCIES. However, DESCRIPTIVES is more efficient because it does not sort values into a frequency table. An additional program, MEANS, computes means and standard deviations for a dependent variable over subgroups of cases defined by independent variables.

CROSSTABS displays cross-classification tables and provides cell counts, row and column percentages, the chi-square test for significance, and all the measures of the strength of the association that have been discussed. The major program for conducting t tests is T-TEST. This program can be used to conduct t tests on independent as well as paired samples. All the non-parametric tests that we have discussed can be conducted by using the NPAR TESTS program.



SAS (www.sas.com)

The main program is UNIVARIATE. In addition to providing a frequency table, this program provides all the associated statistics. Another procedure available is FREQ. For one-way frequency distribution, FREQ does not provide any associated statistics. If only summary statistics are desired, procedures such as MEANS, SUMMARY and TABULATE can be used.

FREQ displays cross-classification tables and provides cell counts, row and column percentages, the chi-square test for significance, and all the measures of the strength of the association that have been discussed. TABULATE can be used for obtaining cell counts and row and column percentages, although it does not provide any of the associated statistics. The program T-TEST can be used. The non-parametric tests may be conducted by using NPAR1WAY. This program will conduct the two-independent-samples test (Mann–Whitney, median and K–S) as well as the Wilcoxon test for paired samples.

Minitab (www.minitab.com)

The main function is stats>descriptive statistics. The output values include the mean, median, mode, standard deviation, minimum, maximum and quartiles. A histogram in a bar chart or graph can be produced from the Graph>Histogram selection. Cross-tabulations (crosstabs) and chi-square are under the Stats>Tables function. Parametric tests available in Minitab in descriptive stat function are *z* test for means, *t* test for mean, and two-sample *t* test. The non-parametric tests can be accessed under the Stat>Time series function. The output includes the one-sample sign, one-sample Wilcoxon, Mann–Whitney, Kruskal–Wallis, median test, Friedman, runs test and pairwise differences.

Excel (www.microsoft.com)

In Excel the output produces the mean, standard error, median, mode, standard deviation, variance, kurtosis, skewness, range, minimum, maximum, sum, count and confidence level. Frequencies can be selected under the histogram function. A histogram can be produced in bar format. The Data>Pivot Table performs crosstabs in Excel. To do additional analysis or customise data, select a different summary function such as max, min, average or standard deviation. In addition, a custom calculation can be selected to calculate based on other cells in the data plane. The chi-square test can be accessed under the Insert>Function>Statistical>Chitest function. The available parametric tests include the *t* test: paired samples 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 for variances of two samples.

SNAP (www.snapsurveys.com)

In SNAP, Descriptive Statistics produces the count, mean, mode, quartiles, median, sum, minimum, maximum, range, standard error, standard deviation, variance, skewness and kurtosis. In either Tables or Charts, the following can be produced displaying absolute values, percentages or means:

- A single question or several questions together.
- Comparing one question against another, or several questions against one or more other questions (cross-tabulation).
- Grids of a body of scaled items, e.g. a body of Likert scales.
- Holecount tables, i.e. a summary of all or any chosen responses to the questionnaire.
- Summary results in counts or percentages in the format of the original questionnaire. Chi-square and *t* tests can be calculated using the Charts option.

Summary



Basic data analysis provides valuable insights and guides the rest of the data analysis as well as the interpretation of the results. A frequency distribution should be obtained for each variable in the data. This analysis produces a table of frequency counts, percentages and cumulative percentages for all the values associated with that variable. It indicates the extent of out-of-range, missing or extreme values. The mean, mode and median of a frequency distribution are measures of central tendency. The variability of the distribution is described by the range, the variance or standard deviation, coefficient of variation and interquartile range. Skewness and kurtosis provide an idea of the shape of the distribution.

Cross-tabulations are tables that reflect the joint distribution of two or more variables. In cross-tabulation, the percentages can be computed either by column, based on column totals, or by row, based on row totals. The general rule is to compute the percentages in the direction of the independent variable, across the dependent variable. Often the introduction of a third variable can provide additional insights. The chi-square statistic provides a test of the statistical significance of the observed association in a cross-tabulation. The phi coefficient, contingency coefficient, Cramer's V and lambda coefficient provide measures of the strength of association between the variables.

Parametric and non-parametric tests are available for testing hypotheses related to differences. In the parametric case, the t test is used to examine hypotheses related to the population mean. Different forms of the t test are suitable for testing hypotheses based on one sample, two independent samples or paired samples. In the non-parametric case, popular one-sample tests include the chi-square, Kolmogorov–Smirnov and binomial tests. For two independent non-parametric samples, the chi-square, Mann–Whitney U , median and Kolmogorov–Smirnov tests can be used. For paired samples, the sign, Wilcoxon matched-pairs signed-ranks, McNemar and chi-square tests are useful for examining hypotheses related to measures of location.

Questions



- Describe the procedure for computing frequencies.
- What measures of location are commonly computed?
- What measures of variability are commonly computed?
- How is the relative flatness or peakedness of a distribution measured?
- What is a skewed distribution? What does it mean?
- What is the major difference between cross-tabulation and frequency distribution?
- What is the general rule for computing percentages in cross-tabulation?
- Define a spurious correlation.
- What is meant by a suppressed association? How is it revealed?
- Discuss the reasons for the frequent use of cross-tabulations. What are some of the limitations?
- Present a classification of hypothesis testing procedures.
- Describe the general procedure for conducting a t test.
- What is the major difference between parametric and non-parametric tests?
- Which non-parametric tests are the counterparts of the two-independent-samples t test for parametric data?
- Which non-parametric tests are the counterparts of the paired samples t test for parametric data?

Exercises



- 1 In each of the following situations, indicate the statistical analysis you would conduct and the appropriate test or test statistic that should be used.
 - a Consumer preferences for Body Shop shampoo were obtained on an 11-point Likert scale. The same consumers were then shown a commercial about the Body Shop. After the commercial, preferences for the Body Shop were again measured. Has the commercial been successful in inducing a change in preferences?
 - b Does the preference for Body Shop shampoo follow a normal distribution?
 - c Respondents in a survey of 1,000 households were classified as heavy, medium, light and non-users of ice cream. They were also classified as being in high-, medium- or low-income categories. Is the consumption of ice cream related to income level?
 - d In a survey of 2,000 households, respondents were asked to rank 10 supermarkets, including Lidl, in order of preference. The sample was divided into small and large households based on a median split of the household size. Does preference for shopping in Lidl vary by household size?

- 2 The current advertising campaign for Red Bull would be changed if less than 30% of consumers like it.
 - a Formulate the null and alternative hypotheses.
 - b Discuss the Type I and Type II errors that could occur in hypothesis testing.
 - c Which statistical test would you use? Why?
 - d A random sample of 300 consumers was surveyed and 84 respondents indicated that they liked the campaign. Should the campaign be changed? Why?

- 3 An electrical goods chain is having a New Year sale of refrigerators. The number of refrigerators sold during this sale at a sample of 10 stores was:

80	110	0	40	70	80	100	50	80	30
----	-----	---	----	----	----	-----	----	----	----

 - a Is there evidence that an average of more than 50 refrigerators per store were sold during this sale? Use $\alpha = 0.05$.
 - b What assumption is necessary to perform this test?

- 4 In a survey pretest, data were obtained from 45 respondents on Benetton clothes. These data are given in the table opposite, which gives the usage, gender, awareness, attitude, preference, intention and loyalty towards Benetton of a sample of Benetton users. Usage was coded as 1, 2, or 3, representing light, medium or heavy users. Gender was coded as 1 for females and 2 for males. Awareness, attitude, preference, intention and loyalty were measured on a 7-point Likert type scale (1 = very unfavourable, 7 = very favourable). Note that 5 respondents have missing values that are denoted by 9.

Analyse the Benetton data to answer the following questions. In each case, formulate the null and alternative hypotheses and conduct the appropriate statistical test(s).

 - a Obtain a frequency distribution for each of the following variables and calculate the relevant statistics: awareness, attitude, preference, intention and loyalty towards Benetton.
 - b Conduct a cross-tabulation of the usage with gender. Interpret the results.
 - c Does the awareness for Benetton exceed 3.0?
 - d Do males and females differ in their awareness for Benetton? Their attitude towards Benetton? Their loyalty for Benetton?
 - e Do the respondents in the pretest have a higher level of awareness than loyalty?
 - f Does awareness of Benetton follow a normal distribution?
 - g Is the distribution of preference for Benetton normal?
 - h Assume that awareness towards Benetton was measured on an ordinal scale rather than an interval scale. Do males and females differ in their loyalty towards Benetton?
 - i Assume that loyalty towards Benetton was measured on an ordinal scale rather than an interval scale. Do males and females differ in their loyalty towards Benetton?
 - j Assume that attitude and loyalty towards Benetton was measured on an ordinal scale rather than an interval scale. Do the respondents have a greater awareness of Benetton than loyalty for Benetton?

Number	Usage	Gender	Awareness	Attitude	Preference	Intention	Loyalty
1	3	2	7	6	5	5	6
2	1	1	2	2	4	6	5
3	1	1	3	3	6	7	6
4	3	2	6	5	5	3	2
5	3	2	5	4	7	4	3
6	2	2	4	3	5	2	3
7	2	1	5	4	4	3	2
8	1	1	2	1	3	4	5
9	2	2	4	4	3	6	5
10	1	1	3	1	2	4	5
11	3	2	6	7	6	4	5
12	3	2	6	5	6	4	4
13	1	1	4	3	3	1	1
14	3	2	6	4	5	3	2
15	1	2	4	3	4	5	6
16	1	2	3	4	2	4	2
17	3	1	7	6	4	5	3
18	2	1	6	5	4	3	2
19	1	1	1	1	3	4	5
20	3	1	5	7	4	1	2
21	3	2	6	6	7	7	5
22	2	2	2	3	1	4	2
23	1	1	1	1	3	2	2
24	3	1	6	7	6	7	6
25	1	2	3	2	2	1	1
26	2	2	5	3	4	4	5
27	3	2	7	6	6	5	7
28	2	1	6	4	2	5	6
29	1	1	9	2	3	1	3
30	2	2	5	9	4	6	5
31	1	2	1	2	9	3	2
32	1	2	4	6	5	9	3
33	2	1	3	4	3	2	9
34	2	1	4	6	5	7	6
35	3	1	5	7	7	3	3
36	3	1	6	5	7	3	4
37	3	2	6	7	5	3	4
38	3	2	5	6	4	3	2
39	3	2	7	7	6	3	4
40	1	1	4	3	4	6	5
41	1	1	2	3	4	5	6
42	1	1	1	3	2	3	4
43	1	1	2	4	3	6	7
44	1	1	3	3	4	6	5
45	1	1	1	1	4	5	3

- 5 In a small group discuss the following issues: 'Why waste time doing basic forms of data analysis? Why not just go straight to performing multivariate analyses – whose outputs from most software packages will include basic analyses?' and 'Why do managers find cross-tabulations so appealing? What would it take to make managers more appreciative of statistical analyses that go beyond the cross-tabulation?'

Notes

- Mathwick, C., Malhotra, N.K. and Rigdon, E., 'The effect of dynamic retail experience on experiential perceptions of value: an Internet and catalog comparison', *Journal of Retailing* 78 (2002), 51–60; Pestervand, T.A., Snyder, DR. and Tsalkis, J.D., 'Influence of catalog vs. store shopping and prior satisfaction on perceived risk', *Journal of the Academy of Marketing Science* (Winter 1986), 28–36.
- Neirrti, L.D., Bosetti, HA, and Teed, K.C., 'Motivation to attend the 1996 Summer Olympic Games', *Journal of Travel Research* 39 (3) (February 2001), 327–331.
- For an application of frequencies and descriptive statistics, see Bitner, M.J., Booms, B.H. and Mohr, L.A., 'Critical service encounters: the employee's viewpoint', *Journal of Marketing* 58 (October 1994), 95–106.
- See any introductory statistics book for a more detailed description of these statistics; for example, see Berenson, M.L., Levine, D.M. and Krehbiel, T.C., *Basic Business Statistics: Concepts and Applications*, 8th edn (Englewood Cliffs, NJ: Prentice Hall, 2002).
- For our purposes, no distinction will be made between formal hypothesis testing and statistical inference by means of confidence intervals.
- Excellent discussions of ways to analyse cross-tabulations can be found in Denham, B.E., 'Advanced categorical statistics: issues and applications in communication research', *Journal of Communication* 52 (1) (March 2002), 162; Hellevik, O., *Introduction to Causal Analysis: Exploring Survey Data by Crosstabulation* (Beverly Hills, CA: Sage, 1984).
- Kivetz, R., Simonson, I., 'Earning the right to indulge: effort as a determinant of customer preferences toward frequency program rewards', *Journal of Marketing Research* 39 (2) (May 2002), 155–170; Feick, L.F., 'Analyzing marketing research data with association models', *Journal of Marketing Research* 21 (November 1984), 376–386. For an application, see Kamakura, W.A. and Wedel, M., 'Statistical data fusion for cross tabulation', *Journal of Marketing Research* 34 (4) (November 1997), 485–498.
- Lenell, W. and Boissoneau, it, 'Using causal-comparative and correlational designs in conducting market research', *Journal of Professional Services Marketing* 13 (2) (1996), 59–69. See also the classic book by Zeisel, H., *Say It with Figures*, 5th edn (New York: Harper & Row, 1968).
- Wright, D.B., *First steps in statistics* (Thousand Oaks, CA: Sage, 2002); Sirkin, R.M., *Statistics for the Social Sciences* (Thousand Oaks, CA: Sage, 1994).
- Higgins, J.J., *Introduction to modern nonparametric statistics*, (Pacific Grove, CA: Duxbury, 2002); Pet, M.A., *Nonparametric Statistics for Health Care Research* (Thousand Oaks, CA: Sage, 1997). For a more extensive treatment see Lancaster, HO., *The Chi Squared Distribution* (New York: Wiley, 1969).
- Berenson, M.L., Levine D.M. and Krehbiel, T.C., *Basic Business Statistics: Concepts and Applications*, 8th edn (Englewood Cliffs, NJ: Prentice Hall, 2002).
- Some statisticians, however, disagree. They feel that a correction should not be applied. See, for example, Overall, J.E., 'Power of chi-square tests for 2×2 contingency tables with small expected frequencies', *Psychological Bulletin* (January 1980), 132–135.
- Significance tests and confidence intervals are also available for either lambda asymmetric or lambda symmetric. See Goodman, L.A. and Kruskal, W.H., 'Measures of association for cross classification: appropriate sampling theory', *Journal of the American Statistical Association* 88 (June 1963), 310–364.
- Hoenig, J.M., 'The abuse of power: the pervasive fallacy of power calculation for data analysis', *American Statistician* 55 (1) (February 2001), 19–24; Cowles, M. and Davis, C., 'On the origins of the .05 level of statistical significance', *American Psychologist* (May 1982), 553–558. See also Kotabe, M., Duhan, D.E., Smith, D.K. Jr and Wilson, RID., 'The perceived veracity of PIMS strategy principles in Japan: an empirical inquiry', *Journal of Marketing* 55 (January 1991), 26–41.
- Technically, a null hypothesis cannot be accepted. It can be either rejected or not rejected. This distinction, however, is inconsequential in marketing research.
- The condition when the variances cannot be assumed to be equal is known as the Behrens–Fisher problem. There is some controversy over the best procedure in this case. For an example, see Heilman, C.M., Nakamoto, K. and Rao, A.G., 'Pleasant surprises: consumer response to unexpected in-store coupons', *Journal of Marketing Research* 39 (2) (May 2002), 242–252.
- Chandler, S., 'Some retailers begin to cater to growing group of aging shoppers', *Knight Ridder Tribune Business News* (17 March 2001), 1; Lumpkin, J.R. and Hunt, J.B., 'Mobility as an influence on retail patronage behavior of the elderly: testing conventional wisdom', *Journal of the Academy of Marketing Science* (Winter 1989), 1–12.
- Hainon, A., 'Skip the ads TV has Madison Ave. upset', *New York Times* (23 May 2002), A1; Dunst, L., 'Is it possible to get creative in 15 seconds?', *Advertising Age* 64 (50) (29 November 1993), 18; Rosenblatt, J.A. and Mainprize, J., 'The history and future of 15-second commercials: an empirical investigation of the perception of ad agency media directors', in Lazer, W., Shaw, E. and Wee, C.H. (eds), *World Marketing Congress*, International Conference Series, Vol. 4 (Boca Raton, FL: Academy of Marketing Science, 1989), 169–177.
- Kanji, G.K., *100 Statistical Tests* (Thousand Oaks, CA: Sage, 1999); Harnett, D.L., *Statistical Methods*, 3rd edn (Reading, MA: Addison-Wesley, 1982).
- Higgins, J.J., *Introduction to modern nonparametric statistics* (Pacific Grove, CA: Duxbury, 2002); Pen, M.A., *Nonparametric Statistics for Health Care Research* (Thousand Oaks, CA: Sage, 1997).
- There is some controversy over whether non-parametric statistical techniques should be used to make inferences about population parameters. See also Cheung, Y.K. and Klotz, J.H., 'The Mann–Whitney–Wilcoxon distribution using linked lists', *Statistica Sinica* 7 (3) (July 1997), 805–813.
- The t test in this case is equivalent to a chi-square test for independence in a 2×2 contingency table. The relationship is

$$\chi^2_{.95}(1) = t^2 \cdot .0005(n_1 + n_2 - 2)$$
 For large samples, the t distribution approaches the normal distribution and so the t test and the z test are equivalent.
- Swaddling, D., 'Good data still worth the investment', *Marketing News* 35 (2) (15 January 2001), 20–21; Krum, J.R.,

- Rau, P.A. and Keiser, S.K., 'The marketing research process: role perceptions of researchers and users', *Journal of Advertising Research* (December-January 1988), 9–21. See also Miller, C., 'Gallup Brothers analyze the research industry', *Marketing News* 31 (1) (6 January 1997), 2.
- 24 Bergmann, R., 'Different outcomes of the Wilcoxon–Mann–Whitney test from different statistics packages', *American Statistician* 54 (1) (February 2000), 72–77; and for an example of Wilcoxon matched-pairs signed-ranks test, see Kalwani, M.U. and Narayandas, N., 'Long-term manufacturer-supplier relationships: do they pay off for supplier firms?', *Journal of Marketing* 59 (January 1995), 1–16.
- 25 Pett, M.A., *Nonparametric Statistics for Health Care Research* (Thousand Oaks, CA: Sage, 1997); Field, J.G., 'The world's simplest test of significance', *Journal of the Market Research Society* (July 1971), 170–172.
- 26 Miles, L., 'Finding a balance in global research', *Marketing* (29 November 2001), 33; de Chernatony, L., Halliburton, C. and Bernath, R., 'International branding: demand or supply driven?', *International Marketing Review* 12 (2) (1995), 9–21.

Visit the *Marketing Research Companion Website* at www.pearsoned.co.uk/malhotra_euro for additional learning resources including annotated weblinks, an online glossary and a suite of downloadable video cases.