Source: © Getty Images

19

Analysis of variance and covariance

Analysis of variance is a straightforward way to examine the differences between groups of responses that are measured on interval or ratio scales.

Objectives

After reading this chapter, you should be able to:

1. discuss the scope of the analysis of variance (ANOVA) technique and its relationship to the *t* test, and regression;
2. describe one-way analysis of variance, including decomposition of the total variation, measurement of effects significance testing and interpretation of results;
3. describe *n*-way analysis of variance and the testing of the significance of the overall effect, the interaction effect and the main effect of each factor;
4. describe analysis of covariance and show how it accounts for the influence of uncontrolled independent variables;
5. explain key factors pertaining to the interpretation of results with emphasis on interactions, relative importance of factors and multiple comparisons;
6. discuss specialised ANOVA techniques applicable to marketing, such as repeated measures ANOVA, non-metric ANOVA, and multivariate analysis of variance (MANOVA).

Overview

In Chapter 18, we examined tests of differences between two means or two medians. In this chapter, we discuss procedures for examining differences between more than two means or medians. These procedures are called analysis of variance and analysis of covariance. These procedures have traditionally been used for analysing experimental data, but they are also used for analysing survey or observational data.

We describe analysis of variance and covariance procedures and discuss their relationship to other techniques. Then we describe one-way analysis of variance, the simplest of these procedures, followed by n-way analysis of variance and analysis of covariance. Special attention is given to issues in interpretation of results as they relate to interactions, relative importance of factors and multiple comparisons. Some specialised topics such as repeated measures analysis of variance, non-metric analysis of variance and multivariate analysis of variance are briefly discussed. We begin with an example illustrating the application of analysis of variance.

Example

Analysis of tourism destinations[1]

A survey conducted by EgeBank in Istanbul, Turkey, focused upon the importance of tour operators and travel agents' perceptions of selected Mediterranean tourist destinations (Egypt, Greece, Italy and Turkey). Operators/travel agents were mailed questionnaires

Image variations of destinations promoted to tour operators and travel agencies

Image items	*Turkey* (*n*=36)	*Egypt* (*n*=29)	*Greece* (*n*=37)	*Italy* (*n*=34)	*Sig.*
Affective (scale 1–7)					
Unpleasant–pleasant	6.14	5.62	6.43	6.50	0.047*
Sleepy–arousing	6.24	5.61	6.14	6.56	0.053
Distressing–relaxing	5.60	4.86	6.05	6.09	0.003*
Gloomy–exciting	6.20	5.83	6.32	6.71	0.061
Perceptual (scale 1–5)					
Good value for money	4.62	4.32	3.89	3.27	0.000*
Beautiful scenery and natural attractions	4.50	4.04	4.53	4.70	0.011*
Good climate	4.29	4.00	4.41	4.35	0.133
Interesting cultural attractions	4.76	4.79	4.67	4.79	0.781
Suitable accommodations	4.17	4.28	4.35	4.62	0.125
Appealing local food (cuisine)	4.44	3.57	4.19	4.85	0.000*
Great beaches and water sports	3.91	3.18	4.27	3.65	0.001*
Quality of infrastructure	3.49	2.97	3.68	4.09	0.000*
Personal safety	3.83	3.28	4.19	4.15	0.000*
Interesting historical attractions	4.71	4.86	4.81	4.82	0.650
Unpolluted and unspoiled environment	3.54	3.34	3.43	3.59	0.784
Good nightlife and entertainment	3.44	3.15	4.06	4.27	0.000*
Standard hygiene and cleanliness	3.29	2.79	3.76	4.29	0.000*
Interesting and friendly people	4.34	4.24	4.35	4.32	0.956

*Significant at 0.05 level.

→

based on the location of tours, broken down as follows: Egypt (53), Greece (130), Italy (150) and Turkey (65). The survey consisted of questions on affective and cognitive evaluations of the 4 destinations. The 4 affective questions were asked on a seven-point semantic differential scale, whereas the 14 cognitive evaluations were measured on a five-point Likert scale (1 = offers very little, 2 = offers somewhat little, 3 = offers neither little nor much, 4 = offers somewhat much, and 5 = offers very much). The differences in the evaluations of the four locations were examined using one-way analysis of variance (ANOVA) as seen in the table on the previous page.

The ANOVA table shows that 'unpleasant–pleasant' and 'distressing–relaxing' affective factors have significant differences among the four destinations. For instance, Greece and Italy were perceived as being significantly more relaxing than Egypt. As for the perceptual factors, 8 of the 14 factors were significant. Turkey was perceived as significantly better value for money than Greece and Italy. Turkey's main strength appears to be 'good value', and the country's tourist agencies should promote this in their marketing strategies. On the other hand, Turkey needs to improve the perception of its infrastructure, cleanliness and entertainment to attract more tour operators and travel agencies.

In this example, *t* tests were not appropriate to examine the overall difference in the four category means, so analysis of variance was used instead.

Relationship among techniques

Analysis of variance (ANOVA)
A statistical technique for examining the differences among means for two or more populations.

Factors
Categorical independent variables in ANOVA. The independent variables must all be categorical (non-metric) to use ANOVA.

Treatment
In ANOVA, a particular combination of factor levels or categories.

One-way analysis of variance
An ANOVA technique in which there is only one factor.

***n*-way analysis of variance**
An ANOVA model where two or more factors are involved.

Analysis of covariance (ANCOVA)
An advanced ANOVA procedure in which the effects of one or more metric-scaled extraneous variables are removed from the dependent variable before conducting the ANOVA.

Analysis of variance and analysis of covariance are used for examining the differences in the mean values of the dependent variable associated with the effect of the controlled independent variables, after taking into account the influence of the uncontrolled independent variables. Essentially, **analysis of variance (ANOVA)** is used as a test of means for two or more populations. The null hypothesis, typically, is that all means are equal. For example, suppose that the researcher was interested in examining whether heavy users, medium users, light users and non-users of yogurt differed in their preference for Müller yogurt, measured on a nine-point Likert scale. The null hypothesis that the four groups were not different in preference for Müller could be tested using ANOVA.

In its simplest form, ANOVA must have a dependent variable (preference for Müller yogurt) that is metric (measured using an interval or ratio scale). There must also be one or more independent variables (product use: heavy, medium, light and non-users). The independent variables must be all categorical (non-metric). Categorical independent variables are also called factors. A particular combination of **factor** levels, or categories, is called a **treatment**. **One-way analysis of variance** involves only one categorical variable, or a single factor. The differences in preference of heavy users, medium users, light users and non-users would be examined by one-way ANOVA. In this, a treatment is the same as a factor level (medium users constitute a treatment). If two or more factors are involved, the analysis is termed ***n*-way analysis of variance**. If, in addition to product use, the researcher also wanted to examine the preference for Müller yogurt of customers who are loyal and those who are not, an *n*-way ANOVA would be conducted.

If the set of independent variables consists of both categorical and metric variables, the technique is called **analysis of covariance (ANCOVA)**. For example, analysis of covariance would be required if the researcher wanted to examine the preference of product use groups and loyalty groups, taking into account the respondents' attitudes towards nutrition and the importance they attached to dairy products. The latter two variables would be measured on nine-point Likert scales. In this case, the categorical independent variables (product use and brand loyalty) are still referred to as factors, whereas the

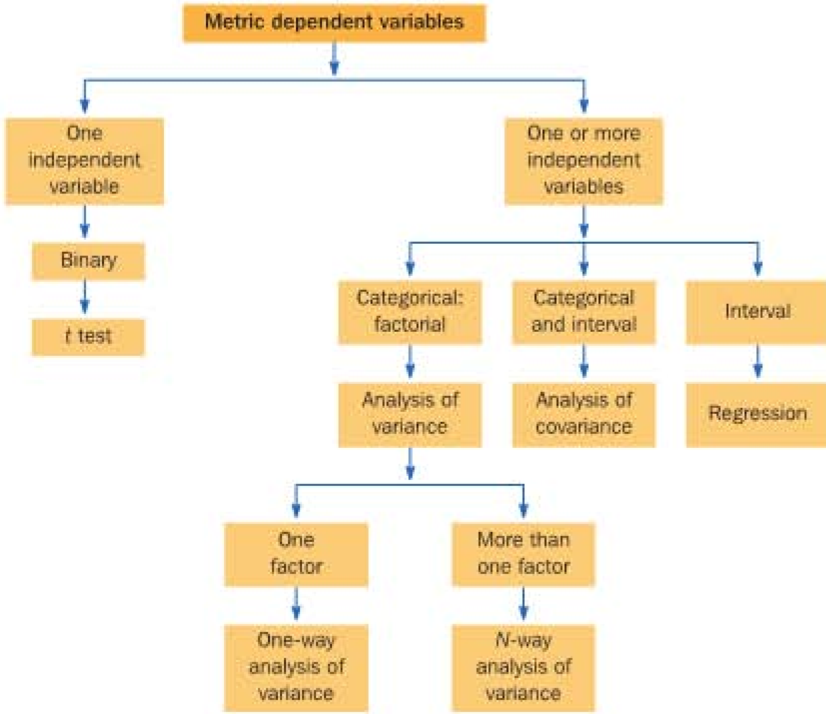


Figure 19.1 Relationship between *t* test, analysis of variance, analysis of covariance and regression

Covariate
A metric-independent variable used in ANCOVA.

metric-independent variables (attitude towards nutrition and importance attached to dairy products) are referred to as **covariates**.

The relationship of ANOVA to *t* tests and other techniques, such as regression (see Chapter 20), is shown in Figure 19.1. These techniques all involve a metric-dependent variable. ANOVA and ANCOVA can include more than one independent variable (product use, brand loyalty, attitude, importance, etc.). Furthermore, at least one of the independent variables must be categorical, and the categorical variables may have more than two categories (in our example, product use has four categories). A *t* test, on the other hand, involves a single, binary independent variable. For example, the difference in the preferences of loyal and non-loyal respondents could be tested by conducting a *t* test. Regression analysis, like ANOVA and ANCOVA, can also involve more than one independent variable. All the independent variables, however, are generally interval scaled, although binary or categorical variables can be accommodated using dummy variables. For example, the relationship between preference for Müller yogurt, attitude towards nutrition, and importance attached to dairy products could be examined via regression analysis, with preference for Müller serving as the dependent variable and attitude and importance as independent variables.

One-way ANOVA

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

- Do various market segments differ in terms of their volume of product consumption?
- Do brand evaluations of groups exposed to different commercials vary?
- Do retailers, wholesalers and agents differ in their attitudes towards the firm's distribution policies?

- How do consumers' intentions to buy the brand vary with different price levels?
- What is the effect of consumers' familiarity with a car manufacturer (measured as high, medium and low) on preference for the car?

The answer to these and similar questions can be determined by conducting one-way ANOVA. Before describing the procedure, we define the important statistics associated with one-way ANOVA.[2]

Statistics associated with one-way ANOVA

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

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

Mean square. This is the sum of squares divided by the appropriate degrees of freedom.

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

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

SS_y. This is the total variation in Y.

Conducting one-way ANOVA

The procedure for conducting one-way ANOVA is described in Figure 19.2. It involves identifying the dependent and independent variables, decomposing the total variation, measuring the effects, testing significance and interpreting the results. We consider these steps in detail and illustrate them with some applications.

Identifying the dependent and independent variables

The dependent variable is denoted by Y and the independent variable by X, and X is a categorical variable having c categories. There are n observations on Y for each category of X, as shown in Table 19.1. As can be seen, the sample size in each category of X is n, and the total sample size $N = n \times c$. Although the sample sizes in the categories of X (the group sizes) are assumed to be equal for the sake of simplicity, this is not a requirement.

Decomposing the total variation

Decomposition of the total variation
In one-way ANOVA, separation of the variation observed in the dependent variable into the variation due to the independent variables plus the variation due to error.

In examining the differences among means, one-way ANOVA involves the **decomposition of the total variation** observed in the dependent variable. This variation is measured by the sums of squares corrected for the mean (SS). ANOVA is so named because it examines the variability or variation in the sample (dependent variable) and, based on the variability, determines whether there is reason to believe that the population means differ.

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

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

where the subscripts *between* and *within* refer to the categories of X. $SS_{between}$ is the variation in Y related to the variation in the means of the categories of X. It represents variation between the categories of X. In other words, $SS_{between}$ is the portion of the sum of squares in Y related to the independent variable or factor X. For this reason, $SS_{between}$ is

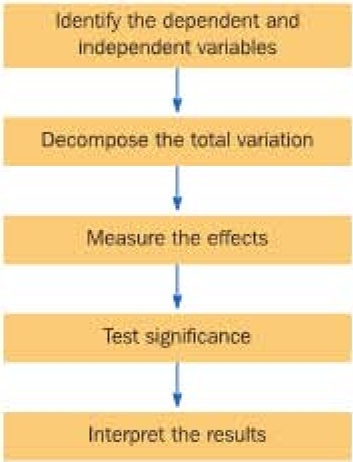


Figure 19.2 Conducting one-way ANOVA

Table 19.1 Decomposition of the total variation: one-way ANOVA

	Independent variable			X
		Categories		*Total sample*
Within-category variation = SS_{within}	X_1	X_2	$X_3 \ldots X_c$	
	Y_1	Y_1	$Y_1 \ldots Y_1$	Y_1
	Y_2	Y_2	$Y_2 \ldots Y_2$	Y_2
	$\vdots$			$\vdots$
	Y_n	Y_n	$Y_n \ldots Y_n$	Y_n
Category mean	$\bar{Y}_1$	$\bar{Y}_2$	$\bar{Y}_3 \ldots \bar{Y}_c$	

Total variation = SS_y

Between-category variation = $SS_{between}$

also denoted as SS_x. SS_{within} is the variation in Y related to the variation within each category of X. SS_{within} is not accounted for by X. Therefore, it is referred to as SS_{error}. The total variation in Y may be decomposed as

$$SS_y = SS_x + SS_{error}$$

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

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

$$SS_{error} = \sum_{j}^{c} \sum_{i}^{n} (\bar{Y}_{ij} - \bar{Y}_j)^2$$

and Y_i = indivdual observation

$\bar{Y}_j$ = mean for category j

$\bar{Y}$ = mean over the whole sample or grand mean

Y_{ij} = ith observation in the jth category

The logic of decomposing the total variation in Y, SS_y into $SS_{between}$ and SS_{within} to examine differences in group means can be intuitively understood. Recall from Chapter 18 that, if the variation of the variable in the population was known or estimated, one could estimate how much the sample mean should vary because of random variation alone. In ANOVA, there are several different groups (e.g. heavy, medium and light users and non-users). If the null hypothesis is true and all the groups have the same mean in the population, one can estimate how much the sample means should vary because of sampling (random) variations alone. If the observed variation in the sample means is more than what would be expected by sampling variation, it is reasonable to conclude that this extra variability is related to differences in group means in the population.

In ANOVA, we estimate two measures of variation: within groups (SS_{within}) and between groups ($SS_{between}$). Within-group variation is a measure of how much the observations, Y values, within a group vary. This is used to estimate the variance within a group in the population. It is assumed that all the groups have the same variation in the population. But because it is not known that all the groups have the same mean, we cannot calculate the variance of all the observations together. The variance for each of the groups must be calculated individually, and these are combined into an 'average' or 'overall' variance. Likewise, another estimate of the variance of the Y values may be obtained by examining the variation between the means. (This process is the reverse of determining the variation in the means, given the population variances.) If the population mean is the same in all the groups, then the variation in the sample means and the sizes of the sample groups can be used to estimate the variance of Y. The reasonableness of this estimate of the Y variance depends on whether the null hypothesis is true. If the null hypothesis is true and the population means are equal, the variance estimate based on between-group variation is correct. On the other hand, if the groups have different means in the population, the variance estimate based on between-group variation will be too large. Thus, by comparing the Y variance estimates based on between-group and within-group variation, we can test the null hypothesis. Decomposition of the total variation in this manner also enables us to measure the effects of X on Y.

Measuring the effects

The effects of X on Y are measured by SS_x. Because SS_x is related to the variation in the means of the categories of X, the relative magnitude of SS_x increases as the differences among the means of Y in the categories of X increase. The relative magnitude of SS_x also increases as the variations in Y within the categories of X decrease. The strength of the effects of X on Y is measured as follows:

$$\eta^2 = \frac{SS_x}{SS_y} = \frac{SS_y - SS_{error}}{SS_y}$$

The value of η^2 varies between 0 and 1. It assumes a value of 0 when all the category means are equal, indicating that X has no effect on Y. The value of η^2 will be 1 when there is no variability within each category of X but there is some variability between categories. Thus, η^2 is a measure of the variation in Y that is explained by the independent variable X. Not only can we measure the effects of X on Y, but we can also test for their significance.

Testing the significance

In one-way ANOVA, the interest lies in testing the null hypothesis that the category means are equal in the population.[3] In other words,

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

Under the null hypothesis, SS_x and SS_{error} come from the same source of variation. In such a case, the estimate of the population variance of Y can be based on either between-category variation or within-category variation. In other words, the estimate of the population variance of Y,

$$S_y^2 = \frac{SS_x}{c-1}$$
$$= \text{mean square due to } X$$
$$= MS_x$$

or

$$S_y^2 = \frac{SS_{error}}{N-c}$$
$$= \text{mean square due to error}$$
$$= MS_{error}$$

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

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

This statistic follows the F distribution, with $(c-1)$ and $(N-c)$ degrees of freedom (df). A table of the F distribution is given as Table 5 in the Appendix of statistical tables at the end of the book. As mentioned in Chapter 18, the F distribution is a probability distribution of the ratios of sample variances. It is characterised by degrees of freedom for the numerator and degrees of freedom for the denominator.[4]

Interpreting results

If the null hypothesis of equal category means is not rejected, then the independent variable does not have a significant effect on the dependent variable. On the other hand, if the null hypothesis is rejected, then the effect of the independent variable is significant. In other words, the mean value of the dependent variable will be different for different categories of the independent variable. A comparison of the category mean values will indicate the nature of the effect of the independent variable. Other salient issues in the interpretation of results, such as examination of differences among specific means, are discussed later.

Illustrative applications of one-way ANOVA

We illustrate the concepts discussed in this section using the data presented in Table 19.2. These data were generated by an experiment in which Renault wanted to examine the effect of direct mail offers and dealership promotions upon the level of sales of new cars. Dealership promotion was varied at three levels: high (1), medium (2) and low (3). Direct mail efforts were manipulated at two levels. Either an exclusive boxed set of DVDs covering the history of Formula One Racing was offered to customers who bought a new car (denoted by 1) or it was not (denoted by 2 in Table 19.2). Dealership promotion and direct mail offer were crossed, resulting in a 3 × 2 design with six cells. Thirty Renault dealerships were randomly selected, and five dealerships were randomly assigned to each treatment condition. The experiment ran for two months. The sales level of new cars were measured, normalised to account for extraneous factors (e.g. dealership size, competitive dealerships in town) and converted to a 1 to 10 scale (10 representing the highest level of sales). In addition, a qualitative assessment was made of the relative affluence of the clientele of each dealership, again using a 1 to 10 scale (10 representing the most affluent client base).

Table 19.2 Direct mail offer, dealership promotion, sales of new cars and clientele rating

Dealership number	Direct mail offer	Dealership promotion	Sales	Clientele rating
1	1	1	10	9
2	1	1	9	10
3	1	1	10	8
4	1	1	8	4
5	1	1	9	6
6	1	2	8	8
7	1	2	8	4
8	1	2	7	10
9	1	2	9	6
10	1	2	6	9
11	1	3	5	8
12	1	3	7	9
13	1	3	6	6
14	1	3	4	10
15	1	3	5	4
16	2	1	8	10
17	2	1	9	6
18	2	1	7	8
19	2	1	7	4
20	2	1	6	9
21	2	2	4	6
22	2	2	5	8
23	2	2	5	10
24	2	2	6	4
25	2	2	4	9
26	2	3	2	4
27	2	3	3	6
28	2	3	2	10
29	2	3	1	9
30	2	3	2	8

To illustrate the concepts of ANOVA, we begin with an example showing calculations done by hand and then by computer. Suppose that only one factor, namely dealership promotion, was manipulated; that is, let us ignore the direct mail efforts for the purpose of this illustration. Renault is attempting to determine the effect of dealership promotion (X) on the sales of new cars (Y). For the purpose of illustrating hand calculations, the data of Table 19.2 are transformed in Table 19.3 to show the dealership (Y_{ij}) for each level of promotion.

The null hypothesis is that the category means are equal:

$$H_0\colon\ \mu_1 = \mu_2 = \mu_3$$

To test the null hypothesis, the various sums of squares are computed as follows:

$$\begin{aligned} SS_y = {} & (10-6.067)^2 + (9-6.067)^2 + (10-6.067)^2 + (8-6.067)^2 + (9-6.067)^2 + (8-6.067)^2 \\ & + (8-6.067)^2 + (7-6.067)^2 + (9-6.067)^2 + (6-6.067)^2 + (5-6.067)^2 + (7-6.067)^2 \\ & + (6-6.067)^2 + (4-6.067)^2 + (5-6.067)^2 + (8-6.067)^2 + (9-6.067)^2 + (7-6.067)^2 \\ & + (7-6.067)^2 + (6-6.067)^2 + (4-6.067)^2 + (5-6.067)^2 + (5-6.067)^2 + (6-6.067)^2 \\ & + (4-6.067)^2 + (2-6.067)^2 + (3-6.067)^2 + (2-6.067)^2 + (1-6.067)^2 + (2-6.067)^2 \\ = {} & 185.867 \end{aligned}$$

$$SS_x = 10(8.3 - 6.067)^2 + 10(6.2 - 6.067)^2 + 10(3.7 - 6.067)^2$$
$$= 106.067$$

$$SS_{error} = (10 - 8.3)^2 + (9 - 8.3)^2 + (10 - 8.3)^2 + (8 - 8.3)^2 + (9 - 8.3)^2 + (8 - 8.3)^2 + (9 - 8.3)^2$$
$$+ (7 - 8.3)^2 + (7 - 8.3)^2 + (6 - 8.3)^2 + (8 - 8.3)^2 + (8 - 6.2)^2 + (7 - 6.2)^2 + (9 - 6.2)^2$$
$$+ (6 - 6.2)^2 + (4 - 6.2)^2 + (5 - 6.2)^2 + (5 - 6.2)^2 + (6 - 6.2)^2 + (4 - 6.2)^2 + (5 - 3.7)^2$$
$$+ (7 - 3.7)^2 + (6 - 3.7)^2 + (4 - 3.7)^2 + (5 - 3.7)^2 + (2 - 3.7)^2 + (3 - 3.7)^2 + (2 - 3.7)^2$$
$$+ (1 - 3.7)^2 + (2 - 3.7)^2$$
$$= 79.8$$

Table 19.3 Effect of dealership promotion on sales of new cars

Dealership number	*Level of dealership promotion*		
	High	*Medium*	*Low*
	Normalised sales		
1	10	8	5
2	9	8	7
3	10	7	6
4	8	9	4
5	9	6	5
6	8	4	2
7	9	5	3
8	7	5	2
9	7	6	1
10	6	4	2
Column totals	83	62	37
Category means: $\bar{Y}_j$	$\frac{83}{10} = 8{\cdot}3$	$\frac{62}{10} = 6{\cdot}2$	$\frac{37}{10} = 3{\cdot}7$
Grand means: $\bar{\bar{Y}} = \frac{83 + 62 + 37}{30} = 6.067$			

It can be verified that

$$SS_y = SS_x + SS_{error}$$

as follows:

$$185.867 = 106.067 + 79.80$$

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

$$\eta^2 = \frac{SS_x}{SS_y}$$
$$= \frac{106.067}{185.897}$$
$$= 0.571$$

In other words, 57.1% of the variation in sales (Y) is accounted for by dealership promotion (X), indicating a modest effect. The null hypothesis may now be tested:

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

$$= \frac{106.067/(3-1)}{79.8/(30-3)}$$

$$= 17.944$$

From Table 5 in the Appendix of statistical tables we see that, for 2 and 27 degrees of freedom, the critical value of F is 3.35 for $\alpha = 0.05$. Because the calculated value of F is greater than the critical value, we reject the null hypothesis. We conclude that the population means for the three levels of dealership promotion are indeed different. The relative magnitudes of the means for the three categories indicate that a high level of dealership promotion leads to significantly higher sales of new car sales.

We now illustrate the ANOVA procedure using a computer program. The results of conducting the same analysis by computer are presented in Table 19.4.

Table 19.4 **One-way ANOVA: effect of dealership promotion on the sale of new cars**

Source of variation	*Sum of squares*	*df*	*Mean square*	*F ratio*	*F probability*
Between groups (dealership promotion)	106.067	2	53.033	17.944	0.000
Within groups (error)	79.800	27	2.956		
Total	185.867	29	6.409		

Cell means

Level of dealership promotion	*Count*	*Mean*
High (1)	10	8.300
Medium (2)	10	6.200
Low (3)	10	3.700
Total	30	6.067

The value of SS_x denoted by main effects is 106.067 with 2 *df*; that of SS_{error} (within-group sums of squares) is 79.80 with 27 *df*. Therefore, $MS_x = 106.067/2 = 53.033$ and $MS_{error} = 79.80/27 = 2.956$. The value of $F = 53.033/2.956 = 17.944$ with 2 and 27 *df*, resulting in a probability of 0.000. Since the associated probability is less than the significance level of 0.05, the null hypothesis of equal population means is rejected. Alternatively, it can be seen from Table 5 in the Appendix that the critical value of F for 2 and 27 *df* is 3.35. Since the calculated value of F (17.944) is larger than the critical value, the null hypothesis is rejected. As can be seen from Table 19.4, the sample means with values of 8.3, 6.2 and 3.7 are quite different. Dealerships with a high level of promotions have the highest average sales (8.3) and dealerships with a low level of promotions have the lowest average sales (3.7). Dealerships with a medium level of promotions have an intermediate level of sales (6.2). These findings seem plausible.

Assumptions in ANOVA

The procedure for conducting one-way ANOVA and the illustrative applications help us understand the assumptions involved. The salient assumptions in ANOVA can be summarised as follows:

1 Ordinarily, the categories of the independent variable are assumed to be fixed. Inferences are made only to the specific categories considered. This is referred to as the *fixed-effects model*. Other models are also available. In the *random-effects model*, the categories or treatments are considered to be random samples from a universe of treatments. Inferences are made to other categories not examined in the analysis. A *mixed-effects model* results if some treatments are considered fixed and others random.[5]
2 The error term is normally distributed, with a zero mean and a constant variance. The error is not related to any of the categories of *X*. Modest departures from these assumptions do not seriously affect the validity of the analysis. Furthermore, the data can be transformed to satisfy the assumption of normality or equal variances.
3 The error terms are uncorrelated. If the error terms are correlated (i.e. the observations are not independent), the *F* ratio can be seriously distorted.

In many data analysis situations, these assumptions are reasonably met. ANOVA is therefore a common procedure.

N-way ANOVA

In marketing research, one is often concerned with the effect of more than one factor simultaneously.[6] For example:

- How do consumers' intentions to buy a brand vary with different levels of price and different levels of distribution?
- How do advertising levels (high, medium and low) interact with price levels (high, medium and low) to influence a brand's sale?
- Do income levels (high, medium and low) and age (younger than 35, 35–55, older than 55) affect consumption of a brand?
- What is the effect of consumers' familiarity with a bank (high, medium and low) and bank image (positive, neutral and negative) on preference for taking a loan out with that bank?

In determining such effects, *n*-way ANOVA can be used as illustrated in the following example. This example involves a comparison of means where there were two factors (independent variables), each of which was varied at two levels.

Example

Electronic shopping risks[7]

Analysis of variance was used to test differences in preferences for electronic shopping for products with different economic and social risks. In a 2 × 2 design, economic risk and social risk were varied at two levels each (high, low). Preference for electronic shopping served as the dependent variable. The results indicted a significant interaction of social risk with economic risk. Electronic shopping was not favoured for high-social-risk products, regardless of the level of economic product risk. It was preferred for low-economic-risk products over high-economic-risk products when the level of social risk was low. Despite the results of this study, year on year, the number of online shoppers has increased significantly. The increase in shoppers can be attributed to bargain-seeking consumers, convenience of using the Internet and, surprisingly, an added sense of safety associated with purchasing online. Improved websites, streamlined order taking and delivery, and assurances of more secure payment systems have increased the flow of new shoppers to the Internet while decreasing the traditional risk associated with online transactions.

Interaction
When assessing the relationship between two variables, an interaction occurs if the effect of X_1 depends on the level of X_2, and vice versa.

A major advantage of *n*-way ANOVA is that it enables the researcher to examine **interactions** between the factors. Interactions occur when the effects of one factor on the dependent variable depend on the level (category) of the other factors. The procedure for conducting n-way ANOVA is similar to that for one-way ANOVA. The statistics associated with n-way ANOVA are also defined similarly. Consider the simple case of two factors X_1 and X_2 having categories c_1 and c_2. The total variation in this case is partitioned as follows:

$$SS_{total} = SS \text{ due to } X_1 + SS \text{ due to } X_2 + SS \text{ due to intreraction of } X_1 \text{ and } X_2 + SS_{within}$$

or

$$SS_y = SS_{x_1} + SS_{x_2} + SS_{x_1x_2} + SS_{error}$$

A larger effect of X_1 will be reflected in a greater mean difference in the levels of X_1 and a larger SS_{x_1}. The same is true for the effect of X_2. The larger the interaction between X_1 and X_2, the larger $SS_{x_1x_2}$ will be. On the other hand, if X_1 and X_2 are independent, the value of $SS_{x_1x_2}$ will be close to zero.[8]

Multiple η^2
The strength of the joint effect of two (or more) factors, or the overall effect.

The strength of the joint effect of two factors, called the overall effect, or **multiple η^2**, is measured as follows:

$$\text{multiple } \eta^2 = (SS_{x_1} + SS_{x_2} + SS_{x_1x_2})/SS_y$$

Significance of the overall effect
A test that some differences exist between some of the treatment groups.

The **significance of the overall effect** may be tested by an F test, as follows:

$$F = \frac{(SS_{x_1} + SS_{x_2} + SS_{x_1x_2})/df_n}{SS_{error}/df_d}$$

$$= \frac{SS_{x_1,x_2,x_1x_2}/df_n}{SS_{error}/df_d}$$

$$= \frac{MS_{x_1,x_2,x_1x_2}}{MS_{error}}$$

where df_n = degrees of freedom for the numerator
$= (c_1 - 1) + (c_2 - 1) + (c_1 - 1)(c_2 - 1)$
$= c_1c_2 - 1$
df_d = degrees of freedom for the denominator
$= N - c_1c_2$
MS = mean square

Significance of the interaction effect
A test of the significance of the interaction between two or more independent variables.

If the overall effect is significant, the next step is to examine the **significance of the interaction effect**.[9] Under the null hypothesis of no interaction, the appropriate F test is:

$$F = \frac{SS_{x_1x_2}/df_n}{SS_{error}/df_d}$$

$$= \frac{MS_{x_1x_2}}{MS_{error}}$$

where $df_n = (c_1 - 1)(c_2 - 1)$
$df_d = N - c_1c_2$

Significance of the main effect of each factor
A test of the significance of the main effect for each individual factor.

If the interaction effect is found to be significant, then the effect of X_1 depends on the level of X_2, and vice versa. Since the effect of one factor is not uniform but varies with the level of the other factor, it is not generally meaningful to test the **significance of the main effect of each factor**. It is meaningful to test the significance of each main effect of each factor, if the interaction effect is not significant.[10]

The significance of the main effect of each factor may be tested as follows for X_1:

$$F = \frac{SS_{x_1}/df_n}{SS_{error}/df_d}$$

$$= \frac{MS_{x_1}}{MS_{error}}$$

where $df_n = c_1 - 1$
$df_d = N - c_1c_2$

The foregoing analysis assumes that the design was orthogonal, or balanced (the number of cases in each cell was the same). If the cell size varies, the analysis becomes more complex.

Returning to the data in Table 19.2, let us now examine the effect of the level of dealership promotion and direct mail efforts on the sales of new cars. The results of running a 3 × 2 ANOVA on the computer are presented in Table 19.5.

Table 19.5 Two-way ANOVA

Source of variation	*Sum of squares*	*df*	*Mean square*	*F*	*Sig. of F*	ω^2
Main effects						
Dealer promotions	106.067	2	53.033	54.862	0.000	0.557
Direct mail	53.333	1	53.333	55.172	0.000	0.280
Combined	159.400	3	53.133	54.966	0.000	
Two-way interaction	3.267	2	1.633	1.690	0.206	
Model	162.667	5	32.533	33.655	0.000	
Residual (error)	23.200	24	0.967			
Total	185.867	29	6.409			
Cell means						
Dealership promotions	*Direct mail*		*Count*		*Mean*	
High	Yes		5		9.200	
High	No		5		7.400	
Medium	Yes		5		7.600	
Medium	No		5		4.800	
Low	Yes		5		5.400	
Low	No		5		2.000	
Factor-level means						
Dealership promotions	*Direct mail*		*Count*		*Mean*	
High			10		8.300	
Medium			10		6.200	
Low			10		3.700	
	Yes		15		7.400	
	No		15		4.733	
Grand mean			30		6.067	

For the main effect of level of promotion, the sum of squares SS_{xp}, degrees of freedom and mean square MS_{xp} are the same as earlier determined in Table 19.4. The sum of squares for direct mail SS_{xd} = 53.333 with 1 *df*, resulting in an identical value for the mean square MS_{xd}. The combined main effect is determined by adding the sum of squares due to the two main effects ($SS_{xp} + SS_{xd} = 106.067 + 53.333 = 159.400$) as well as adding the degrees of freedom (2 + 1 = 3). For the promotions and direct mail interaction effect, the sum of squares SS_{xpxd} = 3.267 with (3 − 1) × (2 − 1) = 2 *df*, resulting in MS_{xpxd} = 3.267/2 = 1.633. For the overall (model) effect, the sum of squares is the sum of squares for promotions main effect, direct mail main effect and interaction effect = 106.067 + 53.333 + 3.267 = 162.667 with 2 + 1 + 2 = 5 *df*, resulting in a mean square of 162.667/5 = 32.533. Note, however, that the error statistics are now different from those in Table 19.4. This is due to the fact that we now have two factors instead of one, SS_{error} = 23.2 with [30 − (3 × 2)] or 24 *df* resulting in MS_{error} = 23.2/24 = 0.967.

The test statistic for the significance of the overall effect is

$$F = \frac{32.533}{0.967} = 33.643$$

with 5 and 24 *df*, which is significant at the 0.05 level.

The test statistic for the significance of the interaction effect is

$$F = \frac{1.633}{0.967} = 1.69$$

with 2 and 24 *df*, which is not significant at the 0.05 level.

As the interaction effect is not significant, the significance of the main effects can be evaluated. The test statistic for the significance of the main effect of promotion is

$$F = \frac{53.033}{0.967} = 54.843$$

with 2 and 24 *df*, which is significant at the 0.05 level.

The test statistic for the significance of the main effect of direct mail is

$$F = \frac{53.333}{0.967} = 55.153$$

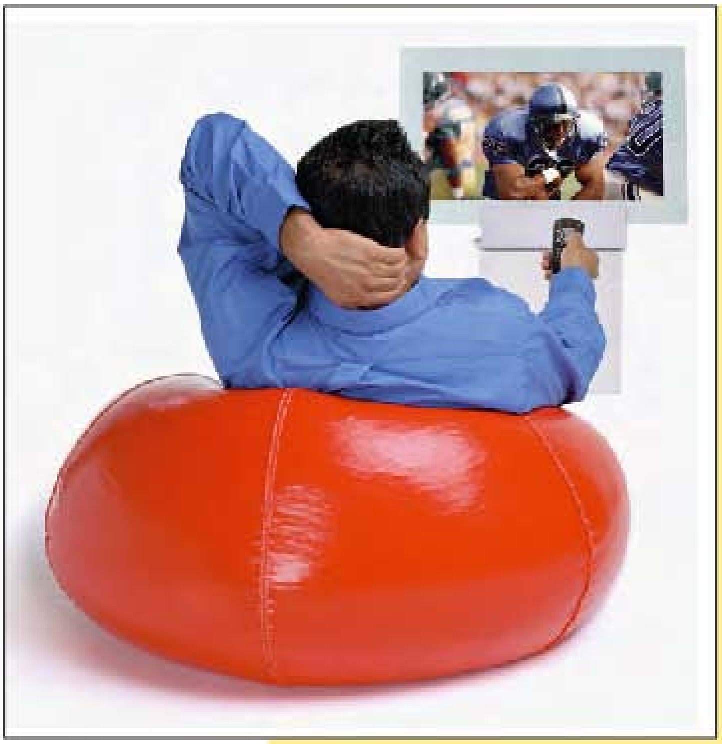
with 1 and 24 *df*, which is significant at the 0.05 level. Thus, higher levels of promotions result in higher sales. The use of a direct mail campaign results in higher sales. The effect of each is independent of the other.

The following example illustrates the use of *n*-way analysis.

Example

Country affects TV reception[11]

A study examined the impact of country affiliation on the credibility of product-attribute claims for TVs. The dependent variables were the following product-attribute claims: good sound, reliability, crisp-clear picture and stylish design. The independent variables which were manipulated consisted of price, country affiliation and store distribution. A 2 × 2 × 2 between-subjects design was used. Two levels of price, 'low' and 'high', two levels of country affiliation, South Korea and Germany, and two levels of store distribution, Kaufhof and without Kaufhof, were specified.



Source: © Getty Images

Data were collected from two shopping centres in a large German city. Thirty respondents were randomly assigned to each of the eight treatment cells for a total of 240 subjects. Table 1 presents the results for manipulations that had significant effects on each of the dependent variables.

The directions of country-by-distribution interaction effects for the three dependent variables are shown in Table 2. Although the credibility ratings for the crisp-clear picture, reliability and stylish design claims are improved by distributing the Korean-made TV set through Kaufhof rather than some other distributor, the same is not true of a German-made set. Similarly, the directions of country-by-price interaction effects for the two dependent variables are shown in Table 3. At the 'high' price level, the credibility ratings for the 'good sound' and 'reliability' claims are higher for the German-made TV set than for its Korean counterpart, but there is little difference related to country affiliation when the product is at the 'low' price.

Table 1 Analyses for significant manipulations

Effect	*Univariate*			
	Dependent variable	*F*	*df*	*p*
Country × price	Good sound	7.57	1.232	0.006
Country × price	Reliability	6.57	1.232	0.011
Country × distribution	Crisp-clear picture	6.17	1.232	0.014
Country × distribution	Reliability	6.57	1.232	0.011
Country × distribution	Stylish design	10.31	1.232	0.002

Table 2 Country-by-distribution interaction means

Country × distribution	*Crisp-clear picture*	*Reliability*	*Stylish design*
South Korea			
Kaufhof	3.67	3.42	3.82
Without Kaufhof	3.18	2.88	3.15
Germany			
Kaufhof	3.60	3.47	3.53
Without Kaufhof	3.77	3.65	3.75

Table 3 Country-by-price interaction means

Country × price	*Good sound*	*Reliability*
Low price		
Kaufhof	3.75	3.40
Without Kaufhof	3.53	3.45
High price		
Kaufhof	3.15	2.90
Without Kaufhof	3.73	3.67

This study demonstrates that credibility of attribute claims, for products traditionally exported to Germany by a company in a newly industrialised country, can be significantly improved if the same company distributes the product through a prestigious German retailer and considers making manufacturing investments in Europe. Specifically, three product-attribute claims (crisp-clear picture, reliability and stylish design) are perceived as more credible when the TVs are made in South Korea if they are also distributed through a prestigious German retailer. Also, the 'good sound' and 'reliability' claims for TVs are perceived to be more credible for a German-made set sold at a higher price, possibly offsetting the potential disadvantage of higher manufacturing costs in Europe.

Analysis of covariance (ANCOVA)

When examining the differences in the mean values of the dependent variable related to the effect of the controlled independent variables, it is often necessary to take into account the influence of uncontrolled independent variables. For example:

- In determining how consumers' intentions to buy a brand vary with different levels of price, attitude towards the brand may have to be taken into consideration.
- In determining how different groups exposed to different commercials evaluate a brand, it may be necessary to control for prior knowledge.
- In determining how different price levels will affect a household's breakfast cereal consumption, it may be essential to take household size into account.

In such cases, ANCOVA should be used. ANCOVA includes at least one categorical independent variable and at least one interval or metric-independent variable. The categorical independent variable is called a *factor*, whereas the metric-independent variable is called a *covariate*. The most common use of the covariate is to remove extraneous variation from the dependent variable, because the effects of the factors are of major concern. The variation in the dependent variable due to the covariates is removed by an adjustment of the dependent variable's mean value within each treatment condition.

An ANOVA is then performed on the adjusted scores.[12] The significance of the combined effect of the covariates, as well as the effect of each covariate, is tested by using the appropriate *F* tests. The coefficients for the covariates provide insights into the effect that the covariates exert on the dependent variable. ANCOVA is most useful when the covariate is linearly related to the dependent variable and is not related to the factors.[13]

Illustrative application of covariance

We again use the data of Table 19.2 to illustrate ANCOVA. Suppose that we wanted to determine the effect of dealership promotion and direct mail on sales while controlling for the affluence of clientele. It is felt that the affluence of the clientele may also have an effect on the sales of new cars. The dependent variable consists of new car sales. As before, promotion has three levels and direct mail has two. Clientele affluence is measured on an interval scale and serves as the covariate. The results are shown in Table 19.6.

Table 19.6 ANCOVA

Source of variation	*Sum of squares*	*df*	*Mean square*	*F*	*Sig. of F*
Covariates					
Clientele	0.838	1	0.838	0.862	0.363
Main effects					
Promotions	106.067	2	53.033	54.546	0.000
Direct mail	53.333	1	53.333	54.855	0.000
Combined	159.400	3	53.133	54.649	0.000
Two-way interaction					
*Promotions*Direct mail*	3.267	2	1.633	1.680	0.208
Model	163.505	6	27.251	28.028	0.000
Residual (error)	22.362	23	0.972		
Total	185.867	29	6.409		
Covariate	Raw coefficient				
Clientele	−0.078				

As can be seen, the sum of squares attributable to the covariate is very small (0.838) with 1 *df* resulting in an identical value for the mean square. The associated *F* value is 0.838/0.972 = 0.862, with 1 and 23 *df*, which is not significant at the 0.05 level. Thus, the conclusion is that the affluence of the clientele does not have an effect on the sales of new Renault cars. If the effect of the covariate is significant, the sign of the raw coefficient can be used to interpret the direction of the effect on the dependent variable.

Issues in interpretation

Important issues involved in the interpretation of ANOVA results include interactions, relative importance of factors, and multiple comparisons.

Interactions

The different interactions that can arise when conducting ANOVA on two or more factors are shown in Figure 19.3.

One outcome is that ANOVA may indicate that there are no interactions (the interaction effects are not found to be significant). The other possibility is that the interaction is significant. An interaction effect occurs when the effect of an independent variable on a dependent variable is different for different categories or levels of another independent

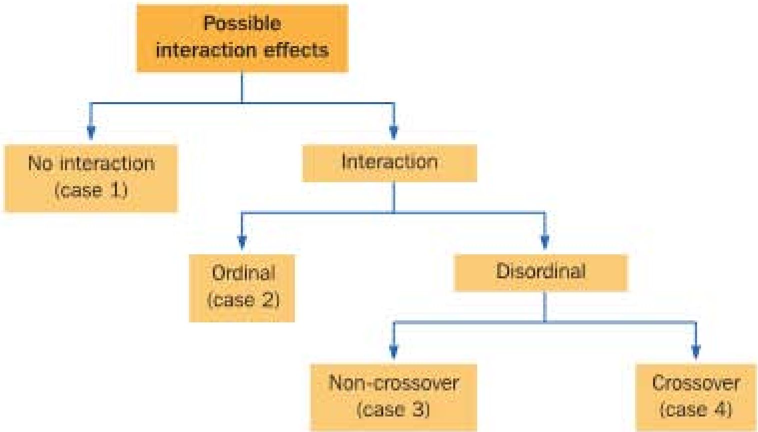


Figure 19.3
A classification of interaction effects

Ordinal interaction
An interaction where the rank order of the effects attributable to one factor does not change across the levels of the second factor.

Disordinal interaction
The change in the rank order of the effects of one factor across the levels of another.

variable. The interaction may be ordinal or disordinal. In **ordinal interaction**, the rank order of the effects related to one factor does not change across the levels of the second factor. **Disordinal interaction**, on the other hand, involves a change in the rank order of the effects of one factor across the levels of another. If the interaction is disordinal, it could be of a non-crossover or crossover type.[14] These interaction cases are displayed in Figure 19.4, which assumes that there are two factors, X_1 with three levels (X_{11}, X_{12} and X_{13}) and X_2 with two levels (X_{21} and X_{22}). Case 1 depicts no interaction.

The effects of X_1 on Y are parallel over the two levels of X_2. Although there is some departure from parallelism, this is not beyond what might be expected from chance. Parallelism implies that the net effect of X_{22} over X_{21} is the same across the three levels of X_1. In the absence of interaction, the joint effect of X_1 and X_2 is simply the sum of their individual main effects.

Case 2 depicts an ordinal interaction. The line segments depicting the effects of X_1 and X_2 are not parallel. The difference between X_{22} and X_{21} increases as we move from X_{11} to X_{12} and from X_{12} to X_{13}, but the rank order of the effects of X_1 is the same over the two levels of X_2. This rank order, in ascending order, is X_{11}, X_{12}, X_{13}, and it remains the same for X_{21} and X_{22}.

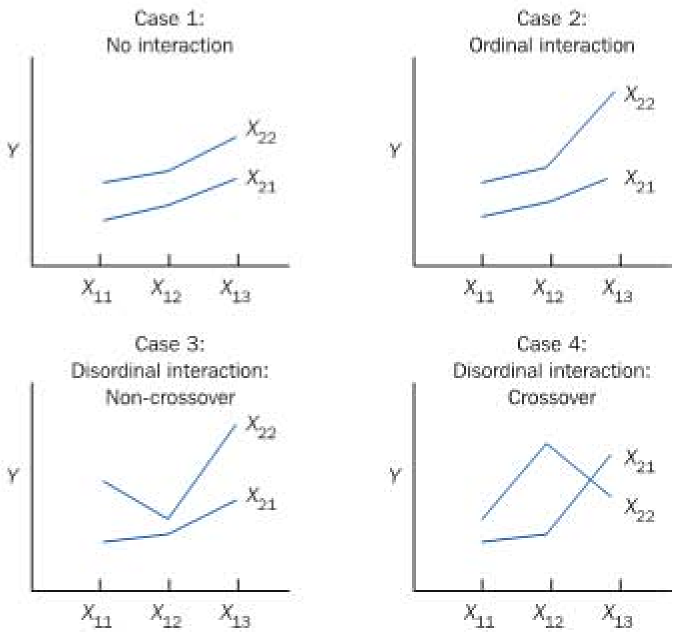


Figure 19.4
Patterns of interaction

Disordinal interaction of a non-crossover type is displayed by case 3. The lowest effect of X_1 at level X_{21} occurs at X_{11}, and the rank order of effects is X_{11}, X_{12}, X_{13}. At level X_{22}, however, the lowest effect of X_1 occurs at X_{12}, and the rank order is changed to X_{12}, X_{11}, X_{13}. Because it involves a change in rank order, disordinal interaction is stronger than ordinal interaction.

In disordinal interactions of a crossover type, the line segments cross each other, as shown by case 4 in Figure 19.4. In this case, the relative effect of the levels of one factor changes with the levels of the other. Note that X_{22} has a greater effect than X_{21} when the levels of X_1 are X_{11} and X_{12}. When the level of X_1 is X_{13}, the situation is reversed, and X_{21} has a greater effect than X_{22}. (Note that in cases 1, 2 and 3, X_{22} had a greater impact than X_{21} across all three levels of X_1.) Hence, disordinal interactions of a crossover type represent the strongest interactions.[15]

Relative importance of factors

Omega squared (ω^2)
A measure indicating the proportion of the variation in the dependent variable that is related to a particular independent variable or factor.

Experimental designs are usually balanced in that each cell contains the same number of respondents. This results in an orthogonal design in which the factors are uncorrelated. Hence, it is possible to determine unambiguously the relative importance of each factor in explaining the variation in the dependent variable.[16] The most commonly used measure in ANOVA is **omega squared (ω^2)**. This measure indicates what proportion of the variation in the dependent variable is related to a particular independent variable or factor. The relative contribution of a factor X is calculated as follows:[17]

$$\omega_x^2 = \frac{SS_x - (df_x \times MS_{error})}{SS_{total} + MS_{error}}$$

Note that the estimated value of ω^2 can be negative, in which case the estimated value of ω^2 is set equal to zero. Normally, ω^2 is interpreted only for statistically significant effects.[18] In Table 19.4, ω^2 associated with level of dealership promotions is calculated as follows:

$$\omega_p^2 = \frac{106.067 - (2 \times 0.967)}{185.867 + 0.967}$$
$$= \frac{104.133}{186.834}$$
$$= 0.557$$

In Table 19.5 note that

$$SS_{total} = 106.067 + 53.333 + 3.267 + 23.2$$
$$= 185.867$$

Likewise, the ω^2 associated with direct mail is

$$\omega^2 = \frac{53.333 - (1 \times 0.967)}{185.867 + 0.967}$$
$$= \frac{52.366}{186.834}$$
$$= 0.280$$

As a guide to interpreting ω^2, a large experimental effect produces an ω^2 of 0.15 or greater, a medium effect produces an index of around 0.06, and a small effect produces an index of 0.01.[19] In Table 19.5, while the effects of promotions and direct mail are both large, the effect of promotions is much larger. Therefore, dealership promotions will be more effective in increasing sales than direct mail.

Multiple comparisons

Contrasts
In ANOVA, a method of examining differences among two or more means of the treatment groups.

A priori contrasts
Contrasts determined before conducting the analysis, based on the researcher's theoretical framework.

A posteriori contrasts
Contrasts made after conducting the analysis. These are generally multiple comparison tests.

Multiple comparison tests
A posteriori contrasts that enable the researcher to construct generalised confidence intervals that can be used to make pairwise comparisons of all treatment means.

The ANOVA F test examines only the overall difference in means. If the null hypothesis of equal means is rejected, we can only conclude that not all the group means are equal. Only some of the means may be statistically different, however, and we may wish to examine differences among specific means. This can be done by specifying appropriate **contrasts**, or comparisons used to determine which of the means are statistically different. Contrasts may be a priori or a posteriori. **A priori contrasts** are determined before conducting the analysis, based on the researcher's theoretical framework. Generally, a priori contrasts are used in lieu of the ANOVA F test. The contrasts selected are orthogonal (they are independent in a statistical sense).

A posteriori contrasts are made after the analysis. These are generally **multiple comparison tests**. They enable the researcher to construct generalised confidence intervals that can be used to make pairwise comparisons of all treatment means. These tests, listed in order of decreasing power, include least significant difference, Duncan's multiple range test, Student–Newman–Keuls, Tukey's alternate procedure, honestly significant difference, modified least significant difference and Scheffé's tests. Of these tests, least significant difference is the most powerful and Scheffé's the most conservative. For further discussion on a priori and a posteriori contrasts, refer to the literature.[20]

Our discussion so far has assumed that each subject is exposed to only one treatment or experimental condition. Sometimes subjects are exposed to more than one experimental condition, in which case repeated measures ANOVA should be used.

Repeated measures ANOVA

In marketing research, there are often large differences in the background and individual characteristics of respondents. If this source of variability can be separated from treatment effects (effects of the independent variable) and experimental error, then the sensitivity of the experiment can be enhanced. One way of controlling the differences between subjects is by observing each subject under each experimental condition (see Table 19.7).

In this sense, each subject serves as its own control. For example, in a survey attempting to determine differences in evaluations of various airlines, each respondent evaluates all the major competing airlines. In a study examining the differences among heavy users, medium users, light users and non-users of a brand, each respondent provides ratings on the relative importance of each attribute. Because repeated measurements are obtained from each respondent, this design is referred to as within-subjects design or **repeated measures ANOVA**. This differs from the assumption we made in our earlier discussion that each respondent is exposed to only one treatment condition, also referred to as between-subjects design.[21] Repeated measures ANOVA may be thought of as an extension of the paired samples t test to the case of more than two related samples.

Repeated measures ANOVA
An ANOVA technique used when respondents are exposed to more than one treatment condition and repeated measurements are obtained.

In the case of a single factor with repeated measures, the total variation, with $n(c-1)$ degrees of freedom, may be split into between-people variation and within-people variation:

$$SS_{total} = SS_{between\ people} + SS_{within\ people}$$

The between-people variation, which is related to the differences between the means of people, has $n-1$ degrees of freedom. The within-people variation has $n(c-1)$ degrees of freedom. The within-people variation may, in turn, be divided into two different sources of variation. One source is related to the differences between treatment means, and the second consists of residual or error variation. The degrees of freedom corresponding to the treatment variation are $c-1$ and those corresponding to residual variation are $(c-1)(n-1)$. Thus,

Table 19.7 Decomposition of the total variation: repeated measures ANOVA

Independent variable X

Subject no.	Categories			Total sample
	X_1	X_2	$X_3 \ldots X_c$	
1	Y_{11}	Y_{12}	$Y_{13} \ldots Y_{1c}$	Y_1
2	Y_{21}	Y_{22}	$Y_{23} \ldots Y_{2c}$	Y_2
	⋮			⋮
n	Y_{n1}	Y_{n2}	$Y_{n3} \ldots Y_{nc}$	Y_N
Category mean	$\bar{Y}_1$	$\bar{Y}_2$	$\bar{Y}_3 \ldots \bar{Y}_c$	$\bar{Y}$

Between-people variation = $SS_{between\ people}$

Total variation = SS_y

Within-people variation = $SS_{within\ people}$

$$SS_{within\ people} = SS_x + SS_{error}$$

A test of the null hypothesis of equal means may now be constructed in the usual way:

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

$$= \frac{MS_x}{MS_{error}}$$

So far we have assumed that the dependent variable is measured on an interval or ratio scale. If the dependent variable is non-metric, however, a different procedure should be used.

Non-metric ANOVA

Non-metric ANOVA
An ANOVA technique for examining the difference in the central tendencies of more than two groups when the dependent variable is measured on an ordinal scale.

***k*-sample median test**
A non-parametric test used to examine differences among more than two groups when the dependent variable is measured on an ordinal scale.

Kruskal–Wallis one-way ANOVA
A non-metric ANOVA test that uses the rank value of each case, not merely its location relative to the median.

Non-metric ANOVA examines the difference in the central tendencies of more than two groups when the dependent variable is measured on an ordinal scale. One such procedure is the ***k*-sample median test**. As its name implies, this is an extension of the median test for two groups, which was considered in Chapter 18. The null hypothesis is that the medians of the k populations are equal. The test involves the computation of a common median over the k samples. Then, a $2 \times k$ table of cell counts based on cases above or below the common median is generated. A chi-square statistic is computed. The significance of the chi-square implies a rejection of the null hypothesis.

A more powerful test is the **Kruskal–Wallis one-way ANOVA**. This is an extension of the Mann–Whitney test (Chapter 18). This test also examines the difference in medians. The null hypothesis is the same as in the k-sample median test, but the testing procedure is different. All cases from the k groups are ordered in a single ranking. If the k populations are the same, the groups should be similar in terms of ranks within each group. The rank sum is calculated for each group. From these, the Kruskal–Wallis H statistic, which has a chi-square distribution, is computed.

The Kruskal–Wallis test is more powerful than the k-sample median test because it uses the rank value of each case, not merely its location relative to the median. If there are a large number of tied rankings in the data, however, the k-sample median test may be a better choice.

Non-metric ANOVA is not popular in marketing research. Another procedure that is also only rarely used is multivariate ANOVA.

Multivariate ANOVA

Multivariate ANOVA (MANOVA)
An ANOVA technique using two or more metric dependent variables.

Multivariate ANOVA (MANOVA) is similar to ANOVA except that instead of one metric-dependent variable we have two or more. The objective is the same, since MANOVA is also concerned with examining differences between groups. Although ANOVA examines group differences on a single dependent variable, MANOVA examines group differences across multiple dependent variables simultaneously. In ANOVA, the null hypothesis is that the means of the dependent variable are equal across the groups. In MANOVA, the null hypothesis is that the vector of the means of multiple dependent variables is equal across groups. MANOVA is appropriate when there are two or more dependent variables that are correlated. If there are multiple dependent variables that are uncorrelated or orthogonal, ANOVA on each of the dependent variables is more appropriate than MANOVA.[22]

As an example, suppose that four groups, each consisting of 100 randomly selected individuals, were exposed to four different commercials about the 'Series 1' BMW. After seeing the commercial, each individual provided ratings on preference for the 'Series 1', preference for BMW, and preference for the commercial itself. Because these three preference variables are correlated, MANOVA should be conducted to determine which commercial is the most effective (produced the highest preference across the three preference variables). The following example illustrates the application of ANOVA and MANOVA in international marketing research.

Example

The commonality of unethical research practices worldwide[23]

A study examined marketing professionals' perceptions of how common unethical practices in marketing research were across different countries, i.e. 'the commonality of unethical marketing research practices'. A sample of marketing professionals was drawn from Australia, the UK, Canada and the USA.

Respondents' evaluations were analysed using MANOVA and ANOVA techniques. The predictor variable was the 'country of respondent' and 15 evaluations of 'commonality' served as the criterion variables. The F values from the ANOVA analyses indicated that only 2 of the 15 commonality evaluations achieved significance ($p < 0.05$ or better). Further, the MANOVA F value was not statistically significant, implying the lack of overall differences in commonality evaluations across respondents of the four countries. It was concluded that marketing professionals in the four countries demonstrate similar perceptions of the commonality of unethical research practices. This finding is not surprising, given other research evidence that organisations in the four countries reflect similar corporate cultures.

Internet and computer applications

The computer packages SPSS and SAS have programs for conducting ANOVA and ANCOVA. In addition to the basic analysis that we have considered, these programs can also perform more complex analysis. Minitab and Excel also offer some programs. Given the importance of ANOVA and ANCOVA, several programs are available in each package.

SPSS

One-way ANOVA can be efficiently performed using the program ONEWAY. This program also allows the user to test a priori and a posteriori contrasts. For performing n-way ANOVA, the program ANOVA can be used. Although covariates can be specified, the program ANOVA does not perform a full ANCOVA. For comprehensive ANOVA or ANCOVA, including repeated measures and multiple dependent measures, the MANOVA procedure is recommended. For non-metric ANOVA, including the *k*-sample median test and Kruskal–Wallis one-way ANOVA, the program NPAR TESTS should be used.

SAS

The main program for performing ANOVA in the case of a balanced design is the program ANOVA. This program can handle data from a wide variety of experimental designs, including MANOVA and repeated measures. Both a priori and a posteriori contrasts can be tested. For unbalanced designs, the more general GLM procedure can be used. This program performs ANOVA, ANCOVA, repeated measures ANOVA and MANOVA. It also allows the testing of a priori and a posteriori contrasts. Whereas GLM can also be used for analysing balanced designs, it is not as efficient as ANOVA for such models. The VARCOMP procedure computes variance components. For non-metric ANOVA, the NPAR1WAY procedure can be used. For constructing designs and randomised plans, the PLAN procedure can be used.

Minitab

ANOVA and ANCOVA can be accessed from the Stats>ANOVA function. This function performs one-way ANOVA, one-way unstacked ANOVA, two-way ANOVA, analysis of means, balanced ANOVA, ANCOVA, general linear model, main-effects plot, interactions plot and residual plots. In order to compute the mean and standard deviation, the CROSSTAB function must be used. To obtain *F* and *p* values, use the balanced ANOVA.

Excel

Both a one-way ANOVA and two-way ANOVA can be performed under the Tools>DATA ANALYSIS function. The two-way ANOVA has the features of a two-factor with replication and a two-factor without replication. The two-factor with replication includes more than one sample for each group of data, while the two-factor without replication does not include more than one sampling per group.

Summary

In ANOVA and ANCOVA, the dependent variable is metric and the independent variables are all categorical and metric variables. One-way ANOVA involves a single independent categorical variable. Interest lies in testing the null hypothesis that the category means are equal in the population. The total variation in the dependent variable may be decomposed into two components: variation related to the independent variable and variation related to error. The variation is measured in terms of the sum of squares corrected for the mean (*SS*). The mean square is obtained by dividing the *SS* by the corresponding degrees of freedom (*df*). The null hypothesis of equal means is tested by an *F* statistic, which is the ratio of the mean square related to the independent variable to the mean square related to error.

N-way ANOVA involves the simultaneous examination of two or more categorical independent variables. A major advantage is that the interactions between the independent variables can be examined. The significance of the overall effect, interaction terms and the main effects of individual factors are examined by appropriate F tests. It is meaningful to the significance of main effects only if the corresponding interaction terms are not significant.

ANCOVA includes at least one categorical independent variable and at least one interval or metric-independent variable. The metric-independent variable, or covariate, is commonly used to remove extraneous variation from the dependent variable.

When ANOVA is conducted on two or more factors, interactions can arise. An interaction occurs when the effect of an independent variable on a dependent variable is different for different categories or levels of another independent variable. If the interaction is significant, it may be ordinal or disordinal. Disordinal interaction may be of a non-crossover or crossover type. In balanced designs, the relative importance of factors in explaining the variation in the dependent variable is measured by omega squared (ω^2). Multiple comparisons in the form of a priori or a posteriori contrasts can be used for examining differences among specific means.

In repeated measures ANOVA, observations on each subject are obtained under each treatment condition. This design is useful for controlling for the differences in subjects that exist prior to the experiment. Non-metric ANOVA involves examining the differences in the central tendencies of two or more groups when the dependent variable is measured on an ordinal scale. MANOVA involves two or more metric-dependent variables.

Questions

1 Discuss the similarities and differences between analysis of variance and analysis of covariance.

2 What is the relationship between analysis of variance and the t test?

3 What is total variation? How is it decomposed in a one-way analysis of variance?

4 What is the null hypothesis in one-way ANOVA? What basic statistic is used to test the null hypothesis in one-way ANOVA? How is this statistic computed?

5 How does n-way analysis of variance differ from the one-way procedure?

6 How is the total variation decomposed in n-way analysis of variance?

7 What is the most common use of the covariate in ANCOVA?

8 What is the difference between ordinal and disordinal interaction?

9 How is the relative importance of factors measured in a balanced design?

10 What is an a priori contrast?

11 What is the most powerful test for making a posteriori contrasts? Which test is the most conservative?

12 What is meant by repeated measures ANOVA? Describe the decomposition of variation in repeated measures ANOVA.

13 What are the differences between metric and non-metric analyses of variance?

14 Describe two tests used for examining differences in central tendencies in non-metric ANOVA.

15 What is multivariate analysis of variance? When is it appropriate?

Exercises

1 A marketing researcher wants to test the hypothesis that there is no difference in the importance attached to shopping by consumers living in Belgium, France, Germany and the Netherlands. A study is conducted and analysis of variance is used to analyse the data. The results obtained are presented in the following table:

Source	*df*	*Sum of squares*	*Mean squares*	*F ratio*	*F probability*
Between groups	3	70.212	23.404	1.12	0.3
Within groups	996	20812.416	20.896		

a Is there sufficient evidence to reject the null hypothesis?
b What conclusion can be drawn from the table?
c If the average importance was computed for each group, would you expect the sample means to be similar or different?
d What was the total sample size in this study?

2 In a pilot study examining the effectiveness of three commercials (A, B, and C), 10 consumers were assigned to view each commercial and rate it on a nine-point Likert scale. The data obtained from the 30 respondents are shown in the following table:

Commercial A	*Commercial B*	*Commercial C*
4	7	8
5	4	7
3	6	7
4	5	6
3	4	8
4	6	7
4	5	8
3	5	8
5	4	5
5	4	6

a Calculate the category means and the grand mean.
b Calculate SS_y into SS_x and SS_{error}.
c Calculate η^2.
d Calculate the value of F.
e Are the three commercials equally effective?

3 An experiment tested the effects of package design and shelf display on the likelihood of buying a breakfast cereal. Package design and shelf display were varied at two levels each, resulting in a 2 × 2 design. Purchase likelihood was measured on a seven-point scale. The results are partially described in the following table:

Source of variation	*Sum of squares*	*df*	*Mean square*	*F*	*Sig. of F*	ω^2
Package design	68.76	1				
Shelf display	320.19	1				
Two-way interaction	55.05	1				
Residual error	176.00	40				

a Complete the table by calculating the mean square, F, significance of F, and ω^2 values.
b How should the main effects be interpreted?

4 In an experiment designed to measure the effect of gender and frequency of travel on preference for long-haul holidays, a 2 (gender) × 3 (frequency of travel) between-subjects design was adopted. Five respondents were assigned to each cell for a total sample size of 30. Preference for long-haul holidays was measured on a nine-point scale (1 = no preference, 9= strong preference). Gender was coded as male = 1 and female = 2. Frequency of travel was coded as light = 1, medium = 2, and heavy = 3. The data obtained are shown in the table.

Number	*Gender*	*Travel group*	*Preference*
1	1	1	2
2	1	1	3
3	1	1	4
4	1	1	4
5	1	1	2
6	1	2	4
7	1	2	5
8	1	2	5
9	1	2	3
10	1	2	3
11	1	3	8
12	1	3	9
13	1	3	8
14	1	3	7
15	1	3	7
16	2	1	6
17	2	1	7
18	2	1	6
19	2	1	5
20	2	1	7
21	2	2	3
22	2	2	4
23	2	2	5
24	2	2	4
25	2	2	5
26	2	3	6
27	2	3	6
28	2	3	6
29	2	3	7
30	2	3	8

Using software of your choice, perform the following analysis.

a Do males and females differ in their preference for long-haul travel?

b Do the light, medium and heavy travellers differ in their preference for long-haul travel?

c Conduct a 2 × 3 analysis of variance with preference for long-haul travel as the dependent variable and gender and travel frequency as the independent variables or factors. Interpret the results.

5 In a small group discuss the following issues: 'Which procedure is more useful in marketing research – analysis of variance or analysis of covariance?' and 'There are few marketing research applications where *t* tests are used; the complexity of marketing phenomena mean that analysis of variance or analysis of covariance are much more commonplace.'

Notes

1 Mangaloglu, S.B.M., 'Tourism destination images of Turkey, Egypt, Greece and Italy as perceived by US based tour operators and travel agents', *Tourism Management* 22 (1) (February 2001), 1–9.

2 For applications of ANOVA, see Sengupta, J., and Gorn, G.J., 'Absence makes the mind grow sharper: effects of element omission on subsequent recall', *Journal of Marketing Research* 39 (2) (May 2002), 186–201; Varki, S. and Rust, R.T., 'Satisfaction is relative', *Marketing Research: A Magazine of Management and Applications* 9 (2) (Summer 1997), 14–19; Deshpande, R. and Stayman, D.M., 'A tale of two cities: distinctiveness theory and advertising effectiveness', *Journal of Marketing Research* 31 (February 1994), 57–64.

3 Janky, D.G., 'Sometimes pooling for analysis of variance hypothesis tests: a review and study of a split level model', *American Statistician* 54 (4) (November 2000), 269–279; Driscoll, W.C., 'Robustness of the ANOVA and Tukey-Kramer statistical tests', *Computers and Industrial Engineering* 31 (1,2) (October 1996), 265–268; Burdick, R.K., 'Statement of hypotheses in the analysis of variance', *Journal of Marketing Research* (August 1983), 320–324.

4 The *F* test is a generalised form of the *t* test. If a random variable is *t* distributed with *N* degrees of freedom, then t^2 is *F* distributed with 1 and *N* degrees of freedom. Where there are two factor levels or treatments, ANOVA is equivalent to the two-sided *t* test.

5 Although computations for the fixed-effects and random-effects models are similar, interpretations of results differ. A comparison of these approaches is found in Turner, J.R. and Thayer, J., *Introduction to Analysis of Variance: Design, Analysis and Interpretation*, (Thousand Oaks, CA: Sage, 2001); Erez, A., Bloom, M.C. and Wells, M.T., 'Using random rather than fixed effects models in meta-analysis: implications for situational specificity and validity generalization', *Personnel Psychology* 49 (2) (Summer 1996), 275–306; Neter, J.W., *Applied Linear Statistical Models*, 4th edn (Burr Ridge, IL: Irwin, 1996).

6 We consider only the full factorial designs, which incorporate all possible combinations of factor levels. For example, see Menon, G., 'Are the parts better than the whole? The effects of decompositional questions on judgments of frequent behaviors', *Journal of Marketing Research* 34 (August 1997), 335–346.

7 Miyazaki, A.D., 'Consumer perceptions of privacy and security risks for online shopping', *Journal of Consumer Affairs* 35 (1) (Summer 2001), 27; Burnett, R., 'As internet sales rise, so do shoppers' complaints', *Knight Ridder Tribune Business News* (20 December 2001), 1.

8 Sengupta, J., and Gorn, G. 'Absence makes the mind grow sharper: effects of element omission on subsequent recall', *Journal of Marketing Research* 39 (2) (May 2002), 186–201; Jaccard, J., *Interaction Effects in Factorial Analysis of Variance* (Thousand Oaks, CA: Sage, 1997); Mayers, J.L., *Fundamentals of Experimental Design*, 3rd edn (Boston, MA: Allyn & Bacon, 1979).

9 Tacq, J., *Multivariate Analysis Techniques in Social Science Research* (Thousand Oaks, CA: Sage, 1997); Daniel, W.W. and Terrell, I.C., *Business Statistics*, 7th edn (Boston, MA: Houghton Muffin, 1995).

10 Nishisato, S., *Measurement and Multivariate Analysis* (New York: Springer-Verlag, New York, 2002).

11 Desai, K.K., 'The effects of ingredient branding strategies on host brand extendibility', *Journal of Marketing* 66 (1) (January 2002), 73–93; Peterson, R.A. and Jolibert, A.J.P., 'A meta-analysis of country-of-origin effects', *Journal of International Business Studies* 26 (4) (Fourth Quarter 1995), 883–900; Chao, P., 'The impact of country affiliation on the credibility of product attribute claims', *Journal of Advertising Research* (April–May 1989), 35–41.

12 Although this is the most common way in which analysis of covariance is performed, other situations are also possible. For example, covariate and factor effects may be of equal interest, or the set of covariates may be of major concern. For applications, see Pham, M.T. and Muthukrishnan, A.V., 'Search and alignment in judgment revision: implications for brand positioning', *Journal of Marketing Research* 39 (1) (February 2002), 18–30; Lane Keller, K. and Aaker, D.A., 'The effects of sequential introduction of brand extensions', *Journal of Marketing Research* 29 (February 1992), 35–50.

13 For more detailed discussions, see Turner, J.R. and Thayer, J., *Introduction to Analysis of Variance: Design, Analysis and Interpretation*, (Thousand Oaks, CA: Sage, 2001); Glantz, S.A. and Slinker, B.K., *Primer of Applied Regression and Analysis of Variance* (Blacklick, OH: McGraw-Hill, 2000); Neter, J.W., *Applied Linear Statistical Models*, 4th edn (Burr Ridge, IL: Irwin, 1996); Wildt, A.R. and Ahtola, O.T., *Analysis of Covariance* (Beverly Hills, CA: Sage, 1978).

14 See Zhang, S. and Schmitt, B.H., 'Creating local brands in multilingual international markets', *Journal of Marketing Research* 38 (3) (August 2001), 313–325; Umesh, U.N., Peterson, R.A., McCann-Nelson, M. and Vaidyanathan, R., 'Type IV error in marketing research: the investigation of ANOVA interactions', *Journal of the Academy of Marketing Science* 24 (1) (Winter 1966), 17–26; Ross W.T. Jr., and Creyer, E.H., 'Interpreting interactions: raw means or residual means', *Journal of Consumer Research* 20 (2) (September 1993), 330–338; Leigh, J.H. and Kinnear, T.C., 'On interaction classification', *Educational and Psychological Measurement* 40 (Winter 1980), 841–843.

15 For an examination of interactions using an ANOVA framework, see Jaccard, J., *Interaction Effects in Factorial Analysis of Variance* (Thousand Oaks, CA: Sage, 1997); Wansink, B., 'Advertising's impact on category substitution', *Journal of Marketing Research* 31 (November 1994), 505–515; Peracchio, L.A. and Meyers-Levy, J., 'How ambiguous cropped objects in ad photos can affect product evaluations', *Journal of Consumer Research* 21 (June 1994), 190–204.

16 Verma, R. and Goodale, J.C., 'Statistical power in operations management', *Journal of Operations Management* 13 (2) (August 1995), 139–152; Wyner, G.A., 'The significance of marketing research', *Marketing Research: A Magazine of Management and Applications* 5 (1) (Winter 1993), 43–45; Sawyer, A. and Peter, J.P., 'The significance of statistical significance tests in marketing research', *Journal of Marketing Research* 20 (May 1983), 125; Beltramini, R.F., 'A meta-analysis of effect sizes in consumer behavior experiments', *Journal of Consumer Research* 12 (June 1985), 97–103.

17 This formula does not hold if repeat measurements are made on the dependent variable. See Fern, E.F. and Monroe, K.B., 'Effect size estimates: Issues and problems in interpretation', *Journal of Consumer Research* 23 (2) (September 1996), 89–105; Dodd, D.H. and Schultz, Jr., R.F., 'Computational

procedures for estimating magnitude of effect for some analysis of variance designs', *Psychological Bulletin* (June 1973), 391–395.

18 The ω^2 formula is attributed to Hays. See Hays, W.L., *Statistics for Psychologists* (New York: Holt, Rinehart & Winston, 1963). For an application, see Ratneshwar, S. and Chaiken, S., 'Comprehension's role in persuasion: the case of its moderating effect on the persuasive impact of source cues', *Journal of Consumer Research* 18 (June 1991), 52–62. For an alternative approach, see also Finn, A. and Kayande, U., 'Reliability assessment and optimisation of marketing measurement', *Journal of Marketing Research* 34 (February 1997), 262–275.

19 Johnson, R.A. and Wichern, D.W., *Applied Multivariate Statistical Analysis* (Paramus, NJ: Prentice Hall, 2001); Fern, E.F. and Monroe, K.B., 'Effect-size estimates: issues and problems in interpretation', *Journal of Consumer Research* 23 (2) (September 1996), 89–105; Cohen, J., *Statistical Power Analysis for the Behavioral Sciences* (New York: Academic Press, 1969).

20 Turner, J.R. and Thayer, J., *Introduction to Analysis of Variance: Design, Analysis and Interpretation* (Thousand Oaks, CA: Sage, 2001); Neter, J.W., *Applied Linear Statistical Models*, 4th edn (Burr Ridge, IL: Irwin, 1996); Winer, B.J., Brown, D.R. and Michels, K.M., *Statistical Principles in Experimental Design*, 3rd edn (New York: McGraw-Hill, 1991).

21 It is possible to combine between-subjects and within-subjects factors in a single design. See, for example, Ahluwalia, R., Unnava, H.R. and Burnkrant, R.E., 'The moderating role of commitment on the spillover effect of marketing communications', *Journal of Marketing Research* 38 (4) (November 2001), 458–470; Mount, M.K., Sytsma, M.A., Hazucha, J.F. and Holt, K.E., 'Rater-ratee effects in developmental performance ratings of managers', *Personnel Psychology* 50 (1) (Spring 1997), 51–69; Broniarczyk, S.M. and Alba, J.W., 'The importance of the brand in brand extension', *Journal of Marketing Research* 31 (May 1994), 214–228; Krishna, A., 'The effect of deal knowledge on consumer purchase behavior', *Journal of Marketing Research* 31 (February 1994), 76–91.

22 See Roehm, M.L. Pullins, E.B. and Roehm, H.A., Jr. 'Designing loyalty-building programs for packaged goods brands', *Journal of Marketing* 39 (2) (May 2002), 202–213; Bray, J.H. and Maxwell, S.E., Multivariate Analysis of Variance (Beverly Hills, CA: Sage, 1985). For an application of MANOVA, see Piercy, N.F., 'Sales manager behaviour control strategy and its consequences: the impact of gender differences', *Journal of Personal Selling & Sales Management* 21(1) (Winter 2001), 39–49.

23 Kimmel, A.J. and Smith, N.C., 'Deception in marketing research: ethical, methodological and disciplinary implications', *Psychology and Marketing* 18 (7) (July 2001), 663–689; Abramson, N.R., Keating, R.J. and Lane, H.W., 'Cross-national cognitive process differences: a comparison of Canadian, American and Japanese managers', *Management International Review* 36 (2) (Second Quarter 1996), 123–147; Akaah, I.P., 'A cross-national analysis of the perceived commonality of unethical practices in marketing research', in Lazer, L., Shaw, E. and Wee, C.-H. (eds), *World Marketing Congress*, International Conference Series, Vol. 4 (Boca Raton, FL: Academy of Marketing Science, 1989), 2–9.

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.