TWELVE
T-TESTS AND ANOVA

The purpose of this chapter is to extend the concept of statistical association learned in the two preceding chapters to the situation where the independent variable is qualitative, and the dependent variable is quantitative. As in the previous cases, two processes are involved: that of establishing that there is an association at the level of the sample, and that of establishing whether it is statistically significant, which means that there is an association at the level of the whole population. In the particular case of an independent qualitative variable and a dependent quantitative variable, the meaning of a statistical association is that the mean scores of the quantitative variable (the dependent variable) differ across the categories of the qualitative variable (the independent variable). If the independent variable has just two categories, the technique used is an extension of that of the *t*-test, seen briefly in Chapter 9. If it has more than two categories, a technique based on the analysis of the variance of the various groups, known as ANOVA, is used instead. ANOVA can also be used when there is more than one independent variable. The chapter will end with general remarks about statistical association and relationships between variables.

After studying this chapter, you should know:

- **how to compare the means of various subgroups on a variable, and how to illustrate the comparison with charts and diagrams;**
- **how to conduct a *t*-test, how to determine the conditions under which it can be used, and how to produce and interpret the relevant SPSS outputs;**
- **how to conduct an ANOVA test (with either one or two factors), how to determine the conditions under which it can be used, and how to produce and interpret the relevant SPSS outputs;**
- **how to measure statistical association when the variables are measured at the ordinal level;**
- **and finally, how to distinguish between the notions of explanation, causal factor, and spurious relationship.**

Introduction

Suppose that we want to analyze the statistical relationship between sex (independent variable, qualitative) and income (dependent variable, quantitative), and

to determine whether men tend to earn more, on the average, than women, or not. Several options are open to us. The simplest is to compute the average income separately for men and women.

Example 12.1

The average income for a sample of 1400 people, consisting of 800 men and 600 women, is $19,400 a year. The average income for women and for men separately is $23,400 for men and $17,300 for women. This would appear to mean that there is a large difference between the incomes of men and women. The average income of men is (23,400 − 17,300) / 17,300 × 100 = 35.2% higher than that of women.

This means that, for the sample under consideration, there is a statistical association between the variables *sex* and *income*; we are not yet generalizing to the whole population. However, the preceding statement does not mean that the variable *sex* is the *cause* of the difference in income. All we can say for the time being is that women make less money than men do. The interpretation of that difference is another matter. It could be due to discrimination (direct or systemic), it could be due to some other intervening variable (if, for instance, the women of this sample are generally younger than the men, and therefore have less working experience) or some other cause.

We will now illustrate the kind of analysis we want to do with a slightly more complex example. Consider Table 12.1, which comes from the *GSS93 subset.sav* data file. This shows the average age at marriage for a number of subgroups: for men, for women, for those who have a college degree and for those who do not, and for all combinations of these two variables. We can read from the table the following information:

Table 12.1 SPSS output for the Compare Means command

Report				
Age When First Married				
Respondent's Sex	College Degree	Mean	N	Std. Deviation
Male	No College degree	23.56	356	4.718
	College degree	25.75	136	4.909
	Total	**24.16**	**492**	**4.867**
Female	No College degree	21.20	580	4.539
	College degree	24.74	129	5.574
	Total	**21.84**	**709**	**4.932**
Total	No College degree	22.10	936	4.745
	College degree	25.26	265	5.258
	Total	**22.79**	**1201**	**5.034**

- Men *in this sample* tend to get married later than women. The difference is: 24.16 − 21.84 = 2.32 years, or roughly 2 years and 4 months.

- Those who hold a college degree tend to get married later than those who do not. The difference is slightly more than 3 years.

- For men, getting a college degree seems to delay the average age of marriage by a little over 2 years; this is found by subtracting the average age of marriage of men without a degree from the corresponding statistic for men with a degree; for women, this delay is more than 3.5 years.

The preceding statements are all formulated in the form of statistical associations based on the sample data. The first one says that, for this sample, the variable *sex* has an effect on the variable *age at first marriage*. The second one says that the variable *college degree* has an effect on *age at first marriage*. The third one is more complicated, as it involves the interaction between *sex* and *college degree*. Indeed, the statement highlights the fact that getting a college degree has a different effect on men and women. This is exactly the meaning of the **interaction** between two variables: the effect of one of the independent variables on the dependent variable differs according to the category of a second independent variable.

To summarize, the notion of **statistical association** between a qualitative independent variable and a quantitative dependent variable is simple: it is a matter of establishing whether the average of the dependent variable is different for the various categories of the independent variable.

A more interesting question is whether this relationship is considered to be statistically significant. You should by now be used to this term: we discussed it at length in the two previous chapters. In plain language, it simply means that the difference in means observed on the sample is the reflection of a similar difference for the whole population. In a more technical language, it means that the observed difference would be quite unlikely (less than a 5% chance for it to occur) if it was not true for the whole population. In this case, it is safer to accept the claim (with a known risk of being wrong) that there is a difference in the average value of the dependent variable between the various groups of the population, corresponding to categories of the independent variable (men and women, graduates and non-graduates, etc.). The risk of being wrong is known as the significance level, and it is given in the SPSS output.

The question now is: how do we assess such a claim? We have seen in Chapter 9 how to do it: when only one independent variable is involved, we can perform a *t*-test to compare any two categories. This technique tests whether the observed difference between the means of two groups could be due to the randomness of the sample, or whether this difference is likely to be found in the population. When we studied it in Chapter 9, we did not formulate it in terms of statistical association: we were simply testing whether two means differ. Now, we interpret this difference as meaning that there is a statistical association between the independent variable (which defines the various categories) and the dependent variable (which is what we measure).

The *t*-test as a test of statistical association

In Chapter 9, you have seen how to conduct a *t*-test to test the value of a mean or the difference between two means. You may want to review the tutorial of that chapter to recall how the commands are given.

But in this chapter we want to interpret the *t*-test differently, as a measure of association between a qualitative and a quantitative variable. The very meaning of a statistically significant association between a qualitative independent variable and a quantitative dependent variable is precisely that the *population* mean scores on the dependent variable vary across the categories of the independent variable. The *t*-test is a test of hypothesis, which can be formulated as follows:

H_0: There is no statistical association between the independent variable and the dependent variable.

H_1: There is a statistical association between the independent variable and the dependent variable.

A second formulation is:

H_0: The *population* mean scores on the dependent variable are identical across the categories of the independent variable.

H_1: The *population* mean scores on the dependent variable differ across the categories of the independent variable.

There is a third possible formulation. Let us denote by μ_1 and μ_2 the means of the dependent variable on the two groups defined by the independent variable. The formulation becomes:

H_0: $\mu_1 = \mu_2$, H_1: $\mu_1 \neq \mu_2$

SPSS will compute the means of the two groups on the sample, and will evaluate whether the observed difference is likely to be obtained under H_0. It will give you the probability p of obtaining a difference as big, or bigger, in the *Sig. (2-tailed)* column of the *t*-test output. But the computation is done in two different ways, depending on whether the variances between the two groups of the independent variable are considered to be equal for the population as a whole. Since we do not have that information, we must first conduct a hypothesis test on the equality of the two variances, called **Levene's test for equality of variance**.

In fact the *t*-test requires three conditions to be valid:

1. that the dependent variable is normally distributed in the various groups defined by the independent variable;

2. that the population variances are equal across the groups;

3. that the samples are chosen randomly and that the scores are independent of each other.

This is why the Levene test is important. If the variances are not equal, the levels of significance are not as reliable. The same conditions are required for the ANOVA procedure, and again, the levels of significance are not reliable if these conditions are violated, but SPSS will give significance levels calculated more conservatively.

To illustrate the method, let us work on the file *GSS93 subset.sav*, and analyze whether there is a *statistically significant* difference between men and women when it comes to age at first marriage. If there is, we will say that there is a statistical association between the variables *Sex* and *Age when first married* at the level of the population. SPSS will first give you a table titled *Group statistics*, which tells you that the average age at first marriage is 24.16 years for men and 21.84 years for women. So you can see that there is a relatively important difference. To determine whether it is statistically significant, you look at the table titled Independent Samples Tests (we will see in the tutorial how to produce these tables). This is reproduced as Table 12.2 (we have modified the column widths to fit the page).

Table 12.2 SPSS output for Independent-samples T-Test command

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
| | | | | | | | | | 95% Confidence Interval of the Difference | |
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Age When First Married	Equal variances assumed	.342	.559	8.066	1200	.000	2.320	.288	1.756	2.885
	Equal variances not assumed			8.085	1064.662	.000	2.320	.287	1.757	2.883

This table gives the result of three tests:

- Levene's test of equality of variances,
- a *t*-test of equality of means if the equality of variances is assumed,
- and a *t*-test of equality of means if the equality of variances is not assumed.

As explained above, Levene's test of equality of variances (two first numerical columns) tells you whether it is safe to assume that the variances of the variable *Age when first married* should be assumed to be equal for men and women. It tells

you, therefore, which of the two *t*-tests you should consider: the one with the equality of variances assumed (first row of Table 12.2) or the one with the equality of variances not assumed (second row of Table 12.2). We see here that the significance level of the Levene's test is 0.559. This means that if we assumed that the variances were different, we would run a 55.9% chance of being wrong. So it is safer to assume that the variances for men and women are equal, and consider, as a consequence, the first row of the table, not the second. Remember: the usual cut-off value for the significance test is 0.05. When the level of significance is less than that, we accept the alternative hypothesis to the effect that there is a difference in the statistics that are measured. When the level of significance is more than 0.05, we assume that there is no difference. In this particular case, the variances of the two groups can be considered to be equal. So, we look at the fist line, and we see that the *Sig. (2-tailed)* column gives a significance level of 0.000. This means that there is less than 1 chance in a 1000 that we will be mistaken if we assert that men and women in the population tend to get married for the first time at different ages. The difference is therefore statistically significant, and we can safely conclude that the association between *Sex* and *Age when first married* is statistically significant, that is, it is valid for the population as a whole, not just for the sample at hand.

The conclusion of this discussion is that *t*-tests provide a way to decide whether the difference observed on a sample between the means of two different groups on a quantitative variable is statistically significant, that is, reflects a real difference at the level of the population, which is the very meaning of saying that the association is statistically significant.

The independent variable need not be dichotomous in order to conduct an *independent samples t-test*. It can contain more than two categories, provided you compare two groups at a time. For instance, you want to see whether individuals in different religious groups tend to marry at different ages. You could do that with a *t*-test, but you can only compare two groups at a time. The tutorial tells you how to do that.

This is a limitation, though. We may want to compare more than two groups at a time. We may also want to examine whether a *combination* of independent variables are statistically related to a dependent variable. The technique known as ANOVA permits such analyses, and it is to this that we now turn.

The analysis of variance

The aim of **analysis of variance**, commonly referred to as **ANOVA**, is to establish whether the observed differences between several means are statistically significant. Typically, these means correspond to the categories of a qualitative variable. The very meaning of the existence of an association, at the level of the whole population, between a qualitative independent variable and a quantitative dependent variable is precisely that the differences in the means of the various groups are statistically significant. In that sense, the technique of ANOVA is a generalization of the *t*-test to the case where the independent variable has more than two categories. If the test is statistically significant, another question can be

asked: which categories differ? Could it be that two of the three categories are similar, and that they both differ from the other category? **One-way ANOVA** allows you to answer such questions. The term 'one-way' refers to the fact that there is only one independent variable, with several categories.

But ANOVA also permits another generalization, to the situation where there is more than one independent variable. You may have two or more independent variables, and in this case you can determine whether each has an effect on the dependent variable, as well as determining whether there is an interaction between the independent variables. It is then called **multi-way ANOVA**. We will now illustrate both of these generalizations. Recall, however, the conditions under which ANOVA can be used, and which are roughly similar to the conditions of validity of *t*-tests explained in the previous section.

One-way ANOVA

We will explain here the basic logic of ANOVA and the interpretation of the ANOVA outputs in SPSS, without going into the computational details, which can easily be found in the literature (suggestions are made in the bibliography).

The technique of ANOVA is based on the comparison between the variances of the different groups to be compared and the total variance when the various groups are all lumped together. A statistic is computed on that basis, called the *F-statistic*. Its distribution is known, and it depends on the *degrees of freedom* of the statistic, which depends, in turn, on the number of groups and the number of cases in each group. A hypothesis testing procedure can be formulated as follows:

H_0: There is no statistical association between the independent variable and the dependent variable.

H_1: There is a statistical association between the independent variable and the dependent variable.

Another formulation is:

H_0: The mean scores on the dependent variable are identical across all categories of the independent variable.

H_1: The mean scores on the dependent variable differ across the categories of the independent variable.

There is a third possible formulation. Let us denote by $\mu_1, \mu_2, \mu_3, \ldots$ the population means of the dependent variable on the various groups defined by the independent variable. The formulation becomes:

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

H_1: At least two of the population means are different.

F will be the test statistic, and its distribution is known. The probability of getting a given value of F under H_0 can be calculated from its distribution. We can therefore determine the cut-off value beyond which the values of F will not occur more than 5% of the time. If a given value of F does fall beyond that cut-off value, we consider this so rare that we question the hypothesis H_0 under which it is obtained, and instead retain H_1, which states that there must be a difference between at least two of the group means. This logic results in a simple decision rule, for a non-mathematically oriented user: look at the significance level in the ANOVA table produced by SPSS: if it is less than 0.05, you reject H_0 and you consider that the observed differences between the group means are statistically significant. In this case, you will have to address another issue: which of the groups differ?

To illustrate the procedure, we will look at the *Age when first married* in the *GSS subset.sav* file, and try to see whether individuals belonging to different religious groups tend to marry at different ages. In order to simplify the analysis, we will focus on the three main religious groups in the file, Protestants, Catholics and Jews, excluding others from the analysis. We have also chosen about 50% of the cases, because when the sample is large, associations, even if weak, tend to be significant as a result of the sheer size of the sample. If we just look at the means of the independent variable for these three groups, we get Table 12.3.

Table 12.3 Comparison of the mean *Age at first marriage* for a sample of 546 individuals belonging to three religious groups

Report			
Age When First Married			
Religious Preference	Mean	N	Std. Deviation
Protestant	22.30	394	5.230
Catholic	23.67	139	4.513
Jewish	24.38	13	3.042
Total	22.70	546	5.051

We see here that, for our sample, there is a difference between Protestants, Catholics and Jews, the first group having the lowest average age at first marriage (22.3 years), and the latter having the highest average age (24.38 years). Two questions arise: are these differences statistically significant (i.e. true for the population at large)? If yes, could it be that two of the groups do not differ much, but differ significantly with the third? Both questions will be answered by the ANOVA procedure. The tutorial will explain how to give the ANOVA commands. Here, we will learn how to read the SPSS output (see Table 12.4).

Basically, as in the *t*-test, the first four numerical columns provide the results of some of the calculations needed to compute the *F*-statistic. The probability of obtaining such an *F*-value under the assumption of equality of means is given in the last column, labelled *Sig.*, which stands for the *level of significance*. We see here

Table 12.4 The ANOVA table in a one-way analysis of variance

		Sum of Squares	df	Mean Square	F	Sig.
ANOVA Table						
Age When First Married * Religious Preference	Between Groups (Combined)	587.718	2	293.859	11.778	.000
	Within Groups	26745.393	1072	24.949		
	Total	27333.111	1074			

that the probability is well under the cut-off point of 0.05. It is less than one chance in 1000. This is so remote that we conclude that the null hypothesis can be safely rejected and we accept the alternative hypothesis of inequality of means between the three groups. ANOVA also gives you an **effect size statistic**, η (a Greek letter pronounced 'eta'), which is a measure of the strength of the association between the two variables. We get Table 12.5.

Table 12.5 The eta and eta squared table in a one-way analysis of variance

Measures of Association	Eta	Eta Squared
Age When First Married * Religious Preference	.147	.022

Eta squared (written η^2) measures the amount of variation in the dependent variable that can be attributed to the variation of the independent variable. It is commonly stated that an η^2 of 0.01 is a weak association, a value of 0.06 denotes a medium association, and a value of 0.14 constitutes a strong association.

Now the next question is: which of the groups really differ? Could it be that some are not *significantly* different? To answer this question, a series of **post-hoc tests** are available in SPSS, which are tests conducted *after* a significant ANOVA association has been found, in order to determine which groups differ, when considered two at a time. Several post-hoc tests are available, and they fall into two broad categories, depending on whether the variances of the different groups are equal or not. So, the first thing to do is to conduct a homogeneity-of-variance test. If the variances are considered equal, a set of post-hoc significance tests are available. If the variances are not equal, another set of tests are available (details are given in the tutorial). Let us consider the output of one of these tests, the Tukey HSD test, which is run if the variances are considered equal. Table 12.6 shows the output.

The table indicates the difference between the sample means of any two groups and the statistical significance of that difference, that is, the risk that one takes in asserting that the groups do differ at the level of the whole population. The differences that are statistically significant are flagged. We see here that the difference between Protestants and Catholics, and between Protestants and Jews are indeed

**Table 12.6 Outputs of the post-hoc tests (in this case, only the Tukey HSD
test was conducted)**

Multiple Comparisons

Dependent Variable Age When First Married Tukey HSD

(I) Religious Preference	(J) Religious Preference	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Protestant	Catholic	−1.381*	.355	.000	−2.21	−.55
	Jewish	−3.399*	1.057	.004	−5.88	−.92
Catholic	Protestant	1.381*	.355	.000	.55	2.21
	Jewish	−2.018	1.086	.151	−4.57	.53
Jewish	Protestant	3.399*	1.057	.004	.92	5.88
	Catholic	2.018	1.086	.151	−.53	4.57

*The mean difference is significant at the 0.05 level.

significant, but that the difference between Catholics and Jews is not statistically significant. The observed difference between the sample means of these latter groups could well be due to chance. The table also indicates the confidence intervals in which these differences are likely to vary at the level of the whole population. You may have noticed that when the confidence interval includes zero, the difference is not statistically significant. Therefore we can formulate our general conclusion in the following way.

> We conducted an ANOVA analysis on a sample of 546 individuals and found that the relationship between religious preference and age when first married is statistically significant (F = 11.778; p < 0.001). Post-hoc tests indicated that the differences between Protestants and Catholics, and between Protestants and Jews, are significant (p < 0.001 and p = 0.004, respectively), but that the difference between Catholics and Jews is not statistically significant (p = 0.151).

In conclusion, the ANOVA technique allows you to determine whether observed differences between several groups are significant or not, and if they are, it allows you to dig further to determine which differences are significant when the groups are taken two at a time.

Two-way and multi-way ANOVA

The ANOVA procedure allows you to analyze the effect of two or more independent variables on a dependent variable. The procedure is sometimes called two-factor or multi-factor ANOVA. In such a situation, the effects of the independent variables on the dependent variable are complex: each independent variable could have an independent effect on the dependent variable. These are called **main effects**. But the various independent variables could also have a combined effect,

which cannot be reduced to the effect of each factor separately. Such effects are called **interactions**. For instance, if you develop a new teaching method, it could have an effect by itself on the learning process, in comparison with the old method, but it could also have a distinct effect on men and women. In this case, there is an interaction between the sex and teaching method variables. This is what is meant by interaction. Finally, if you had two new teaching methods that you want to compare to an old method, you may want to study how men and women separately react to each of these methods. These are called **simple main effects**: these are the effects of one of the independent variables on the dependent variable, studied separately for the various levels of some other independent variable.

In this section, we will not give a full treatment of the procedure and its complexities. Indeed, the multi-way ANOVA requires many conditions in order to be used.

Multi-way ANOVA is based on the same mathematical approach as the one-way ANOVA. An *F*-statistic is computed by comparing the variances between the various groups with the total variance. Its distribution depends on the degrees of freedom of the statistics, as explained above. On this basis, several hypotheses are tested: one for the overall effect of the model, one for each main effect, one for each interaction, etc. In each case, if the level of significance of the *F*-statistic is less than 0.05, the observed differences between the groups are considered statistically significant.

Here is an illustration. Looking at our example of *Age when first married*, we see in Table 12.7 the mean age of various subgroups defined by a combination of the variables *Sex* and *Religious preference*. Take a minute to examine the various means that appear in this table.

Table 12.7 Age at first marriage broken down by sex and by religious preference

Descriptive Statistics				
Dependent Variable Age When First Married				
Respondent's Sex	Religious Preference	Mean	Std. Deviation	N
Male	Protestant	23.69	4.869	309
	Catholic	24.85	4.833	95
	Jewish	27.07	4.305	14
	Total	24.06	4.889	418
Female	Protestant	21.33	4.892	478
	Catholic	22.95	4.981	170
	Jewish	23.44	2.833	9
	Total	21.78	4.943	657
Total	Protestant	22.25	5.014	787
	Catholic	23.63	5.004	265
	Jewish	25.65	4.141	23
	Total	22.67	5.045	1075

There are evidently differences between the various religious groups, but we cannot tell, at first sight, whether they are significant. Table 12.4 told us that the effect of the variable *Religious preference* was significant overall, and Table 12.6 told us which of the religious groups displayed a statistically significant difference in the average age at first marriage. A *t*-test also indicated that gender differences are statistically significant. But we do not know whether the interaction between religious preference and sex is statistically significant, which means that the religious preference affects differently men and women, when *Age when first married* is considered. Table 12.8 will help us settle the issue.

Table 12.8   Detailed analysis of main effects and interaction effects

Tests of Between-Subjects Effects

Dependent Variable Age When First Married

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1924.762[a]	5	384.952	16.196	.000
Intercept	100572.455	1	100572.455	4231.363	.000
sex	304.466	1	304.466	12.810	.000
relig	483.373	2	241.687	10.168	.000
sex * relig	19.944	2	9.972	.420	.657
Error	25408.349	1069	23.768		
Total	579614.000	1075			
Corrected Total	27333.111	1074			

a. R Squared = .070 (Adjusted R Squared = .066)

We are dealing here with a model that states that the *Age when first married* is affected by the variables *Sex*, *Religious preference*, and their interaction, and the table gives us five *F*-values, together with their significance levels. The first is for the model as a whole, and we see that the model is statistically significant, but we do not know which components are also statistically significant. The intercept refers to a constant that should appear in an equation that links the independent variables and the dependent variable, but we will not deal with this equation here. The effect of each individual independent variable is statistically significant, but we see that the *Sig.* value for the interaction is 0.657, which is not statistically significant. The conclusion can be formulated as follows:

> Each of the variables *sex* and *religious preference* has a statistically significant effect on age when first married ($p < 0.001$ in each case). But the effect of their interaction is not statistically significant ($p = 0.657$). This means that the differences between men and women, across the various religious groups, are not large enough to allow us to conclude that gender differences are linked to religious preference for the population at large.

We see, therefore, that the SPSS output represented in Table 12.8 allows us to establish whether the model as a whole is valid, and whether the main and interaction effects are significant. It is possible to go further, and engage in the analysis of simple main effects, and to contrast the various groups within each sex category. We will leave these issues for a more advanced course on ANOVA.

One last remark is in order. We have purposely used sometimes *Sig.* and sometimes p to denote the significance level. This is so because the significance level is equal to the probability of obtaining such an F-value under the null hypothesis.

Graphical representations of the various kinds of effects

Figure 12.1 illustrates the various kinds of effects, showing the box plots of a quantitative dependent variable with two qualitative independent variables, IV_1 (three categories represented on the X-axis) and IV_2 (two categories that define the clusters). Thus, there are three groups which may differ on IV_1 alone, or on IV_2 alone, or on both without interaction, or on both with interaction. Let us examine them one at a time.

In Figure 12.1(a) there is a noticeable difference between the three categories of IV_1, but there seems to be very little difference within each of the three clusters. The two groups defined by IV_2 within each of the clusters are very much alike. This is an illustration of a situation where the effect of IV_1 would be statistically significant, but neither IV_2 nor their interactions would have a statistically significant effect.

In Figure 12.1(b) we do not see a noticeable difference between the three categories of IV_1, but there is a comparable difference between the two groups within each of the three clusters. The two groups defined by IV_2 within each of the clusters differ noticeably in very comparable ways: the second group within each cluster scores higher than the first group, and the difference between them seems comparable across the three clusters of IV_1. This is an illustration of a situation where the effect of IV_1 is not statistically significant, but the effect of IV_2 is. The interaction between them does not seem to have a statistically significant effect.

In Figure 12.1(c) there is a noticeable difference between the three categories of IV_1, and there seems to be a big difference within each of the three clusters as well. The two groups defined by IV_2 within each of the clusters differ from each other in very comparable ways. This is an illustration of a situation where both the effects of IV_1 and of IV_2 are significant, but their interactions are not.

In Figure 12.1(d) we see a noticeable difference between the three categories of IV_1, and also a difference within each of the three clusters defined by IV_2. Therefore, the effects of both variables are statistically significant. Look carefully: in the first two clusters, the second group scores lower than the first group. In the third cluster, the second group scores higher. This means that the effect of IV_2 varies across the categories of IV_1, sometimes being associated with a lower score for the second group, and at other times being associated with a higher score for the
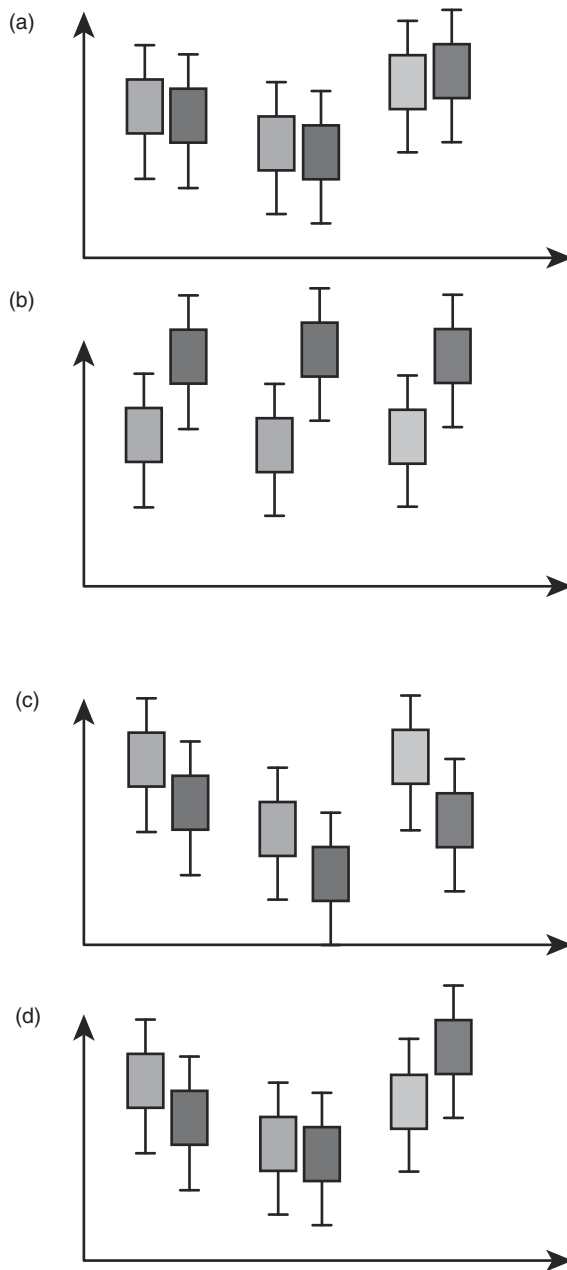
Figure 12.1 Graphical representations of the different types of effects

second group. This is an illustration of a situation where both variables and their interaction have a statistically significant effect.

The exercises and the SPSS tutorial will allow you to master these concepts.

Ordinal variables

There are specific methods for establishing whether or not there is a statistical association between ordinal variables, and whether it is significant. Such methods take into account the ranking of each individual on one of the variables in comparison to his ranking on the other variable. They will not be treated here. Ordinal variables are often treated as quantitative variables and correlations are computed. The results of such computations are sometimes difficult to interpret. A classical presentation of this topic is given by Hildebrand et al. (1977), while Agresti (2010) provides a more advanced and more extensive coverage.

Statistical association as a qualitative relationship

In this chapter and the two previous ones, you have seen how to determine and measure a statistical association, and how to determine whether it is statistically significant, which means that it reflects, with a reasonably large probability, an association at the level of the whole population, not just the sample. You have also seen that the method used to determine whether there is an association and whether it is significant depends on the scale of measurement: nominal, ordinal or quantitative.

We are now faced with a more difficult question: if a statistical association is found, what does it really mean? In qualitative terms, does it indicate that one variable explains the other? Predicts the other? Is a cause of the other? This is the task we will now address: the qualitative interpretation of statistical associations.

To illustrate the set of questions that arise, let us consider again the situation where the two variables were the level of socialization of workers with their peers in a factory and their desire to stay or quit their job. We found that these two variables were associated statistically. But there could be several possible interpretations of that statistical association.

First interpretation. We can interpret the statistical association to mean that a high level of socialization induces people to want to stay in that job. The explanation could be that the job is therefore more enjoyable, and people want to continue working there. In a way, the high level of socialization can be considered to be a cause for staying in the job, and a low level of socialization a reason to leave. So, we are now talking about more than a statistical association: we are talking about a **relationship** between variables. This situation can be represented by the following diagram:

Level of socialization	$\Rightarrow$	Desire to stay or quit

In symbolic terms, if we denote the level of socialization by X, and the desire to quit the job by Y, we could write:

$$X \Rightarrow Y$$

We could go a little further in that interpretation. If, in our theoretical frame-work, we had used the variable *Satisfaction with the job*, denoted by Z, as a general concept, and the level of socialization as one indicator of that concept, we could now conclude the following relationships:

| Level of socialization | ⟹ | Satisfaction with job | ⟹ | Desire to stay or quit |
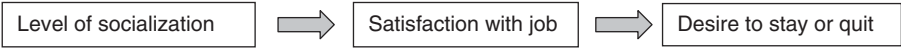
The following pattern illustrates the situation:

$$X \Rightarrow Z \Rightarrow Y$$

In other words, the level of socialization is used as an **explanatory** variable, to explain why people are more inclined to quit their jobs. Notice that this interpretation does not follow from the statistical analysis of the association between the two variables. This is clearly an interpretation, and it is not the only possible interpretation as we will see in what follows.

A more sophisticated interpretation can be thought of. There could be several factors that explain why people are satisfied with the jobs. For instance, one could think of the following factors influencing the variable *Satisfaction with job*: good pay, socialization with peers, good work environment, existence of a gym at the place of work, prospects for professional advancement. This would give the following model:


Second interpretation. We could reverse the preceding interpretation and say that if individuals tend to quit their job (they may perhaps want a better salary, or a more challenging job), they will not invest a lot of energy in socializing with their peers, since they know they are going to quit soon. Here the model is reversed:

$$Y \Rightarrow X$$

In other words, the desire to quit the job is used to explain why people do not socialize a lot with their peers. This interpretation, like the previous one, does not follow automatically from the statistical association between the two variables. The statistical association allows such an interpretation, but it does not prove it. In fact, this could be one among many factors that explain why people do not socialize with their peers. There could definitely be other reasons as well.

Third interpretation. The results of the statistical analysis are consistent with yet another interpretation, which asserts that both the desire to quit and the lack of socialization are the result of a third variable, such as *Desire to get a better salary*. If people think that their present salary is too low, and that they can get a better salary if they find another job, they may plan to quit and also they may decide not to invest too much energy and time in socializing with their peers. The model proposed here for explaining the statistical association is the following:

$$Z \nearrow X \\ \searrow Y$$
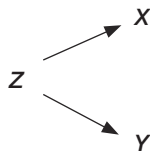
Fourth interpretation. We could consider both variables as indicators of the general concept *Satisfaction with job*. This concept could be measured by several indicators: level of socialization, intention to stay, satisfaction with the salary level, pleasant atmosphere in the office, relationship of support and cooperation with the management, etc. In this interpretation, the key concept is the overall satisfaction with the job. When people are globally satisfied, they are more likely to socialize with their peers, to consider staying in the job for a long time, etc.

We see that statistics does not tell you which interpretation to adopt. In fact, *such interpretations must precede and inform the statistical analysis*. This is what allows you to formulate a hypothesis, which you can then test statistically.

Sometimes the qualitative relationship between two correlated variables is said to be **spurious**. This means that there is no logical link between the two variables, and that the statistical association is misleading. Such statistical association is often due to a third variable, and the conceptual links between each of the two correlated variables and the third one are completely unrelated. A classical example is that of height and salary. It could turn out that there is a statistical association between the height of an individual and his or her salary for a given sample. But if we break down the sample studied into men and women, we find that within each group there is no relationship. What happens is that, on one hand, men tend to be taller than women, and on the other hand, in most societies, the social structure favors men over women and the former end up tending to have higher salaries. The two kinds of associations (sex and

height; gender and salary) follow logics that are totally unrelated to each other, hence our conclusion that the statistical association between height and salary is spurious. However, it is not always clear whether two sets of causal relationships are related or not, and one should be quite careful in interpreting a statistical association as spurious or as meaningful.

A few more points need to be raised here. The first has to do with the definition of the various concepts and their operationalization. A definition is a statement that explains what is meant by the concept, what kinds of phenomena are designated by it. The **operationalization** of the concept is a concrete way of observing and measuring it. This measure must be both valid and reliable, as explained in Chapter 1. Unless you have a good definition of the concepts and you have thoroughly examined their operationalization, seeking advice and experience in the literature, you cannot come up with reasonable ways of studying the relationships between variables.

But there is something else, which is much more important. We see here that in order to start thinking about relationships between variables, we drew a series of models of these relationships. A **model** is an abstract representation of a set of variables, together with a set of links between the variables. The links are qualitative: a variable could be a dimension, or an indicator, of another variable. Or it could be one of its causes. Sometimes, in formulating the model, we are assuming that certain relationships exist between the variables. We are here formulating a theory. All of this *precedes* statistical analysis. And this is precisely the point I would like to emphasize:

> **Any statistical analysis must be preceded by a clear and reasoned formulation of the theoretical framework on which it is based.**

This is a fundamental idea that should never be forgotten. The theoretical framework precedes any statistical analysis. This is why we can perform hypothesis testing: we have good (theoretical) reasons to believe that certain relationships exist, and we have a conceptual pattern of these relationships, and then we test whether, statistically, they are valid.

These comments allow us now to introduce one more idea into the discussion. When we measure a statistical association, we are really interested in knowing whether, conceptually, it is valid. There are always associations between variables in a sample. The real issue is whether they reflect something more general, more abstract, which is valid for the whole population under study. When you study the relationship between a new teaching method in a class and the improvement in efficiency of learning that results from using this tool, you are not interested just in the particular class where you tested it. You want to establish that it is valid in general, that using it somewhere else will produce comparable results. So, *statistical associations are interesting only when they are statistically significant*. This is why some authors consider that there is a statistical association only when the observed links are statistically significant. We opted instead for the position that the notion of a *link* between two variables, as observed on a sample, must be clearly defined,

independently of the fact that they can or cannot be generalized to the whole population.

The discussion above should help us understand better two distinct concepts, the concept of *statistical association* and the concept of *relationship between variables*.

Statistical association is something that can be observed objectively and measured, as we have seen in the examples above. Basically, it means that if you know the score of an individual on a variable X you can make a better guess of his/her score on another variable Y than if you did not know the score on X. The method for establishing whether there is a statistical association or not depends on the level of measurement of the variables, which depends partly on the type of variables.

- For quantitative variables measured by a numerical scale, statistical association is called correlation. Two quantitative variables are correlated when the values of one of them can be predicted with some precision from the values of the other variable. For linear correlation, the points representing the individuals are close to a straight line, called the regression line. If the association is strong, the points are very close to the line, the correlation coefficient *r* is close to 1 or −1, and the predictions based on the regression line are generally good and they involve only a small error.

- For qualitative variables measured by a nominal scale, statistical association is analyzed with the help of a contingency table, also called a two-way table or cross-tabulation. Statistical association means that individuals who are in a given category of the independent variable are more likely to be in a specific category of the dependent variable than in other categories. There are ways of measuring the strength of the association, but they will not be discussed here.

- If we want to study the statistical association between a qualitative variable X and a quantitative variable Y (measured by a numerical scale), we compare the average scores of Y across the various categories determined by X.

This situation is summarized in Figure 12.2. In each of these situations, there are specific ways of determining whether the observed association is statistically significant, which means that it is likely to be valid for the population as a whole and not just for the sample.

Relationship between variables. This notion is used to describe the *logical link* between variables. The independent variable could be a *cause* of the dependent variable, or an *explanatory factor* of the dependent variable; they could both be effects of some other variable; or they may be two indicators of a concept, or even two aspects of the same phenomenon. Relationship between variables is a qualitative notion. It is a matter of interpretation, and it depends on the theoretical framework used in the research and on the research question or the research hypothesis. For instance, if there is a statistically significant difference in the

How to measure statistical association?
It depends on the level of measurement of the variable

LEVEL OF MEASUREMENT OF THE VARIABLES

PROCEDURE FOR ESTABLISHING THE ASSOCIATION

Scale vs Scale
(two quantitative variables)

Correlation and regression
The value of r, the correlation coefficient, tells us whether the association is strong or weak, and whether it is positive or negative. The regression line (given by an equation as well as on a graph) helps us predict how an individual scores on the dependent variable when we know the score on the independent variable. When predicting the values of a dependent variable, there is always an error, which is small when the correlation is strong.
CHAPTER 10

Nominal vs Nominal
(two qualitative variables)

Crosstabs and chi-squared
We compare the row percentages across the categories of the independent variable. If the difference is big we say that there is a statistical association.
CHAPTER 11

Nominal vs Scale
(one qualitative and one quantitative variable)

Compare means, t-tests and ANOVA
We compute the mean of the quantitative variable for each category defined by the nominal variable separately. We compare these means to see if there is a big difference across categories.
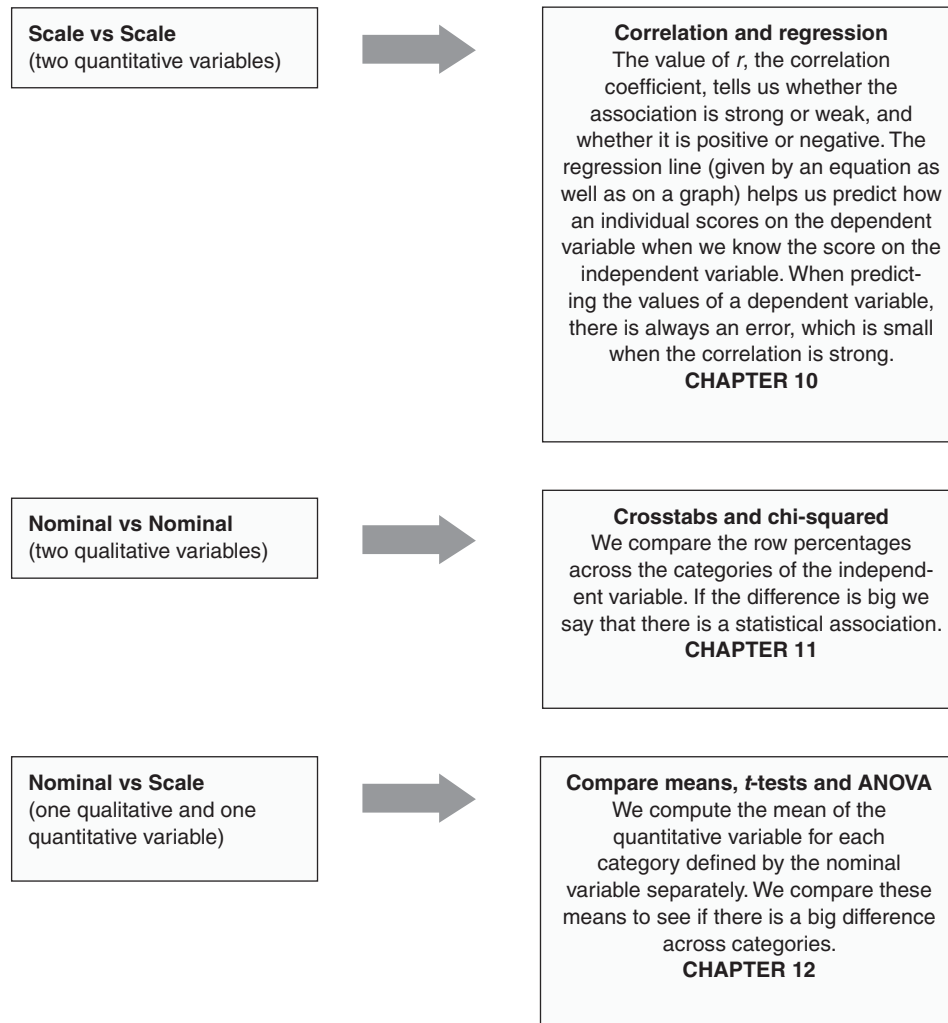**CHAPTER 12**

Figure 12.2   Method for measuring statistical association as a function of the measurement level of the variables

average age when first married between people who do not have a college degree and those who do, it may mean that getting a college degree does have something to do with the fact people get married later on in their lives than the rest of the

population. But statistical association should not be automatically interpreted as meaning a causal link.

Summary

This chapter developed the notion of statistical association for the case where the independent variable is qualitative and the dependent variable is quantitative. All the tests seen here are variations on the approach used in t-tests. If you want to establish whether a relationship exists between a qualitative independent variable and a quantitative dependent variable, you start by computing the average scores on the dependent variable across the groups defined by the independent variable. If you notice a difference, this tells you that there is a relationship between the two variables *at the level of the sample*, but this relationship may not be *statistically significant*. To be statistically significant, the differences observed between group averages should not be due to the randomness of sampling. In order to establish whether this is the case, you compute the F-statistic (or get SPSS to do it for you). You also calculate the level of significance, which is the probability of getting this specific F-value under the null hypothesis of no difference between the means. If the significance level is less than 0.05, you conclude that such an F-value is so rare under the null hypothesis that you prefer instead to accept the alternative hypothesis of a difference between the population means of the various groups.

In practical terms, several procedures can be followed to establish whether observed differences are statistically significant, depending on the number of independent variables and the number of classes in the independent variables. We considered three situations:

One qualitative independent variable, with two categories only	**Independent samples *t*-test. Levene's test of equality of variances** tells you whether to assume equal variances between the two groups or not.
One qualitative independent variable, with several categories	**One-way ANOVA**, with post-hoc tests if the ANOVA test is statistically significant, to determine which of the groups differ significantly.
More than one qualitative independent variable, with possibly more than two categories each	**General Linear Model** → **Univariate**. Post-hoc tests are conducted for the variables that have more than two categories. Main and interaction effects can be estimated.
All cases	You can use the **Compare Means** procedure and check the option that gives you ANOVA tables, but the output is less detailed than in the other procedures.

In all these cases, a measure of the strength of the association, called the *effect size*, can be computed. When comparing means, the effect size is measured by *eta squared*. The association is considered weak if eta squared is 0.01, moderate if it is 0.06, and strong if it is 0.14.

SPSS allows you to do many more tests. The principles learned in the chapter can easily be extended to understand the various adjustments made, when for instance the individuals in the samples are paired. This changes slightly the procedure and the calculation done by SPSS, but it does not alter the basic approach: you observe a difference between groups, and then you use some statistic (usually the F-statistic) to determine whether that difference is statistically significant.

This chapter also made some remarks on the notion of statistical association in general. Statistical association, independently of the level of measurement of the variables, is not to be confused with a qualitative relationship between the variables. Sometimes the relationship is spurious. But it could be a causal relationship, or some other kind of qualitative relationship. This depends on the theoretical framework that informs the statistical exploration. This is perhaps the most important lesson to remember. Statistics by themselves are meaningless. They can only be interpreted within a given conceptual or theoretical framework. Sometimes this framework is implicit: we take it so much for granted that we do not think about it, and we do not question it. Our conclusion suggests that you should think about it and make it explicit, for this is what will give meaning and relevance to your statistics.

Key words

- Statistical association between a qualitative independent variable and a quantitative dependent variable

- Prediction

- Dependent variable

- Independent variable

- Effect size

- Interaction between independent variables

- t-test

- Independent samples t-test

- ANOVA (one-way, two-way, multi-way, or one-factor, two-factor, multi-factor)

- F-statistic

- Main effect

- Simple main effect

- Interaction effect

- Post-hoc tests

- Levene's test of equality of variance

SPSS tutorial

We will illustrate the t-tests and ANOVA procedures with the *ANOVA_ex1.sav* data set, available on the website of this book. The data is fictitious and designed simply to illustrate the procedures. It contains three quantitative variables: *Grade in English*, *Grade in Psychology*, and *High School Average*. It also contains two qualitative variables: *Sex* and *Class* (three categories). Here are the basic procedures used to analyze the relationship between one or two qualitative independent variables and one quantitative dependent variable.

The Compare Means procedure

Let us start with one independent variable.

1. Select **Analyze → Compare Means → Means**….

2. Place the variable *Grade in Psychology* in the **Dependent List** box, and the variable *Sex* in the **Independent List** box.

3. Click on **Options**…. Keep the Mean, Number of Cases, and Standard Deviation in the **Cell Statistics** box (they should be there already; if not, move them from the Statistics box on the left to the **Cell Statistics** box). At the bottom of the table, check the box for **Anova table and eta**, and click on Continue.

4. Click on **OK**.

This produces Tables 12.9–12.11.

Table 12.9 The report table of the Compare Means procedure

Report			
Grade in Psychology			
Sex	Mean	N	Std. Deviation
Male	58.24	34	17.722
Female	70.66	35	16.823
Total	64.54	69	18.250

Table 12.10 The ANOVA table of the Compare Means procedure with *Sex* as the independent variable

ANOVA Table							
			Sum of Squares	df	Mean Square	F	Sig.
Grade in Psychology * Sex	Between Groups	(Combined)	2661.156	1	2661.156	8.921	.004
	Within Groups		19986.003	67	298.299		
	Total		22647.159	68			

Table 12.11 The Measures of Association table of the Compare Means procedure

Measures of Association		
	Eta	Eta Squared
Grade in Psychology * Sex	.343	.118

Table 12.9 tells you that female students in this example have a much higher average in psychology than men, indicating that, at the level of the sample, there is a statistical association between *Sex* and *Grade in Psychology*, but the table does not tell you whether such a difference is statistically significant. Table 12.10 does. It gives you the ANOVA results: the *F*-statistic is 8.921, and it tells you that a value that large, or larger, would occur with a probability less than 0.004. Therefore, you accept the conclusion that this difference is statistically significant – that it is valid for the population at large. And in drawing this conclusion, you could be wrong at most 4 times in a thousand. The eta squared statistic is equal to 0.118, which indicates a rather strong effect.

In step 4 above, you could have pasted the commands into the Syntax window instead of clicking on **OK**. You would have obtained the following commands:

```
MEANS TABLES=gradepsy BY sex
   /CELLS MEAN COUNT STDDEV
   /STATISTICS ANOVA.
```

The commands are useful because one could run the procedure again with other variables by making minimal changes in the commands.

For instance, you could replace the variable *Sex* by the variable *Group*, and run the command. You would see the variation in the average grade for the three groups (72.79, 69.36, and 48.60, respectively), and that these differences are statistically significant ($F = 15.686$; $p < 0.001$). Eta squared is 0.322, which indicates a very strong effect of the *Group* variable on the average grades in psychology.

The Compare Means procedure with two independent variables

The same procedure could be run with two independent variables. The procedure is similar to the previous **Compare Means** procedure:

1. Select **Analyze → Compare Means → Means….**

2. Place the variable *Grade in Psychology* in the **Dependent List** box, and the variable *Sex* in the **Independent List** box**.**

3. In the **Layer** box, click on **Next** and place the variable *Class* in the **Independent List** box.

4. Click on **Options**…. Keep the Mean, Number of Cases, and Standard Deviation in the **Cell Statistics** box (they should be there already; if not, move them from the Statistics box on the left to the **Cell Statistics** box). At the bottom of the table, check the box for **Anova table and eta**, and click on **Continue**.

5. Click on **Paste**.

The commands you get are slightly different from the previous case:

```
MEANS TABLES=gradepsy BY sex BY group
   /CELLS MEAN COUNT STDDEV
   /STATISTICS ANOVA.
```

Table 12.12 shows the resulting Report table.

Table 12.12 **The Report table of the Compare Means procedure when there are two layers of independent variables**

Report				
Grade in Psychology				
Sex	Class	Mean	N	Std. Deviation
Male	1	66.73	11	12.042
	2	60.44	9	16.591
	3	50.14	14	19.529
	Total	58.24	34	17.722
Female	1	78.08	12	13.283
	2	74.38	16	10.714
	3	49.43	7	17.897
	Total	70.66	35	16.823
Total	1	72.65	23	13.703
	2	69.36	25	14.494
	3	49.90	21	18.550
	Total	64.54	69	18.250

Table 12.12 shows the resulting Report table. It gives you the average for each category of the two independent variables, as well as for their combination. For instance, you can read that males scored on average 58.24 in psychology, and that males in group 1 scored 66.73 on average. The other tables are the same as for the case of one independent variable and are interpreted the same way.

The *t*-test for independent samples

The **Compare Means** procedure does not allow you to have outputs as detailed as the *t*-test or the ANOVA procedures. The *t*-test is formulated as a hypothesis testing procedure, and you can select the desired level of confidence. Here is how the *t*-test procedure works for independent samples. Let us take again the variables *Sex* and *Grades in Psychology* in the same data file as above.

1. Select **Analyze → Compare Means → Independent-Samples T Test**.

2. Place the variable *Grade in Psychology* in the **Test Variable(s)** box, and the variable *Sex* in the **Grouping Variable** box.

3. Click on the **Define Groups...** button and input 1 for **Group 1** and 2 for **Group 2**. This button actually allows you to compare any two groups when your variable has several grouping

categories. But you can only compare two at a time. You could also use a quantitative variable as the grouping variable, and choose a cut-point to separate your sample into two groups, above and below this cut-point.

4. Click on the **Options…** button to determine the required level of confidence (by default it is 95%), and click on **Continue**.

5. Click on **Paste** if you want to save the commands for future use and run them, or **OK** if you just want to run the procedure directly.

The commands are as follows:

```
T-TEST GROUPS=sex(1 2)
    /MISSING=ANALYSIS
    /VARIABLES=gradepsy
    /CRITERIA=CI(.95).
```

If you run them, you get Tables 12.13 and 12.14.

Table 12.13   The Group Statistics table of the Independent-Samples T Test procedure

Group Statistics					
	Sex	N	Mean	Std. Deviation	Std. Error Mean
Grade in Psychology	Male	34	58.24	17.722	3.039
	Female	35	70.66	16.823	2.844

Table 12.14   The Independent Samples Test table of the Independent-Samples T Test procedure

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Grade in Psychology	Equal variances assumed	.070	.792	−2.987	67	.004	−12.422	4.159	−20.723	−4.121
	Equal variances not assumed			−2.985	66.559	.004	−12.422	4.162	−20.730	−4.113

As explained before, we first look at the Levene's test, and we conclude from it that we can assume that the variances between the groups are equal. The *t*-test proper gives a significance level of 0.004, which is the same as before and has the same interpretation. We can assert, therefore, that there is a statistically significant association between the variables *Sex* and *Grades in Psychology*, that is, that the average grade in psychology for women is different from the average for men.

The one-way ANOVA procedure

We will now see the ANOVA procedure to test the hypothesis that the grades in Psychology differ significantly between the three classes.

1. Select **Analyze → Compare Means → One-Way ANOVA.**

2. Place the variable *Grade in Psychology* in the **Dependent List** box, and the variable *Class* in the **Factor** box. Click on the **Post Hoc…** button. You get the window shown in Figure 12.3.
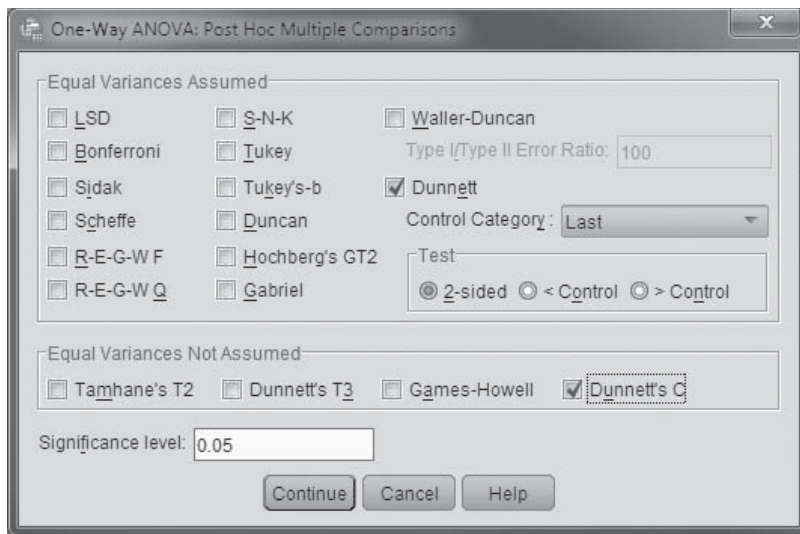

Figure 12.3   The Post Hoc dialog box

3. You have four choices to make. You must choose a significance test for the case where the variances are equal. In this case, you must decide whether to contrast the groups with the first one or the last one. Here, it is the last that has been chosen. You must then choose which test to conduct if the variances were not equal. And finally, you must choose the desired level of significance. We will not discuss the advantages of each of these tests: you can find such a discussion in the SPSS Help menu. Make the choices as indicated in Figure 12.3.

4. Click on the **Options…** button and check **Descriptive** and **Homogeneity** of **variance test**. Click on **Continue**.

5. Click on **Paste** to see what the commands look like and run them.

Here are the commands:

```
ONEWAY gradepsy BY group
    /STATISTICS DESCRIPTIVES HOMOGENEITY
    /MISSING ANALYSIS
    /POSTHOC=C DUNNETT ALPHA(0.05).
```

The first table in the output gives you the mean grade for every group, and several other statistics, including a confidence interval for every population group mean. Notice that the confidence intervals for groups 1 and 2 overlap, but none of them overlaps with that of group 3. This means that the difference in population means between groups 1 and 2 are not statistically significant, but the differences with group 3 are. This will be confirmed by the other outputs. The test of homogeneity of variance gives us a Sig. of 0.157, which means that we can assume that the variances of the three populations (from which the three groups come) are equal. The ANOVA table gives a Sig. of 0.000, which means you can safely assume that the three population means are indeed different. You now need to examine the differences of two groups at a time, which are given in the Multiple Comparisons table (Table 12.15).

Post-hoc tests are done only if the overall ANOVA test is statistically significant. The aim of post-hoc tests is to identify which groups differ significantly when we compare them two at a time. The first remark is that there are two sets of results. The Dunnett *C*-test is run when the variances are not equal. For this reason, while differences that are considered significant are flagged, the precise risk of error in making that assessment is not provided. The Dunnett *t*-test is run when variances are equal. It does tell us that groups 1 and 2 differ significantly

Table 12.15 The Multiple Comparisons table of the post-hoc tests

			Mean Difference			95% Confidence Interval	
	(I) Class	(J) Class	(I – J)	Std. Error	Sig.	Lower Bound	Upper Bound
Dunnett C	1	2	3.292	4.070		−6.90	13.49
		3	22.747*	4.955		10.24	35.25
	2	1	−3.292	4.070		−13.49	6.90
		3	19.455*	4.979		6.91	32.00
	3	1	−22.747*	4.955		−35.25	−10.24
		2	−19.455*	4.979		−32.00	−6.91
Dunnett t (2-sided)	1	3	22.747*	4.707	.000	12.13	33.37
	2	3	19.455*	4.617	.000	9.04	29.87

The mean difference is significant at the 0.05 level.

Dependent Variable Grade in Psychology — Multiple Comparisons

from group 3, but it does not tell us whether they differ from each other. The overlap of the confidence intervals tells us they do not differ significantly. In this case, since the Levene test told us that the variances could be considered equal, we only need to look at the Dunnett *t*-test.

Try to rerun the test with the first group considered as the reference group. All the conclusions drawn above should be confirmed.

Conducting a two-way ANOVA

We will now examine the combined effect of the variables *Sex* and *Class* on grades in psychology. We know from the previous analyses that each of these variables does have an effect on grades in psychology. But we do not know whether they interact, which means that the variable *Sex* affects each of the three groups differently, or, to put it another way, whether the variable *Class* has a different effect on men and women. In other words, we have an implicit model here, which states that:

Grades in Psychology = **Constant factor + effect of** *Sex* **+ effect of** *Class* **+ effect of the interaction** *Sex*Class* **+ other effects, not accounted for.**

Here is the procedure:

1. Select **Analyze → General Linear Model → Univariate….**

2. Place the variable *Grade in Psychology* in the **Dependent Variable** box, and the variables *Class* and *Sex* in the **Fixed Factor(s)** box. Click on the **Post Hoc…** button. You get the window shown in Figure 12.4.
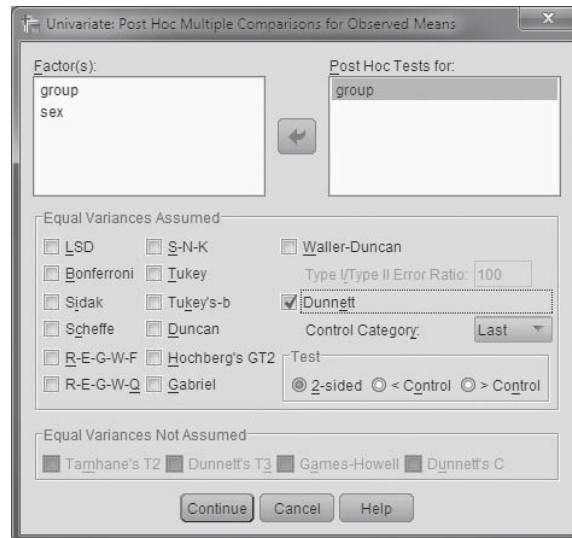

Figure 12.4   The Post Hoc dialog window for the univariate ANOVA test. Notice that the variable group should be placed in the Post Hoc Tests for: box, on the right

3. You must choose a significance test for the case where the variances are equal, and you must decide whether to contrast the groups with the first or last one. In this example, we have chosen the Dunnett test, and we have chosen the **Last** category as the Control Category, as shown in Figure 12.4.

4. Click on the **Options…** button, place the (Overall) factor in the **Display Means for:** box, and check **Descriptive** and **Homogeneity of variance test**. Click on **Continue**.

5. Click on **Paste** to see what the command looks like and run it.

You should be able to easily interpret the first three tables of the output, explained before (and not reproduced here). The fourth one is the most important one. It tells you the effect of each one of the explanatory factors (a constant, *Sex*, *Class*, *Sex*Class*, and Error, that is, other factors not taken into account). Here it is.

Table 12.16 The two-way ANOVA table that tests the significance and the effect of each of the factors

Tests of Between-Subjects Effects

Dependent Variable Grade in Psychology

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	9272.368[a]	5	1854.474	8.735	.000
Intercept	242148.942	1	242148.942	1140.607	.000
group	6453.992	2	3226.996	15.200	.000
sex	689.066	1	689.066	3.246	.076
group * sex	988.676	2	494.338	2.329	.106
Error	13374.791	63	212.298		
Total	310027.000	69			
Corrected Total	22647.159	68			

a. R Squared = .409 (Adjusted R Squared = .363)

We can see here that the model as a whole is significant. We can also see that only one of the two independent variables, the **group**, has a significant impact on the dependent variable. **Sex** and the interaction between **sex** and **group** do not. The Sig. value for variable **sex** is 0.076, which is more than the threshold of 0.05. The Sig of the interaction is 0.106, a risk far too high to say that the interaction does have an effect on grades in psychology.

You may have noticed something intriguing. When you run a test on the variable **sex** alone, it turns out to have a significant effect. But when you include both **sex** and **group**, only the latter has a significant effect. How come the effect of the factor **sex** disappears (ceases to become significant) when you include the **group** in the analysis? The answer is that men and women are not assigned to the three groups in equal proportions. So, part of the effect of the variable **sex** manifests itself through the variable **group**, thus rendering the variable **sex** less useful in the explanation.

Exercise 1. Repeat all the analyses proposed above for the variable *Grade in English*, then, separately, for the variable *High School Average*. Determine whether each of the two independent variables explains the variations when analyzed on its own, then use the two-way ANOVA procedure to study the two independent variables together and their interaction. A clarification is in order. What does it mean to study the effect of the class you are in when attending college on your high school average? Of course this is the wrong way to ask the question. A statistical association of the class variable, if found, would not mean that there is an effect of the class you are in on your high school average. It would mean, rather, that the groups under study are not formed of students having the same degree of preparedness, since their past performance in high school differed significantly.

Suggestions for further reading

Agresti, Alan (2010) *The Analysis of Ordinal Categorical Data* (2nd edn). Hoboken, NJ: Wiley.

Hamburg, Morris and Young, Peg (1996) *Statistical Analysis for Decision Making* (6th edn). New York: Harcourt Brace Jovanovich.

Hildebrand, David K., Laing, James D., and Rosenthal, Howard (1977) *Analysis of Ordinal Data*. Beverly Hills, CA: Sage.

Howell, David C. (2011) *Fundamental Statistics for the Behavioral Sciences* (7th edn). Belmont, CA: Wadsworth-Cengage Learning.

Manly, Brian F.J. (2005) *Multivariate Statistical Methods: A Primer* (3rd edn). London: Chapman & Hall.