THREE

# ACCOUNTING FOR DIFFERENCES IN A COMPLEX WORLD

This chapter provides a ladder that leads from the foundations laid in the first two chapters to the higher levels of data analysis where we find multivariate techniques. The ideas and techniques we have discussed so far enable analysts to find patterns in one or two sets of differences, that is, to engage in univariate and bivariate analysis. However, it is obvious that the world is not neatly packaged into isolated pairs of variables. So if attempts are made to account for differences only in a two-by-two fashion, it is inevitable that the resulting pictures will be at best limited or at worst distorted. Multivariate techniques are tools that help the analyst reduce these limitations and distortions by capturing more complex portions of the world inside one analytic frame.

In Section 3.1 we examine the fundamental limitations of bivariate analysis when it is applied to data representing three or more variables. These limitations stem partly from the increasing number of variables as such, but also from the more complex patterns of relationships that now become possible and even likely. This latter concern will lead us to introduce the distinctive relational patterns of confounding, moderating, and mediating relationships. Section 3.2 introduces the strategy that lies at the heart of multivariate data analysis, and discusses in general terms how this strategy combats the limitations of bivariate analysis identified in Section 3.1. The strategy involves combining variables into a composite by "weighting" each variable and adding it into the composite. The notions of composites and weights are so fundamental that we will examine their meaning and interpretation in some detail, but always stressing conceptual rather than statistical issues. To ensure that these central ideas are firmly grounded, we begin the section with a review of simple regression; extend this into multiple regression, which accommodates more than one independent variable; and finally show how the multiple regression approach can be generalized into a broad MDA strategy. In Section 3.3 we step back from the details and try to gain a more critical perspective on what multivariate data analysis has to offer. Its strengths are considerable, but they can easily be exaggerated or at least misinterpreted. This happens partly because of the language employed, when statistical terms such as "explanation" and "prediction" appear to promise a lot but actually have much more constrained meanings than they do in general research discourse. With these final reflections in mind, we should then be ready to approach the techniques themselves in Part 2 in an informed and balanced way.

## 3.1 LIMITATIONS OF BIVARIATE ANALYSIS

In Chapter 1 we used simple regression to explore the possibility that differences in positive affect might account for differences in well-being. Since we also had data on life satisfaction, we could have repeated the analysis to see whether differences in satisfaction also account for differences in well-being. What are the limitations of addressing these two relationships in separate bivariate analyses? They can be organized into three categories: descriptive, inferential, and relational limitations.

From a descriptive perspective, we would have no indication of how positive affect and satisfaction are *jointly* related to well-being. For example, we would have no ready answer to the question of how much variance in well-being is explained *in total* by these two attributes. Under certain circumstances an answer to this question can be derived from bivariate analyses but, as we will see, these circumstances are likely to be rare. More powerful answers to this type of descriptive question are to be found in multivariate analysis.

Turning to statistical inference, in Chapter 2 we found that the outcome of a null hypothesis test depends critically on alpha, the threshold probability that must be met if the null hypothesis is to be rejected. The value of alpha is our protection against inappropriately rejecting the null hypothesis (Type I error) more often than we would wish—usually 5% of the time. One feature of this procedure that we passed over at the time is that the calculations are predicated on a "one-off" sampling logic. If we keep returning to the data set to test more and more hypotheses that are linked by common variables, the actual value of alpha inflates invisibly. So, while we may think that a protective alpha of .05 is in place, its real value may be much higher. Put another way, the more hypotheses we test about the same variables, the more likely we are to believe erroneously that we have found an effect. A wide range of techniques have been developed to counter this inflationary consequence of multiple tests, and we will explore some of them in Chapter 6. These are valuable, but it would be helpful to also have at our disposal another approach that enabled us to test a *set* of hypotheses all at once. We will find that multivariate analysis provides exactly this facility so that, for example, we could evaluate whether positive affect and satisfaction have any statistically significant relationship with well-being using just one multivariate test.

Although the limitations of bivariate analysis with respect to indexing the total effect of a set of variables and to reducing the number of hypothesis tests are important, those concerning complex relationships among variables easily outweigh them. Here we return to the central idea that the world is not composed of isolated pairs of variables. It is unlikely, for example, that any effects of positive affect and satisfaction on well-being are independent of each other, or indeed of other variables so far unmentioned. If this is true, we need to think systematically about how three or more variables might be related in principle and then about how such patterns might be statistically analyzed. Clearly, by definition, bivariate analysis will not be up to the task, and so we reveal another motive for turning to multivariate analysis.

Three types of multivariable relational patterns are commonly considered by analysts: the confounding, moderating, and mediating patterns. We can illustrate all of these with our two independent variables and one dependent variable, but we need to appreciate that these are the simplest, three variable versions. As more variables are added into an analysis, the potential for ever more complex relationships grows at an alarming rate. Also, this basic list of patterns is not exhaustive, but it covers many analytic situations and provides building blocks for other possibilities.

In the **confounding** pattern the individual effects of two independent variables on a dependent variable are distorted because the independent variables are themselves related. So, for example, if positive affect and satisfaction are systematically related in some way, then it may prove difficult to untangle the effect of each on well-being: They are potentially confounded with each other. This means that a search for confounds should focus on variables that might be systematically related to an independent variable *and* to a dependent variable: a *triangle* of relationships. It is important to appreciate, though, that even when such a triangle is suspected or demonstrated, distortion due to confounding will not necessarily occur. The relationship triangle is a necessary but not sufficient condition for confounding. It is also important to note that distortions due to confounding can manifest in a variety of ways. Depending on the strength and direction of the relationships in the triangle, effects may appear, disappear, increase or decrease in magnitude, or change their sign. As a consequence, it is desirable not only to consider and measure potential confounding variables, but also to conduct analyses that reveal where confounding is actually occurring and to exercise appropriate control over it. Multivariate analysis provides the means to achieve this, using what is known as statistical control. Later we will use this type of control to see if positive affect and satisfaction are confounded and to obtain estimates of their unconfounded effects on well-being.

The **moderating** pattern is not a form of potential distortion as such, but it opens up the way for more sophisticated theorizing even with only three variables. In essence it suggests that the relationship between an independent variable and a dependent variable differs according to the *level* of a third variable. It is thus suggested that the third variable *moderates* the relationship. In our example we might suggest that the impact of positive affect on well-being will be stronger for individuals who enjoy a higher level of satisfaction. The key word here is “stronger” as this specifies how the effect changes at higher levels of the moderating variable. This type of pattern in which a bivariate relationship is thought to differ according to the level of a third variable is also known as an **interaction effect.** In the example positive affect and satisfaction are proposed to have an interactive or joint effect on well-being, which is additional to whatever individual effects each may have. Here we have a so-called *two-way interaction effect* since two independent variables are implicated. As more independent variables are added to the picture, more two-way and higher-order interaction effects become possibilities. Finally, note that the analyst is free to specify any form of moderating relationship as long as there is theoretical or empirical justification. The particular form suggested here, whereby the effect of positive affect *increases* at *higher* levels of satisfaction, has just been drawn from the air, however much it might appeal to common sense.

The **mediating** pattern can be thought of as a causal chain and as another way in which theorizing might be made more sophisticated. In this pattern an independent variable is seen as having an effect on a dependent variable *through* another independent variable. Put another way, if we imagine a causal chain linking three variables, the middle variable is said to mediate the effect of the first variable on the third. We might theorize that the effect of satisfaction on well-being is mediated by positive affect. In this case, satisfaction produces positive affect that in turn enhances well-being. It is possible in principle to distinguish between total and partial mediation. Applying the total mediation scenario to our example would suggest that the *only* way in which satisfaction influences well-being is through positive affect. Partial mediation would suggest that positive affect is only one pathway by which satisfaction might have an impact on well-being. Choosing between total and partial mediation effects is again the prerogative of the analyst and an important issue as it will guide expectations about the patterns that should emerge in multivariate analyses.

Later in the following section we will discuss in general terms how confounding, moderating, and mediating relationships can be viewed from a multivariate analysis perspective. For now, it is important mainly to appreciate the form of each type of relationship and the distinctions among them. It is also important to appreciate that the types of relationship are not mutually exclusive. All three may occur in the same analysis and require appropriate analytic strategies. Whatever else is clear at this point, it should be evident that bivariate strategies in and of themselves cannot begin to deal with the complexities we have introduced. They do, however, provide the building blocks from which multivariate strategies can be constructed.

## 3.2 THE MULTIVARIATE STRATEGY

We are now finally ready to approach the multivariate question: How do we set about accounting for differences when the analysis contains three or more variables? As we saw in the last section, ideally the answer to this question should enable the analyst to treat variables in sets or subsets, minimize the number of statistical tests, and provide ways of capturing confounding, moderating, and mediating patterns of relationships. Our discussion of the answer will begin in Subsection 3.2.1 with a review and some further elaboration of the technique of simple, or bivariate, regression analysis that we introduced in Chapter 1. Then, in Subsection 3.2.2 we will see how this can be easily extended into **multiple regression,** when there are two or more independent variables. After exploring in general terms how multiple regression can be used to analyze confounding, moderating, and mediating relationships, we will discuss in Subsection 3.2.3 how the regression approach can be seen as just one example of the generalized strategy that lies at the heart of multivariate analysis. So simple regression and multiple regression are used in this section simply as a vehicle to introduce the key concepts of multivariate analysis. A more detailed treatment of multiple regression can be found in Chapter 4.

### 3.2.1 A Review of Regression Building Blocks

Imagine that 5 individuals, identified as A–E, provide scores on measures of positive affect, satisfaction, and well-being. This time each score is on an interval scale with a possible range of 1 to 10. The imaginary scores appear in Table 3.1.

**Table 3.1** Three Sets of Scores for 5 Individuals (A–E)

| *Individual* | *Positive Affect* | *Satisfaction* | *Well-Being* |
|---|---|---|---|
| A | 3 | 3 | 3 |
| B | 5 | 6 | 5 |
| C | 6 | 5 | 4 |
| D | 7 | 4 | 5 |
| E | 9 | 7 | 8 |

To what extent are differences in well-being due to positive affect in these data? As discussed in Chapter 1, the relationship between two interval-level variables can be summarized with a regression equation, which has two unknowns: the slope and the Y intercept. Calculating these for the present data gives a slope of .75 and a Y intercept of .5. These numbers can be used to address the research question from the perspective of individual differences and of group differences. But before we look at these, it is helpful to begin by focusing on just one individual such as person E. The regression equation can be used to predict Person E's well-being score on the basis of E's positive affect score of 9, by multiplying the latter score by the slope and adding the Y intercept, as follows:

**E's predicted well-being score = slope times E's positive affect score + Y intercept**
**= .75(9) + .5**
**= 7.25**

How much in error is this prediction for person E? This can be calculated by simply subtracting the predicted from the actual well-being score of 8: 8 − 7.25 = .75, an error value known as the residual. So, in effect, the regression equation can be used to generate two new scores for person E or for any other individual in the sample: a predicted score on the dependent variable and a residual. The first score of 7.25 represents that part of the dependent variable that can be predicted from the independent variable, and the second residual score of .75 represents the remaining unpredictable part. Moreover, obviously these two new scores add up to the dependent variable score:

**E's dependent variable score = E's predicted score + E's residual score**
**8 = 7.25 + .75**

Since each person has a predicted and a residual score, we can summarize across the 5 cases and derive statistics that capture individual differences in the usual way. All of the relevant statistics appear in the ANOVA summary table in Table 3.2. The sum of squares for the predicted scores is otherwise known as the regression sum of squares that we encountered in Chapter 1, and dividing it by the number of independent variables produces the regression variance (mean square). In the present example the regression sum of squares and variance are both 11.25. How can this be turned into a more interpretable statistic that indicates how much of the individual differences in well-being can be accounted for by differences in positive affect?

**Table 3.2** ANOVA Summary Table Showing the Relationship Between Positive Affect and Well-Being

| *Source of Differences* | *Sum of Squares* | *Degrees of Freedom* | *Variance* |
|---|---|---|---|
| Regression | 11.25 | 1 | 11.25 |
| Residual | 2.75 | 3 | 0.92 |
| Total | 14.00 | 4 | |

The regression sum of squares of 11.25 indicates how much of the individual differences in well-being can be predicted from differences in positive affect. The *total* amount of individual differences in well-being regardless of any other variable is indicated by the sum of squares for well-being, which is 14. These two figures turn out to be directly comparable because of the Y intercept, about which we have had little good to say up to this point. The Y intercept can be seen as an adjustment to the regression calculations, which ensures that the mean of the predicted dependent variable is equal to the mean of the dependent variable itself. It then follows that our two sums of squares numbers are comparable because they are based on deviations around the same mean. As a result, we can simply divide the regression sum of squares by the total sum of squares and arrive at $r^2$: the proportion of variance in the dependent variable accounted for by the independent variable. For the present data $r^2$ is 11.25/14 = .804, which says that positive affect accounts for 80.4% of variance in well-being.

Turning now to the five residual scores, we see that these can also be turned into a sum of squares and variance. Following from the logic of $r^2$, the residual sum of squares divided by the total sum of squares will indicate the variance in well-being that is *not* accounted for by positive affect. This number is 2.75/14 = .196, which says that positive affect fails to account for 19.6% of the variance in well-being. As we would expect, explained variance and unexplained variance add up to 1 or 100%. It is also possible to view unexplained variance in terms of a standard deviation. If we take the square root of the residual variance, we produce the standard error of estimate, which in the present case is .96. This is directly comparable with the standard deviation of the dependent variable, which is 1.87. One way to make this comparison is in terms of a proportional reduction in error. If we subtract the standard error of estimate from the standard deviation of the dependent variable and divide it by the latter, we can see proportionately how much error is reduced by taking account of the independent variable as opposed to just using the mean of the dependent variable. The calculation here is (1.87 −.96)/1.87 = .487. This means that the amount of error made in predicting individual well-being scores is reduced by 48.7% if we take account of positive affect scores.

We have used the regression equation to generate predicted happiness scores and residuals for each individual and then derived summary statistics such as $r^2$ and the standard error of estimate to quantify how far *individual* differences in positive affect account for *individual* differences in well-being.

Now we turn to the question of how far *group* differences in positive affect account for *group* differences in well-being. This is simply answered by reference to the slope statistic in the regression equation. The value of .75 says that groups who differ by 1 on the positive affect measure differ by .75 on the well-being measure, on average. Note that because we are now taking a group perspective, we are concerned with what's happening "on average."

This completes the review of key statistics in simple regression. We have taken time to revisit and elaborate on them because all of them carry over into multiple regression. So, having a good grasp of them at this point will smooth the transition we are about to make from bivariate to multivariate analysis.

### 3.2.2 The Composite Variable in Regression

In Subsection 3.2.1 we divided up the relationship between two variables, well-being and positive affect, into additive components. We saw that:

**well-being score = predicted well-being score**
**+ residual**

and decomposed the right-hand side of the equation into:

**predicted well-being score = slope(positive affect score)**
**+ Y intercept**

It is this regression equation that provides the vehicle for dealing with more than one independent variable. All that is required is to extend the equation by literally adding in a new term for each additional independent variable. So, for the data in Table 2.1 the equation would be:

**predicted well-being score = slope(positive affect)**
**+ slope(satisfaction)**
**+ Y intercept**

To explore the details of this new multiple regression equation, we will focus again on person E in Table 3.1. The regression equation for person E is:

**E's predicted well-being score = (.5)(9) + (.5)(7) + (−.5)**

The scores for positive affect (9) and satisfaction (7) are taken from the last row of Table 3.1. The values for the slopes (which happen to be the same) and Y intercept have been calculated using the SPSS computer package. (As we move into multivariate analysis proper we will no longer refer to the calculations as such as they become too complex and are best left to a computer.) The predicted well-being score for person E is 7.5. Since this person's actual well-being score is 8 (see Table 3.1), the residual score is .5. Note that the predicted score is now based on two independent variables rather than just the one we used earlier and, for this case, has reduced the residual from .75 to .5. Adding in information about person E's satisfaction has increased predictive power or, equivalently, reduced predictive error. This may be true for this person E, but how well are we now accounting for individual differences for all of the cases?

Although the regression equation has been extended, the predicted happiness scores and residuals are no different in form from those in simple regression. So we can proceed straight to summary statistics derived from these scores. In terms of explained variance, the computer calculations result in a value of .893. Positive affect and satisfaction together account for 89.3% of individual differences in well-being for these 5 cases, compared with the 80.4% we achieved with just positive affect. In multiple regression this statistic becomes the multivariate version of $r^2$ and is now referred to as **multiple *$R^2$*,** with an uppercase *R,* but its interpretation remains unchanged. The other individual difference statistic we reviewed earlier was the standard error of estimate: an indicator of unexplained variability. When only positive affect was in the equation, the standard error of estimate was .96. When satisfaction is added in, the computer calculations show that the standard error of estimate drops to .87—a further reduction in error.

Reflecting on the nature of $R^2$ takes us straight to the core concept in multivariate analysis. It can be thought of as the squared correlation between the actual and predicted well-being scores. Since the predicted scores are generated by the independent variables packaged up inside the multiple regression equation, $R^2$ can also be seen as the squared correlation between the dependent variable and a *composite variable* that contains all of the independent variables. This forming of a composite variable in order to analyze many variables all at once is the strategy that lies at the heart of multivariate analysis. Here we see the strategy in action in the form of a multiple regression equation, but later we will generalize the idea to reveal its analytic power and scope more fully.

The composite variable not only generates the predicted and residual scores for each individual in the analysis, but also contains information that indicates in what way *group* differences in the independent variables relate to *group* differences in the dependent variable. As in simple regression, this information is to be found in the slopes. The simple regression we conducted earlier on positive affect and well-being resulted in a slope of .75, whereas the positive affect slope in the multiple regression is .5. Something has shifted, but what exactly? In a multiple regression a slope is more properly known as a **partial slope.** Roughly speaking, it quantifies the impact of an independent variable over and above the impacts of all other independent variables in the equation. The notion of a partial slope is not intuitively obvious, and it is fundamental to multivariate analysis. So it will be beneficial to explore various ways in which a partial slope can be interpreted.

The most general way to conceptualize a partial slope is in terms of statistical control. The partial slope of .5 indicates the impact of positive affect on well-being when satisfaction is statistically controlled. Similarly, the partial slope for satisfaction (which also happens to be .5) indicates its impact on well-being when positive affect is controlled. But *what* is being controlled? The answer is the relationship between positive affect and satisfaction. The correlation between these two independent variables turns out to be .71, that is, they share just over 50% of their variance ($.71^2 = .504$). This means that their effects on well-being may be confounded, and so some sort of confound detection and control is needed. This is exactly what multiple regression offers by way of partial slopes. The shift from a slope of .75 in the simple regression to a partial slope of .5 in the multiple regression shows confound detection and control in action for the positive affect variable. The first number indicates the effect of positive affect on well-being, whereas the second indicates the same effect once the confounding relationship with satisfaction has been controlled or neutralized.

Another way to approach the partial slope is in terms of holding a potentially confounding variable constant. The strategy of controlling a variable by holding it constant is a staple of experimental design. If we wanted to study the causes of a behavior that varied systematically with the time of day, for example, then it would be wise to always conduct the study at the same time. Holding time constant in this way would ensure that it could not confound the effects of other variables of interest. Turning a variable into a constant guarantees that it cannot covary with any other variable and so cannot be a confound. The confound triangle of relationships has been broken. This strategy of literally holding something constant is obviously not always a practicable or ethical option. The partial slope can be seen as another way of achieving the same end in a nonliteral, statistical fashion. So the partial slope for positive affect indicates its impact on well-being when satisfaction is held constant. Put another way, the partial slope shows how much well-being differs across groups who differ by a score of 1 on the positive affect measure, but who do not differ on the satisfaction measure.

A third way to think about the partial slope is in terms of adjusted variables. From this perspective the partial slope shows the impact of an independent variable on the dependent variable after the independent variable has been adjusted to take account of its relationships with the other independent variables. Synonymously, analysts also talk about "correcting" for or "partialing" out other variables. So the partial slope for positive affect captures the impact of positive affect on happiness after adjusting for, correcting for, or partialing out positive affect's relationship with satisfaction. Although we will not delve into the statistical mechanics of how this is achieved, it is instructive to look a little more closely at how partialing is performed. One way to do this is to think of multiple regression as a series of simple regressions.

To find the partial slope for positive affect, we could first conduct a simple regression with positive affect as the dependent variable and satisfaction as the independent variable. This would generate two new variables: a predicted positive affect score and a residual score for each case. The residuals variable contains those differences in positive affect that are *not* predicted by differences in satisfaction. So the residuals variable can be treated as an adjusted positive affect variable, that is, adjusted to exclude any differences due to satisfaction. The positive affect variable has been "residualized" with respect to satisfaction. This residualized variable would then become the independent variable in another simple regression with well-being as the dependent variable. The slope for positive affect from this simple regression would be equal to the partial slope from the multiple regression since the positive affect slope has been adjusted for the relationship between positive affect and satisfaction. We could do the same parallel procedure for satisfaction. First there would be a simple regression, this time with satisfaction as the dependent variable and positive affect as the independent variable. Then there would be another simple regression with the residuals variable from the first regression as the independent variable (residualized satisfaction) and well-being as the dependent variable. This would produce a slope indicating the impact of satisfaction having adjusted for positive affect.

Using the regression equation to form a composite variable provides solutions to some of the limitations we discussed in Section 3.1. By generating the predicted dependent variable, we have a device for calculating statistics such as $R^2$ that refers to the independent variables *as a set*. The multiple regression shows that positive affect and satisfaction *together* explain 89.3% of the variance in well-being. It is important to appreciate that this $R^2$ figure is different from the sum of the two $r^2$ values that are generated by the two separate simple regressions. These simple regressions show that positive affect explains 80.4%, and satisfaction explains 71.4% of variance in well-being. However, because of the relationship between the two independent variables ($r = .71$), the relationship of each with the dependent variable is inflated in this case. The positive affect variable's predictive ability is confounded with that of satisfaction and vice versa. The statistical control provided by multiple regression not only produces partial slopes, but also produces an $R^2$ value that is adjusted to take account of correlations among independent variables. The only time that a multiple regression $R^2$ is equal to the sum of the simple regressions' $r^2s$ is when the independent variables are uncorrelated, and so there is nothing to adjust for. Since multivariate techniques are most commonly adopted in order to deal with correlated independent variables, this is a rare situation.

The second limitation of conducting bivariate analyses on multivariate data that we identified in Section 3.1 was the problem of the inflation of Type I error (alpha) produced by multiple hypothesis tests. This, too, is countered by the composite variable and its products. When a multiple regression is carried out, it is usual to begin by testing the null hypothesis that $R^2$ is zero in the population sampled using an $F$ test. This is equivalent to testing whether *any* of the independent variables has a statistically significant relationship with the dependent variable. Since all of the relationships are tested simultaneously, this is often referred to as an **omnibus test.** This is a very efficient approach in the sense that if the test of $R^2$ is *not* statistically significant, that is, the null hypothesis is accepted, there is no need to conduct further hypothesis tests on individual independent variables. In our example the $F$ of 8.33 has an associated $p$ value of .11. Assuming that we adopt a conventional alpha of .05, the null hypothesis is therefore accepted, and no further tests are justified. Despite the large sample value for $R^2$ of .893, it is still consistent with the hypothesis that the population $R^2$ is zero. Note in passing that this outcome is highly influenced by the tiny sample size and the consequent lack of statistical power.

If the test of $R^2$ had been statistically significant, we could have proceeded to test hypotheses about each of the partial slopes. This type of analysis tests the hypothesis that a partial slope is zero in the population using either a $t$ or an $F$ test. For example, the partial slope for positive affect is .5, its $t$ value is 1.83, and its $p$ value is .209. As the $R^2$ test already suggested, there is no significant relationship between positive affect and well-being, and a $t$ test for the satisfaction slope would lead to the same type of conclusion. Hypothesis testing will play an important role in the multivariate techniques that we explore in Part 2. For now, the key point to note is that the use of a composite variable allows the analyst to test multiple hypotheses simultaneously and thereby reduce the number of tests in any one analysis. This in turn helps to combat the problem of inflating Type I error, which occurs when multiple, linked hypotheses are tested in sequence. Further, hypotheses can also be tested about the partial slopes that capture the effects of independent variables after they have been adjusted for confounding relationships among the independent variables.

The third limitation of bivariate analyses is that of handling multivariable patterns such as confounding, moderation, and mediation. By now it should be clear how multiple regression provides a tool for detecting and controlling confounding relationships. It can also be used to analyze moderating and mediating relationships, and the general ways in which this is accomplished bring to the surface two further strengths of multivariate analysis. These are the capacity to represent complex effects as single terms in a regression equation and to analyze complex relationships using series of regression equations.

In our earlier discussion of moderating relationships, we suggested the possibility that satisfaction might moderate the effect of positive affect on well-being. Put another way, this is the possibility that positive affect and satisfaction might have an *interactive* effect on well-being *over and above* their individual effects. This additional effect can be thought of as another independent variable that is literally the product of the positive affect and satisfaction variables. The new variable can be added to the multiple regression equation so that it now takes the form:

**predicted well-being score = slope(positive affect) + slope(satisfaction)**
**+ slope(positive affect X satisfaction)**
**+ Y intercept**

Values for the partial slopes in this multiple regression equation can be found in the usual way. Our particular interest would be in the partial slope for the interaction variable. If this partial slope was statistically significant, we would have evidence that there is an interaction effect over and above the individual effects of the two independent variables. Further analysis would then be required to find out whether the particular form of the interaction was in line with expectations. Statistical details aside, the important conceptual point here is that moderating relationships can be literally added into a composite variable and tested in a controlled fashion, just like any other independent variable. Generalizing this point further, we will see that a variable in a regression equation can be constructed to represent all sorts of effects, which gives the technique enormous scope for capturing complex relationships.

Regarding mediation relationships, it was suggested earlier that the effect of satisfaction on well-being may be mediated by positive affect. This envisages a causal chain in which satisfaction leads to increased positive affect, which in turn enhances well-being. This pattern can be evaluated with a *series* of regressions that decompose the chain of relationships. Such an approach is called **hierarchical or sequential regression,** and in the present example would proceed as follows. A first (simple) regression would be run to provide a slope that showed the impact of satisfaction on well-being: the first and last variables in the chain. A second (multiple) regression would then be run to provide the partial slopes that showed the impact of satisfaction and positive affect on well-being. If it is true that satisfaction can only have an impact on well-being through positive affect (total mediation), then breaking the chain by holding positive affect constant should remove any relationship between satisfaction and well-being. From a statistical standpoint, this means that we would expect a significant positive slope for satisfaction in the first regression, but a nonsignificant partial slope in the second regression when positive affect is held constant.

There are many other issues surrounding the analysis of mediation effects, some of which we will pursue in later chapters. The main purpose of this discussion is to introduce the notion of sequential regression and to give an initial sense of how it might be used to analyze mediation relationship patterns. It is worth noting that in practice sequential regression is also used to analyze moderation or interaction effects. In the earlier example a multiple regression would first be run to examine the independent effects of positive affect and satisfaction on well-being: the so-called **main effects.** Then a second multiple regression would be run that included the main effects variables and the interaction variable. Again our interest would be in comparing the results of the two regressions, this time in terms of $R^2$. Specifically, we would want to know whether there was a noteworthy increase in $R^2$ when the interaction variable was added into the regression equation. This increase can itself be tested using an *F* test. If it were significant, we would have evidence that there was an interaction effect of the two independent variables over and above their separate effect. As these two brief examples show, sequential regression provides a powerful tool for the analysis of complex patterns.

### 3.2.3 Generalizing the Composite Variable

Multiple regression involves expressing the relationship between one dependent variable (DV) and multiple independent variables (IV1, IV2, etc.) into various additive components. The components can be written:

**DV = [slope 1(IV1) + slope 2(IV2) + . . . . . . . .**
**+ Y intercept]**
**+ residual**

Slopes 1 and 2 represent the effects of IV1 and IV2, respectively, on the dependent variable. The dotted line indicates that we can add more independent variables, separately and/or in combination. The terms within the square brackets form a composite variable and are combined as shown to generate the predicted dependent variable score. Within this composite, the Y intercept can be seen just as an adjustment to ensure that the predicted and actual dependent variables are aligned by having the same mean. Finally, the residual captures that part of the dependent variable that is not predicted by the composite variable.

So far all of these multivariate ideas have been expressed in terms of multiple regression, a particular multivariate technique. It will be helpful now to reexpress the equation above in more general terms so that we will be able to use it each time we examine a particular technique in Part 2. This common framework will do more than any other device to help us make sense of the unity of multivariate techniques. The re-expression involves no more than the replacement of some of the words in the equation as follows:

**DV = [coefficient 1(effect 1) + coefficient 2(effect 2) + . . . . . . . .**
**+ constant]**
**+ residual**

The IV terms have been replaced by the more general term "effect." An effect may be a single variable, a so-called main effect; or it may be two or more variables in combination—an interaction effect. What were slopes are now referred to as **coefficients,** which is just a general instruction to multiply the variable value that follows by the coefficient value. Coefficients are also referred to as weights, so, for example, regression slopes may be described as regression coefficients or regression weights. The composite variable in the square brackets is often described as a weighted linear sum since it involves summing a set of variable scores, each of which has been weighted by a multiplier or coefficient. Rather than "weight," though, we will adopt the term "coefficient" since it is the most commonly used across multivariate techniques. The term "Y intercept" is specific to regression so it has been replaced by the more generic term "**constant.**" This word is also useful because it carries the connotation of a fixed adjustment. The terms "residual" and "error" are used synonymously. However, since error also has many other meanings, using the more precise "residual" is preferable.

With these terms in place, we can now reflect on some aspects of this cornerstone of multivariate analysis. The composite variable is often seen as a device for building a statistical "model" that accounts for patterns in data, such as the differences found in a dependent variable. As we have seen, this modeling can be viewed from the perspective of the individual—what is the predicted value on the dependent variable for this person; of individual differences—how much of the individual differences in the dependent variable can be explained by differences on the independent variables; and of group differences—how are group differences on the dependent variable related to group differences on the independent variables? The term "modeling" is helpful because it reminds us that it is up to the analyst to choose which effects should be included in the model. The statistical framework allows any number of variables in principle, singly or in combination, and in their original or some transformed state, so the choice can be immense. If the analyst is to account for differences in parsimonious and helpful ways, many effects have to be excluded, but this must be achieved without loss of important information. Well-developed theories should play an important role in guiding these choices. The data analyses themselves, though, also have a role to play by allowing the analyst to *compare* different models of the same data set. This model-comparison approach is one we will encounter repeatedly in Part 2.

Although the composite variable will provide a common framework within which we can locate all of the multivariate techniques in Part 2, we will find that the way in which coefficients are generated will differ across techniques. As we saw in Chapter 1, regression coefficients are calculated according to the principle of least squares. This means that coefficient values are found that minimize prediction error, that is, the sum of squares of the residuals. Another equivalent way to express the least square criterion is that it produces coefficient values that maximize the correlation between actual and predicted scores on the dependent variable. Each time we encounter a new multivariate technique, we will ask what criterion is used to generate the coefficients. The criterion will always be in the form of minimizing or maximizing something, but the something will vary by technique.

This completes our introduction to the core ideas of multivariate analysis. The composite variable, made up of a sum of weighted variables, is clearly a powerful tool with the potential to provide answers to a wide range of complex research questions. However well an analysis has been executed, though, the problem remains that the answers may be misinterpreted. In Part 2 particular issues surrounding the correct interpretation of results from specific techniques will loom large. In addition to these issues, there are what could be called "generic" misinterpretations, in the sense that they arise often and with respect to all multivariate techniques. In the following section we discuss the nature of these generic misinterpretations as a final preliminary before turning to the techniques themselves.

## 3.3 COMMON MISINTERPRETATIONS OF MULTIVARIATE ANALYSES

For our purposes, generic misinterpretations can be organized around four questions:

- What does "accounting for" differences mean?
- To whom or what do the results apply?
- What do the results of hypothesis tests mean?
- What does statistical control actually achieve?

The first three questions are not specific to multivariate analyses in that they can be raised for simpler forms of analysis as well. However, they are so fundamental that they deserve discussion in any introduction to data analysis. Moreover, the sophisticated and complex appearance of multivariate analyses can mislead the unwary into believing that the problems to which these questions point have been solved in some magical statistical way. In fact, multivariate analysis only helps to deal with the issues surrounding the last question and then only in a limited fashion, as we will see.

Answering these four questions leads quickly into deep statistical and philosophical waters. The aim of this section is not to dive into these waters, but to give some general sense of the ways in which misinterpretations can arise, so that they can be avoided. The references at the end of the chapter provide lively and accessible treatments of the underlying issues for those who wish to explore further.

### 3.3.1 What Does "Accounting for" Differences Mean?

The vague expression "accounting for" appears throughout this book and was chosen because it carries fewer potentially misleading connotations than many of the synonyms that are used in the language of data analysis. As we have seen, data analysts talk in terms such as *predicting* scores on, and *explaining* variance in, the dependent variable. It is also common to define the regression slope as the amount of *change* in the dependent variable when the independent variable changes by one unit. So it appears that analyses are able to provide predictions, explanations, and verification of causal processes. However, data analyses in and of themselves provide none of these.

The term "prediction" is a slippery one that has at least three meanings in the research context. In its strongest sense it means making a claim about what will happen in the future, and doing so successfully is seen as a mark of scientific progress. In another sense prediction is a synonym for constructing hypotheses about differences or relationships, without any reference to time. In the final meaning a predictor is just a statistical synonym for an independent variable. The presence of a "predictor" in an analysis suggests that the analyst believes that it may help to account for differences in the dependent variable and usually implies a formal hypothesis to that effect. However, even when it is shown that the "predictor" is significantly related to the dependent variable, this does not demonstrate any predictive power in the strong sense. The only basis for this interpretation would be if the data had been gathered longitudinally, where the independent variable is measured at one time and the dependent variable some time later. So interpreting "accounting for" in terms of prediction is justified by the design and analysis of a study, not by the analysis itself.

The term "explanation" is much more slippery and multidimensional than "prediction," so we will focus more on what it is not than on what it is. The first thing to note is that it is different from prediction. It is possible to make successful predictions without being able to explain why these predictions work. Similarly, the workings of a phenomenon may be well explained, but predicting its future states may be impossible because of the many other factors that enable or prevent the occurrence of these states. Turning back to data analyses, if we detect "explained" variance in a dependent variable it simply means that the independent and dependent variables covary in some systematic way. Why this covariation occurs remains open to explanation. Explanation lies in the realm of theory, not data analysis. Good theories generate testable hypotheses, and the results from the subsequent data analyses may be consistent with the hypotheses or not. There is clearly an important connection between theory and data analysis, but one should not be confused with the other. Moreover, this connection usually comprises a series of assumptions about measurement, sampling, and so forth, any of which may be faulty. So theory and data analysis not only are separate, but also have a loose linkage. All of this indicates that whatever "accounting for" means in data analysis, it is not equivalent to explanation.

A special and valued type of scientific explanation is that which provides understandings of causal processes. Once again, the fact that some terms in data analysis appear to refer to causal processes does not mean that statistical results, however sophisticated, can provide evidence of causation in and of themselves. Such evidence requires a reliance on theory, research design, and data analyses, tied together with explicit arguments. The most compelling evidence typically comes from a combination of an elaborated theory from which clear causal hypotheses can be rigorously deduced, a strong experimental design that mimics the productive or generative aspect of causation and that ensures tight control of confounding variables, and an appropriate analysis for the type of data that is generated. These are strong requirements that clearly cannot be replaced by data analyses alone, however complex they may be.

Other problems may also undermine attempts to treat statistical results as explanations of causal processes. Without a compelling theoretical explanation, a potential cause may be confused with a *marker* of that cause. As a facetious example, having white hair is strongly related to the incidence of many diseases, but no one suggests hair dye as an intervention to avoid these diseases. Hair color is simply a marker for age, which *is* causally implicated in the occurrence of disease. Statistical analyses are blind to the distinction between cause and marker, so other, additional ways must be found to develop accurate causal accounts.

Another problem in the statistical search for causes is that even complex statistical techniques are only designed to detect patterns consistent with what Lieberson (1985) calls symmetric or reversible causation. Referring to an independent variable or cause as X and a dependent variable or effect as Y will help to clarify this idea. Symmetric or reversible causation assumes that the amount of increase in Y produced by a unit increase in X will be the same as the amount of decrease in Y consequent on a unit decrease in X. But this symmetric behavior of increases and decreases may not hold. Once achieved, an increase in Y due to X may be irreversible whatever subsequently happens to X. Or a decrease in X may produce a partial reversal where Y returns to an intermediate value. Lieberson provides convincing examples of behavioral and social phenomena that display asymmetric causal processes and teases out the disturbing consequences for the study of such phenomena. For present purposes, the noteworthy consequence is that some causal processes cannot be captured by conventional data analyses. Most forms of analysis assume a billiard-ball-type of causation, but this is only one form of many. Again we arrive at the conclusion that while data analyses may contribute to the production of causal explanations under some circumstances, they cannot do so in isolation from other research activities, especially the activity of theorizing.

To complete this subsection, an even more fundamental issue about causality may be raised. The view of causal process served by conventional data analyses locates the process in the relations between variables. On this view, causal propositions are tested by evaluating the presence/absence, magnitude, and direction of relationships among variables. This is a task to which multivariate analysis is well suited, as we have seen. However, there are other views of causation that do not fit so well. For example, the realist conception locates causal forces in agents, not in relationships among variables (Sayer, 1992). So a realist causal account will explain how the causal powers of agents bring about changes. Further, such an account will refer to ways in which these causal powers may be enabled or constrained. A fundamental consequence of enablements and constraints is that a causal process will not necessarily be evident in any consistent way in patterns of relationships. Sometimes you see them, sometimes you don't. Deep waters are close by, so we will not pursue this further. The general point to note is that the causal accounts to which multivariate analyses can contribute are but one way to develop scientific explanations. So, to interpret their results meaningfully, the analyst has to buy into and, if necessary, defend a particular conception of causation, which itself is problematic.

### 3.3.2 To Whom or What Do Statistical Results Apply?

At first glance, this question appears to refer to the issue of generalizing beyond the data available. However, this is not the focus of this subsection. The issue here is that of being clear about what "units" are the object of an analysis at any given point. The potential problem of misinterpretation this raises is that of applying results to the wrong unit of analysis. In the analyses conducted in Part 1, we have consistently distinguished three "levels" of analysis: the individual, individual differences, and group differences. Analysts in different disciplines focus on different levels in this sense. A sociologist, for example, might conduct an analysis containing communities that aggregate into towns or cities. But in any statistical analysis there will always be multiple levels, whatever they may be, and the points to be raised here apply regardless of the nature of those levels or units. For the sake of consistency, the discussion will continue to be framed in terms of individuals and groups.

The most fundamental point to note is that since statistical analyses typically aggregate across individuals, the results do not refer or apply to any single individual. As we have seen, aggregate statistics, such as regression slopes, may be used to generate a prediction about an individual, but they do not quantify any attribute of any particular individual. This may seem self-evident, and most analysts are aware of it. However, there is a subtle version of this slide from the aggregate to the individual that is common, at least in psychology. Valsiner (1986b) has provided an intriguing demonstration of how even experienced researchers misinterpret simple correlations. The description of the correlation typically begins by referring to groups or an averaged relationship, but it quickly slides into a discourse about an idealized individual. This can be partly explained by the tension in psychology between a disciplinary focus on the individual and a general reliance on aggregate data to help us understand the individual. But Valsiner argues convincingly that many other cognitive processes lie behind the misinterpretation, and there is no reason to believe that these processes afflict only psychologists.

Once pointed out, the potential for interpretive confusion between group and individual is clear though not always easily avoided. The distinction between group differences and aggregated individual differences is harder to grasp but important because it provides another common cause for misinterpretation. The distinction can be sharpened up by considering again two bivariate statistics that refer to group differences and individual differences, respectively: the slope and $r^2$. A regression slope indicates how *group* differences on the dependent variable are related to *group* differences on the independent variable, where the groups are defined by values of the independent variable. In contrast, the $r^2$ statistic indicates how *individual* differences in a dependent variable are related to *individual* differences in an independent variable. So while the slope can be interpreted as the averaged effect of the independent variable on the dependent variable, $r^2$ cannot be interpreted in this way. The reason for highlighting this is that many research reports express a theoretical interest in group effects but then give pride of place to $r^2$-type statistics. This has the unfortunate effect of implicitly shifting research objectives away from examining group effects to maximizing the capture of individual differences on the dependent variable. As we will see in Chapter 4, since $r^2$ is highly dependent on variances and therefore unstable across samples, it has very limited uses. But the more fundamental concern is that focusing on $r^2$ at the expense of the slope can distort research objectives and lead to misinterpretations.

### 3.3.3 What Do the Results of Hypothesis Tests Mean?

In Chapter 2 null hypothesis testing was introduced as the predominant method used by social scientists to evaluate the role of chance in their results. Statistical significance with an alpha of no more than 5% is usually the license required for results to be treated as worthy of interpretation. This approach to setting a threshold for interpretability, based on a rejection that the pattern of results is due to chance, continues to play an important role throughout the realm of multivariate analysis, as we will see in Part 2. Over many decades methodologists have debated the meaning and worth of null hypothesis testing, usually from a critical perspective. Of all the procedures explored in this book, null hypothesis testing is the one that has received the most critical attacks. Yet despite this, it remains a cornerstone of data analysis in the mainstream social sciences. So it is important for any data analyst or user to gain some sense of the criticisms and to find a comfortable personal position.

Frank Schmidt, a very distinguished methodologist in psychology, has provided a memorable thumbnail sketch of null hypothesis testing that certainly pulls no punches. He wrote:

> If we were clairvoyant and could enter the mind of a typical researcher, we might eavesdrop on the following thoughts:
>
> > . . . If my findings are not significant, then I know that they probably just occurred by chance and that the true difference [or relationship] is probably zero. If the result is significant, then I know I have a reliable finding. The *p* values from the significance tests tell me whether the relationships in my data are large enough to be important or not. I can also determine from the *p* value what the chances are that these findings would replicate if I conducted a new study.
>
> Every one of these thoughts about the benefits of significance testing is false.
>
> —Schmidt (1996, p.126)

This arresting paragraph with its knockout punch line raises a host of issues too complex to pursue here in any depth. A few comments, though, may help to explain Schmidt's conclusion. The outcome of a null hypothesis test turns on the *p* value, and it is misinterpretations of this probability that lay the foundations for further misconceptions. As we noted in Chapter 2, the *p* value represents the probability of finding a sample difference or relationship at least as large as that calculated *if the null hypothesis were true.* The italicized words highlight two important points. First, the probability is a *conditional* probability, not simply the probability of a particular sample value occurring at all. Second, the probability refers to the sample value, not to the hypothesis under test. The probability of a sample value conditional on a hypothesis being true is different from the probability of a hypothesis being true conditional on a sample value. So the *p* value tells us nothing directly about the probable truth of the hypothesis. The plot thickens further when it is appreciated that the null hypothesis usually under test, what Cohen (1994) calls the "nil hypothesis," is usually if not always false as a matter of fact. The notion that a particular population difference or relationship is precisely zero seems an odd assumption, and yet it provides the precisely defined start and end point for the testing process. Finally, if there is no probabilistic basis for the sample, either by probability sampling or random assignment, any inferences from the *p* value lack a clear point of reference. All of this suggests that while a chance interpretation based on the *p* value may be the most viable of those listed by Schmidt, the exact nature of that interpretation is not straightforward or always clear.

Schmidt also dismissed the claims that statistical significance implies reliability and replicability of results. This can again be appreciated as a consequence of the abstract nature and origin of the *p* value. In null hypothesis testing, the sample value is seen as one of an infinite set of possible sample values. This set of values has a frequency distribution—the sampling distribution—that defines the relative frequency or probability of any given value occurring. It is by reference to a particular sampling distribution, representing the null hypothesis, that the *p* value is derived. Given this highly abstract framework, it is hard to see how any implications about the consistency of future sample values could be drawn. The reliability and replicability of results can be based only on the cross-validation provided by repeated analyses, either on subsamples within the same study or on samples from different studies. The latter option has been greatly enhanced in recent years by the advent of meta-analysis, whereby results from different studies can be statistically amalgamated in a rigorous fashion.

The remaining claim, that statistical significance has implications for the importance of a result, is probably the most widespread misinterpretation, despite repeated warnings in methods texts. Even if we assume that a chance interpretation of statistical significance is defensible, it still has no direct implications for other sorts of significance—theoretical, practical, or otherwise. One way to see this is first to remember that the *p* value can be interpreted as the probability of committing a Type I error—rejecting the null hypothesis when it is actually true. Now add to this a further point from Chapter 2, that an effective way to reduce Type I error is to increase the sample size. This means then that sample size is one of the determinants of the *p* value. This is borne out in practice in large sample surveys in which even tiny differences and relationships turn out to be statistically significant. At best, statistical significance may be seen as a limited indicator that a result is unlikely to be due to sampling error and that it is worthy of interpretation. It is this *subsequent* interpretation, using criteria outside the analysis, which forms a basis for claims of importance.

The cumulative effect of these concerns has led an increasing number of methodologists to recommend the abandonment of null hypothesis testing. Schmidt (1996) advocates this not only on the grounds of the indefensibility of the procedure in his view, but also because of the damage it has done. He believes that progress, at least in psychology, has been retarded by widespread Type II error. In other words, many hypotheses and their parent theories have been wrongly dismissed on the statistical basis that the null hypothesis had to be accepted, when in fact it was probably false. The most common recommended alternative is to base chance interpretations of results on point and interval estimation, as discussed in Subsection 2.3.2 of Chapter 2. It is further recommended that replication interpretations should be based on cross-validations within studies and **meta-analyses** across studies. The latter are statistical methods for combining results across studies in order to evaluate the size and reliability of effects. Issues of substantive interpretation or importance should remain outside the domain of statistics. The meta-analysis recommendation has clearly taken hold as such analyses are now a commonplace in the research literature. However, chance interpretations continue to be based on null hypothesis testing in the main—reason enough to gain an understanding of its nature and limitations.

### 3.3.4 What Does Statistical Control Actually Achieve?

As we noted earlier, all of the preceding interpretive problems can arise in almost any analysis, and multivariate analyses are certainly not exempt. Finally, we turn to an issue that only arises when the relationships among three or more variables are being analyzed: the issue of statistical control. A naive interpretation of results that have been statistically controlled would suggest that the results must be somehow definitive or "correct" since they have been "corrected." In fact, there are various ways in which the exercise of statistical control may lead to distortions or, at best, limitations.

The first problem for the researcher is to decide *which* variables might be confounds and should therefore be included in addition to the chosen independent and dependent variables. In principle, the list of possible confounds is infinite, especially when we start to think about what variables might be confounded with confounds! The threat of the missing confound, and consequent distortion due to undercontrol, is ever present. An understandable response to this problem is to be overinclusive and control for a gigantic set of possible confounds. Unfortunately, this not only reduces the statistical power of the analysis, making it even more hungry for cases, but also runs the risk of distortion due to overcontrol. The very effects being sought may be obliterated or distorted by the complex adjustments required by the presence of numerous possible confounds. Once again we arrive at the conclusion that analyses need to be designed with reference to external sources, particularly theory and existing evidence. Moreover, as Cohen (1990, pp. 1304–1305) has observed: generally speaking in data analysis "less is more" and "simple is better." There is no easy answer to the question of which variables should be controlled in an analysis, and the results are always contingent on which choices have been made. Interpretations of results have to always keep their contingent nature in mind; in no sense can they be definitive.

The second problem is one we have already encountered in Chapter 2—that of measurement quality. Whatever control variables, i.e., potential confounds, are included, they must be measured with adequate reliability and validity. If they are not, the statistical control process will be undermined and may produce unpredictable distortions of the relationships between the independent and dependent variables. It is very easy to treat control variables as second-class citizens not deserving of the measurement efforts expended on the "real" variables. However, the consequences of such neglect may be disastrous, especially since the random error produced by the unreliable measurement of a single variable can ripple through the network of relationships in a multivariate analysis, causing widespread contamination. All of this means that any report of an analysis should include evidence to reassure the reader that interpretations of the results are not threatened by poor measurement of *any* of the variables.

The final problem concerns the consequences of statistical control for the interpretation of variables themselves. As we saw in Section 3.2, statistical control involves adjusting an independent variable by removing the variance it shares with a possible confound. So the original independent variable is replaced in the analysis by an adjusted or residualized version. This raises two interpretive issues. The first is that the nature of this adjusted variable may be different according to which other variables have been adjusted for. Accordingly, the results of two different analyses focusing on a particular independent variable may not be comparable if the set of control variables is not the same in each analysis. The second, deeper issue is the question of how an adjusted variable is to be interpreted at all. If the variable of positive affect, say, has been adjusted for satisfaction, what exactly is left? It is tempting to resort to some notion of uncovering the essence of a variable by stripping away those features it somehow shares with other variables. But given the inherent interrelatedness of so many variables, this does not seem very convincing.

Once again, we are led to appreciate the contingent nature of multivariate results and the cautious interpretations that are therefore required.

After this catalog of sticky problems, it may be tempting to stop reading and give up on multivariate analysis, perhaps even on simpler analyses. But it is important to reiterate that Section 3.3 has been all about general *limitations* on interpretations of results from multivariate analyses. It is not an attack on multivariate techniques themselves, but an attempt to encourage a critical and balanced approach to them. Every technique can be misused, and the more complex the technique, the more scope for unwitting misuse.

## 3.4 FURTHER READING

A good general orientation to the multivariate perspective can be found in Cohen, Cohen, West, & Aiken (2003, Chapter 1), while Darlington (1990, Chapters 1 and 4) provides an excellent discussion of the nature of confounding and statistical control. Baron and Kenny's (1986) discussion of moderating and mediating relationships has become the classic reference for this topic. The texts by Lieberson (1985) and Sayer (1992, especially Chapter 6) provide accessible critical accounts of some of the conceptual limitations of multivariate analysis. Tacq (1997) pays extensive attention to the question of how multivariate analytic frameworks map onto the structure of research problems. Runkel (1990) and Valsiner (1986a) examine in depth the issue of the relationship between single case and aggregate analyses. Good critical accounts of null hypothesis testing are available in Cohen (1994) and Schmidt (1996), and a contrasting view can be found in Frick (1996).