13 GLM 2: Comparing means adjusted for other predictors (analysis of covariance)

13.1 What will this chapter tell me?

My road to rock stardom had taken a bit of a knock with my unexpected entry to an all-boys' grammar school (rock bands and grammar schools really didn't go together). I needed to be inspired and I turned to the masters: Iron Maiden. I first heard Iron Maiden at the age of 11 when a friend lent me *Piece of Mind* on a cassette and told me to listen to 'The Trooper'. It was, to put it mildly, an epiphany. I became their smallest (I was 11) biggest fan and obsessed about them in the unhealthiest of ways. I bombarded the man who ran their fan club (a guy called Keith) with letters, and, bless him, he replied to them all. Eventually my stalking paid off and Keith arranged for me to go backstage when they played what was then (and to me always will be) the Hammersmith Odeon in London on 5 November 1986 (*Somewhere on Tour*, in case you're interested). Not only was it the first time I had seen them live, but I got to meet them too. It is difficult to convey how exciting and anxiety-provoking that night was. It was all quite overwhelming. I was so utterly awe-struck that I managed to say precisely nothing to any of the band (but I do have some good photos where my speechlessness is tangible; see Figure 13.1). Soon to become a theme in my life, a social situation had provoked me to make an utter fool of myself.[1] When it was over I was in no doubt that this was the best day of my life. In fact, I thought, I should just kill myself there and then because nothing would ever be as good.[2] This may be true, but I have subsequently had other very nice experiences, so who is to say that they were not better? I could compare experiences to see which one is the best, but there is an important confound: my age. At the age of 13, meeting Iron Maiden was bowel-weakeningly exciting, but adulthood (sadly) dulls your capacity for this kind of unqualified excitement. To really see which experience was best, I would have to take account of the variance in enjoyment that is attributable to my age at the time. Doing so will give me a purer measure of how much variance in my enjoyment is attributable to the event itself.

1 In my teens I met many bands I liked, and Iron Maiden were by far the nicest.

2 Apart from my wedding day, as it turned out.

This chapter extends the previous one to look at situations in which you want to compare groups means, but also adjust those means for another variable (or variables) that you expect to affect the outcome. This involves a linear model in which an outcome is predicted from dummy variables representing group membership but one or more other predictors (usually continuous variables) are included. These additional predictors are sometimes labelled covariates, and this configuration of the linear model is sometimes known as *analysis of covariance*.

Figure 13.1 Dave Murray (guitarist from Iron Maiden) and me backstage in

London in 1986 (my grimace reflects the utter terror I was feeling at meeting my hero)

13.2 What is ANCOVA?

In the previous chapter we saw how we can compare multiple group means with the linear model by using dummy variables to code group membership. In addition, in Chapter 9 we saw how the linear model can incorporate several continuous predictor variables. It should, therefore, be no surprise that the linear model to compare means can be extended to include one or more continuous variables that predict the outcome (or dependent variable). When the main focus of the model is to compare means (perhaps from different experimental groups) then these additional predictors in the model are sometimes referred to as **covariates**. Also, this form of the linear model is sometimes referred to as **analysis of covariance** (or **ANCOVA** for short).[3]

3 As we've discussed before, these labels for special cases of the linear model (such as one-way independent ANOVA in the previous chapter, and ANCOVA here) reflect historical divisions in methods (see Misconception Mutt 12.1). They are unhelpful because they create the impression that we're using distinct statistical models when we're not. I want you to focus on the general linear model that underpins these special cases, but I can't really avoid using the ANOVA/ANCOVA labels now and again so that when your supervisor tells you to do ANOVA/ANCOVA you can find the relevant part of the book!

In the previous chapter, we used an example about the effects of puppy therapy

on happiness. Let's think about things other than puppy therapy that might influence happiness. Well, the obvious one is how much you like dogs (a dog phobic is going to be about as happy after puppy therapy as I would be after tarantula therapy), but there are other things too such as individual differences in temperament. If these variables (the covariates) are measured, then it is possible to adjust for the influence they have on the outcome variable by including them in the linear model. From what we know of hierarchical regression (see Chapter 9) it should be clear that if we enter the covariate into the model first, and then enter the dummy variables representing the group means (e.g., the experimental manipulation), we can see what effect a predictor variable has, *adjusting for* the effect of the covariate. In essence, rather than predicting the outcome from group means, we predict it from group means that have been adjusted for the effect of covariate(s). There are two main reasons to include covariates in ANOVA:

- **To reduce within-group error variance**: When we predict an outcome from group means (e.g., when these represent the effect of an experiment), we compute an *F*-statistic by comparing the amount of variability in the outcome that the experiment can explain against the variability that it cannot explain. If we can attribute some of this 'unexplained' variance (SS_R) to other measured variables (covariates), then we reduce the error variance, allowing us to assess more sensitively the difference between group means (SS_M).
- **Elimination of confounds**: In any experiment, there may be unmeasured variables that confound the results (i.e., variables other than the experimental manipulation that affect the outcome variable). If any variables are known to influence the outcome variable being measured, then including them as covariates can remove these variables as potential explanations for the effect of interest.

13.3 ANCOVA and the general linear model 

The researchers who conducted the puppy therapy study in the previous chapter suddenly realized that a participant's love of dogs would affect whether puppy therapy would affect happiness. Therefore, they repeated the study on different participants, but included a self-report measure of love of puppies from 0 (I am a weird person who hates puppies, please be deeply suspicious of me) to 7 (puppies are the best thing ever, one day I might marry one). The data are in Table 13.1 and in the file **Puppy Love.sav**, which contains the variables **Dose** (1 = control, 2 = 15 minutes, 3 = 30 minutes), **Happiness** (the person's happiness

on a scale from 0 to 10), and **Puppy_love** (love of puppies from 0 to 7).

Use IBM SPSS Statistics to find the means and standard deviations of both happiness and love of puppies across all participants and within the three groups. (Answers are in Table 13.2.)

In the previous chapter, we characterized this experimental scenario as equation (12.2), and knowing what we know about the linear model (Chapter 9) you can hopefully see that that equation can be extended to include the covariate as follows:

$$\text{Happiness}_i = b_0 + b_1\text{Long}_i + b_2\text{Short}_i + b_3\text{Covariate}_i + \varepsilon_i$$
$$\text{Happiness}_i = b_0 + b_1\text{Long}_i + b_2\text{Short}_i + b_3\text{Puppy\_love}_i + \varepsilon_i \quad (13.1)$$

We can compare the means of different groups using a linear model (see Section 12.2) in which groups are coded as the dummy variables **Long** and **Short**: **Long** takes the value of 1 only for the 30-minute group, **Short** takes a value of 1 only for the 15-minute group, and in all other situations they have a value of 0. We can add a covariate as a predictor to the model to test the difference between group means *adjusted for the covariate*. Let's look at a practical example.

Add two dummy variables to the file **Puppy Love.sav** that compare the 15-minute group to the control (**Dummy 1**) and the 30-minute group to the control (**Dummy 2**) – see Section 12.2 for help. If you get stuck use **Puppy Love Dummy.sav.**

Fit a hierarchical regression with **Happiness** as the outcome. In the first block enter love of puppies (**Puppy_love**) as a predictor, and then in a second block enter both dummy variables (forced entry) – see Section 9.10 for help.

Table 13.1 Data from **Puppy Love.sav**

Dose	Participant's happiness	Love of puppies
Control	3	4
	2	1
	5	5
	2	1
	2	2
	2	2
	7	7
	2	4
	4	5
15 minutes	7	5
	5	3
	3	1
	4	2
	4	2
	7	6
	5	4
	4	2
30 minutes	9	1
	2	3
	6	5
	3	4
	4	3
	4	3
	4	2
	6	0
	4	1
	6	3
	2	0
	8	1
	5	0

Table 13.2 Means (and standard deviations) from **Puppy Love.sav**

Dose	Participant's happiness	Love of puppies
Control	3.22 (1.79)	3.44 (2.07)
15 minutes	4.88 (1.46)	3.12 (1.73)
30 minutes	4.85 (2.12)	2.00 (1.63)
Total	**4.37 (1.96)**	**2.73 (1.86)**

The summary of the model resulting from the self-test (Output 13.1) shows us the goodness of fit of the model first when only the covariate is used in the model, and second when both the covariate and the dummy variables are used. The difference between the values of R^2 (0.288 − 0.061 = 0.227) represents the individual contribution of puppy therapy to predicting happiness. Puppy therapy accounted for 22.7% of the variation in happiness, whereas love of puppies

accounted for only 6.1%. This additional information provides some insight into the substantive importance of puppy therapy. The next table is the ANOVA table, which is also divided into two sections. The top half represents the effect of the covariate alone, whereas the bottom half represents the whole model (i.e., covariate and puppy therapy included). Notice at the bottom of the ANOVA table (the bit for model 2) that the entire model (love of puppies and the dummy variables) accounts for 31.92 units of variance (SS_M), there are 110.97 units in total (SS_T) and the unexplained variance (SS_R) is 79.05.

The interesting bit is the table of model coefficients (Output 13.2). The top half shows the effect when only the covariate is in the model, and the bottom half contains the whole model. The *b*-values for the dummy variables represent the difference between the means of the 15-minute group and the control group (**Dummy 1**) and the 30-minute group and the control group (**Dummy 2**) – see Section 12.2 for an explanation of why. The means of the 15-and 30-minute groups were 4.88 and 4.85 respectively, and the mean of the control group was 3.22. Therefore, the *b*-values for the two dummy variables should be roughly the same (4.88 − 3.22 = 1.66 for **Dummy 1** and 4.85 − 3.22 = 1.63 for **Dummy 2**). The astute among you might notice that the *b*-values in Output 13.2 are not only very different from each other (which shouldn't be the case because the 15-and 30-minute groups means are virtually the same), but also different from the values I've just calculated. Does this mean I've been lying to you for the past 50 pages about what the beta values represent? I'm evil, but I'm not *that* evil. The reason for this apparent anomaly is that with a covariate present, the *b*-values represent the differences between the means of each group and the control *adjusted for the covariate(s)*. In this case, they represent the difference in the means of the puppy therapy groups adjusted for the love of puppies.

Output 13.1

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.246[a]	.061	.027	1.929
2	.536[b]	.288	.205	1.744

a. Predictors: (Constant), Love of puppies (0–7)

b. Predictors: (Constant), Love of puppies (0–7), Dummy 1: (control vs. 15 mins), Dummy 2: (control vs. 30 mins)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.734	1	6.734	1.809	.189[b]
	Residual	104.232	28	3.723		
	Total	110.967	29			
2	Regression	31.920	3	10.640	3.500	.030[c]
	Residual	79.047	26	3.040		
	Total	110.967	29			

a. Dependent Variable: Happiness (0–10)

b. Predictors: (Constant), Love of puppies (0–7)

c. Predictors: (Constant), Love of puppies (0–7), Dummy 1: (control vs. 15 mins), Dummy 2: (control vs. 30 mins)

These **adjusted means** come directly from the model. If we replace the *b*-values in equation (13.1) with the values in Output 13.2, our model becomes:

$$\text{Happiness}_i = 1.789 + 2.225\text{Long}_i + 1.786\text{Short}_i + 0.416\text{Puppy\_love}_i \quad (13.2)$$

Remember that **Long** and **Short** are dummy variables such that **Long** takes the value of 1 only for the 30-minute group, and **Short** takes a value of 1 only for the 15-minute group; in all other situations they have a value of 0. To get the adjusted means, we use this equation, but rather than replacing the covariate with an individual's score, we replace it with the mean value of the covariate from Table 13.2 (2.73) because we're interested in the predicted value for each group at the mean level of the covariate. For the control group, the dummy variables are both coded as 0, so we replace **Long** and **Short** in the model with 0. The adjusted mean will, therefore, be 2.925:

$$\begin{aligned}\overline{\text{Happiness}}_{\text{Control}} &= 1.789 + (2.225 \times 0) + (1.786 \times 0) + \left(0.416 \times \bar{X}_{\text{Puppy\_love}}\right) \\ &= 1.789 + (0.416 \times 2.73) \\ &= 2.925\end{aligned} \quad (13.3)$$

For the 15-minute group, the dummy variable **Short** is 1 and **Long** is 0, so the adjusted mean is 4.71:

$$\begin{aligned}\overline{\text{Happiness}}_{15\text{ mins}} &= 1.789 + (2.225 \times 0) + (1.786 \times 1) + \left(0.416 \times \bar{X}_{\text{Puppy\_love}}\right) \\ &= 1.789 + 1.786 + (0.416 \times 2.73) \\ &= 4.71\end{aligned} \tag{13.4}$$

For the 30-minute group, the dummy variable **Short** is 0 and **Long** is 1, so the adjusted mean is 5.15:

$$\begin{aligned}\overline{\text{Happiness}}_{30\text{ mins}} &= 1.789 + (2.225 \times 1) + (1.786 \times 0) + \left(0.416 \times \bar{X}_{\text{Puppy\_love}}\right) \\ &= 1.789 + 2.225 + (0.416 \times 2.73) \\ &= 5.15\end{aligned} \tag{13.5}$$

We can now see that the *b*-values for the two dummy variables represent the differences between these *adjusted* means (4.71 − 2.93 = 1.78 for **Dummy 1** and 5.15 − 2.93 = 2.22 for **Dummy 2**). These adjusted means are the average amount of happiness for each group at the mean level of love of puppies. Some people think of this kind of model (i.e., ANCOVA) as 'controlling' for the covariate, because it compares the predicted group means at the average value of the covariate, so the groups are being compared at a level of the covariate that is the same for each group. However, as we shall see, the 'controlling for the covariate' analogy is not a good one.

Output 13.2

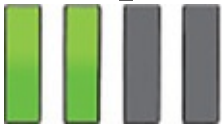
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	3.657	.634		5.764	.000
	Love of puppies (0-7)	.260	.193	.246	1.345	.189
2	(Constant)	1.789	.867		2.063	.049
	Love of puppies (0-7)	.416	.187	.395	2.227	.035
	Dummy 1: (control vs. 15 mins)	1.786	.849	.411	2.102	.045
	Dummy 2: (control vs. 30 mins)	2.225	.803	.573	2.771	.010

a. Dependent Variable: Happiness (0-10)

13.4 Assumptions and issues in ANCOVA 

Including covariates doesn't change the fact we're using the general linear model, so all the sources of potential bias (and counteractive measures) discussed in Chapter 6 apply. There are two additional considerations: (1) independence of the covariate and treatment effect; and (2) homogeneity of regression slopes.

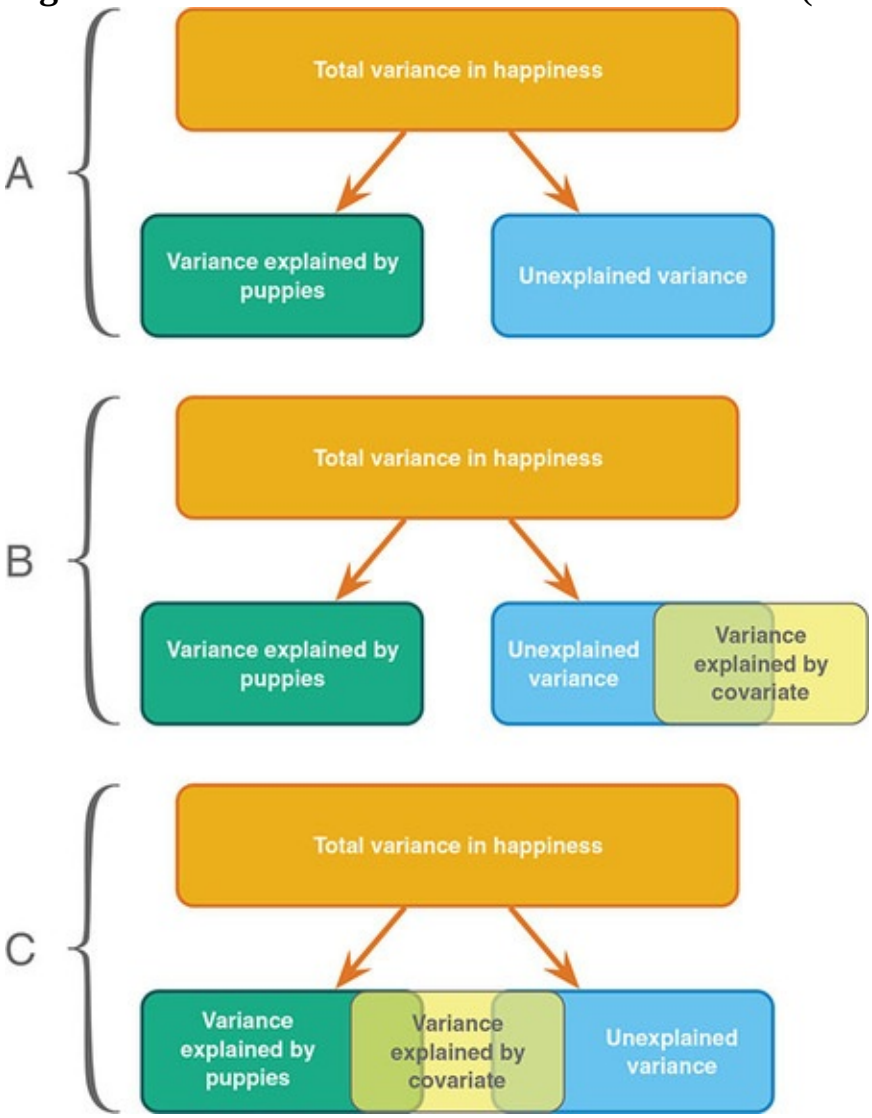
13.4.1 Independence of the covariate and treatment effect 

I said in the previous section that covariates can be used to reduce within-group error variance if the covariate explains some of this error variance, which will be the case if the covariate is independent of the experimental effect (group means). Figure 13.2 shows three different scenarios. Part A shows a basic model that compares group means (it is like Figure 12.5). The variance in the outcome (in our example happiness) can be partitioned into two parts that represent the experimental or treatment effect (in this case the administration of puppy therapy) and the error or unexplained variance (i.e., factors that affect happiness that we haven't measured). Part B shows the ideal scenario when including a covariate, which is that the covariate shares its variance only with the bit of happiness that is currently unexplained. In other words, it is completely independent of the treatment effect (it does not overlap with the effect of puppy therapy at all). Some argue that this scenario is the only one in which ANCOVA is appropriate (Wildt & Ahtola, 1978). Part C shows a situation in which the effect of the covariate overlaps with the experimental effect. In other words, the experimental effect is confounded with the effect of the covariate. In this situation, the covariate will reduce (statistically speaking) the experimental effect because it explains some of the variance that would otherwise be attributable to the experiment. When the covariate and the experimental effect (independent variable) are not independent, the treatment effect is obscured, spurious treatment effects can arise, and at the very least the interpretation of the ANCOVA is seriously compromised (Wildt & Ahtola, 1978).

The problem of the covariate and treatment sharing variance is common and is ignored or misunderstood by many people (Miller & Chapman, 2001). Miller and Chapman are not the only people to point this out, but their paper is very readable and they cite many examples of people misapplying ANCOVA. Their main point is that when treatment groups differ on the covariate, putting the covariate into the analysis will not 'control for' or 'balance out' those differences (Lord, 1967, 1969). This situation arises mostly when participants are not randomly assigned to experimental treatment conditions. For example, anxiety and depression are closely correlated (anxious people tend to be depressed), so if you wanted to compare an anxious group of people against a non-anxious group on some task, the chances are that the anxious group would also be more depressed than the non-anxious group. You might think that by adding depression as a covariate into the analysis you can look at the 'pure' effect of anxiety, but you can't. This situation matches part C of Figure 13.2 because the effect of the covariate (depression) would contain some of the variance from the effect of anxiety. Statistically speaking, all that we know is that anxiety and depression share variance; we cannot separate this shared

variance into ‘anxiety variance’ and ‘depression variance’, it will always be ‘shared’. Another common example is if you happen to find that your experimental groups differ in their ages. Placing age into the analysis as a covariate will not solve this problem – it is still confounded with the experimental manipulation. The use of covariates cannot solve this problem (see Jane Superbrain Box 13.1).

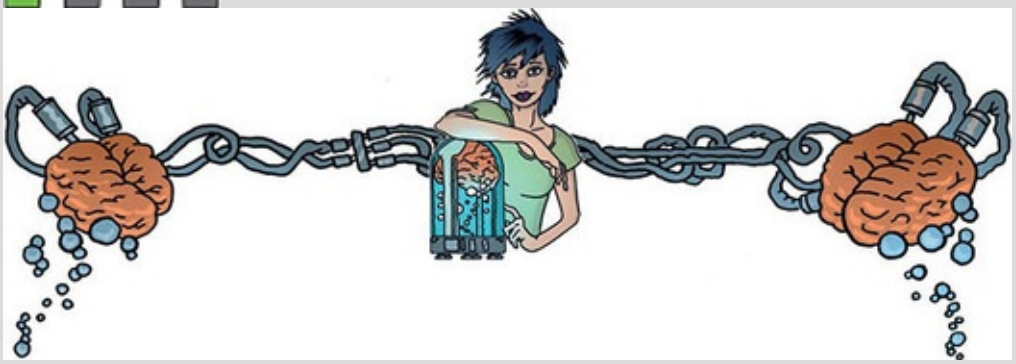
Figure 13.2 The role of the covariate in ANCOVA (see text for details)



This problem can be avoided by randomizing participants to experimental groups, or by matching experimental groups on the covariate (in our anxiety example, you could try to find participants for the low-anxiety group who score high on depression). We can see whether this problem is likely to be an issue by checking whether experimental groups differ on the covariate before fitting the model. To use our anxiety example again, we could test whether our high-and low-anxiety groups differ on levels of depression. If the groups do not

significantly differ then we might consider it reasonable to use depression as a covariate.

Jane Superbrain 13.1 An interpretational or statistical requirement?

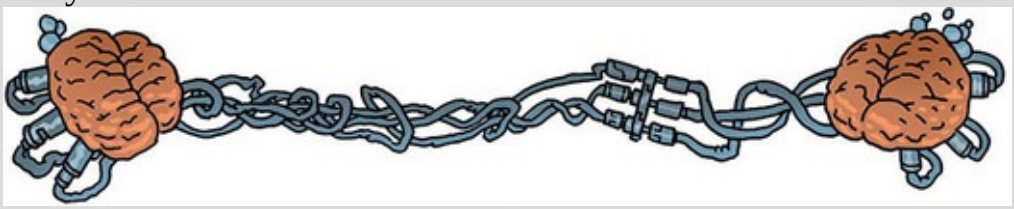


The treatment effect and covariate are simply predictor variables in a general linear model, yet despite several hundred pages discussing linear models, I haven't before mentioned that predictors should be completely independent. I've said that they shouldn't overlap too much (e.g., collinearity) but that's quite different from saying that they shouldn't overlap at all. If, in general, we don't care about predictors being independent in linear models, why should we care now? The short answer is we don't – there is no *statistical* requirement for the treatment variable and covariate to be independent.

However, there are situations in which ANCOVA can be biased when the covariate is not independent of the treatment variable. One situation, common in medical research, has been discussed a lot: an outcome (e.g., hypertension) is measured at baseline, and after a treatment intervention (with participants assigned to a treatment or control group). This design can be analysed using an ANCOVA in which treatment effects on post-intervention hypertension are analysed while covarying baseline levels of hypertension. In this scenario the independence of treatment and covariate variables means that baseline levels of hypertension are equal in the different treatment groups. According to Senn (2006), the idea that ANCOVA is biased unless treatment groups are equal on the covariate applies only when there is *temporal additivity*. To use our hypertension example, temporal additivity is the assumption that both treatment groups would experience the same change in hypertension over time

if the treatment had no effect. In other words, had we left the two groups alone, their hypertension would change by exactly the same amount. Given that the groups have different overall levels of hypertension to begin with, this assumption might not be reasonable, which undermines the argument for requiring group equality in baseline measures.

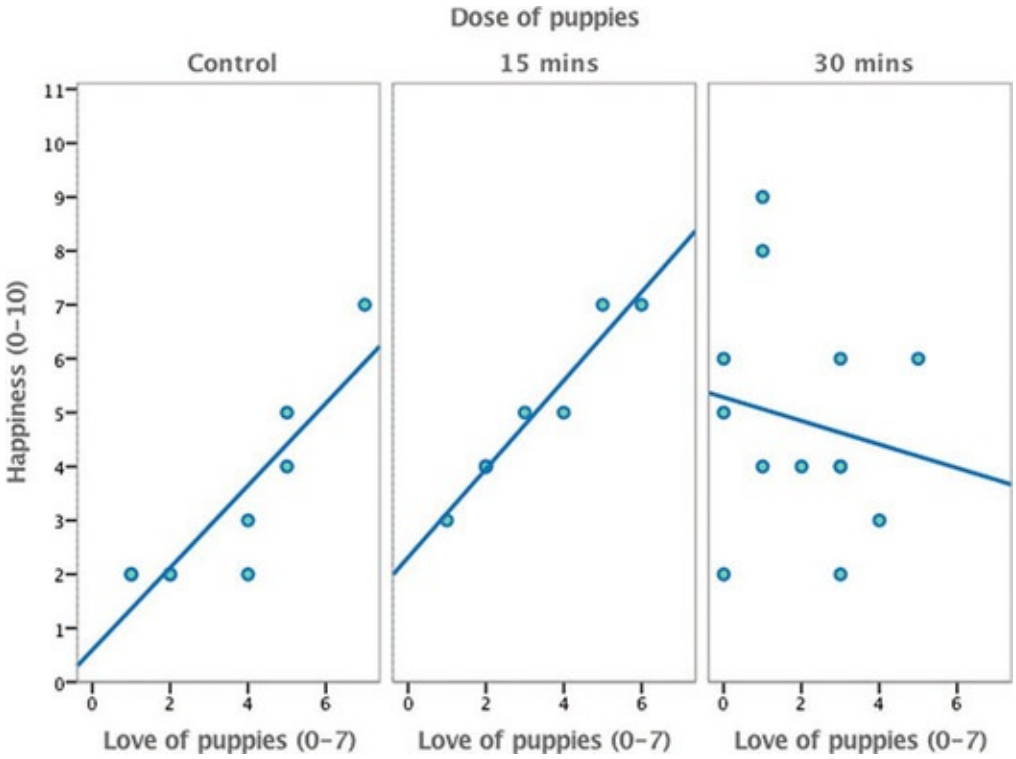
To sum up, the independence of the covariate and treatment makes interpretation more straightforward but is not a statistical requirement. ANCOVA can be unbiased when groups differ on levels of the covariate, but as Miller and Chapman point out, it creates an interpretational problem that ANCOVA cannot magic away.



13.4.2 Homogeneity of regression slopes

When a covariate is used we look at its overall relationship with the outcome variable: we ignore the group to which a person belongs. We assume that this relationship between covariate and outcome variable holds true for all groups of participants, which is known as the assumption of **homogeneity of regression slopes**. Think of the assumption like this: imagine a scatterplot for each group of participants with the covariate on one axis, the outcome on the other, and a regression line summarizing their relationship. If the assumption is met then the regression lines should look similar (i.e., the values of *b* in each group should be equal).

Figure 13.3 Scatterplot and regression lines of happiness against love of puppies for each of the experimental conditions

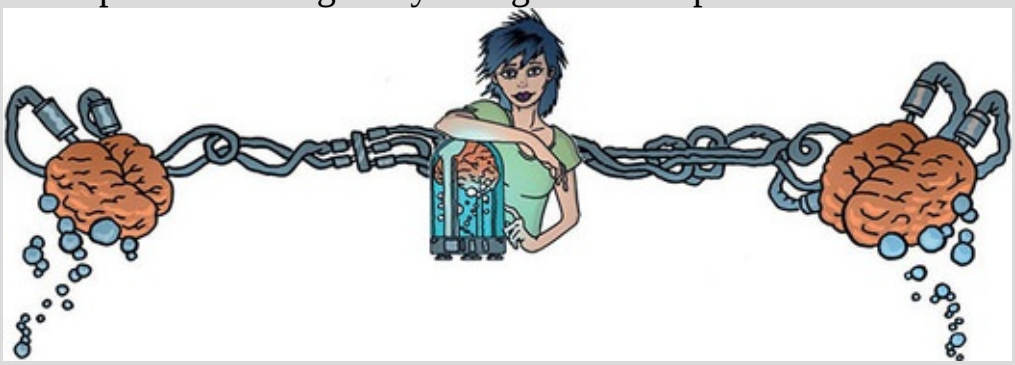


Let's make this concept a bit more concrete. Remember that the main example in this chapter looks at whether different doses of puppy therapy affect happiness when including love of puppies as a covariate. The *homogeneity of regression slopes* assumption means that the relationship between the outcome (dependent variable) and the covariate is the same in each of our treatment groups. Figure 13.3 shows a scatterplot with regression line that summarizes this relationship (i.e., the relationship between love of puppies, the covariate, and the outcome, participant's happiness) for the three experimental conditions (shown in different panels). There is a positive relationship (the regression line slopes upwards from left to right) between love of puppies and participant's happiness in both the control (left panel) and 15-minute conditions (middle panel). In fact, the slopes of the lines for these two groups are very similar, showing that the relationship between happiness and love of puppies is very similar in these two groups. This situation is an example of *homogeneity* of regression slopes. However, in the 30-minute condition (right panel) there is a slightly negative relationship between happiness and love of puppies. The slope of this line differs from the slopes in the other two groups, suggesting *heterogeneity of regression slopes* (because the relationship between happiness and love of puppies is different in the 30-minute group compared to the other two groups).

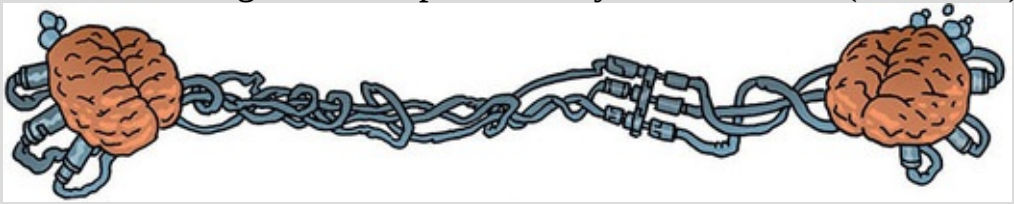
Although in a traditional ANCOVA heterogeneity of regression slopes is a bad thing (Jane Superbrain Box 13.2), there are situations where you might expect regression slopes to differ across groups and that variability may be interesting.

For example, when research is conducted across different locations, you might expect the effects to vary across those locations. Imagine you had a new treatment for backache, and you recruit several physiotherapists to try it out in different hospitals. The effect of the treatment is likely to differ across these hospitals (because therapists will differ in expertise, the patients they see will have different problems and so on). As such, heterogeneity of regression slopes is not a bad thing *per se*. If you have violated the assumption of homogeneity of regression slopes, or if the variability in regression slopes is an interesting hypothesis in itself, then you can explicitly model this variation using multilevel linear models (see Chapter 21).

Jane Superbrain 13.2 What are the consequences of violating the assumption of homogeneity of regression slopes?



When the assumption of homogeneity of regression slopes is met the resulting *F*-statistic can be assumed to have the corresponding *F*-distribution; however, when the assumption is not met it can't, meaning that the resulting *F*-statistic is being evaluated against a distribution different than the one that it actually has. Consequently, the Type I error rate of the test is inflated and the power to detect effects is not maximized (Hollingsworth, 1980). This is especially true when group sizes are unequal (Hamilton, 1977) and when the standardized regression slopes differ by more than 0.4 (Wu, 1984).



13.4.3 What to do when assumptions are violated

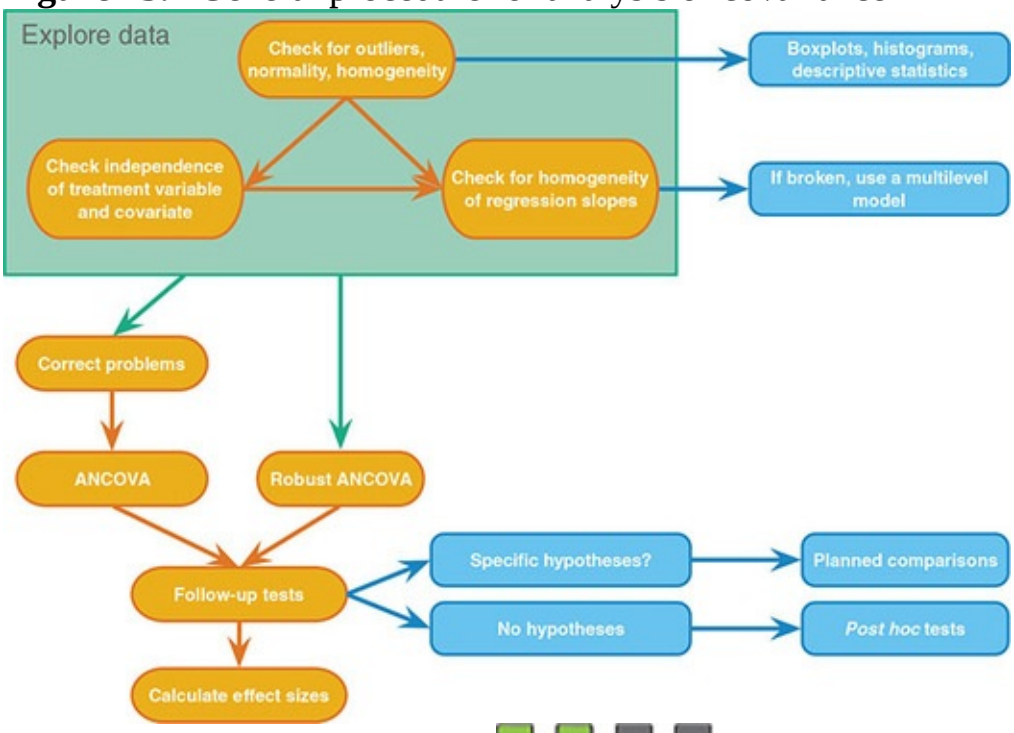
A bootstrap for the model parameters and *post hoc* tests can be used so that these, at least, are robust (see Chapter 6). The bootstrap won't help for the *F*-tests though. There is a robust variant of ANCOVA that can be implemented using R, and we'll discuss this in Section 13.8.

13.5 Conducting ANCOVA using SPSS Statistics

13.5.1 General procedure

The general procedure is much the same as for any linear model, so remind yourself of the steps for fitting a linear model (Chapter 9). Figure 13.4 shows a simpler overview of the process that highlights some of the specific issues for ANCOVA-style models. As with any analysis, begin by graphing the data and looking for and correcting sources of bias.

Figure 13.4 General procedure for analysis of covariance



13.5.2 Inputting data

We have already looked at the data (Table 13.1) and the data file (**Puppy Love.sav**). To remind you, the data file is set out like Table 13.1 and contains three columns: a coding variable called **Dose** (1 = control, 2 = 15 minutes, 3 = 30 minutes), a variable called **Happiness** containing the scores for the person's happiness, and a variable called **Puppy_love** containing the scores for love of puppies from 0 to 7. The 30 rows correspond to each person's scores on these three variables.

13.5.3 Testing the independence of the treatment variable and covariate

In Section 13.4.1, I mentioned that if the covariate and group means (independent variable) are independent then the interpretation of ANCOVA models is a lot more straightforward. In this case, the covariate is love of puppies, so we'd want to check that the mean level of love of puppies is roughly equal across the three puppy therapy groups by fitting a linear model with **Puppy_love** as the outcome and **Dose** as the predictor.

Fit a model to test whether love of puppies (our covariate) is independent of the dose of puppy therapy (our independent variable).

Output 13.3 shows that the main effect of dose is not significant, $F(2, 27) = 1.98$, $p = 0.16$, which shows that the average level of love of puppies was roughly the same in the three puppy therapy groups. In other words, the means for love of puppies in Table 13.2 are not significantly different across the control, 15-and 30-minute groups. This result is good news for using love of puppies as a covariate in the model.

Output 13.3

ANOVA

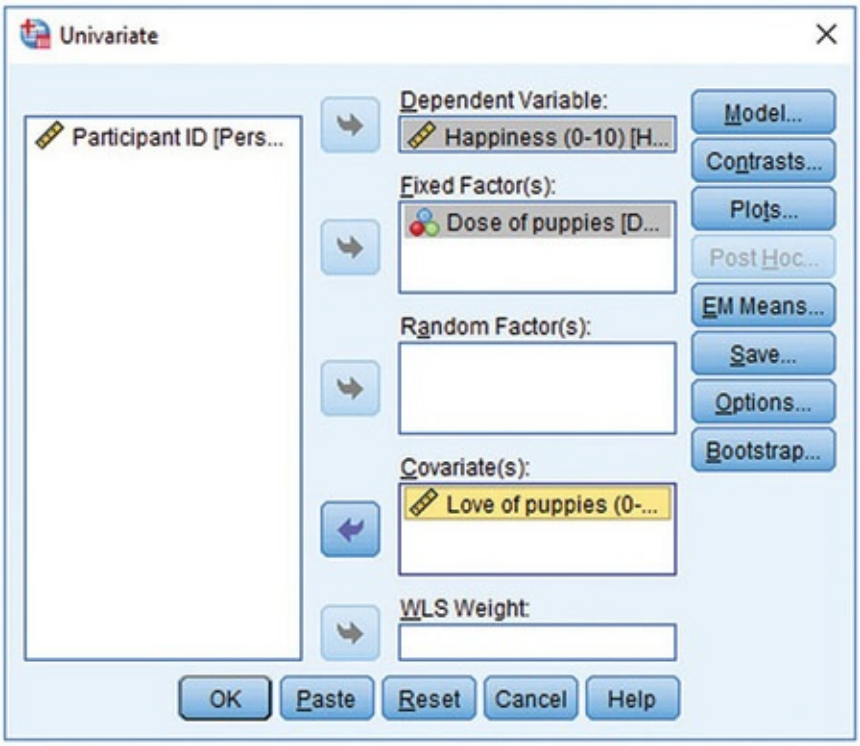
Love of puppies (0-7)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	12.769	2	6.385	1.979	.158
Within Groups	87.097	27	3.226		
Total	99.867	29			

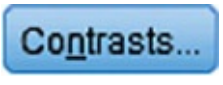
13.5.4 The main analysis

Most of the *General Linear Model* (GLM) procedures in SPSS Statistics contain the facility to include one or more covariates. For designs that don't involve repeated measures it is easiest to include covariates by selecting *Analyze*  *General Linear Model* to activate the dialog box in Figure 13.5. Drag the variable **Happiness** into the box labelled *Dependent Variable* (or click ), drag **Dose** into the box labelled *Fixed Factor(s)* and drag **Puppy_love** into the box labelled *Covariate(s)*.

Figure 13.5 Main dialog box for GLM univariate



13.5.5 Contrasts 

There are various dialog boxes that can be accessed from the main dialog box. If a covariate is selected, the *post hoc* tests are disabled because the tests that we used in the previous chapter are not designed for models that include covariates. However, comparisons can be done by clicking  to access the

Contrasts dialog box in Figure 13.6. You cannot enter codes to specify user-defined contrasts (but see SPSS Tip 13.1); instead you can select one of the standard contrasts that we met in Table 12.6. In this example, there was a control condition (coded as the first group), so a sensible set of contrasts would be simple contrasts comparing each experimental group to the control (this results in the same contrasts as dummy coding). Click the drop-down list (None) and select a type of contrast (in this case *Simple*) from this list. For simple contrasts you need to specify the reference category (i.e., the category against which all other groups are compared). By default the last category is used, which for our data is the 30-minute group. We need to change the reference category to be the control group, which is the first category (assuming that you coded control as 1). We make this change by selecting First. Having selected a contrast, click Change to register the selection. Figure 13.6 shows the completed dialog box. Click Continue to return to the main dialog box.

13.5.6 Other options

You can get a limited range of *post hoc* tests by clicking EM Means... to access the *Estimated Marginal Means* dialog box (see Figure 13.7). To specify *post hoc* tests, drag the independent variable (in this case **Dose**) from the box labelled *Estimated Marginal Means: Factor(s) and Factor Interactions* to the box labelled *Display Means for* (or click Save...). Once a variable has been transferred, you'll be able to select Compare main effects to activate the drop-down list (LSD(none)) of *post hoc* tests. The default is to perform a Tukey LSD *post hoc* test which makes no adjustment for multiple tests (and which I don't recommend). The other options are a Bonferroni *post hoc* test (recommended) and a **Šidák correction**, which is like the Bonferroni correction but is less conservative and so should be selected if you are concerned about the loss of power associated with Bonferroni. For this example we'll use the Šidák correction just for variety (we have used Bonferroni in previous examples). As well as producing *post hoc* tests for the **Dose** variable, the options we've selected will create a table of estimated marginal means for this variable: these are the

group means adjusted for the effect of the covariate. Click Continue.

Clicking Options... opens a dialog box containing the options described in Jane Superbrain Box 13.3. The most useful are (in my opinion) descriptive statistics, parameter estimates, residual plot and HC4 robust standard errors (see Figure 13.7).

Figure 13.6 Options for standard contrasts in GLM univariate

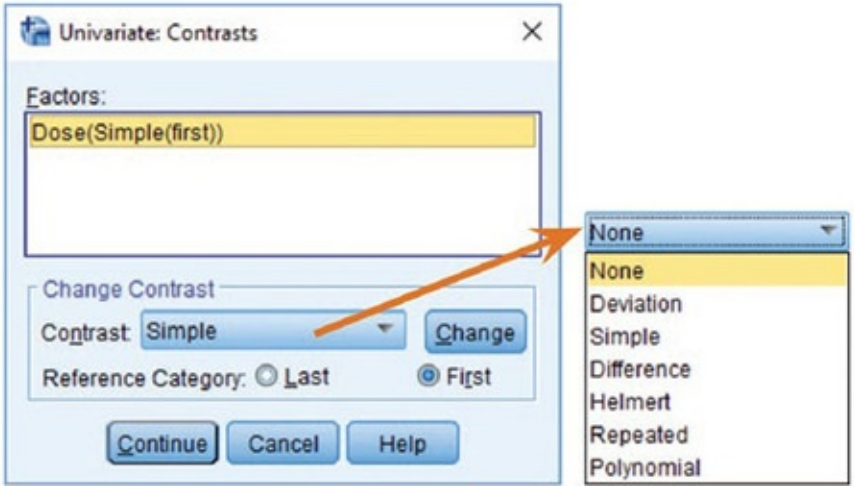
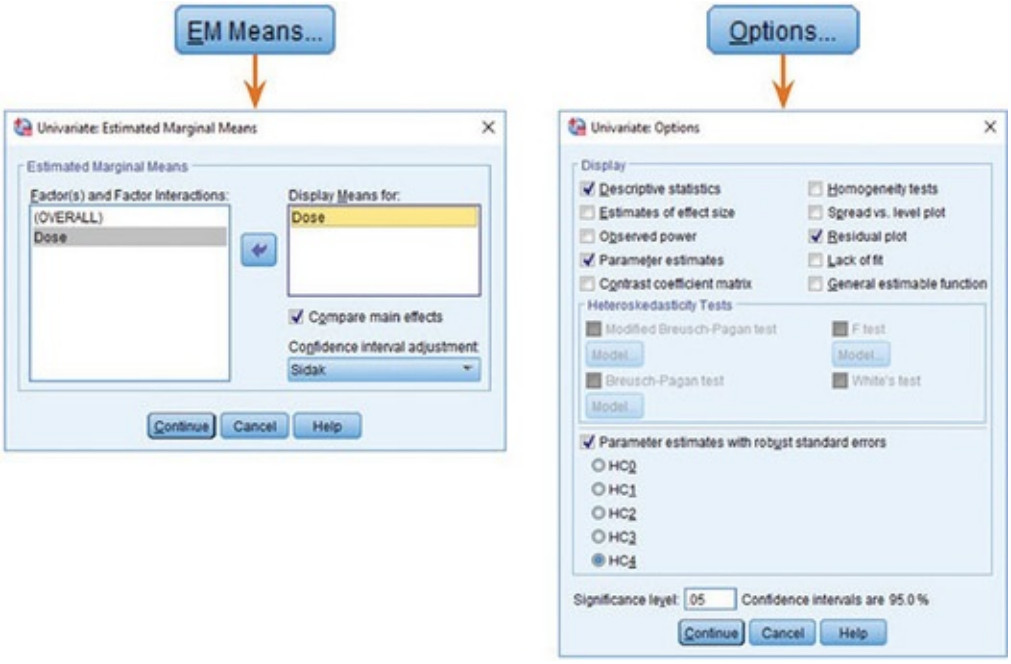
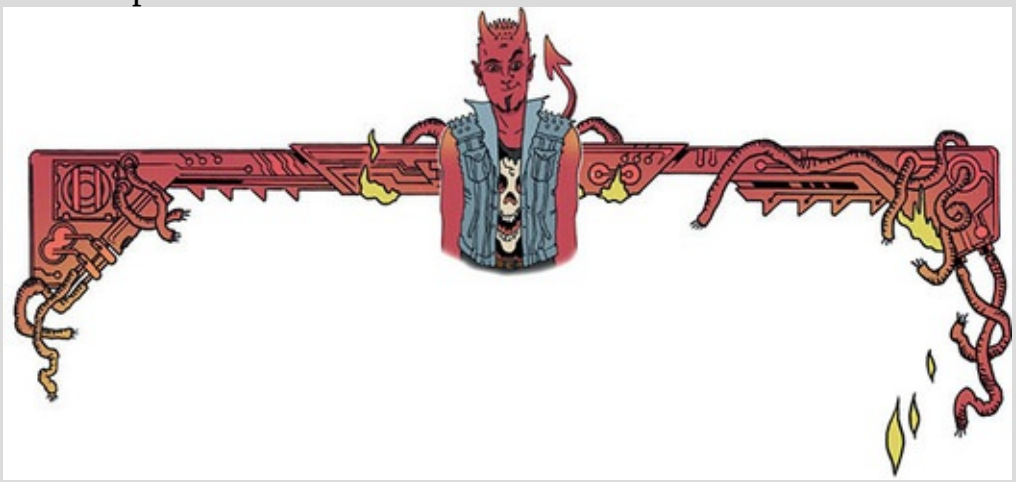


Figure 13.7 *Estimated marginal means* and *Options* dialog boxes for GLM univariate



SPSS Tip 13.1 Planned contrasts for ANCOVA



There is no option for specifying planned contrasts like we used in the previous chapter (see Section 12.6.2). However, these contrasts can be done if we fit the model using the regression menu. Imagine you chose some planned contrasts as in Chapter 12, in which the first contrast compared the control group to all doses of puppy therapy, and the second contrast then compared the 30-and 15-minute groups (see Section 12.4). We saw in Sections 12.4 and 12.6.2 that we specify these contrasts with codes. For the first contrast we discovered that an appropriate set of codes was −2 for the control group and then 1 for both the 30-and 15-minute groups. For the second contrast the codes were 0 for the control group, −1 for the 15-minute group and 1 for the 30-minute group (see Table 12.4). To do these contrasts when a covariate is included in the model, enter these values as two dummy variables. In other words, add a column called **Dummy1** in which every person in the control group has a value of −2 and all other participants have a value of 1. Add a second column called **Dummy2**, in which everyone in the control group has the value 0, everyone in the 15-minute group has the value −1 and those in the 30-minute group have a value of 1. The file **Puppy Love Contrast.sav** includes these dummy variables.

Output 13.4

Coefficients[a]

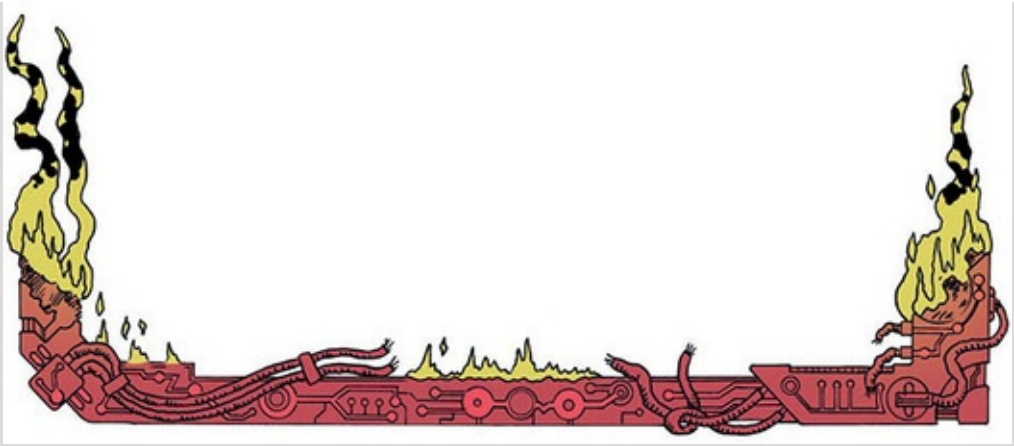
Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	3.657	.634		5.764	.000
	Love of puppies (0-7)	.260	.193	.246	1.345	.189
2	(Constant)	3.126	.625		5.002	.000
	Love of puppies (0-7)	.416	.187	.395	2.227	.035
	Dummy 1: puppies vs. control	.668	.240	.478	2.785	.010
	Dummy 2: 15 mins vs. 30 mins	.220	.406	.094	.541	.593

a. Dependent Variable: Happiness (0-10)

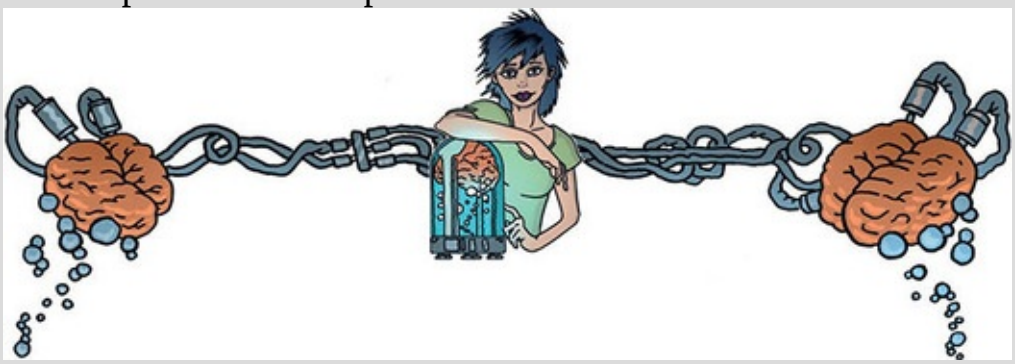
Run the analysis as described in Section 13.3. The model summary and ANOVA table for the model will be identical to Output 13.1 (because we've done the same thing as before; the only difference is how the model variance is subsequently broken down with the contrasts). The *b*-values for the dummy variables will be different than before because we've specified different contrasts. Output 13.4 shows the model parameters. The first dummy variable compares the control group with the 15-and 30-minute groups. As such, it compares the adjusted mean of the control group (2.93) with the average of the adjusted means for the 15-and 30-minute groups ((4.71+5.15)/2 = 4.93). The *b*-value for the first dummy variable should reflect the difference between these values: 4.93 − 2.93 = 2. We discovered in a rather complex and boring bit of Section 12.4.2 that this value gets divided by the number of groups within the contrast (i.e., 3) and so will be 2/3 = 0.67 (as in Output 13.4).[4] The associated *t*-statistic is significant ($p = 0.010$), indicating that the control group was significantly different from the combined adjusted mean of the puppy therapy groups.

4 The output contains the value 0.668 rather than 0.67. This difference is because we've rounded values to 2 decimal places in our caculations whereas SPSS retains all decimal places in its calculations.

The second dummy variable compares the 15-and 30-minute groups, and so the *b*-value should reflect the difference between the adjusted means of these groups: 5.15 − 4.71 = 0.44. In Section 12.4.2 we discovered that this value gets divided by the number of groups within the contrast (i.e., 2) and so will be 0.44/2 = 0.22 (as in Output 13.4). The associated *t*-statistic is not significant ($p = 0.593$), indicating that the 30-minute group did not produce significantly higher happiness than the 15-minute group after adjusting for love of puppies.



Jane Superbrain 13.3 Options for ANCOVA

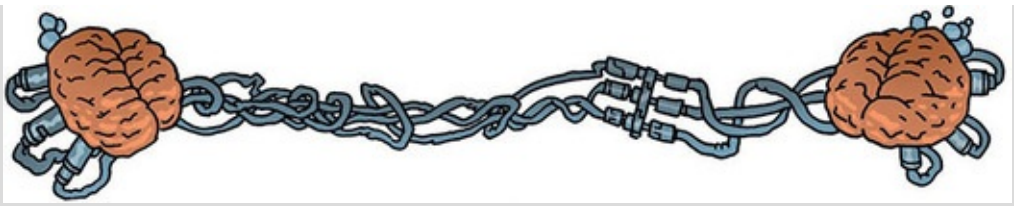


The remaining options in this dialog box are as follows:

- *Descriptive statistics*: This option produces a table of means and standard deviations for each group.
- *Estimates of effect size*: This option produces the value of partial eta squared (partial η^2) – see Section 13.10.
- *Observed power*: This option provides an estimate of the probability that the statistical test could detect the difference between the observed group means (see Section 2.9.7). This measure is pointless because if the *F*-test is significant then the probability that the effect was detected will, of course, be high. Likewise, if group differences were small, the observed power would be low. Do power calculations before the experiment is conducted, not after (see Section 2.9.8).
- *Parameter estimates*: This option produces a table of model parameters (*b*-values) and their tests of significance for the variables in the model (see Section 13.6.2).
- *Contrast coefficient matrix*: This option produces matrices of

the coding values used for any contrasts in the analysis, which is useful for checking which groups are being compared in which contrast.

- *Homogeneity tests*: This option produces Levene's test of the homogeneity of variance assumption (see Section 9.3). You'll have seen by now that I think this test needs to be taken with a pinch of salt.
- *Spread vs. level plot*: This option produces a chart that plots the mean of each group of a factor (x-axis) against the standard deviation of that group (y-axis). This plot is useful to check that there is no relationship between the mean and standard deviation. If a relationship exists then the data may need to be stabilized using a logarithmic transformation (see Chapter 6).
- *Residual plot*: This option produces a matrix scatterplot of all combinations of pairs of the following variables: observed values of the outcome, predicted values from the model, standardized residuals from the model. These plots can be used to assess the assumption of homoscedasticity. In particular, the plot of the standardized residuals against the predicted values from the model can be interpreted in a similar way to the zpred vs. zresid plot that we have discussed before.
- *Heteroskedasticity tests*: There are four tests for heteroscedasticity that you can select (two variants of the Breusch-Pagan test, White's test and an F-test). For the same reasons that I don't recommend Levene's test, I also don't recommend these (that is, because they are significance tests your decisions based on them will be confounded by your sample size).
- *Parameter estimates with robust standard errors*: This produces one of 5 methods (HC0 to HC4) to estimate standard errors (and, therefore, confidence intervals) for the model parameters that are robust to heteroscedasticity. These methods are described clearly in Hayes and Cai (2007). In short, HC3 has been shown to outperform HC0 to HC2 (Long & Ervin, 2000) but HC4 outperforms HC3 in some circumstances (Cribari-Neto, 2004). Basically choose HC3 or HC4.



Oditi's Lantern ANCOVA

'I, Oditi, have discovered that covariates give us greater control. I like control, especially controlling people's minds and making them worship me, erm, I mean controlling people's minds for the benevolent purpose of helping them to seek truth and personal enlightenment. As long as they are personally enlightened to worship me. In any case, stare into my lantern to discover more about using covariates and ANCOVA.'

13.5.7 Bootstrapping and plots

There are other options available from the main dialog box. For example, if you have several independent variables you can plot them against each other (which is useful for interpreting interaction effects – see Section 14.7). There's also the Bootstrap... button, which you can use to activate bootstrapping. Selecting this option will bootstrap confidence intervals around the estimated marginal means, parameter estimates (*b*-values) and *post hoc* tests, but not the main *F*-statistic.

Select the options described in Section 6.12.3 and click OK in the

main dialog box to run the analysis.

13.6 Interpreting ANCOVA

13.6.1 What happens when the covariate is excluded?

Output 13.5 shows (for illustrative purposes) the ANOVA table for these data when the covariate is not included. It is clear from the significance value, which is greater than 0.05, that puppy therapy seems to have no significant effect on happiness. Note that the total amount of variation in happiness (SS_T) was 110.97 (*Corrected Total*), of which the therapy condition accounted for 16.84 units (SS_M), while 94.12 were unexplained (SS_R).

Output 13.5

Tests of Between-Subjects Effects

Dependent Variable: Happiness (0-10)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	16.844[a]	2	8.422	2.416	.108
Intercept	535.184	1	535.184	153.522	.000
Dose	16.844	2	8.422	2.416	.108
Error	94.123	27	3.486		
Total	683.000	30			
Corrected Total	110.967	29			

a. R Squared = .152 (Adjusted R Squared = .089)

Output 13.6

Tests of Between-Subjects Effects

Dependent Variable: Happiness (0-10)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	31.920[a]	3	10.640	3.500	.030
Intercept	76.069	1	76.069	25.020	.000
Puppy_love	15.076	1	15.076	4.959	.035
Dose	25.185	2	12.593	4.142	.027
Error	79.047	26	3.040		
Total	683.000	30			
Corrected Total	110.967	29			

a. R Squared = .288 (Adjusted R Squared = .205)

Fit the model without the covariate to see whether the three groups

differ in their levels of happiness.

13.6.2 The main analysis

The format of the ANOVA table in Output 13.6 is largely the same as without the covariate, except that there is an additional row of information about the covariate (**Puppy_love**). Looking first at the significance values, the covariate significantly predicts the dependent variable (p = 0.035, which is less than 0.05). Therefore, the person's happiness is significantly influenced by their love of puppies. What's more interesting is that when the effect of love of puppies is removed, the effect of puppy therapy is significant (p = 0.027, which is less than 0.05). The amount of variation accounted for by puppy therapy has increased to 25.19 units and the unexplained variance (SS_R) has been reduced to 79.05 units. Notice that SS_T has not changed; all that has changed is how that total variation is partitioned.[5]

[5] I am often asked what the *Corrected Model* represents in this table. It is the fit of the model overall (i.e., the model containing the intercept, **Puppy_love** and **Dose**). Note that the SS of 31.92, *df* of 3, F of 3.5 and p of 0.03 are identical to the values in Output 13.1 (model 2), which tested the overall fit of this model when we ran the analysis as a regression.

This example illustrates how covariates can help us to exert stricter experimental control by taking account of confounding variables to give us a 'purer' measure of effect of the experimental manipulation. Looking back at the group means from Table 13.1, you might think that the significant F-statistic reflects a difference between the control group and the two experimental groups – because the 15-and 30-minute groups have very similar means (4.88 and 4.85) whereas the control group mean is much lower at 3.22. However, we can't use these group means to interpret the effect because they have not been adjusted for the effect of the covariate. These original means tell us nothing about the group differences reflected by the significant F. Output 13.7 gives the adjusted values of the group means (which we calculated in Section 13.3), and we use these values for interpretation (this is why we selected *Display Means for* in Section 13.5.6). From these adjusted means you can see that happiness increased across

the three doses.

Output 13.7

Estimates

Dependent Variable: Happiness (0–10)

Dose of puppies	Mean	Std. Error	95% Confidence Interval		Bootstrap for Mean[gn]			
							BCa 95% Confidence Interval	
			Lower Bound	Upper Bound	Bias	Std. Error	Lower	Upper
Control	2.926[a]	.596	1.701	4.152	.030	.446	2.111	4.125
15 mins	4.712[a]	.621	3.436	5.988	.033[go]	.392[go]	3.988[go]	5.620[go]
30 mins	5.151[a]	.503	4.118	6.184	.041	.651	3.923	6.771

a. Covariates appearing in the model are evaluated at the following values: Love of puppies (0–7) = 2.73.

gn. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

go. Based on 999 samples

Output 13.8 shows the parameter estimates selected in the *Options* dialog box and their bootstrapped confidence intervals and *p*-values (bottom table). These estimates result from **Dose** being coded using two dummy coding variables. The dummy variables are coded with the last category (the category coded with the highest value in the data editor, in this case the 30-minute group) as the reference category. This reference category (labelled Dose=3 in the output) is coded with a 0 for both dummy variables (see Section 12.2 for a reminder of how dummy coding works). Dose=2, therefore, represents the difference between the group coded as 2 (15 minutes) and the reference category (30 minutes), and Dose=1 represents the difference between the group coded as 1 (control) and the reference category (30 minutes). The *b*-values represent the differences between the adjusted means in Output 13.7 and the significances of the *t*-tests tell us whether these adjusted group means differ significantly. The *b* for Dose=1 in Output 13.8 is the difference between the adjusted means for the control group and the 30-minute group, 2.926 − 5.151 = −2.225, and the *b* for Dose=2 is the difference between the adjusted means for the 15-minute group and the 30-minute group, 4.712 − 5.151 = −0.439.

The degrees of freedom for the *t*-test of the *b*-values are $N - k - 1$ (see Section 9.2.5), in which N is the total sample size (in this case 30) and k is the number of predictors (in this case 3, the two dummy variables and the covariate). For these data, $df = 30 - 3 - 1 = 26$. Based on the bootstrapped significance and confidence intervals (remember you'll get different values than me because of how bootstrapping works), we could conclude that the 30-minute group differs significantly from the control group, $p = 0.021$ (Dose=1 in the table), but not from the 15-minute group, $p = 0.558$, (Dose=2 in the table).

The final thing to note is the value of *b* for the covariate (0.416), which is the same as in Output 13.2 (when we ran the analysis through the regression menu). This value tells us that if love of puppies increases by one unit, then the person's happiness should increase by just under half a unit (although there is nothing to suggest a causal link between the two); because the coefficient is positive we

know that as love of puppies increases so does happiness. A negative coefficient would mean the opposite: as one increases, the other decreases.

Output 13.9 repeats the parameter estimates from Output 13.8 but with standard errors, *p*-values and confidence intervals robust to heteroscedasticity (the HC4 estimates that we asked for). We can interpret the effects for **Dose** in the same way as for the regular and Bootstrap *p*-values and confidence intervals. For the effect of puppy love, the HC4 robust confidence interval and *p*-value supports the conclusion from the non-robust model: the *p*-value is 0.038, which is less than 0.05, and the confidence interval does not contain zero (0.025, 0.807). However, the bootstrap confidence interval (Output 13.8) contradicts this conclusion because it contains zero (−0.023, 0.698) and has a $p = 0.052$ (again, we're reminded of how daft it is to have a threshold that yields such opposing conclusions from such small differences in a value).

Output 13.8

Parameter Estimates

Dependent Variable: Happiness (0-10)

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	4.014	.611	6.568	.000	2.758	5.270
Puppy_love	.416	.187	2.227	.035	.032	.800
[Dose=1]	-2.225	.803	-2.771	.010	-3.875	-.575
[Dose=2]	-.439	.811	-.541	.593	-2.107	1.228
[Dose=3]	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

Bootstrap for Parameter Estimates

Dependent Variable: Happiness (0-10)

Parameter	B	Bootstrap[a]				
		Bias	Std. Error	Sig. (2-tailed)	BCa 95% Confidence Interval	
					Lower	Upper
Intercept	4.014	.091[b]	.843[b]	.003[b]	1.969[b]	5.949[b]
Puppy_love	.416	-.029[b]	.202[b]	.052[b]	-.023[b]	.698[b]
[Dose=1]	-2.225	-.011[b]	.760[b]	.021[b]	-3.753[b]	-.823[b]
[Dose=2]	-.439	-.008[b]	.745[b]	.558[b]	-1.937[b]	.935[b]
[Dose=3]	0	0[b]	0[b]		.[b]	.[b]

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

b. Based on 999 samples

Output 13.9

Parameter Estimates with Robust Standard Errors

Dependent Variable: Happiness (0–10)

Parameter	B	Robust Std. Error[a]	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	4.014	.805	4.989	.000	2.360	5.668
Puppy_love	.416	.190	2.187	.038	.025	.807
[Dose=1]	-2.225	.690	-3.226	.003	-3.642	-.807
[Dose=2]	-.439	.695	-.632	.533	-1.868	.990
[Dose=3]	0[b]	.	.	.	.	.

a. HC4 method

b. This parameter is set to zero because it is redundant.

Output 13.10

Contrast Results (K Matrix)

Dose of puppies Simple Contrast[a]			Dependent Variable Happiness (0–10)
Level 2 vs. Level 1	Contrast Estimate		1.786
	Hypothesized Value		0
	Difference (Estimate – Hypothesized)		1.786
	Std. Error		.849
	Sig.		.045
	95% Confidence Interval for Difference	Lower Bound	.040
		Upper Bound	3.532
Level 3 vs. Level 1	Contrast Estimate		2.225
	Hypothesized Value		0
	Difference (Estimate – Hypothesized)		2.225
	Std. Error		.803
	Sig.		.010
	95% Confidence Interval for Difference	Lower Bound	.575
		Upper Bound	3.875

a. Reference category = 1

Output 13.11

Pairwise Comparisons

Dependent Variable: Happiness (0-10)

(I) Dose of puppies	(J) Dose of puppies	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
Control	15 mins	-1.786	.849	.130	-3.953	.381
	30 mins	-2.225*	.803	.030	-4.273	-.177
15 mins	Control	1.786	.849	.130	-.381	3.953
	30 mins	-.439	.811	.932	-2.509	1.631
30 mins	Control	2.225*	.803	.030	.177	4.273
	15 mins	.439	.811	.932	-1.631	2.509

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Sidak.

Bootstrap for Pairwise Comparisons

Dependent Variable: Happiness (0-10)

(I) Dose of puppies	(J) Dose of puppies	Mean Difference (I-J)	Bootstrap[a] Bias	Std. Error	Sig. (2-tailed)	BCa 95% Confidence Interval Lower	Upper
Control	15 mins	-1.786	-.003[b]	.535[b]	.003[b]	-2.778[b]	-.765[b]
	30 mins	-2.225	-.011	.760	.021	-3.752	-.832
15 mins	Control	1.786	.003[b]	.535[b]	.003[b]	.663[b]	2.879[b]
	30 mins	-.439	-.008[b]	.745[b]	.558[b]	-1.937[b]	.935[b]
30 mins	Control	2.225	.011	.760	.021	.686	3.923
	15 mins	.439	.008[b]	.745[b]	.558[b]	-.938[b]	1.945[b]

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

b. Based on 999 samples

13.6.3 Contrasts

Output 13.10 shows the result of the contrast analysis specified in Figure 13.6 and compares level 2 (15 minutes) against level 1 (control) as a first comparison, and level 3 (30 minutes) against level 1 (control) as a second comparison. The group differences are displayed: a difference value, standard error, significance value and 95% confidence interval. These results show that both the 15-minute group (contrast 1, $p = 0.045$) and 30-minute group (contrast 2, $p = 0.010$) had significantly different happiness compared to the control group (note that contrast 2 is identical to the parameter for Dose=1 in the previous section). Output 13.11 shows the results of the Šidák corrected *post hoc* comparisons that were requested in Section 13.5.6. The bottom table shows the bootstrapped significance and confidence intervals for these tests and because these will be robust we'll interpret this table (again, remember your values will differ because of how bootstrapping works). There is a significant difference between the control group and both the 15- ($p = 0.003$) and 30-minute ($p = 0.021$) groups. The 30-and 15-minute groups did not significantly differ ($p = 0.558$). It is interesting that the significant difference between the 15-minute and control groups when bootstrapped ($p = 0.003$) is not present for the normal *post hoc* tests ($p = 0.130$). This anomaly could reflect properties of the data that have biased

the non-robust version of the *post hoc* test.

13.6.4 Interpreting the covariate

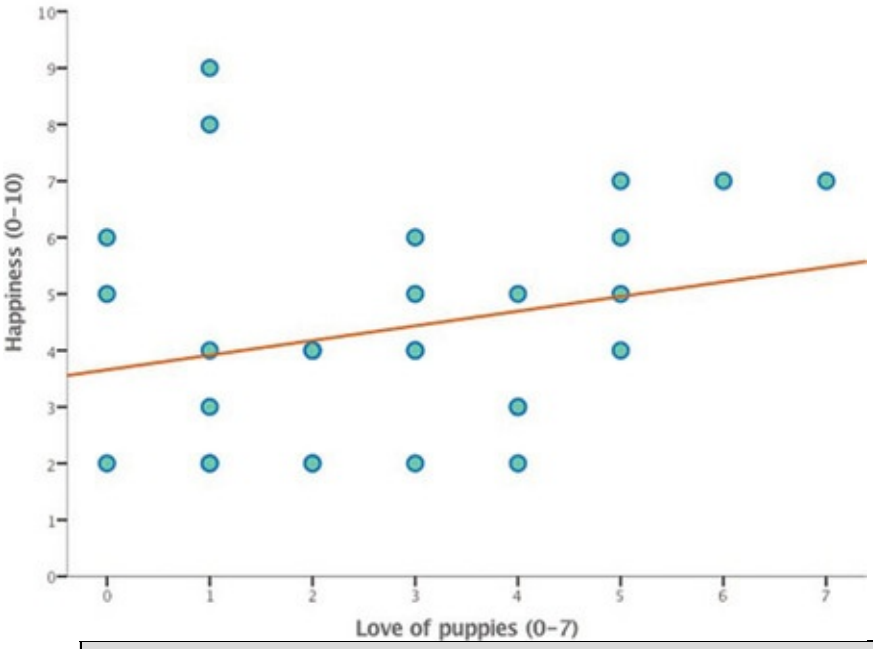
I've already mentioned that the parameter estimates (Output 13.8) tell us how to interpret the covariate: the sign of the *b*-value tells us the direction of the relationship between the covariate and outcome variable. For these data the *b*-value was positive, indicating that as the love of puppies increases, so does the participant's happiness. Another way to discover the same thing is to draw a scatterplot of the covariate against the outcome.

SELF TEST

Produce a scatterplot of love of puppies (horizontal axis) against happiness (vertical axis).

Figure 13.8 confirms that the effect of the covariate is that as love of puppies increases, so does the participant's happiness (as shown by the slope of the line).

Figure 13.8 Scatterplot of happiness against love of puppies



Labcoat Leni's Real Research 13.1 Space invaders

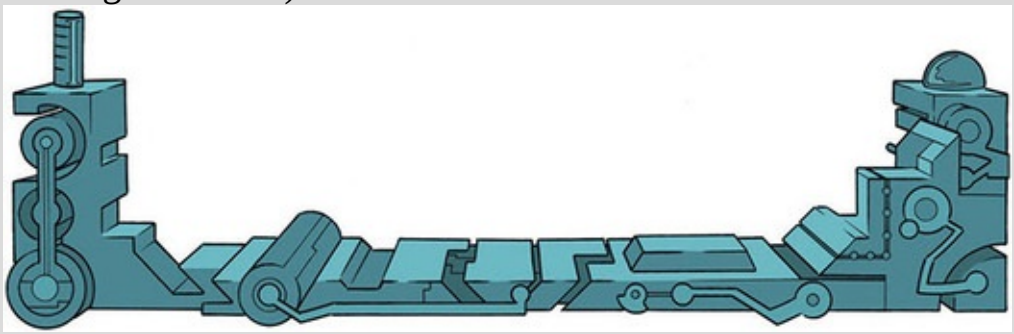
Muris, P., *et al.* (2008). *Child Psychiatry and Human Development, 39*(4), 469–480.

Anxious people tend to interpret ambiguous information in a negative way. For example, being highly anxious myself, if I overheard a student saying 'Andy Field's lectures are really *different*,' I would assume that 'different' meant rubbish, but it could also mean 'refreshing' or 'innovative'. Muris, Huijding, Mayer, and Hameetman (2008) addressed how these interpretational biases develop in children. Children imagined that they were astronauts who had discovered a new planet. They were given scenarios about their time on the planet (e.g., 'On the street, you encounter a spaceman. He has a toy handgun and he fires at you …') and the child had to decide whether a positive ('You laugh: it is a water pistol and the weather is fine anyway') or negative ('Oops, this hurts! The pistol produces a red beam which burns your skin!') outcome occurred. After each response the child was told whether their choice was correct. Half of the children were *always* told that the negative interpretation was correct, and the reminder were told that the positive interpretation was correct.

Over 30 scenarios children were trained to interpret their experiences on the planet as negative or positive. Muris *et al.* then measured interpretational biases in everyday life to see whether the training had created a bias to interpret things negatively. In doing so, they could ascertain whether children might learn interpretational biases through feedback (e.g., from parents).

The data from this study are in the file **Muris et al (2008).sav**. The independent variable is **Training** (positive or negative) and the outcome is the child's interpretational bias score (**Interpretational_Bias**) – a high score reflects a tendency to interpret situations negatively. It is important to adjust for the **Age**

and **Gender** of the child and also their natural anxiety level (which they measured with a standard questionnaire of child anxiety called the **SCARED**) because these things affect interpretational biases also. Labcoat Leni wants you to fit a model to see whether **Training** significantly affected children's **Interpretational_Bias** using **Age, Gender** and **SCARED** as covariates. What can you conclude? Answers are on the companion website (or look at pages 475–476 in the original article).



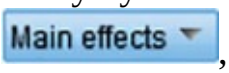
13.7 Testing the assumption of homogeneity of regression slopes

Remember that the assumption of homogeneity of regression slopes means that the relationship between the covariate and outcome variable (in this case **Puppy_love** and **Happiness**) should be similar at different levels of the predictor variable (in this case in the three **Dose** groups). Figure 13.3 shows that the relationship between **Puppy_love** and **Happiness** looks comparable in the 15-minute and control groups, but seems different in the 30-minute group.

To test the assumption of homogeneity of regression slopes we need to refit the model but customize it to include the interaction between the covariate and categorical predictor. Access the main dialog box as before and place the variables in the same boxes as before (the finished dialog box should look like Figure 13.5). To customize the model, click 

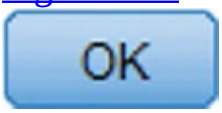
to access the dialog box in Figure 13.9 and select 

. The variables specified in the main dialog box are listed on the left-hand side. We need a model that includes the interaction between the covariate and grouping variable. To test this interaction term it's important to also include the main effects otherwise variance in the outcome (happiness) may be attributed to the interaction term that would otherwise be attributed to the main effects. To begin with, then, select **Dose** and

Puppy_love (you can select both simultaneously by holding down *Ctrl*, or *Cmd* on a Mac), change the drop-down menu to 

, and click  to transfer the main effects of **Dose** and **Puppy_love** to the box labelled *Model*. Next specify the interaction term by selecting **Dose** and **Puppy_love** simultaneously (as just described), change the drop-down menu to 

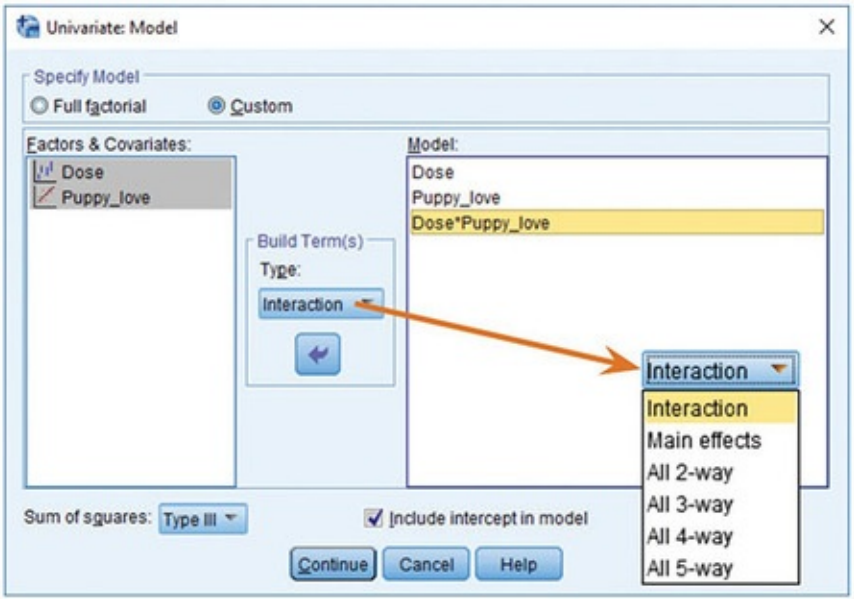
and click  to transfer the interaction of **Dose** and **Puppy_love** to the box labelled *Model*. The finished dialog box should look like Figure 13.9. Click 

to return to the main dialog box and 

to run the analysis.

Output 13.11 shows the main summary table for the model including the interaction term. The effects of the dose of puppy therapy and love of puppies are still significant, but so is the covariate by outcome interaction (**Dose** × **Puppy_love**), implying that the assumption of homogeneity of regression slopes is not realistic (p = 0.028). Although this finding is not surprising given the pattern of relationships shown in Figure 13.3, it raises concerns about the main analysis.

Figure 13.9 *Model* dialog box for GLM univariate



Output 13.12

Tests of Between-Subjects Effects

Dependent Variable: Happiness (0-10)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	52.346[a]	5	10.469	4.286	.006
Intercept	53.542	1	53.542	21.921	.000
Dose	36.558	2	18.279	7.484	.003
Puppy_love	17.182	1	17.182	7.035	.014
Dose * Puppy_love	20.427	2	10.213	4.181	.028
Error	58.621	24	2.443		
Total	683.000	30			
Corrected Total	110.967	29			

a. R Squared = .472 (Adjusted R Squared = .362)

Cramming Sam's Tips Covariates

- When the linear model is used to compare several means adjusted for the effect of one or more other variables (called *covariates*) it can be referred to as analysis of covariance (ANCOVA).
- Before the analysis check that the covariate(s) are independent

of any independent variables by seeing whether those independent variables predict the covariate (i.e., the covariate should not differ across groups).

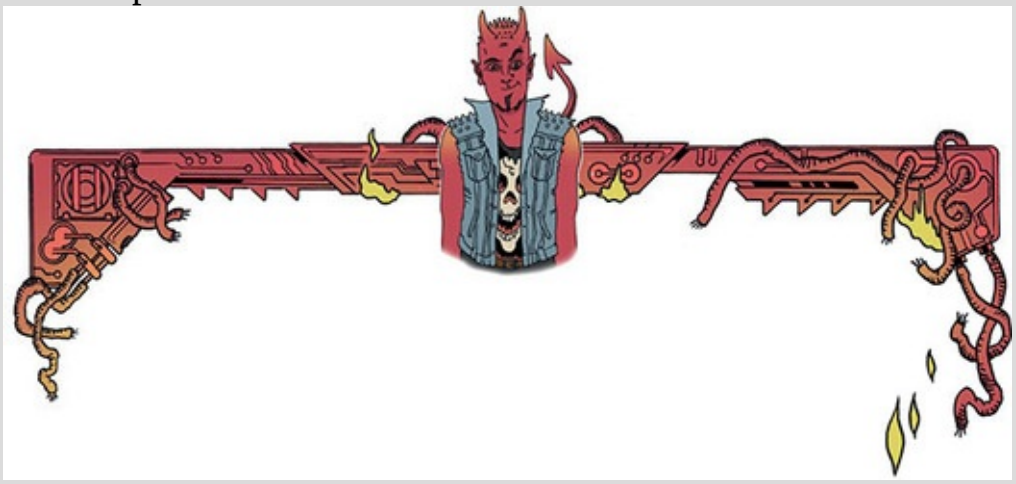
- In the table labelled *Tests of Between-Subjects Effects*, assuming you're using an alpha of 0.05, look to see if the value in the column labelled *Sig.* is below 0.05 for both the covariate and the independent variable. If it is for the covariate then this variable has a significant relationship to the outcome variable; if it is for the independent variable then the means (adjusted for the effect of the covariate) are significantly different across categories of this variable.
- If you have generated specific hypotheses before the experiment use planned contrasts; if not, use *post hoc* tests.
- For parameters and *post hoc* tests, look at the columns labelled *Sig.* to discover if your comparisons are significant (they will be if the significance value is less than 0.05). Use bootstrapping to get robust versions of these tests.
- In addition to the assumptions in Chapter 6, test for *homogeneity of regression slopes* by customizing the model to look at the independent variable × covariate interaction.

13.8 Robust ANCOVA

We have already looked at robust confidence intervals and *p*-values for the model parameters that were computed using bootstrapping and heteroscedasticity robust standard errors (Section 13.6.2). In addition, the companion website contains a syntax file (**robustANCOVA.sps**) for running a robust variant of ANCOVA (*ancboot*) that works on trimmed means and is described by Wilcox (2017). We need the *Essentials for R* plugin and *WRS2* package installed (Section 4.13). This test is limited to the situation where the independent variable (the categorical predictor) has two categories and there is one covariate. But it does enable you to ignore assumptions and get on with your life. Because this syntax only works when you have two groups, I have provided a data file called **PuppiesTwoGroup.sav**, which contains the example data for

this chapter but excluding the 15-minute condition, so it compares the control (no puppies) with the 30-minute group (**Dose**), and has the scores for the love of puppies covariate too (**Puppy_love**). The syntax to run the robust test is as follows: BEGIN PROGRAM R.
library(WRS2)
mySPSSdata = spssdata.GetDataFromSPSS(factorMode = "labels")
ancboot(Happiness ~ Dose + Puppy_love, data = mySPSSdata, tr = 0.2, nboot = 1000) END PROGRAM.
Select and run these five lines of syntax (see SPSS Tip 10.3). As Output 13.13 shows, the test works by identifying values of the covariate for which the relationship between the covariate and outcome are comparable in the two groups. In this example it identifies five values of **Puppy_love** (2, 3, 5, 6, and 8) for which the relationship between love of puppies and happiness is comparable. At each of these design points, we're told the number of cases for the two groups (*n1* and *n2*) that have a value of the covariate (**Puppy_love**) close to these design points (not exactly *x*, but close to it). Based on these two samples, trimmed means (20% by default) are computed and the difference between them tested. This difference is stored in the column *Diff* along with the boundaries of the associated 95% bootstrap confidence interval (corrected to control for doing five tests) in the next two columns. The test statistic comparing the difference is in the column *statistic*, with its *p*-value in the final column. Output 13.12 shows no significant differences between trimmed means for any of the design points (all *p*-values are greater than 0.05).

SPSS Tip 13.2 Robust ANCOVA



The function *ancboot* takes the form:

ancboot(outcome ~ categorical predictor + covariate, data =

```
        mySPSSdata, tr = 0.2, nboot = 1000)
```

In our example we would replace outcome with **Happiness,** categorical predictor with **Dose** and covariate with **Puppy_love**. *tr* relates to the amount of trimming to the data (by default 0.2, or 20%, so change it if you want to use a different proportion). *nboot* refers to the number of bootstrap samples which I have set to a reasonable 1000, but feel free to increase it.

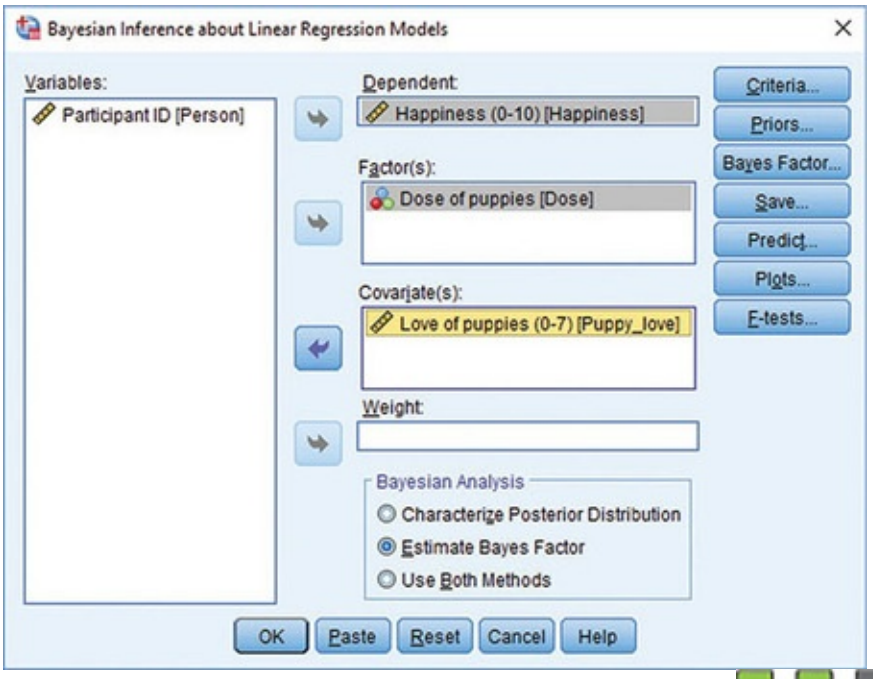
Output 13.13

	n1	n2	diff	lower CI	upper CI	statistic	p-value
Puppy_love = 2	22	13	-1.0873	-3.1547	0.9801	-1.6952	0.098
Puppy_love = 3	27	15	-1.0719	-2.8097	0.6659	-1.9881	0.058
Puppy_love = 5	30	22	-0.6508	-2.4220	1.1204	-1.1843	0.250
Puppy_love = 6	23	21	-0.9846	-3.3281	1.3589	-1.3542	0.207
Puppy_love = 8	12	13	-1.5278	-4.3223	1.2667	-1.7622	0.119

13.9 Bayesian analysis with covariates

Because the model we have fitted is a linear model with a categorical predictor and a continuous predictor, you can use what you learned in Section 9.13 to run a Bayesian regression. You would need to manually create dummy variables (as in the file **Puppy Love Dummy.sav**) and drag these to the box labelled *Factor(s)* and drag **Puppy_Love** to the box labelled *Covariate(s)* see Figure 13.10. You would interpret in the same way as the model we fitted in Section 9.13.

Figure 13.10



13.10 Calculating the effect size

In the previous chapter we used eta squared, η^2, as an effect size measure when comparing means (Section 12.10). When we include a covariate too we have more than one effect and we could calculate eta squared for each effect. We can also use an effect size measure called **partial eta squared (partial η2)**. This differs from eta squared in that it looks not at the proportion of total variance that a variable explains, but at the proportion of variance that a variable explains that *is not explained by other variables in the analysis*. Let's look at this with our example. Suppose we want to know the effect size of the dose of puppy therapy. Partial eta squared is the proportion of variance in happiness that the dose of puppy therapy shares that is not attributed to love of puppies (the covariate). If you think about the variance that the covariate cannot explain, there are two sources: it cannot explain the variance attributable to the dose of puppy therapy, $SS_{\text{puppy therapy}}$, and it cannot explain the error variability, SS_R. Therefore, we use these two sources of variance instead of the total variability, SS_T, in the calculation. The difference between eta squared and partial eta squared is illustrated by comparing the following two equations:

$$\eta^2 = \frac{SS_{\text{Effect}}}{SS_{\text{Total}}} \quad (13.6)$$

$$\text{Partial } \eta^2 = \frac{SS_{\text{Effect}}}{SS_{\text{Effect}} + SS_{\text{Residual}}} \quad (13.7)$$

SPSS Statistics will produce partial eta squared for us (see Jane Superbrain Box

13.3), but to illustrate its calculation look at equation (13.8), where we use the sums of squares in Output 13.6 for the effect of dose (25.19), the covariate (15.08) and the error (79.05):

$$\text{Partial } \eta^2_{\text{Dose}} = \frac{SS_{\text{Dose}}}{SS_{\text{Dose}} + SS_{\text{Residual}}} = \frac{25.19}{25.19 + 79.05} = \frac{25.19}{104.24} = 0.24$$
$$\text{Partial } \eta^2_{\text{Puppy\_love}} = \frac{SS_{\text{Puppy\_love}}}{SS_{\text{Puppy\_love}} + SS_{\text{Residual}}} = \frac{15.08}{15.08 + 79.05} = \frac{15.08}{94.13} = 0.16 \quad (13.8)$$

These values show that **Dose** explained a bigger proportion of the variance not attributable to other variables than **Puppy_love**.

Rerun the analysis but select ☑ Estimates of effect size in Figure 13.7. Do the values of partial eta squared match the ones we have just calculated?

You can also use omega squared (ω^2). However, as we saw in Section 12.8, this measure can be calculated only when we have equal numbers of participants in each group (which is not the case in this example). So, we're a bit stumped! Not all is lost, though, because, as I've said many times already, the overall effect size is not nearly as interesting as the effect size for more focused comparisons. These are easy to calculate because we selected to see the model parameters (see Output 13.8) and so we have *t*-statistics for the covariate and comparisons between the 15-and 30-minute groups and the control and 30-minute group. These *t*-statistics have 26 degrees of freedom (see Section 13.6.1). We can use the same equation as in Section 10.9.5:[6]

6 Strictly speaking, we should use a slightly more elaborate procedure when groups are unequal. It's a bit beyond the scope of this book, but Rosnow, Rosenthal, and Rubin (2000) give a very clear account.

$$r_{\text{Contrast}} = \sqrt{\frac{t^2}{t^2 + df}} \quad (13.9)$$

Therefore, we get (using *t* from Output 13.8) values of 0.40 for the covariate, and 0.48 and 0.11 respectively for the comparison of the 30-minute group and control, and the 15-and 30-minute groups:

$$r_{\text{Covariate}} = \sqrt{\frac{2.23^2}{2.23^2 + 26}} = \sqrt{\frac{4.97}{30.97}} = 0.40$$
$$r_{30\text{ mins vs. control}} = \sqrt{\frac{(-2.77)^2}{(-2.77)^2 + 26}} = \sqrt{\frac{7.67}{33.67}} = 0.48 \quad (13.10)$$
$$r_{30\text{ vs. }15\text{ mins}} = \sqrt{\frac{(-0.54)^2}{(-0.54)^2 + 26}} = \sqrt{\frac{0.29}{26.29}} = 0.11$$

For the effect of the covariate and the difference between the 30-minute and control groups the effects are not only statistically significant but also substantive in size. The difference between the 30-and 15-minute groups was a fairly small effect.

13.11 Reporting results 

When using covariates you can report the model in much the same way as any other. For the covariate and the experimental effect give details of the *F*-statistic and the degrees of freedom from which it was calculated. In both cases, the *F*-statistic was derived from dividing the mean squares for the effect by the mean squares for the residual. Therefore, the degrees of freedom used to assess the *F*-statistic are the degrees of freedom for the effect of the model (df_M = 1 for the covariate and 2 for the experimental effect) and the degrees of freedom for the residuals of the model (df_R = 26 for both the covariate and the experimental effect) – see Output 13.6. The correct way to report the main findings would be:

> ✓ The covariate, love of puppies, was significantly related to the participant's happiness, $F(1, 26) = 4.96$, $p = 0.035$, $r = 0.40$. There was also a significant effect of puppy therapy on levels of happiness after controlling for the effect of love of puppies, $F(2, 26) = 4.14$, $p = 0.027$, partial $\eta^2 = 0.24$.

We can also report some contrasts (see Output 13.8):

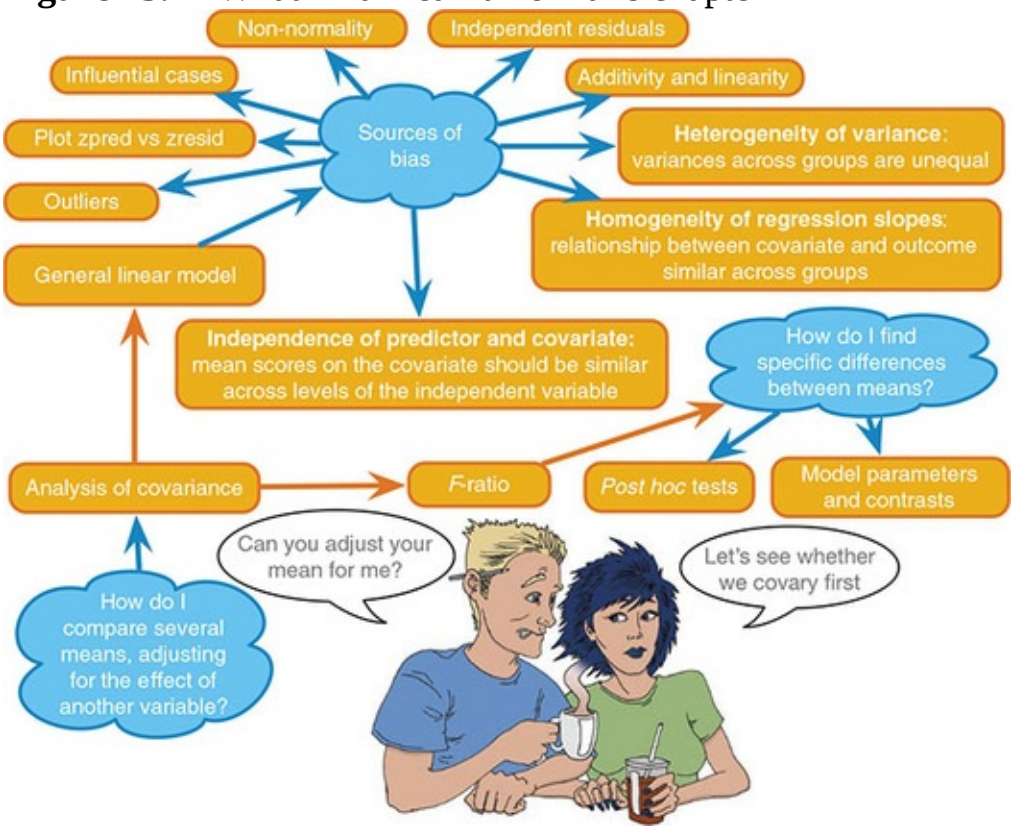
> ✓ Planned contrasts revealed that having 30 minutes of puppy therapy significantly increased happiness compared to having a control, $t(26) = -2.77$, $p = 0.01$, $r = 0.48$, but not compared to having 15 minutes, $t(26) = -0.54$, $p = 0.59$, $r = 0.11$.

13.12 Brian's attempt to woo Jane 

The encounter in Blow Your Speakers had been beyond weird. Jane felt terrible. This Brian guy was so nice to her, and she'd just told him where to go – again! It

had been easy to dismiss Brian at first, he'd seemed like a loser, a waste of her time. But there was more to him than that: he'd been working hard to learn statistics, and he'd made impressive progress. She liked how awkward he was around her, and how he always defaulted to talking stats. It was endearing. It could derail her research, though, and he could never know about that. She was a monster, and if he found out the truth it would be another let-down. Best to keep her distance.

Figure 13.11 What Brian learnt from this chapter



The phone rang. It was her brother, Jake. She loved and admired Jake like no one else. Until he left home, he'd been her sanity in the madhouse that they grew up in. Their parents, both highly successful academics, were at home only long enough to pile the pressure on them both to succeed. Jane reacted by spending her youth in books, in a futile pursuit of their attention. Every set of straight As was met with 'these are just a step towards the exams that really matter, you'll need to up your game'. She was tired of trying to impress them. Jake was her opposite – he'd realized early on that he could never win. He let the pressure roll off him, and left home as soon as he could. But he always looked out for Jane.

'Mum is in hospital,' he said as the blood drained from Jane's legs.

'I don't care,' she replied, but she did. She also wanted to see Brian, because he was the closest thing she had to a friend in this town.

13.13 What next? 

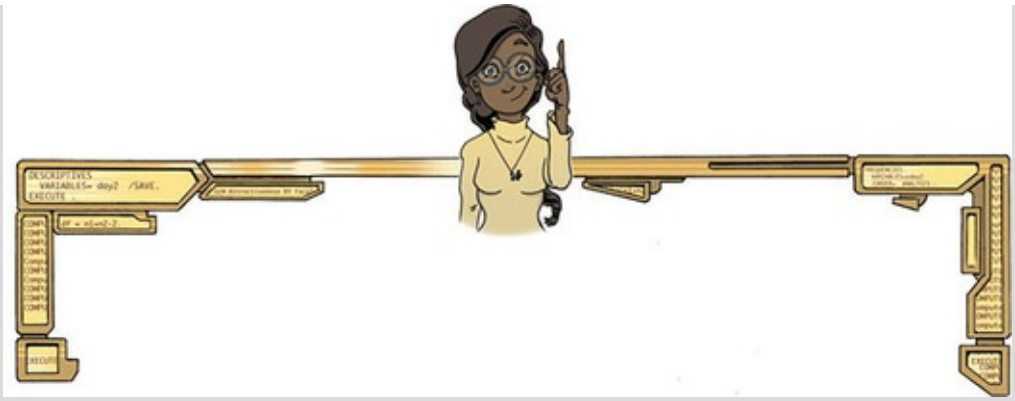
At the age of 13 I met my heroes, Iron Maiden, and very nice they were too. I've met them a couple of times since (not because they're my best buddies or anything exciting like that, but over the years the fan club put on various events where you were allowed to stand next to them and gibber like a fool while they humoured you politely). You'll notice that the photo at the start of this chapter is signed by Dave Murray. This wasn't possible because I had my own darkroom installed backstage at the Hammersmith Odeon in which I could quickly process photographs, or because I had access to time travel (sadly), but because I took the photo with me when I met him in 2000. I recounted the tale of how terrified I had been about meeting him in 1986. If he thought I was some strange stalker he certainly didn't let on. Uncharacteristic of most people who've sold millions of albums, they're lovely blokes.

Anyway, having seen Iron Maiden in their glory, I was inspired. They still inspire me: I still rate them as the best live band I've ever seen (and I've seen them over 35 times, so I ought to know). Although I had briefly been deflected from my destiny by the shock of grammar school, I was back on track. I *had* to form a band. There was just one issue: no one else played a musical instrument. The solution was easy: through several months of covert subliminal persuasion I convinced my two best friends (both called Mark, oddly enough) that they wanted nothing more than to start learning the drums and bass guitar. A power trio was in the making.

13.14 Key terms that I've discovered

Adjusted mean
Analysis of covariance (ANCOVA)
Covariate
Homogeneity of regression slopes
Partial eta squared (partial η^2)
Šidák correction

Smart Alex's tasks

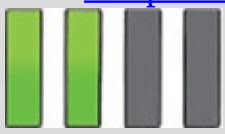


- **Task 1**: A few years back I was stalked. You'd think they could have found someone a bit more interesting to stalk, but apparently times were hard. It could have been a lot worse, but it wasn't particularly pleasant. I imagined a world in which a psychologist tried two different therapies on different groups of stalkers (25 stalkers in each group – this variable is called **Group**). To the first group he gave cruel-to-be-kind therapy (every time the stalkers followed him around, or sent him a letter, the psychologist attacked them with a cattle prod). The second therapy was psychodyshamic therapy, in which stalkers were hypnotized and regressed into their childhood to discuss their penis (or lack of penis), their father's penis, their dog's penis, the seventh penis of a seventh penis and any other penis that sprang to mind. The psychologist measured the number of hours stalking in one week both before (**stalk1**) and after (**stalk2**) treatment (**Stalker.sav**). Analyse the effect of therapy on stalking behaviour after therapy, covarying for the amount of stalking behaviour before therapy. 
- **Task 2**: Compute effect sizes for Task 1 and report the results. 
- **Task 3**: A marketing manager tested the benefit of soft drinks for curing hangovers. He took 15 people and got them drunk. The next morning as they awoke, dehydrated and feeling as though they'd licked a camel's sandy feet clean with their tongue, he gave five of them water to drink, five of them Lucozade (a very nice glucose-based UK drink) and the

remaining five a leading brand of cola (this variable is called **drink**). He measured how well they felt (on a scale from 0 = I feel like death to 10 = I feel really full of beans and healthy) two hours later (this variable is called **well**). He measured how **drunk** the person got the night before on a scale of 0 = as sober as a nun to 10 = flapping about like a haddock out of water on the floor in a puddle of their own vomit (**HangoverCure.sav**). Fit a model to see whether people felt better after different drinks when covarying for how drunk they were the night before. 

- **Task 4**: Compute effect sizes for Task 3 and report the results. 
- **Task 5**: The highlight of the elephant calendar is the annual elephant soccer event in Nepal (google search it). A heated argument burns between the African and Asian elephants. In 2010, the president of the Asian Elephant Football Association, an elephant named Boji, claimed that Asian elephants were more talented than their African counterparts. The head of the African Elephant Soccer Association, an elephant called Tunc, issued a press statement that read 'I make it a matter of personal pride never to take seriously any remark made by something that looks like an enormous scrotum'. I was called in to settle things. I collected data from the two types of elephants (**elephant**) over a season and recorded how many goals each elephant scored (**goals**) and how many years of experience the elephant had (**experience**). Analyse the effect of the type of elephant on goal scoring, covarying for the amount of football experience the elephant has (**Elephant Football.sav**). 
- **Task 6**: In Chapter 4 (Task 6) we looked at data from people who had been forced to marry goats and dogs and measured their life satisfaction and also how much they like animals (**Goat or Dog.sav**). Fit a model predicting life satisfaction from

the type of animal to which a person was married and their animal liking score (covariate). 

- **Task 7**: Compare your results for Task 6 to those for the corresponding task in Chapter 11. What differences do you notice and why? 
- **Task 8**: In Chapter 10 we compared the number of mischievous acts (**mischief2**) in people who had invisibility cloaks to those without (**cloak**). Imagine we also had information about the baseline number of mischievous acts in these participants (**mischief1**). Fit a model to see whether people with invisibility cloaks get up to more mischief than those without when factoring in their baseline level of mischief (**Invisibility Baseline.sav**). 

Answers & additional resources are available on the book's website at **https://edge.sagepub.com/field5e**