
Analysis of Covariance

CONTENTS

- 7.1 Introduction
- 7.2 Purposes of Covariance
- 7.3 Adjustment of Posttest Means
- 7.4 Reduction of Error Variance
- 7.5 Choice of Covariates
- 7.6 Numerical Example
- 7.7 Assumptions in Analysis of Covariance
- 7.8 Use of ANCOVA with Intact Groups
- 7.9 Computer Example for ANCOVA
- 7.10 Alternative Analyses
- 7.11 An Alternative to the Johnson–Neyman Technique
- 7.12 Use of Several Covariates
- 7.13 Computer Example with Two Covariates
- 7.14 Summary

7.1 INTRODUCTION

In Chapter 4 we examined the effect of two or more independent variables (factors) in explaining variation on the dependent variable. We set up an experimental design, and thus this method is called experimental control. In this chapter we consider explaining variation on the dependent variable by measuring the subjects on some other variable(s), called covariates, that are correlated with the dependent variable. Recall that the square of a correlation can be interpreted as “proportion of variance accounted for.” Thus, if we find that I.Q. is correlated with achievement

(dependent variable), say .60, we will be able to attribute 36% of the within group variance on the dependent variable to variability on I.Q. In analysis of covariance (ANCOVA), this part of the variance is removed from the error term, and yields a more powerful test. This method of explaining variation is called statistical control. We now consider an example to illustrate how ANCOVA can be very useful in an experimental study in which the subjects have been randomly assigned to groups.

Example

Suppose an investigator is comparing the effects of two treatments on achievement in science. He assesses achievement through the use of a 50 item multiple choice test. He has 24 students and is able to randomly assign 12 of them to each of the treatments. I.Q. scores are also available for these subjects. The data are as follows:

	Treat. 1		Treat. 2	
	I.Q.	Ach.	I.Q.	Ach.
	100	23	96	19
	113	31	108	26
	98	35	122	31
	110	28	103	22
	124	40	132	36
	135	42	120	38
	118	37	111	31
	93	29	93	25
	120	34	115	29
	127	45	125	41
	115	33	102	27
	104	25	107	21
Means	113.08	33.5	111.17	28.83

The investigator feels no need to use the I.Q. data for analysis purposes since the groups have been “equated” on all variables because of the random assignment. He therefore runs a t test for independent samples on achievement at the .05 level. He finds $t = 1.676$, which is not significant because the critical values are ± 2.074 .

Because of small sample size we have a power problem. The estimated effect size is $\hat{d} = (33.5 - 28.83) / 6.83 = .68$ (cf. Section 3.2), which is undoubtedly of practical significance since the groups differ by about two-thirds of a standard deviation. We have not detected it because of the power problem *and* because there is considerable within group variability on achievement. In fact, the pooled within correlation of I.Q. with achievement for the above data is about .80. This means that 64% of the variation in achievement test scores is associated with variation

(individual differences) on I.Q. An analysis of covariance removes that portion from the error term and yields a t value significant at the .05 level ($t = 2.25$). Actually it comes out as an F statistic, so you need to take the square root. Recall that $F = t^2$ for two groups. After reading this chapter, the reader will be able to verify the above t value by running the ANCOVA on SAS or SPSS.

The above example showed that analysis of covariance is very useful in creating a more powerful test in an experimental study. ANCOVA is also used to reduce bias when comparing intact or self-selected groups, such as males and females, Head Start and non-Head Start. A classical use is adjusting posttest means on the dependent variable for any initial differences that may have been present on a pretest. Another typical use is in teaching methods studies that use intact classrooms. If the average I.Q.'s for the classrooms differ by 10 points, then an adjustment of the posttest achievement is done. Although the use of analysis of covariance in this context may seem reasonable, it is quite controversial, which we discuss in detail in Section 7.8.

The first 10 sections of this chapter cover the basics for ANCOVA with one covariate. We discuss the purposes of covariance, the underlying concepts, the assumptions, interpretation of results, the relationship of ANOVA and ANCOVA, and the running of ANCOVA on SAS and SPSS. The last five sections are more advanced, especially the section on the Johnson–Neyman technique, and may be skipped without loss of continuity. Much has been written about analysis of covariance, and the reader should at least be aware of two classic review articles by Cochran (1957) and Elashoff (1969), and a very comprehensive and thorough book on covariance and alternatives by Huitema (1980).

7.2 PURPOSES OF COVARIANCE

Analysis of covariance is related to the following two basic objectives in experimental design:

1. Elimination of systematic bias.
2. Reduction of within group or error variance.

Systematic bias means that the groups differ systematically on some key variable(s) that are related to performance on the dependent variable. If the groups involve treatments, then a significant difference on a posttest at the end of treatments will be confounded (mixed in with) with initial differences on a key variable. It would not be clear whether the treatments were making the difference, or whether initial differences simply transferred to posttest means. A simple example is a teaching methods study with initial differences between groups on I.Q. Suppose two methods of teaching algebra are compared (same teacher for both methods)

with two classrooms in the same school. The following summary data, means for the groups, are available:

	Method 1	Method 2
I.Q.	120.2	105.8
Posttest	73.4	67.5

If the t test for independent samples on the posttest is significant, then it isn't clear whether it was method 1 that made the difference, or the fact that the children in that class were "brighter" to begin with, and thus would be expected to achieve higher scores.

As another example, suppose we are comparing the effect of four stress situations on blood pressure (the dependent variable). It is found that situation 3 is significantly more stressful than the other three situations. However, we note that the blood pressure of the subjects in group 3 under minimal stress is greater than for the subjects in the other groups. Then, it isn't clear that situation 3 is necessarily most stressful. We need to determine whether the blood pressure for group 3 would still be higher if the posttest means for all 4 groups were "adjusted" in some way to account for initial differences in blood pressure. We see later that the posttest means are adjusted in a linear fashion to what they would be if all groups started out equally on the covariate, that is, at the grand mean.

The best way of dealing with systematic bias is to randomly assign subjects to groups. Then we can be confident, within sampling error, that the groups don't differ systematically *on any variables*. Of course, in many studies random assignment is not possible, so we look for ways of at least partially equating groups. One way of partially controlling for initial differences is to match on key variables. Of course, then we can only be sure the groups are equivalent on those matched variables. Analysis of covariance is a *statistical* way of controlling on key variables. Once again, as with matching, ANCOVA can only *reduce* bias, and not eliminate it.

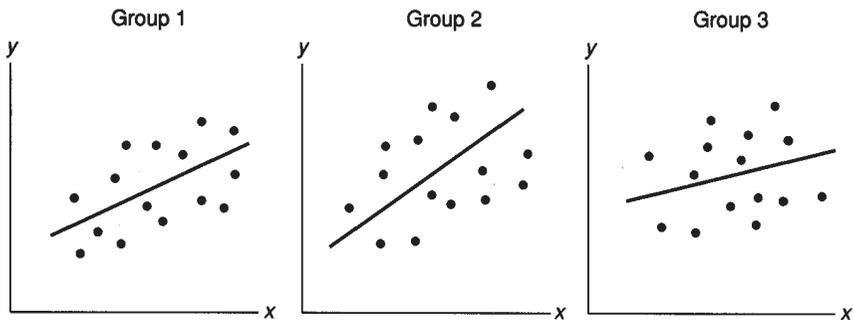
Why is reduction of error variance, the second purpose of analysis of covariance, important? Recall from Chapter 2 on one way ANOVA that the F statistic was $F = MS_b/MS_w$, where MS_w was the estimate of error. If we can make MS_w smaller, then F will be larger and we will obtain a more sensitive or powerful test. And from Chapter 3 on power, remember that power is generally poor in small or medium sample size studies. Thus the use of perhaps 2 or 3 covariates in such studies should definitely be considered. The use of covariates that have relatively low correlations with each other are particularly helpful because each covariate removes a somewhat different part of the error variance from the dependent variable.

Analysis of covariance is a *statistical* way of reducing error variance. There are several other ways of reducing error variance. One way is through sample selec-

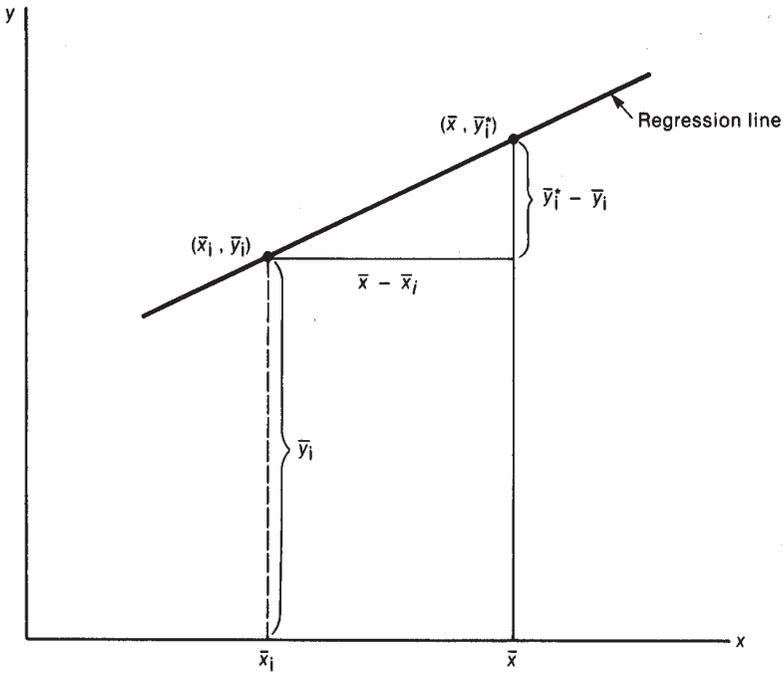
tion; subjects who are more homogeneous vary less on the dependent measure. Another way, discussed in Chapter 4 on factorial designs, was to block on a variable, or consider it as another factor in the design.

7.3 ADJUSTMENT OF POSTTEST MEANS

As mentioned earlier, analysis of covariance adjusts the posttest means to what they would be if all groups started out equally on the covariate; at the grand mean. In this section we derive the general equation for linearly adjusting the posttest means for one covariate. Before we do that, however, it is important to discuss one of the assumptions underlying the analysis of covariance. That assumption for one covariate requires *equal population regression slopes* for all groups. Consider a three group situation, with 15 subjects per group. Suppose that the scatterplots for the 3 groups looked as given below.



Recall from beginning statistics that the x and y scores for each subject determine a point in the plane. Requiring that the slopes be equal is equivalent to saying that the nature of the linear relationship is the *same* for all groups, or that the rate of change in y as a function of x is the same for all groups. For the above scatterplots the slopes are different, with the slope being the largest for group 2 and smallest for group 3. But the issue is whether the *population* slopes are different, and whether the sample slopes differ sufficiently to conclude that the population values are different. With small sample sizes as in the above scatterplots, it is dangerous to rely on visual inspection to determine whether the population values are equal, because of considerable sampling error. Fortunately there is a statistic for this, and later we indicate how to obtain it on SPSS and SAS. In deriving the equation for the adjusted means we are going to assume the slopes are equal. What if the slopes are not equal? Then ANCOVA is *not* appropriate, and we indicate alternatives later on in the chapter.



$$\text{Slope of Straight Line} = b = \frac{\text{change in } y}{\text{change in } x}$$

$$b = \frac{\bar{y}_i^* - \bar{y}_i}{\bar{x} - \bar{x}_i}$$

$$b(\bar{x} - \bar{x}_i) = \bar{y}_i^* - \bar{y}_i$$

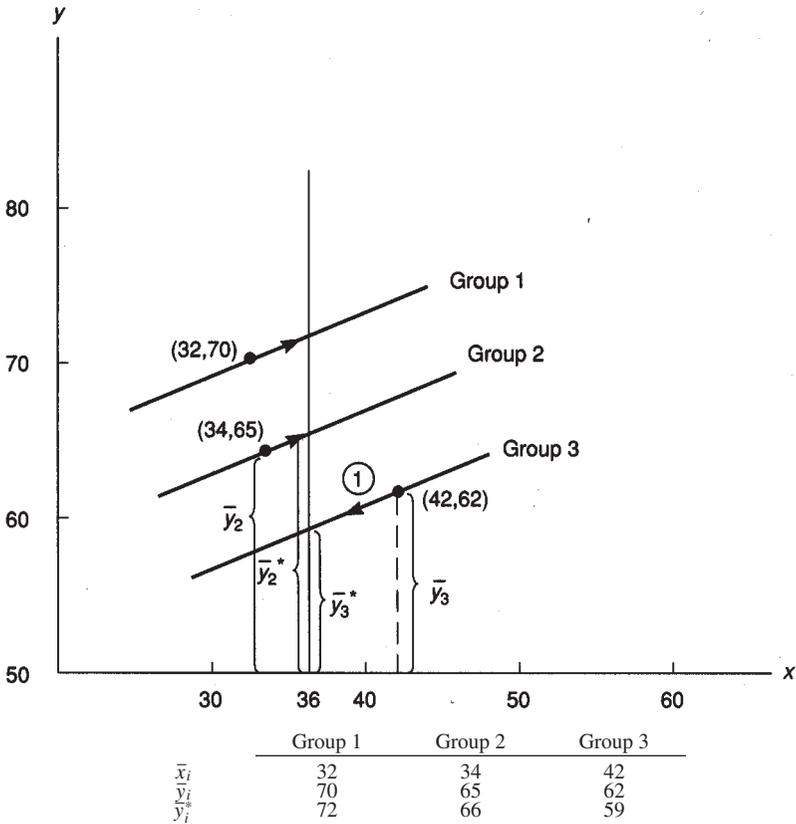
$$\bar{y}_i^* = \bar{y}_i + b(\bar{x} - \bar{x}_i)$$

$$\bar{y}_i^* = \bar{y}_i - b(\bar{x}_i - \bar{x})$$

FIGURE 7.1 Deriving the General Equation for Adjusted Means in Covariance

The details of obtaining the adjusted mean for the *i*th group (i.e., any group) are given in Figure 7.1. The general equation follows from the definition for the slope of a straight line and some basic algebra.

In Figure 7.2 we show the adjusted means geometrically for a hypothetical 3 group data set. A positive correlation is assumed between the covariate and the dependent variable, so that a higher mean on *x* implies a higher mean on *y*. Note that since group 1 scored below the grand mean on the covariate, its mean is adjusted upward. On the other hand, since the mean for group 3 on the covariate is *above* the



A common slope = .5 is being assumed here.

$$\bar{y}_1^* = 70 - .5(32 - 36), \bar{y}_2^* = 65 - .5(34 - 36), \bar{y}_3^* = 62 - .5(42 - 36)$$

① The arrows on the regression lines indicate that the means are adjusted linearly upward or downward to what they would be if the groups had started out at the grand mean on the covariate.

FIGURE 7.2 Means and Adjusted Means for Hypothetical Three Group Data Set

grand mean, covariance estimates that it would have scored lower on y if its mean on the covariate was lower (at grand mean), and therefore the mean for group 3 is adjusted downward.

7.4 REDUCTION OF ERROR VARIANCE

It is relatively simple to derive the approximate error term for covariance. Denote the correlation between the covariate (x) and the dependent variable (y) by r_{xy} . The square of a correlation can be interpreted as “proportion of variance accounted for.” The within group variance for ANOVA is MS_w . Thus, the part of the within group variance on y that is accounted for by the covariate is $r_{xy}^2 MS_w$. The within variability left, after the portion due to the covariate is removed, is

$$MS_w - MS_w r_{xy}^2 = MS_w (1 - r_{xy}^2) \quad (1)$$

and this becomes our new error term for the analysis of covariance, which we denote by MS_w^* . Technically, there is an additional part to the adjusted error term:

$$MS_w^* = MS_w (1 - r_{xy}^2) [1 - 1/(f_e - 2)]$$

where f_e is the error degrees of freedom. However, the effect of this additional factor is slight as long as $N > 50$.

To show how much of a difference a covariate can make in increasing the sensitivity of an experiment, we consider a hypothetical study. An investigator runs a one-way ANOVA (3 groups and 20 subjects per group), and obtains $F = 200/100 = 2$, which is not significant, because the critical value at .05 is 3.18. He pretested the subjects, but didn't use the pretest as a covariate (even though the correlation between covariate and posttest was .71) because the groups didn't differ significantly on the pretest. This is a common mistake made by some researchers who are unaware of the other purpose of covariance, that of reducing error variance. The analysis is redone by another investigator using ANCOVA. Using the equation we just derived she finds

$$MS_w^* \approx 100[1 - (.71)^2] = 50$$

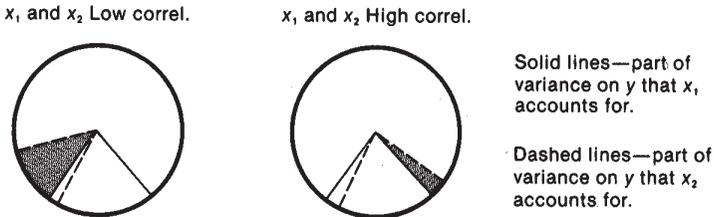
Thus, the error term for the ANCOVA is only half as large as the error term for ANOVA. It is also necessary to obtain a new MS_b^* for ANCOVA, call it MS_b^* . In Section 7.6 we show how to calculate MS_b^* . Let us assume here that the investigator obtains the following F ratio for the covariance analysis:

$$F^* = 190/50 = 3.8$$

This is significant at the .05 level. Therefore, the use of covariance can make the difference between finding and not finding significance. Finally, we wish to note that MS_b^* can be smaller or larger than MS_b although in a randomized study the expected values of the two are equal.

7.5 CHOICE OF COVARIATES

In general, any variables that theoretically should correlate with the dependent variable, or variables that have been shown to correlate on similar types of subjects, should be considered as possible covariates. The ideal is to choose as covariates variables that of course are significantly correlated with the dependent variable *and* have low correlations among themselves. If two covariates are highly correlated (say .80), then they are removing much of the *same* error variance from y ; x_2 will not have much incremental validity. On the other hand, if two covariates (x_1 and x_2) have a low correlation (say .20), then they are removing relatively distinct pieces of the error variance from y , and we will obtain a much greater total error reduction. This is illustrated graphically below using Venn diagrams, where the circle represents error variance on y .



The shaded portion in each case represents the incremental validity of x_2 , that is, the part of error variance on y it removes that x_1 did not.

Huitema (1980, p. 161) has recommended limiting the number of covariates to the extent that the ratio

$$\frac{[C + (J - 1)]}{N} < .10$$

where C is the number of covariates, J is the number of groups, and N is total sample size. Thus, if we had a four group problem with a total of 80 subjects, then $(C + 3)/80 < .10$ or $C < 5$. Less than 5 covariate should be used. If the above ratio is $> .10$, then the adjusted means are likely to be unstable.

7.6 NUMERICAL EXAMPLE

We now consider an example to illustrate how to calculate an ANCOVA and to make clear what the null hypothesis is that is being tested. We use the following 3 group data set from Myers (1979, p. 417), where x indicates the covariate:

Group 1		Group 2		Group 3	
x	y	x	y	x	y
12	26	11	32	6	23
10	22	12	31	13	35
7	20	6	20	15	44
14	34	18	41	15	41
12	28	10	29	7	28
11	26	11	31	9	30

Recall that in the one way ANOVA the null hypothesis was $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ (population means are equal). But in analysis of covariance we are adjusting the means (Section 7.3), so that the null hypothesis becomes $H_0: \mu_1^* = \mu_2^* = \dots = \mu_k^*$, that is, the *adjusted* population means are equal. In the above example, the specific null hypothesis is $H_0: \mu_1^* = \mu_2^* = \mu_3^*$. In ANCOVA we adjust sums of squares corresponding to the sums of squares total, within and between from ANOVA. We denote these adjusted sums of squares by SS_t^* , SS_w^* , and SS_b^* respectively. SS_b^* is obtained by subtracting SS_w^* from SS_t^* .

An ANOVA on the above Myers data, as the reader should check, yields a within cells sum of squares of 666.83 and a group sum of squares of 172.11. We will need these results in obtaining the ANCOVA. Recall that SS_t from ANOVA measures variability of the subjects scores about the grand mean;

$$SS_t = \sum (x_{ij} - \bar{x})^2$$

Let r_{xy} denote the correlation between the covariate and the dependent variable for all the scores, disregarding group membership. Remember that r_{xy}^2 can be interpreted as proportion of variance accounted for. Thus, $r_{xy}^2 SS_t$ represents the amount of variability on y that is accounted for by its relationship with the covariate. Therefore, the remaining variability on y , or the adjusted total sum of squares, is given by

$$SS_t^* = (1 - r_{xy}^2)SS_t \quad (2)$$

Now consider the pooled within correlation for x and y , that is, where group membership is taken into account. Although not strictly true, this correlation can be thought of as the average (or weighted average for unequal group sizes) of the correlations within the groups. Denote this correlation by $r_{xy(w)}$. Then the amount of within group variability on y accounted for by the covariate is given by $r_{xy(w)}^2 SS_w$. Therefore, the remaining within variability on y , or the adjusted within sum of squares, is given by

$$SS_w^* = (1 - r_{xy(w)}^2)SS_w \quad (3)$$

Finally, the adjusted between sum of squares is obtained as the difference between the adjusted total and adjusted within:

$$SS_b^* = SS_t^* - SS_w^* \quad (4)$$

The F ratio for analysis of covariance is then given by

$$F^* = (SS_b^* / (k - 1)) / SS_w^* / (N - K - C) = MS_b^* / MS_w^* \quad (5)$$

where C is the number of covariates. Note that one degree of freedom for error is lost for each covariate used.

This method of computing the ANCOVA is conceptually fairly simple, and importantly shows its direct linkage with the results from an ANOVA on the same data. The SAS GLM control lines for running the ANCOVA are presented in Table 7.1, along with selected printout. The total correlation is .85286 and the within group correlations are gp 1: .9316, gp 2: .9799, and gp 3: .9708. Using these results, the F ratio for the ANCOVA is easily obtained. First, from Equation 2 we have that

$$SS_t^* (1 - (.85286)^2) 838.94 = 228.72$$

TABLE 7.1
SAS GLM Control Lines and Selected Printout for ANCOVA on Myers
Data and SPSS Windows 12.0 Interactive Plots and Regression Lines

Placeholder for T0701 from p. 315 of previous edition

TABLE 7.1
(Continued)

Place holder for T0701-b from p. 316 of previous edition.

Now, using the average of the within correlations as a rough estimate of the pooled within correlation, we find that $r = (.9316 + .9799 + .9708)/3 = .9608$ (the actual pooled correlation is .965). Now, using Equation 3 we find the adjusted within sum of squares:

$$SS_w^* - (1 - (.965)^2)(666.83) = 45.86$$

Therefore, the adjusted between sum of squares is:

$$SS_b^* - 228.72 - 45.86 = 182.86$$

and the F ratio for the analysis of covariance is

$$F^* = (182.86/2)/45.86/(18-3-1) = 27.87$$

ANCOVA as a Special Case of Multiple Regression

Since analysis of covariance involves both analysis of variance and regression analysis, we can do an ANCOVA using multiple regression. Recall that in the last chapter on regression analysis we showed that ANOVA was a special case of regression analysis. We dummy coded group membership and used these dummy variables to predict the dependent variable.

We will illustrate how an ANCOVA can be done using multiple regression with the Myers data. First, we shall check the homogeneity of regression slopes assumption. For a one way design, as Myers and Well (1991, p. 567) point out, “Performing an ANCOVA on a design that has a single factor A can now be seen as determining whether A has effects over and above those of the covariate x . “Thus, we *force* the covariate in and then determine whether group membership has an effect above and beyond the covariate. Since we have 3 groups here, we will need two dummy variables to code group membership (we denote them by DUM1, DUM2). Recall that a violation of the slopes assumption meant there was a group by covariate interaction. Thus, we set up an interaction effect and test it for significance. We create the group by covariate interaction effects by multiplying (we denote them by COVDUM1 and COVDUM2) and then test these for significance. The complete control lines for testing homogeneity of slopes and doing the ANCOVA are presented in Table 7.2.

Selected printout from SPSS for Windows 12.0 is presented in Table 7.3. Note that the assumption of equal regression slopes is tenable ($F = .354$), and that the ANCOVA is significant ($F = 27.886$).

7.7 ASSUMPTIONS IN ANALYSIS OF COVARIANCE

Analysis of covariance rests on the same assumptions as the analysis of variance *plus* three additional assumptions regarding the regression part of the covariance analysis. ANCOVA also assumes

1. A linear relationship between the dependent variable and the covariate(s).

TABLE 7.2
Syntax Command File for Homogeneity of Slopes Test and ANCOVA on Myers Data Using SPSS for Windows 12.0

```

TITLE 'MULT. REG ON MYERS DATA--ANCOVA'.
DATA LIST FREE/COVAR DEP DUM1 DUM2 COVDUM1 COVDUM2.
BEGIN DATA.
12 26 1 0 12 0 10 22 1 0 10 0 7 20 1 0 7 0 14 34 1 0 14 0 12 28 1 0 12 0 11 26 1 0 11 0
11 32 0 1 0 11 12 31 0 1 0 12 6 20 0 1 0 6 18 41 0 1 0 18 10 29 0 1 0 10 11 31 0 1 0 11
6 23 0 0 0 0 13 35 0 0 0 0 15 44 0 0 0 0 15 41 0 0 0 0 7 28 0 0 0 0 9 30 0 0 0 0
END DATA.
LIST.
REGRESSION DESCRIPTIVES = DEFAULT/
VARIABLES = COVAR TO COVDUM2/
DEPENDENT = DEP/
① ENTER COVAR DUM1 DUM2/TEST (COVDUM1 COVDUM2) /.
VARIABLES = COVAR TO DUM2/
DEPENDENT = DEP/
② ENTER COVAR/TEST(DUM1 DUM2) /.

```

① This is the statement which yields the homogeneity of slopes test.

② This is testing the main hypothesis in ANCOVA, whether the adjusted population means are equal.

COVAR	DEP	DUM1	DUM2	COVDUM1	COVDUM2
12.00	26.00	1.00	.00	12.00	.00
10.00	22.00	1.00	.00	10.00	.00
7.00	20.00	1.00	.00	7.00	.00
14.00	34.00	1.00	.00	14.00	.00
12.00	28.00	1.00	.00	12.00	.00
11.00	26.00	1.00	.00	11.00	.00
11.00	32.00	.00	1.00	.00	11.00
12.00	31.00	.00	1.00	.00	12.00
6.00	20.00	.00	1.00	.00	6.00
18.00	41.00	.00	1.00	.00	18.00
10.00	29.00	.00	1.00	.00	10.00
11.00	31.00	.00	1.00	.00	11.00
6.00	23.00	.00	.00	.00	.00
13.00	35.00	.00	.00	.00	.00
15.00	44.00	.00	.00	.00	.00
15.00	41.00	.00	.00	.00	.00
7.00	28.00	.00	.00	.00	.00
9.00	30.00	.00	.00	.00	.00

2. Homogeneity of the regression slopes (for one covariate); parallelism of the regression planes for two covariates and for more than 2 covariates homogeneity of the regression hyperplanes.
3. The covariate is measured without error.

Since covariance rests on the same assumptions as ANOVA, any violations that are serious in ANOVA (like dependent observations) are also serious in ANCOVA. Violation of *all 3* of the above regression assumptions can also be serious. For example, if the relationship between the covariate and the dependent variable is curvilinear, then the adjustment of the means will be improper.

There is always measurement error for the variables that are typically used as covariates in social science research. In randomized designs this reduces the power of the ANCOVA, but treatment effects are not biased. For non-randomized designs the treatment effects can be seriously biased.

A violation of the homogeneity of regression slopes can also yield quite misleading results. To illustrate this, we present in Figure 7.3 the situation where the assumption is met and two situations where the slopes are unequal. Notice that with equal slopes the estimated superiority of group 1 at the grand mean is a totally accurate estimate of group 1's superiority for *all* levels of the covariate, since the lines are parallel. For Case 1 of unequal slopes there is a *covariate by treatment interaction*. That is, how much better group 1 is depends on which value of the covariate we specify. This is analogous to the concept of interaction in a factorial design. For Case 2 of heterogeneous slopes the use of covariance would be totally misleading. Covariance estimates no difference between the groups, while for $x = c$, group 2 is quite superior, and for $x = d$, group 1 is quite superior. Later in the chapter we show how to test the assumption of equal slopes on SPSS and on SAS.

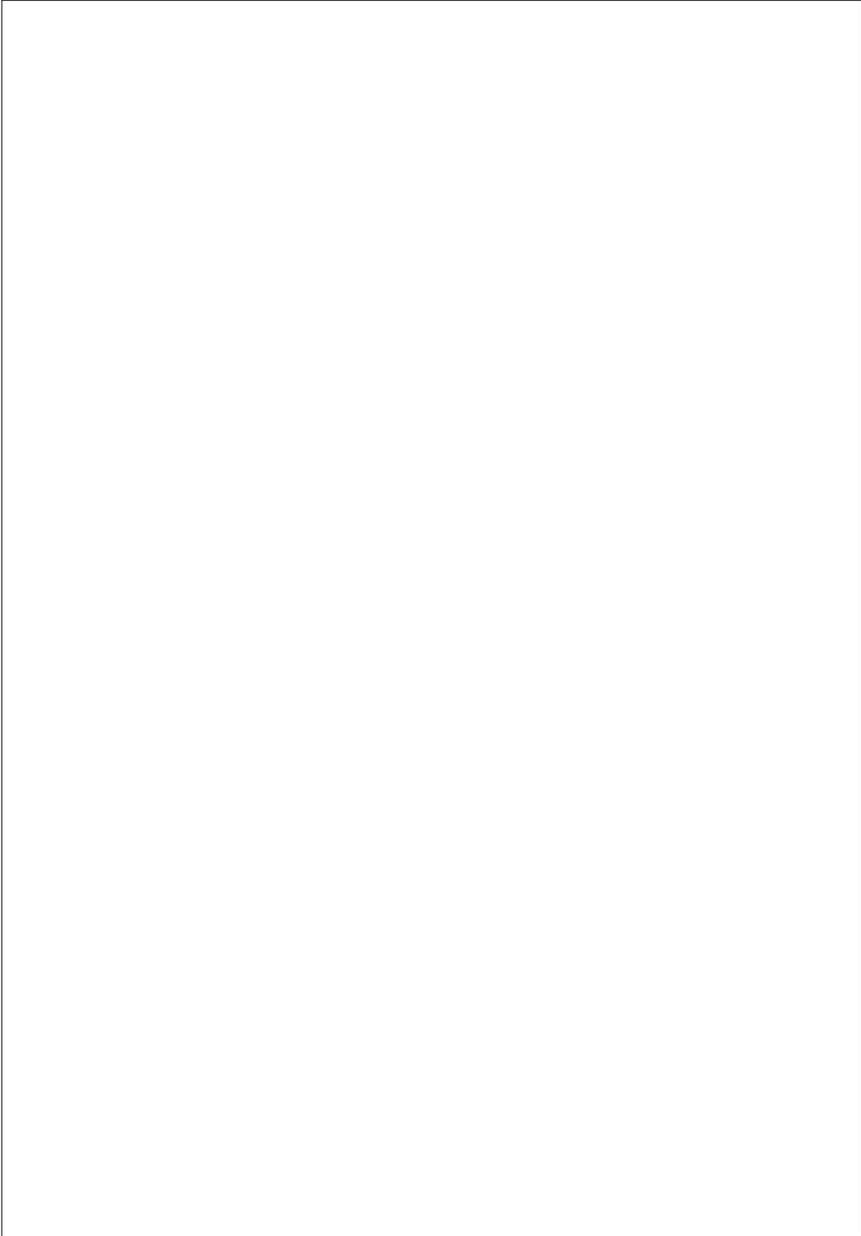
Therefore, in examining printout from the statistical packages it is important to *first* make two checks to determine whether analysis of covariance is appropriate:

1. **Check to see whether there is a linear relationship between the dependent variable and the covariate.**
2. **Check to determine whether the homogeneity of the regression slopes is tenable.**

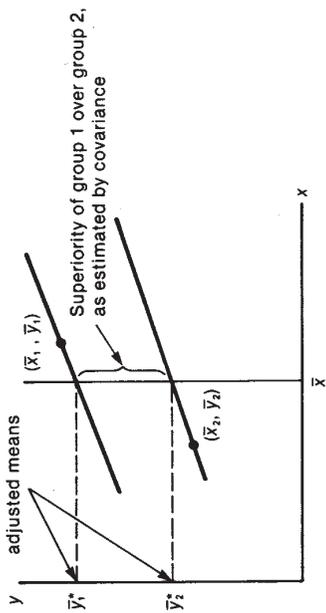
If the above assumptions are met, then there is not any debate about the appropriateness of ANCOVA in randomized studies in which the subjects have been randomly assigned to groups. For intact groups, there is a debate, and we discuss that in the next section.

If either of the above assumptions is not satisfied, then covariance is not appropriate. In particular, if (2) is not met, then one should consider using the

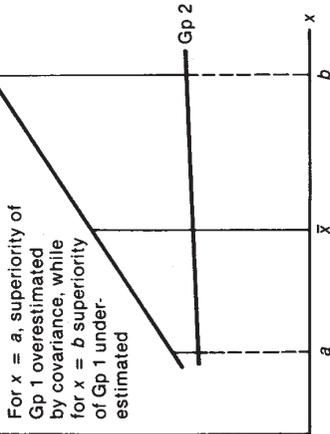
TABLE 7.3
Selected Printout From SPSS for Windows 12.0 for ANCOVA on Myers
Data Using Multiple Regression



-
- ① This is the test for a significant regression on y on the covariate.
 - ② This is the test for homogeneity of the regression slopes.
 - ③ This is the main test in covariance; whether the adjusted population means are equal.



Heterogeneous Slopes
Case 1



Heterogeneous Slopes
Case 2

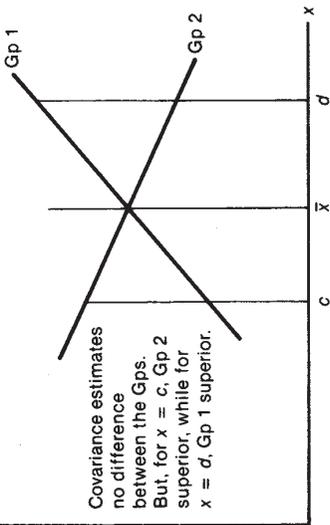


FIGURE 7.3 Effect of Heterogeneous Slopes on Interpretation in ANCOVA

Johnson–Neyman (1936) technique. For extended discussion on the Johnson–Neyman technique see Rogosa (1977, 1980).

7.8 USE OF ANCOVA WITH INTACT GROUPS

It should be noted that some researchers (Anderson, 1963; Lord, 1969) have argued strongly against using analysis of covariance with intact groups. Although we do not take this position, it is important that the reader be aware of the several limitations and/or possible dangers when using ANCOVA with intact groups. First, even the use of several covariates will *not* equate intact groups, and one should never be deluded into thinking it can. The groups may still differ on some unknown important variable(s). Also, note that equating groups on one variable may result in accentuating their differences on other variables.

Second, recall that ANCOVA adjusts the posttest means to what they would be if all the groups had started out equal on the covariate(s). You then need to consider whether groups that are equal on the covariate would ever exist in the real world. Elashoff (1969) gives the following example. Teaching methods *A* and *B* are being compared. The class using *A* is composed of high ability students, whereas the class using *B* is composed of low ability students. A covariance analysis can be done on the posttest achievement scores holding ability constant, as if *A* and *B* had been used on classes of equal and average ability. But, as Elashoff notes, “It may make no sense to think about comparing methods *A* and *B* for students of average ability, perhaps each has been designed specifically for the ability level it was used with, or neither method will, in the future, be used for students of average ability” (p. 387).

Third, the assumptions of linearity and homogeneity of regression slopes need to be satisfied for ANCOVA to be appropriate.

A fourth issue that can confound the interpretation of results is differential growth of subjects in intact or self selected groups on some dependent variable. If the natural growth is much greater in one group (treatment) than for the control group and covariance finds a significance difference, after adjusting for any pretest differences, then it isn’t clear whether the difference is due to treatment, differential growth, or part of each. Bryk and Weisberg (1977) discuss this issue in detail and propose an alternative approach for such growth models.

A fifth problem is that of measurement error. Of course this same problem is present in randomized studies. But there the effect is merely to attenuate power. In non-randomized studies measurement error can seriously bias the treatment effect. Reichardt (1979), in an extended discussion on measurement error in ANCOVA, states,

Measurement error in the pretest can therefore produce spurious treatment effects when none exist. But it can also result in a finding of no intercept difference when a true treatment effect exists, or it can produce an estimate of the treatment effect which is in the opposite direction of the true effect. (p. 164)

It is no wonder then that Pedhazur (1982, p. 524), in discussing the effect of measurement error when comparing intact groups, says,

The purpose of the discussion here was only to alert you to the problem in the hope that you will reach two obvious conclusions: (1) that efforts should be directed to construct measures of the covariates that have very high reliabilities and (2) that ignoring the problem, as is unfortunately done in most applications of ANCOVA, will not make it disappear. (p. 524)

Porter (1967) has developed a procedure to correct ANCOVA for measurement error, and an example illustrating that procedure is given in Huitema (1980, pp. 315–316). This is beyond the scope of the present text.

Given all of the above problems, the reader may well wonder whether we should abandon the use of covariance when comparing intact groups. But other statistical methods for analyzing this kind of data (such as matched samples, gain score ANOVA) suffer from many of the same problems, such as seriously biased treatment effects. The fact is that inferring cause-effect from intact groups is treacherous, regardless of the type of statistical analysis. Therefore, the task is to do the best we can and exercise considerable caution, or as Pedhazur (1982) put it: “But the conduct of such research, indeed all scientific research, requires sound theoretical thinking, constant vigilance, and a thorough understanding of the potential and limitations of the methods being used” (p. 525).

7.9 COMPUTER EXAMPLE FOR ANCOVA

To illustrate how to run an ANCOVA, while at the same time checking the critical assumptions of linearity and homogeneity of slopes, we consider part of a Sesame Street data set from Glasnapp and Poggio (1985), who present data on many variables, including 12 background variables and 8 achievement variables for 240 subjects. Sesame Street was developed as a television series aimed mainly at teaching preschool skills to 3- to 5-year-old children. Data was collected at 5 different sites on many achievement variables both before (pretest) and after (posttest) viewing of the series. We consider here only the achievement variable of knowledge of numbers. The maximum possible score is 54 and the content of the items included recognizing numbers, naming numbers, counting, addition, and subtraction. We use ANCOVA to determine whether the posttest knowledge of numbers for the

TABLE 7.4
SPSS MANOVA Control Lines for Analysis of Covariance on Sesame
Street Data

```
TITLE 'ANALYSIS OF COVARIANCE ON SESAME DATA'.
DATA LIST FREE/SITE PRENUMB POSTNUMB.
BEGIN DATA.

    DATA (ON CD)
END DATA.
MANOVA PRENUMB POSTNUMB BY SITE(1,3) /
① ANALYSIS POSTNUMB WITH PRENUMB /
② PRINT = PMEANS /
    DESIGN /
③ ANALYSIS = POSTNUMB /
    DESIGN = PRENUMB, SITE, PRENUMB BY SITE /
④ ANALYSIS = PRENUMB/.
```

- ① The covariate(s) follow the keyword WITH.
- ② This PRINT subcommand is needed to obtain the adjusted means, which is what we are testing for significance.
- ③ This ANALYSIS subcommand and the following DESIGN subcommand are needed to test the homogeneity of the regression slopes assumption.
- ④ This ANALYSIS subcommand is used to test whether the sites differed significantly on the pre-test.

children at the first 3 sites differed after adjustments are made for any pretest differences.

In Table 7.4 we give the complete control lines for running the ANCOVA on SPSS MANOVA. Table 7.5 gives selected annotated output from that run. We indicate *which* of the F tests are checking the assumptions of linearity and homogeneity of slopes, and which F addresses the main question in covariance (whether the adjusted population means are equal).

7.10 ALTERNATIVE ANALYSES

When comparing two or more groups with pretest and posttest data, the following other modes of analysis have been used by many researchers:

1. An ANOVA is done on the difference or gain scores (posttest—pretest).
2. A two way repeated measures (this is covered in Chapter 5) ANOVA is done. This is also called a one between (the grouping variable) and one within (pretest-posttest part) factor ANOVA.

TABLE 7.5
Selected Printout from SPSS MANOVA for ANCOVA on Sesame Street
Data

Placeholder for T0705 from p. 236 of the previous edition.

① This indicates there is a significant correlation between the dependent variable and the covariate(PRENUMB), or equivalently a significant regression of POSTNUMB on PRENUMB.

② This test indicates that homogeneity of regression slopes is tenable at the .05 level, since the p value is .607.

③ This F is testing the main result in ANCOVA; whether the adjusted population means are equal. This is rejected at the .05 level, indicating SITE differences.

④ These are the adjusted means. Since the estimated common regression slope is .686 (given on the printout but not presented here), the adjusted mean for SITE 1 is

$$\bar{y}_1^* = 30.083 - .686(22.4 - 21.67) = 29.58$$

⑤ This test indicates the subjects at the 3 sites differ significantly on the pretest, i.e., on PRENUMB.

Huck and McLean (1975) and Jennings (1988) have compared the above two modes of analysis along with the use of ANCOVA for the pretest-posttest control group design, and conclude that ANCOVA is the preferred method of analysis. Several comments from the Huck and McLean article are worth mentioning. First, they note that with the repeated measures approach it is the *interaction F* that is indicating whether the treatments had a differential effect, and not the treatment main effect. We consider two patterns of means below to illustrate.

	<i>Situation 1</i>			<i>Situation 2</i>	
	Pretest	Posttest		Pretest	Posttest
Treat.	70	80	Treat	65	80
Control	60	70	Control	60	68

In situation 1 the treatment main effect would probably be significant, because there is a difference of 10 in the row means. However, the difference of 10 on the posttest just transferred from an initial difference of 10 on the pretest. There is not a differential change in the treatment and control groups here. On the other hand, in situation 2 even though the treatment group scored higher on the pretest, it increased 15 points from pre to post while the control group increased just 8 points. That is, there was a *differential* change in performance in the two groups. But, recall from Chapter 4 that one way of thinking of an interaction effect is as a “difference in the differences.” This is exactly what we have in situation 2, hence a significant interaction effect.

Second, Huck and McLean (1975) note that the interaction *F* from the repeated measures ANOVA is *identical* to the *F* ratio one would obtain from an ANOVA on the gain (difference) scores. Finally, whenever the regression coefficient is not equal to 1 (generally the case), the error term for ANCOVA will be smaller than for the gain score analysis and hence the ANCOVA will be a more sensitive or powerful analysis.

Although not discussed in the Huck and McLean paper, we would like to add a measurement caution against the use of gain scores. It is a fairly well known measurement fact that the reliability of gain (difference) scores is generally not good. To be more specific, *as the correlation between the pretest and posttest scores approaches the reliability of the test, the reliability of the difference scores goes to 0*. The following table from Thorndike and Hagen (1977) quantifies things:

Correlation between tests	Average reliability of two tests					
	.50	.60	.70	.80	.90	.95
.00	.50	.60	.70	.80	.90	.95
.40	.17	.33	.50	.67	.83	.92
.50	.00	.20	.40	.60	.80	.90
.60		.00	.25	.50	.75	.88
.70			.00	.33	.67	.83
.80				.00	.50	.75
.90					.00	.50
.95						.00

If our dependent variable is some noncognitive measure, or a variable derived from a nonstandardized test (which could well be of questionable reliability), then a reliability of about .60 or so is a definite possibility. In this case, if the correlation between pretest and posttest is .50 (a realistic possibility), the reliability of the difference scores is only .20! On the other hand, the above table also shows that if our measure is quite reliable (say .90), then the difference scores will be reliable for moderate pre-post correlations. For example, for reliability = .90 and pre-post correlation = .50, the reliability of the differences scores is .80.

7.11 AN ALTERNATIVE TO THE JOHNSON–NEYMAN TECHNIQUE

We consider hypothetical data from Huitema (1980, p. 272). The effects of two types of therapy are being compared on an aggressiveness score. The covariate (x) are scores on a sociability scale. Since the Johnson–Neyman technique is still (18 years after the first edition of this text) not available on SAS or SPSS, we consider an alternative analysis that does shed some light. Recall that a violation of the homogeneity of regression slopes assumption meant there was a covariate by group interaction. Thus, one way of investigating the nature of this interaction would be to set up a factorial design, with groups being one of the factors and two or more levels for the covariate (other factor), and run a regular two way ANOVA. This procedure is not as desirable as the Johnson–Neyman technique for two reasons: (1) the Johnson–Neyman technique is more powerful, and (2) the Johnson–Neyman technique enables us to determine where the group differences are for *all* levels of the covariate, whereas the factorial approach can only check for differences for the levels of the covariate included in the design. Nevertheless, at least most researchers can easily do a factorial design, and this does yield useful information.

For the Huitema data, although there is a strong linear relationship in each group, the assumption of equality of slopes is not tenable (Figure 7.4 shows why). Therefore, covariance is not appropriate, and we split the subjects into three levels

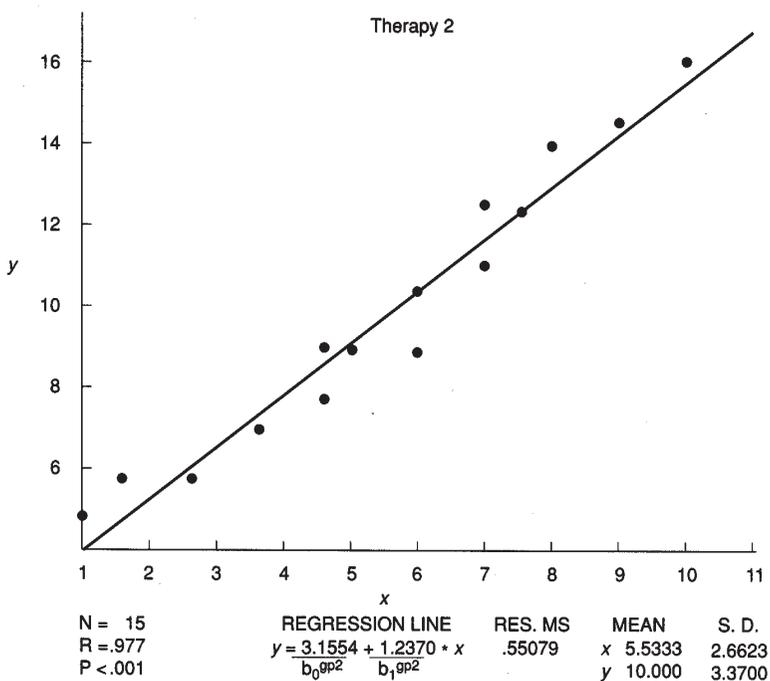
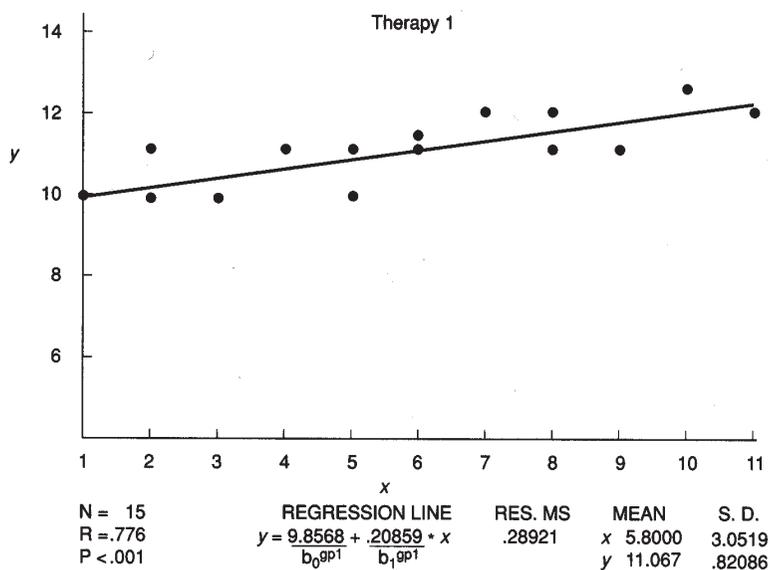


FIGURE 7.4 Scatterplots and Summary Statistics for Each Therapy Group

TABLE 7.6
Results From SPSS for Windows 12.0
for 2×3 Factorial Design on Huitema Data

Placeholder for T0706 from p. 331 of previous edition

for sociability: low (1–4), medium (4.5–7.5) and high (8–11), and set up the following 2×3 ANOVA on aggressiveness:

	SOCIABILITY		
	LOW	MEDIUM	HIGH
THERAPY 1			
THERAPY 2			

Results from the resulting run on SPSS for Windows 12.0 are presented in Table 7.6. They show, as expected, that there is a significant sociability by therapy interaction ($F = 19.735$). The nature of this interaction can be gauged by examining the means for the SOCIAL*THERAPY table. These show that for low sociability therapy group 1 is more aggressive, whereas for high sociability therapy group 2 is more aggressive. The results from the Johnson–Neyman analysis for this data, presented in the first edition of this text (p. 179), show that more precisely there is *no* significant difference in aggressiveness for sociability scores between 6.04 and 7.06.

7.12 USE OF SEVERAL COVARIATES

What is the rationale for using several covariates? First, the use of several covariates will result in greater error reduction than can be obtained with just one covariate. The error reduction will be substantially greater if there are low intercorrelations among the covariates. In this case each of the covariates will be removing a somewhat different part of the error variance from the dependent variable. Also, with several covariates we can make a better adjustment for initial differences among groups.

Recall that with one covariate simple linear regression was involved. With several covariates (predictors), multiple regression is needed. In multiple regression the linear combination of the predictors that is *maximally correlated* with the dependent variable is found. The multiple correlation (R) is a maximized Pearson correlation between the observed scores on y and their predicted scores, $R = r_{yy}$. Although R is more complex it is a correlation and hence R^2 can be interpreted as “proportion of variance accounted for.” Also, we will have regression coefficients for each of the covariates (predictors). Below we present a table comparing the single and multiple covariate cases:

TABLE 7.7
SPSS MANOVA Control Lines for ANCOVA on Sesame Street Data
With Two Covariates

```
TITLE 'SESAME ST. DATA-2 COVARIATES'.
DATA LIST FREE/ SITE PRENUMB PRERELAT POSTNUMB.
BEGIN DATA.

DATA LINES

END DATA.
MANOVA PRENUMB PRERELAT POSTNUMB BY SITE(1,3) /
ANALYSIS POSTNUMB WITH PRENUMB PRERELAT /
PRINT = PMEANS /
DESIGN /
ANALYSIS = POSTNUMB /
DESIGN = PRENUMB+PRERELAT, SITE, PRENUMB BY SITE+
PRERELAT BY SITE /
ANALYSIS = PRENUMB PRERELAT / .
```

	One Covariate	Multiple Covariates
Error Reduction	primarily determined by simple correlation r_{yx}^2 – within variance on y accounted for by x	determined by the multiple correlation R^2 – within variance on y accounted for by the set of covariates
Adjustment of Means	$y_i^* = y_i - b(\bar{x}_i - \bar{x})$, b is assumed common slope	$y_j^* = y_j - b_1(\bar{x}_{1j} - \bar{x}_1) - b_2(\bar{x}_{2j} - \bar{x}_2) - \dots - b_k(\bar{x}_{kj} - \bar{x}_k)$

where the b_i are the regression coefficients, \bar{x}_{1j} is the mean for covariate 1 in group j , \bar{x}_{2j} is the mean for covariate 2 in group j , etc., and the \bar{x}_i are the grand means for the covariates.

7.13 COMPUTER EXAMPLE WITH TWO COVARIATES

To illustrate running an ANCOVA with more than one covariate, we reconsider the Sesame Street data set used in Section 7.9. Again we shall be interested in site differences on POSTNUMB, but now we use *two* covariates: PRENUMB and PRERELAT (pretest on knowledge of relational terms—amount, size, and position relationship—maximum score of 17). Before we give the control lines for running the analysis, we need to discuss in more detail how to set up the lines for testing the homogeneity assumption. For one covariate this is equality of regression slopes. For two covariates it is parallelism of the regression planes, and for more than two covariates it involves equality of regression hyperplanes.

TABLE 7.8
Printout from SPSS MANOVA for Sesame Data with Two Covariates

Placeholder for T 0708 from p. 334 of previous edition.

-
- ① This test indicates significant SITE differences at .05 level.
 - ② These are the regression coefficients
 - ③ These are the adjusted means, which would be obtained as follows:

$$\begin{aligned}\bar{y}_3^* &= \bar{y}_3 - b_1(\bar{x}_{13} - \bar{x}_1) - b_2(\bar{x}_{23} - \bar{x}_2) \\ &= 25.437 - .564(16.563 - 21.670) - .622(8.563) - 10.14 \\ &= 29.30\end{aligned}$$

- ④ This test indicates parallelism of the regression planes is tenable at the .05 level.

It is important to recall that a violation of the assumption means there is a covariate by treatment (group) interaction. If the assumption is tenable this means the interaction will *not* be significant. Therefore, what one does in SPSS MANOVA is to set up an effect involving the interaction (for one covariate), and then test whether this effect is significant. If the effect is significant, it means the assumption is not tenable.

For more than one covariate, as in the present case, there is an interaction term for each covariate. The effects are lumped together and then we test whether the combined interactions are significant. Before we give a few examples, note that **BY** is the keyword used by SPSS to denote an interaction, and **+** is used to lump effects together.

We show the control lines for testing the homogeneity assumption for two covariates and for three covariates. Denote the dependent variable by y , the covariates by x_1 and x_2 and the grouping variable by gp . The control lines are

```
ANALYSIS = Y /
DESIGN = X1+X2,GP,X1 BY GP+X2 BY GP /
```

Now, suppose there were three covariates. Then the control lines will be:

```
ANALYSIS = Y /
DESIGN = X1+X2+X3,GP,X1 BY GP+X2 BY GP+X3 BY GP /
```

The control lines for running the ANCOVA on the Sesame Street data with the covariates of **PRENUMB** and **PRERELAT** are given in Table 7.7. In Table 7.8 we present selected output from the SPSS analysis of covariance.

7.14 SUMMARY

1. In analysis of covariance a linear relationship is assumed between the dependent variable and the covariate(s).

2. ANCOVA is directly related to the two basic objectives in experimental design of (a) eliminating systematic bias and (b) reduction of error variance.

While ANCOVA does not eliminate bias, it can reduce bias. The use of several covariates with low intercorrelations will substantially reduce error variance.

3. Limit the number of covariates (C) so that

$$\frac{C + (J - 1)}{N} < .10$$

where J is the number of groups and N is total sample size.

4. A numerical example is given to show the intimate relationship between ANCOVA and the results for ANOVA on the same data.

TABLE 7.9

Placeholder for T0709 from p. 337 of previous edition.

5. Measurement error on the covariate causes loss of power in randomized designs, and can lead to seriously biased treatment effects in non-randomized designs.

6. In examining printout from the statistical packages, first make two checks to determine whether covariance is appropriate: (1) check that there is a linear relationship between the covariate and the dependent variable and (2) check that the regression slopes are equal. If *either* of these is not true, then covariance is not appropriate. In particular, if (2) is not true then the Johnson–Neyman technique should be considered.

7. Several cautions are given concerning the use of analysis of covariance with intact groups.

8. Three ways of analyzing a k group pretest-posttest design are: ANOVA on the difference scores, analysis of covariance, and a two way repeated measures ANOVA. Articles by Huck and McLean (1975) and by Jennings (1988) show that ANCOVA is generally the preferred method of analysis.

9. Although the Johnson–Neyman technique is preferred when the slopes are not equal, it is still not available on SAS or SPSS. *A violation of the equal slopes assumption means there is a group by covariate interaction effect.* Because of this, we illustrated, in Section 7.12, use of a two way ANOVA to get at the nature of this interaction.

10. We showed how ANCOVA can be done using multiple regression. By dummy coding group membership and appropriate multiplication we obtained both the test for homogeneity of regression slopes and the ANCOVA.

EXERCISES

1. A social psychological study by Novince (1977) examined the effect of behavioral rehearsal, and behavioral rehearsal plus cognitive restructuring (combination treatment) on reducing anxiety and facilitating social skills for female college freshmen. The 33 subjects were randomly assigned (11 each) to either BH, a control group (group 2), or BH + CR. The subjects were pretested and posttested on several variables. The scores for the avoidance variable are given as follows:

BEHAVIORAL REHEARSAL		CONTROL		BEHAVIORAL REHEARSAL + COGNITIVE RESTRUCTURING	
Avoid	Preavoid	Avoid	Preavoid	Avoid	Preavoid
91	70	107	115	121	96
107	121	76	77	140	120
121	89	116	111	148	130
86	80	126	121	147	145
137	123	104	105	139	122
138	112	96	97	121	119
133	126	127	132	141	104
127	121	99	98	143	121
114	80	94	85	120	80
118	101	92	82	140	121
114	112	128	112	95	92

Table 7.9 shows selected printout from an ANCOVA on SPSS for Windows 7.5 (top two thirds of printout).

- (a) Is ANCOVA appropriate for this data? Explain.
 - (b) If ANCOVA is appropriate, then do we reject the null hypothesis of equal adjusted population means at the .05 level?
 - (c) The bottom portion of the printout shows the results from an ANOVA on just avoidance. Note that the error term is 280.07. The error term for the ANCOVA is 111.36. How are the two error terms fundamentally related?
2. (a) Run an ANOVA on the difference scores for the data in exercise 1.
(b) Compare the error term for that analysis vs the error term for the ANCOVA on the same data. Relate these results to the discussion in Section 7.10.
 3. This question relates the use of a pretest as covariate to experimental design considerations. Suppose in a counseling study eight subjects were randomly assigned to each of three groups. The subjects were pretested and posttested on client satisfaction, which served as the dependent variable.
 - (a) What is the main reason for using the pretest here as a covariate?
 - (b) In what other way might the covariate be useful?
 - (c) What effect would the possibility of pretest sensitization have on your decision to use a pretest in this study?
 4. An analysis of variance is run on three intact groups and a significant difference is found at the .05 level. The pattern of means is

	GP 1	GP 2	GP 3
COVARIATE	120	100	110
DEP. VAR.	70	60	65

A few days later the investigator, after talking to a colleague, runs an ANCOVA on this data and no longer finds significance at the .05 level. The correlation between the dependent variable and the covariate is .61 and the homogeneity of regression slopes assumption is found to be tenable. Explain what has happened here, and relate this to the discussion in section 7.3.

5. A study by Huck and Bounds (1972) examined whether the grade assigned an essay test is influenced by handwriting neatness. They hypothesized that an interaction effect would occur, with graders who have neat handwriting lowering the essay grade while graders with messy handwriting will not lower the grade. Students in an Educational Measurement class at the University of Tennessee served as subjects. Sixteen were classified as having neat handwriting and 18 were classified as messy handwriters. Each of these 34 subjects received two one page essays. A person with average handwriting neatness copied the first (better) essay. The second essay was copied by two people, one having neat handwriting and one having messy handwriting. Each subject was to grade each of the two essays on a scale from 0 to 20. Within the neat handwriters, half of them were randomly assigned to receive a neatly written essay to grade and the other half a messy essay. The same was done for the messy handwriters who were acting as graders. The grade assigned to essay 1 served as the covariate in this study. Means and adjusted means are given below for groups:

		Neat Essay		Messy Essay
Neat	Essay 1	$\bar{x} = 14.75$	Essay 1	$\bar{x} = 15.00$
Writer	Essay 2	$\bar{x} = 13.00$	Essay 2	$\bar{x} = 9.75$
		$sd = 2.51$		$sd = 3.62$
	Adj. Mean	$= 13.35$	Adj. Mean	$= 9.98$
Messy	Essay 1	$\bar{x} = 16.33$	Essay 1	$\bar{x} = 15.78$
Writer	Essay 2	$\bar{x} = 12.11$	Essay 2	$\bar{x} = 12.44$
		$sd = 3.14$		$sd = 2.07$
	Adj. Mean	$= 11.70$	Adj. Mean	$= 12.30$

The following is from their RESULTS section (Huck & Bounds, 1972):

Prior to using analysis of covariance, the researchers tested the assumption of homogeneous within-group regression coefficients. Since this preliminary test proved to be nonsignificant ($F = 1.76, p > .10$), it was appropriate to use the conventional covariance analysis.

Results of the 2×2 analysis of covariance revealed that neither main effect was significant. However, an interaction between the legibility of the essay and the handwriting neatness of the graders was found to be significant ($F = 4.49, p < .05$). To locate the precise nature of this interaction, tests of simple main effects (Kirk, 1968, p. 481) were used to compare the two treatment conditions, first for graders with neat handwriting and then a second time for graders with messy handwriting. Results indicated that neat writers gave higher grades to the neat essay than to the messy essay ($F = 6.13, p < .05$), but that messy handwriters did not differentiate significantly between the two essays. (pp. 281–82)

- (a) From what is mentioned in the above RESULTS section, can we be confident that analysis of covariance is appropriate? Explain.
 - (b) What is the main reason for using analysis of covariance in this study?
 - (c) Should the investigators have been concerned about the homogeneity of cell population variances in this study? Why, or why not?
 - (d) Estimate the effect size for the interaction effect (see Section 4.6). Is it large or fairly large? Relate this to the sample size in the study and the significance that was found for the interaction effect.
6. Determine whether ANCOVA is appropriate for the HEADACHE data, using UNCOMF as the dependent variable and the PREUNCOMF as the covariate. What checks did you make?
 7. What is the main reason for using ANCOVA in a randomized study?
 8. Cochran, in his 1957 review article on ANCOVA, made the statement that ANCOVA will not be useful when the correlation between the dependent variable and the covariate is less than .3 in absolute value. Why did he say this?