Chapter 13
ANCOVA

13.1 Performing a One-Way ANCOVA with One Covariate

13.1.1 Visual Inspection of the Data

The following is the R code I used in looking at the Lyster data:

Get the SPSS file Lyster.Oral.sav and name it lysterO.

Code for parallel coordinate plots:

```
library(lattice)
parallel(~lysterO[2:4]|lysterO$Cond)
```

Code for mean and standard deviation summary:

```
numSummary(lysterO[,c("PreObjectID", "PostObjectID",
"DelayObjectID")],groups=lysterO$Cond)
```

Code for scatterplots found on p. 360 of the SPSS book (*A Guide to Doing Statistics in Second Language Research Using SPSS*):

```
library(lattice)
xyplot(PreObjectID~PostObjectID|Cond,layout=c(4,1),col="black",type=c("p","smooth",
"r"),data=lysterO)
#the "type" argument determines what to draw; "p"=points, "smooth"=Loess line,
"r"=regression line; I used all of them!
```

13.1.2 Checking the Assumptions for the Lyster (2004) Data

For the first assumption that there may be a strong correlation between covariates, for Lyster's data, since there is only one covariate, I do not need to worry about correlation between covariates.

For the second assumption of linearity between the covariate and the dependent variable, the scatterplots shown by using the code given above are a good way to test this. The Loess lines imposed on each of the scatterplots by group do not match the regression lines very well except in the case of the FFI only group. Clearly the slopes of the lines for the groups are not parallel, which will violate the third assumption of homogeneity of regression. This data set seems to be a good candidate for a robust analysis, and we will look at that later in the chapter. Remember, what we risk losing by continuing to conduct a parametric analysis when the data do not meet the assumptions is the power to find differences that do exist.

Another way to test whether there is **homogeneity of regression slopes** is to test for the presence of an interaction between the covariate and the treatment or grouping variable. If the interaction is not statistical, I can proceed with the normal model, according to Tabachnick and Fidell (2001, p. 292). We can do this by setting up a normal regression model for the Lyster (2004) variable of Object Identification and then adding in the covariate *plus* an interaction between the dependent variable and the covariate. Basically this will mean setting up a full factorial ANOVA with the two variables in the case of the one-way ANOVA, like this:

objectID.m1=aov(PostObjectID~Cond*PreObjectID, data=lysterO)
summary(objectID.m1)

```
                Df Sum Sq Mean Sq F value    Pr(>F)
Cond             3 169.25  56.417  4.9041   0.00427 **
PreObjectID      1 214.64 214.643 18.6582 6.455e-05 ***
Cond:PreObjectID 3  41.64  13.879  1.2065   0.31585
Residuals       56 644.22  11.504
```

The ANOVA output shows that the interaction (Cond:PreObjectID) is not statistical (p=.32). This is one of those times when we are hoping the p-value will be *larger* than p=0.05. If it is, we can conclude that the slopes of the groups on the covariate are parallel enough and that there is homogeneity of regression. If there were a statistical interaction, then you can see that that would mean that the groups performed differently on the covariate. In the case of checking for the assumption of homogeneity of slopes, the interaction (Cond:PreObjectID) is the only row of the ANOVA that you need to look at; you can ignore the other parts of the output.

13.1.3 Performing a One-Way ANCOVA with One Covariate

Performing an ANCOVA in R requires no more effort than simply including an extra term in the regression equation for an ANOVA. For example, this would be the regression equation if we simply wanted to perform a one-way ANOVA to examine the effect of group on post-test scores for the Lyster variable of Object Identification:

aov(PostObjectID~Cond, data=lysterO)

Now we can just add the PreObjectID (the covariate) term to the right of the tilde:

object.m1=aov(PostObjectID~Cond+PreObjectID, data=lysterO)
object.m0=aov(PostObjectID~Cond, data=lysterO)

We could use either aov() or lm() to model the regression; the only difference is in the format of the output. I will show the aov() modeling here. Although the syntax in R is simple, the idea for ANCOVA is that the maximal model will have different slopes and intercepts for each level of the factor. Since we have only one independent variable (as we are looking at a one-way ANOVA) there will be no need to try to simplify terms.

anova(object.m1)

```
           Df Sum Sq Mean Sq F value    Pr(>F)
Cond        3 169.25  56.417  4.8532  0.004383 **
PreObjectID 1 214.64 214.643 18.4644 6.566e-05 ***
Residuals  59 685.86  11.625
```

The summary shows that pre-test scores have a statistical effect on the post-test score. However, we also now have evidence that the condition has a statistical effect on post-test scores even with the effect of the pre-test factored out. Including the covariate means that we can factor out the influence of that variable.

Note that order can matter in the regression equation (Crawley, 2007), so that we actually get a different outcome if we model with the covariate first:

object.m2=aov(PostObjectID~PreObjectID + Cond, data=lysterO)
anova(object.m2)

```
Response: PostObjectID
            Df Sum Sq Mean Sq F value    Pr(>F)
PreObjectID  1 299.32 299.318 25.7484 4.155e-06 ***
Cond         3  84.57  28.192  2.4251  0.07456 .
Residuals   59 685.86  11.625
```

Now the factor of condition is not statistical.

After you have considered your ANCOVA, it might turn out that you will need to conduct pairwise comparisons on the variables. In this case, we know that Condition is statistical for the Lyster data, but don't know which groups differ from each other. We want to conduct pairwise comparison with the effect of the pre-test factored out, however. For this, use the glht() command we have seen previously (in the online document "One way ANOVA.One-way ANOVA test"), but use it on the ANCOVA model.

summary(glht(object.m1, linfct=mcp(Cond="Tukey")))

Assumptions for ANOVA can be checked in the normal way:

plot(object.m1)

Diagnostic plots show evidence of heteroscedasticity, evidence of some departure from normality at the ends of the Q-Q plot, and several influential outliers (just what we've come to expect from our real-life data sets!).

Performing a One-Way ANCOVA with One Covariate

1. Model your ANOVA with the aov() command (lm() can also be used, but a summary will result in regression output, not ANOVA tables). Add your covariate to your equation (do not have it enter into interactions with the variables of interest). Here is an example where scores on the Object Identification task (measure on a post-test) are modeled according to the condition that the subjects were in with a covariate of scores on the Object Identification task in a pre-test:

object.m1=aov(PostObjectID~Cond+PreObjectID, data=lysterO)

2. Examine output using the anova() command. Whether you pay attention to the statistical significance of the covariate depends on your question. You will, however, want to note whether your variable of interest is statistical even when the effects of the covariate are factored out.

3. If you would like to perform post-hoc comparisons, use the regression model in the glht() procedure to compare adjusted means.

13.2 Performing a Two-Way ANCOVA with Two Covariates

In my own research I wanted to examine the effect of studying English at a younger age if the amount of input was minimal (four hours or less a week). I collected data on both grammaticality judgments in English (gjtscore) and phonemic discrimination ability (RLWtest). However, I also assumed that the total amount of input in English over however many years the student had studied could factor into their abilities, and I also thought perhaps the participants' own language aptitude would affect their scores. So I decided to factor these variables out of the equation by measuring them and using them as covariates. But before I can proceed with a two-way ANCOVA, I need to check whether this data fulfills the requirements for doing an ANCOVA.

13.2.1 Checking the Assumptions for the Larson-Hall (2008) Data

For the first assumption that there may be a strong correlation between covariates, for my own study (Larson-Hall, 2008) I used two covariates, and we would check the correlation between the two variables of language aptitude (aptscore) and total amount of input in English (totalhrs) by simply following the menu STATISTICS > SUMMARIES > CORRELATION MATRIX in R Commander and choosing those two variables. The correlation between these two covariates is $r=0.08$. This is not cause for worry at all.

For the assumption that the regression slopes are equal we can test for the presence of an interaction between the covariates and the grouping variable (here, erlyexp). We'll test for an interaction between each covariate and the grouping variable one at a time.

```
check1=aov(gjtscore~erlyexp* aptscore, data=larsonhall2008)
summary(check1)
check2=aov(gjtscore~erlyexp* totalhrs, data=larsonhall2008)
summary(check2)
```

In neither case was the interaction statistical, so I may proceed with the ANCOVA analysis.

13.2.2 Performing a Two-Way ANCOVA with Two Covariates

Conceptually there is not much difference between performing a one-way or a two-way ANCOVA in R, but this section will illustrate how it can be done. If you are following along with me, import the SPSS file LarsonHall2008.sav as larsonhall2008. In this study I looked at whether exposure in childhood to formal English lessons resulted in Japanese college students being able to perform better on a test of the English R/L/W contrast and on a grammaticality judgment test in English. Since Japanese learners of English who began studying English at a younger age might have more hours of exposure, I wanted to use this variable as a covariate. In other words, I wanted to factor this out of the comparison between the group that had had early exposure and the group that hadn't. I also thought it was possible language aptitude might be involved, so I measured that in order to statistically take that factor out of the equation. And that is what we will do here—try taking both the factor of hours of exposure and language aptitude out of the equation!

We'll look at the question of scores on the grammaticality judgment test. My original research question was whether the group which had early exposure differs from the group which did not, but in order to illustrate a two-way ANCOVA I will also include the independent variable of the sex of the participant. Here is the ANOVA model in R:

lh.m1=aov(gjtscore~erlyexp*sex+totalhrs+aptscore, data=larsonhall2008)

Is this the only possible model? No. The covariates will not enter into any interactions in the model, but the two independent variables are presumed to have an interaction in this first model. Let's check out whether this assumption is justified.

summary(lh.m1)

```
            Df  Sum Sq Mean Sq F value   Pr(>F)
erlyexp      1   251.1  251.07  2.4986 0.115576
sex          1   937.1  937.14  9.3263 0.002576 **
totalhrs     1   429.9  429.94  4.2787 0.039919 *
aptscore     1    47.2   47.24  0.4702 0.493727
erlyexp:sex  1     0.6    0.61  0.0060 0.938158
Residuals  194 19493.8  100.48
```

The summary shows that there is a statistical main effect for sex, but not for early exposure or the interaction between sex and early exposure, at least when the effects of aptitude and hours of exposure are factored out. Let's try removing the interaction between sex and early exposure.

lh.m2=update(lh.m1,~.-erlyexp:sex, data=larsonhall2008)
summary(lh.m2)

```
           Df  Sum Sq Mean Sq F value   Pr(>F)
erlyexp     1   251.1  251.07  2.5114 0.114646
sex         1   937.1  937.14  9.3741 0.002511 **
totalhrs    1   429.9  429.94  4.3006 0.039412 *
aptscore    1    47.2   47.24  0.4726 0.492619
Residuals 195 19494.4   99.97
```

Nothing has changed—the early exposure variable is still not statistical. Shall we keep this simpler model?

anova(lh.m1, lh.m2)

```
Model 1: gjtscore ~ erlyexp * sex + totalhrs + aptscore
Model 2: gjtscore ~ erlyexp + sex + totalhrs + aptscore
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    194 19494
2    195 19494 -1  -0.60642 0.006 0.9382
```

Yes. There is no difference between the models, so we will keep the simpler one. At this point I don't want to eliminate the variable of early exposure, however, because my main question is whether this variable is statistical when the effect of aptitude and input is factored out. My analysis has answered that question—it is not.

Performing a Two-Way ANCOVA with Two Covariates

1. Model your ANOVA with the aov() command. Add your covariates to your equation (but do not have them enter into interactions with the variables of interest). Here is an example where scores on a grammaticality judgment test are modeled according to whether the participant studied English at an early age (erlyexp) and the sex of the participant (sex), but the equation is modified by including a covariate of total hours of input and scores on an aptitude test:

lh.m1=aov(gjtscore~erlyexp*sex+totalhrs+aptscore, data=larsonhall2008)

2. Examine output using the anova() command.

13.3 Performing a Robust ANCOVA in R

In a robust ANCOVA we would like to eliminate the assumption of parametric models that the regression lines of all groups are parallel (Wilcox, 2005). The method I will present in this section does not make any assumptions about the linearity of regression lines, allows heteroscedasticity, and performs well even if data is non-normal. The general idea of the method is to use a Loess smoother and then make comparisons between groups at specific points along the smooth line. This method also trims the data (20% is recommended) and bootstraps it with the recommended bootstrap-t. The function will pick five arbitrary points at first, but a researcher can specify the points themselves as well, and will probably want to after looking at scatterplots with smooth lines that will result from the output.

Wilcox's (2005) function ancboot from the WRS library can only be used with one covariate, and it can only examine two groups at a time. In other words, the only time this function would work without having to first subset your data would be if you have an independent variable with only two groups and one covariate. This function has been shown to have good power when ANCOVA assumptions are not met, but another advantage, Wilcox says, it that "[e]ven when ANCOVA has relatively good power, an advantage of using trimmed means with a running-interval smoother is that the goal is to determine where the regression lines differ and by how much, and this is done without assuming that the regression lines are straight" (2005, p. 526).

To use the ancboot function, first the data must be subsetted into separate data frames for each group as shown below (an example of my published work using this function can be

found in Larson-Hall, 2008). We'll continue to work with the Object Identification variable from the previous section, so use the Lyster.Oral.sav file that you imported as lysterO.

```
lyster.prompt=subset(lysterO,subset=Cond=="FFIprompt")
lyster.recast=subset(lysterO,subset=Cond=="FFIrecast")
lyster.only=subset(lysterO,subset=Cond=="FFIonly")
lyster.comp=subset(lysterO,subset=Cond=="Comparison")
```

Here is an example of ancboot() used to compare the FFI prompt and FFI recast groups on the pre-test and immediate post-test scores. Notice that x1 is pre-test scores from one group and x2 is pre-test scores from the other group. Then y1 is post-test scores from the first group and y2 is post-test scores from the second group.

```
x1=lyster.prompt$PreObjectID
y1=lyster.prompt$PostObjectID
x2=lyster.recast$PreObjectID
y2=lyster.recast$PostObjectID
```

ancboot(x1,y1,x2,y2,fr1=1,fr2=1,tr=.2, nboot=599, plotit=T, pts=NA)	
ancboot (x1, y1, x2, y2)	Performs an ANCOVA using trimming and a percentile bootstrap method; the data for group 1 are stored in x1 and y1, and the data for group 2 are stored in x2 and y2.
fr1=1, fr2=1	The values of the span (f) for the groups, defaults to 1; span refers to how frequently data are sampled by the running-interval smoother, meaning how smooth the smooth line looks.
tr=.2	Specifies amount of trimming, default is 20%.
nboot=599	Specifies number of bootstrap samples to use; 599 is default.
plotit=T	Plots a scatterplot with Loess smoothers drawn in.
pts=NA	If not specified, the function will automatically choose five points at which to compare groups. The points can be specified by the user, such as pts=c(1,3,5).

The output returns the points where data are compared, confidence intervals for the test that trimmed means of the groups are equivalent at that point, and a *p*-value for the test. Wilcox says that the function determines "a critical value based on the Studentized maximum modulus distribution" (2005, p. 527). The plot that is returned is also very useful. Here is the text output from the command run above:

```
[1] "Note: confidence intervals are adjusted to control FWE"
[1] "But p-values are not adjusted to control FWE"
[1] "Taking bootstrap samples. Please wait."
$output
      X n1 n2 DIF      TEST    ci.low   ci.hi    p.value
[1,] 11 14 13 0.8 0.3953362 -5.371443 6.971443 0.7262104
[2,] 11 14 13 0.8 0.3953362 -5.371443 6.971443 0.7045075
[3,] 11 14 13 0.8 0.3953362 -5.371443 6.971443 0.6944908
[4,] 11 14 13 0.8 0.3953362 -5.371443 6.971443 0.6978297
[5,] 12 14 14 0.6 0.3126638 -5.252441 6.452441 0.7679466

$crit
[1] 3.049744
```

The results show that the FFI prompt and FFI recast groups were tested only at two points on the pre-test (we set it up so that the pre-test was on the x-axis and the post-test on the y-axis), points 11 and 12 (we know this because only 11 and 12 are listed in the column with "X"). The n1 and n2 specify how many subjects were tested at each of those points. If we decide we want to test at more intervals, we find that the number of participants in those points is smaller, which is why the automatic procedure only tested at points 11 and 12.

ancboot(x1,y1,x2,y2,fr1=1,fr2=1,tr=.2, nboot=599, plotit=T, pts=c(8,10,12,16,18))

```
        X n1 n2         DIF        TEST        se     ci.low     ci.hi   p.value
[1,]    8 10  9  1.8333333  0.8003339  2.290710  -18.49807  22.16473 0.4156928
[2,]   10 12 10  1.5000000  0.6642112  2.258318  -18.54390  21.54390 0.5342237
[3,]   12 14 14  0.6000000  0.3126638  1.918994  -16.43220  17.63220 0.7328881
[4,]   16  7 10 -1.0666667 -0.6330564  1.684947  -16.02156  13.88823 0.4958264
[5,]   18  6  8 -0.4166667 -0.2104634  1.979759  -17.98819  17.15485 0.7929883
```

The "DIF" column gives the difference in scores at that point, and the "TEST" gives the test statistic for the hypothesis that the mean of one group is equal to the mean of the other at that point. The "se" column gives the standard error for the test statistic, and the "p.value" column gives the *p*-value for the test. None of the points show any statistical difference between the groups. The "ci.low" and "ci.hi" columns give the lower and upper ends of the 95% confidence interval for the difference between the groups. The last part of the text output lists the critical value used to evaluate whether the test is statistical. If the value in the "TEST" column is *below* this critical value, the test will not be statistical. Wilcox notes that this critical value will be adjusted if fewer than five points appropriate for testing can be found.

The plot that is called for is extremely helpful in understanding what is happening, and shows the points with the smooth lines on them (Figure 13.1).
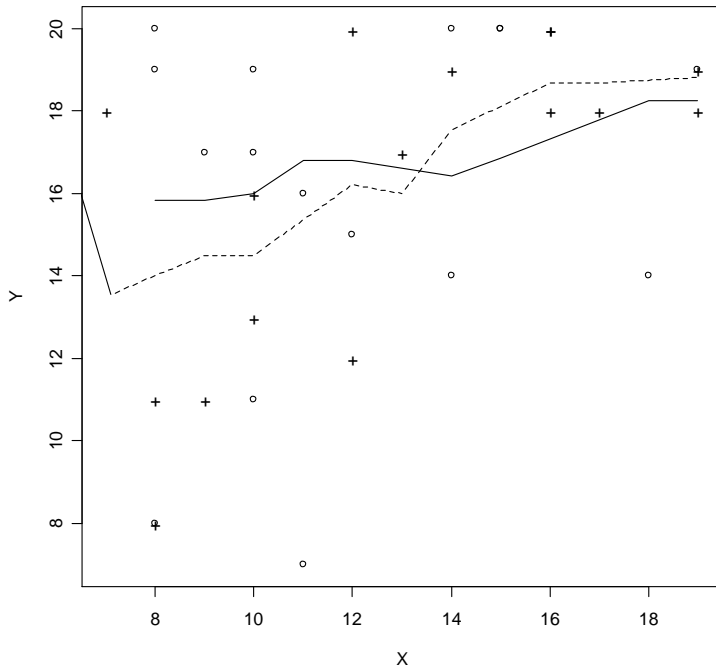
Figure 13.1 Output from Wilcox's ancboot function showing Loess smooth lines for ANCOVA comparison.

When I tried to make a comparison between the FFI prompt and FFI only group, I got the following error message:

```
Warning in min(sub[vecn >= 12]) :
  no non-missing arguments to min; returning Inf
Warning in max(sub[vecn >= 12]) :
  no non-missing arguments to max; returning -Inf
Warning in near(x1, x1[isub[i]], fr1) : NAs introduced by coercion
Warning in near(x2, x1[isub[i]], fr2) : NAs introduced by coercion
Error in var(y) : 'x' is empty
```

Like many error messages, this one was confusing, as there were no NAs in the data sets. However, I got the idea that maybe the program was having trouble picking out points to test at, and when I specified just points 10 and 12 the function worked, although at both of these points one group had only five points. This is probably what was problematic, as Wilcox says "the points among the covariates at which the groups will be compared are determined by the function; it finds a point among the x1 values that has the deepest halfspace depth, plus the points on the .5 depth contour, and the groups are compared at these points, provided that the corresponding sample sizes are at least 10" (2005, p. 533).

Because only two groups can be compared at one time on two variables, this robust ANCOVA is more limited that the parametric version, but I think that, when it can work, it is an informative function.

Performing a Robust ANCOVA

1. Wilcox (2005) lists several possible commands for performing robust ANCOVAs. In this section I focus on just one, ancboot(), which does not assume that groups have parallel regression lines, and uses means trimming and also a bootstrap-t procedure.

2. Data need to be arranged so that data for only one group are in a data frame.

3. The command is:

ancboot(x1,y1,x2,y2,fr1=1,fr2=1,tr=.2, nboot=599, plotit=T, pts=NA)

where x1=data for a particular variable from group 1 at time 1, and y1=data for the same variable from group 1 at time 2. Similarly, x2=data for the same variable as x1 but from group 2 at time 1, and y2=data for the same variable from group 2 at time 2. The rest of the specifications are default but can be changed.

4. Errors might arise if there are too few data points at specific points of comparison, but this can be controlled via the pts term of the command. A scatterplot called from this command helps show visually what points of comparison are of interest.

13.4 Application Activities for ANCOVA

1. Class Time. Import the data set I have called ClassTime.sav as classtime (this data set was taken from Howell, 2002, p. 629, but I adapted it to reflect a design that will be associated with the second language research field). Let's pretend that a researcher who is in charge of teaching Arabic at the university level notices that there seems to be a difference in how students in her 8 a.m. class respond to her teaching versus how students in the later classes respond. At the start of a new school year she gives them an initial test of their enthusiasm and motivation for learning Arabic. There are 30 items, which contain a ten-point Likert scale, where a higher score is more positive about the class. The researcher averages their answers together for a score out of 10. She then administers the same test at the end of the semester. The researcher has five classes, one at 8 a.m., one at 10 a.m., one at 11 a.m., one at 1 p.m., and one at 2 p.m. This study could be analyzed with an RM ANOVA (if the data were arranged in the "wide" format), but the researcher decides to analyze it with an ANCOVA using the pre-test scores as a covariate so that any differences among the post-test scores due to variability in pre-test scores will be controlled. Use PreTestScores as the covariate, PostTestScores as the dependent variable, and TimeOfClass as the independent variable. First check the special assumptions for ANCOVA. Even if the data violates the assumptions, go ahead and perform the ANCOVA. What are the results of the parametric ANCOVA? Then run a robust ANCOVA on the comparison between the 10 a.m. and 1 p.m. class. Do the results differ from the parametric ANCOVA?

2. Larson-Hall (2008). Import the SPSS file LarsonHall2008.sav as larsonhall2008. In the online document "ANCOVA: Two-way ANCOVA with two covariates" you saw an analysis with two covariates (aptscore and totalhrs) with the dependent variable of grammaticality judgment test scores. Perform the same analysis using the dependent variable of the phonemic discrimination test scores (rlwscore). Start by seeing whether the model satisfies the special ANCOVA assumptions. If it does, start with the maximal model and simplify if

273

possible. Is early exposure a statistical factor when the effect of aptitude and input is factored out?