Lab Tutorial 7: Designs with Covariates: Analysis of Covariance (ANCOVA)

Chris M. Fiacconi

Overview

To this point in the course, we have discussed how we can utilize the general linear model to predict an individual's score based on a known predictor variable(s) that is *discrete* or *categorical* (i.e., IV such as Drug). However, sometimes one of the predictor variables of interest is *continuous* - that is, it can take on any value within a wide range (e.g., Age). Often, such predictor variables are investigated simultaneously with a second categorical IV, where the goal is to evaluate the effect of the categorical IV after statistically removing the influence of the continuous predictor variable. In such cases, the continuous predictor variable is known as a *covariate* because it shares a relationship with the DV.

Typically, the covariate is not of primary interest. Rather, the covariate is included in the analysis either to reduce error variability and thus make it easier to observe an effect of the categorial IV, or to control for a potential confound in cases where there are pre-existing differences between experimental groups on the covariate. In general, there are two approaches to dealing with designs where covariates are of concern: **1)** Blocking, and **2)** Analysis of Covariance (ANCOVA).

Blocking

One approach to analyzing designs with covariates is known as *blocking*. In this procedure, participants who share similar scores on the covariate are grouped together to form blocks. For example, imagine you were interested in testing older adults (aged 60 - 80+) on a motor control task to see if different emotional states affect their performance. You might randomly assign your older participants to three groups: positive induction, negative induction or neutral induction. Prior to the motor control task, you could show participants in each group pictures from one of these three emotion categories to induce the corresponding emotional state. You could then examine how different emotional states influence motor control by comparing performance across each group.

You might expect, however, that motor control is affected by age independently of emotional state. Therefore, if you relied solely on random assignment of participants to the 3 emotion conditions, it is possible that participants' age would not be evenly distributed across the three groups. This could confound your interpretation of any effects of your primary manipulation (i.e., emotion) on motor control performance. Instead you might divide participants evenly into 3 sub-groups (blocks) based on their age (i.e., 60-69, 70-79, 80+) and then pseudo-randomize assignment such that each emotion induction condition contains an equal number of participants from each block. By dividing participants into blocks based on age, we can include Block as a factor in our analysis which will increase the likelihood of detecting an effect of emotion by statistically removing orthogonal variability associated with age. To illustrate this procedure, I've used the data from MDK Chapter 9 Table 11.

```
library(tidyverse)
options(contrasts=c("contr.sum","contr.poly"))

age.motor.data<-read.table(file="Ch9T11.txt",header=T)
str(age.motor.data)
```

```
## 'data.frame':    18 obs. of  4 variables:
```

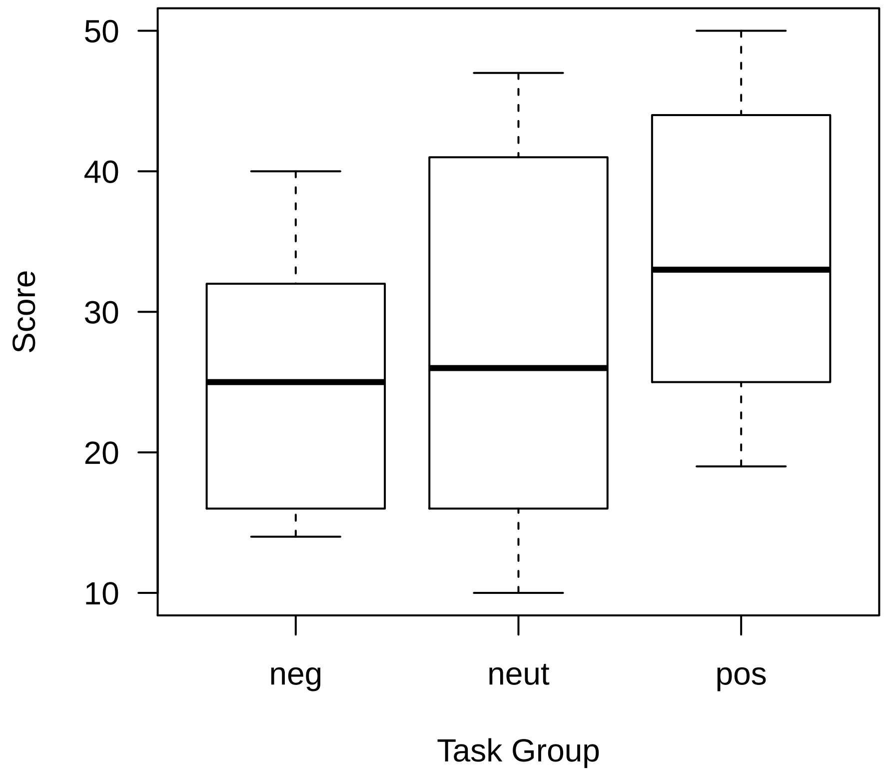

Figure 1: Boxplot depicting data taken from MDK Chapter 9 Table 11

```
##  $ Block: int  1 1 1 1 1 1 2 2 2 2 ...
##  $ Task : Factor w/ 3 levels "neg","neut","pos": 1 1 2 2 3 3 1 1 2 2 ...
##  $ X    : int  60 69 62 66 63 67 74 76 71 78 ...
##  $ Y    : int  14 24 10 16 19 25 16 26 22 30 ...
```

Here, we have a total of 18 observations divided into three groups of subjects (neg, neut, pos) each with 2 participants from the three age blocks. Let's take a look at the data separated by the Task factor:

```
boxplot(Y~Task,data=age.motor.data,las=1,ylab="Score",xlab="Task Group")
```

First, let's get the means separated by Task and Block:

```
age.motor.means<-age.motor.data %>% group_by(Task,Block) %>% summarize(Mean=mean(Y))
print(age.motor.means)
```

```
## # A tibble: 9 x 3
## # Groups:   Task [?]
##   Task  Block  Mean
##   <fct> <int> <dbl>
## 1 neg       1    19
## 2 neg       2    21
## 3 neg       3    36
## 4 neut      1    13
## 5 neut      2    26
## 6 neut      3    44
```

```
## 7 pos        1    22
## 8 pos        2    33
## 9 pos        3    47
```

Now, let's test whether our Task manipulation had any effect on motor control using a one-way ANOVA:

```
motor.aov<-aov(Y~Task,data=age.motor.data)
print(summary(motor.aov))
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Task         2  241.3   120.7   0.824  0.458
## Residuals   15 2196.7   146.4
```

Even though it appeared from the group means and our boxplot that our Task manipulation seemed to affect motor control performance, the ANOVA suggests that the group means do not differ. Although we might be tempted to conclude that emotion does not appear to influence motor control among older adults, inspection of the means for each of our Blocks suggests that there are substantial differences in motor control between each age group. In the preceeding analysis, we ignored these differences such that they effectively increased the error variance, $SS_{Residuals}$. By including Block as a factor in our analysis we can remove this variability thereby reducing the error variance and making it easier to "see" the effect of Task:

```
age.motor.data$Block<-as.factor(age.motor.data$Block) # Ensure Block is treated as a factor
age.motor.aov<-aov(Y~Block*Task,data=age.motor.data)
print(summary(age.motor.aov))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Block        2 1825.3   912.7  32.339 7.78e-05 ***
## Task         2  241.3   120.7   4.276   0.0495 *
## Block:Task   4  117.3    29.3   1.039   0.4385
## Residuals    9  254.0    28.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Now that we've removed the variability associated with our Block variable, there is indeed a statistically significant effect of Task, $F(2, 9) = 4.28, p = .0495$. Note that in each ANOVA, $SS_{Task} = 241.3$. In other words, by including Block as a factor we did not increase SS_{Task}. Rather we *reduced* $SS_{Residuals}$ which had the effect of increasing F.

The lesson here is that creating homogenous blocks of participants that have similar scores on a covariate (i.e., age), we can remove what otherwise would be error variance that might obscure the effect of primary interest. Although blocking can be useful, it is often better to use an ANCOVA to analyze the data where possible. I will illustrate how to do this in the next section.

Analysis of Covariance (ANCOVA)

Using the *blocking* approach above, we transformed a continuous variable (i.e., age) into a categorical variable (i.e, block) by discretizing age into 3 bins. When use this approach, we are effectively losing information because we treat ages 63 and 67 as essentially the same (both belong to bin #1). An alternative approach is to take advantage of the full range of age values to further reduce our error variance by using an ANCOVA in which we compare the following full model:

$$Y_{ij} = \mu + \alpha_j + \beta X_{ij} + \epsilon_{ij}$$

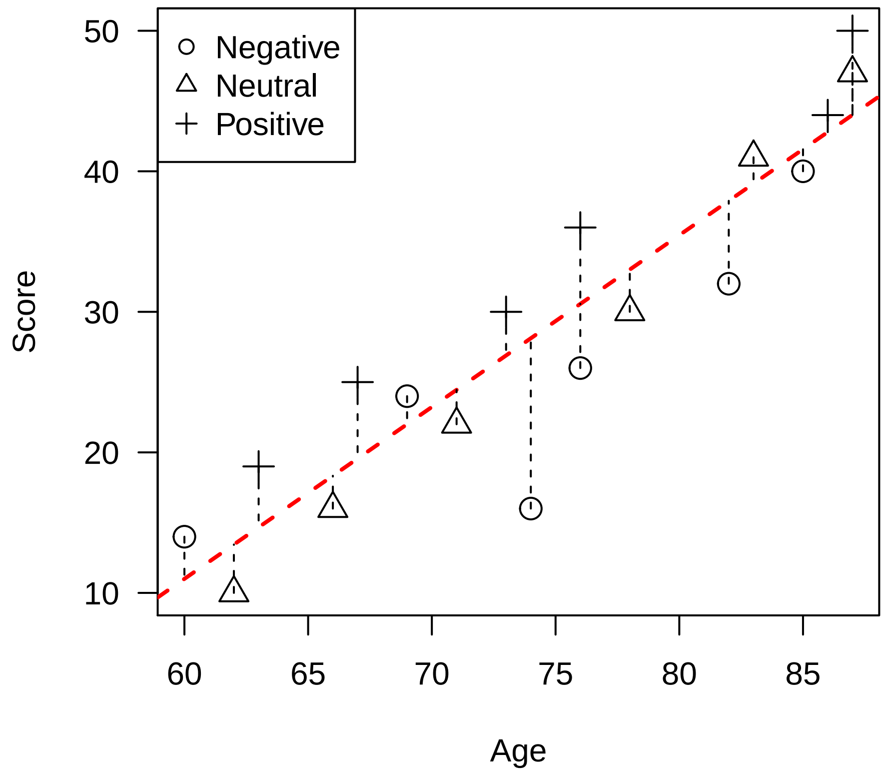

Figure 2: Plot showing residuals based on restricted ANCOVA model. Data taken from from MDK Chapter 9 Table 11

to the following restricted model:

$$Y_{ij} = \mu + \beta X_{ij} + \epsilon_{ij}$$

In the case of ANCOVA, μ represents the value of Y when $X = 0$ (Y-intercept), α_j represents the effect of being in group j, β represents the slope relating the covariate (X) to the DV, and ϵ_{ij} is the leftover prediction error for individual i in group j.

The relative fit of these models is computed using an F-statistic of the form we've seen before:

$$F = \frac{(E_R - E_F)/(df_R - df_F)}{E_F/df_F}$$

We can visualize the influence of including the covariate in our model by plotting how the residuals are computed in the restricted and full models, respectively. For this illustration, we'll again use the data from MDK Chapter 9 Table 11.

```
with(age.motor.data,plot(x=X,y=Y,type="p",las=1,pch=as.numeric(Task),cex=1.5,xlab="Age",ylab="Score"))
legend("topleft",legend=c("Negative","Neutral","Positive"),pch=c(1,2,3))

lm.restricted<-lm(Y~X,data=age.motor.data)
abline(lm.restricted,lty=2,lwd=2,col="red")

segments(x0=age.motor.data$X,y0=age.motor.data$Y,x1=age.motor.data$X,
    y1=(lm.restricted$coefficients[2]*age.motor.data$X + lm.restricted$coefficients[1]),lty=2,lwd=1)
```

As you can see, in the restricted model, the residuals represent the distance from each point to a single regression line that captures the relationship between Age and the DV. Any potential effect of Task is ignored in this model.

To account for the effect of Task, we can perform an ANCOVA which is relatively straightforward to implement in R. Let's first do the ANCOVA by comparing nested models using the **lm** and **anova** commands:

```
restricted.model<-lm(Y~1+X,data=age.motor.data)
full.model<-lm(Y~1+X+Task,data=age.motor.data)

anova(restricted.model,full.model)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ 1 + X
## Model 2: Y ~ 1 + X + Task
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     16 387.62
## 2     14 210.80  2    176.82 5.8718 0.01407 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

You can also perform an ANCOVA using the **aov** command:

```
age.motor.ancova<-aov(Y~X+Task,data=age.motor.data)
print(summary(age.motor.ancova))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## X            1 2050.4  2050.4 136.176 1.34e-08 ***
## Task         2  176.8    88.4   5.872   0.0141 *
## Residuals   14  210.8    15.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As you can see, the effect of Task is statistically significant, $F(2, 14) = 5.87, p = .014$. Notice also that the obtained F-value for Task is larger when using an ANCOVA than when using the blocking approach. This increased senstivity reflects the fact that the ANCOVA considers all values of the covariate rather than just the discretized bins that result from the formation of blocks.

Let's now visualize what the residuals look like in the full model that accounts for the effect of Task:

```
# Sub-divide data based on Task
neg.data<-filter(age.motor.data,Task=="neg")
neut.data<-filter(age.motor.data,Task=="neut")
pos.data<-filter(age.motor.data,Task=="pos")

# Get full model coefficients
full.coef<-dummy.coef(age.motor.ancova)
print(full.coef)
```

```
## Full coefficients are
##
## (Intercept):    -61.02348
## X:                1.204775
## Task:                 neg      neut       pos
##                 -3.198143 -1.065606  4.263748
```

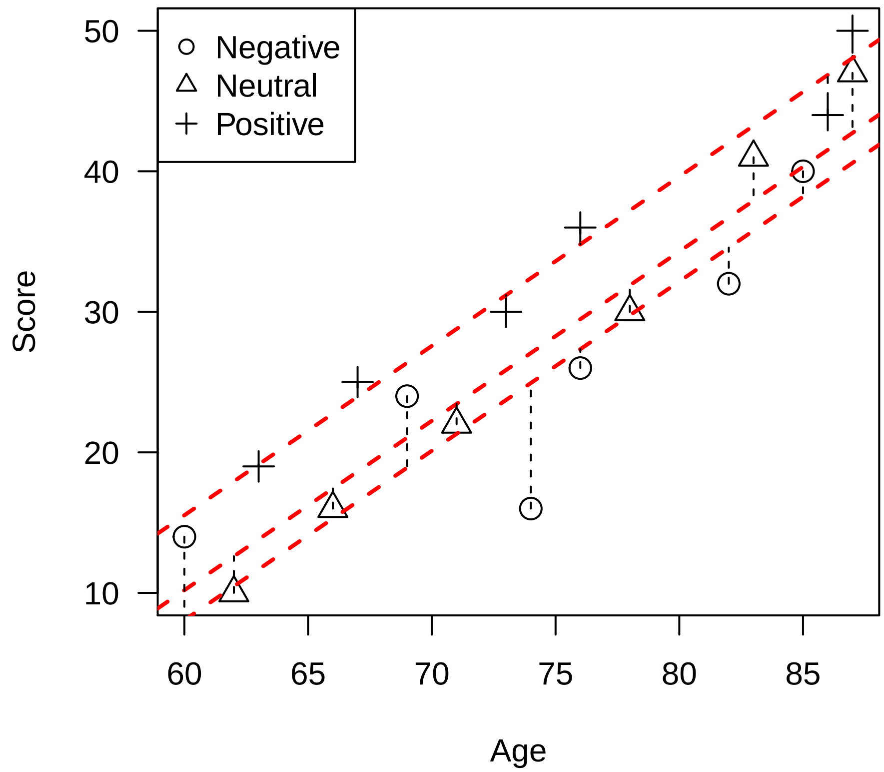

Figure 3: Plot showing residuals based on the full ANCOVA model. Data taken from from MDK Chapter 9 Table 11

```
# Define slope and intercept from full model
slope<-full.coef[[2]]
mod.intercept<-full.coef[[1]]

with(age.motor.data,plot(x=X,y=Y,type="p",las=1,pch=as.numeric(Task),cex=1.5,xlab="Age",ylab="Score"))
legend("topleft",legend=c("Negative","Neutral","Positive"),pch=c(1,2,3))

abline(mod.intercept+full.coef$Task[1],slope,lty=2,lwd=2,col="red") # Draw slope for negative group
abline(mod.intercept+full.coef$Task[2],slope,lty=2,lwd=2,col="red") # Draw slope for neutral group
abline(mod.intercept+full.coef$Task[3],slope,lty=2,lwd=2,col="red") # Draw slope for positive group

# Draw dotted lines depicting residuals for negative group
segments(x0=neg.data$X,y0=neg.data$Y,x1=neg.data$X,
    y1=(slope*neg.data$X + mod.intercept+full.coef$Task[1]),lty=2,lwd=1)

# Draw dotted lines depicting residuals for neutral group
segments(x0=neut.data$X,y0=neut.data$Y,x1=neut.data$X,
    y1=(slope*neut.data$X + mod.intercept+full.coef$Task[2]),lty=2,lwd=1)

# Draw dotted lines depicting residuals for positive group
segments(x0=pos.data$X,y0=pos.data$Y,x1=pos.data$X,
    y1=(slope*pos.data$X + mod.intercept+full.coef$Task[3]),lty=2,lwd=1)
```

You should notice that the residuals associated with the full model are much smaller than those for the restricted model. This reduction in the residuals is the reason why ANCOVA provides a sensitive test of our manipulation of interest.

Also notice that the slopes (β) are constrained to be equal for all 3 groups. The slope is determined by pooling the individual best-fitting β's for each group. You should also be aware that ANCOVA captures only the *linear* relationship between the covariate and the DV, although it can be adapted to account for non-linear relationships as well.

We can contrast ANCOVA with ANOVA in which the covariate, Age, is ignored and only the main effect of Task is examined. The full model for such an ANOVA is:

$$Y_{ij} = \mu + \alpha_j + \epsilon_{ij}$$

and the residuals for such a model can be visualized as follows:

```r
with(age.motor.data,plot(x=X,y=Y,type="p",las=1,pch=as.numeric(Task),cex=1.5,xlab="Age",ylab="Score"))
legend("topleft",legend=c("Negative","Neutral","Positive"),pch=c(1,2,3))

nCond<-3
cond.means<-age.motor.data %>% group_by(Task) %>% summarize(Mean=mean(Y))

abline(h=cond.means$Mean[1],lty=2,lwd=2,col="red")
abline(h=cond.means$Mean[2],lty=2,lwd=2,col="red")
abline(h=cond.means$Mean[3],lty=2,lwd=2,col="red")

sorted.data<-arrange(age.motor.data,Task) # Sort data by Task

ii<-1
for (kk in 1:nCond){
  segments(x0=sorted.data$X[ii:(ii+5)],y0=sorted.data$Y[ii:(ii+5)],x1=sorted.data$X[ii:(ii+5)],
           y1=rep(cond.means$Mean[kk],times=6),lty=2,lwd=1)
  ii<-ii+6
}
```

As can be seen, failing to include the covariate in the model greatly increases the residuals because variation in Age is left unaccounted for and treated as error variance. In an ANCOVA, the slopes for each group are not restricted to 0, meaning that variability in Age can be separated from error variance resulting in a reduction of the latter.

Adjusted Means

While an ANOVA evaluates the null hypothesis that the raw group means are equal, an ANCOVA evaluates the null hypothesis that the *adjusted* group means are equal. These adjusted means represent what each group mean should be if all groups had the same mean value on the covariate, $\bar{X}$.

We can calculate the adjusted means using the following formula:

$$\bar{Y}_j' = \mu + \beta\bar{X} + \alpha_j$$

where $\bar{Y}_j'$ is the adjusted mean for group j, and $\bar{X}$ is the grand mean for the covariate, X.

First let's look at the raw group means, and then we'll calculate the adjusted means:

```r
# Look at raw group means
print(cond.means)
```

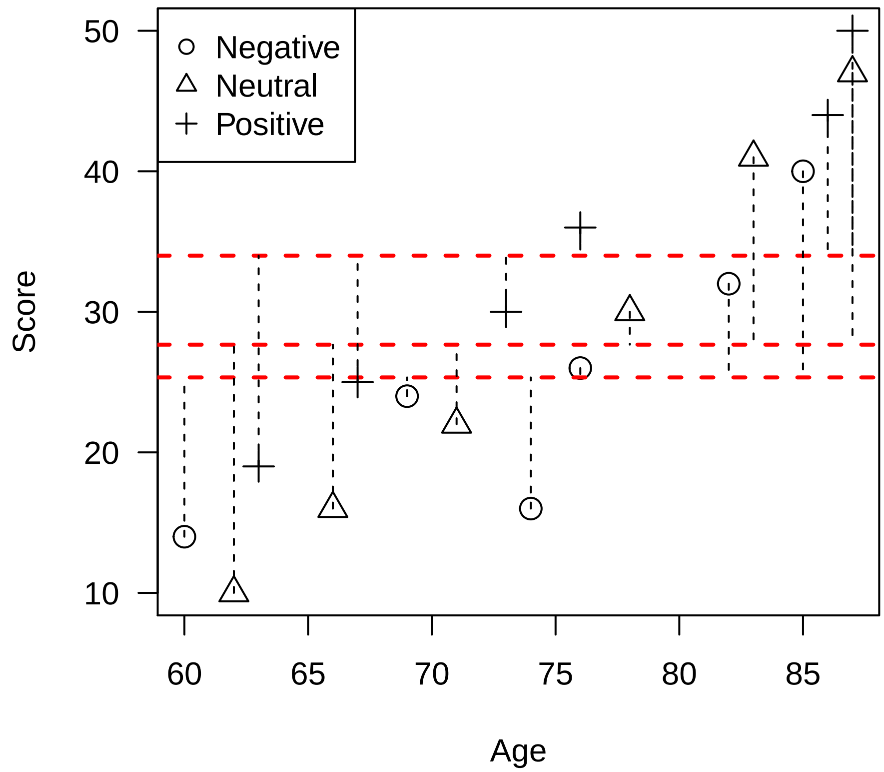

Figure 4: Plot of residuals based on full ANOVA model with no covariate

```
## # A tibble: 3 x 2
##   Task   Mean
##   <fct> <dbl>
## 1 neg    25.3
## 2 neut   27.7
## 3 pos    34
```

```
# Calculate adjusted group means
mean.X<-mean(age.motor.data$X) # Calculate grand mean of covariate scores
print(full.coef)
```

```
## Full coefficients are
##
## (Intercept):     -61.02348
## X:                1.204775
## Task:                  neg      neut       pos
##                  -3.198143 -1.065606  4.263748
```

```
adj.neg<-(-61.02) + 1.20*mean.X + (-3.20)
print(adj.neg)
```

```
## [1] 25.44667
```

```r
adj.neut<-(-61.02) + 1.20*mean.X + (-1.07)
print(adj.neut)
```

```
## [1] 27.57667
```

```r
adj.pos<-(-61.02) + 1.20*mean.X + (4.26)
print(adj.pos)
```

```
## [1] 32.90667
```

Comparisons Among Group Means

Unlike the follow-up comparisons/contrasts that follow a significant one-way ANOVA, follow-up tests for a significant main effect of the group factor in the case of ANCOVA should be performed on the **adjusted means**.

To perform a linear contrast, the calculation of psi (ψ) is similar to how we've been doing it all along except that the adjusted means ($\bar{Y}_j'$) rather than the raw means ($\bar{Y}_j$) are used:

$$\psi = \sum_{j=1}^{a} c_j \bar{Y}_j'$$

However, calculating the variance for the denominator of our F-statistic (s_ψ^2) is considerably more complicated. However, the overall F-statistic for the linear contrast bears a similar form to what we've seen thus far:

$$F = \frac{\psi^2}{s_\psi^2}$$

To save you the headache of calculating s_ψ^2, we can use the **psi.adj** function also written by Patrick J. Bennett (McMaster University) to do the linear contrast for us. Let's first load the function into R.

```r
source("psi_adj.R") # Make sure psi_adj.R file is in your working directory!
```

Now we can perform a linear contrast on the adjusted means. If such comparisons are complex and post-hoc, Scheffe's Method of controlling $\alpha_{FW} = .05$ should be used. Looking at the adjusted means, we can see that the positive emotion group seems to score the highest on motor control, whereas the negative and neutral groups appear to perform similarly. Let's test this hypothesis using a linear contrast with weights (-1, -1, 2). To perform this contrast, we also need to get the $MS_{Residuals}$ from the full ANCOVA model (*age.motor.ancova*) using the **print** and **summary** commands.

```r
print(summary(age.motor.ancova))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## X            1 2050.4  2050.4 136.176 1.34e-08 ***
## Task         2  176.8    88.4   5.872   0.0141 *
## Residuals   14  210.8    15.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
MS.resid<-15.1
weights<-c(-1,-1,2)
y.adj<-c(adj.neg,adj.neut,adj.pos)
X<-age.motor.data$X
G<-age.motor.data$Task

psi.adj(y.adj=y.adj,x=X,group=G,c.weights=weights,ms.error=MS.resid)
```

```
## $psi
## [1] 12.79
##
## $psi.var
## [1] 15.14054
##
## $F
## [1] 10.80437
##
## $p.value
## [1] 0.005400166
```

Because this contrast was selected post-hoc, *after* inspecting the pattern of adjusted means, we need to use the Scheffe Method to control $\alpha_{FW} = .05$.

```
F.crit<-qf(p=.95,df1=3-1,df2=18-3-1) # df1 = a - 1, df2 = N - a - 1
F.scheffe<-(3-1)*F.crit
print(F.scheffe)
```

```
## [1] 7.477784
```

Note that our observed $F = 10.80 > F$ scheffe $= 7.48$, so we reject the null hypothesis that the adjusted mean for the positive group is equal to the average of the adjusted means for the negative and neutral groups.

Effect Size

While so far we have emphasized the benefits of including a covariate in our analysis, measures of the effect size of group are better calculated *ignoring* the covariate. This is because generally our intent is not to estimate the size of the group effect based on the adjusted means, but rather the raw means. Otherwise our effect size measure would only apply for a subset of the population where the scores on the covariate, X, are equal.

Therefore, we can calculate η^2 based on the following full and restricted models:

$$Y_{ij} = \mu + \alpha_j + \epsilon_{ij}$$

$$Y_{ij} = \mu + \epsilon_{ij}$$

Note that we're not including the covariate in these models and so the $SS_{Residuals}$ term in the η^2 calculation will include variability that would otherwise be attributed to the covariate.

η^2 can be calculated as follows:

$$\eta^2 = \frac{SS_{Task}}{SS_{Residuals}}$$

```r
print(summary(motor.aov)) # Get SS values from Task ANOVA without covariate
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Task         2  241.3   120.7   0.824  0.458
## Residuals   15 2196.7   146.4
```

```r
eta.sq<-241.3/2196.7
print(round(eta.sq,digits=2))
```

```
## [1] 0.11
```

Homogeneity of Slopes Assumption

In order for the results of an ANCOVA to be valid, it is assumed that the slopes relating scores on the covariate, X, are the same for all groups in the experiment. This assumption is necessary because in an ANCOVA model, the slopes for each group are restricted to be equal.

An easy way to test whether the homogeneity of slopes assumption holds is to examine the interaction between the covariate, X, and the group factor of interest (in this case, *Task*). A non-significant interaction would suggest it is reasonable to assume that the slopes are indeed homogenous across different groups. Here is how to perform this test using *R*:

```r
interaction.aov<-aov(Y~X*Task,data=age.motor.data)
print(summary(interaction.aov))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## X            1 2050.4  2050.4 165.552 2.22e-08 ***
## Task         2  176.8    88.4   7.138  0.00907 **
## X:Task       2   62.2    31.1   2.510  0.12283
## Residuals   12  148.6    12.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The above analysis shows that there is (not surprisingly) a significant effect of Age (X), a significant effect of Task after controlling for Age, but no significant Age X Task interaction. Therefore, we are safe to assume homogeneity of slopes, and an ANCOVA would be considered appropriate for these data.