Chapter 11 – Factorial ANOVA

**Repeated measures two-way ANOVA**

Students’ levels of anxiety were measured based on two sources, examinations and bungee jumping. Both factors were measured pre- and post-activity. Each measurement used a scale of 1 to 10, 1 being the lowest and 10 being the highest. One of the factors is the source of anxiety; the other is time before and after the activity.

Perform a repeated measures two-way ANOVA to analyze the individual and interaction effects of activity and time on the students’ anxiety levels (on the Repeated-measures two-way ANOVA.csv file).

|  |  |  |  |
| --- | --- | --- | --- |
| PreExam\_Anxiety | PostExam\_Anxiety | PreBungee\_Anxiety | PostBungee\_Anxiety |
| 6 | 5 | 9 | 7 |
| 9 | 6 | 6 | 4 |
| 5 | 3 | 8 | 5 |
| 6 | 2 | 5 | 5 |
| 6 | 5 | 9 | 6 |
| 3 | 3 | 7 | 5 |
| 9 | 6 | 7 | 5 |
| 4 | 2 | 4 | 3 |
| 8 | 5 | 6 | 5 |
| 7 | 2 | 8 | 4 |

As the repeated measures ANOVA is quite complicated in R, all code will be presented.

First comes the creation of the data frame.

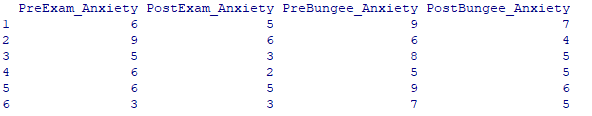
file = read.csv("Repeated-Measures two-way ANOVA.csv")

phases = with(file, data.frame(PreExam\_Anxiety, PostExam\_Anxiety,

PreBungee\_Anxiety, PostBungee\_Anxiety))

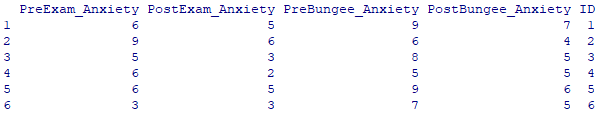
phases = na.omit(phases)

head(phases) # check that the data frame is correct



phases$ID = seq.int(nrow(phases)) # adds numbers in variable ID

head(phases)



library(reshape2)

data = melt(phases, id="ID", measured =

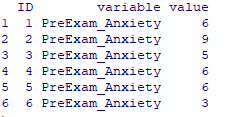
c("PreExam\_Anxiety "," PostExam\_Anxiety ", " PreBungee\_Anxiety ", " PostBungee\_Anxiety "

))

head(data)

# In some versions of the book, head(data) is on the same line as a previous expression

# head(data) needs to be on a fresh line



We now have a single observation for each value, but 'variable' needs to be split. Now we check to see that the important parts of the names are correctly placed prior to splitting into two factors:

summary(data$variable)



We want to have the split looking correct, so we want to equalize the phrases. Our problem in this case is that Pre only covers the first three spaces, while Post covers four. So we'll had an underline to Pre. We may as well add a couple of underlines to Exam so that it matches with Bungee.

Use underlines – EZANOVA doesn't like hyphens ( - ). It also doesn't like 'curly' punctuation. Use only straight apostrophes and speech marks ( ' " ).

library(car)

data$variable = Recode(data$variable,

" 'PreExam\_Anxiety' = 'Pre\_Exam\_\_\_Anxiety';

'PreBungee\_Anxiety' = 'Pre\_BungeeAnxiety' ")

summary(data$variable)



The next paragraph creates new variables data$activity and data$time.

data$activity = substr(data$variable, 5,8)

# Looking for Bungee and Exam\_\_

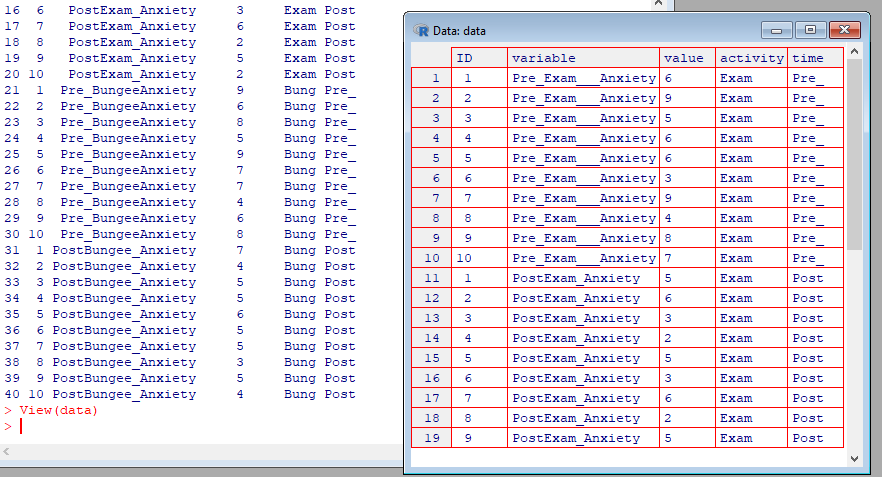
data$time = substr(data$variable, 1, 4)

# Looking for Pre\_ and Post

To check these are correct:

data

View(data)



And finally, we get to the business!

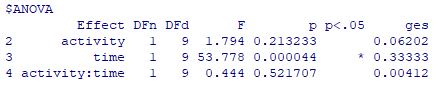
library(ez)

model = ezANOVA(data, dv=value, ID,

within = .(activity, time), between= NULL)

options(scipen=999, digits=3)

model



The results show that only time has a significant effect; its *p* value is smaller than 0.05. In addition, it has a large effect size (Generalized Eta-Squared). There is no reason to reject the null hypotheses for the interaction and activity effects.

Note that no Mauchly's W statistic appears. Sphericity is only an issue when there are more than two levels (conditions) in a factor. In this case, both factors contain only two conditions.

Adapt the instructions in the book if you wish to find the partial eta squared statistic for effect size. This is often preferred for factorial repeated measures ANOVA.

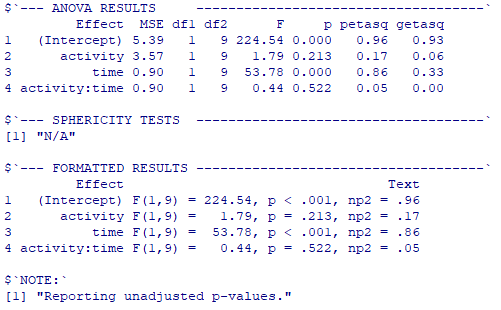
model = ezANOVA(data, dv=value, ID,

within = .(activity, time), between= NULL,

detailed=TRUE, return\_aov=TRUE) # changed

library(schoRsch)

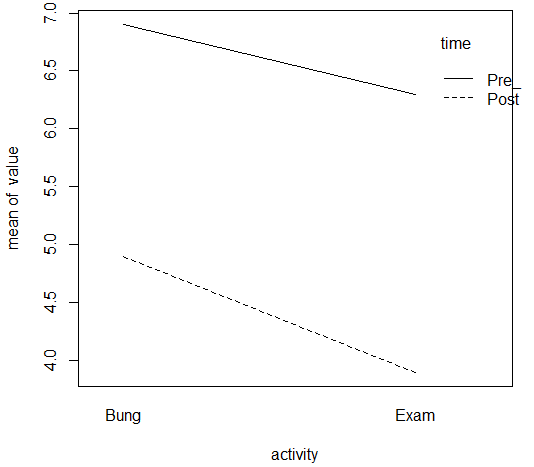
effects = anova\_out(model)



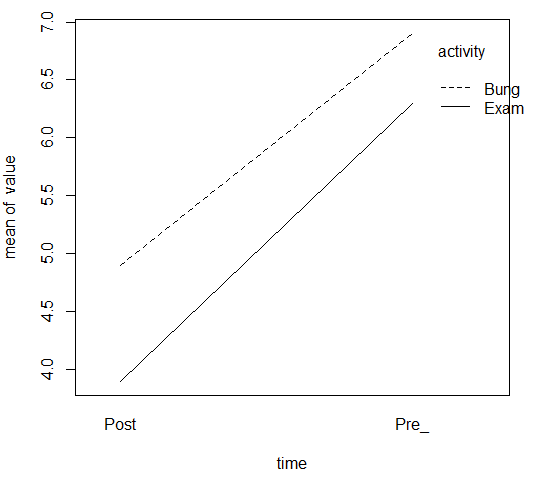
This indicates a very large effect size for the time variable.

For visualisation, you may take a simpler approach to that in the book:

with(data, interaction.plot(activity, time, value))



with(data, interaction.plot(time, activity, value))



**Between-Subjects ANOVA**

A fiber company is analyzing the breaking strength of their product on several production batches. Four production machines are chosen, and the operator was also noted. The results are as follows (on the Between Subjects ANOVA.csv file):

|  |  |  |
| --- | --- | --- |
| Breaking strength | Machine | Operator |
| 109 | 1 | 1 |
| 110 | 1 | 1 |
| 110 | 1 | 2 |
| 112 | 1 | 2 |
| 116 | 1 | 3 |
| 114 | 1 | 3 |
| 110 | 2 | 1 |
| 115 | 2 | 1 |
| 110 | 2 | 2 |
| 111 | 2 | 2 |
| 112 | 2 | 3 |
| 115 | 2 | 3 |
| 108 | 3 | 1 |
| 109 | 3 | 1 |
| 111 | 3 | 2 |
| 109 | 3 | 2 |
| 114 | 3 | 3 |
| 119 | 3 | 3 |
| 110 | 4 | 1 |
| 108 | 4 | 1 |
| 114 | 4 | 2 |
| 112 | 4 | 2 |
| 120 | 4 | 3 |
| 117 | 4 | 3 |

Test to see if there is a difference in breaking strength according to the factors considered.

file = read.csv("Between Subjects ANOVA.csv")

fiberStudy = with(file, data.frame(Breaking.strength, Machine, Operator))

fiberStudy = na.omit(fiberStudy)

data = fiberStudy # for a variable that is easier to type for reuse

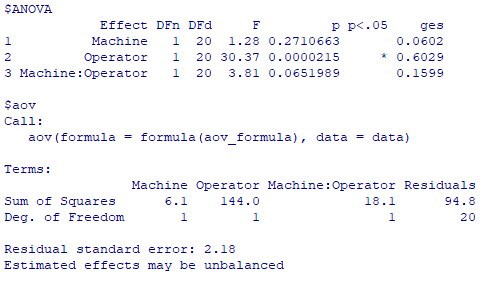
data$ID = seq.int(nrow(data)) # creates ID variable, adding cases

library(ez)

model = ezANOVA(data, Breaking.strength, wid = ID, within = NULL,

between = .( Machine, Operator), return\_aov=TRUE )

model



There is reason to consider the operator to be the crucial factor here.

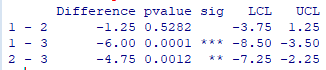
Here is just one of the possible comparison tests:

library(agricolae)

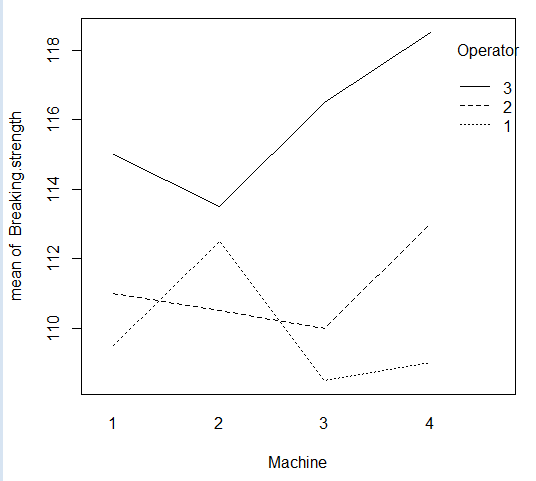
scheffe = scheffe.test(model$aov, "Operator",

group=FALSE)

scheffe$comparison



For a visualisation, just change the variables in the coding provided in the book.



The ANOVA results show that the average breaking strength differs based upon the operator (the *p* value is lower than 0.05). Moreover, it has a very large effect size at 0.60. Since there are three operators, a post hoc test may be useful: operators 1 and 3 as well as 2 and 3 have significantly different effects on breaking strength. As shown in the plot, operator 3 has the highest average breaking strength on all machines..

**Mixed ANOVA**

The quality control department of a fabric dye company is analyzing the difference in quality of dyes produced by three machines. In addition, the department wants to study if the dyeing temperature affects the quality; different temperatures can be tried on the same machine. Below are the results (also on the Mixed ANOVA.csv file):

|  |  |  |  |
| --- | --- | --- | --- |
| Cloth | Machine | 300\_Temperature | 350\_Temperature |
| 1 | 1 | 23 | 24 |
| 2 | 1 | 24 | 23 |
| 3 | 1 | 25 | 28 |
| 4 | 2 | 30 | 38 |
| 5 | 2 | 28 | 36 |
| 6 | 2 | 26 | 35 |
| 7 | 3 | 31 | 34 |
| 8 | 3 | 32 | 36 |
| 9 | 3 | 29 | 39 |

Conduct a mixed ANOVA to identify which factors are associated with dye score.

file = read.csv("Mixed ANOVA.csv")

dyeStudy = with(file, data.frame(Cloth, Machine,

X300\_Temperature, X350\_Temperature))

dyeStudy = na.omit(dyeStudy)

dyeStudy

library(reshape2)

data = melt(dyeStudy,

id.vars= c("Cloth", "Machine"), # The rationale is explained in the book

measure.vars= c("X300\_Temperature", "X350\_Temperature"),

variable.name="Temperature", value.name="Score")

data[1:14, ]

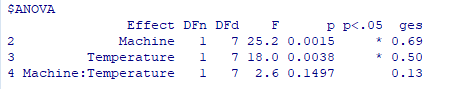
options(digits=2)

library(ez)

model = ezANOVA(data, Score, wid=Cloth,

within = Temperature, between = Machine)

model



The within subject and between subject effects are both significant. In the case of the interaction between machine and temperature, the *p* value is not small enough for us to say that there is evidence to reject the null hypothesis.

The Scheffé is one of the tests used in the book.

modelBETWEEN = ezANOVA(data, Score, wid=Cloth,

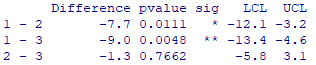
between = Machine, return\_aov=TRUE)

library(agricolae)

scheffe = scheffe.test(modelBETWEEN$aov,

"Machine", group=FALSE)

scheffe$comparison



Simple visualisations of the machines run at different temperatures, will show that Machine 1 has a different performance from the others:

with(file, plot(Machine, X300\_Temperature))

with(file, plot(Machine, X350\_Temperature))

