**Homework Week 2 (due Nov 23, 2015)**

As last week, you'll end up with several files from your homework:

* An R script file
* A word doc with answers to the questions and some figures
* 2 csv files

***Paste the R script content into the word file (at the end); print that whole thing out, don't forget to put your name on top of the document, and bring it to the next class.***

**(A) Finishing the rest from what you started last week with the data file REGRESS.SAV**

**STEP 3: Investigate Bivariate Associations**

**OLD SPSS instructions**

Once univariate distributions have been inspected; associations between all pairs of variables should be examined. If variables are dichotomous or interval-level (continuous), Pearson Product-Moment (interclass) correlations are used to investigate linear associations. Calculate the correlations using SPSS (in the *Analyze* menu) and answer the following question.

**7. How are the four variables related to one another? ~~Provide a thorough, yet concise description of the Pearson correlations IN APA STYLE. Be sure to describe the direction and magnitude of all associations.~~**

timedrs phyheal menheal stress

timedrs \*\*\*\*\* 0.440 0.256 0.287

phyheal <0.001 \*\*\*\*\* 0.505 0.306

menheal <0.001 <0.001 \*\*\*\*\* 0.370

stress <0.001 <0.001 <0.001 \*\*\*\*\*

**SCRIPT**

**library(ltm)**

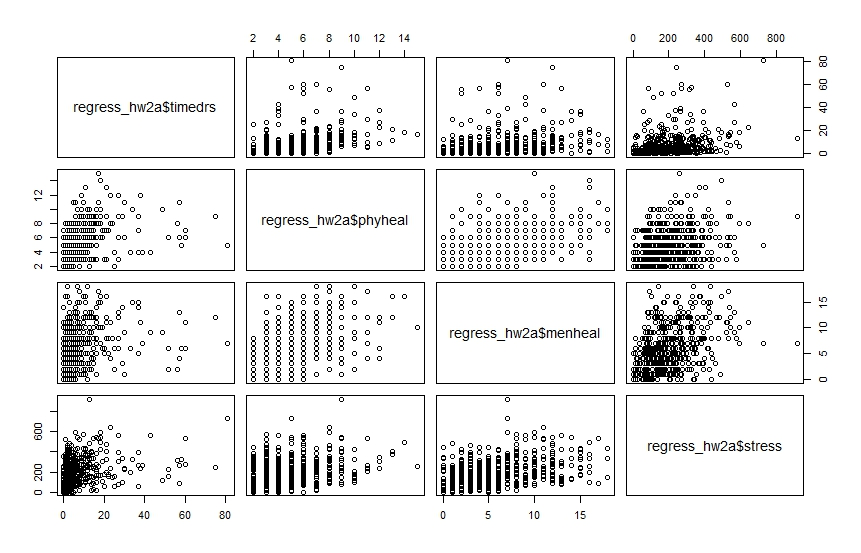
**rcor.test(regress\_hw2a[,2:5])**

**pairs(~ regress\_hw2a$timedrs + regress\_hw2a$phyheal + regress\_hw2a$menheal + regress\_hw2a$stress)**

**For R/Stats II:**

You don't have to do all of this (that's why I crossed out some and deleted some stuff), but DO compute the pairwise correlations, using Pearson correlations.

Also, plot all pairwise correlations, using either plot() or pairs() or some fancier way (see, e.g., here: http://www.statmethods.net/graphs/scatterplot.html)



**(B) New Stuff!**

***[NOTE: Do NOT simply copy/paste my code! Do it from your own notes, getting help from the slides and the book etc, but without looking at my exact code.]***

1. Create a data frame in "wide" format which consists at least of

* 10 rows of data (or more) [if you have data from your thesis etc, feel free to use those data!]
* 5 columns, of which
  + 1 is the participant code
  + 1 is gender or some other categorical variable with two levels
  + 3 (or more) are some kind of repeated-measures variables (continuous, i.e., numeric)

**SCRIPT**

**gender<-c(1,0,1,1,0,0,0,1,1,1)**

**score1<-c(60,65,80,72,74,69,62,63,64,71)**

**score2<-c(time1+1\*code/2)**

**score3<-c(time2\*\*2)**

**gender<-factor(gender, levels = c(0,1), labels = c("male", "female"))**

**wide\_1<-data.frame(id = c('1':'10'), gender, score1, score2, score3)**

**rcor.test(wide\_1[,2:5])**

2. Compute all possible pairwise correlations

gender score1 score2 score3

gender \*\*\*\*\* 0.068 0.102 0.116

score1 0.851 \*\*\*\*\* 0.971 0.977

score2 0.780 <0.001 \*\*\*\*\* 0.999

score3 0.751 <0.001 <0.001 \*\*\*\*\*

upper diagonal part contains correlation coefficient estimates

lower diagonal part contains corresponding p-values

3. Compute: (a) all possible paired t tests; (b) at least one t test using your categorical variable as between-subjects factor.

> ttest\_1

Welch Two Sample t-test

data: wide\_1$score1 by wide\_1$gender

t = -0.2092, df = 7.914, p-value = 0.8395

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-10.034699 8.368032

sample estimates:

mean in group male mean in group female

67.50000 68.33333

> ttest\_2

Welch Two Sample t-test

data: wide\_1$score2 by wide\_1$gender

t = -0.3116, df = 7.908, p-value = 0.7634

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-10.518063 8.018063

sample estimates:

mean in group male mean in group female

70.00 71.25

> ttest\_3

Welch Two Sample t-test

data: wide\_1$score3 by wide\_1$gender

t = -0.3549, df = 7.915, p-value = 0.732

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-1517.065 1113.065

sample estimates:

mean in group male mean in group female

4920.625 5122.625

> ttest\_4

Paired t-test

data: wide\_1$score1 and wide\_1$score2

t = -5.7446, df = 9, p-value = 0.0002782

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-3.832925 -1.667075

sample estimates:

mean of the differences

-2.75

> ttest\_5

Paired t-test

data: wide\_1$score1 and wide\_1$score3

t = -17.5387, df = 9, p-value = 0.00000002882

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-5615.353 -4332.297

sample estimates:

mean of the differences

-4973.825

> ttest\_6

Paired t-test

data: wide\_1$score3 and wide\_1$score2

t = 17.5329, df = 9, p-value = 0.0000000289

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

4329.688 5612.462

sample estimates:

mean of the differences

4971.075

**SCRIPT**

**ttest\_1<-t.test(wide\_1$score1 ~ wide\_1$gender, paired = FALSE)**

**ttest\_2<-t.test(wide\_1$score2 ~ wide\_1$gender, paired = FALSE)**

**ttest\_3<-t.test(wide\_1$score3 ~ wide\_1$gender, paired = FALSE)**

**ttest\_4<-t.test(wide\_1$score1, wide\_1$score2, paired = TRUE)**

**ttest\_5<-t.test(wide\_1$score1, wide\_1$score3, paired = TRUE)**

**ttest\_6<-t.test(wide\_1$score3, wide\_1$score2, paired = TRUE)**

4. Reshape the data frame to long format, do it once using melt() and do it a second time using reshape()

long\_1 <- melt(wide\_2, id = c('gender', 'id2'), measured=c('score1','score2','score3'))

str(wide\_2)

long\_2 <- reshape(wide\_2, idvar = 'id2', varying = c('score1', 'score2', 'score3'), timevar = 'scorenumbers', v.names = 'score', direction = 'long')

5. Do at least one of the paired t test again, this time using the long format data frame

Paired t-test

data: long\_2$scorenumbers and long\_2$score

t = -3.8784, df = 29, p-value = 0.000556

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-2634.437 -815.280

sample estimates:

mean of the differences

-1724.858

> ttest\_7\_long\_2

Welch Two Sample t-test

data: long\_2$score by long\_2$gender

t = -0.07422, df = 24.382, p-value = 0.9414

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-1958.164 1822.108

sample estimates:

mean in group male mean in group female

1686.042 1754.069

> ttest\_3long1

Welch Two Sample t-test

data: long\_1$value by long\_1$gender

t = -0.07422, df = 24.382, p-value = 0.9414

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-1958.164 1822.108

sample estimates:

mean in group male mean in group female

1686.042 1754.069

ttest\_4\_long\_2<-t.test(long\_2$scorenumbers, long\_2$score, paired = TRUE)

ttest\_7\_long\_2<-t.test(long\_2$score ~ long\_2$gender, paired = FALSE)

ttest\_3long1<-t.test(long\_1$value ~ long\_1$gender, paired = FALSE)

6. Reshape the long format data frame back to wide format using cast()

cast1 <- cast(long\_1, id2 + gender ~ variable)

7. Save both the wide and the long format data frame as csv file

regress\_hw2<-read.csv("C:\\Users\\André\\Google Drive\\Master\\period 2\\R\\hw2\\hw2.csv", sep = ',')

regress\_hw2a<-regress\_hw2 [1:465,]

install.packages("psych")

install.packages("ltm")

install.packages("msm")

#to run correlations an plot them

library(ltm)

rcor.test(regress\_hw2a[,2:5])

pairs(~ regress\_hw2a$timedrs + regress\_hw2a$phyheal + regress\_hw2a$menheal + regress\_hw2a$stress)

#creating database

pcode<-c(1,2,3,4,5,6,7,8,9,10)

names(pcode)<-c('p1', 'p2', 'p3', 'p4', 'p5', 'p6', 'p7', 'p8', 'p9', 'p10')

gender<-c(1,0,1,1,0,0,0,1,1,1)

score1<-c(60,65,80,72,74,69,62,63,64,71)

score2<-c(score1+1\*pcode/2)

score3<-c(score2\*\*2)

id <- c('1':'10')

gender<-factor(gender, levels = c(0,1), labels = c("male", "female"))

id2<-factor(id, levels = c(1:10), labels = c('p1', 'p2', 'p3', 'p4', 'p5', 'p6', 'p7', 'p8', 'p9', 'p10'))

wide\_1<-data.frame(id, gender, score1, score2, score3)

wide\_2<-data.frame(id2, gender, score1, score2, score3)

rcor.test(wide\_1[,2:5])

#t test

#como es dummy, no es la misma distribucion, por eso false, aparte el~ dice que score 1 es una funcion de gender.

#para el otro las dos son continuas so, true... solo checas como score 1 es diferente de socre 2.

options (scipen = 10) #option to show the scienytific notation.

ttest\_1<-t.test(wide\_1$score1 ~ wide\_1$gender, paired = FALSE)

ttest\_2<-t.test(wide\_1$score2 ~ wide\_1$gender, paired = FALSE)

ttest\_3<-t.test(wide\_1$score3 ~ wide\_1$gender, paired = FALSE)

ttest\_4<-t.test(wide\_1$score1, wide\_1$score2, paired = TRUE)

ttest\_5<-t.test(wide\_1$score1, wide\_1$score3, paired = TRUE)

ttest\_6<-t.test(wide\_1$score3, wide\_1$score2, paired = TRUE)

install.packages("reshape")

library(reshape)

#para transformar de wide a long

long\_1 <- melt(wide\_2, id = c('gender', 'id2'), measured=c('score1','score2','score3'))

str(wide\_2)

long\_2 <- reshape(wide\_2, idvar = 'id2', varying = c('score1', 'score2', 'score3'), timevar = 'scorenumbers', v.names = 'score', direction = 'long')

write.csv(regress\_hw2, file = "C:\\Users\\André\\Google Drive\\Master\\period 2\\R\\hw2\\hw2long.csv")

ttest\_4\_long\_2<-t.test(long\_2$scorenumbers, long\_2$score, paired = TRUE)

ttest\_7\_long\_2<-t.test(long\_2$score ~ long\_2$gender, paired = FALSE)

ttest\_3long1<-t.test(long\_1$value ~ long\_1$gender, paired = FALSE)

cast1 <- cast(long\_1, id2 + gender ~ variable)

write.csv(regress\_hw2, file = "C:\\Users\\André\\Google Drive\\Master\\period 2\\R\\hw2\\hw2wide.csv")

**(C) Some reading**

FMF book

* finish (and perhaps re-read) Chapter 3: do all the exercises etc that are described in there
* read chapter 4 in preparation for next week's class
* read chapter 5: part of it is in preparation for the class on linear regression, part of that chapter covers things we covered this and last week

***Don't look at the next page! It has some hints. (So look only at it if you really get stuck.)***

**Some hints**

**re (A)**

The ltm package has the function rcor.test()

You can use it, e.g., like this:

rcor.test(regress\_2[,2:5])

The psych package also has a correlation function, called corr.test()

corr.test(regress\_2[,2:5])

You can also have a look at the example R script from week 2 that I posted on Blackboard.

**re (B)**

Hmm, no hints here, really. Except, again: Try to first do it only with the help of the built-in help function in R, the slides, and the book. Only if you get really stuck, have a look at my example R script from week 2.