HW 7

Gaea Daniel

12/4/2017

### 1. Perform a repeated measures analysis of variance (RM-ANOVA) for the 5 CESD measurements across time by treatment group.

#### a. treat time as a continuous variable (not as a factor) - this is your WITHIN group effect

#### b. treat the treatment group treat as a factor - this is your BETWEEN group effect

#### c. TABLE: present the table of the intercept, time, treat and time\*treat interaction effects including the tests of significance. [Remember this significance might change depending on the treatment group coding - try flipping the 0 and 1 and run the model again to see if the significance changes] ####d. FIGURE: make a plot of the CESD means across time by group - if you can make it an error bar plot which has the means and CI’s (confidence intervals) or SE’s (standard errors)

library(tidyverse)

## Loading tidyverse: ggplot2  
## Loading tidyverse: tibble  
## Loading tidyverse: tidyr  
## Loading tidyverse: readr  
## Loading tidyverse: purrr  
## Loading tidyverse: dplyr

## Conflicts with tidy packages ----------------------------------------------

## filter(): dplyr, stats  
## lag(): dplyr, stats

library(haven)  
  
helpdat <- haven::read\_spss("helpmkh.sav")  
  
h1 <- helpdat %>%  
select(treat, cesd, cesd1, cesd2, cesd3, cesd4)  
  
#Correlation matrix between the 5 cesd measurements over time  
library(psych)

##   
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

psych::corr.test(h1[,2:6], method="pearson")

## Call:psych::corr.test(x = h1[, 2:6], method = "pearson")  
## Correlation matrix   
## cesd cesd1 cesd2 cesd3 cesd4  
## cesd 1.00 0.39 0.42 0.40 0.33  
## cesd1 0.39 1.00 0.55 0.64 0.49  
## cesd2 0.42 0.55 1.00 0.74 0.60  
## cesd3 0.40 0.64 0.74 1.00 0.74  
## cesd4 0.33 0.49 0.60 0.74 1.00  
## Sample Size   
## cesd cesd1 cesd2 cesd3 cesd4  
## cesd 453 246 209 248 266  
## cesd1 246 246 150 157 165  
## cesd2 209 150 209 155 154  
## cesd3 248 157 155 248 197  
## cesd4 266 165 154 197 266  
## Probability values (Entries above the diagonal are adjusted for multiple tests.)   
## cesd cesd1 cesd2 cesd3 cesd4  
## cesd 0 0 0 0 0  
## cesd1 0 0 0 0 0  
## cesd2 0 0 0 0 0  
## cesd3 0 0 0 0 0  
## cesd4 0 0 0 0 0  
##   
## To see confidence intervals of the correlations, print with the short=FALSE option

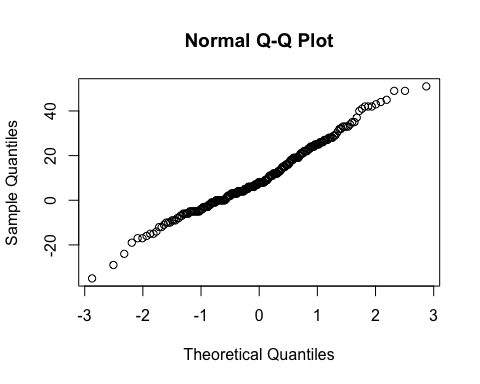
#PAIRED t-test of first 2 timepoints to see if the scores are significantly changing across time WITHIN individuals  
t.test(h1$cesd, h1$cesd1, paired=TRUE)

##   
## Paired t-test  
##   
## data: h1$cesd and h1$cesd1  
## t = 10.512, df = 245, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 8.020506 11.719331  
## sample estimates:  
## mean of the differences   
## 9.869919

#Compute the change scores and compare the difference scores to 0  
h1 <- h1 %>%  
 mutate(diff\_cesd\_bl\_1=cesd - cesd1)  
t.test(h1$diff\_cesd\_bl\_1, mu=0)

##   
## One Sample t-test  
##   
## data: h1$diff\_cesd\_bl\_1  
## t = 10.512, df = 245, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 8.020506 11.719331  
## sample estimates:  
## mean of x   
## 9.869919

qqnorm(h1$diff\_cesd\_bl\_1)



#Repeated measures analysis of variance (RM-ANOVA) for the 5 CESD measurements across time by treatment group  
# add rowid to h1  
h1 <- h1 %>%  
 mutate(rowid=as.numeric(rownames(h1)))  
   
h1long <- h1 %>%  
 gather(key=item,  
 value=value,  
 -c(treat,diff\_cesd\_bl\_1,rowid))

## Warning: attributes are not identical across measure variables;  
## they will be dropped

#Add a time variable to long format  
h1long <- h1long %>%  
 mutate(time=c(rep(0,453),  
 rep(1,453),  
 rep(2,453),  
 rep(3,453),  
 rep(4,453)))  
  
h1long\_bl1 <- h1long %>%  
 filter(time<2) %>%  
 select(rowid,value,time,treat)  
  
library(car)

##   
## Attaching package: 'car'

## The following object is masked from 'package:psych':  
##   
## logit

## The following object is masked from 'package:dplyr':  
##   
## recode

## The following object is masked from 'package:purrr':  
##   
## some

rm1 <- aov(value~factor(time)+Error(factor(rowid)),   
 data = h1long\_bl1)  
summary(rm1)

##   
## Error: factor(rowid)  
## Df Sum Sq Mean Sq F value Pr(>F)   
## factor(time) 1 4422 4422 21.17 5.46e-06 \*\*\*  
## Residuals 451 94191 209   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Error: Within  
## Df Sum Sq Mean Sq F value Pr(>F)   
## factor(time) 1 11982 11982 110.5 <2e-16 \*\*\*  
## Residuals 245 26567 108   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Compare the 2 changes from BL to 6m for cesd and cesd1 between the 2 treat groups  
bartlett.test(h1$diff\_cesd\_bl\_1~h1$treat)

##   
## Bartlett test of homogeneity of variances  
##   
## data: h1$diff\_cesd\_bl\_1 by h1$treat  
## Bartlett's K-squared = 0.19749, df = 1, p-value = 0.6568

t.test(h1$diff\_cesd\_bl\_1~h1$treat, var.equal=TRUE)

##   
## Two Sample t-test  
##   
## data: h1$diff\_cesd\_bl\_1 by h1$treat  
## t = 0.35478, df = 244, p-value = 0.7231  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -3.044544 4.382214  
## sample estimates:  
## mean in group 0 mean in group 1   
## 10.226087 9.557252

#RM-ANOVA for the changes from BL to 6m BETWEEN the 2 treat groups; a comparison of the time\*treat effect to the t-test above for the difference scores  
rm2 <- aov(value~(factor(time)\*factor(treat))+  
 Error(factor(rowid)/(treat)),   
 data = h1long\_bl1)

## Warning in aov(value ~ (factor(time) \* factor(treat)) +  
## Error(factor(rowid)/(treat)), : Error() model is singular

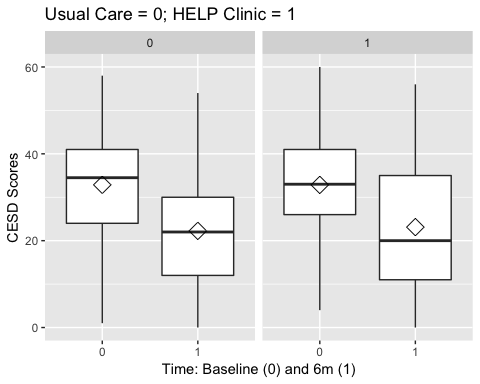
summary(rm2)

##   
## Error: factor(rowid)  
## Df Sum Sq Mean Sq F value Pr(>F)   
## factor(time) 1 4422 4422 21.087 5.71e-06 \*\*\*  
## factor(treat) 1 18 18 0.088 0.767   
## factor(time):factor(treat) 1 18 18 0.087 0.768   
## Residuals 449 94154 210   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Error: Within  
## Df Sum Sq Mean Sq F value Pr(>F)   
## factor(time) 1 11982 11982 110.104 <2e-16 \*\*\*  
## factor(time):factor(treat) 1 14 14 0.126 0.723   
## Residuals 244 26553 109   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Plot of cesd and cesd1 scores by group to get an idea of trend across time (plot is cross sectional, not paired)  
ggplot(h1long\_bl1, aes(x=factor(time), y=value)) +   
 geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=5, size=4) +   
 xlab("Time: Baseline (0) and 6m (1)") +  
 ylab("CESD Scores") +  
 facet\_wrap(~treat) +  
 ggtitle("Usual Care = 0; HELP Clinic = 1")

## Warning: Removed 207 rows containing non-finite values (stat\_boxplot).

## Warning: Removed 207 rows containing non-finite values (stat\_summary).

 ###2. Repeate the “repeated measures/longitudinal” analysis using a random intercepts MLM model ####a. REMEMBER to restructure the data from WIDE to LONG format ####b. TABLE: present the table of the intercept, time, treat and time\*treat interaction effects including the tests of significance. [Remember this significance might change depending on the treatment group coding - try flipping the 0 and 1 and run the model again to see if the significance changes]

library(tidyverse)  
library(haven)  
  
helpdat <- haven::read\_spss("helpmkh.sav")  
h1 <- helpdat %>%  
 select(id, treat, cesd, cesd1, cesd2, cesd3, cesd4)  
  
# restructure into long format  
  
h1long <- h1 %>%  
 gather(key=item,  
 value=value,  
 -c(id,treat))

## Warning: attributes are not identical across measure variables;  
## they will be dropped

names(h1long) <- c("id","treat","cesditem","cesdvalue")  
  
# add a time variable to long format  
h1long <- h1long %>%  
 mutate(time=c(rep(0,453),  
 rep(1,453),  
 rep(2,453),  
 rep(3,453),  
 rep(4,453)))  
  
# from the cookbook for R  
## Gives count, mean, standard deviation, standard error of the mean, and confidence interval (default 95%).  
## data: a data frame.  
## measurevar: the name of a column that contains the variable to be summariezed  
## groupvars: a vector containing names of columns that contain grouping variables  
## na.rm: a boolean that indicates whether to ignore NA's  
## conf.interval: the percent range of the confidence interval (default is 95%)  
summarySE <- function(data=NULL, measurevar, groupvars=NULL, na.rm=FALSE,  
 conf.interval=.95, .drop=TRUE) {  
 library(plyr)  
   
 # New version of length which can handle NA's: if na.rm==T, don't count them  
 length2 <- function (x, na.rm=FALSE) {  
 if (na.rm) sum(!is.na(x))  
 else length(x)  
 }  
   
 # This does the summary. For each group's data frame, return a vector with  
 # N, mean, and sd  
 datac <- ddply(data, groupvars, .drop=.drop,  
 .fun = function(xx, col) {  
 c(N = length2(xx[[col]], na.rm=na.rm),  
 mean = mean (xx[[col]], na.rm=na.rm),  
 sd = sd (xx[[col]], na.rm=na.rm)  
 )  
 },  
 measurevar  
 )  
   
 # Rename the "mean" column   
 datac <- rename(datac, c("mean" = measurevar))  
   
 datac$se <- datac$sd / sqrt(datac$N) # Calculate standard error of the mean  
   
 # Confidence interval multiplier for standard error  
 # Calculate t-statistic for confidence interval:   
 # e.g., if conf.interval is .95, use .975 (above/below), and use df=N-1  
 ciMult <- qt(conf.interval/2 + .5, datac$N-1)  
 datac$ci <- datac$se \* ciMult  
   
 return(datac)  
}  
  
h1long\_nomiss <- na.omit(h1long)  
table(h1long\_nomiss$time)

##   
## 0 1 2 3 4   
## 453 246 209 248 266

h1se <- summarySE(h1long\_nomiss,   
 measurevar="cesdvalue",   
 groupvars=c("time","treat"))

## -------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.  
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:  
## library(plyr); library(dplyr)

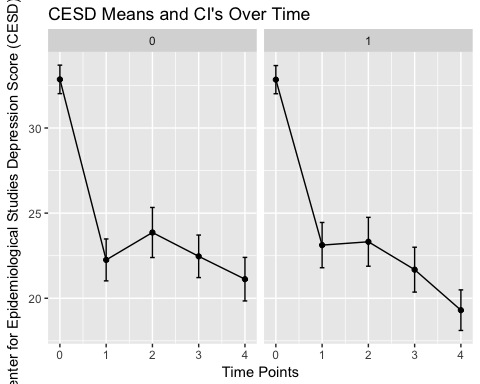
## -------------------------------------------------------------------------

##   
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':  
##   
## arrange, count, desc, failwith, id, mutate, rename, summarise,  
## summarize

## The following object is masked from 'package:purrr':  
##   
## compact

ggplot(h1se, aes(x=time, y=cesdvalue)) +   
 geom\_errorbar(aes(ymin=cesdvalue-se, ymax=cesdvalue+se), width=.1) +  
 geom\_line() +  
 geom\_point() +  
 xlab("Time Points") +  
 ylab("Center for Epidemiological Studies Depression Score (CESD)") +  
 ggtitle("CESD Means and CI's Over Time") +  
 facet\_wrap(~treat)



# use nlme package  
library(nlme)

##   
## Attaching package: 'nlme'

## The following object is masked from 'package:dplyr':  
##   
## collapse

lme1 <- lme(cesdvalue ~ time\*treat,  
 data=h1long,  
 random= ~1 | id,  
 method="REML",  
 na.action=na.omit)  
# get summary - model coefficients  
# tests coefficients not equal to 0  
summary(lme1)

## Linear mixed-effects model fit by REML  
## Data: h1long   
## AIC BIC logLik  
## 11208.56 11240.1 -5598.28  
##   
## Random effects:  
## Formula: ~1 | id  
## (Intercept) Residual  
## StdDev: 9.655901 10.11562  
##   
## Fixed effects: cesdvalue ~ time \* treat   
## Value Std.Error DF t-value p-value  
## (Intercept) 30.778989 0.8699977 967 35.37824 0.0000  
## time -2.626044 0.2683068 967 -9.78747 0.0000  
## treat 0.165015 1.2295777 451 0.13420 0.8933  
## time:treat -0.412528 0.3719693 967 -1.10904 0.2677  
## Correlation:   
## (Intr) time treat   
## time -0.473   
## treat -0.708 0.334   
## time:treat 0.341 -0.721 -0.478  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -2.41023282 -0.62607170 -0.04357131 0.58091139 3.26478926   
##   
## Number of Observations: 1422  
## Number of Groups: 453

# get anova tables - both  
# of these yield type III Sums of Squares  
anova.lme(lme1, type="marginal")

## numDF denDF F-value p-value  
## (Intercept) 1 967 1251.6201 <.0001  
## time 1 967 95.7946 <.0001  
## treat 1 451 0.0180 0.8933  
## time:treat 1 967 1.2300 0.2677

car::Anova(lme1, type="III")

## Analysis of Deviance Table (Type III tests)  
##   
## Response: cesdvalue  
## Chisq Df Pr(>Chisq)   
## (Intercept) 1251.620 1 <2e-16 \*\*\*  
## time 95.795 1 <2e-16 \*\*\*  
## treat 0.018 1 0.8932   
## time:treat 1.230 1 0.2674   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# FYI - sequential SS or Type II SS  
anova.lme(lme1, type="sequential")

## numDF denDF F-value p-value  
## (Intercept) 1 967 2384.5592 <.0001  
## time 1 967 233.9617 <.0001  
## treat 1 451 0.2026 0.6528  
## time:treat 1 967 1.2300 0.2677

car::Anova(lme1, type="II")

## Analysis of Deviance Table (Type II tests)  
##   
## Response: cesdvalue  
## Chisq Df Pr(>Chisq)   
## time 233.6719 1 <2e-16 \*\*\*  
## treat 0.2026 1 0.6526   
## time:treat 1.2300 1 0.2674   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# flip treat and run again  
h1long <- h1long %>%  
 mutate(treat\_flip = as.numeric(treat==0))  
  
lme2 <- lme(cesdvalue ~ time\*treat\_flip,  
 data=h1long,  
 random= ~1 | id,  
 method="REML",  
 na.action=na.omit)  
# get summary - model coefficients  
# tests coefficients not equal to 0  
summary(lme2)

## Linear mixed-effects model fit by REML  
## Data: h1long   
## AIC BIC logLik  
## 11208.56 11240.1 -5598.28  
##   
## Random effects:  
## Formula: ~1 | id  
## (Intercept) Residual  
## StdDev: 9.655901 10.11562  
##   
## Fixed effects: cesdvalue ~ time \* treat\_flip   
## Value Std.Error DF t-value p-value  
## (Intercept) 30.944004 0.8688874 967 35.61336 0.0000  
## time -3.038572 0.2576289 967 -11.79437 0.0000  
## treat\_flip -0.165015 1.2295777 451 -0.13420 0.8933  
## time:treat\_flip 0.412528 0.3719693 967 1.10904 0.2677  
## Correlation:   
## (Intr) time trt\_fl  
## time -0.483   
## treat\_flip -0.707 0.341   
## time:treat\_flip 0.334 -0.693 -0.478  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -2.41023282 -0.62607170 -0.04357131 0.58091139 3.26478926   
##   
## Number of Observations: 1422  
## Number of Groups: 453

# get anova tables - both  
# of these yield type III Sums of Squares  
anova.lme(lme2, type="marginal")

## numDF denDF F-value p-value  
## (Intercept) 1 967 1268.3116 <.0001  
## time 1 967 139.1073 <.0001  
## treat\_flip 1 451 0.0180 0.8933  
## time:treat\_flip 1 967 1.2300 0.2677

car::Anova(lme2, type="III")

## Analysis of Deviance Table (Type III tests)  
##   
## Response: cesdvalue  
## Chisq Df Pr(>Chisq)   
## (Intercept) 1268.312 1 <2e-16 \*\*\*  
## time 139.107 1 <2e-16 \*\*\*  
## treat\_flip 0.018 1 0.8932   
## time:treat\_flip 1.230 1 0.2674   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### 3. Compare the results between the 2 approaches

#### a. compare the sample size differences

The sample size for #1 (n=1422) is greater than the sample size for #2 (n=453).

#### b. why do you think the results are different or are similar?

The sample sizes are different because #1 is looking at each individual sample and #2 is looking at groups of samples.