Lab 2

Names of group members

# Lab 2 Instructions:

Working in a group of 2, 3, or 4 people, complete the following questions. Turn in a single *PDF* *from your word document* for the group *with all group member names on it* after knitting this document with your answers “in-line” (after the questions).

The initial header contains some packages we will use frequently this semester. In addition to packages from CRAN, we need two special packages available from my github repository. In order to access them, you need to download and load the remotes package. Then uncomment the lines that contain #remotes::install\_github("greenwood-stat/ggResidpanel") and #remotes::install\_github("greenwood-stat/catstats2"). Do this to download the packages and then re-comment those lines as you will not need to do that again unless you update your version of R or change computers. Do NOT use install.packages("ggResidpanel") as this will install a different version. Then you can uncomment the #library(ggResidpanel) line for this lab and future uses.

## White matter lesions (continued):

bpdata <- read\_csv("bipolardata.csv")

library(tidyverse)  
  
bpdata <- bpdata %>% dplyr::rename(IllnessDuration = 'Illness duration',  
 NormVolume = 'nWML\_sum',  
 SubDep = 'substance dependancy',  
 AlcDep = 'alcohol dependancy',  
 AnxDis = 'anxiety disorder',  
 smoking = 'smoking\_yes\_no') %>%   
 mutate(group = factor(group),  
 sex = factor(sex),  
 YMRS = factor(YMRS),  
 DM = factor(DM),  
 smoking = factor(smoking),  
 HYPERT = factor(HYPERT),  
 group = fct\_recode(group,   
 patient = "1",   
 HC = "2"),  
 sex = fct\_recode(sex,   
 male = "1",  
 female = "2"),  
 smoking = fct\_recode(smoking,  
 no = "0",  
 yes = "1"),  
 GroupSexCombs = factor(str\_c(group, sex)) #For later use  
 )  
  
favstats(NormVolume ~ group, data = bpdata)

## group min Q1 median Q3 max mean  
## 1 patient 0.06890985 0.1727848 0.2394119 0.3435528 3.6152561 0.3359443  
## 2 HC 0.05615140 0.1127530 0.1405374 0.2071285 0.7028555 0.1941617  
## sd n missing  
## 1 0.4092558 100 0  
## 2 0.1486240 54 0

inline\_test <- mean(NormVolume ~ group, data = bpdata)

1. Make an enhanced stripchart for the NormVolume variable by BD and HC groups. Report your code and output. Try to edit/improve the x and y-axis labels by adding and editing + labs(y = "Something better for y", x = "Something better for x"). No discussion.

e1 <- enhanced\_stripchart(NormVolume ~ group, data = bpdata)+  
 labs(y = 'Something for y',  
 x = 'Something for x')

1. Normalized Volume is “unitless” since it is the ratio of two volumes (both are assumed to be in mm). One would assume that the volume of white matter lesions (numerator of the ratio) is less than the *total* intracranial volume (denominator) that they divided by. What about the previous plot seems odd about the results given this information?

*Normalized should be between 0 and 1, so the range of patient is incorrect.*

1. Using the provided favstats and the plot, what is the estimated difference in the mean Normalized Volume between patients and health controls? Make sure you make it clear which one is larger/smaller in writing a sentence to report this (“*difference* of …” is not read by different people the same way). Note that Normalized Volume is “unitless” since it is the ratio of two volumes and one would assume that it should be between 0 and 1

\*The mean of the patient inline\_test[1] and the mean of the HC is inline\_test[2]. The difference in these sample means is inline\_test[1] - inline\_test[2].

1. Fit an lm with NormVolume and as the response and group as the predictor. Generate a base R model summary and generate a nice table of results using %>% tbl\_regression(intercept = T) %>% add\_global\_p(), and make an effects plot using the effects package (so something like plot(allEffects(modelname))). Which level of group was the baseline/reference level and how can you tell?

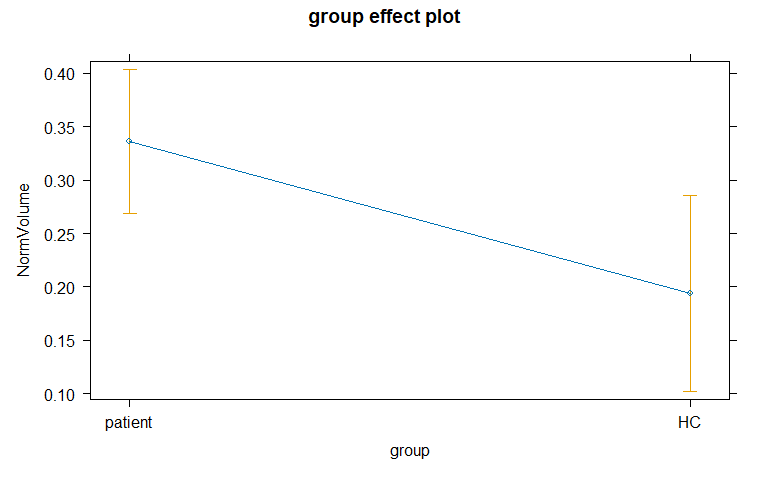
model1 <- lm(NormVolume ~ group, bpdata)  
summary(model1)

##   
## Call:  
## lm(formula = NormVolume ~ group, data = bpdata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.2670 -0.1249 -0.0730 0.0112 3.2793   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 0.33594 0.03417 9.830 <2e-16  
## groupHC -0.14178 0.05771 -2.457 0.0151  
##   
## Residual standard error: 0.3417 on 152 degrees of freedom  
## Multiple R-squared: 0.03819, Adjusted R-squared: 0.03186   
## F-statistic: 6.035 on 1 and 152 DF, p-value: 0.01515

model1 %>% tbl\_regression(intercept = T) %>% add\_global\_p()

| **Characteristic** | **Beta** | **95% CI***1* | **p-value** |
| --- | --- | --- | --- |
| (Intercept) | 0.34 | 0.27, 0.40 | <0.001 |
| group |  |  | 0.015 |
| patient | — | — |  |
| HC | -0.14 | -0.26, -0.03 |  |
| *1*CI = Confidence Interval | | | |

plot(allEffects(model1))



1. Write out the estimated model, defining the indicator variable you used.

* where is 1 for HC and 0 for patient.

1. Interpret the “size” of the slope coefficient for the group row in the model summary, finding and reporting a 95% confidence in parentheses as part of that single sentence. Hint: the confint function is an easy way to obtain a confidence interval or it is in the tbl\_regression output.

*For an HC group member the estimated mean of NormVolume changes by -0.14 (95% CI:-0.26, -0.03).*

“95% confident if we did this over and over that our mean would fall between …”

1. Report the “evidence” sentence for the group row in the model summary. Remember to include a conclusion where you declare what you think is going on here and/or what you would do based on the result: Do you think there is a difference or not/would you retain this term in the model?

There is strong evidence against the null hypothesis that there is no differnece in the true mean of NormVolume between patient and HC groups (t152 = -2.457)

1. It’s useful to record some information about the version of R you are using. When you Knit this documentation, it will report on the version of R that you are using. It should say 4.4.1 in your compiled word document:

* R version (short form): 4.4.1

1. Document any resources used outside of your fellow group members and course provided resources. If you do not use any, report “NONE” to get credit for this question.