The\_final\_code

Anders Hjulmand, Morten Gade, Sigrid Nielsen and Gustav Helms

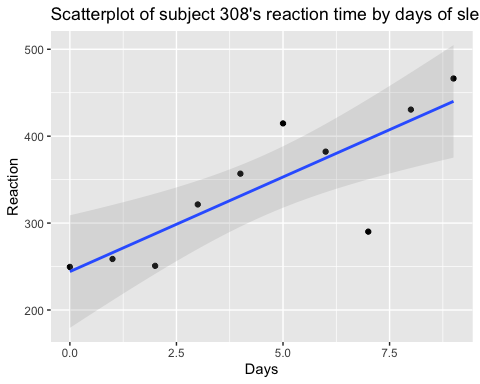
2/3/2020

#Setup  
library(pacman)  
pacman::p\_load("tidyverse")  
  
#Loading the data  
df <- read.csv("sleepstudy.csv", header = T)  
df$Subject <- as.factor(df$Subject)

## Task 1

### 1.a: Get the data from one participant, e.g. using subset(). Make a linear regression for reaction time as a function of days of sleep deprivation, e.g. using lm(). Report the F-statistics.

#Making a subset df  
participan1 <- subset(df, df$Subject == 308)  
  
#Making a scatterplot of subject 308  
ggplot(participan1, aes(Days, Reaction))+  
 geom\_point()+  
 geom\_smooth(method = lm, alpha = 0.2)+  
 ggtitle("Scatterplot of subject 308's reaction time by days of sleep deprivation")



#Linear regression model  
summary(lm(participan1$Reaction~participan1$Days))

##   
## Call:  
## lm(formula = participan1$Reaction ~ participan1$Days)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -106.397 -4.098 9.688 22.269 61.674   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 244.19 28.08 8.695 2.39e-05 \*\*\*  
## participan1$Days 21.77 5.26 4.137 0.00326 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 47.78 on 8 degrees of freedom  
## Multiple R-squared: 0.6815, Adjusted R-squared: 0.6417   
## F-statistic: 17.12 on 1 and 8 DF, p-value: 0.003265

A linear regression analysis was used to test if the amount of days of sleep deprivation significantly predicted reaction time. The results of the regression indicated that the predictor explained 64.17% of the variance in the reaction time (Adjusted R2 =.6417, F(1,8)= 17.12, p<.01). It was found that a larger duration of sleep deprivation significantly predicted reaction time tendencies (β = 21.77, SE =5.26, t = 4.137, p<.01)

### 1.b: How many degrees of freedom does the relevant F-distribution have?

The relevant F-distribution have df=1 for predictors and df=8 for subjects.

### 1.c: At which F-value does a regression with this distribution become statistically significant (p<0.05)?

#Getting the F-score for the 95% significance cutoff using the quantile function qf()  
cutoff<-qf(0.95,1,8)  
'p=0.05 cutoff value with df(1,8)'

## [1] "p=0.05 cutoff value with df(1,8)"

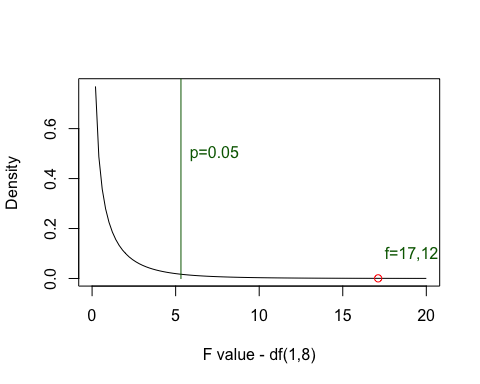
cutoff

## [1] 5.317655

A F-distributio with df(1,8) becomes statistically significant when F>5.318 at a significance level of 0.05.

### 1.d: Make a plot of the F-distribution.

#The F statistics for the effect of days on reaction time  
model<-lm(participan1$Reaction~participan1$Days)  
res=anova(model)  
#Making a string of 100 numbers beteen 0 and 20 to use when finding the F distribution  
nn<-seq(0,20,len=100)  
#Getting the F distribution using df()  
fdist<-df(nn,1,8)  
plot(nn,fdist, type='l',xlab='F value - df(1,8)',ylab='Density')  
#plotting a vertical line at the cutoff  
lines(c(cutoff,cutoff),c(0,1),col='darkgreen')  
#Add explanation for the line  
text(cutoff+2,0.5,'p=0.05',col='darkgreen')  
#draw F-value as point on the curve  
points(res$`F value`[1],df(res$`F value`[1],1,8),col='red')  
#add text for the f-value  
text(res$`F value`[1]+2,0.1,'f=17,12',col='darkgreen')



## Task 2. For all participant in the experiment

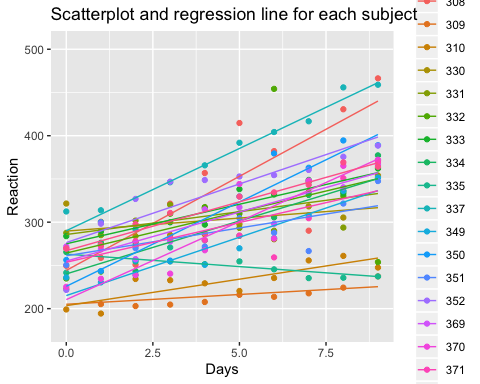
### 2.a: Find the coefficients (slope and intercept) for the regression for reaction time as a function of days of sleep deprivation (a hint for the solution: use group\_by() in tidyverse or this function here: <https://stat.ethz.ch/R-manual/R-devel/library/nlme/html/lmList.html,hint2>: pool=FALSE)

all\_subjects <- df %>% group\_by(Subject) %>% summarise("Intercept" = lm(Reaction~Days)$coefficients[1],  
 "Slope" = lm(Reaction~Days)$coefficients[2])  
all\_subjects

## # A tibble: 18 x 3  
## Subject Intercept Slope  
## <fct> <dbl> <dbl>  
## 1 308 244. 21.8   
## 2 309 205. 2.26  
## 3 310 203. 6.11  
## 4 330 290. 3.01  
## 5 331 286. 5.27  
## 6 332 264. 9.57  
## 7 333 275. 9.14  
## 8 334 240. 12.3   
## 9 335 263. -2.88  
## 10 337 290. 19.0   
## 11 349 215. 13.5   
## 12 350 226. 19.5   
## 13 351 261. 6.43  
## 14 352 276. 13.6   
## 15 369 255. 11.3   
## 16 370 210. 18.1   
## 17 371 254. 9.19  
## 18 372 267. 11.3

### 2.b: Combine both scatter plot and regression line in the same figure. You may also include all participants in one plot.

#Making the scatterplot  
ggplot(df, aes(Days,Reaction, colour = Subject))+  
 geom\_point()+  
 geom\_smooth(method = lm, alpha = 0, size = 0.5)+  
 ggtitle("Scatterplot and regression line for each subject")



### 2.c: Collect and report the inferential statistics for each participant in a table using t-statistics, including t-value, df and p-value.

### 2.d: How many individual participants display a statistically significant effect of sleep deprivation (p-values uncorrected for mulitple comparisons)?

#Adding the inferential t-statistics to the df  
all\_subjects <- df %>% group\_by(Subject) %>% summarise("Intercept" = lm(Reaction~Days)$coefficients[1],  
 "t-value\_intercept" = summary(lm(Reaction~Days))$coefficients[1,3],  
 "p-value\_intercept" = summary(lm(Reaction~Days))$coefficients[1,4],  
 "Slope" = lm(Reaction~Days)$coefficients[2],  
 "t-value\_slope" = summary(lm(Reaction~Days))$coefficients[2,3],  
 "p-value\_slope" = summary(lm(Reaction~Days))$coefficients[2,4],  
 "df" = lm(Reaction~Days)$df.residual)  
  
#Adding a collumn with marked "\*" if p<0.05.  
all\_subjects$significant <- ifelse(all\_subjects$`p-value\_slope`<0.05,"\*"," ")   
  
print.data.frame(all\_subjects)

## Subject Intercept t-value\_intercept p-value\_intercept Slope  
## 1 308 244.1927 8.695486 2.385022e-05 21.764702  
## 2 309 205.0549 39.311440 1.927496e-10 2.261785  
## 3 310 203.4842 28.100452 2.778115e-09 6.114899  
## 4 330 289.6851 22.105427 1.852824e-08 3.008073  
## 5 331 285.7390 20.749421 3.050571e-08 5.266019  
## 6 332 264.2516 7.382976 7.744761e-05 9.566768  
## 7 333 275.0191 37.557051 2.772320e-10 9.142045  
## 8 334 240.1629 19.877588 4.275274e-08 12.253141  
## 9 335 263.0347 39.295935 1.933560e-10 -2.881034  
## 10 337 290.1041 30.242212 1.551288e-09 19.025974  
## 11 349 215.1118 26.118064 4.959747e-09 13.493933  
## 12 350 225.8346 15.772768 2.608897e-07 19.504017  
## 13 351 261.1470 19.523721 4.923090e-08 6.433498  
## 14 352 276.3721 18.426353 7.749702e-08 13.566549  
## 15 369 254.9681 27.400119 3.393323e-09 11.348109  
## 16 370 210.4491 14.845934 4.174433e-07 18.056151  
## 17 371 253.6360 17.211684 1.321176e-07 9.188445  
## 18 372 267.0448 40.264853 1.592623e-10 11.298073  
## t-value\_slope p-value\_slope df significant  
## 1 4.137485 3.264657e-03 8 \*  
## 2 2.314848 4.931443e-02 8 \*  
## 3 4.508107 1.980757e-03 8 \*  
## 4 1.225416 2.552687e-01 8   
## 5 2.041462 7.550229e-02 8   
## 6 1.426926 1.914426e-01 8   
## 7 6.664912 1.583426e-04 8 \*  
## 8 5.414117 6.352350e-04 8 \*  
## 9 -2.297764 5.064731e-02 8   
## 10 10.588367 5.530467e-06 8 \*  
## 11 8.746570 2.285006e-05 8 \*  
## 12 7.272169 8.617903e-05 8 \*  
## 13 2.567717 3.324544e-02 8 \*  
## 14 4.828776 1.306668e-03 8 \*  
## 15 6.510472 1.860407e-04 8 \*  
## 16 6.799987 1.378251e-04 8 \*  
## 17 3.328717 1.040424e-02 8 \*  
## 18 9.094290 1.716323e-05 8 \*

## Task 3: Across participants:

### 3.a: Use the slopes you found for each participant in exercise 2 as a new dataset. Test the hypothesis that the slopes are larger than zero against the null-hypothesis that the slopes are zero (i.e. no differences in response time exist as a function of time).

#Testing for normality   
hist(all\_subjects$Slope)



#By visual inspection the data approximates normal distribution even though the small sample size of n=18.

#Making a 1-sample t-test to test if the slopes significantly differs from 0.   
m <- t.test(all\_subjects$Slope)  
m

##   
## One Sample t-test  
##   
## data: all\_subjects$Slope  
## t = 6.7715, df = 17, p-value = 3.264e-06  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 7.205956 13.728615  
## sample estimates:  
## mean of x   
## 10.46729

### 3.b: Justify your use of test statistics.

T-test is used to test whether two means are statically significant different from each other. We used a 1-sample t-test of the slopes to test if they differed from 0, and in that way if tere is a significant of sleep deprivation on reaction time.

### 3.c: Report inferential statistics.

Using an 1-sample t-test, we found that the mean gradient of the slopes significantly differed from 0, t(17) = 6.77, p<0.001, (M slope = 10.47, M = 0), which indicates that sleep deprication significantly modulates reaction time.

### 3.d: Make a plot with the mean reaction time and standard errror bars for each day across participants and plot the averaged regression line in the same figure.

#Making a plot  
ggplot(df, aes(Days, Reaction))+  
 stat\_summary(fun.y = mean, geom = "point", colour = "black")+  
 stat\_summary(fun.data = mean\_se, geom = "errorbar", width = 0.3)+  
 geom\_smooth(method = lm, alpha = 0)+  
 ggtitle("Mean reaction time pr. day across participants with se-bars")

