portfolio 1 experimental methods 2

Søren Orm (same as; Astrid Olling, Daniel Christensen, Karoline Klitgaard, Pernille Jensen)

3/2/2020

```
sleepstudy <- read.csv("sleepstudy.csv")
```

##Tasks ####1. Plot the data

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().

```
#We choose one random participant from the original sleepstudy dataframe
our_guy <- subset(sleepstudy, sleepstudy$Subject == 352)
```

```
#linear regression on reaction time by days of sleep deprivation on our random participant
#guy_reg <-
summary(lm(Reaction ~ Days, our_guy))
```

```
##
## Call:
## lm(formula = Reaction ~ Days, data = our_guy)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -54.695  -9.763   2.455  15.732  29.784
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   276.37      15.00  18.426 7.75e-08 ***
## Days           13.57       2.81   4.829  0.00131 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.52 on 8 degrees of freedom
## Multiple R-squared:  0.7445, Adjusted R-squared:  0.7126
## F-statistic: 23.32 on 1 and 8 DF,  p-value: 0.001307
```

Report the F-statistics.

The F-statistic is 23.32 significant at p-value = 0.001307

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

The degrees of freedom for the model is 1 and the degrees of freedom for the residuals is 8

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

```
#we calculate the cut-off value
qf(0.95, 1, 8)
```

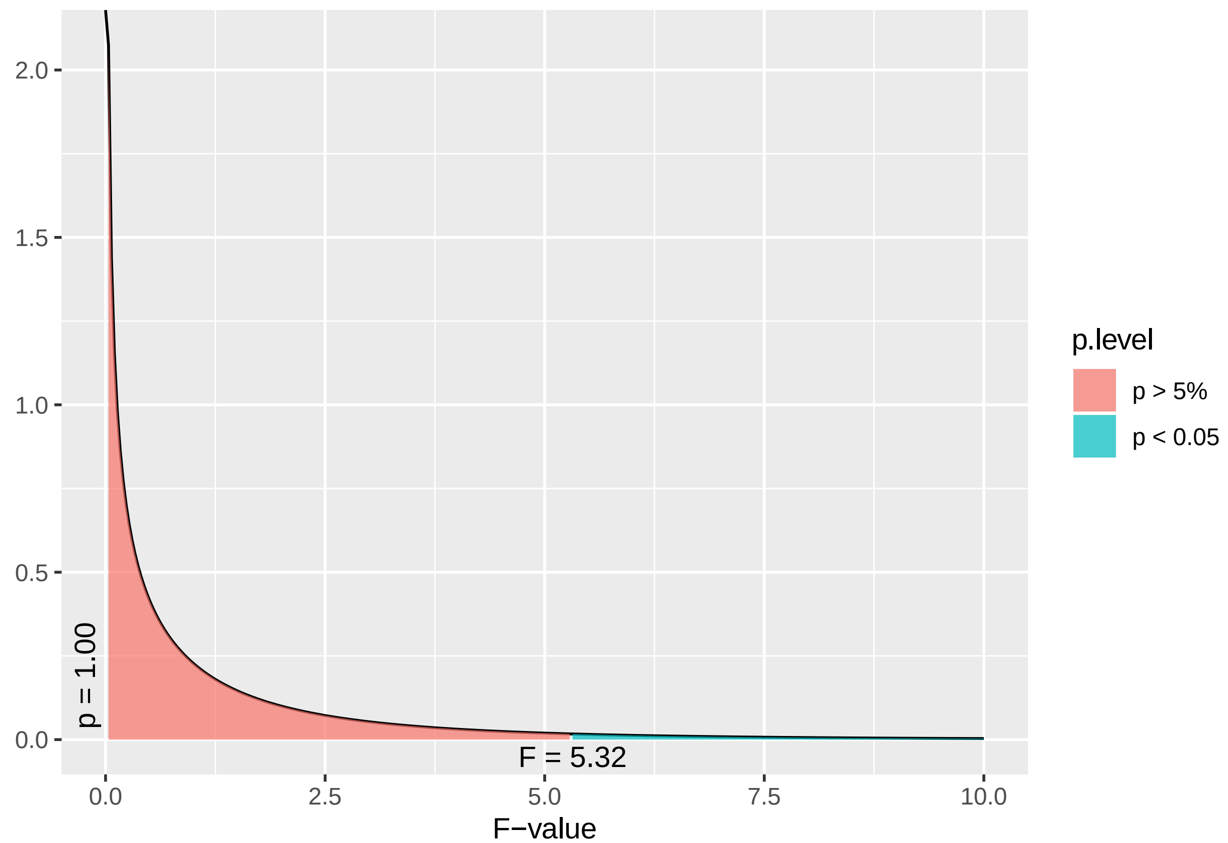
```
## [1] 5.317655
```

```r
#cutoff = 5.317655
```

A regression with this distribution becomes statistically significant at F-value 5.317655

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

```r
dist_f(f=0, deg.f1 = 1, deg.f2 = 8,xmax=10)
```



####2.For all participants 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

```r
#We convert the subject values to factors
sleepstudy$Subject <- as.factor(sleepstudy$Subject)

#Slopes and intercepts for each participants
list <- lmList(Reaction ~ Days | Subject, data = sleepstudy, pool = FALSE)

list
```

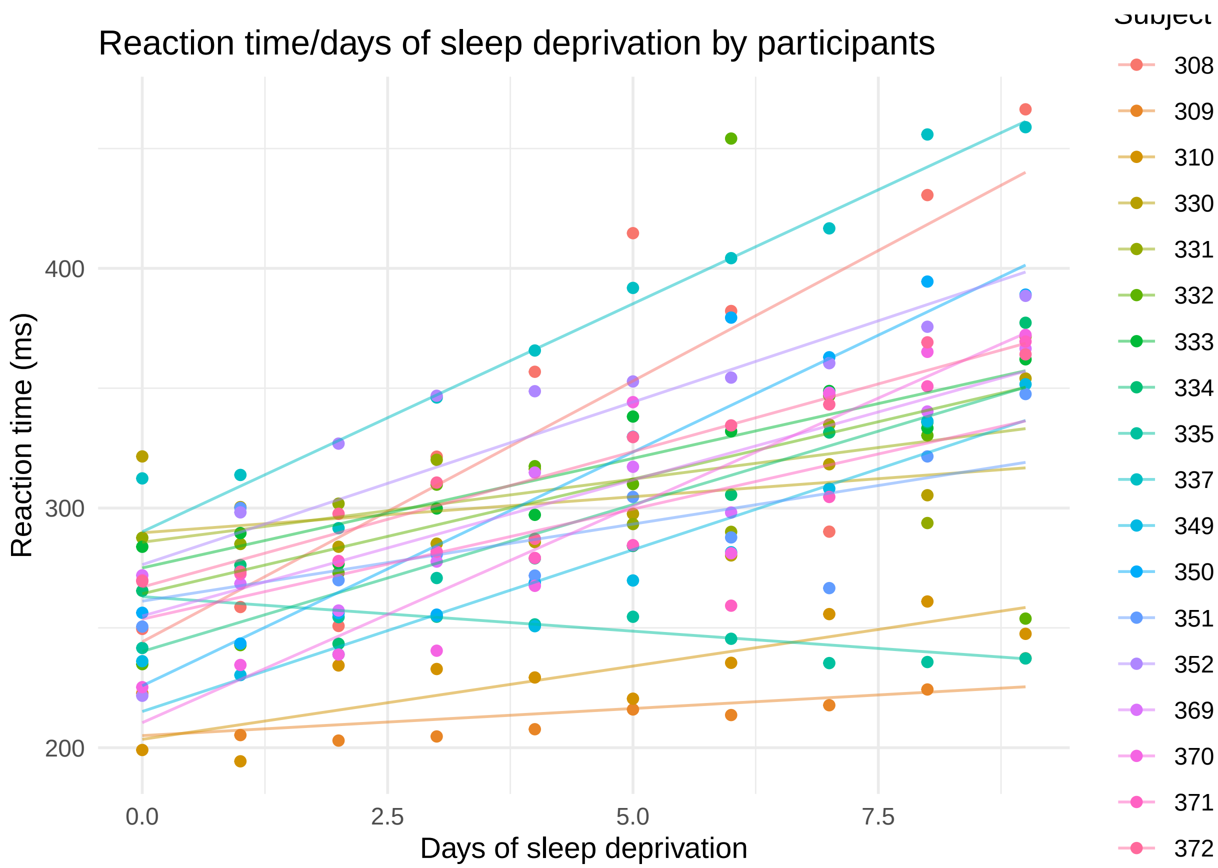
```
## Call:
##   Model: Reaction ~ Days | Subject
##    Data: sleepstudy
##
## Coefficients:
```

```
##     (Intercept)      Days
## 308    244.1927 21.764702
## 309    205.0549  2.261785
## 310    203.4842  6.114899
## 330    289.6851  3.008073
## 331    285.7390  5.266019
## 332    264.2516  9.566768
## 333    275.0191  9.142045
## 334    240.1629 12.253141
## 335    263.0347 -2.881034
## 337    290.1041 19.025974
## 349    215.1118 13.493933
## 350    225.8346 19.504017
## 351    261.1470  6.433498
## 352    276.3721 13.566549
## 369    254.9681 11.348109
## 370    210.4491 18.056151
## 371    253.6360  9.188445
## 372    267.0448 11.298073
```

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

```
#group by subject
grouped_pers <- group_by(sleepstudy, Subject)

#creating a plot
ggplot(grouped_pers, aes(Days, Reaction, color = Subject)) +
  geom_point() +
  stat_smooth(method = lm, geom='line', alpha=0.5, se=FALSE)+
  theme_minimal()+
    labs(title = "Reaction time/days of sleep deprivation by participants", x = "Days of sleep deprivati
```



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

```
#Saving a summary of the output from the linear model:
listsummary <- summary(list)

#Creating a data frame from the output:
final <- data.frame(listsummary[["coefficients"]])

#with relevant columns
final <- data.frame(row.names(final), final$t.value.Days, final$Pr...t...Days)

#Adding columns with degrees of freedom:
final$df_model <- 1
final$df_res <- 8

#Changing the column names:
colnames(final)[1] <- "ID"
colnames(final)[2] <- "t values"
colnames(final)[3] <- "p values"
```

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

When looking at the p-values from the output it seems that sleep deprivation has a significant effect on reaction time for all participants. It should hovever be emphasized that one of the participants seems to have decreased reaction time by days of sleep deprivation.

####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

```
#dataframe with only the coefficients
all_coefs <- coef(list, data = sleepstudy)

#t-test for testing the hypothesis
t.test(all_coefs$Days, alternative = "greater", paired = FALSE)
```

```
##
##  One Sample t-test
##
## data:  all_coefs$Days
## t = 6.7715, df = 17, p-value = 1.632e-06
## alternative hypothesis: true mean is greater than 0
## 95 percent confidence interval:
##  7.778221      Inf
## sample estimates:
## mean of x
##  10.46729
```

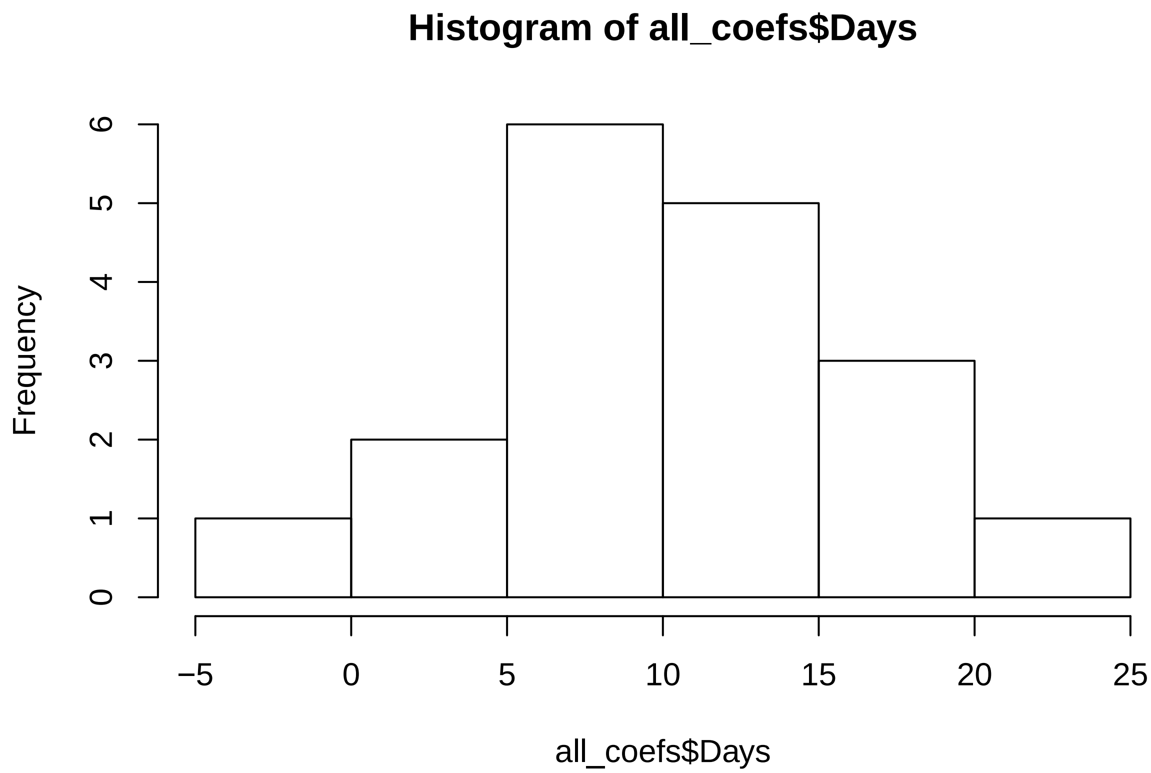
3.b: Justify your use of test statistics.

```
#checking if the data is normally distributed

#numeric inspection of data distribution
round(stat.desc(all_coefs$Days, norm = TRUE, basic = FALSE), digits=2)
```

```
##       median         mean      SE.mean CI.mean.0.95          var
##        10.43        10.47         1.55         3.26        43.01
##      std.dev     coef.var     skewness     skew.2SE     kurtosis
##         6.56         0.63        -0.07        -0.06        -0.82
##     kurt.2SE   normtest.W   normtest.p
##        -0.40         0.98         0.94
```

```
#visual inspection of data distribution
hist(all_coefs$Days)
```



We made a one tailed t-test which compares the alternative hypothesis to the null hypothesis. The data is normally distributed (the shape of the histogram, the Shapiro Wilks test for normality, and the skew.2SE and kurt.2SE indicates normality)

3.c: Report inferential statistics.

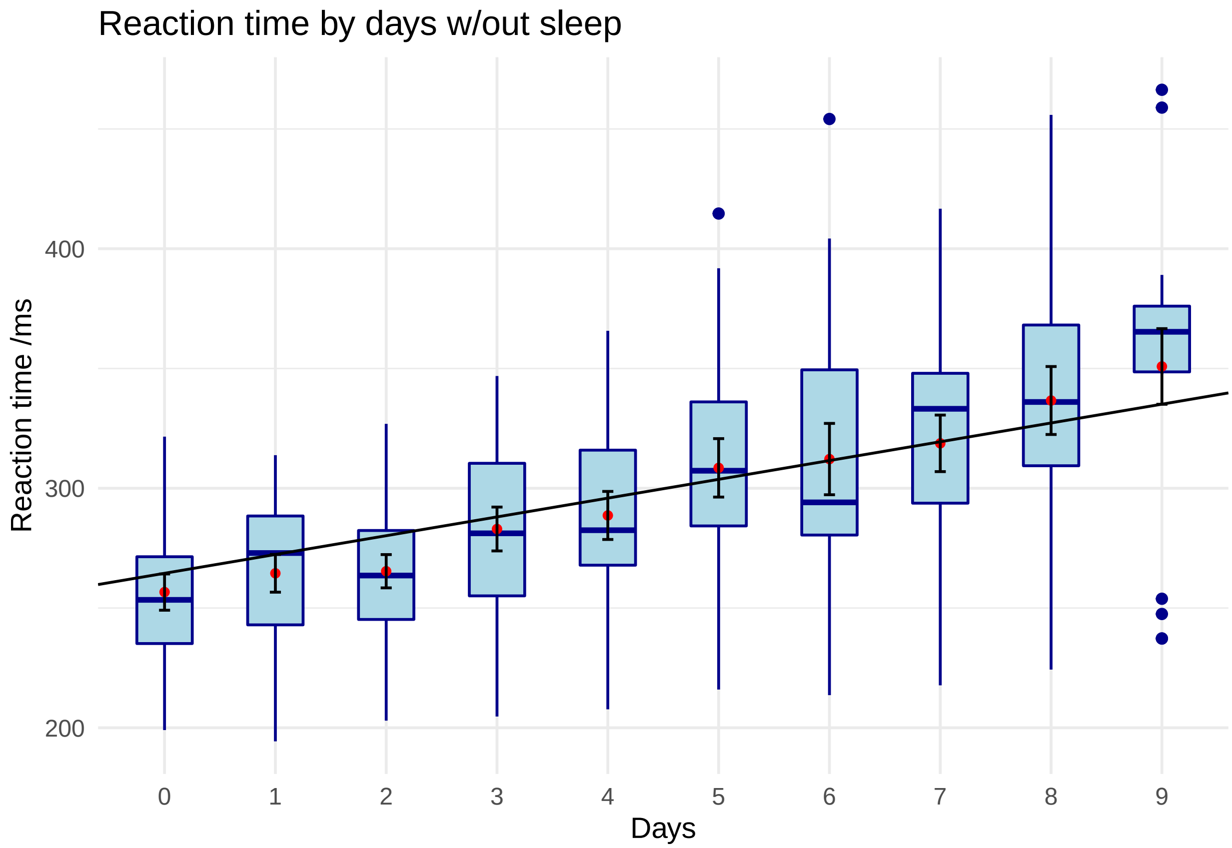
The t-test is significant at $t(17) = 6.7715$, p-value = 1.632e-06 Which leaves us to reject the null hypothesis that the sample mean is equal to zero

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.

```
#days as factor
sleepstudy$Days <- as.factor(sleepstudy$Days)

#boxplot for mean reaction time by days
reg_line <- coef(lm(Reaction ~ Days, data = sleepstudy))

ggplot(sleepstudy, aes(x = Days, y = Reaction)) +
  geom_boxplot(width=0.5, color = 'darkblue', fill = 'lightblue') +
  stat_summary(fun.y = mean, geom = "point", shape = 16, color="red") +
  stat_summary(fun.data = mean_se, geom = "errorbar", color = 'black', width = 0.1) +
  geom_abline(intercept = reg_line[1], slope = reg_line[2]) +
  ggtitle("Reaction time by days w/out sleep") +
  labs(x = "Days", y = "Reaction time /ms") +
  theme_minimal()
```



Reaction time by days w/out sleep
Reaction time /ms
400
300
200
0
1
2
3
4
5
6
7
8
9
Days