Portfolio 1 - Regression

Nanna Bernth, Sebastian Scott, Roxana Petrache, Fredrik Sejr & Signe M. R. Holdgaard

Study Group 2

2/7/2018

Teacher Comment:
tip: the last graph is a bit confusing with two lines. I would suggest using "geom="bar"" in the stat_summary function call.

Sleep deprivation exercise

Data + packages

```
#Load data
setwd("~/Dropbox/Uni/2 semester/Ekseperimental Methods 2/R-code")
sleepstudy <- read.csv("sleepstudy.csv")
str(sleepstudy)
```

```
## 'data.frame':    180 obs. of  4 variables:
##  $ X       : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Reaction: num  250 259 251 321 357 ...
##  $ Days    : int  0 1 2 3 4 5 6 7 8 9 ...
##  $ Subject : int  308 308 308 308 308 308 308 308 308 308 ...
```

```
#Load packages
library(lmerTest) #in order to make linear regression
library(sjPlot) #in order to plot F-distribution
library(nlme)
library(tidyverse)
```

1. Plot the data

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

#We chose participant 331 and use subset to create a dataframe only containing par
ticipant 331
participant331 <- subset(sleepstudy, Subject == 331)

#print the dataframe
participant331
```

```
##     X Reaction Days Subject
## 41 41 287.6079    0     331
## 42 42 285.0000    1     331
```

```
## 43 43 301.8206    2     331
## 44 44 320.1153    3     331
## 45 45 316.2773    4     331
## 46 46 293.3187    5     331
## 47 47 290.0750    6     331
## 48 48 334.8177    7     331
## 49 49 293.7469    8     331
## 50 50 371.5811    9     331
```

```
#We make linear regression using lm. First we create the model with Reaction as ou
tcome and days as predictor
model_participant331 <- lm(Reaction ~ Days, participant331)

#We summarise the model
summary(model_participant331)
```

```
##
## Call:
## lm(formula = Reaction ~ Days, data = participant331)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.120 -15.564   3.709  11.531  38.448
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  285.739     13.771  20.749 3.05e-08 ***
## Days           5.266      2.580   2.041   0.0755 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.43 on 8 degrees of freedom
## Multiple R-squared:  0.3425, Adjusted R-squared:  0.2603
## F-statistic: 4.168 on 1 and 8 DF,  p-value: 0.0755
```

1.a Response:

Days was found not to significantly predict reaction time for participant 331, $\beta = 5.266$ (SE = 2.58), $t = 2.04$, $p = 0.08$ and $F(1,8) = 4.17$.

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

1.b Response:

An F-distribution has two degrees of freedom - So in our case the degrees of freedom are (1,8)

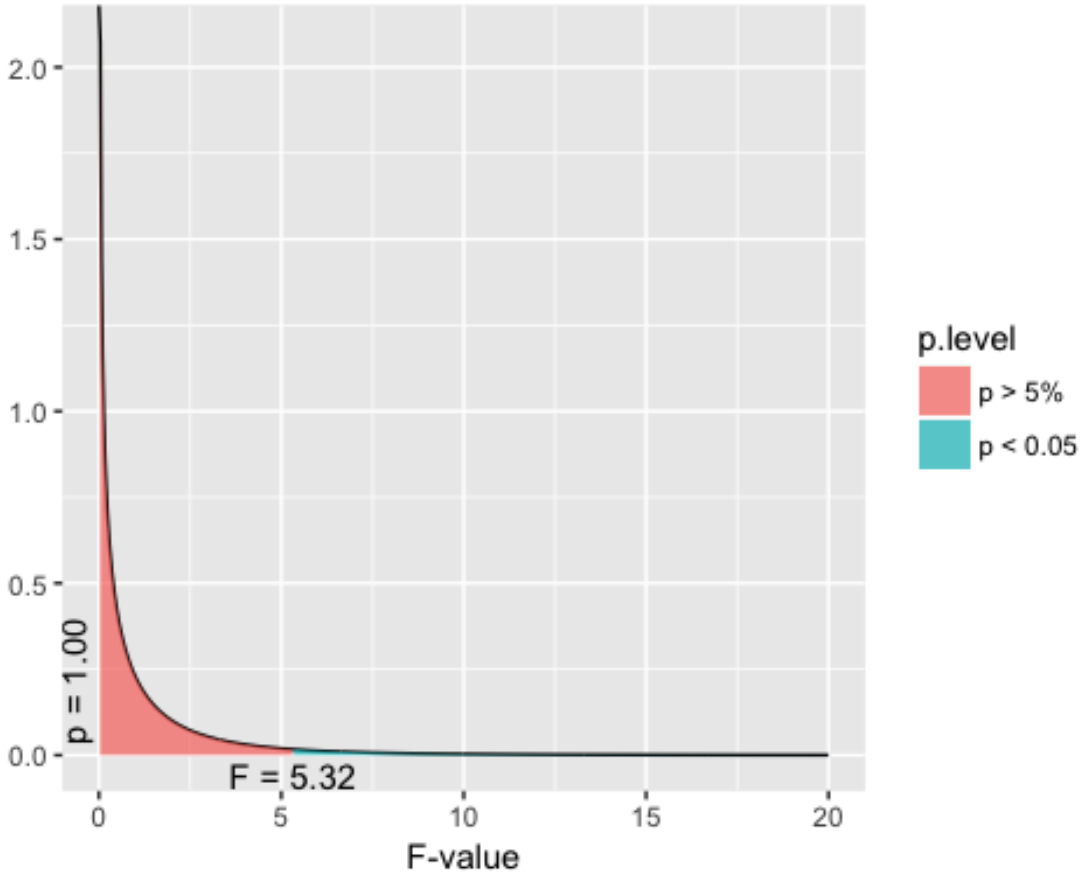
```r
#1.c: At which F-value does a regression with this distribution become statistical
ly significant (p<0.05)?

#We use the qf function to find the f-value that would have given us a statistical
ly significant p-value (< 0.05)
qf(.95, df1 = 1, df2 = 8)
```

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

1.c Response:

At the f-value 5.32 and above the regression would have become statistically significant ($p < 0.05$)

```r
#1.d: Make a plot of the F-distribution
#Use the sjPlot package to use the dist_f function in order to make a plot of the
F-distribution.
#We set f=0 and the max at 20. We add the degrees of freedom (1,8)
dist_f(f=0, deg.f1 = 1, deg.f2 = 8, xmax=20)
```

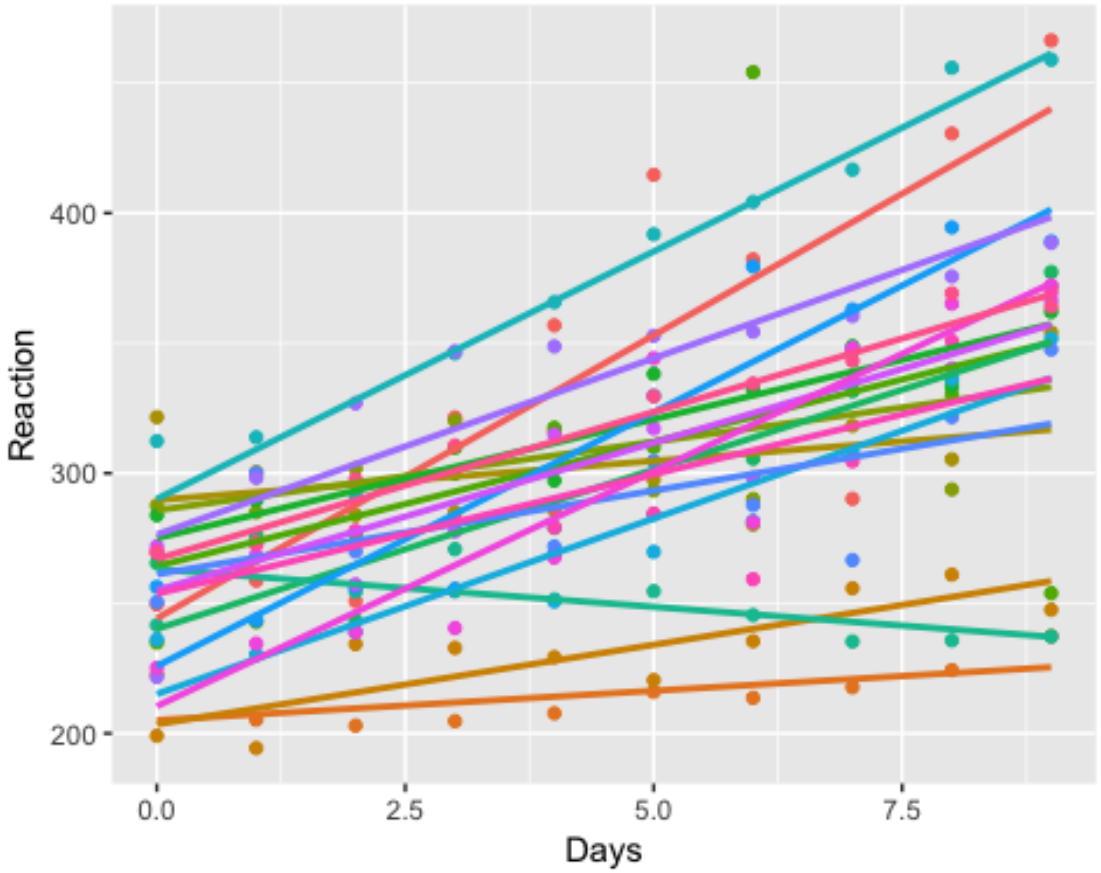


2. For all participants in the experiment

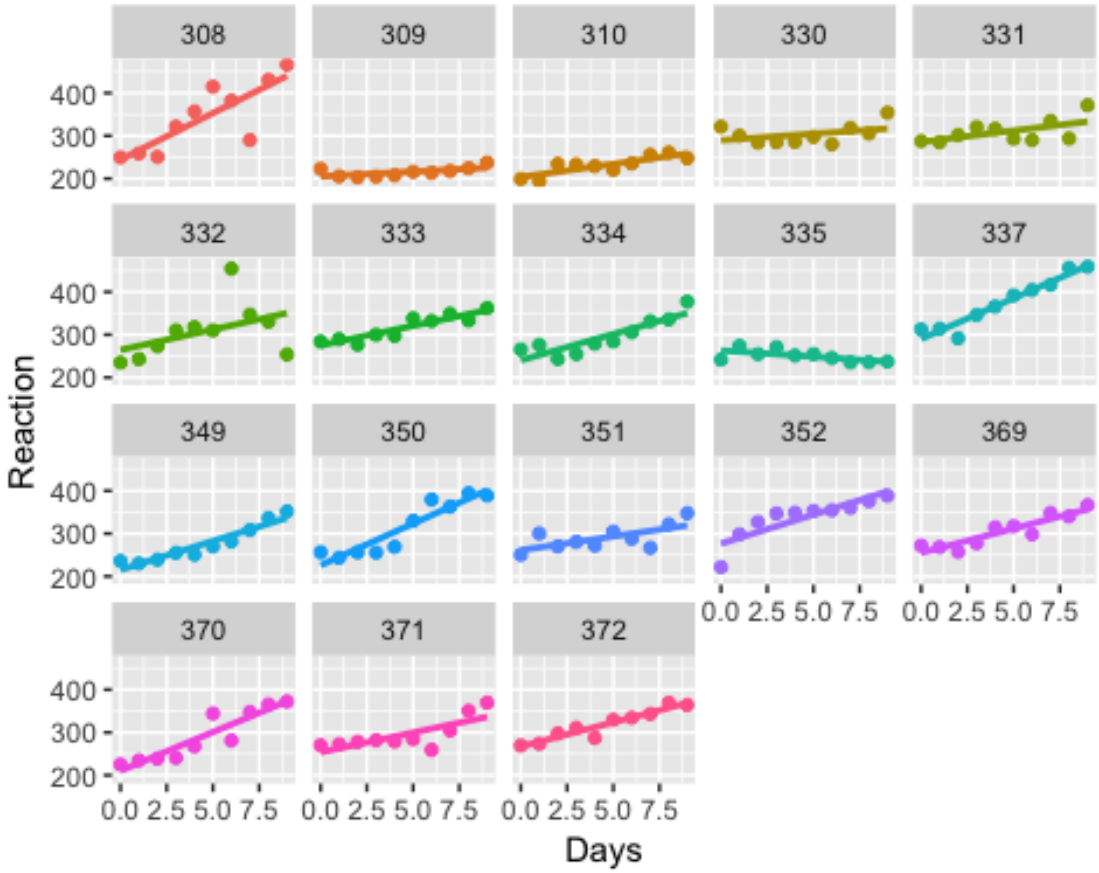
```
#2.a: Find the coefficients (slope and intercept) for the regression for reaction
time as a function of days of sleep deprivation

#use lmList to make a model with the coefficients with Reaction as outcome and pre
dicted by days with Subject as random intercept
model_coefficients <- lmList(Reaction ~ Days | Subject, sleepstudy, pool = FALSE)

#summarise model
sum_model <- summary(model_coefficients)

#print
sum_model
```

```
## Call:
##   Model: Reaction ~ Days | Subject
##    Data: sleepstudy
##
## Coefficients:
##    (Intercept)
##     Estimate Std. Error   t value     Pr(>|t|)
## 308 244.1927  28.082693  8.695486 2.385022e-05
## 309 205.0549   5.216165 39.311440 1.927496e-10
## 310 203.4842   7.241315 28.100452 2.778115e-09
## 330 289.6851  13.104705 22.105427 1.852824e-08
## 331 285.7390  13.770937 20.749421 3.050571e-08
## 332 264.2516  35.792017  7.382976 7.744761e-05
## 333 275.0191   7.322702 37.557051 2.772320e-10
## 334 240.1629  12.082095 19.877588 4.275274e-08
## 335 263.0347   6.693687 39.295935 1.933560e-10
## 337 290.1041   9.592689 30.242212 1.551288e-09
## 349 215.1118   8.236130 26.118064 4.959747e-09
## 350 225.8346  14.318007 15.772768 2.608897e-07
## 351 261.1470  13.375883 19.523721 4.923090e-08
## 352 276.3721  14.998740 18.426353 7.749702e-08
## 369 254.9681   9.305367 27.400119 3.393323e-09
## 370 210.4491  14.175538 14.845934 4.174433e-07
## 371 253.6360  14.736271 17.211684 1.321176e-07
## 372 267.0448   6.632206 40.264853 1.592623e-10
##    Days
##      Estimate Std. Error   t value     Pr(>|t|)
## 308 21.764702  5.2603704  4.137485 3.264657e-03
## 309  2.261785  0.9770772  2.314848 4.931443e-02
## 310  6.114899  1.3564226  4.508107 1.980757e-03
## 330  3.008073  2.4547361  1.225416 2.552687e-01
## 331  5.266019  2.5795328  2.041462 7.550229e-02
## 332  9.566768  6.7044590  1.426926 1.914426e-01
```

```
## 333  9.142045  1.3716678  6.664912 1.583426e-04
## 334 12.253141  2.2631838  5.414117 6.352350e-04
## 335 -2.881034  1.2538425 -2.297764 5.064731e-02
## 337 19.025974  1.7968753 10.588367 5.530467e-06
## 349 13.493933  1.5427685  8.746570 2.285006e-05
## 350 19.504017  2.6820084  7.272169 8.617903e-05
## 351  6.433498  2.5055324  2.567717 3.324544e-02
## 352 13.566549  2.8095214  4.828776 1.306668e-03
## 369 11.348109  1.7430549  6.510472 1.860407e-04
## 370 18.056151  2.6553215  6.799987 1.378251e-04
## 371  9.188445  2.7603564  3.328717 1.040424e-02
## 372 11.298073  1.2423260  9.094290 1.716323e-05

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

#Use ggplot to make a scatter plot of the data points
plot <- ggplot(sleepstudy, aes(Days, Reaction, colour = factor(Subject))) + #add a
stetics
  geom_point() + #creates an extra layer to the plot with the data points
  geom_smooth(method = lm, se = FALSE) + #add a linear model to the plot
  theme(legend.position ="none") #remove legends

#plot all participants in one plot
plot
```




```
#Plot individual plot by facet wrapping each subject
plot + facet_wrap(~ Subject)
```



2. b response

See plots

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

#use summary information from our model and add to a new data frame
stat_data <- data.frame(sum_model$coefficients)

#We find the degrees of freedom for the residuals for one of the participants, bec
ause we know, that it will be the same for all participants
df.residual(model_participant331)

## [1] 8

#we are still missing degrees of freedom in our table, but this can be added to th
e dataframe by using the function mutate()
stat_data <- mutate(stat_data, df = 8)
```

```r
#print dataframe with the degress of freedom
stat_data
```

```
##    Estimate..Intercept. Std..Error..Intercept. t.value..Intercept.
## 1              244.1927              28.082693            8.695486
## 2              205.0549               5.216165           39.311440
## 3              203.4842               7.241315           28.100452
## 4              289.6851              13.104705           22.105427
## 5              285.7390              13.770937           20.749421
## 6              264.2516              35.792017            7.382976
## 7              275.0191               7.322702           37.557051
## 8              240.1629              12.082095           19.877588
## 9              263.0347               6.693687           39.295935
## 10             290.1041               9.592689           30.242212
## 11             215.1118               8.236130           26.118064
## 12             225.8346              14.318007           15.772768
## 13             261.1470              13.375883           19.523721
## 14             276.3721              14.998740           18.426353
## 15             254.9681               9.305367           27.400119
## 16             210.4491              14.175538           14.845934
## 17             253.6360              14.736271           17.211684
## 18             267.0448               6.632206           40.264853
##    Pr...t....Intercept. Estimate.Days Std..Error.Days t.value.Days
## 1          2.385022e-05     21.764702       5.2603704     4.137485
## 2          1.927496e-10      2.261785       0.9770772     2.314848
## 3          2.778115e-09      6.114899       1.3564226     4.508107
## 4          1.852824e-08      3.008073       2.4547361     1.225416
## 5          3.050571e-08      5.266019       2.5795328     2.041462
## 6          7.744761e-05      9.566768       6.7044590     1.426926
## 7          2.772320e-10      9.142045       1.3716678     6.664912
## 8          4.275274e-08     12.253141       2.2631838     5.414117
## 9          1.933560e-10     -2.881034       1.2538425    -2.297764
## 10         1.551288e-09     19.025974       1.7968753    10.588367
## 11         4.959747e-09     13.493933       1.5427685     8.746570
## 12         2.608897e-07     19.504017       2.6820084     7.272169
## 13         4.923090e-08      6.433498       2.5055324     2.567717
## 14         7.749702e-08     13.566549       2.8095214     4.828776
## 15         3.393323e-09     11.348109       1.7430549     6.510472
## 16         4.174433e-07     18.056151       2.6553215     6.799987
## 17         1.321176e-07      9.188445       2.7603564     3.328717
## 18         1.592623e-10     11.298073       1.2423260     9.094290
##    Pr...t...Days df
## 1   3.264657e-03  8
## 2   4.931443e-02  8
## 3   1.980757e-03  8
## 4   2.552687e-01  8
## 5   7.550229e-02  8
## 6   1.914426e-01  8
```

```
## 7   1.583426e-04  8
## 8   6.352350e-04  8
## 9   5.064731e-02  8
## 10  5.530467e-06  8
## 11  2.285006e-05  8
## 12  8.617903e-05  8
## 13  3.324544e-02  8
## 14  1.306668e-03  8
## 15  1.860407e-04  8
## 16  1.378251e-04  8
## 17  1.040424e-02  8
## 18  1.716323e-05  8
```

```
#2.d: How many individual participants display a statistically significant effect
of sleep deprivation (p-values uncorrected for mulitple comparisons)?
significant_sleep <- filter(stat_data, Pr...t...Days < 0.05)

#print results
significant_sleep
```

```
##    Estimate..Intercept. Std..Error..Intercept. t.value..Intercept.
## 1              244.1927              28.082693            8.695486
## 2              205.0549               5.216165           39.311440
## 3              203.4842               7.241315           28.100452
## 4              275.0191               7.322702           37.557051
## 5              240.1629              12.082095           19.877588
## 6              290.1041               9.592689           30.242212
## 7              215.1118               8.236130           26.118064
## 8              225.8346              14.318007           15.772768
## 9              261.1470              13.375883           19.523721
## 10             276.3721              14.998740           18.426353
## 11             254.9681               9.305367           27.400119
## 12             210.4491              14.175538           14.845934
## 13             253.6360              14.736271           17.211684
## 14             267.0448               6.632206           40.264853
##    Pr...t....Intercept. Estimate.Days Std..Error.Days t.value.Days
## 1          2.385022e-05     21.764702       5.2603704     4.137485
## 2          1.927496e-10      2.261785       0.9770772     2.314848
## 3          2.778115e-09      6.114899       1.3564226     4.508107
## 4          2.772320e-10      9.142045       1.3716678     6.664912
## 5          4.275274e-08     12.253141       2.2631838     5.414117
## 6          1.551288e-09     19.025974       1.7968753    10.588367
## 7          4.959747e-09     13.493933       1.5427685     8.746570
## 8          2.608897e-07     19.504017       2.6820084     7.272169
## 9          4.923090e-08      6.433498       2.5055324     2.567717
## 10         7.749702e-08     13.566549       2.8095214     4.828776
## 11         3.393323e-09     11.348109       1.7430549     6.510472
## 12         4.174433e-07     18.056151       2.6553215     6.799987
```

```
## 13         1.321176e-07      9.188445       2.7603564     3.328717
## 14         1.592623e-10     11.298073       1.2423260     9.094290
##    Pr...t...Days df
## 1   3.264657e-03  8
## 2   4.931443e-02  8
## 3   1.980757e-03  8
## 4   1.583426e-04  8
## 5   6.352350e-04  8
## 6   5.530467e-06  8
## 7   2.285006e-05  8
## 8   8.617903e-05  8
## 9   3.324544e-02  8
## 10  1.306668e-03  8
## 11  1.860407e-04  8
## 12  1.378251e-04  8
## 13  1.040424e-02  8
## 14  1.716323e-05  8
```

2. d response

14 out of 18 have shows a significant effect 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-hypoth
esis that the slopes are zero (i.e. no differences in response time exist as a fun
ction of time).

#make a t-test and compare our results with zero
t.test(stat_data$Estimate.Days, mu=0)

##
##  One Sample t-test
##
## data:  stat_data$Estimate.Days
## 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.a Response

See results

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

3.b Response

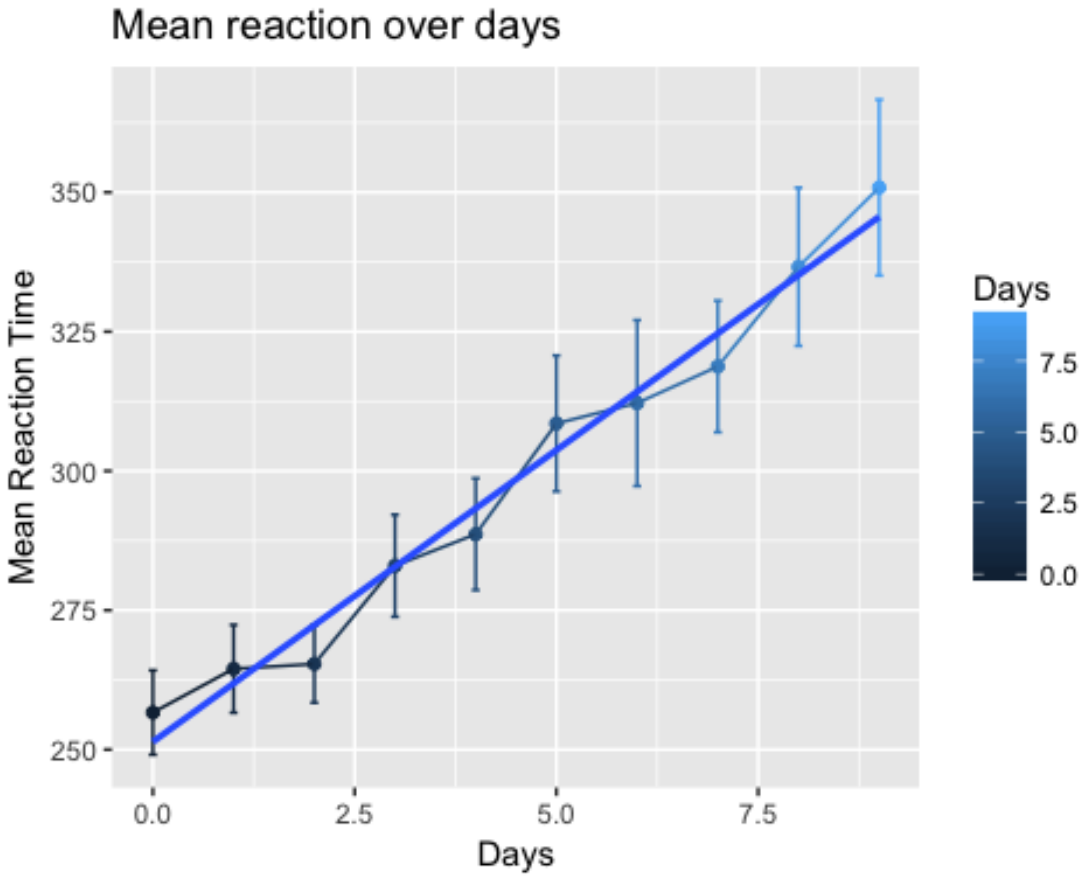
We conduct a one-sample t-test. T-tests are used when comparing if the means significantly differs from each other - in our case we are comparing our model's slopes to the null-hypothesis - a slope of 0.

```
#3.c: Report inferential statistics.
```

3.c Response

The t-test shows that there is a significant difference in response time as a function of days compared to the null-hypothesis, $t(17) = 6.77$, $p < .001$.
We have confidence intervals that does not cross 0, which means it is a positive relationship.

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

ggplot(sleepstudy, aes(x=Days, y=Reaction, color = Days))+
  geom_point(stat = "summary", fun.y=mean)+
  stat_summary(fun.y=mean, geom="line")+
  geom_errorbar(stat = "summary", fun.data = mean_se, width = 0.1)+ #add errorbars
with standard error
  labs(title = "Mean reaction over days", x = "Days", y = "Mean Reaction Time") +
geom_smooth(method = "lm", se =F) #add title
```


Mean reaction over days
Mean Reaction Time
350
325
300
275
250
0.0
2.5
5.0
7.5
Days
Days
7.5
5.0
2.5
0.0