# portfolio 1 experimental methods 2

*Astrid Olling, Daniel Christensen, Karoline Klitgaard, Pernille Jensen, Søren Orm*

*3/2/2020*

Setting up knitr and loading the relevant packages:

Loading the sleep data to a data frame:

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

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

1.a: Get the data from one participant:

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

Make a linear regression for reaction time as a function of days of sleep deprivation:

```
#linear regression on reaction time by days of sleep deprivation on our random participant

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:

An increase in number of days with sleep deprivation was found to significantly predict an increase in reaction time at $F(1,8) = 23.32$, $p > 0.01$

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 ($p<0.05$)?

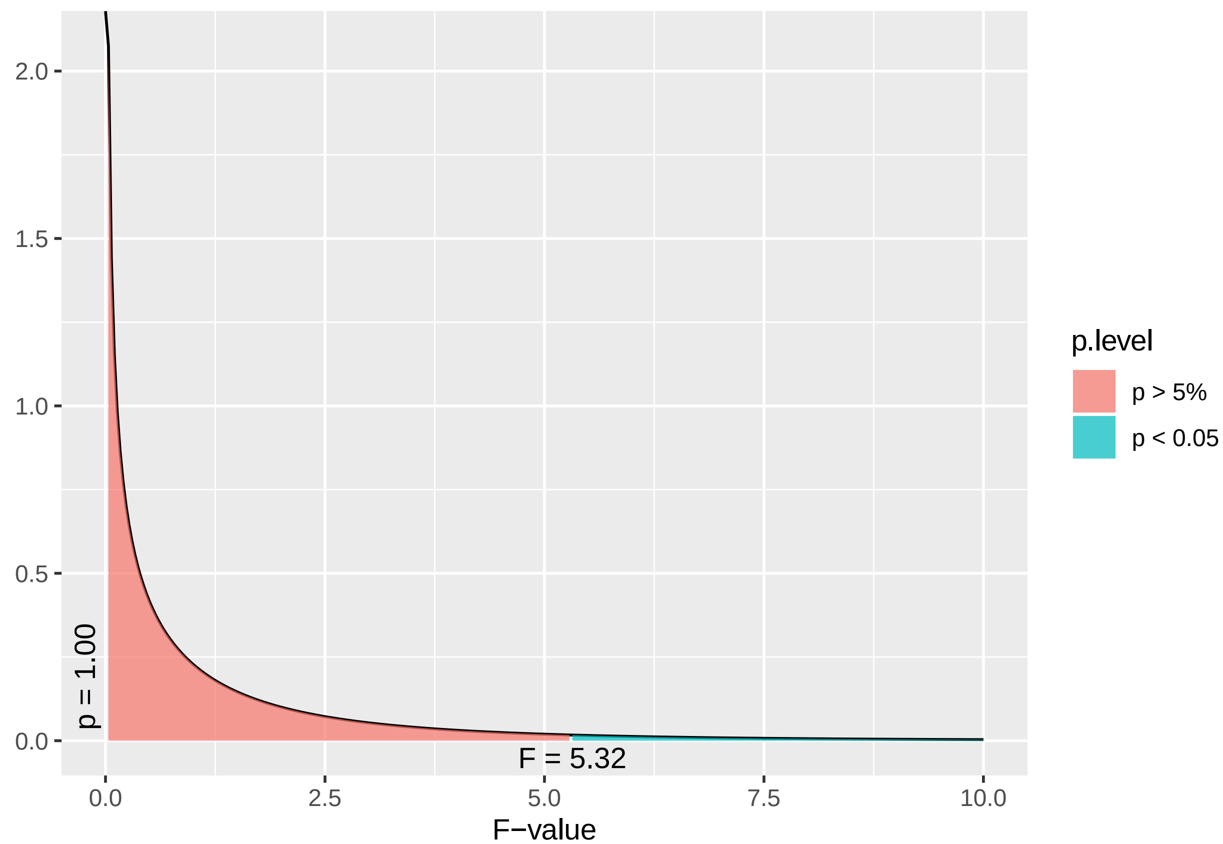
```r
#we calculate the cut-off value at 0.95
qf(0.95, 1, 8)
```

```
## [1] 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

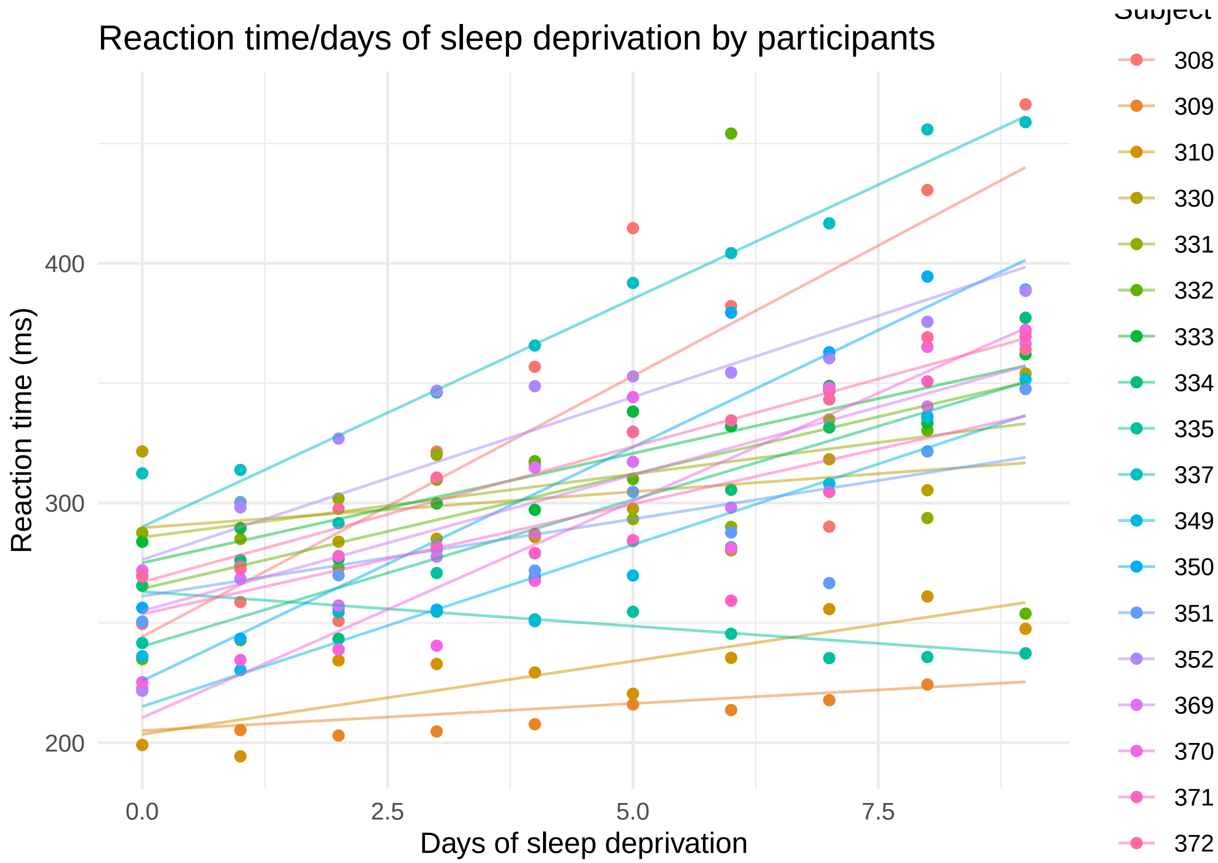
```r
#We convert the subject values to factors
sleepstudy$Subject <- as.factor(sleepstudy$Subject)

#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.

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

#creating a plot with all participants
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 deprivat
```




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)?

An increase in number of days with sleep deprivation is found to significanty predict an increase in reaction time for all participants except one. For subject (find the number) an increase in days with sleep deprivation is instead found to predict a decrease in reaction time. For all participants the slopes are significant at $p > 0.05$.

####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
```

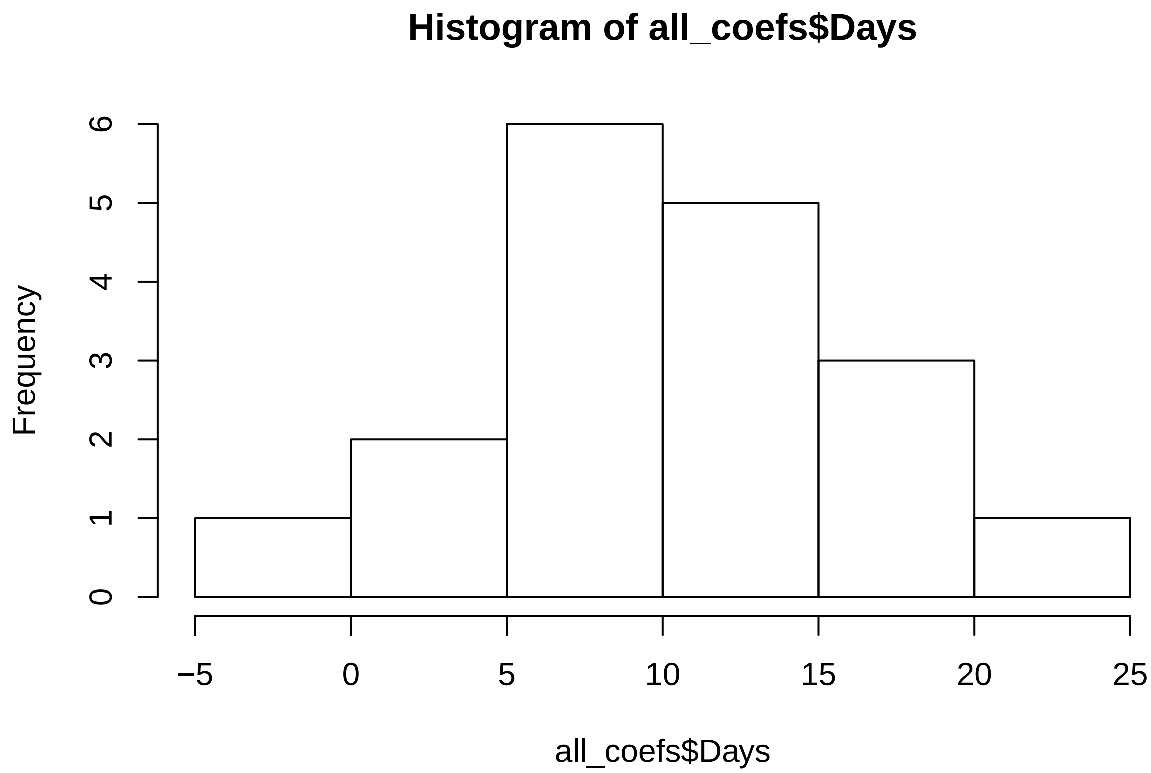
3.b: Justify your use of test statistics.

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)
```



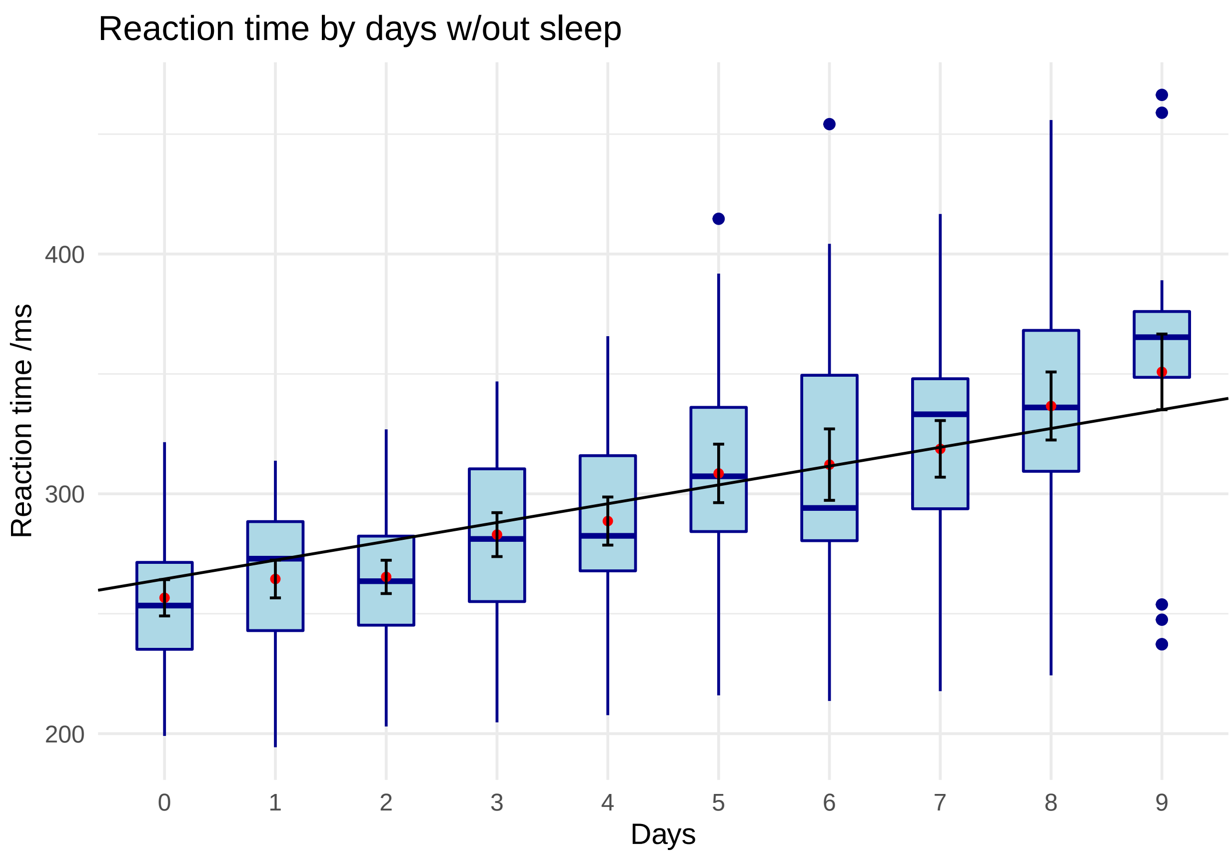

We conducted a one-tailed t-test testing the alternative hypothesis against the null hypothesis. This test can be used because the data is sufficiently normally distributed: the shape of the histogram, the Shapiro Wilks test for normality, and the skew.2SE and kurt.2SE indicates normality. It should however be noticed that a bigger sample size would make the test more reliable.

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()
```




####Voluntory bonus task Add 10% white/uniform noise to the Reaction data. What happens to the single participant coefficients statistics and to the group effects?

```
sleepstudy$noise_reaction = sleepstudy$Reaction+sleepstudy$Reaction*runif(10, min = -0.05, max = 0.05)

sleepstudy$Subject <- as.factor(sleepstudy$Subject)

sleepstudy$Days <- as.numeric(sleepstudy$Days)

list_noise <- lmList(noise_reaction ~ Days | Subject, data = sleepstudy, pool = FALSE)
summary(list_noise)
```

```
## Call:
##   Model: noise_reaction ~ Days | Subject
##    Data: sleepstudy
##
## Coefficients:
##    (Intercept)
##     Estimate Std. Error   t value     Pr(>|t|)
## 308 228.2487  33.100005  6.895730 1.250661e-04
## 309 206.0646   7.745271 26.605211 4.284613e-09
## 310 200.6960  10.628394 18.883000 6.396661e-08
## 330 291.3360  16.627233 17.521619 1.149203e-07
## 331 285.2059  18.675840 15.271382 3.353082e-07
## 332 258.6717  39.766474  6.504767 1.871615e-04
## 333 270.4622   9.871515 27.398250 3.395158e-09
```

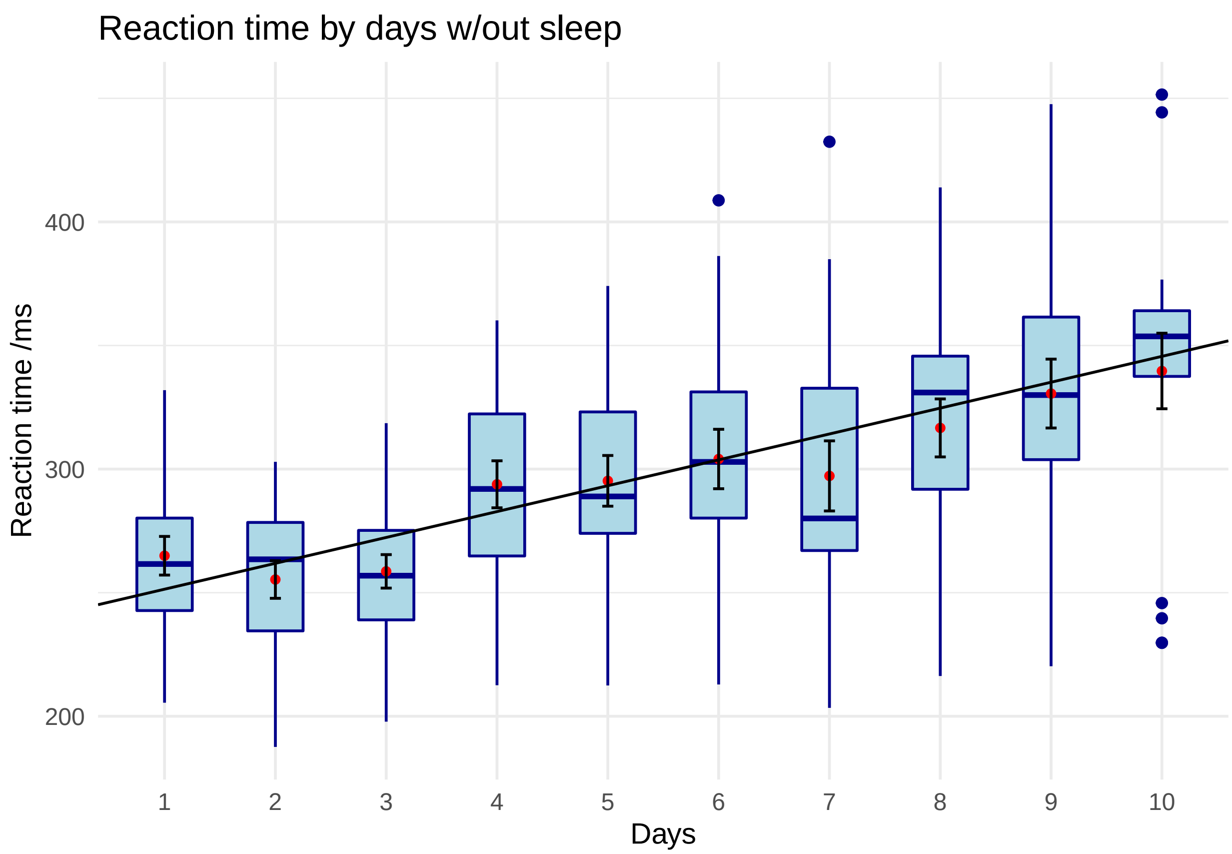
```
## 334 232.3838  13.898355 16.720236 1.656369e-07
## 335 268.9263   8.661099 31.049899 1.258391e-09
## 337 277.1044  14.514850 19.091097 5.869737e-08
## 349 206.0299  11.095592 18.568630 7.296437e-08
## 350 211.1395  14.684867 14.378030 5.348625e-07
## 351 258.4445  13.964632 18.507074 7.488784e-08
## 352 266.9275  18.886019 14.133605 6.106349e-07
## 369 248.3762  14.250920 17.428786 1.197923e-07
## 370 196.7034  17.212184 11.428151 3.107517e-06
## 371 248.9707  18.317168 13.592207 8.252741e-07
## 372 260.3478   9.235317 28.190462 2.708563e-09
##    Days
##      Estimate Std. Error    t value     Pr(>|t|)
## 308 20.059694   5.334549  3.7603356 5.541671e-03
## 309  1.324445   1.248263  1.0610299 3.196552e-01
## 310  5.131985   1.712921  2.9960433 1.717495e-02
## 330  1.682223   2.679721  0.6277606 5.476572e-01
## 331  3.929500   3.009884  1.3055319 2.280036e-01
## 332  8.275689   6.408948  1.2912711 2.326684e-01
## 333  7.759415   1.590939  4.8772559 1.228621e-03
## 334 10.894408   2.239923  4.8637430 1.249852e-03
## 335 -3.806315   1.395863 -2.7268555 2.597181e-02
## 337 17.255484   2.339280  7.3764084 7.793687e-05
## 349 12.190668   1.788217  6.8172210 1.354259e-04
## 350 17.984085   2.366681  7.5988644 6.311672e-05
## 351  5.228612   2.250604  2.3232039 4.867537e-02
## 352 12.203343   3.043758  4.0093019 3.898851e-03
## 369  9.971789   2.296744  4.3417074 2.472700e-03
## 370 16.714074   2.773995  6.0252728 3.144249e-04
## 371  7.871721   2.952079  2.6665008 2.851654e-02
## 372  9.895811   1.488406  6.6485961 1.610417e-04
```

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

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

```
##
##  One Sample t-test
##
## data:  all_coefs_noise$Days
## t = 6.0849, df = 17, p-value = 6.068e-06
## alternative hypothesis: true mean is greater than 0
## 95 percent confidence interval:
##  6.528827      Inf
## sample estimates:
## mean of x
##  9.142591
```

```
reg_line2 <- coef(lm(Reaction ~ Days, data = sleepstudy))

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

```r
ggplot(sleepstudy, aes(x = Days, y = noise_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_line2[1], slope = reg_line2[2]) +
  ggtitle("Reaction time by days w/out sleep") +
  labs(x = "Days", y = "Reaction time /ms") +
  theme_minimal()
```




With the 10% white noise, the coefficients change slightly (see the table) and they are still significant.