Homework06

Markus Heiervang

19/10/2021

Homework (in book: 4.10.1)

In this exercise, you will perform the analysis corresponding to Figure 4.1 on page 70.

1. Load in the data set `ELP_frequency.csv` into your R session.

```
df = read.csv("ELP_frequency.csv")
df
```

```
##        Word  Freq     RT
## 1     thing 55522 621.77
## 2      life 40629 519.56
## 3      door 14895 507.38
## 4     angel  3992 636.56
## 5      beer  3850 587.18
## 6  disgrace   409 705.00
## 7    kitten   241 611.26
## 8     bloke   238 794.35
## 9     mocha    66 725.04
## 10    gnome    32 809.87
## 11 nihilism     4 763.50
## 12 puffball     4 877.53
```

2. Use `mutate()` to apply the `log10()` function to the frequency column (`Freq`) (logarithms will be explained in Chapter 5).

```
df = mutate(df, Freq = log10(Freq))
df
```

```
##        Word     Freq     RT
## 1     thing 4.744465 621.77
## 2      life 4.608836 519.56
## 3      door 4.173041 507.38
## 4     angel 3.601191 636.56
## 5      beer 3.585461 587.18
## 6  disgrace 2.611723 705.00
## 7    kitten 2.382017 611.26
## 8     bloke 2.376577 794.35
## 9     mocha 1.819544 725.04
## 10    gnome 1.505150 809.87
## 11 nihilism 0.602060 763.50
## 12 puffball 0.602060 877.53
```

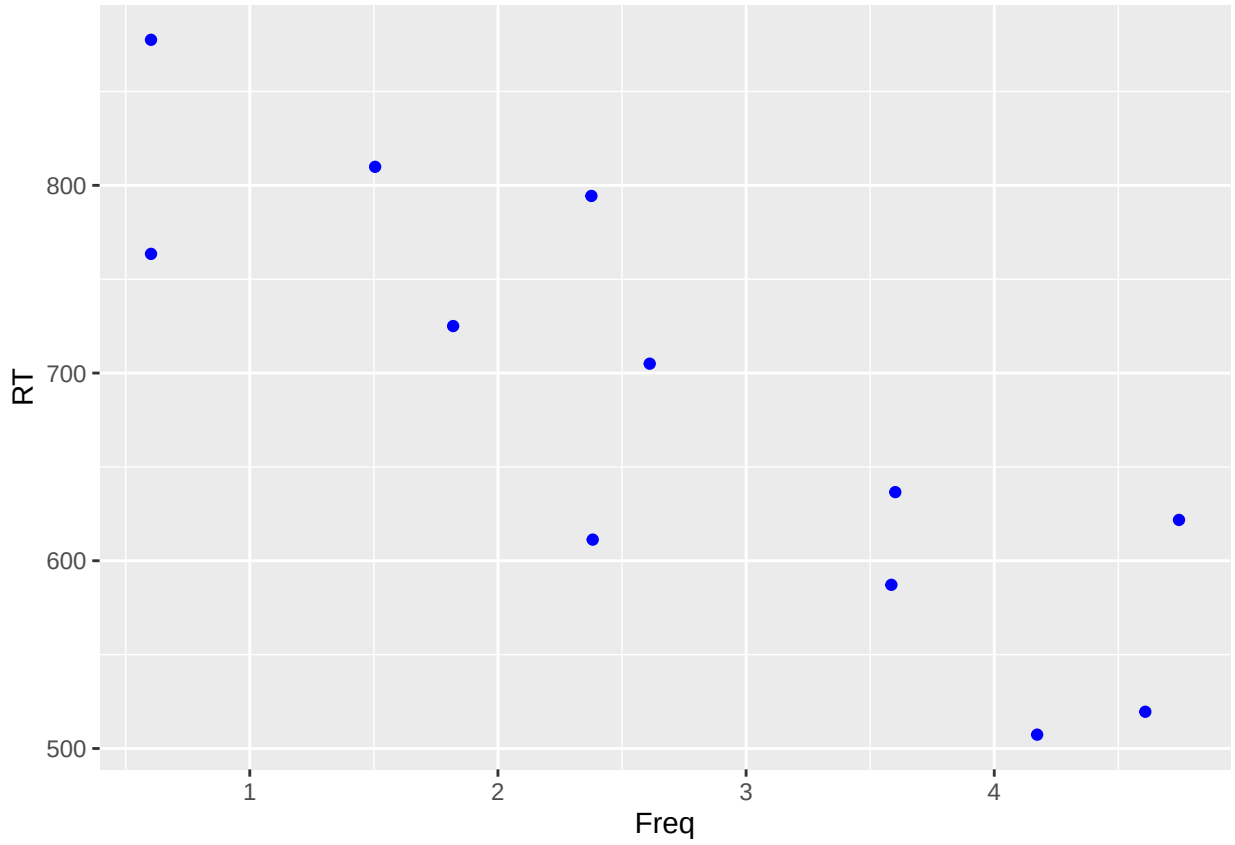
3. Fit a linear model in which response time (`RT`) are modeled as a function of log frequencies.

```r
# The log10 function is already applied in-place on the dataframe
# So we don't need to apply it again
model = lm(df$RT ~ df$Freq, data = df)
summary(model)
```

```
##
## Call:
## lm(formula = df$RT ~ df$Freq, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -92.246 -40.088  -0.179  45.790  90.462
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   870.91      40.42  21.548 1.03e-09 ***
## df$Freq       -70.28      13.26  -5.299 0.000348 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 63.34 on 10 degrees of freedom
## Multiple R-squared:  0.7374, Adjusted R-squared:  0.7111
## F-statistic: 28.08 on 1 and 10 DF,  p-value: 0.0003482
```

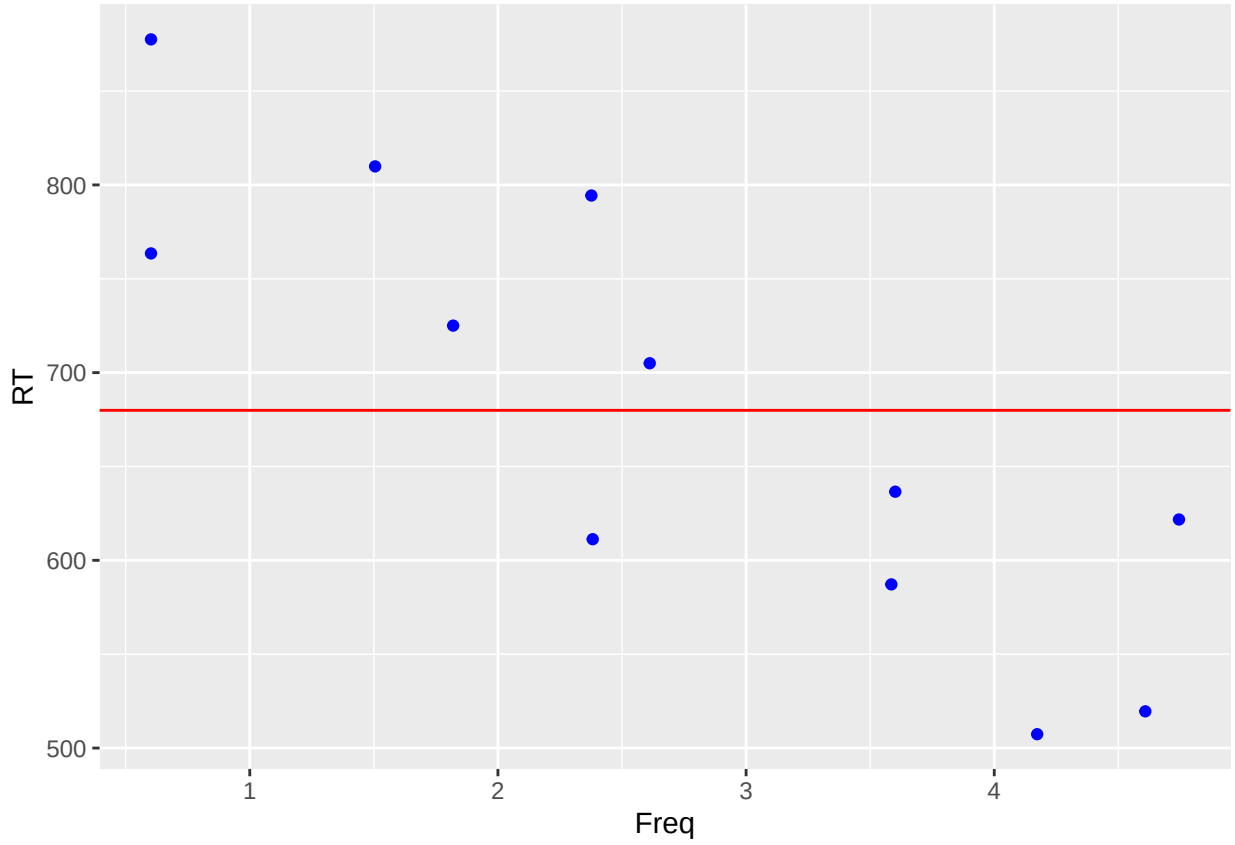
4. Create a plot for the relationship between these two variables.

```r
ggplot(df, aes(x = Freq, y = RT)) + geom_point(color = "blue")
```



5. Can you add a horizontal line showing the mean response duration using `geom_hline()` and the `yintercept` aesthetic?

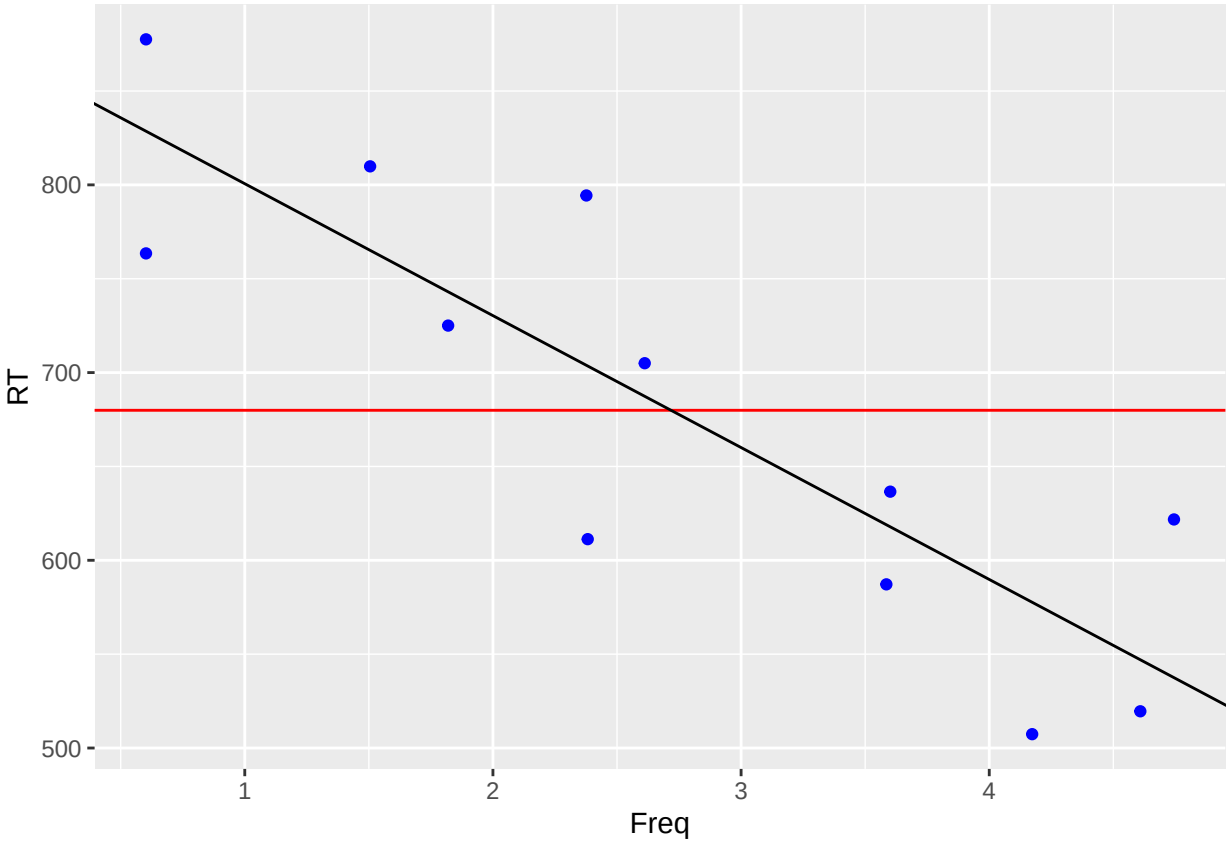
```
ggplot(df, aes(x = Freq, y = RT)) +
  geom_point(color = "blue") +
  geom_hline(yintercept = mean(df$RT), color = "red")
```



6. Can you add the regression line from your model?

```
coefs = coefficients(model)
intercept = coefs[["(Intercept)"]]
slope = coefs[["df$Freq"]]

ggplot(df, aes(x = Freq, y = RT)) +
  geom_point(color = "blue") +
  geom_hline(yintercept = mean(df$RT), color = "red") +
  geom_abline(intercept = intercept, slope = slope)
```



(7. Upload your Rmarkdown file (`.Rmd`) and a knitted document (`.html` or `.pdf`) to Canvas. (deadline: 19.10.2021, 23.59h CET))