Report Template coursework assignment A - 2018

CS4125 Seminar Research Methodology for Data Science

Þórunn Arna Ómardóttir (4917499), Nathan Buskulic (4947916), Mitchell Deen(4396340)

4/3/2019

Contents

1 Part 1 - Design and set-up of true experiment

1.1 The motivation for the planned research.

(Max 250 words) The coffee is today the most consumed drink in the world and it is told to increase your performance and concentration. We want to challenge this idea and verify scientifically if this is a valid idea. We want to test how coffee consumption (and the level of caffeine inside) affect the result of an IQ test. We are most interesting and seeing what the affect is on TU Delft students like ourselves. So the participants will be recruited from the TU Delft student body. # Add that we are doing that on tudelft student

1.2 The theory underlying the research.

(Max 250 words) Preferable based on theories reported in literature There is a large body of literature available on the effects of caffeine on the performance in cognitive tasks. Literature generally supports the idea that coffee improves this performance, see e.g. (Jarvis, 1993; Nehlig, 2010; Rogers et al., 2008). In a brief survey of the relevant literature we did not find any studies specifically addressing students. We would like to investigate this part of the population in more detail.

Jarvis, M. J. (1993). Does caffeine intake enhance absolute levels of cognitive performance?. Psychopharmacology, 110(1-2), 45-52. Rogers, P. J., Smith, J. E., Heatherley, S. V., & Pleydell-Pearce, C. W. (2008). Time for tea: mood, blood pressure and cognitive performance effects of caffeine and theanine administered alone and together. Psychopharmacology, 195(4), 569. Nehlig, A. (2010). Is caffeine a cognitive enhancer?. Journal of Alzheimer's Disease, 20(s1), S85-S94.

1.3 Research questions

The research question that will be examined in the experiment (or alternatively the hypothesis that will be tested in the experiment) Does coffee consumption increases IQ test score #Should we add the caffeine level ?

1.4 The related conceptual model

This model should include: *Independent variable(s) -> Coffee consumption* Dependent variable -> score at IQ test. *Mediating variable (at least 1) -> sleepingness feeling* Moderating variable (at least 1) -> amount of caffeine, prior coffee consumption habit/caffeine tolerance

1.5 Experimental Design

Note that the study should have a true experimental design The experiment is a two groups, post test only, randomized controlled trail.

1.6 Experimental procedure

Describe how the experiment will be executed step by step The participants will be separated into two groups randomly. One group will do the IQ test without any prior coffee consumption while the second group will do the test half an hour after coffee consumption. In the coffe consumption groups, participants will be separated in three subgroups where they will get coffe with different caffeine level. This will allow us to measure the general impact of drinking coffee on an IQ test but it will also allow us to test the difference between each caffeine level.

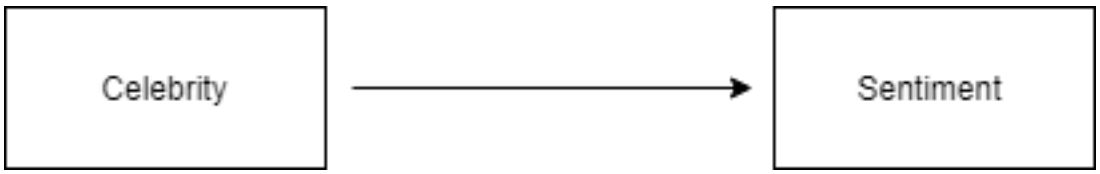


Figure 1: Conceptual model

1.7 Measures

Describe the measure that will be used The Coffe consumption will be measured in ml. The perfomance in an IQ test will in a simple integer number on the scale from 0-200 where the mean is around 100. Sleepingness will be given by the participants on the scale from 0-10 where 10 means the highest level of sleepingness. The amount of caffeine will be measured in mg. Prior coffeedrinking habits will be given by participants. They will be asked how much coffee they typically drink on a normal day.

1.8 Participants

Describe which participants will recruit in the study and how they will be recruited

Since we are just going to make this experiment on the effects of coffee consumption on students at TU Delft we need to find participants from that group of people. Emails will be sent out to the student body explaining the theory of the experiments and willing volunteers asked to fill in a form. We will try to contact an external company of some sort to get some credit or coupons that we can give to participants as a reward for helping out.

1.9 Suggested statistical analyses

Describe the statistical test you suggest to care out on the collected data We will use a one way Analysis of Variance (ANOVA) test between group. Indeed, since the IQ test is follows a gaussian distribution, we just want to compare the mean of each group.

2 Part 2 - Generalized linear models

2.1 Question 1 Twitter sentiment analysis (Between groups - single factor)

2.1.1 Collecting tweets, and data preparation

We collected Tweets for the Three celebrities Beyonce, Madonna and Kanye West. The code can be found in the markdown file.

2.1.2 Conceptual model

Make a conceptual model for the following research question: Is there a difference in the sentiment of the tweets related to the different celebrities?

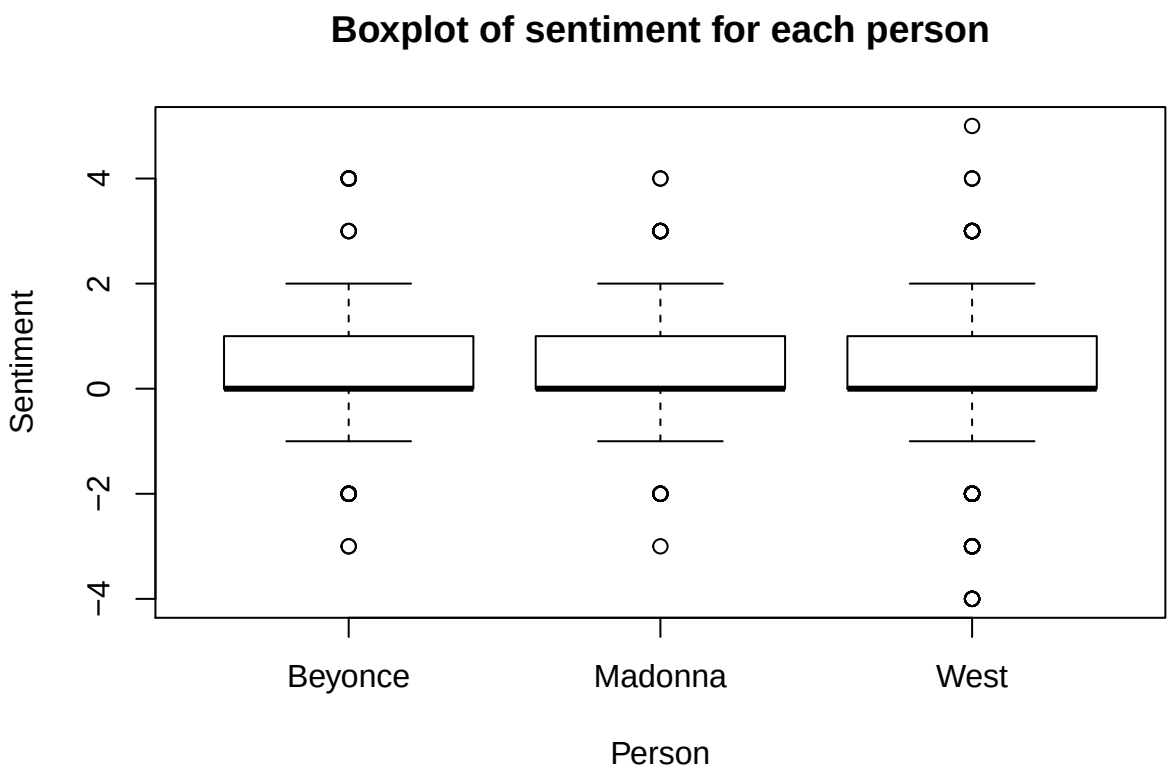
We can see that the sentiment of tweets related to different celebrity is directly connected to the celebrity itself. Therefor the conceptual model is very simple consisting of two variables, “Celebrity” and “Sentiment”.

2.1.3 Homogeneity of variance analysis

Analyze the homogeneity of variance of sentiments of the tweets of the different celebrities

Lets start by looking at how the boxplot looks for each person and the relevant sentiment that has been analysed.

```r
#this was not here in the intermediate report.
#include your code and output in the document
boxplot(score ~ Person, data=semFrame, main="Boxplot of sentiment for each person",
        xlab="Person", ylab="Sentiment")
```




```r
leveneTest( semFrame$score,semFrame$Person, center = median)
```

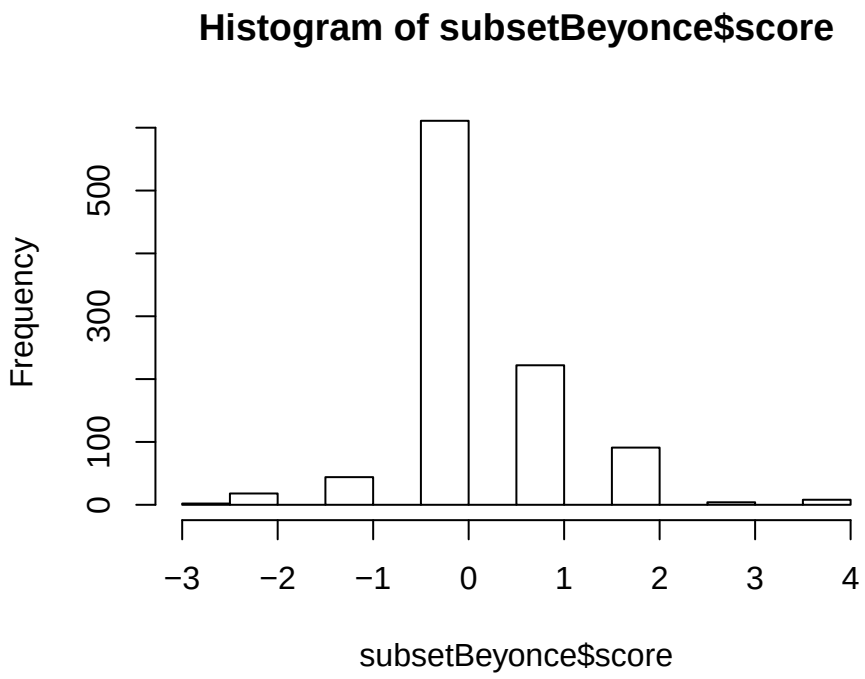
```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value    Pr(>F)
## group    2  9.7131 6.241e-05 ***
##       2997
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Levene test results in a very low p-value., therefore the hypothesis of equal variances is rejected and it is concluded that there is a difference between the variances in the population. Therefore the variance is not considered to be homogeneous.

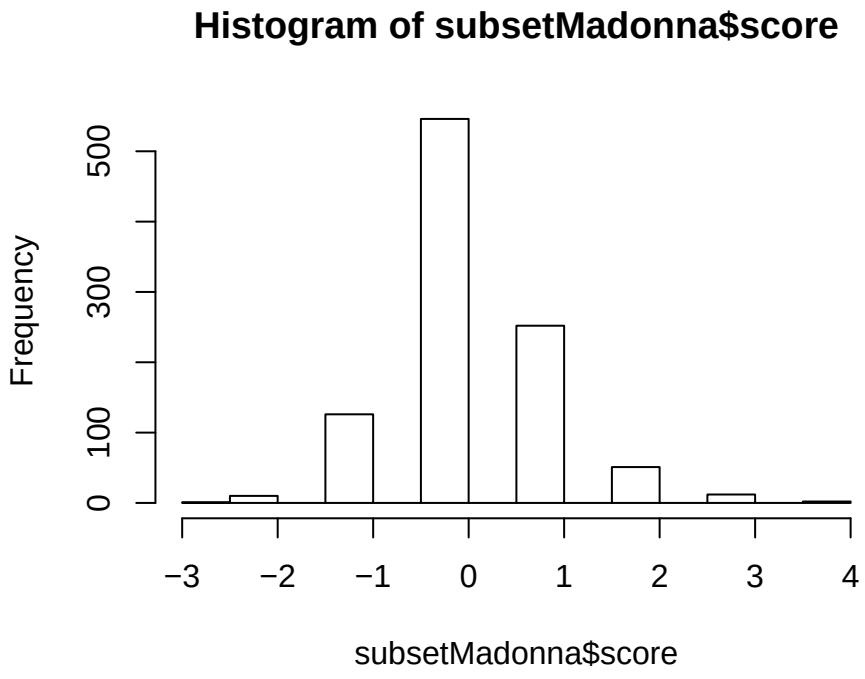
2.1.4 Visual inspection

Graphically examine the variation in tweets' sentiments for each celebrity (e.g. histogram, density plot)

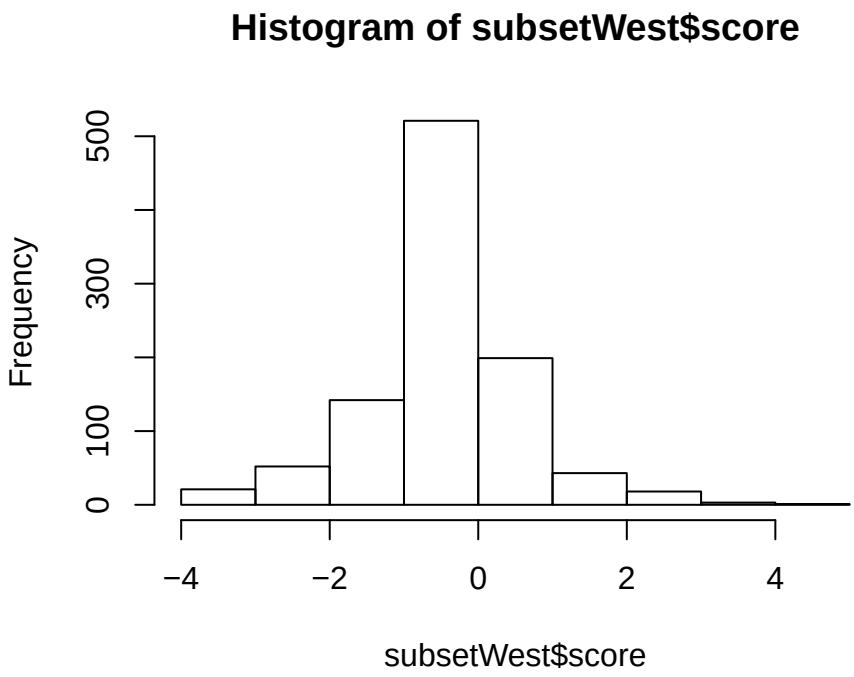
```r
#include your code and output in the document

hist(subsetBeyonce$score)
```




```
hist(subsetMadonna$score)
```




```
hist(subsetWest$score)
```

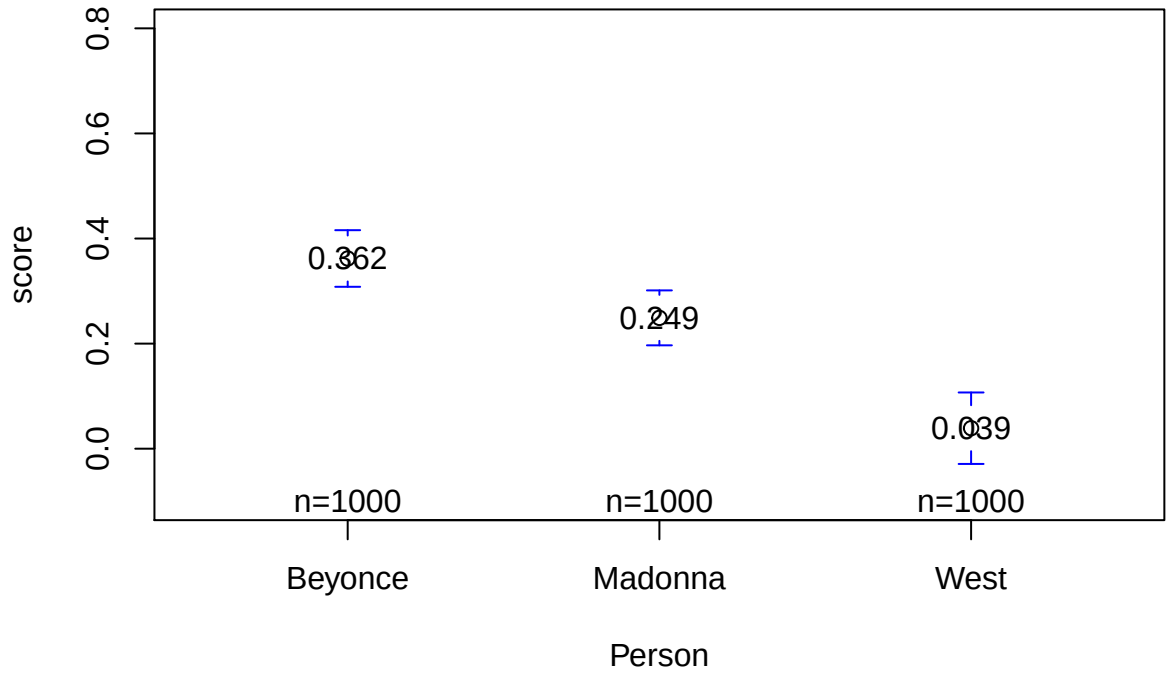



```
meanBeyonce<-mean(subsetBeyonce$score)
meanMadonna<-mean(subsetMadonna$score)
meanWest<-mean(subsetWest$score)
```

2.1.5 Mean sentiments

Here below we plot the means of each class using plotmeans from the package gplots. We can see that the mean for Beyonce is 0.362, for Madonna is 0.249 and for Kanye West it is 0.039. Where a lower value means that the sentiment analysis is more negative.

```
plotmeans(score ~ Person, data = semFrame,mean.labels = TRUE,connect = FALSE,ylim = c(-0.1,0.8))
```



2.1.6 Linear model

```
#include your code and output in the document
 model0<- lm(score ~ 1, data = semFrame) #model without predictor
 model1<- lm(score ~ Person, data = semFrame) #model with predictor
 AnovaResults <-anova(model0,model1)
```

The calculated f-value, F(2,2997) is 30.2537658 and the p- value is $9.8136898 \times 10^{-14}$. since the p- value is so small we can assume that the sentiment of tweets is significantly different depending on what celebrity is mentioned in the tweet.

2.1.7 Post Hoc analysis

If a model that includes the celebrity is better in explaining the sentiments of tweets than a model without such predictor, conduct a post-hoc analysis with e.g. Bonferroni correction, to examine which of celebrity tweets differ from the other celebrity tweets

```
#include your code and output in the document
BonferroniResults <- pairwise.t.test(semFrame$score, semFrame$Person, paired = FALSE, p.adjust.method =
BonferroniP <- BonferroniResults$p.value
BonferroniP
```

```
##              Beyonce      Madonna
## Madonna 2.212015e-02           NA
## West    7.235678e-14 1.984914e-06
```

We chose to use the Bonferroni correction to conduct this post-hoc analysis. There the p-values are multiplied bu the number of comparisons.

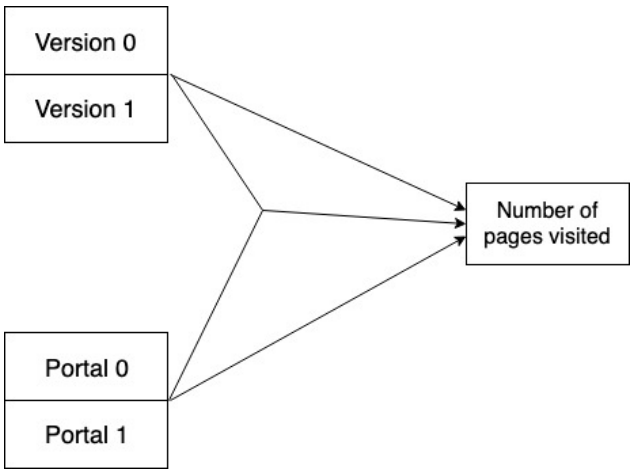


Figure 2: Model for the webvisit question

According to results here above: Madonna vs West has significance that is adjusted to bonferroni of 1.9849142×10^{-6},Madonna vs Beyonce has significance that is adjusted to bonferroni of 0.0221202 and Beyonce vs West has significance that is adjusted to bonferroni of $7.2356777 \times 10^{-14}$. We can see that the bonferroni adjustment leads to all te p-values to be higher than the p-value obtained in the Anova test above. Since the value for Beyonce and Madonna is rather close to 1 it tells us that there really is not much difference on the sentiment for the two. But th other two pairs still have a significant difference, that it the sentiment is significantly different between the two persons in the pair.

2.1.8 Report section for a scientific publication

According to the analysis we have seen that the sentiment of Tweets is not always significantly correlated to the celebrity.

First we visualized the results a bit by looking at the plot of the means of the tweet sentiment per person. We saw that the means for Madonna and Beyonce are closest to each other.

A linear model was fitted on the number of the sentiment score, comparing the difference when taking the relative celbrity in account and not. We first conducted an Anova test and obtained the p value of $9.8136898 \times 10^{-14}$ which hinted there was significant difference in the sentiment, depending on which celebrity it is for. Then A Post Hoc analysis by the means of Bonferroni was conducted. There we saw that the difference of sentiment is not significant for Beyonce and Madonna(p-value of 0.0221202), but is significant for Kanye West in relation to the other two.(p-values: $7.2356777 \times 10^{-14}$,$1.9849142 \times 10^{-6}$).

2.2 Question 2 - Website visits (between groups - Two factors)

2.2.1 Conceptual model

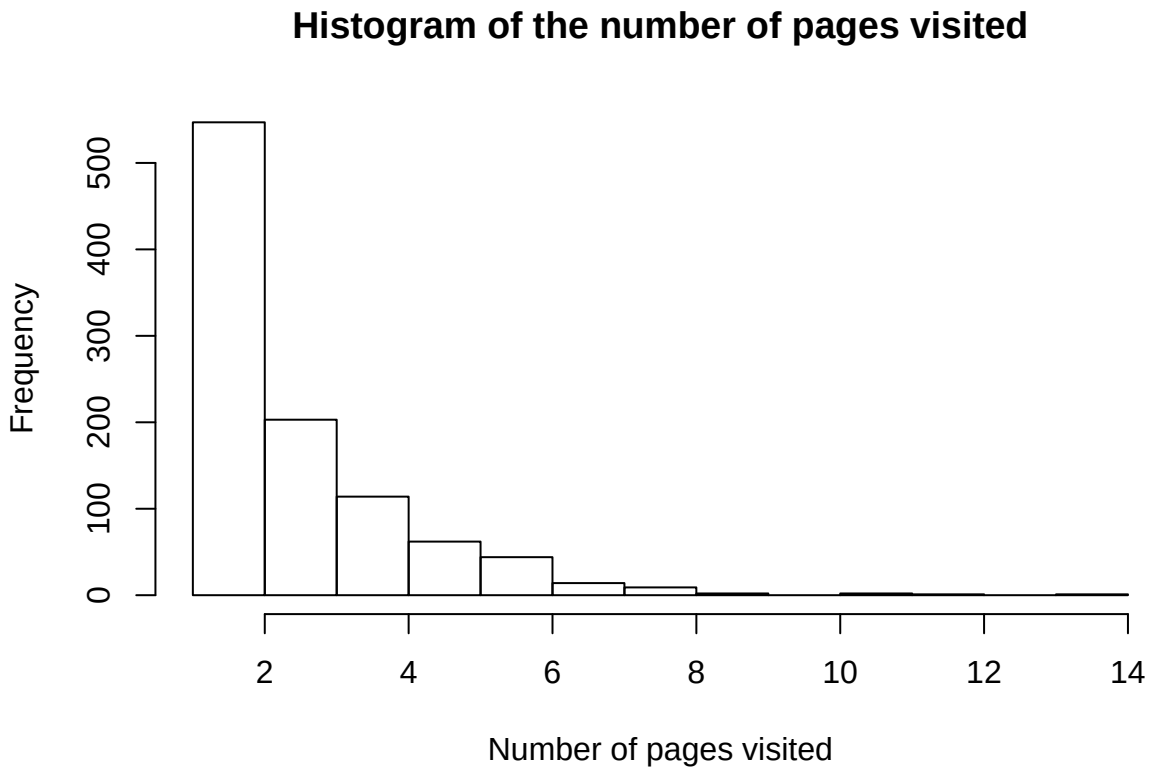
The model can be found in the figure below.

2.2.2 Visual inspection

Graphically examine the variation in page visits for different factors levels (e.g. histogram, density plot etc.)

```
myData <- read.csv("webvisita.csv",header=TRUE)
# We transform into factors what need to be.
myData$user <- factor(myData$user)
myData$version <- factor(myData$version, levels=c(0:1), labels=c("old","new"))
myData$portal <- factor(myData$portal, levels=c(0:1),labels=c("consummer","company"))
```

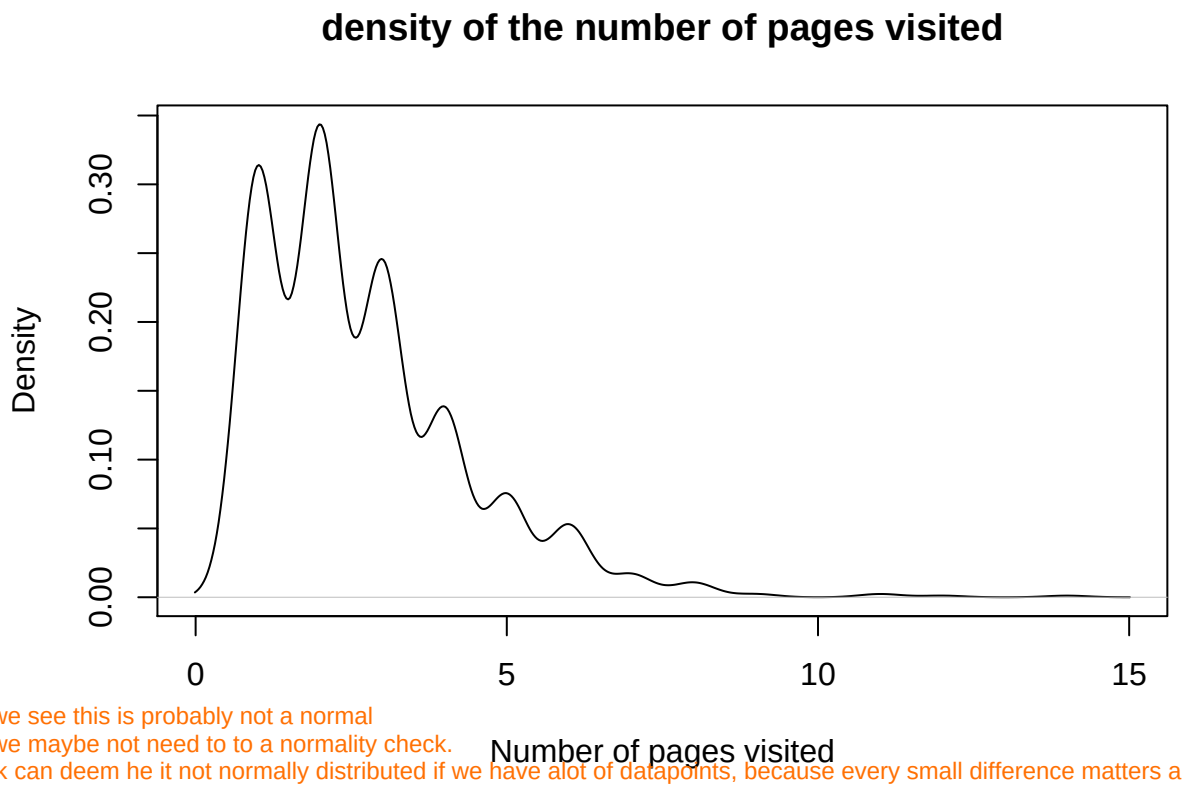
```r
hist(myData$pages, xlab="Number of pages visited", main = "Histogram of the number of pages visited")
```




```r
plot(density(myData$pages), xlab="Number of pages visited", main = "density of the number of pages visi
```

make a graph of the means to see the interaction effects.

make density plot for each subset (4)
have meaningful legends on figures!



looking at this we see this is probably not a normal distibution. so we maybe not need to to a normality check.
Normality check can deem he it not normally distributed if we have alot of datapoints, because every small difference matters a lot then.

2.2.3 Normality check

Statistically test if variable page visits deviates from normal distribution

We can see that the data does not seems to come from a normal distribution, thus we will do a normality test.

```
shapiro.test(myData$pages)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  myData$pages
## W = 0.8436, p-value < 2.2e-16
```

Not normally distributed!
what other distribution is better?

The really small p-value indcates here that there is a high probability that this data do not come from a normal distribution.

2.2.4 Model analysis

linear model cannot be applied on a data that is nor normally distributed!
mentioned on the end of the 4th lecture(genaralized linear model)

Conduct a model analysis, to examine the added values of adding 2 factors and interaction between the factors in the model to predict page visits.

```
library(pander)
# We create all the different models
model0 <- lm(pages ~ 1, data=myData)
model1 <- lm(pages ~ version, data=myData)
model2 <- lm(pages ~ portal, data=myData)
model3 <- lm(pages ~ version + portal, data=myData)
```

we should not be making linear models here!

```
model4 <- lm(pages ~ version + portal + version:portal, data = myData)
```

```
pander(anova(model0,model1,test="F"),caption = "Version as main effect on the number of pages visited")
```

Table 1: Version as main effect on the number of pages visited

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
998	2858	NA	NA	NA	NA
997	2838	1	19.9	6.99	0.008324

```
pander(anova(model0,model2,test="F"),caption = "Portal as main effect on the number of pages visited")
```

Table 2: Portal as main effect on the number of pages visited

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
998	2858	NA	NA	NA	NA
997	2695	1	162.9	60.28	2.033e-14

```
pander(anova(model3,model4,test="F"),caption = "Interaction effect on top of the two main effect")
```

Table 3: Interaction effect on top of the two main effect

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
996	2679	NA	NA	NA	NA
995	2593	1	86.47	33.18	1.118e-08

```
pander(anova(model4),caption = "Effect of version, portal and interaction effect on the number of pages
```

Table 4: Effect of version, portal and interaction effect on the number of pages visited

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
version	1	19.9	19.9	7.636	0.005828
portal	1	158.6	158.6	60.86	1.546e-14
version:portal	1	86.47	86.47	33.18	1.118e-08
Residuals	995	2593	2.606	NA	NA

We see a significant two-way interaction effect, we will thus perform a simple effect analysis to better understand this interaction effect.

2.2.5 Simple effect analysis

If the analysis shows a significant two-way interaction effect, conduct a Simple Effect analysis to explore this interaction effect in more detail.It helps first to look at the means in a figure

```
myData$simple <- interaction(myData$version, myData$portal) #merge two factors
```

```r
contrastConsummer <-c(1,-1,0,0) #Only the consummer portal data
contrastCompany <-c(0,0,1,-1) #Only the company portal data

SimpleEff <- cbind(contrastConsummer,contrastCompany)
contrasts(myData$simple) <- SimpleEff #now we link the two contrasts with the factor simple
pander(simpleEffectModel <-lm(pages ~ simple , data = myData, na.action = na.exclude), caption = "Simple
```

Table 5: Simple effect analysis

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.689	0.05118	52.54	2.964e-289
simplecontrastConsummer	-0.1701	0.07241	-2.349	0.01904
simplecontrastCompany	0.4196	0.07235	5.799	8.94e-09
simple	0.7676	0.1024	7.498	1.428e-13

2.2.6 Report section for a scientific publication

Write a small section for a scientific publication, in which you report the results of the analyses, and explain the conclusions that can be drawn.

A linear model was fitted on the number of pages visited of a website, taking the version of the website (an old version and a new one) and the portal that was used to access the website (a portal for consumers and one for companies) as independent variables, and including a two-way interaction between these variables. The analysis found a significant main effect for the version (F(1,995) = 7.6357, p. < 0.006)and for the portal (F(1,995) = 60.8571, p. < 10^-13). It also found a significant two-way interaction effect (F(1, 995) = 33.1827, p. < 10^-7) between these two variables. ((((However, since a lot of data was used, one can put in doubt the significance result of the version, especially compared to the F values of the portal and the interaction between the variables.)))) The two-way interaction was further examined by a Simple Effect analysis. It found a significant difference for the version in the portal for consumer (t = -2.349, p. = 0.019) as well as in the portal for companies (t = 5.799, p. < 9e-09).

2.3 Question 3 - Linear regression analysis

2.3.1 Conceptual model

For this assignment we retrieved a data set from http://www.stat.ufl.edu/~winner/datasets.html. The dataset contains facts about 153 hybrid cars, including their price, year built, acceleration data and fuel consumption; those are the four quantitative variables that will be the subject of the linear model in this question. We would like to predict the price of the car (response variable), using data on acceleration rate of the car, the fuel consumption and the year that it was built. The conceptual model for this research looks like this:

2.3.2 Visual inspection

The distribution of the independent variable is displayed

```r
# Reading in the necessary packages
library(readr)
d <- read_csv("hybrid_reg.csv")
mpg <- d$mpg
year <- d$year
accel <- d$accelrate
price <- d$msrp

# Histogram of the distribution of the price variable
hist(price)
```

make sure to explain the data well, if it is interval, fixed values or what.

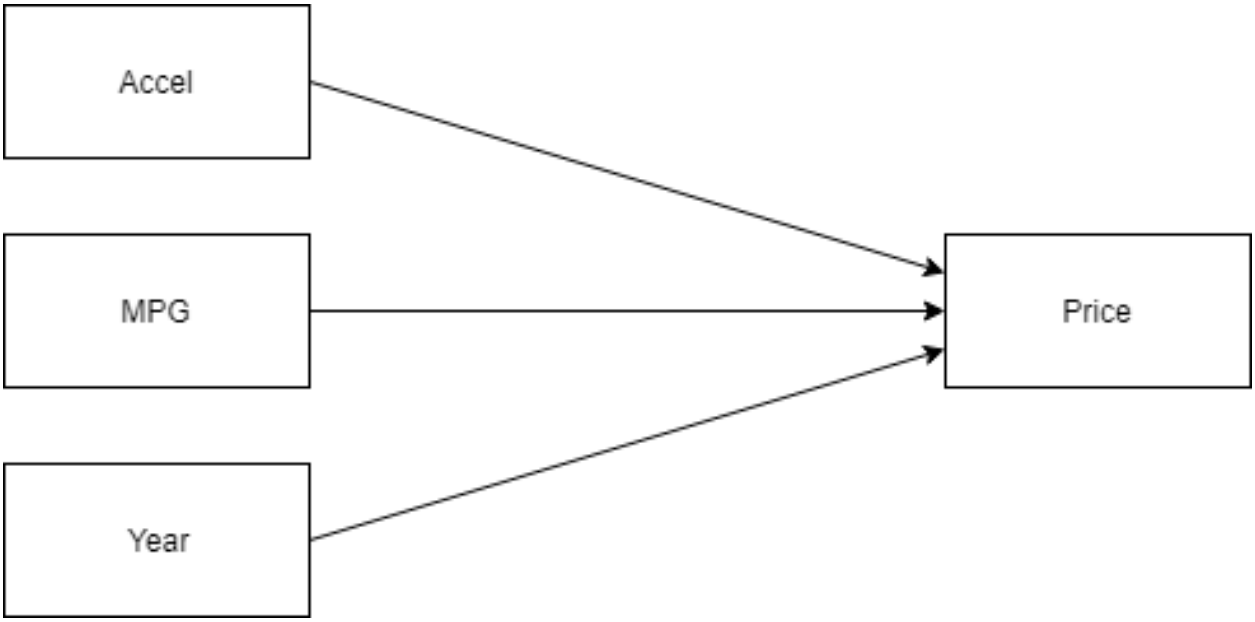
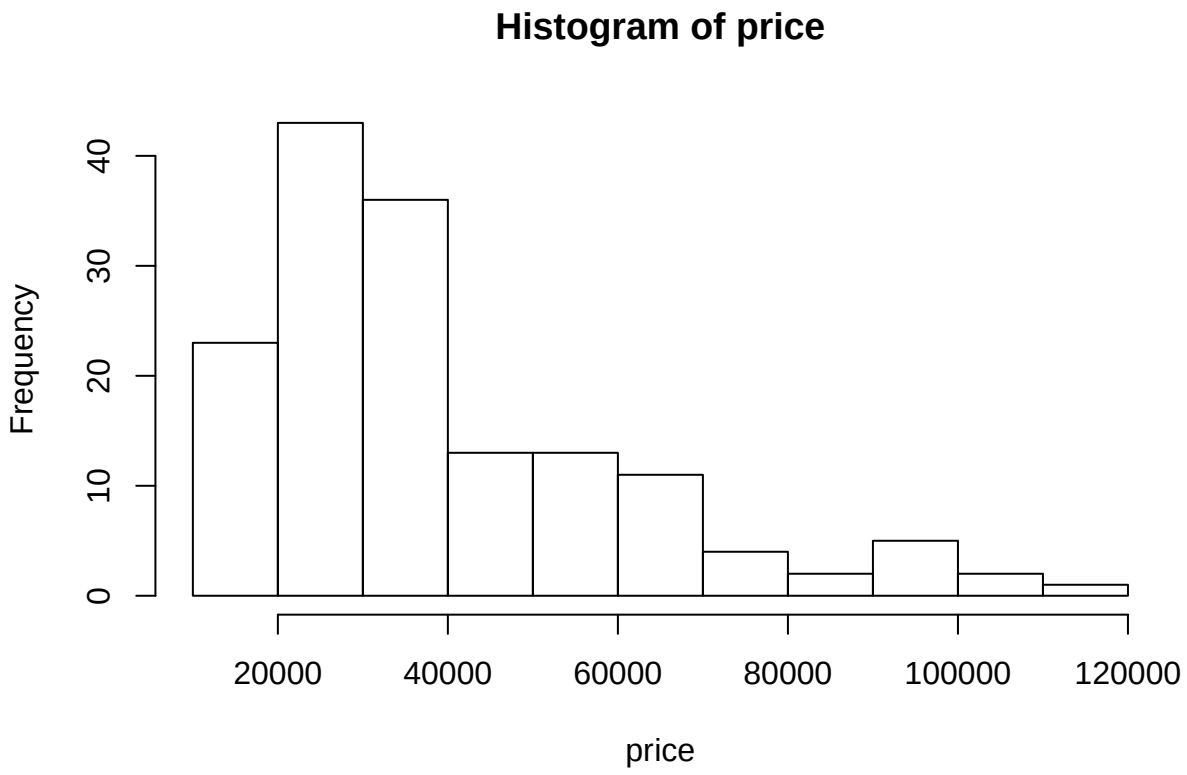
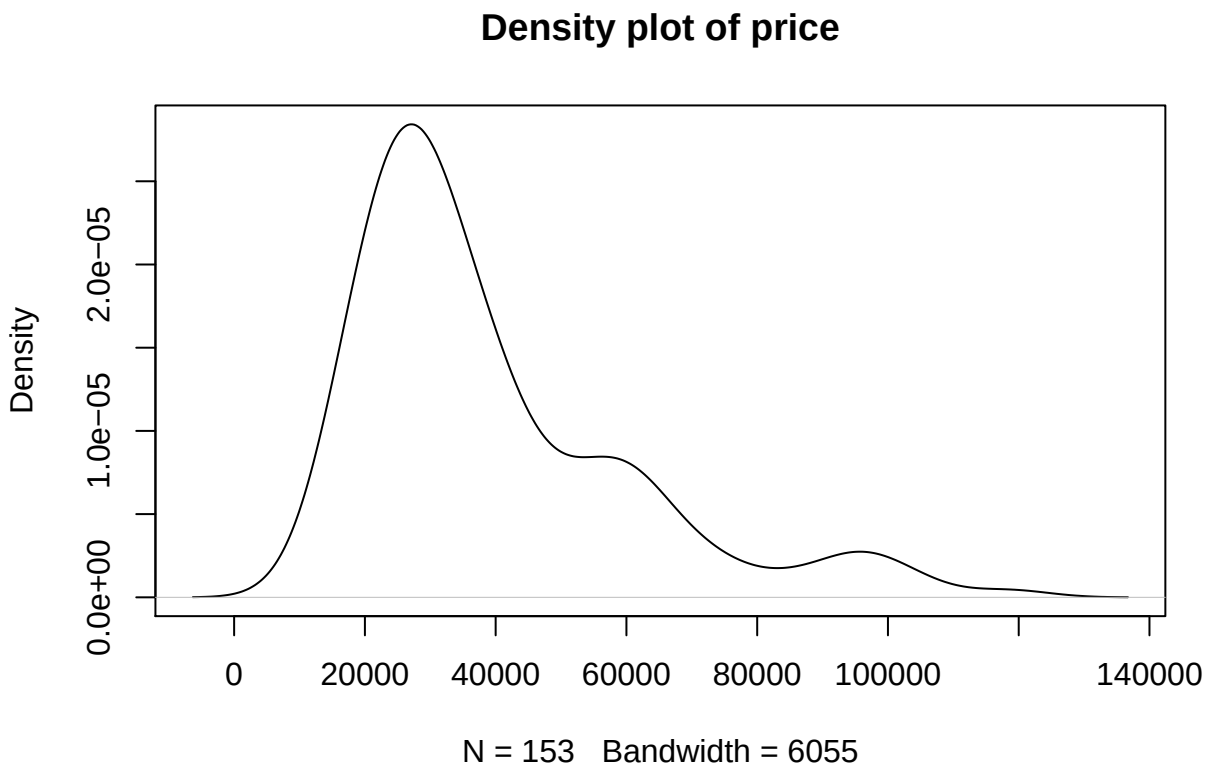


Figure 3: Conceptual model of the four considered variables

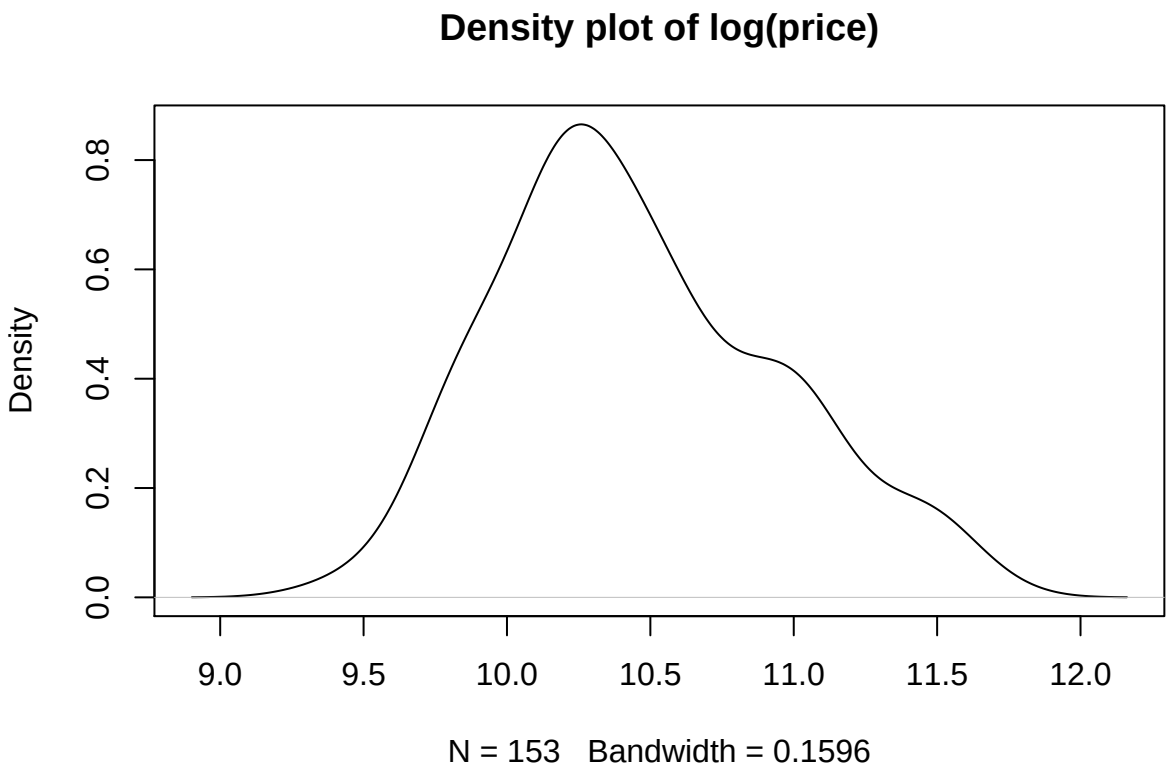



```
# Density plot of the price variable
plot(density(price),main="Density plot of price")
```



Visual inspection of the plots reveals that the distribution of price deviates from a normal distribution. Especially the right tail of the density distribution has more mass than it should have. Since the distribution is right skewed, a logarithmic transformation is effective in increasing the normality; the result can be seen in the figure below.

```
# Density plot of the transformed price variable
plot(density(log(price)),main="Density plot of log(price)")
```




```
shapiro.test(price)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  price
## W = 0.85261, p-value = 4.345e-11
```

```
shapiro.test(log(price))
```

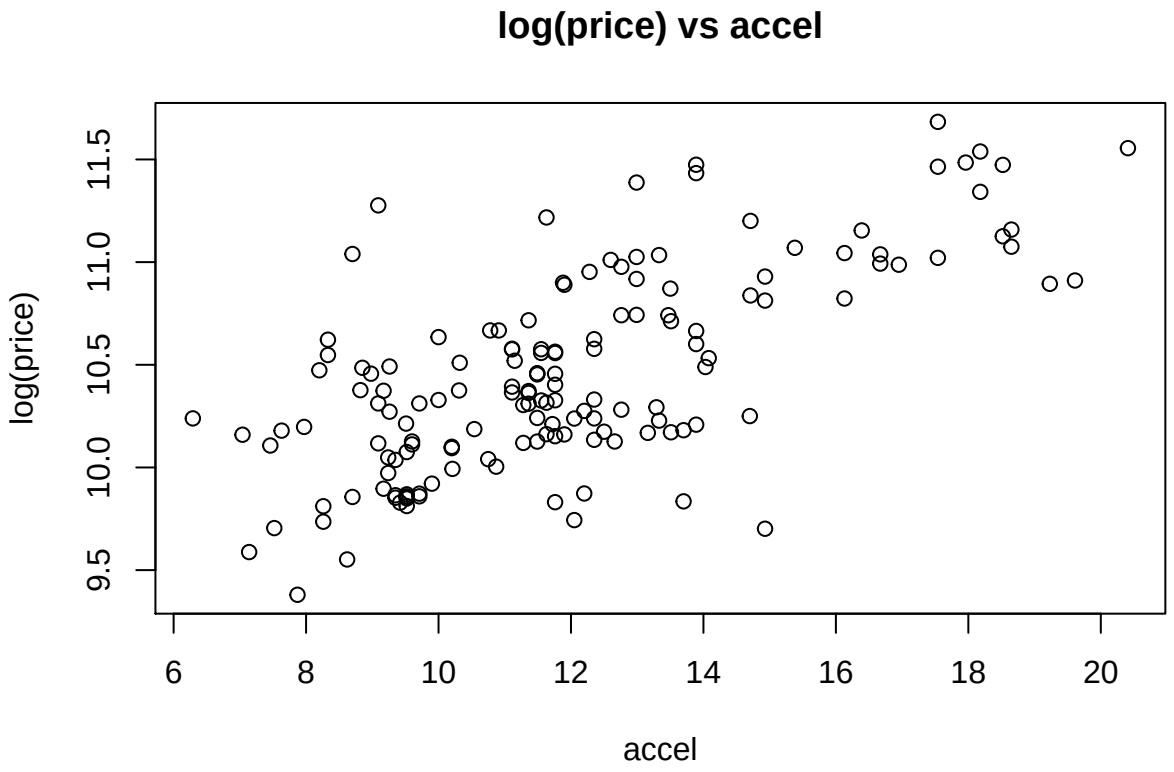
```
##
##  Shapiro-Wilk normality test
##
## data:  log(price)
## W = 0.97322, p-value = 0.004424
```

As we will see later on, the logarithmic transformation is also necessary to justify the choice of linear regression, as without it the assumptions for linear regression do not hold.

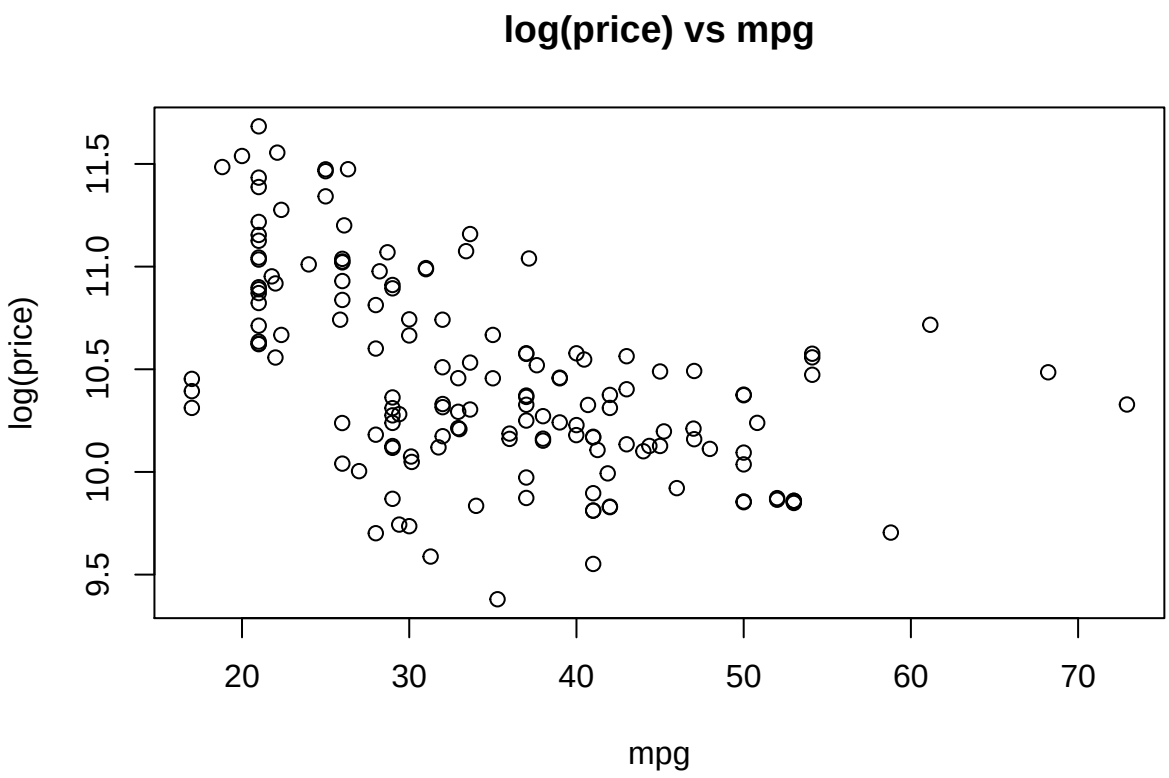
2.3.3 Scatter plot

Using scatter plots we can visually examine the relationship between two variables. The following figures show the scatter plots of the response variable price paired with each of the predictors.

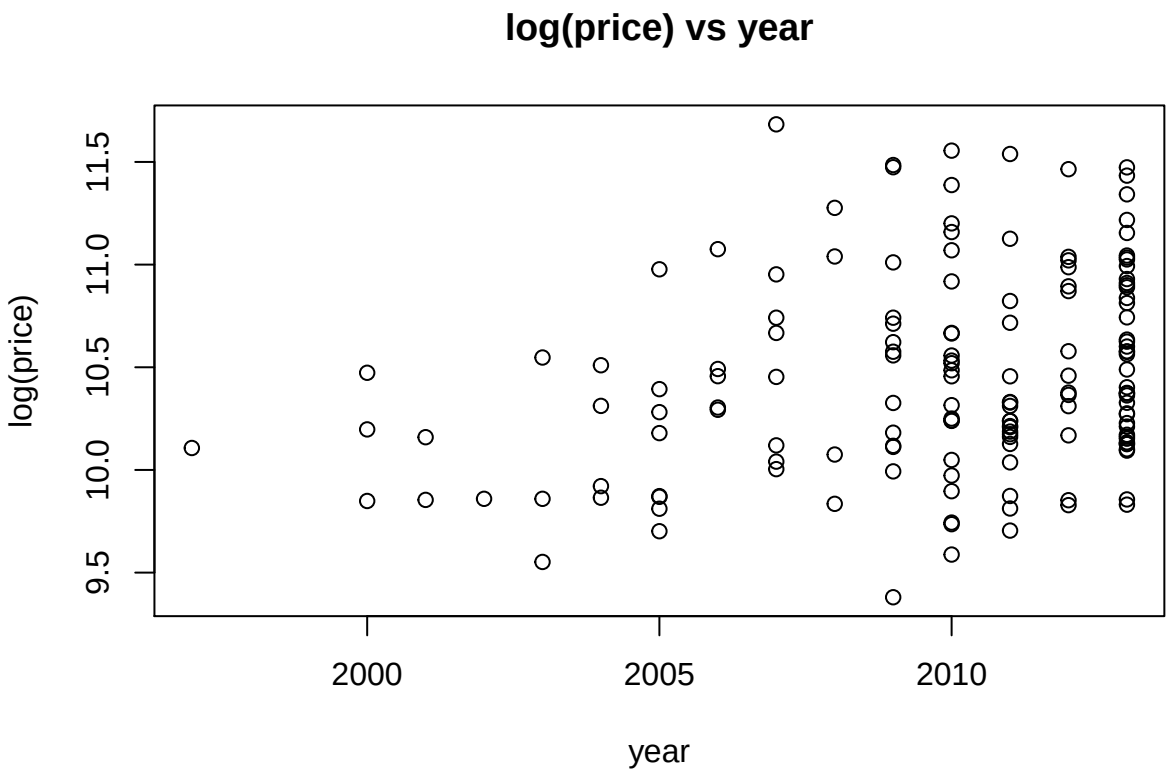
```r
plot(log(price) ~ accel,main="log(price) vs accel")
```




```r
plot(log(price) ~ mpg,main="log(price) vs mpg")
```




```
plot(log(price) ~ year,main="log(price) vs year")
```



2.3.4 Linear regression

From the anova table we can see that the accel and mpg variables are able to significantly improve the model. The year variable is not able to explain any additional significant variance in the price variable. Therefore we exclude the price variable from further analysis.

```
library(pander)
model0 <- lm(log(price) ~ 1)
model1 <- lm(log(price) ~ accel)
model2 <- lm(log(price) ~ accel + mpg)
model3 <- lm(log(price) ~ accel + mpg + year)

pander(anova(model0,model1,model2,model3),
       caption = "Model comparison to predict the price of a car")
```

Table 6: Model comparison to predict the price of a car

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
152	35.77	NA	NA	NA	NA
151	18.83	1	16.94	146.9	5.776e-24
150	17.19	1	1.639	14.22	0.0002343
149	17.18	1	0.01013	0.08785	0.7673

```
library(QuantPsyc)
pander(confint(model2),
```

```
       caption = "#95% confidence interval of the estimates")
```

Table 7: #95% confidence interval of the estimates

	2.5 %	97.5 %
(Intercept)	9.328	10.13
accel	0.07142	0.1142
mpg	-0.0167	-0.00524

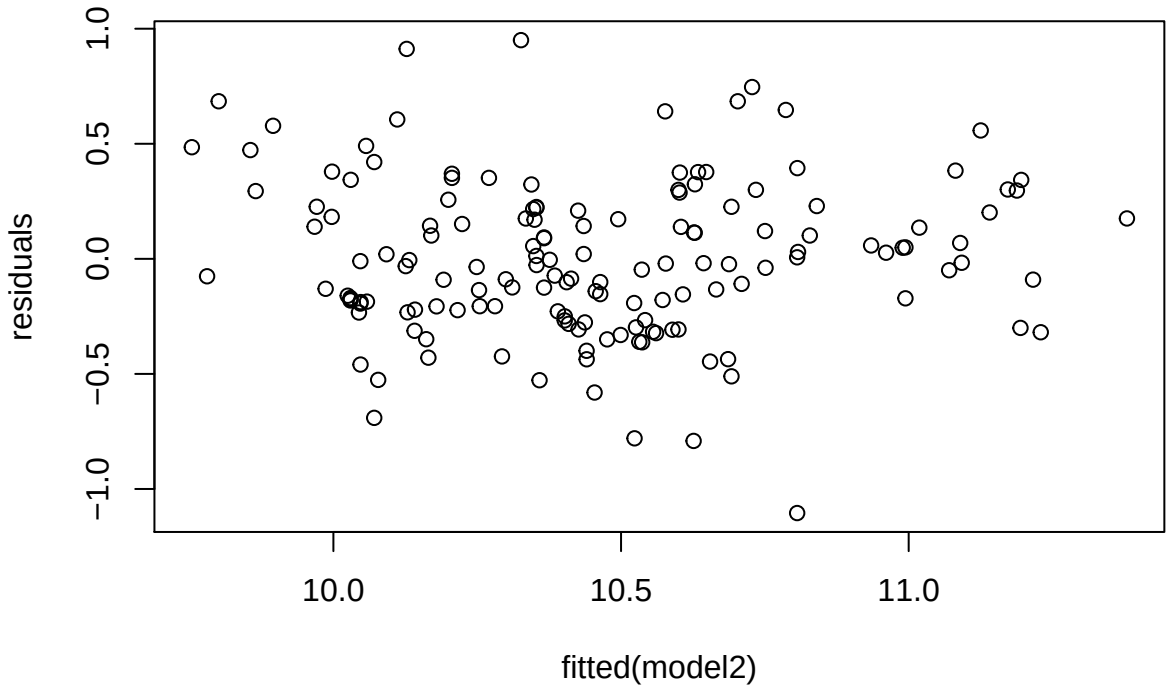
```
pander(lm.beta(model2),
       caption = "standardised regression coefficients") # standardised regression coefficients
```

accel	mpg
0.5626	-0.2482

2.3.5 Examine assumption

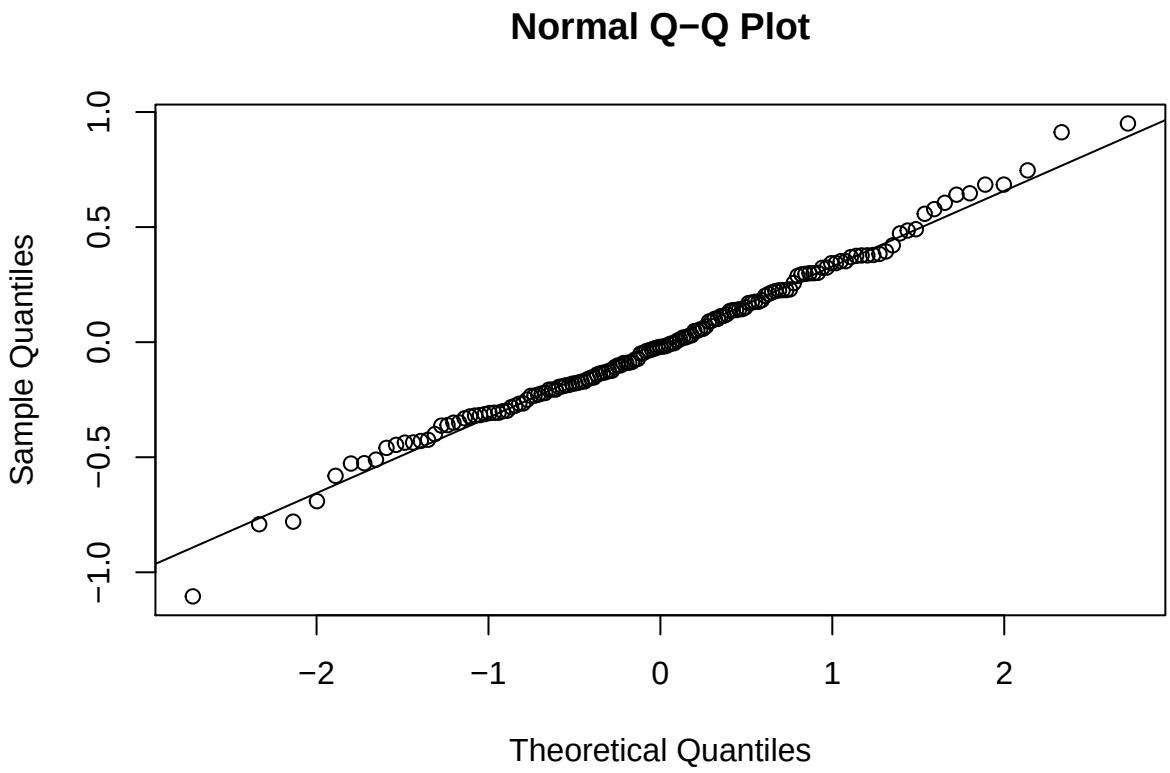
The residuals vs fitted plot is a useful tool in examining the linearity and equal variances assumptions.

```
residuals = resid(model2)
plot(residuals ~ fitted(model2))
```



For normality we can check the qq-plot:

```
library(car)
qqnorm(residuals)
qqline(residuals)
```



Multicollinearity:

```
vif(model2)
```

```
##   accel     mpg
## 1.34428 1.34428
```

```
1/vif(model2) # Tolerance
```

```
##     accel       mpg
## 0.7438928 0.7438928
```

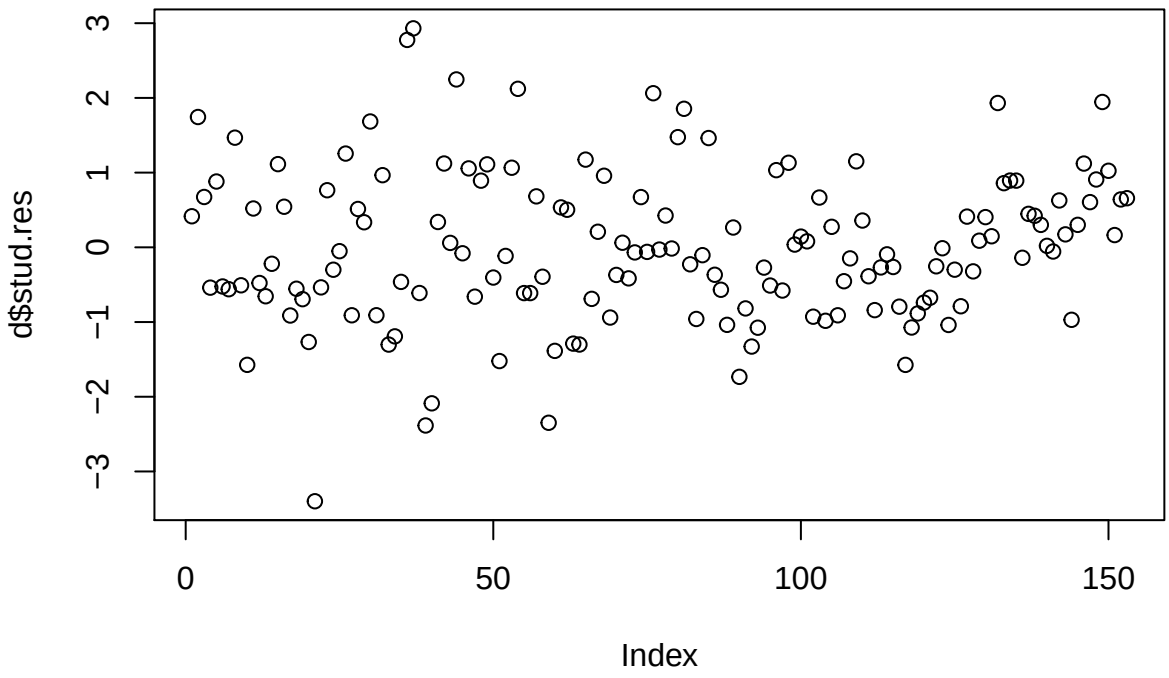
Autocorrelation:

```
durbinWatsonTest(model2)
```

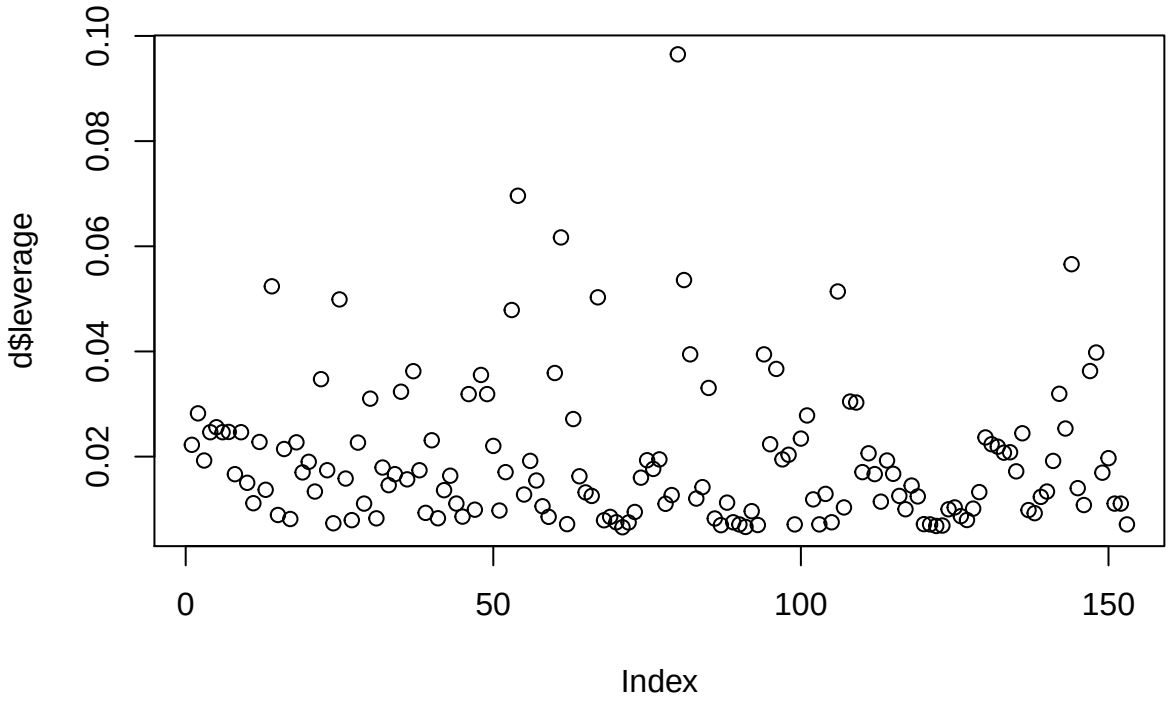
```
##  lag Autocorrelation D-W Statistic p-value
##    1       0.2303107      1.535374   0.004
##  Alternative hypothesis: rho != 0
```

2.3.6 Impact analysis of individual cases

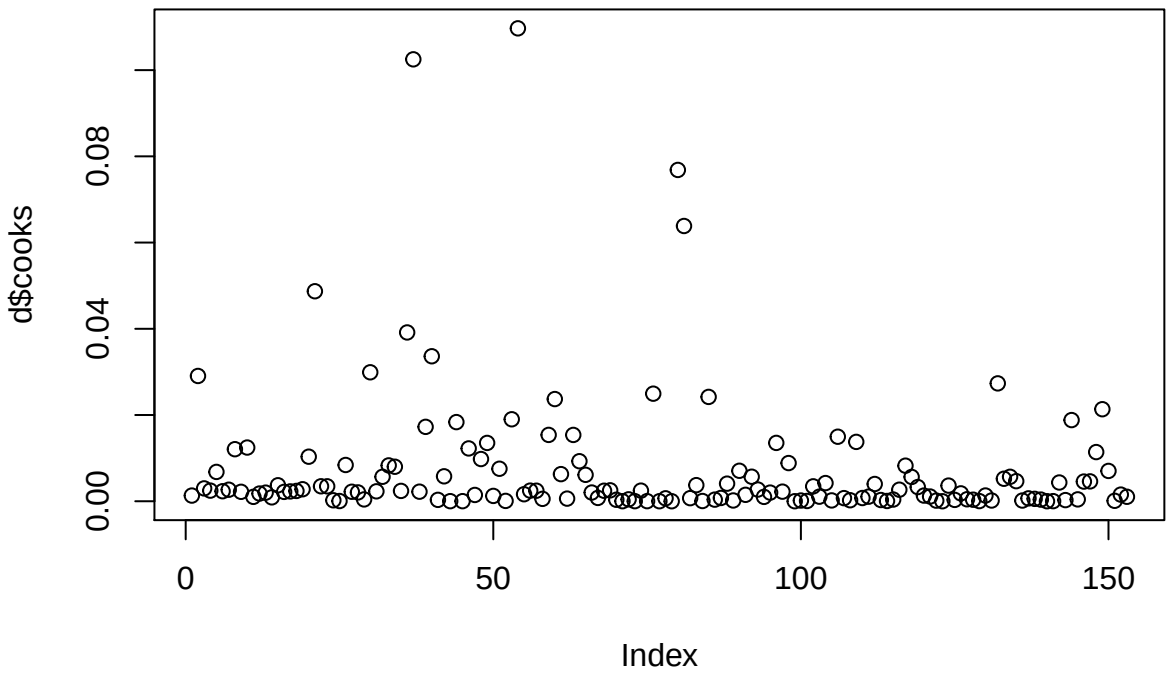
```
d$stud.res<-rstudent(model2)
plot(d$stud.res)
```




```
d$leverage<-hatvalues(model2)
plot(d$leverage)
```




```
d$cooks<-cooks.distance(model2)
plot(d$cooks)
```



2.3.7 Report section for a scientific publication

In this section we briefly present the results of fitting a linear regression model in order to predict the price of cars using data on their acceleration rate, their fuel efficiency in miles per gallon, and their build year.

First of all we can conclude that a logarithmic transformation was necessary to increase normality of the distribution of the price variable. While the distribution still significantly differs from a normal one (W = 0.973, p = 0.004), it is an improvement over the original distribution (W = 0.853, p = 4.3e-11). Furthermore, the transformation was necessary to justify the choice of performing linear regression, as without it the assumptions for linear regression do not hold, especially the linearity assumption.

Inspecting the scatter plots, it becomes clear that a linear relationschip between the natural logarithm of price and the year the car was built is absent. The scatter plots of the other two variables show some indication that a relationship might exist.

Fitting the model revealed that the year variable is indeed not able to explain any additional variance in the price variable on top of the accel and mpg variables (F = 0.088, p = 0.77). Based on this result we decided to exclude the independent variable year from the model. Hence we end up with the following model:

log(price) = 9.73 + 0.093 x accel - 0.011 x mpg

Checking the assumptions of the linear regression, we found that the distribution of the residuals is normal with expected value 0 and (roughly) constant variance. Testing for independence showed a violation of the assumption (D-W = 1.54, p = 0.008). Violating the independence assumption is quite problematic, but for the sake of the exercise we will continue the analysis. Additionally, no multicollinearity could be found in our model.

Analysis of influential and leverage points revealed no severe outlying cases that undermine the linear regression model.

The interpretation of the coefficients is slightly tricky, since we are dealing with a transformed dependent variable. Instead of additive, the model becomes multiplicative, and each coefficients has to be interpreted as an exponent (i.e. the intercept becomes e^9.73 = 16,815). The standardized coefficients tell us that the effect of one higher standard deviation in the acceleration rate has about twice the effect on the price of one higher standard deviation in the fuel efficiency.

To conclude, we were able to formulate a linear model that is to some extent able to predict the price of a car based on its acceleration rate and its fuel efficiency. Since not all assumptions of linear regression were met, interpretation of the results requires caution.

2.4 Question 4 - Logistic regression analysis

```
#include your code and output in the document
Data <- read.csv("port_taiwan.csv",header=TRUE)

Data$year <- factor(Data$year, levels=c(2003:2006), labels=c("year2003","year2004","year2005","year2006

#remove one port so out data can pass as dichotomous
PortData <- subset(Data,(port != "3"))
PortData$port <- factor(PortData$port, levels=c(1:2),labels=c("1","2"))
```

2.4.1 Conceptual model

Make a conceptual model underlying this research question

2.4.2 Logistic regression

Conduct a logistic regression, examine whether adding individual indicators in the model improves the model compared to Null model. Make a final model with only significant predictor(s). For this model, calculate the pseudo R-square. Calculate the odd ratio for the predictors and their confidence interval

```
#include your code and output in the document
```

2.4.3 Crosstable predicted and observed responses

Make a crosstable of the predicted and observed response

```
#include your code and output in the document
```

2.4.4 Report section for a scientific publication

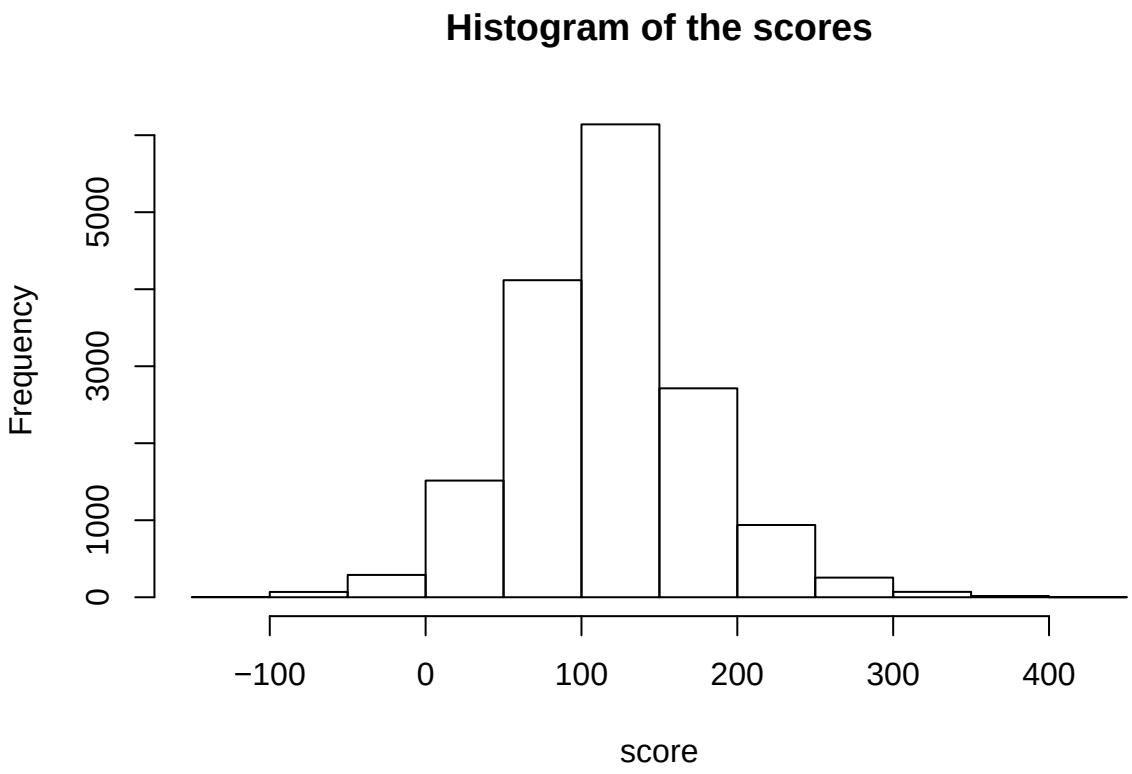
Write a small section for a scientific publication, in which you report the results of the analyses, and explain the conclusions that can be drawn.

3 Part 3 - Multilevel model

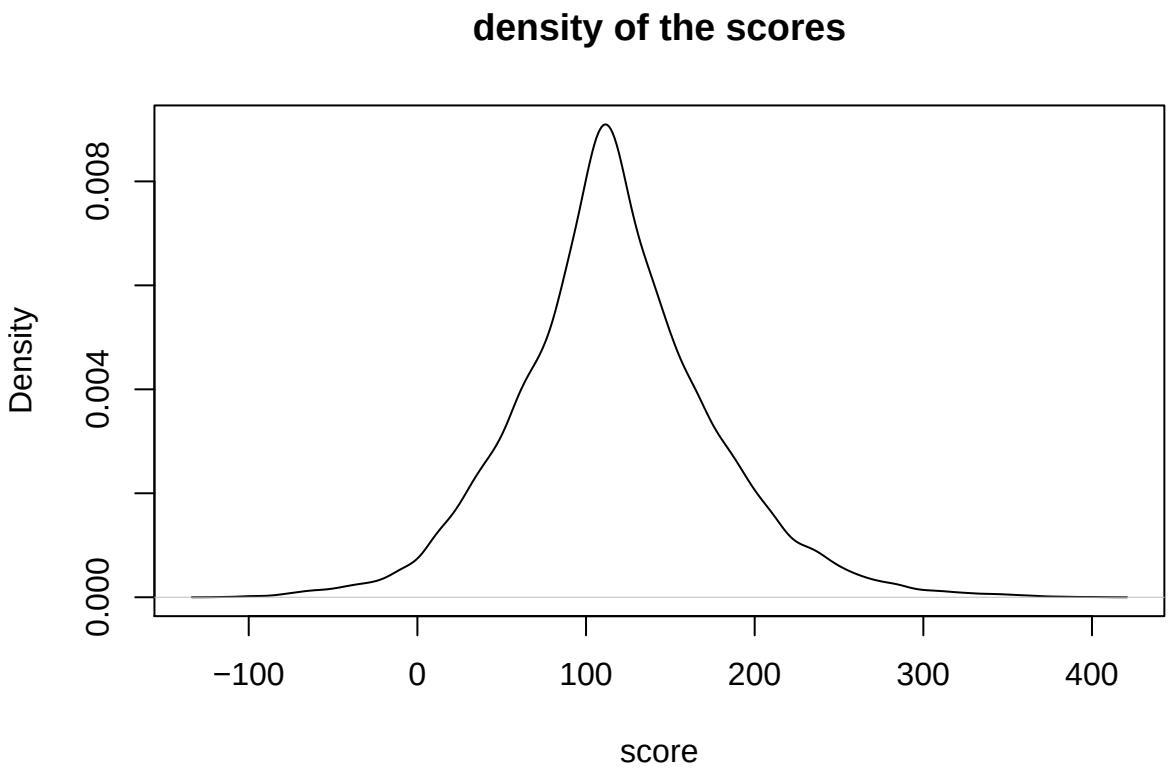
3.1 Visual inspection

Use graphics to inspect the distribution of the score, and relationship between session and score

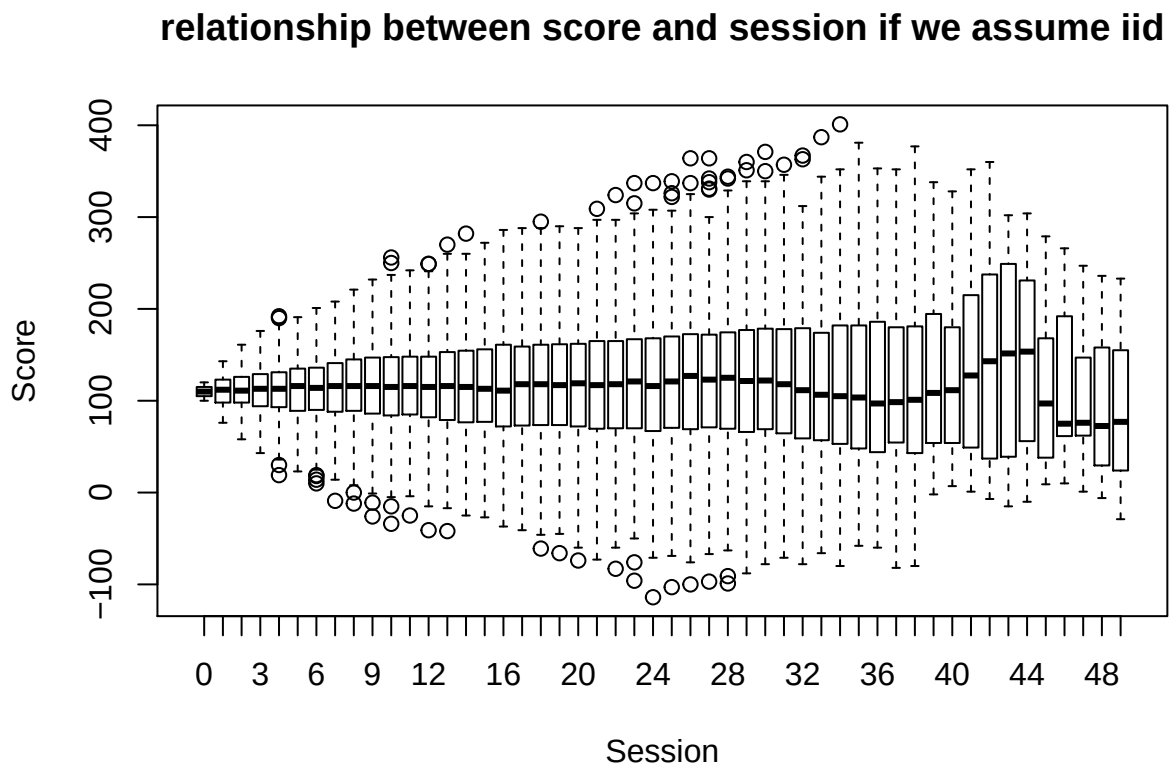
```
set1 <- read.csv("set1.csv", header = TRUE)
set1$Subject <- factor(set1$Subject)
# Should we consider the session as factor ?
#set1$session <- factor(set1$session)

hist(set1$score, xlab="score", main="Histogram of the scores")
```




```
plot(density(set1$score), xlab="score", main="density of the scores")
```




```
#Plot of the relationship between session and score
# Assuming iid of the variables which is not true, shouldn't do that.
boxplot(score ~ session, data=set1, xlab="Session", ylab="Score", main="relationship between score and s
```



3.2 Multilevel analysis

Conduct multilevel analysis and calculate 95% confidence intervals, determine:

- If session has an impact on people score

```
library(nlme)
library(pander)

randomIntercept <- lme(score ~ 1, data = set1, random = ~1|Subject, method="ML")
addSession <- update(randomIntercept, .~. + session)
pander(anova(randomIntercept,addSession), caption="comparisons of models when session is added as a fix
```

Table 9: comparisons of models when session is added as a fixed factor (continued below)

	call	Model	df	AIC
randomIntercept	lme.formula(fixed = score ~ 1, data = set1, random = ~1 | Subject, method = "ML")	1	3	162711
addSession	lme.formula(fixed = score ~ session, data = set1, random = ~1 | Subject, method = "ML")	2	4	162545

	BIC	logLik	Test	L.Ratio	p-value
randomIntercept	162734	-81352		NA	NA
addSession	162576	-81269	1 vs 2	167.7	2.317e-38

The addition of the session significantly improves the model, we will now verify the 95% confidence bound.

```
intervals(addSession)
```

```
## Approximate 95% confidence intervals
##
##  Fixed effects:
##                   lower        est.       upper
## (Intercept) 106.8665678 111.0675622 115.2685566
## session       0.3126229   0.3682005   0.4237781
## attr(,"label")
## [1] "Fixed effects:"
##
##  Random Effects:
##   Level: Subject
##                     lower    est.    upper
## sd((Intercept)) 43.67471 46.5146 49.53914
##
##  Within-group standard error:
##    lower     est.    upper
## 34.68269 35.06933 35.46028
```

We see that for the fixed effect, the session deviates from 0 in the 95% interval.

- If there is significant variance between the participants in their score

```
intervals(randomIntercept)
```

```
## Approximate 95% confidence intervals
##
##  Fixed effects:
##                lower     est.    upper
## (Intercept) 112.7019 116.8139 120.9259
## attr(,"label")
## [1] "Fixed effects:"
##
##  Random Effects:
##   Level: Subject
##                     lower     est.    upper
## sd((Intercept)) 43.68633 46.52747 49.55338
##
##  Within-group standard error:
##    lower     est.    upper
## 34.86891 35.25763 35.65067
```

We can see that in the Random effect, the standard deviation of the intercept do not include 0 in the 95% interval, thus there is a significant variance between the participants in their score

3.3 Report section for a scientific publication

Write a small section for a scientific publication, in which you report the results of the analyses, and explain the conclusions that can be drawn.

From our analysis it appears that there is a significant main effect of the session over participants score (L.ratio = 167.7, p. $< 10\hat{}$-37) when compared to the baseline at a 95% confidence interval (sd(session) = [0.32,0.42]). We can also show that there is a significant variance between participants in their score (sd(intercept) = [43.7,49.6]) at a 95% confidence interval. From these results, we can draw the conclusion that the session in which the participant is has an impact on his score which can be interpreted as improvement over each exercise session. We can also conclude that each participant is different and that indeed, we cannot consider the observations to be independent.