**The effects of lifestyle activities on reaction time**

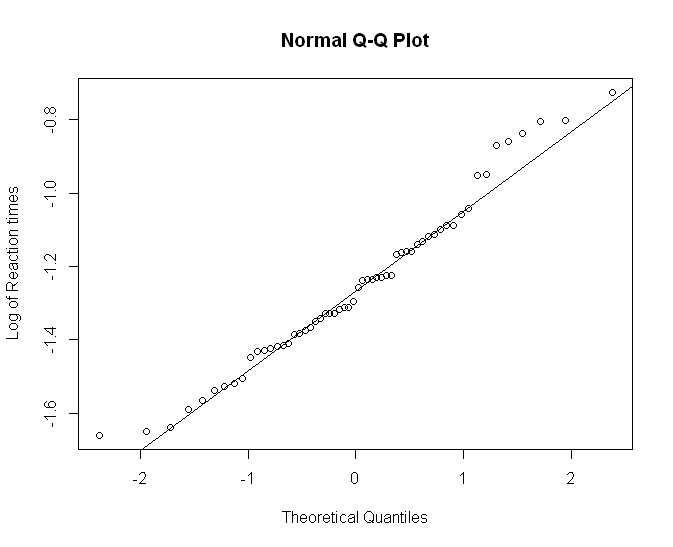
In this experiment, 59 UT Students above the age of 18 and under the age of 28 were surveyed about their lifestyle activities and how long they spend on them, and is tested for their reaction times, measured in milliseconds, through the website humanbenchmark.com. Students were surveyed outside of the Perry-Castañeda Library as they were exiting or leaving, with no particular biases in the selection of volunteers. The same setup was used for everyone, by using a supplied mouse and computer to record data, and questions were asked in regards to activity and hours spent. The reaction time is predicted by certain activities or how long is spent on an activity. The reaction time is the response variable, and the two predicting explanatory variables are time spent on an activity and what activity is being done. Sedentary, gaming, and sport lifestyles are compared among each other as different categories that organize the explanatory variable. The duration spent on an activity is a numerical variable measured by hours in a week, and its effect on reaction time is investigated.

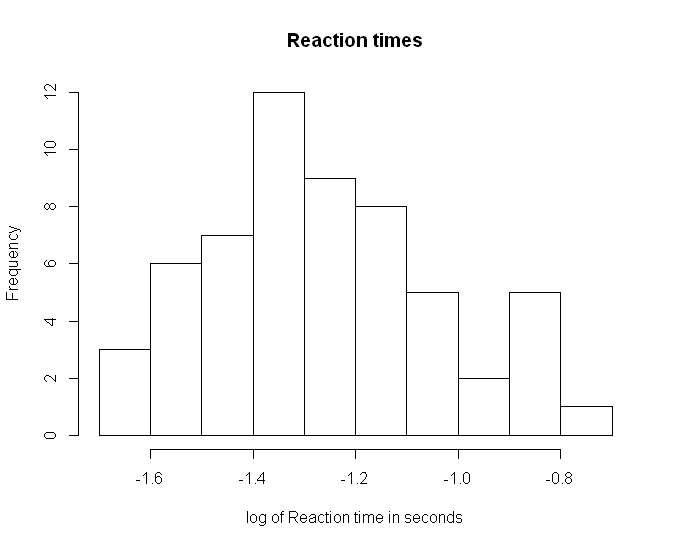
**Descriptive Analysis of the Response Variable**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Minimum | 1st Quartile | Median | 3rd Quartile | Maximum |
| -1.661 | -1.415 | -1.277 | -1.122 | -0.7257 |

After applying a logarithmic transformation to the vector of reaction times, two visual representations of normality can be seen with a QQ plot and a histogram. A few outliers can be seen on the QQ plot, potentially disrupting normality - however, we can formally confirm normality of the log of the reaction times by applying the Shapiro-Wilks test to the log-transformed vector.

W = 0.9737, p-value = 0.2387. As the p-value is greater than 0.05 (the alpha), the log of the data is normal.





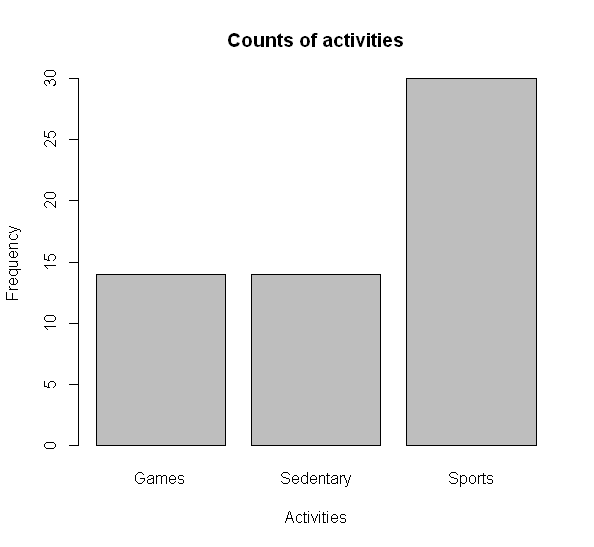
Even with what appears to be outliers in the data set on the right side of both the histogram and QQ plot, with a p-value greater than the alpha calculated by the Shapiro-Wilks test, no removals are necessary.

**Descriptive Analysis of the Two Explanatory Variables**

**Activity:** Nominal categorical variable

|  |  |  |
| --- | --- | --- |
| **Sedentary** | **Video Games** | **Sport** |
| 14 | 14 | 30 |

Survey volunteers were asked what recreational activity they like to engage in, and answers had large variety in describing drawing, basketball, video gaming, and hobbyist programming. These categories were collapsed into three primary groups: Sedentary activities that do not engage in active engagement of motor skills but may require intricate training, thinking, and processing; Video Games, which is a primarily all-encompassing category of video games but mostly included competitive e-sport games or online reaction-based shooters, and sports; sports, which typically engaged in full or most use of motor functions, acting in competitive sports or body training.

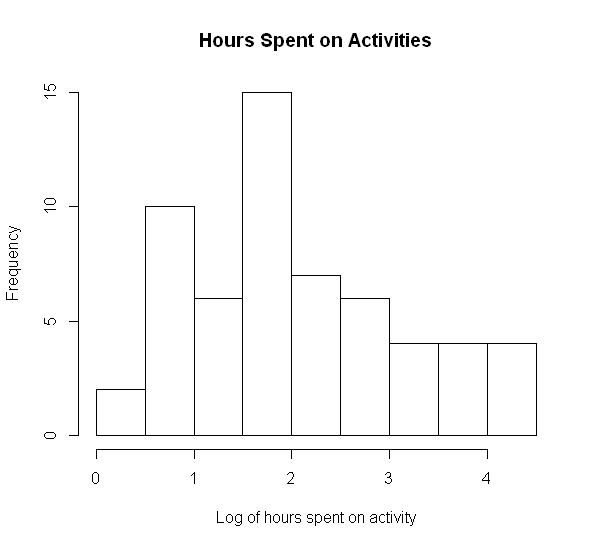
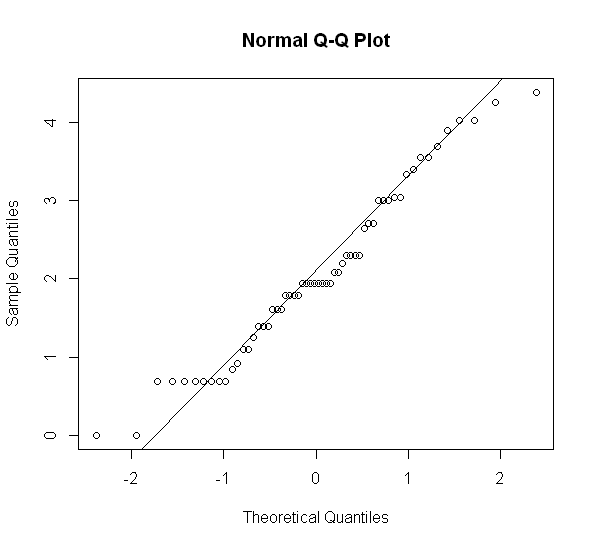


Since each category has more than 5 inputs in it, there collapsing the data further beyond these categories will not be necessary.

**Hours spent on Activity:** Continuous numeric variable

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Minimum | 1st Quartile | Median | 3rd Quartile | Maximum |
| 0 | 1.286 | 1.946 | 2.924 | 4.382 |

A QQ plot and histogram of the log of hours spent on activities is represented below, showing normal trends with some outliers. Since the graph is not clearly normal, we can confirm its normality with a Shapiro-Wilk normality test. Calculated, we receive W = 0.966, and p-value = 0.1033. With that p-value in mind above the alpha, we know that the data is normal.



**R Code for the Response Variable**

First, “reaction.csv” was imported into RStudio.

model <- reaction$Response.time //Created a vector “model” representing the response times. Plotting for histogram and performing the Shapiro-Wilk test for normality, we can see the data is not normal. So, we transform the data set.

logmodel <- log(model) //Transformed data in order to normalize.

hist(logmodel, main='Reaction times', xlab='log of Reaction time in seconds') New histogram seen. Looks normal.

shapiro.test(logmodel) //Confirmed normality with p-value > 0.05.

Additional QQ plot is created for further proof of data normality.

qqnorm(logmodel, ylab = 'Log of Reaction times')

qqline(logmodel)

Obtained five number summary.

summary(logmodel)

**R code regarding the Explanatory Variable of Activity**

Next, a table of the different activities is created to count how many people answered for each respective activity.

table(reaction$Activity)

With that data in mind, we can create a bar plot representing how many people answered for each activity.

barplot(table(reaction$Activity), xlab = 'Activities', ylab='Frequency',main='Counts of activities')

**R code regarding the Explanatory Variable of Hours spent on activity**

Lastly, we investigate the hours spent on each activity. A vector is created of how many hours is spent on an activity.

hours <- reaction$Hours

hist(hours)

We see that the data is not quite ready without a transformation after creating a histogram, so we transform the data.

loghours <- log(hours)

hist(loghours, xlab='Log of hours spent on activity', main='Hours Spent on Activities')

shapiro.test(loghours)

After transforming the data with a logarithm, normality is confirmed by performing the Shapiro-Wilk test for normality, and passes. A QQ plot is created to provide more insight on the normality of the data.

qqnorm(loghours)

qqline(loghours)

Obtained five number summary.

summary(loghours)