INE2002 Term Project

Students' Name & Surname, ID: Alize USTA 1600256
Sude KOSVALI 1506601
Yousef ELBAYOUMI 1722326

Music in Our Life

TABLE OF CONTENT

1. AIM OF THE PROJECT……………………………………………….………..…..3

2. DATA COLLECTION…………………………………………………...……3

2.1 Data Collection Description…………………………………………….……3

2.2 Deciding on the sample……………………………………………………….….3

2.3 Data Summaries……………………………………………………….….….….…4

2.4 The Plots Taken from RStudio……………………………………………..........5

2.5 The Plots Taken from Microsoft Excel………………………………........…6

2.6 Correlation……………………………………….……………………..........…8

3. DATA ANALYSIS…………………………………………………….……9

3.1 Point Estimations……………….……………………….…………….........…9

3.2 Confidence Interval………….………………………………………........…10

3.3 Hypothesis Testing………….……………………………………….........…11

3.4 Goodness of fit……………………………..………..………………….………13

4. MODELLING…………………………………………………….……14

4.1 Linear Regression Models……………………………………………….…….14

5. CONCLUSION………………………………………..…………….…..19

6. REFERENCES…………………………….…………………………......20

7. APPENDIX……………………………………………….……..……..…..21

1. Aim of the project

The principal goal of this project is to collect many common questions related to music and their answers and analyze them, in other words, to show as many statistics and numbers as possible for frequently asked questions related to music.

For example, is it true that listening to music daily will improve a person's mood? (Happiness). What is the estimated time of listening to music for a person? How much does a person pay for music during the month from subscriptions and etc., and other questions will be covered in our research.

2. Data collection

2.1. Data collection description

Our method of collecting data was through questionnaires, because we cannot confirm the authenticity of the numbers on the Internet and because we want the numbers at the highest level of credibility, so we prepared two google surveys one in English and the other in Turkish and we shared it to friends, family and people who we know will provide real and true answers. Most of the people who answered the survey answered as they use Spotify as a music player, therefore we can say this research's values and statistics are taken mostly from Spotify's users.

2.2. Deciding on the sample

After we collected all the answers from the questionnaires, we got about 98 answers, but some of them were repeated answers. Therefore, after filtering, we got answers from 84 different person, so this is basically our population. For determining the sample, we decided to use random sample method because it doesn't really matter which method to use, therefore, we selected the easiest one. From the following photo we used R to determine our sample.

We created excel file and inserted the answers from the surveys, our population was in cells between 3 and 86 (total of 84 person) and we decided to take 25 samples, therefore, R gave us random 25 person and we used their answers as a sample for our research. Also, we decided to use Excel too for making diagrams because it's easier and they look clearer on Excel.

```
> sample(seq(3,86),25)
 [1] 51 38  9  7 42 85 70  5 84 83 11 82 55 65 59 26 37 80 33 75 54 44 57 40 30
```

2.3. Data summaries

The surveys we did contained 13 questions, however, we decided that 6 questions are the most important ones, 5 of them are analyzable and they are shown in the tables below, the remaining question's details will be shown in the following pages in the research.

Quantitative Attributes:

1. **How many hours per day do you listen to music?**
2. **From a 1-100 scale, how much music make you happy?**
3. **How much do you spend on music in TL per month? (subscribes, etc.)**
4. **How many songs do you have in your own playlists? approximately**
5. **What was your age when you started listening to music much?**

Population's data summaries:

	1.	2.	3.	4.	5.
MIN	0.5	10	0	0	1
1Q	1	80	0.25	0	6
MEDIAN	2	90	15	26	10
MEAN	2.515	84.31	16.26	94.07	11.10
3Q	3	100	20	100	15
MAX	12	100	100	600	55

Table 2.3.1: Summary Statistics for each variable for Population

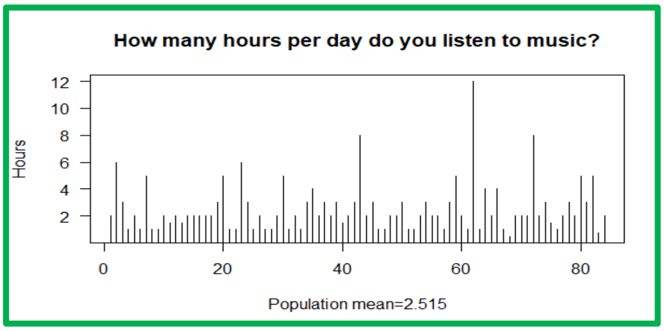
Sample's data summaries:

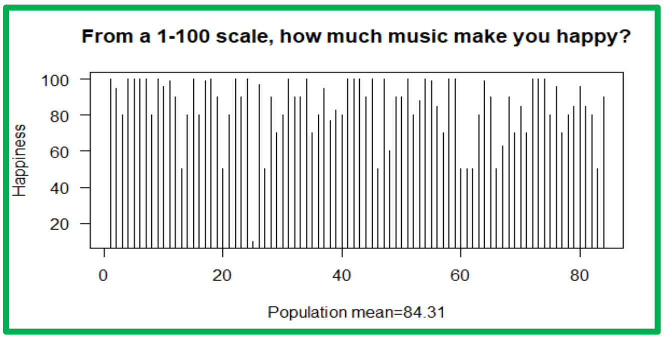
	1.	2.	3.	4.	5.
MIN	0.5	50	0	0	1
1Q	1	80	11	0.5	7
MEDIAN	2	88	20	30	11
MEAN	2.27	86.6	21.16	82.24	13.44
3Q	3	100	20	100	15
MAX	5	100	100	600	55

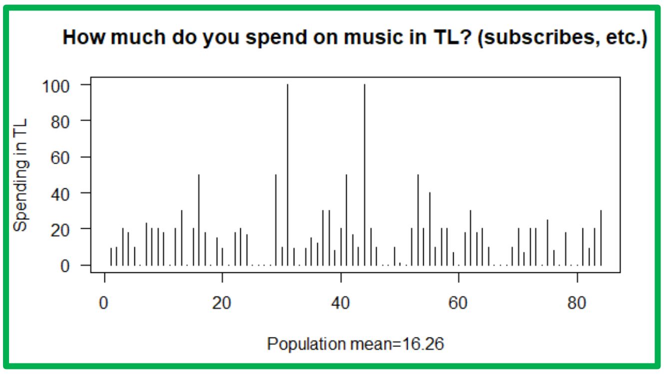
Table 2.3.2: Summary Statistics for each variable for Sample

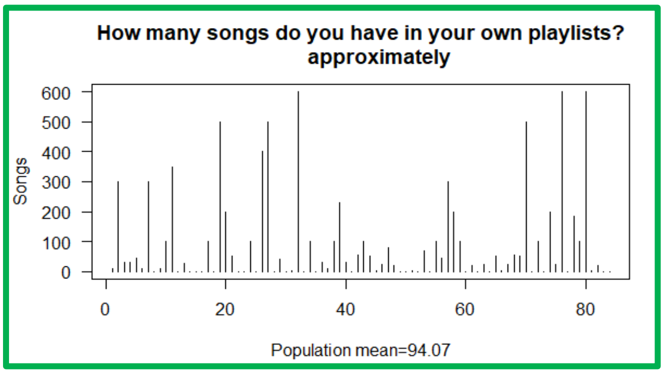
2.4. The Plots Taken from RStudio

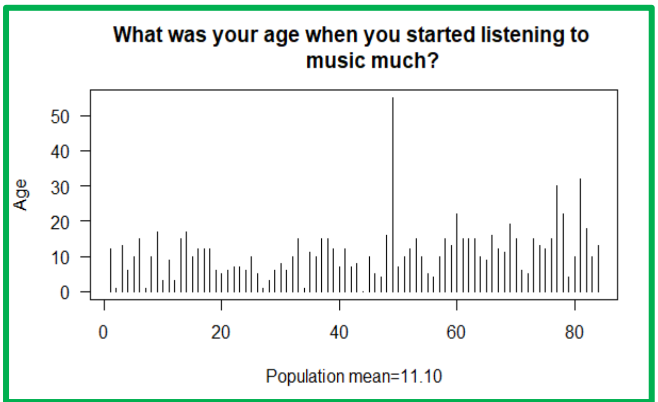
Population's plots:



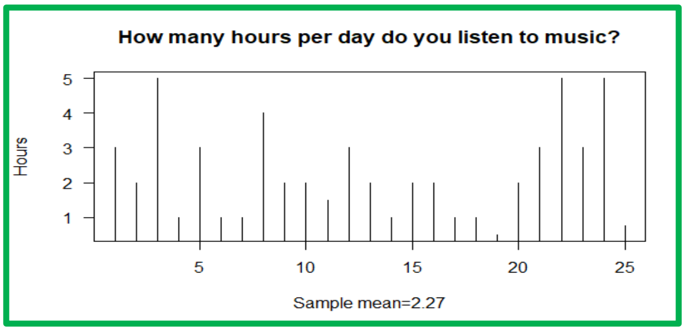


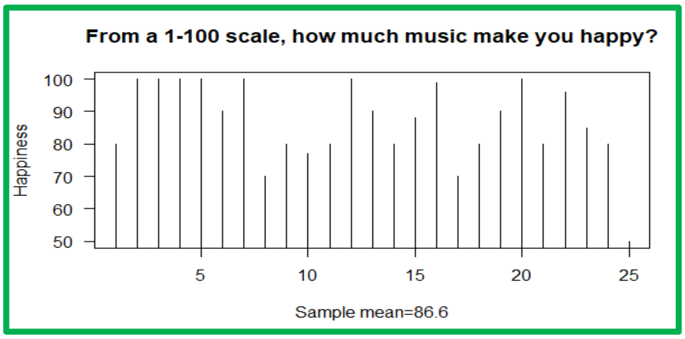


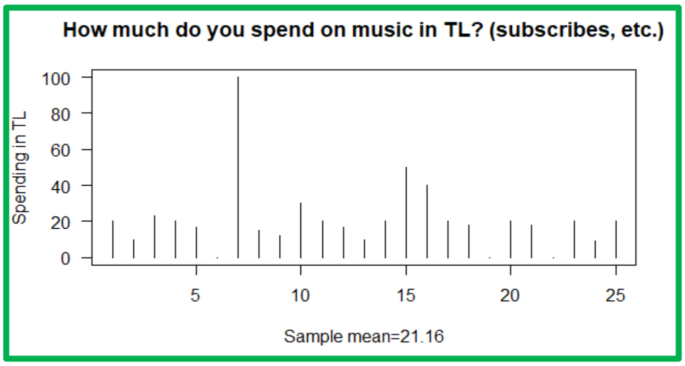


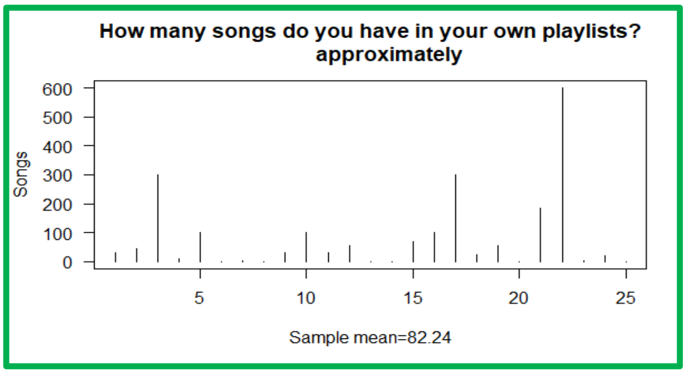


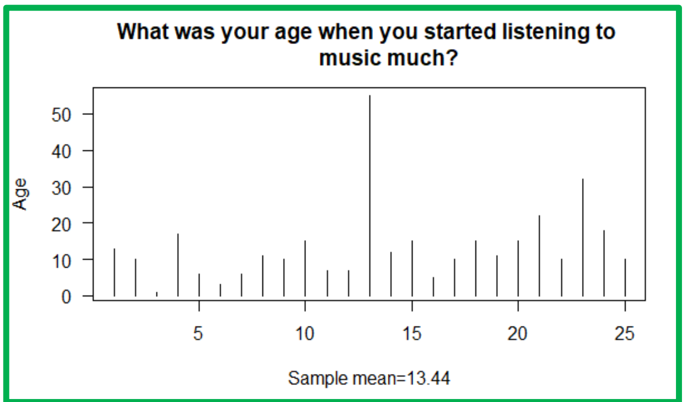
Sample's plots:





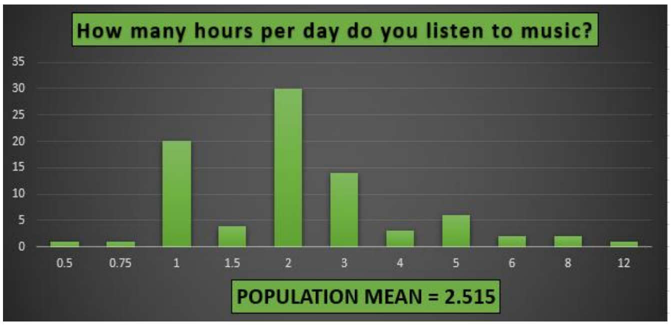


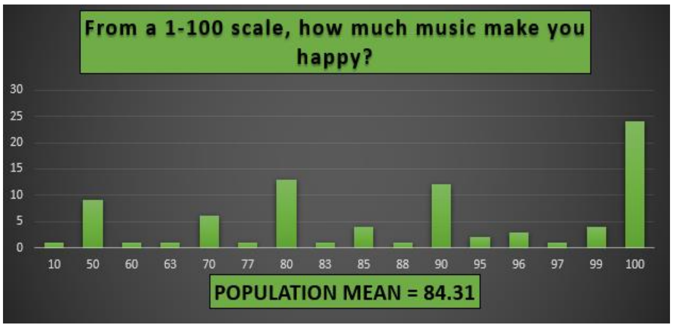


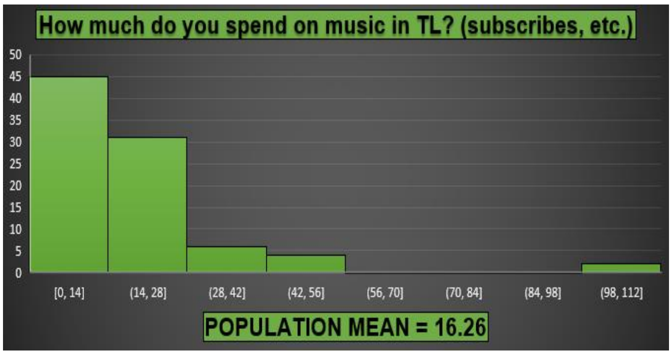


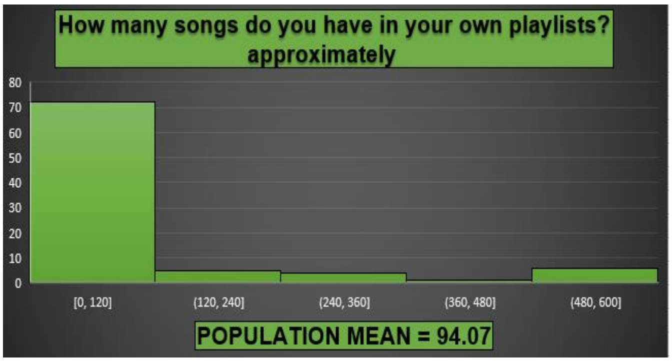
2.5. The Plots Taken from Microsoft Excel

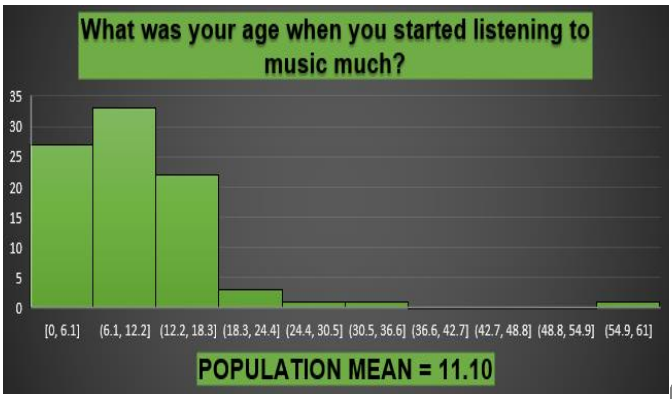
Population's plots:



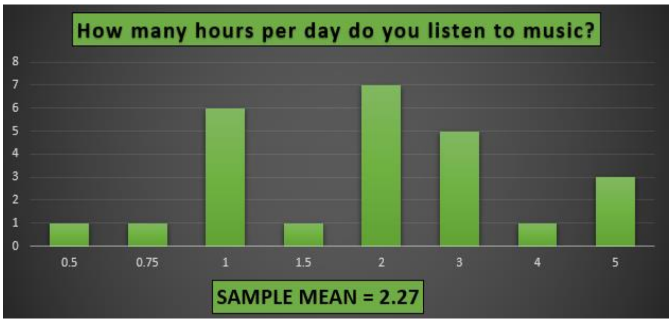


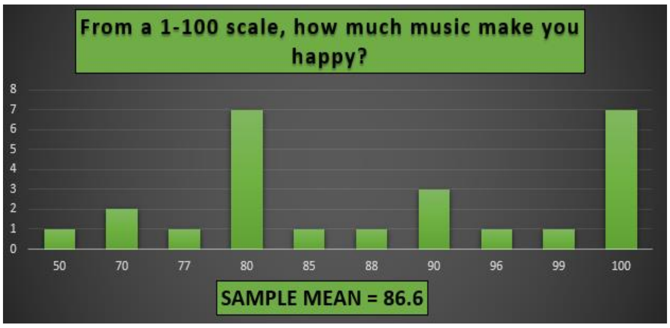


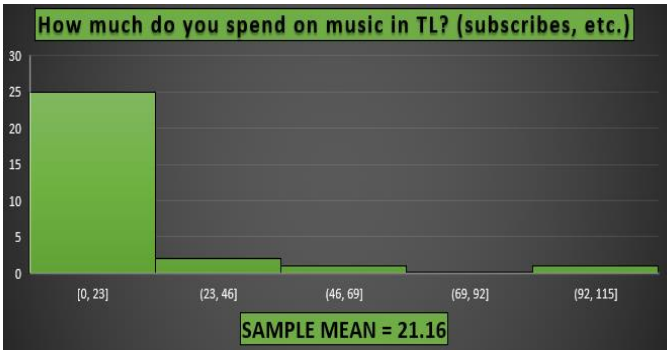


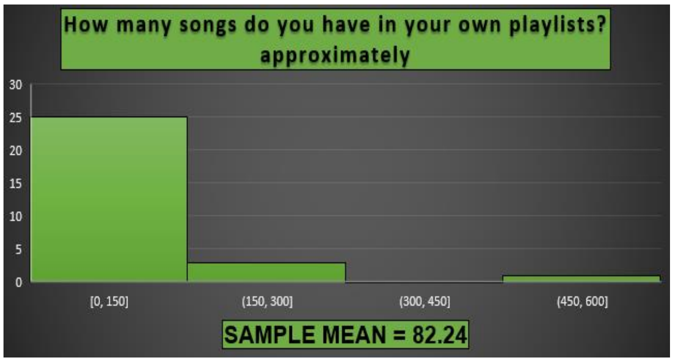


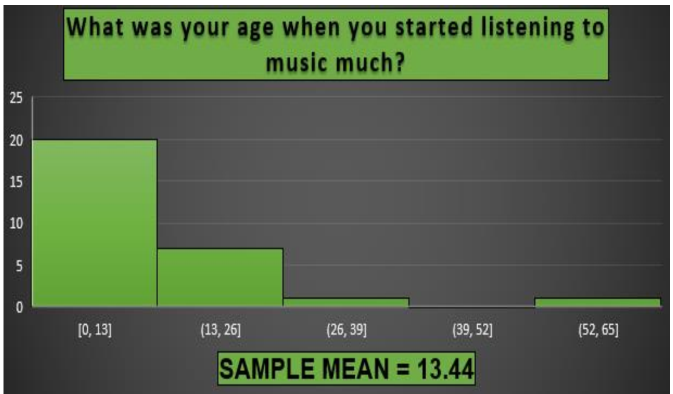
Sample's plots:











One more thing about plots:

An important question has been asked in our surveys which is:

(Which time do you prefer listening to music?)

But because the data for this question is little bit complicated and has different types of answers. For example, **numbers** such as: **13.00**, and **in words** such as: **afternoon**.

Therefore, the data for this question has been collected and analyzed appreciatively, in other words, the mean for this data has been estimated from the chart by the eyes. The charts have been created using Microsoft Excel because we couldn't analyze using RStudio.

Population's plot:

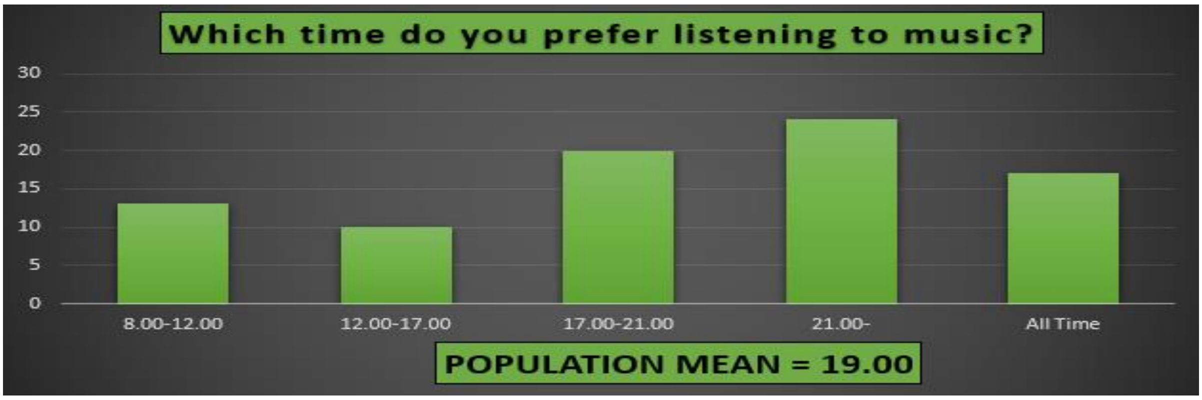


Figure 2.5.1: Plot for special case question on the population

Sample's plot:

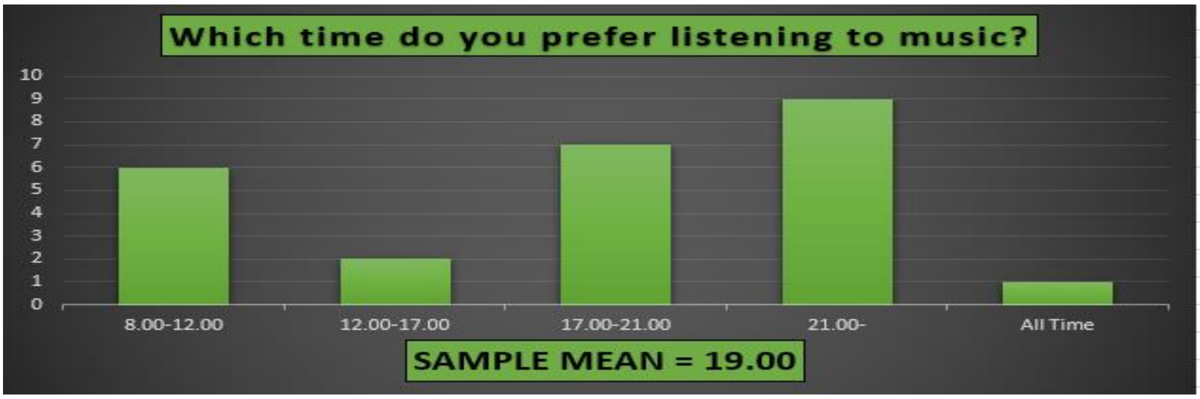


Figure 2.5.2: Plot for special case question on the sample

In analyzing, "All Time" means we increase the other bars by the same value "All Time" have, the other way we estimate without looking at "All Time" bar.

By both ways we approximately estimated both population mean and sample mean as 19.00.

2.6. Correlation

2.6.1. Correlation Matrix

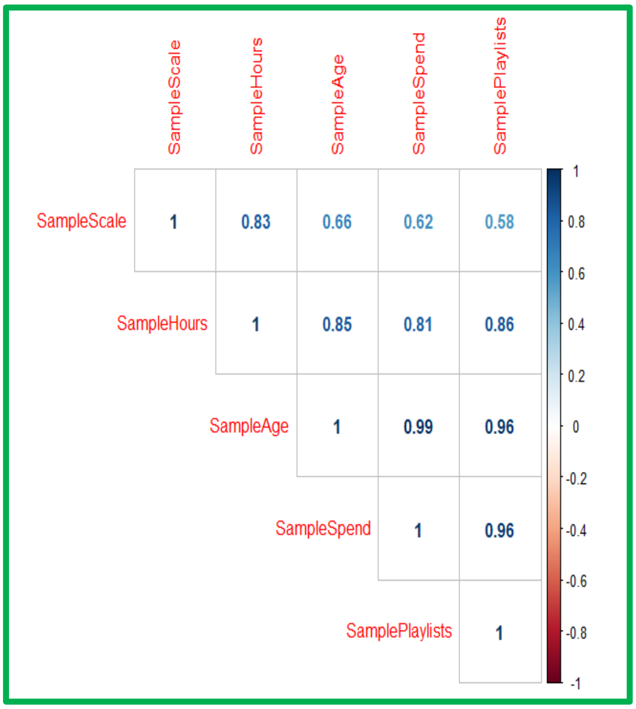


Figure 2.6.1.1: Correlation Matrix in numbers

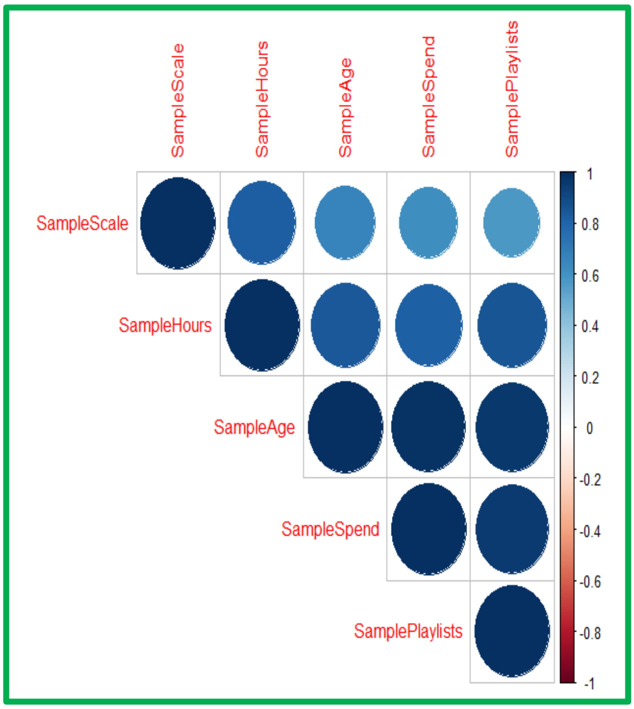


Figure 2.6.1.2: Correlation Matrix in circles

SampleScale = Happiness scale. **SampleHours = Averagely listening to music per day.**

SampleAge = Interested to music at what age. **SampleSpend = Spending money in TL per month.**

SamplePlaylists = Number of songs in created playlists.

2.6.2. Some correlations between the quantitative

With using the correlation matrix, we will have clue about the relationships that our variables have between each other. For example, the relationship between **Number of songs in created playlists** and **Interested to music at what age** looks strong and positive, almost close to 1. Another positive linear relationship would be between **Interested to music at what age** and **Averagely listening to music per day**, and between **Spending money in TL per month** and **Happiness scale** there is a relationship but not so strong. Let's check the numbers:

Correlation between Number of songs in created playlists and Interested to music at what age
0.9645156

Table 2.6.2.1: Correlation result table

Correlation between Interested to music at what age and Averagely listening to music per day
0.8482773

Table 2.6.2.2: Correlation result table

Correlation between Spending money in TL per month and Happiness scale
0.6151995

Table 2.6.2.3: Correlation result table

3. Data Analysis

3.1. Point Estimations

Population's:

Variables	Mean	Standard Deviation
Averagely listening to music per day	2.515	1.8470
Happiness scale	84.31	17.8158
Spending money in TL per month	16.26	17.9988
Number of songs in created playlists	94.07	153.7999
Interested to music at what age	11.10	7.5445

Table 3.1.1: Point Estimations for Quantitative Variables for population

Here, we see that the average happiness scale for this study is approximately 84.31. The average spending money in TL per month value is approximately 16.26. Also, we see that happiness scale standard deviation is approximately 17.8158, similar with spending money in TL per month standard deviation is approximately 17.9988. And so on.

Sample's:

Variables	Mean	Standard Deviation
Averagely listening to music per day	2.27	1.3577
Happiness scale	86.6	12.6260
Spending money in TL per month	21.16	19.8025
Number of songs in created playlists	82.24	137.1697
Interested to music at what age	13.44	10.7822

Table 3.1.2: Point Estimations for Quantitative Variables for sample

Here, we see that the average happiness scale for this study is approximately 86.6. The average spending money in TL per month value is approximately 21.16. Also, we see that happiness scale standard deviation is approximately 12.6260, similar with spending money in TL per month standard deviation is approximately 19.8025. And so on.

3.2. Confidence Interval

Question: What is the confidence interval for Averagely listening to music per day range at %95 confidence level?

Normality assumption for this data was previously checked and met.

Here, we can say with %95 confidence level that the data lie between 1.709571 and 2.830429.

One Sample t-test	
95 percent confidence interval:	
t = 8.3598	df = 24
True mean is not equal to 0	Mean of x = 2.27
1.709571	2.830429

Table 3.2.1: One Sample t-test

Question: What is the confidence interval for Happiness scale at %99 confidence level?

For normality assumption:

H_0: Happiness scale is distributed normally.

H_1: Happiness scale is not distributed normally.

From the output, the p-value > 0.0226 implying that the distribution of the data are not significantly different from normal distribution.

In other words, we can assume the normality.

Shapiro–Wilk normality test	
W = 0.87102	**p-value = 0.04535**

Table 3.2.2: Shapiro-Wilk Normality Test

Here, we can say with %99 confidence level that the data lie between 79.53715 and 93.66285.

One Sample t-test	
99 percent confidence interval:	
t = 34.294	**df = 24**
True mean is not equal to 0	**Mean of x = 86.6**
79.53715	**93.66285**

Table 3.2.3: One Sample t-test

3.3. Hypothesis Testing

Question: Is the averagely listening to music per day is higher&around the average time listening to music per day for people in Turkey?

According to DAILY SABAH Website, the 2020 report on digital media in Turkey by social creative agency have shown that turks spend 1 hour and 21 minutes listening to music per day. Thus, our hypothesis is as follows:

$H_{0:}$ $\mu_{\text{laveragely istening to music per day}} > 1.21$

$H_{1:}$ $\mu_{\text{laveragely istening to music per day}} \leq 1.21$

First things first, we check the normality for the data, therefore, our hypothesis test for it is:

H_0: Averagely listening to music per day is distributed normally.

H_1: Averagely listening to music per day is not distributed normally.

From the output, the p-value > 0.033 implying that the distribution of the data are not significantly different from normal distribution.

In other words, we can assume the normality.

Shapiro–Wilk normality test	
W = 0.87907	p-value = 0.06613

Table 3.3.1: Shapiro-Wilk Normality Test

For t-test, the p-value is smaller than 0.05. We accept the null hypothesis that the averagely listening to music per day is more than 1.21, In other words, it's true that people in turkey listen to music around 1 hour and 21 minutes daily.

One Sample t-test		95 percent confidence interval:
t = 3.9037	df = 24	p-value = 0.003359
Mean of x		
2.27		

Table 3.3.2: One Sample t-test

Question: Is the Happiness scale is higher&around the average happiness for a person?

According to OECD Better Life Website, on a scale from 0 to 100, people on average across the OECD gave it a 65. Thus, our hypothesis is as follows:

$H_{0:}$ $\mu_{\text{happiness scale}} > 65$

$H_{1:}$ $\mu_{\text{happiness scale}} \leq 65$

First things first, we check the normality for the data, therefore, our hypothesis test for it is:

H_0: Happiness scale is distributed normally.

H_1: Happiness scale is not distributed normally.

From the output, the p-value > 0.0226 implying that the distribution of the data are not significantly different from normal distribution.

In other words, we can assume the normality.

Shapiro–Wilk normality test	
W = 0.87102	p-value = 0.04535

Table 3.3.3: Shapiro-Wilk Normality Test

For t-test, the p-value is smaller than 0.05. We accept the null hypothesis that the happiness scale is more than 65, In other words, it's true that the average happiness for a person is around 65 out of 100.

One Sample t-test		95 percent confidence interval:
t = 8.5538	df = 24	p-value = 0.000473
Mean of x		
86.6		

Table 3.3.4: One Sample t-test

Question: Is the Spending money in TL per month is less than the average spending on music for students in Turkey?

According to Study in Turkey Website, different type of data has been collected from various students to give good view for whoever thinking about studying in Turkey, data such as books, mobile fees, transportation, and the average for music subscriptions collected and calculated as 25 TL. Thus, our hypothesis is as follows:

$H_{0:}\ \mu_{\text{spending money in tl per month}} < 25$

$H_{1:}\ \mu_{\text{spending money in tl per month}} \geq 25$

First things first, we check the normality for the data, therefore, our hypothesis test for it is:

H_0: Spending money in TL per month is distributed normally.

H_1: Spending money in TL per month is not distributed normally.

From the output, the p-value > 0.0248 implying that the distribution of the data are not significantly different from normal distribution.

In other words, we can assume the normality.

Shapiro–Wilk normality test	
W = 0.68748	p-value = 0.04967

Table 3.3.5: Shapiro-Wilk Normality Test

For t-test, the p-value is smaller than 0.05. We accept the null hypothesis that the spending money in TL per month is less than 25, In other words, it's true that person in turkey spends more than 21.16 TL per month in music, 25 TL actually.

One Sample t-test		95 percent confidence interval:
t = -0.96957	df = 24	p-value = 0.0171
Mean of x		
21.16		

Table 3.3.6: One Sample t-test

3.4. Goodness of Fit

For goodness of fit test, we decided to use the special question we talked about previously.

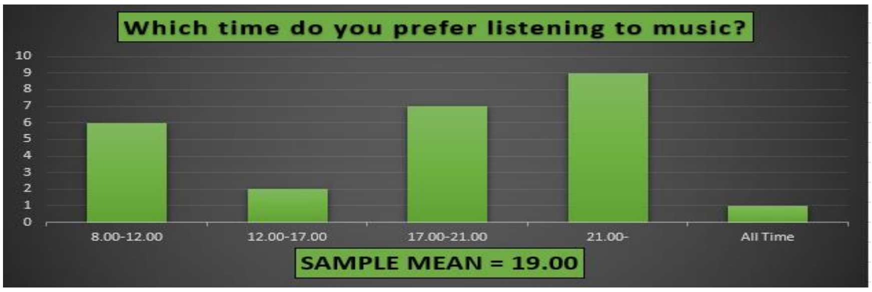


Figure 3.4.1: Plot for special case question on the sample

As we discussed before, we will assume that we have 24 observations (without All Time observation) to make good estimation.

N. of Defects	Observed Frequency	Probability	Expected Frequency	Chi-square
0	6	4.8836	11.72066	2.79216
1	2	0.35000	8.40020	4.87638
2	7	0.12542	3.879128	2.51083
3	9	0.03620		**SUM = 10.17938**
N = 24	Poisson distribution = 0.7167		**Chi-square test statistic is k – p – 1**	
Alpha = 0.05			**= 3 – 1 -1 = 1**	

Table 3.4.1: Goodness of Fit table for calculating Chi-square

The expected value for 2nd defect = 3.01021272 and for 3rd defect = 0.868915. Therefore, we took the sum which is 3.879128

The sum of the Chi-squares is = 2.79216+4.87638+2.51083 = **10.17938**

H_0: The form of the distribution of defects is Poisson

H_1: The form of the distribution of defects is not Poisson

From chi-square table $X^2_{0.05,23} = 35.17$, and since 10.179 < 35.17 we will reject the null hypothesis in favor of the alternative hypothesis. Also from p-value perspective, we will get p-value as 0.0005588, and since p-value < 0.05 we will reject the null hypothesis.

4. Modelling

4.1. Linear Regression Models

Question: Is there a relationship between [Averagely listening to music per day] and [happiness scale]?

For answering such a questions, we can apply many ways. First of all, we better visualize the question so we can have a general idea.

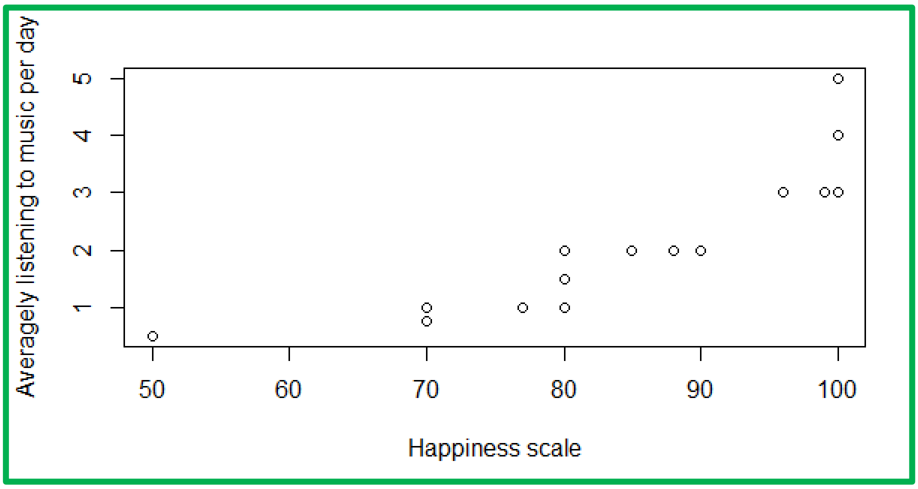


Figure 4.1.1: Plot of averagely listening to music per day and happiness scale

As we discussed previously in correlation section, the 0.828 correlation means we have a strong positive linear relationship between the two variables.

Correlation between averagely listening to music per day and happiness scale
0.8287254

Table 4.1.1: Correlation result table

After checking the correlation, we have a general idea and a good starting point, now, let's apply linear regression model to determine the relationship between these two variables.

First Model

In order to construct the first model, let's use averagely listening to music per day as our response variable and happiness scale as our regressor.

lm(formula = Averagely listening to music per day ~ Happiness scale)				
Residuals:				
Min	**1Q**	**Median**	**3Q**	**Max**
-0.6818	-0.5730	-0.3750	0.3181	1.5359

Table 4.1.2: First Model Summary table

Coefficients:				
	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	-5.44727	1.09773	-4.962	0.0000512
Happiness scale	0.08911	0.01255	7.102	0.0000 *
p-value: 0.0000 *		Residual standard error: 0.7762 on 23 degree of freedom		
Multiple R-squared: 0.6868		Adjusted R-squared: 0.6732		

Table 4.1.3: First Model Summary table

The coming two graphs are to see the adequacy of the model, even though we can see that both $ß_0$ and the $ß_1$ are significant.

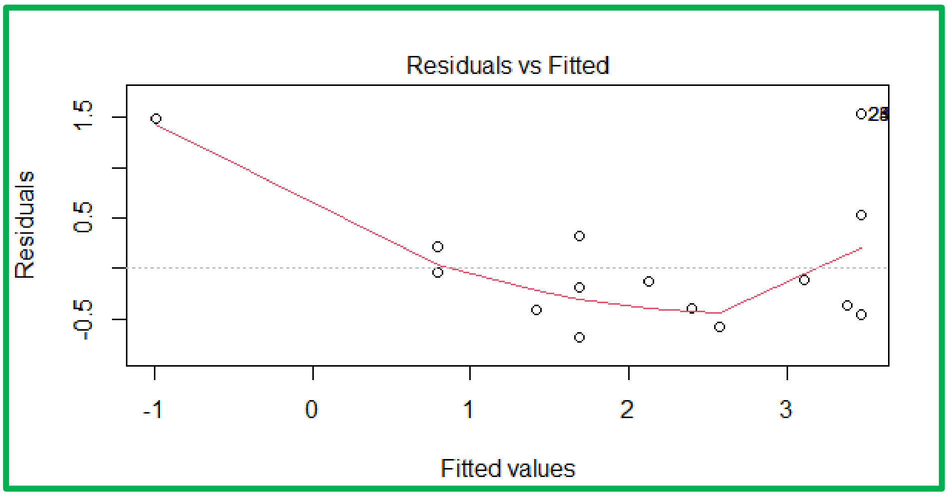


Figure 4.1.2: Residuals VS. Fitted Values

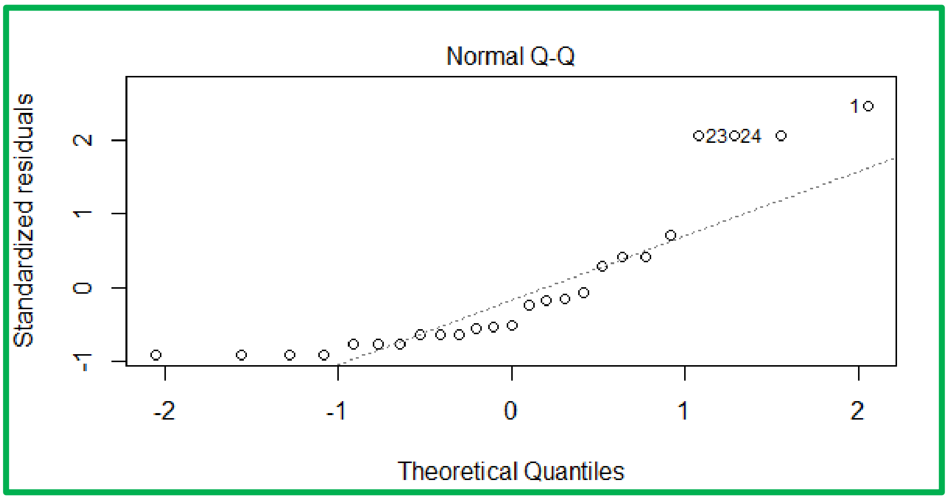


Figure 4.1.3: Standardized Residuals

We can see the problem in Residual VS. Fitted Values graph, which is the graph has a trend, obviously we don't want that in order to satisfy constant variance. Another problem is that normality is not satisfied for errors. Therefore, we will apply a transformation method.

Second Model

Our plan in the second model is to use transformation method in order to handle the unequal error variances, also we had a problem which is normality is not satisfied for errors, so we are hoping to solve it too. Therefore, we will take the power of 0.5 (square root) of our response (averagely listening to music per day) in this model.

lm(formula = Averagely listening to music per day^0.5 ~ Happiness scale)				
Residuals:				
Min	**1Q**	**Median**	**3Q**	**Max**
-0.23802	-0.13383	-0.07182	0.14195	0.39913

Table 4.1.4: Second Model Summary table

Coefficients:				
	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	-1.242106	0.301179	-4.124	0.000143
Happiness scale	0.031002	0.003443	9.005	0.0000 *
p-value: 0.0000 *		Residual standard error: 0.213 on 23 degree of freedom		
Multiple R-squared: 0.779		Adjusted R-squared: 0.7694		

Table 4.1.5: Second Model Summary table

The tables are trolling again, both $ß_0$ and the $ß_1$ are significant again, and R-squared can be considered high, **BUT** what will happen if we check the plots ?? let's check

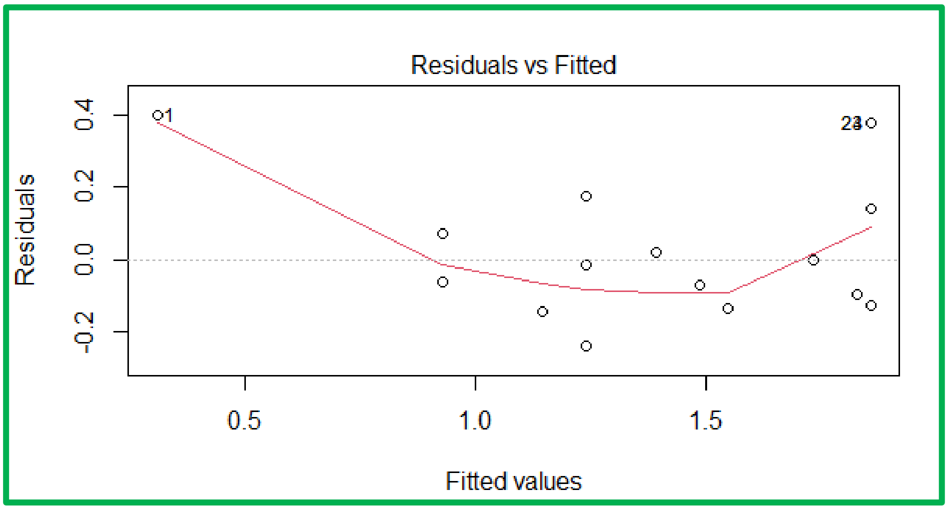


Figure 4.1.4: Residuals VS. Fitted Values

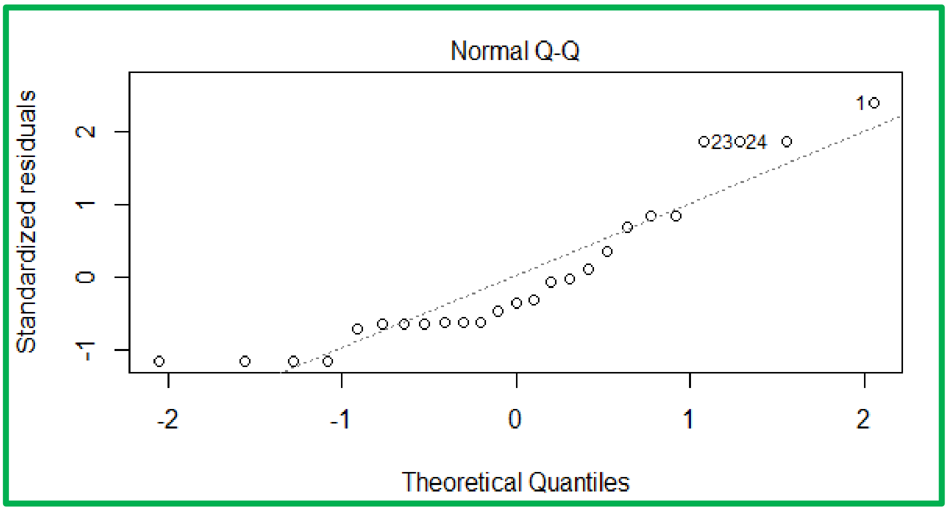


Figure 4.1.5: Standardized Residuals

We still need to do another transformation, yes we fixed the normality of errors but we still have non-constant variance. Therefore, we will do transformation on x.

<u>Third Model</u>

lm(formula = Averagely listening to music per day^0.5 ~ I(Happiness scale^0.5))				
Residuals:				
Min	**1Q**	**Median**	**3Q**	**Max**
-0.26117	-0.14029	-0.08282	0.15305	0.45871

Table 4.1.6: Third Model Summary table

Coefficients:				
	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	-3.57463	0.60880	-5.872	0.00000552
I(Happiness scale^0.5)	0.54066	0.06542	8.264	0.0000 *
p-value: 0.0000 *		Residual standard error: 0.2274 on 23 degree of freedom		
Multiple R-squared: 0.7481		Adjusted R-squared: 0.7371		

Table 4.1.7: Third Model Summary table

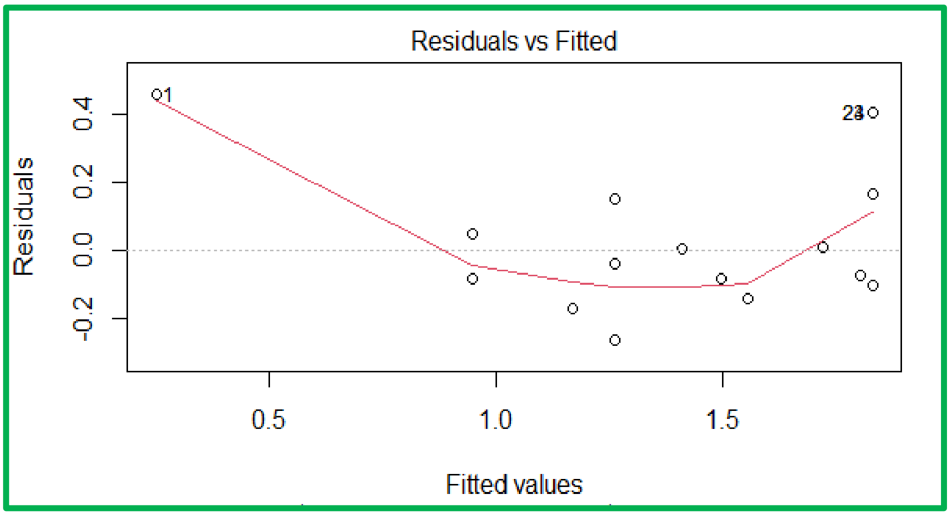


Figure 4.1.6: Residuals VS. Fitted Values

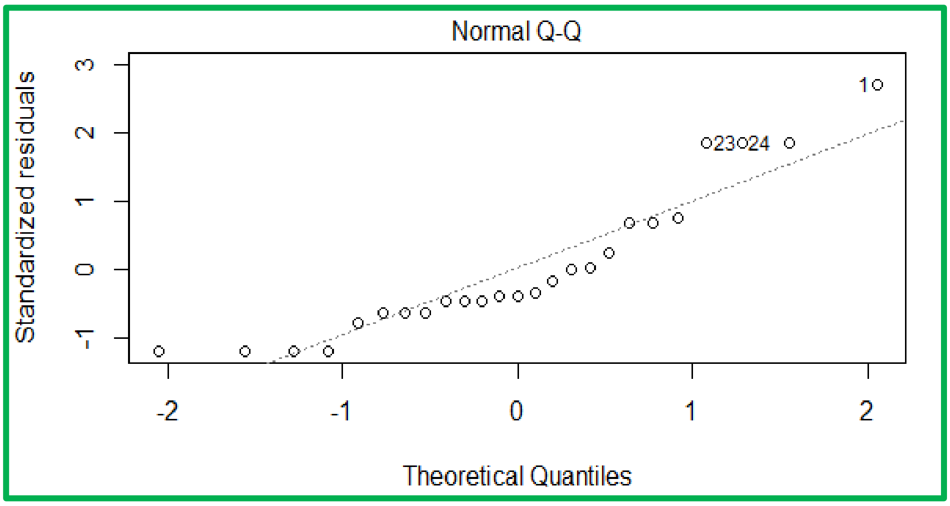


Figure 4.1.7: Standardized Residuals

After the third model we can say that we got the better results so far, our variance is stabilized, and the errors are approximately normally distributed. Unfortunately, we have some outliers (most importantly 1, 23, and 24th observations) which has some affects on the effectiveness of our model.

We can do hypothesis test and check the p-value since we're talking about simple linear regression.

H_0 : Model is not significant. ($ß_1 = 0$)

H_1 : Model is significant. ($ß_1 \neq 0$)

Since p-value < 0.05, Reject H_0. And with full of confidence we can say that our model is significant at 0.95 confidence interval (0.05 significant level) and B_1 is significant as well.

Our model accounts approximately for 74% of the variability in the data since we got R-Squared approximately as 0.74.

The fitted values for the model are:

$\sqrt{(y_i)}$ = -3.57463 + 0.54066$\sqrt{(x_i)}$

where y: **Averagely listening to music per day**

x: **Happiness scale**

Interpretation of Coefficients:

- One unit increase in the square root of **Happiness scale** will result in 0.54066 increase in the square root of **Averagely listening to music per day**.
- The average square root of **Averagely listening to music per day** is -3.57463 if the range of **Happiness scale** covers 0.

5. Conclusion

In conclusion for our research about music with taking the data from Spotify's users, many different statistics methods and tests has been applied such as confidence level test, hypothesis test, goodness of fit, correlation, simple linear regression. We discussed many important questions and we found answers for them. We found the 95% and 99% confidence intervals for some data in CI section. We saw that our assumptions can be lower than the ones on the internet such as third question in the hypothesis testing. We found that our data are nicely connected and related with each other as we saw in correlation, of course we had some medium correlations but most of them have high correlations. We saw in linear regression models how we try over and over until we got stabilized variance and the errors are approximately normally distributed.

6. References

Google Survey we created in English language

https://docs.google.com/forms/d/e/1FAIpQLSfqIeKnAjLxmU9XkoxeR1LPB-3hlFhP-M1QcIwdRI8mQgNc9w/viewform?usp=sf_link

Google Survey we created in Turkish language

https://docs.google.com/forms/d/e/1FAIpQLSckShih6aNl7_2aei9kcDqVbIHz-e8zmSRdpA5CzkBJFBO_eQ/viewform

We created this survey in the beginning of our research to make sure that our data is so sufficient and new, we created in both English and Turkish languages to make sure that we get as most answers as possible.

Daily Sabah Website, retrieved from

https://www.dailysabah.com/life/health/5050-turks-spend-equal-time-sleeping-and-on-the-internet-study-shows

From this website the average time listening to music per day has been taken as 1.21 hours.

OECD Better Life Index Website, retrieved from

http://www.oecdbetterlifeindex.org/topics/life-satisfaction/

From this website the happiness scale has been taken as 65 for the normal person.

Study in Turkey Website, retrieved from

http://www.studyinturkey.gov.tr/StudyinTurkey/ShowDetail?rID=haH8OKhk97k=&&cId=PE4Nr0mMoY4=

From this website the Spending money per month in TL has been taken as 25 TL for the students in Turkey, which makes it perfect as source because we collected our data mostly from students in Turkey too.

7. Appendix

```
#Libraries used during the analysis
library(readxl)
library(corrplot)

#Deciding the sample
sample(seq(3,86),25)

#Reading the files from Excel
library(readxl)
Test1 <- read_excel("C:/Users/Yousef El-bayoumi/Desktop/Test1.xlsx")
View(Test1)
library(readxl)
Test2 <- read_excel("C:/Users/Yousef El-bayoumi/Desktop/Test2.xlsx")
View(Test2)

#Summary statistics for the data
     #Population:
  x1 = Test1$PopulationHours
  x2 = Test1$PopulationScale
  x3 = Test1$PopulationSpend
  x4 = Test1$PopulationPlaylists
  x5 = Test1$PopulationAge
     # Sample:
  x6 = Test2$SampleHours
  x7 = Test2$SampleScale
  x8 = Test2$SampleSpend
  x9 = Test2$SamplePlaylists
  x10 = Test2$SampleAge
  plot(x1,type="h",las=1,bty="o",
       main="How many hours per day do you listen to music?",
       xlab="Population mean=2.515",ylab="Hours")
  plot(x2,type="h",las=1,bty="o",
```

```
      main="From a 1-100 scale, how much music make you happy?",
      xlab="Population mean=84.31",ylab="Happiness")
  plot(x3,type="h",las=1,bty="o",
      main="How much do you spend on music in TL? (subscribes, etc.)",
      xlab="Population mean=16.26",ylab="Spending in TL")
  plot(x4,type="h",las=1,bty="o",
      main="How many songs do you have in your own playlists?
      approximately",
      xlab="Population mean=94.07",ylab="Songs")
  plot(x5,type="h",las=1,bty="o",
      main="What was your age when you started listening to
      music much?",
      xlab="Population mean=11.10",ylab="Age")
  plot(x6,type="h",las=1,bty="o",
      main="How many hours per day do you listen to music?",
      xlab="Sample mean=2.27",ylab="Hours")
  plot(x7,type="h",las=1,bty="o",
      main="From a 1-100 scale, how much music make you happy?",
      xlab="Sample mean=86.6",ylab="Happiness")
  plot(x8,type="h",las=1,bty="o",
      main="How much do you spend on music in TL? (subscribes, etc.)",
      xlab="Sample mean=21.16",ylab="Spending in TL")
  plot(x9,type="h",las=1,bty="o",
      main="How many songs do you have in your own playlists?
      approximately",
      xlab="Sample mean=82.24",ylab="Songs")
  plot(x10,type="h",las=1,bty="o",
      main="What was your age when you started listening to
      music much?",
      xlab="Sample mean=13.44",ylab="Age")
#Showing the correlation between the data
  cor(Test2$SampleSpend, Test2$SampleScale)
  cor(Test2$SampleAge, Test2$SampleHours)
```

```
cor(Test2$SamplePlaylists, Test2$SampleAge)
cor.test(Test2$SampleSpend, Test2$SampleScale)
cor.test(Test2$SampleAge, Test2$SampleHours)
cor.test(Test2$SamplePlaylists, Test2$SampleAge)
cor(Test2)
cor2 = cor(Test2)
install.packages("corrplot")
require(corrplot)
corrplot(cor2,method = "circle", type = "upper", order = "AOE")
corrplot(cor2,method = "number", type = "upper", order = "AOE")

#Point estimations
mean(x1)
mean(x2)
mean(x3)
mean(x4)
mean(x5)
mean(x6)
mean(x7)
mean(x8)
mean(x9)
mean(x10)
sd(x1)
sd(x2)
sd(x3)
sd(x4)
sd(x5)
sd(x6)
sd(x7)
sd(x8)
sd(x9)
sd(x10)
```

```
 #Confidence Interval
t.test(Test2$SampleHours, conf.level = 0.95)
  shapiro.test(Test2$SampleScale)
  t.test(Test2$SampleScale, conf.level=0.99)

#Hypothesis Testing
  shapiro.test(Test2$SampleHours)
  t.test(Test2$SampleHours,alternative = "greater",mu=1.21)
  shapiro.test(Test2$SampleScale)
  t.test(Test2$SampleScale,alternative = "greater",mu=65)
  shapiro.test(Test2$SampleSpend)
  t.test(Test2$SampleSpend,alternative = "less",mu=25)

#Goodness of Fit
  Times <- c(10, 14.5, 19, 22.5)
  names(Times) <- c("Morning", "Afternoon", "Evening", "Late time")
  probability <- c(6/24, 2/24, 7/24, 9/24)
  chisq.test(Times, p=probability)

#Modeling
     #Simple Linear Regression
  plot(y=Test2$SampleHours,x=Test2$SampleScale,
  ylab = "Averagely listening to music per day",
  xlab = "Happiness scale")
  cor(Test2$SampleHours,Test2$SampleScale)
  #First Model
  Model1<-lm(Test2$SampleHours~Test2$SampleScale)
  summary(Model1)
  plot(Model1,1:2)
  #Second Model
Model2<-lm(Test2$SampleHours^0.5~Test2$SampleScale)
  summary(Model2)
  plot(Model2,1:2)
```

```
#Third Model

Model3<-lm(Test2$SampleHours^0.5~I(Test2$SampleScale^0.5))

summary(Model3)

plot(Model3,1:2)
```