Graphical user interface, text

Description automatically generated

**Hospital Service Quality**

**A FINAL PROJECT REPORT SUBMITTED**

**IN FULFILMENT OF THE REQUIREMENTS FOR COURSE STAT 250**

**APPLIED STATISTICS**

**DEPARTMENT OF STATISTICS OF**

**MIDDLE EAST TECHNICAL UNIVERSITY**

**BY**

**Ömer Adsız**

**Zülcan Duygu Benzer**

**Muaz Ozpolat**

**Alper Tunahan Öztürk**

**Çağla Şen**

**June 2021**

Table of Contents

[ABSTRACT 3](#_Toc75489521)

[1. Introduction 3](#_Toc75489522)

[1.1. Data description 3](#_Toc75489523)

[1.2. Research questions 4](#_Toc75489524)

[1.3. Aim of the study 4](#_Toc75489525)

[2. Methodology/Analysis 4](#_Toc75489526)

[3. Results and Findings 5](#_Toc75489527)

[Research Question 1: Which variable plays the most critical role in patient satisfaction? 5](#_Toc75489528)

[Research Question 2: Does patient satisfaction differ among genders? 10](#_Toc75489529)

[Research Question 3: Do the anxiety levels of the patients have any relationship to their gender? 13](#_Toc75489530)

[Research Question 4: Does patient satisfaction fluctuate for the number of days the patient stays in the hospital or is the patient a surgical patient? 15](#_Toc75489531)

[4. Discussion/Conclusion 19](#_Toc75489532)

[References 19](#_Toc75489533)

# ABSTRACT

This report is a detailed analysis associated with hospital patients at an anonymous hospitalization. The hospital is looking to utilize a curriculum to improve its service quality and productivity. In order to progress, this process undoubtedly starts with and ends with the patients in the hospital. The objective of this report is to break down by primarily dissecting the multiple factors related to the patients and seeking the correlation between those factors and the overall hospital productivity. In order to get to the root of the dilemma, several statistical methods were used to approach the problem, such as hypothesis tests (t-test), multiple linear regression and a nonparametric test. Based on the results of these findings, various statistical representations such as histograms, boxplots, scatterplots and correlation graph were applied. By the end of this report, conclusions from the data can be drawn vividly to answer the unknown.

# 1. Introduction

Having an exceptional hospital and phenomenal service is very essential in today’s day and age, hence examining these patients’ frustrations is very important. The original dataset contains 95 observations (patients in this case). The sample was selected randomly, without any pattern whatsoever. The exact numbers of representations used were two boxplots, four histograms, seven scatterplots and one correlation graph. For the scatter plots and correlation graphs, linear regression method was used. As for the corresponding histogram and boxplots, the t-test and non-parametric test were applied.

## 1.1. Data description

Out of all the factors that could possibly affect these patients, seven variables were drawn from the original dataset. Five of these variables are numerical (age, severity, anxiety, satisfaction and day), and the remaining two are categorical (gender and surgical). The surgical variable has two categories, and they are whether a patient is a surgical or a medical patient. Gender has two categories as well, respectfully. The severity variable refers to how severe the patients’ condition is, meanwhile the anxiety variable is attributed to the patients’ anxiety level. The day variable is assigned to how many days the patient has been in the hospital and the satisfaction variable points out how content the patient is with the hospital service and quality. The age variable is self-explanatory.

## 1.2. Research questions

**1)** Which variable plays the most critical role in patient satisfaction?

**2)** Does patient satisfaction differ among genders?

**3)** Do the anxiety levels of the patients have any relationship to their gender?

**4)** Does patient satisfaction fluctuate for the number of days the patient stays in hospital or is the patient a surgical patient?

## 1.3. Aim of the study

The objective of the analysis was to correlate the assigned variables and analyze how they contributed to the hospitals overall service quality and production. The goals to achieve were to see how the selected categorical and numerical variables had an effect on the patients’ situation and finding out whether multiple certain variables had a linear relationship, based on the research questions.

# 2. Methodology/Analysis

The three essential statistical methods utilized to interpret the data were hypothesis testing (t-testing in particular), a nonparametric test and linear regression (both simple and multiple). By using R-studio, the representations were then created from the tests. Linear regression was used to see any possible linear relationships between the variables and correlation as well. The variables that multiple linear regression was applied to were patient satisfaction. The t-test was used again for patient satisfaction but this time with the gender variable. Lastly, the non-parametric test was used for the anxiety levels of patients` along with again the gender variable.

# 3. Results and Findings

## Research Question 1: Which variable plays the most critical role in patient satisfaction?

In this research question multiple linear regression is used. In multiple linear regression patient satisfaction is taken as response and all the other variables are taken as covariates.

* One of the assumptions of multiple linear regression is multicollinearity, in order to check multicollinearity the VIF values are checked.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Age | severity | surgical1 | anxiety | day | gender1 |
| 1.0762 | 1.0228 | 1.0629 | 1.0532 | 1.0268 | 1.0135 |

Since all the VIF values are less than 10, there is no multicollinearity problem.

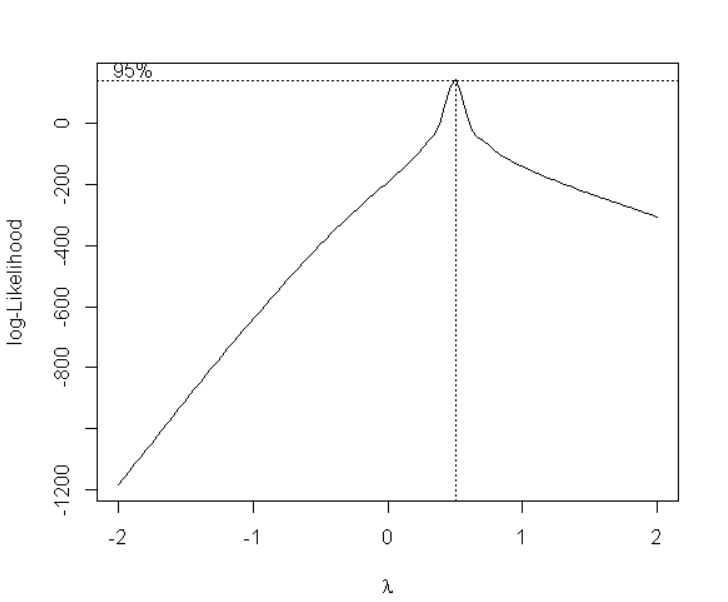
* In order to check the normality of the response, Shapiro-Wilk normality test is used.

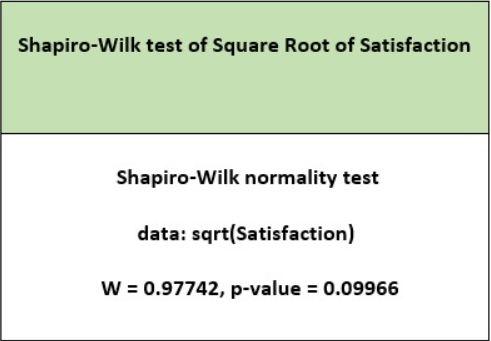
Graphical user interface, text, application, email

Description automatically generated

Since p-value is lower than alpha, which is 0.05, The Shapiro-Wilk test concludes that the response is not normally distributed.

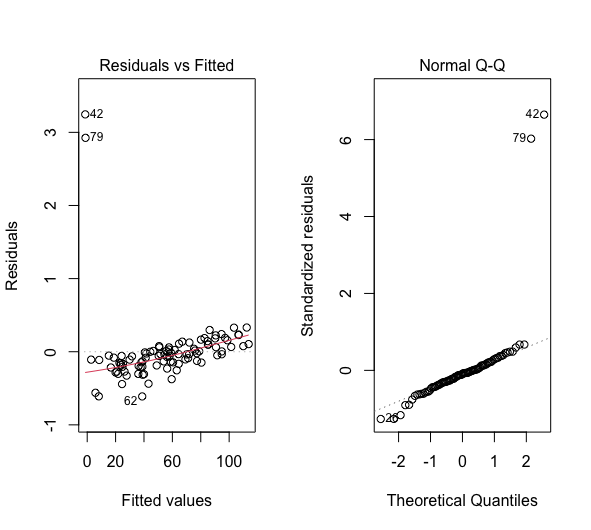
* To overcome non-normality, the Box-Cox Transformation was applied. Since λ is close to 0.5, the square root of the model was taken then the transformed data was able to be found.





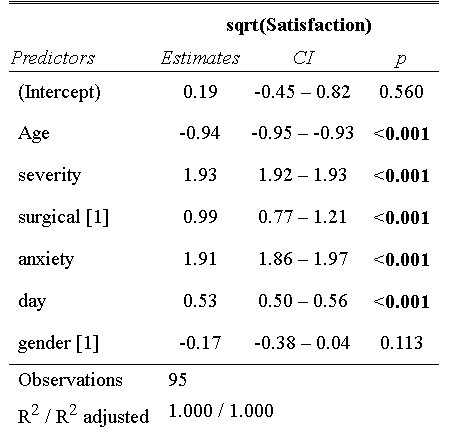
After the transformation the normality problem is solved.

* Then the plots are checked to see how they fit the model.



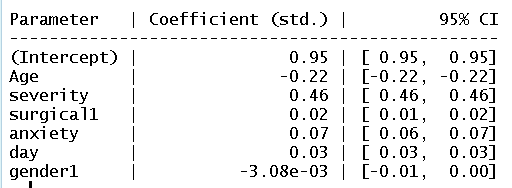
There is no pattern in the residuals vs fitted plot, so the residuals have the constant variance and the normal Q-Q plot does not have a problem.

* Summarizing the data by using summary function



The model has significant variables which are age, severity, surgical, anxiety and day. The gender variable is the only variable which is not significant.

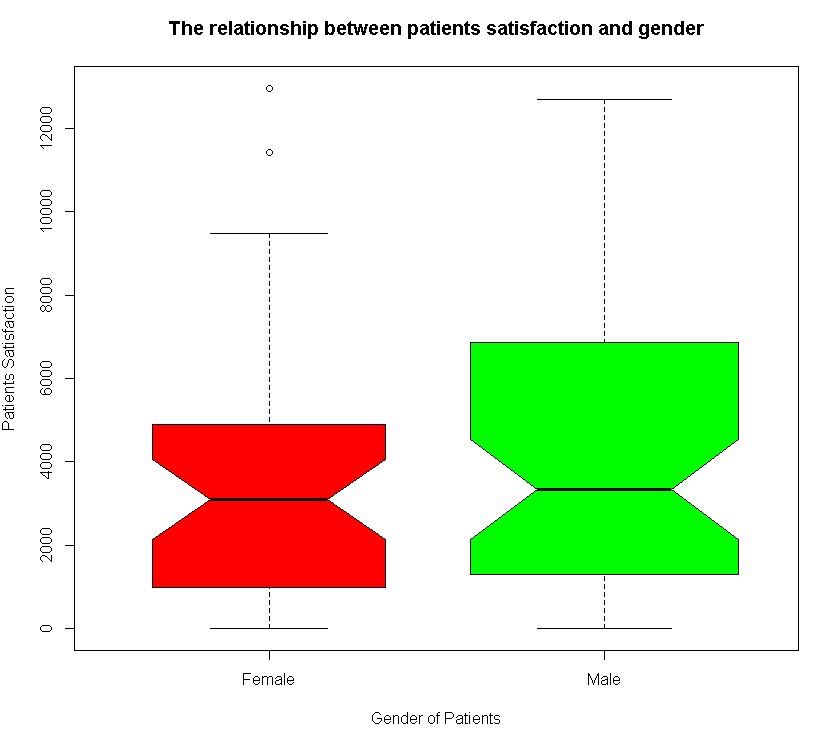
To understand which variables are contributed most, the model is standardized.



It can be seen in the graph that severity has the most massive contribution to the model. After that, age and the rest of the variables moderately follow the severity variable.

## Research Question 2: Does patient satisfaction differ among genders?

In this research question T-test is used. Patient satisfaction is taken as a response and genders as dependent.

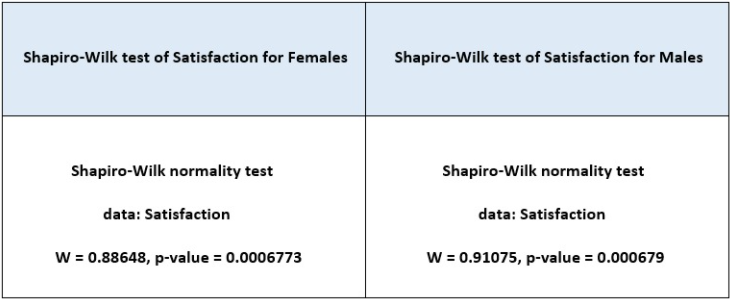


H0: Patient satisfaction does not differ among genders.

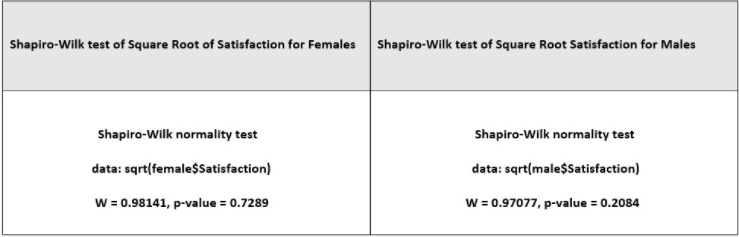
H1: Patient satisfaction differs for at least one age group.

To apply the T-test, assumptions are checked.

Firstly, the normality is checked.



As can be seen above there is a normality problem. To solve this normality problem, transformation is used.



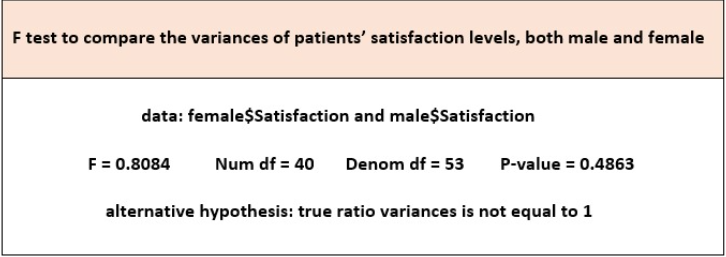
After transformation is used, the response variable becomes normal.

Second of all, in order to apply the t-test second assumption, which is the variances of patients’ satisfaction levels, both male and female should be checked.

H0: The variances of patient satisfactions for both genders are equal.

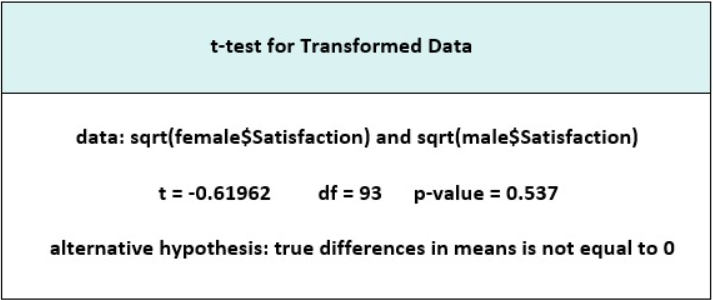
H1: The variances of patient satisfactions for both genders are not equal.

In order to check variances of the patients` satisfaction levels for both male and female, the F-test is used.

****

Since the p-value is bigger than 0.05, we do not reject H0 which means that the variances of patients` satisfaction levels are equal for males and females.

Hence, both assumptions are satisfied. The t-test is used to answer the question.



Since p-value is bigger than alpha which is 0.05, the null hypothesis is not rejected. Thus, it can be said that patients' satisfaction does not differ among genders.

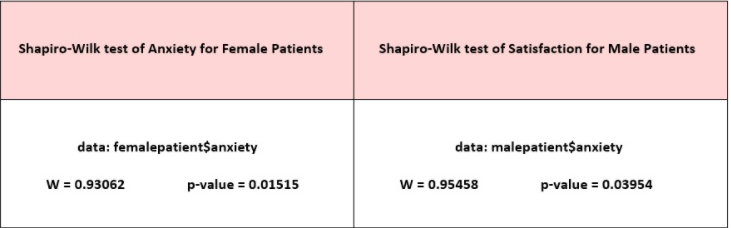
## Research Question 3: Do the anxiety levels of the patients have any relationship to their gender?

For this research question a non-parametric test is used. As a response the anxiety level is taken, for the covariate variable the gender is taken.

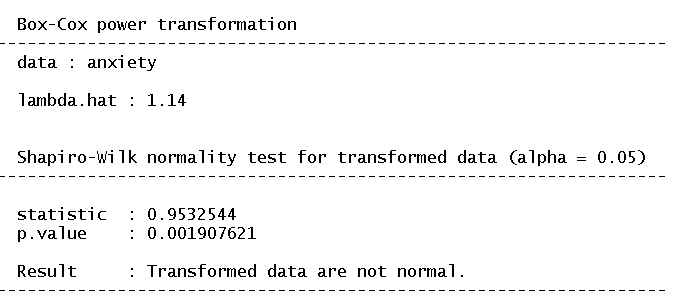
H0: Anxiety levels of patients do not differ among genders.

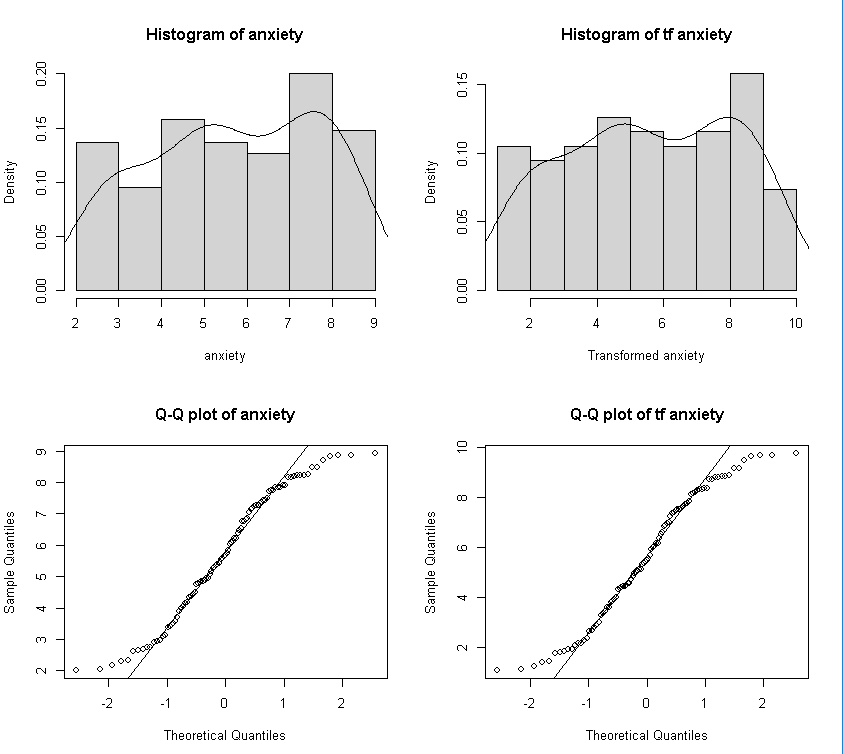
H1: Anxiety levels of patients differ among genders.

In order to test the difference between anxiety levels according to the patient's gender, a z-test can be applied. There are two assumptions to apply to the z-test that should be checked. To check the normality, the Shapiro-Wilk test is used.

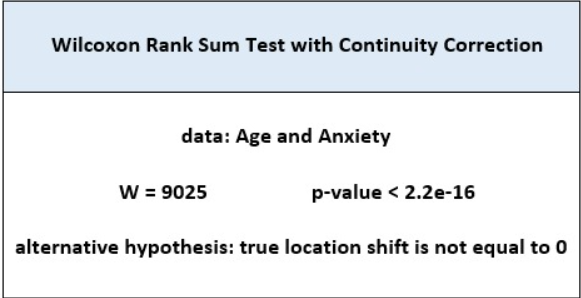


Since both p-values are smaller than 0.05, the data is not normally distributed. Then Box-Cox transformation is applied in order to make the data normal.





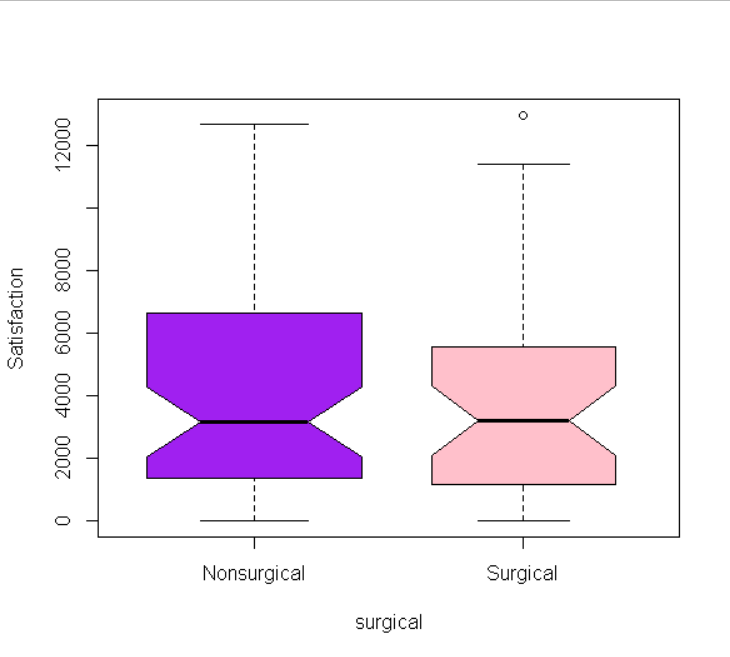
Even when the transformation is applied, the data is still not normalized and the Q-Q plot is not properly displayed. The non-parametric version of the z-test, which is the Wilcoxon Rank Sum test is used.

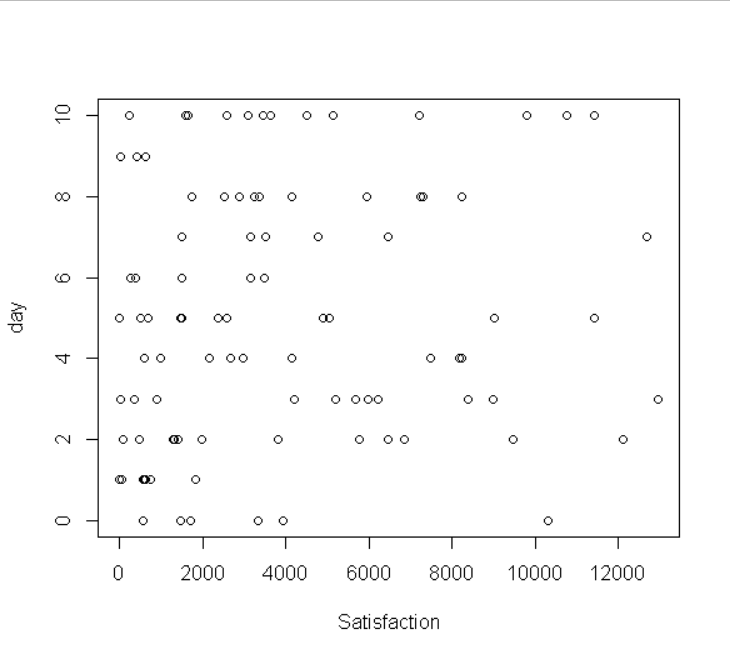


Since p-value is lower than 0.05, the null hypothesis is rejected. It can be concluded that the mean anxiety levels for males is not equal to the mean anxiety levels of females. Thus, it can be said that anxiety levels of patients differ among genders.

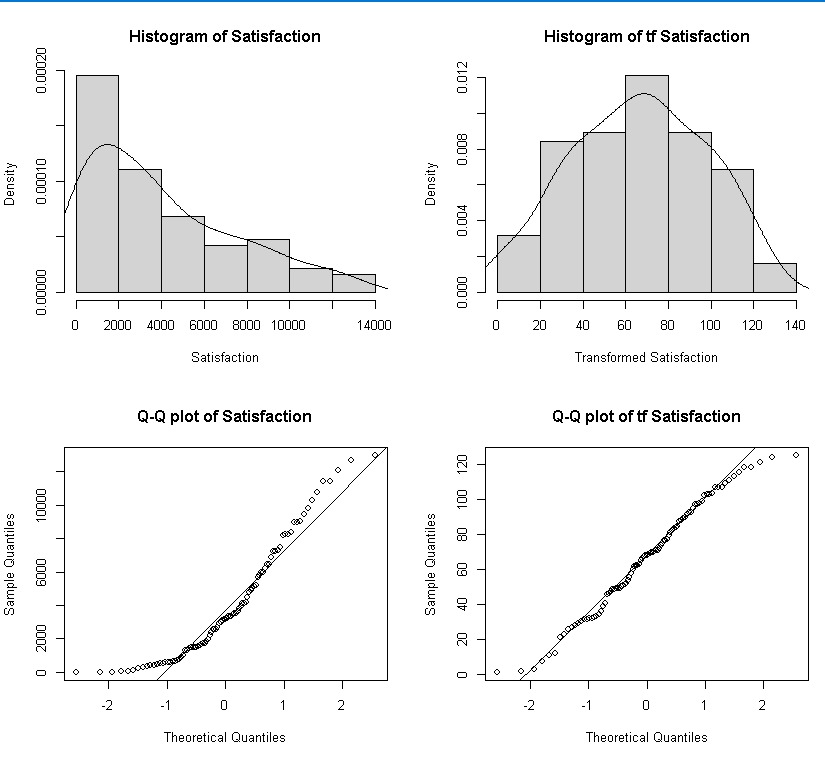
## Research Question 4: Does patient satisfaction fluctuate for the number of days the patient stays in the hospital or is the patient a surgical patient?

In order to find out the effect of the number of days the patient stays in the hospital or surgical patients on satisfaction, multiple linear regression is conducted. In multiple linear regression, the patient satisfaction is taken as a dependent variable and the number of days the patient stays in the hospital and surgical patients are taken as independent variables.

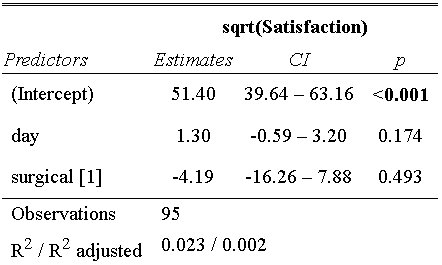




It's known that the satisfaction variable is not normally distributed. Because of that, the Box-Cox transformation is applied.



It seems that the transformation fits the data well when the histogram of the response and Q-Q plot is checked. It can be concluded that the day and the surgical variables do not play an important role in patients` satisfaction since the coefficients of the model shows that they are insignificant.



The transformed model gives insignificant estimates for both variables. So, the patience satisfaction does not differ for the number of days the patient stays in the hospital or is a surgical patient.

# 4. Discussion/Conclusion

To begin with, out of 6 variables, 5 have an important relationship which are age, severity, surgical, anxiety and the number of days the patient stays in the hospital with the satisfaction level. The most contributed variable to the satisfaction level is severity. The patient satisfaction does not differ among genders. On the other hand, the anxiety level of patients seems to differ among both males and females. Patient satisfaction does not fluctuate for the number of days the patient stays in hospital or is the patient a surgical patient.

# References

No references were used in the analysis of this data.