Indian Liver Patient Records

Wilson Moreno

1/8/2020

# Introduction

# Methods/Analysis

## Step 0: require package

if (!require(package)) install.packages('psych', repos = "http://cran.us.r-project.org")  
if (!require(package)) install.packages('knitr', repos = "http://cran.us.r-project.org")  
if (!require(package)) install.packages('ggplot2', repos = "http://cran.us.r-project.org")  
library(knitr)  
library(ggplot2)  
library(psych)

## **Step 1:** Load the data base

database <- read.csv("Data/indian\_liver\_patient.csv")

## **Step 2:** Exploratory Data Analysis

### Summay Statistics

The following table shows the descriptive statistics for all the variables in the database.

round(data.frame( describeBy(database, digits= 2)),1)

## Warning in describeBy(database, digits = 2): no grouping variable requested

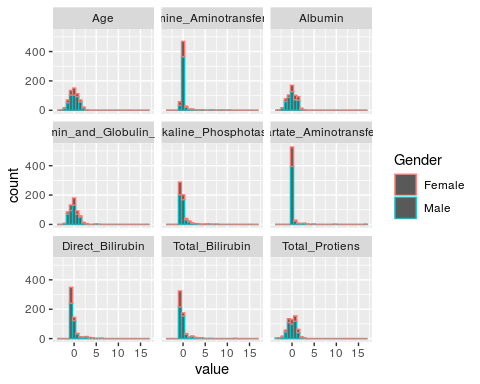
## vars n mean sd median trimmed mad min max  
## Age 1 583 44.7 16.2 45.0 44.8 17.8 4.0 90.0  
## Gender\* 2 583 1.8 0.4 2.0 1.8 0.0 1.0 2.0  
## Total\_Bilirubin 3 583 3.3 6.2 1.0 1.7 0.4 0.4 75.0  
## Direct\_Bilirubin 4 583 1.5 2.8 0.3 0.7 0.3 0.1 19.7  
## Alkaline\_Phosphotase 5 583 290.6 242.9 208.0 238.4 74.1 63.0 2110.0  
## Alamine\_Aminotransferase 6 583 80.7 182.6 35.0 43.9 22.2 10.0 2000.0  
## Aspartate\_Aminotransferase 7 583 109.9 288.9 42.0 56.8 31.1 10.0 4929.0  
## Total\_Protiens 8 583 6.5 1.1 6.6 6.5 1.0 2.7 9.6  
## Albumin 9 583 3.1 0.8 3.1 3.1 0.9 0.9 5.5  
## Albumin\_and\_Globulin\_Ratio 10 579 0.9 0.3 0.9 0.9 0.3 0.3 2.8  
## Dataset 11 583 1.3 0.5 1.0 1.2 0.0 1.0 2.0  
## range skew kurtosis se  
## Age 86.0 0.0 -0.6 0.7  
## Gender\* 1.0 -1.2 -0.6 0.0  
## Total\_Bilirubin 74.6 4.9 36.7 0.3  
## Direct\_Bilirubin 19.6 3.2 11.2 0.1  
## Alkaline\_Phosphotase 2047.0 3.7 17.5 10.1  
## Alamine\_Aminotransferase 1990.0 6.5 50.0 7.6  
## Aspartate\_Aminotransferase 4919.0 10.5 149.1 12.0  
## Total\_Protiens 6.9 -0.3 0.2 0.0  
## Albumin 4.6 0.0 -0.4 0.0  
## Albumin\_and\_Globulin\_Ratio 2.5 1.0 3.2 0.0  
## Dataset 1.0 0.9 -1.1 0.0

# tmp <- describeBy(database,  
# group = database$Gender,  
# digits= 1)

### Visualization

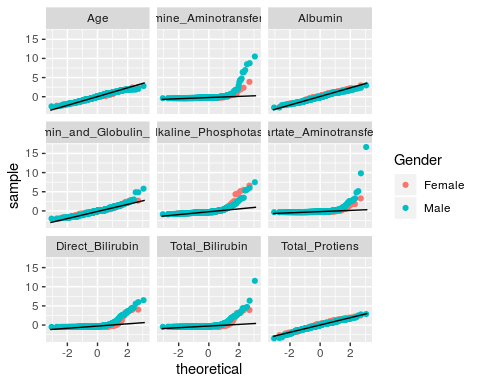
As can be seen in the visual analysis of the data, they were divided into 2 groups by gender, in all the graphs the rates are much higher in women than in men, and tend to follow the same distribution. Variables: Age, Albumin, Albumin\_and\_Globulin\_Ratio and Total\_Protiens, are suspected to follow a normal distribution.

scale\_database <- database %>% select(-Gender,-Dataset) %>% scale() %>% as.data.frame() %>%   
 cbind(Gender = database$Gender)  
  
database.gathered <- scale\_database %>% as.data.frame() %>%   
 gather(key = "variable", value = "value", - Gender)  
  
ggplot(data = database.gathered , mapping = aes(x = value, color = Gender)) +  
 geom\_histogram() +  
 facet\_wrap(facets = vars(variable ))



This graph certifies the suspicion that the aforementioned variables follow a normal distribution.

ggplot(data = database.gathered , mapping = aes(sample = value, color = Gender)) +  
 stat\_qq() + stat\_qq\_line(color = "black") +  
 facet\_wrap(facets = vars(variable ))



# Results

### Summay Statistics

### Visualization

# Conclusion

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars)

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.