**Probability Project**

**Statistical Analysis of Dataset:**

**Dataset Description:**

Dataset is based on KRK problem (King-rook-king), its basically the endgame position of chess with three pieces on board; White king, white rook, black king where **white** has to checkmate **black** with its remaining pieces that are **king** and **rook.**

**Number of Attributes:** 7

**Attribute information:**

* **WKF:** white king file (column)
* **WKR:** white king rank (row)
* **WRF:** white rook file
* **WRR:** white rook rank
* **BKF:** black king file
* **BKR:** black king rank
* **Result:** draw (game will be draw), #moves (**win** in number of moves mentioned)

**Instances:** 28056

**Associated Tasks:** Classification

**Code & Analysis:**

#Reading File

df<-read.csv("krkopt.data")

#displaying table

table(df$result)

#sum

sum(table(df$result))

#maximum element

max(table(df$result))

#minimum element

min(table(df$result))

#calculating baseline accuracy based on most frequent class occurring

ba<-(ma/s)\*100

ba

#Function to Calculate frequency distribution table

freqdist=function(x, freqorder=F)

{

counts = table(x)

n = sum(counts)

if(freqorder) ord=order(-counts)

else ord=1:length(counts)

data.frame(

row.names = row.names(counts[ord]),

Count=as.vector(counts[ord]),

Percent=100\*as.vector(counts[ord])/n,

CumCount=cumsum(as.vector(counts[ord])),

CumPercent=100\*cumsum(as.vector(counts[ord]))/n

)

}

#Frequency Distribution Graph

freqdist(df$result)

table(df$result)

#mean

mean(table(df$result))

#median

median(table(df$result))

#Variance

var(table(df$result))

#Standard Deviation

sd(table(df$result))

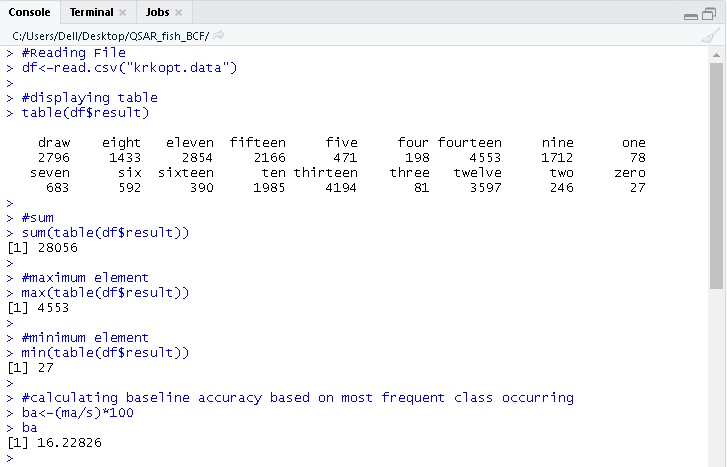
#bar chart

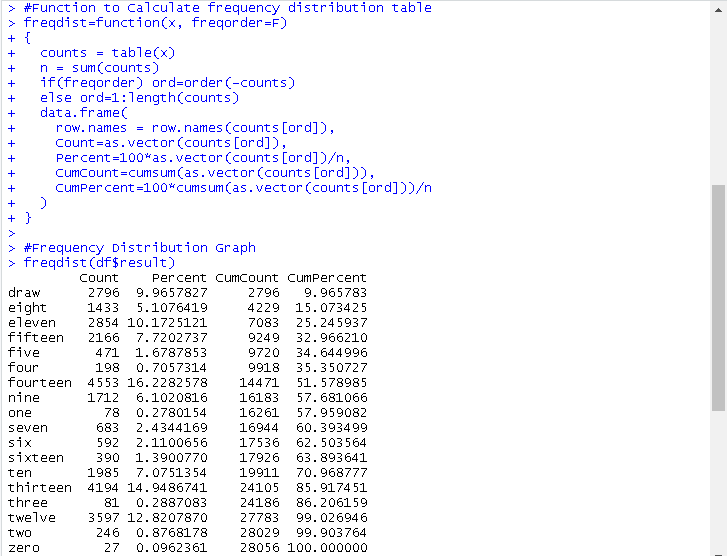
barplot(table(df$result))

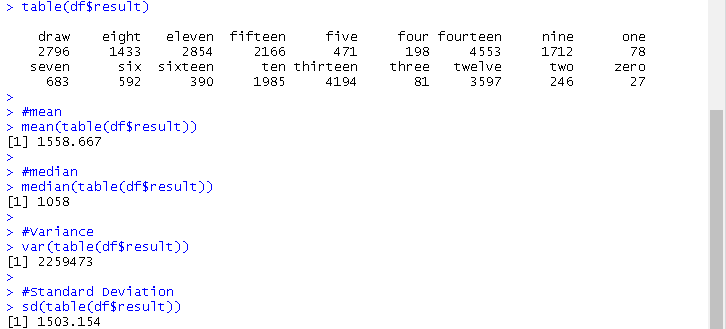
#histogram

hist(table(df$result),10)

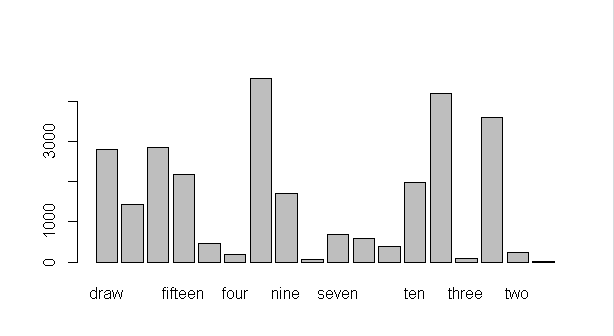
**Console Screen:**



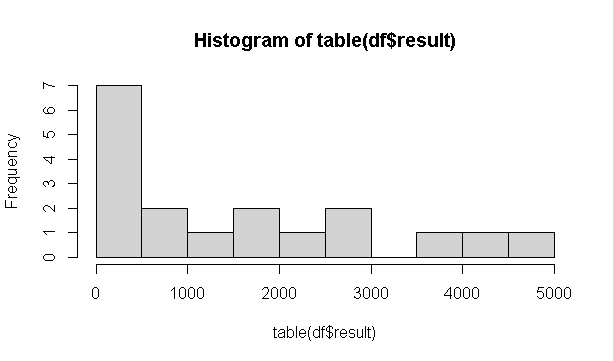




**Bar Chart / Bar Plot:**



**Histogram:**



**Conclusion & discussion:**

This analysis will help in machine learning methods to devise algorithms which will determine which moves to play in given situation. Providing the basis for such an endeavor, the **King-Rook-King** dataset consists of **28,056** board positions with their respective classification as we can see above The classification, the attribute of ultimate interest, is given either as the minimum number of moves for White to win **(0 to 16 moves)**, or alternately as “**draw**.” The board positions are not uniformly distributed across these 18 classes. The 2 distribution ranges from **27** board positions for the smallest class up to **4,553** for the most frequent class. This suggests a baseline accuracy of **16.2%** by guessing the most frequent class. If we revalue this analysis in terms of **frequency distribution table** we will find in relative frequency percentage or percentagethat it invokes more interesting assumptions I-e shows that there is **16.2% chance** that whenever this **KRK(King-rook-King)** situation or problem will occur, the positioning of pieces will fall into the category/class distribution of **Fourteen moves**. Similarly, for other distributions chances can be seen in **Percentage** column of **Frequency distribution table**.

­