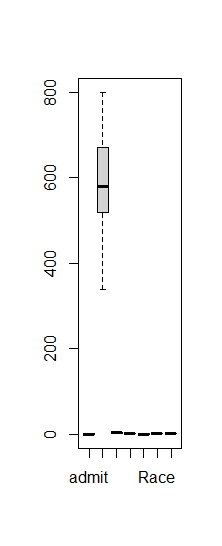
**Project : College Admissions**

######################importing the data####################################

college =read.csv('college\_admission.csv')

str(college)

summary(college)

is.na(college)

sum(is.na(college))

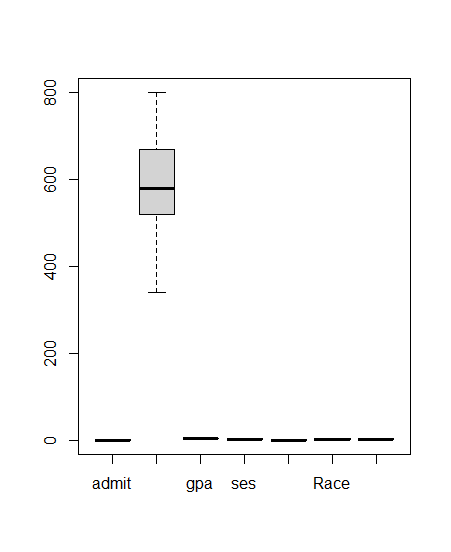
college$gre <- as.numeric(college$gre)

str(college)

boxplot(college, main= "Checking the outliers", horizontal = TRUE)$out

#########################removing the outliers################################

boxplot(college$gre, plot=FALSE)$out



outliers <- boxplot(college$gre, plot=FALSE)$out

outliers

x<- college

x<- x[-which(x$gre %in% outliers),]

x

boxplot(college$gpa, plot=FALSE)$out

outliers1 <- boxplot(x$gpa, plot=FALSE)$out

x1<- x

x1<- x1[-which(x1$gpa %in% outliers1),]

boxplot(x1)

#########################transforming the data################################

str(x1)

x1$admit <- as.factor(x1$admit)

x1$ses <- as.factor(x1$ses)

x1$Gender\_Male <- as.factor(x1$Gender\_Male)

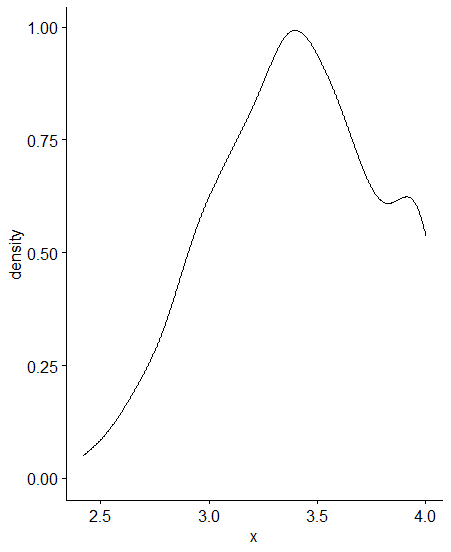
x1$Race <- as.factor(x1$Race)

x1$rank <- as.factor(x1$rank)

str(x1)

##########################Normal Distribution#################################

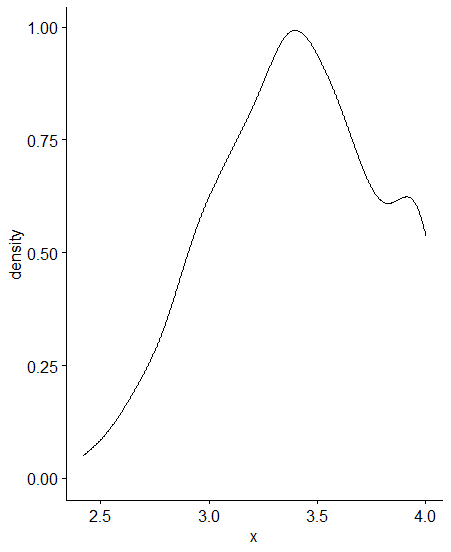
install.packages('dplyr')

install.packages('ggpubr')

library(dplyr)

library(ggpubr)

ggdensity(x1$gre, main= "Gre & Normality", xlab= 'gre')

ggdensity(x1$gpa)

shapiro.test(x1$gre)

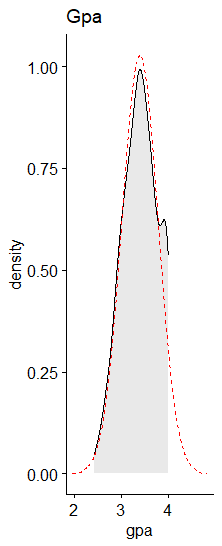
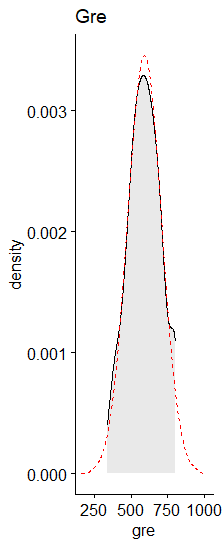
shapiro.test(x1$gpa)

ggdensity(x1, x = "gre", fill = "lightgray", title = "Gre") +

stat\_overlay\_normal\_density(color = "red", linetype = "dashed")

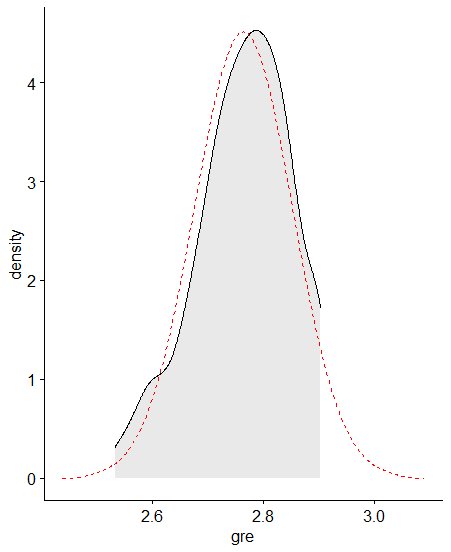
ggdensity(x1, x = "gpa", fill = "lightgray", title = "Gpa") +

stat\_overlay\_normal\_density(color = "red", linetype = "dashed")



skewness(x1$gre, na.rm = TRUE)

skewness(x1$gpa, na.rm= TRUE)

x1$gre <- log10(x1$gre)

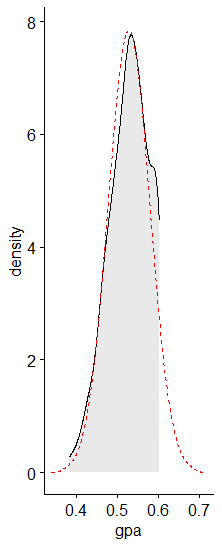
n1 <- log10(max(x1$gre+1)-x1$gre)

x1$gpa <- log10(x1$gpa)

n2 <- log10(max(x1$gpa+1)-x1$gpa)

ggdensity(x1,x= 'gre', fill = 'lightgray') +

stat\_overlay\_normal\_density(color = "red", linetype = "dashed")

ggdensity(x1,x= 'gpa', fill = 'lightgray') +

stat\_overlay\_normal\_density(color = "red", linetype = "dashed")

library('moments')

install.packages('moments')

skewness(x1$gre, na.rm= TRUE)

skewness(x1$gpa, na.rm = TRUE)

shapiro.test(x1$gre)

shapiro.test(x1$gpa)

###############################logistic model################################

install.packages('caret')

library(caret)

install.packages('dplyr')

library(dplyr)

install.packages('tidyverse')

library(tidyverse)

library(caTools)

install.packages('caTools')

set.seed(123)

intrain <- sample(nrow(x1), 0.70\*nrow(x1))

train <- x1[intrain, ]

test <- x1[-intrain]

logistic <- glm(formula = admit ~ ., data= train, family = 'binomial' )

summary(logistic)

#########################dropping insignifcant variables#####################

df <- x1[ -c(4:6) ]

logistic1 <- glm(formula = admit ~ ., data = df, family = 'binomial')

summary(logistic1)

1-pchisq(494.62,394)

1-pchisq(453.87,389)

1-pchisq(494.62-453.87, 394-389)

p1<-predict(logistic1, newdata = df)

p1

#########################Decision Tree######################################

install.packages('e1071')

install.packages('rpart')

install.packages('mlbench')

install.packages('plyr')

install.packages('caret')

library(e1071)

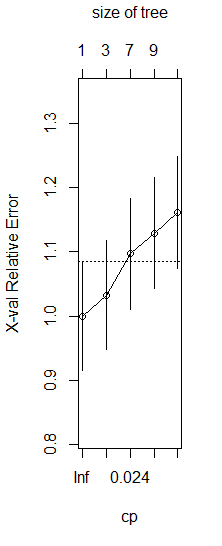
library(rpart)

library(mlbench)

library(plyr)

library(caret)

###############################building the model############################

tree\_model <- rpart(admit~., data= train, method= 'class')

tree\_model

summary(tree\_model)

plotcp(tree\_model)

new <- svm(admit~., train)

confusionMatrix(train$admit, predict(new), positive = '1')