```r
#Carregamento de bibliotecas

library(MASS)

library(class)


help(biopsy)


# sumário estatístico

summary(biopsy)


# Variáveis com outliers V4, V5, V7, V8, V9

par(mfrow=c(1,5))

boxplot(biopsy$V4)

boxplot(biopsy$V5)

boxplot(biopsy$V7)

boxplot(biopsy$V8)

boxplot(biopsy$V9)


str(biopsy);


new_biopsy = biopsy;


#is.na(biopsy$class);


#is.na(biopsy$v6);


# Será substituído pela média da variável

#new_biopsy[is.na(biopsy$V6),]$V6 <- round(mean(biopsy$V6, na.rm = TRUE), 0);


# Eliminação de linhas com NA

new_biopsy <- new_biopsy[complete.cases(new_biopsy),]
```

```r
# Eliminação da coluna com NA
#new_biopsy <- new_biopsy[, c(2:6, 8:11)]


#is.na(new_biopsy$v6);


# 1/3 de dados de teste
L <- sample(1:nrow(new_biopsy), round(nrow(new_biopsy)/3));
test <- new_biopsy[L, 2:10];
#test <- new_biopsy[L, 1:8];


# Restantes dos dados são de treinamento
train <- new_biopsy[-L, 2:10];
#train <- new_biopsy[-L, 1:8];


# classificação
classe <- factor(new_biopsy[-L, 11]);
#classe <- factor(new_biopsy[-L, 9]);


# Execução com outliers
acc = c(1:100) * 0


for (x in 1:100){
  fit <- knn(train, test, classe, k = 5);
  conf_matrix <- table(fit[1:length(L)], factor(new_biopsy[L, 11]));
  accuracy <- sum(diag(conf_matrix))/sum(conf_matrix) * 100;
  acc[x] = accuracy
  cat('Accuracy: ', accuracy, '% | Exec = ', x)
  print(conf_matrix);
  print('=======================');
}
```

```r
cat('Mean Accuracy - outliers: ', mean(acc), '%')




# Execução sem outliers

#acc = c(1:100) * 0


#for (x in 1:100){

#  fit <- knn(train, test, classe, k = 2);

#  conf_matrix <- table(fit[1:length(L)], factor(new_biopsy[L, 11]));

#  accuracy <- sum(diag(conf_matrix))/sum(conf_matrix) * 100;

#  acc[x] = accuracy

#  cat('Accuracy: ', accuracy, '% | Exec = ', x)

#  print(conf_matrix);

#  print('=======================');

#}


#cat('Mean Accuracy - no outliers: ', mean(acc), '%')
```