UNIVERSIDADE FEDERAL DA PARAÍBA

DEPARTAMENTO DE COMPUTAÇÃO CIENTIFICA

CENTRO DE INFORMÁTICA

THOMAS RIBEIRO DE ARAÚJO

**ANÁLISE DE DADOS NA LINGUAGEM DE PROGRAMAÇÃO R: Transporte aéreo – horas voadas – voos internacionais**

JOÃO PESSOA

2019

THOMAS RIBEIRO DE ARAÚJO

**ANÁLISE DE DADOS: Transporte aéreo - horas voadas - voos internacionais - Hora - Agência Nacional de Aviação Civil (Anac)**

Relatório apresentado ao Curso de Bacharelado em Matemática Computacional da Universidade Federal da Paraíba como requisito parcial para a conclusão da disciplina de Cálculo das probabilidades l, sob orientação da Profa Andrea Vanessa Rocha.

JOÃO PESSOA

2019

Neste breve relatório demonstrarei o programa desenvolvido para calcular os dados estatísticos, e realizarei um comentário sobre os dados obtidos.

Utilizei uma base de dados extraída do Instituto de Pesquisa Econômica Aplicada (Ipea), disponível > macroeconômico > temas > transporte > Anac > Transporte aéreo – horas voadas – voos internacionais.

Scripty em R:

#importando planilha no formato excel

library(readxl)

#ler arquivo em excel

ipeadatavoos <- read\_excel("ipeadatavoos.xlsx")

#mostrando ao usuario a tabela em uma melhor exibição

View(ipeadatavoos)

#exibindo alguns valores básicos

summary(ipeadatavoos)

# Armazenando valores em variaveis para cálculos estátisticos

# soma

soma <- sum(ipeadatavoos[2])

soma

#tamanho de linhas de dados da tabela

tamanhotable <- length(ipeadatavoos$Data)

tamanhotable

#media

media <- soma/tamanhotable

media

#mediana

mediana <- median(ipeadatavoos$`Transporte aéreo - horas voadas - voos internacionais - Hora - Agência Nacional de Aviação Civil (Anac) - ANAC\_HVINT -`)

mediana

#desviopadrão

desviopadrao <- sd(ipeadatavoos$`Transporte aéreo - horas voadas - voos internacionais - Hora - Agência Nacional de Aviação Civil (Anac) - ANAC\_HVINT -`)

desviopadrao

#variança

varianca <- sqrt(desviopadrao)

varianca

# valor maximo

valmax <- max(ipeadatavoos$`Transporte aéreo - horas voadas - voos internacionais - Hora - Agência Nacional de Aviação Civil (Anac) - ANAC\_HVINT -`)

valmax

#valor minimo

valmim <- min(ipeadatavoos$`Transporte aéreo - horas voadas - voos internacionais - Hora - Agência Nacional de Aviação Civil (Anac) - ANAC\_HVINT -`)

valmim

#coeficiente de variação

cvar <- (desviopadrao/media)\*100

cvar

#Moda

Moda <- "AMODAL"

Moda

#Quartis

quartis <- quantile(ipeadatavoos$`Transporte aéreo - horas voadas - voos internacionais - Hora - Agência Nacional de Aviação Civil (Anac) - ANAC\_HVINT -`)

quartis

#exibindo gráfico de histograma

hist(ipeadatavoos$`Transporte aéreo - horas voadas - voos internacionais - Hora - Agência Nacional de Aviação Civil (Anac) - ANAC\_HVINT -`)

#Observando a frequência de horas de voos absolutas

freq <- table (ipeadatavoos$`Transporte aéreo - horas voadas - voos internacionais - Hora - Agência Nacional de Aviação Civil (Anac) - ANAC\_HVINT -`)

freq

#tabela de frequência relativas

freq\_rel <- prop.table(freq)

freq\_rel

#porcentagem (100\*freq\_rel\_table)

porcentagem\_freq\_rel <- 100\* prop.table(freq\_rel)

porcentagem\_freq\_rel

#frequencia acumulada

freq\_ac <- cumsum(ipeadatavoos$`Transporte aéreo - horas voadas - voos internacionais - Hora - Agência Nacional de Aviação Civil (Anac) - ANAC\_HVINT -`)

freq\_ac

#inserir dados na tabela e criar linha de total

freq <- c(freq,sum(freq))

freq\_rel <- c(freq\_rel, sum(freq\_rel))

porcentagem\_freq\_rel <- c(porcentagem\_freq\_rel, sum(porcentagem\_freq\_rel))

names(freq)[43] <- "TOTAL"

#variação das horas de voo

title(main = "Horas de voo por ano")

barplot(ipeadatavoos$`Transporte aéreo - horas voadas - voos internacionais - Hora - Agência Nacional de Aviação Civil (Anac) - ANAC\_HVINT -`, names.arg = ipeadatavoos$Data)

#tabela final, organizando por coluna, e usando um arredondamento de valores"round"

tabelaf <- cbind (freq, freq\_rel= round(freq\_rel,digits = 2), porcentagem\_freq\_rel= round(porcentagem\_freq\_rel, digits = 2),freq\_ac= round(freq\_ac, digits = 2))

tabelaf

Resumo:

Ao calcular a média obtivemos um resultado de 134.039 horas de voo, entre os anos de 1969-2010. Porém esse dado não foi muito significante, levando em consideração a discrepância entre horas de voo e ano.

Em caso da avaliadora não conseguir importar os dados da tabela, coloque o arquivo excel na pasta que indica ao executar o comando >getwd(). Saíra o endereço onde deve está o arquivo.

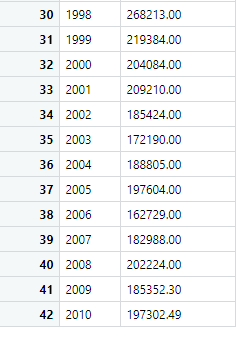
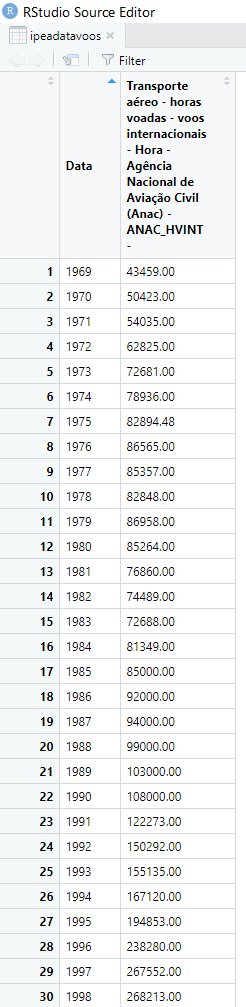
Com isso precisamos realizar outras medidas para exibir um dado mais exato. Comprovando isso mostramos o desvio padrão que varia de 63.580 horas, um número muito alto em relação aos dados, pois temos um valor máximo de 269.213 e um valor mínimo de 43.459 horas de voo. Além disso, a variância demostra a falta de uniformidade nos dados analisados, com o valor de 252.151.

Exibindo um quartil, conseguimos mostrar a dispersão dos dados analisados.



A seguir será mostrado gráficos e tabelas sobre a base estudada.

Mostrando a base de dados trabalhada:



Mostrando o histograma sobre a base de dados trabalhada:

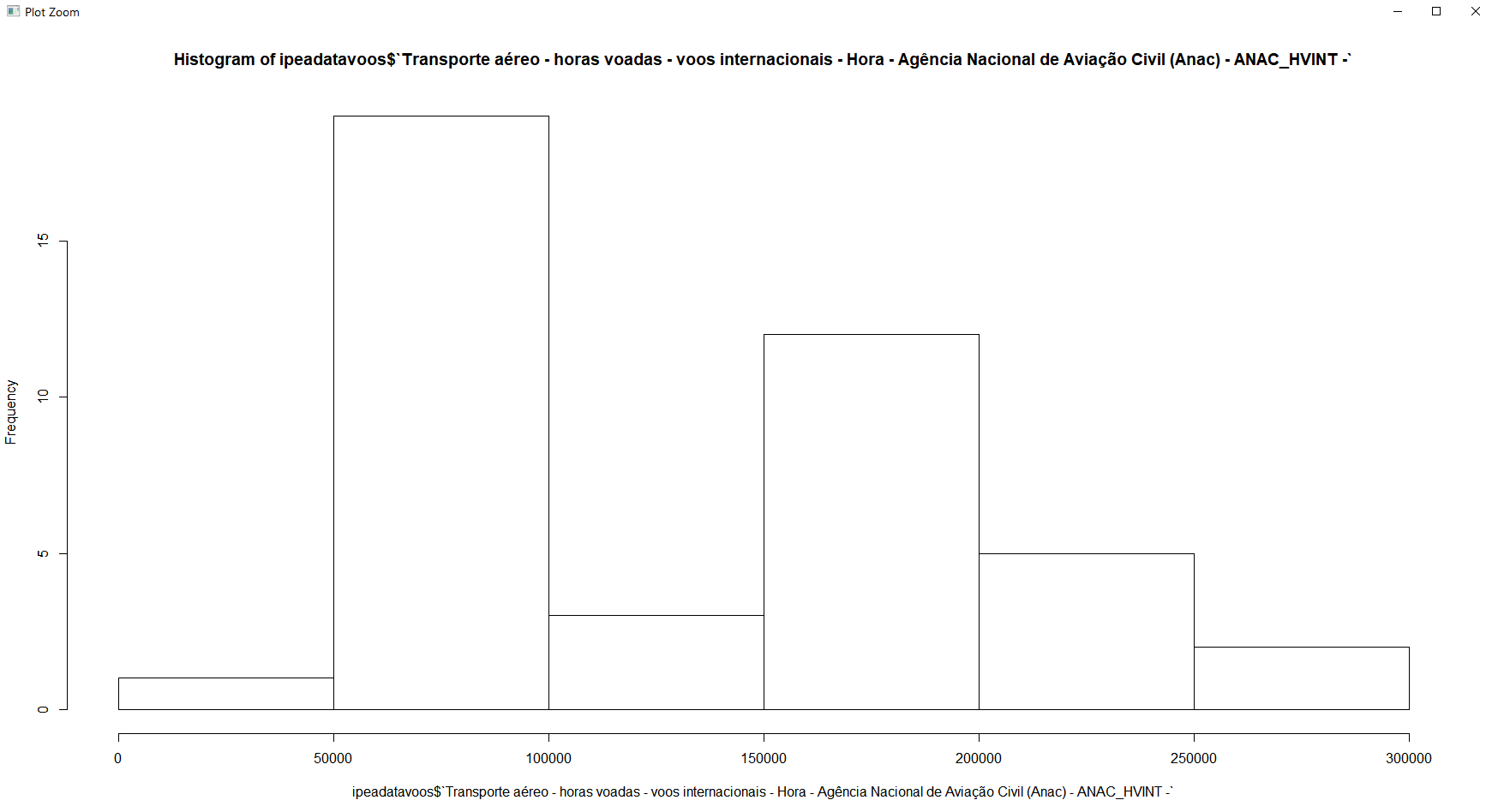
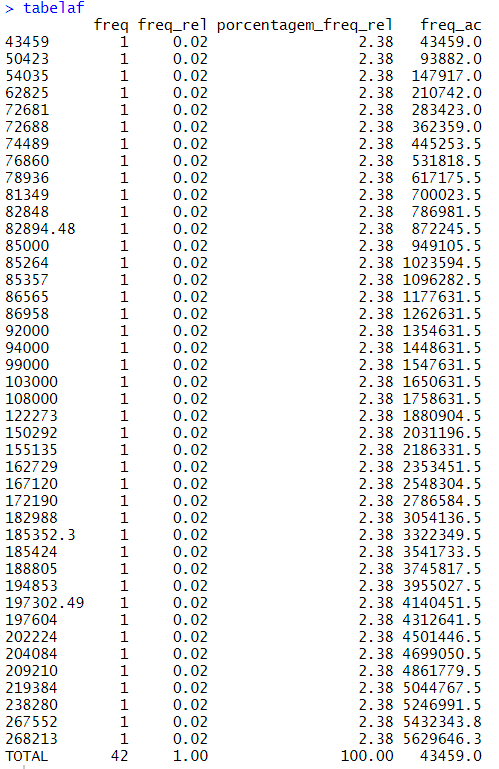
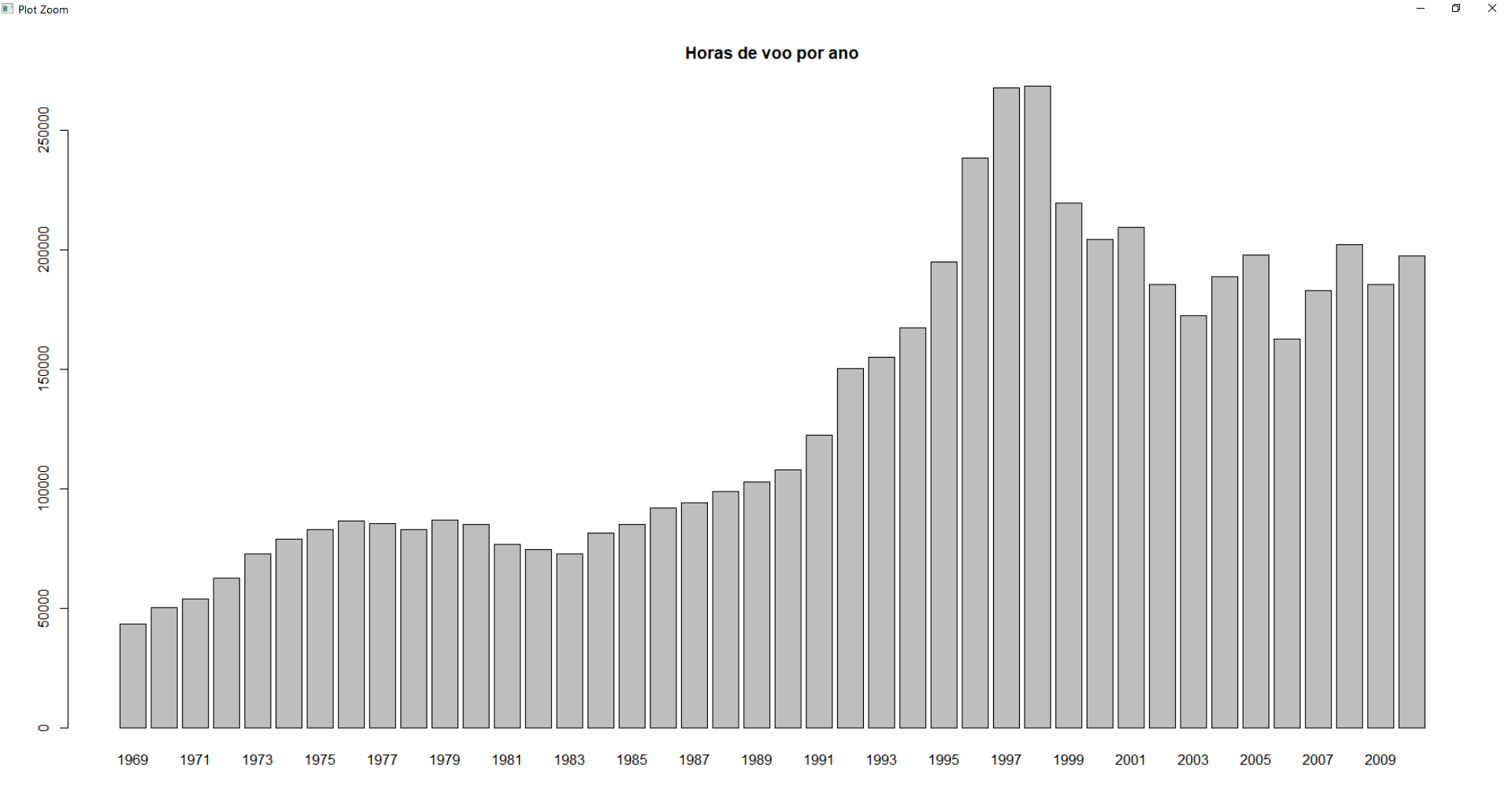


Tabela frequência:



Variação de horas de voo com o ano:



Referências

Instituto de Pesquisa Econômica Aplicada - IPEA, (2012). **Horas voadas em voos regulares internacionais**. Disponível em: <<http://www.ipeadata.gov.br/Default.aspx>>, acessado em: 01 de julho de 2019.

Azevedo A.F. **ANALISE DE DADOS COM SOFTWARE R:** MÉTODOS ESTATISTICOS,COMPUTACIONAIS E ECONOMETRICOS. Disponivel em < <http://rstudio-pubs-static.s3.amazonaws.com/7342_3aee84b4bc9549adb3080f06c69174e1.html>> acessado em> 02 de julho de 2019.