

```

## Trabalho
#
# Script de exemplo para o trabalho final de bioestatística 2022.2
# - Gráficos
# - Testes estatísticos
# - Análise de regressão linear
# - Análise de regressão logística
#
# Eric Kauati Saito
# 28/10/2022

# Carregar dados do arquivo

dados <- read.table(file="D:\\Macao Disciplinas\\2022\\R\\healthcare-
dataset-stroke-data.csv", sep=";", header=T)

# Trabalhando com os dados
## Primeiras análises
dados_reduzidos = dados[1:400,]

glicose = dados_reduzidos$avg_glucose_level

### Estatísticas
mean(dados_reduzidos$avg_glucose_level)
var(dados_reduzidos$avg_glucose_level)
sd(dados_reduzidos$avg_glucose_level)

media_glicose = mean(dados_reduzidos$avg_glucose_level)

summary(dados_reduzidos$avg_glucose_level) #5 estatísticas

## Gráficos
### Gráfico de Barras
table(dados_reduzidos$hypertension, dados_reduzidos$gender)
barplot(table(dados_reduzidos$hypertension, dados_reduzidos$gender), beside
= TRUE)

### Histograma
hist(dados_reduzidos$avg_glucose_level)
hist(dados_reduzidos$avg_glucose_level, breaks = 6)
hist(dados_reduzidos$avg_glucose_level, freq = TRUE)
hist(dados_reduzidos$avg_glucose_level,
     freq = TRUE,
     ylim = c(0, 60),
     main = "Histograma do nível médio de glicose",
     xlab="Glicose (mg/dL)",
     ylab = "Frequência Relativa (%)")
)

### Boxplot
# Boxplot do nível de glicose em relação a hipertensão
boxplot(dados_reduzidos$avg_glucose_level ~ dados_reduzidos$hypertension,
        main = "Nível de glicose e Hipertensão",
        ylab = "Nível de glicose (mg/dL)",
        xlab= "Hipertensão")

boxplot(dados_reduzidos$avg_glucose_level ~ dados_reduzidos$hypertension,
        main = "Nível de glicose e Hipertensão",

```

```

        ylab = "Nível de glicose (mg/dL)",
        xlab= "Hipertensão",
        names = c('Não', 'Sim'),
        col = c('orange','blue')
    )

## Testes estatísticos

### Intervalo de confiança
n_hipertenso = dados_reduzidos[dados_reduzidos$hypertension==0,]
hipertenso = dados_reduzidos[dados_reduzidos$hypertension==1,]

t.test(hipertenso$avg_glucose_level, conf.level = 0.9) #Intervalo de
confiança para 90%

### teste t
t.test(hipertenso$avg_glucose_level, n_hipertenso$avg_glucose_level,
alternative = "two.sided") #less, greater

### teste t pareado
# t.test(hipertenso$avg_glucose_level, n_hipertenso$avg_glucose_level,
alternative = "two.sided", paired = TRUE)

### Wilcoxon rank sum test
wilcox.test(hipertenso$avg_glucose_level, n_hipertenso$avg_glucose_level,
alternative = "two.sided")

### ANOVA
#bmi -> erro
oneway.test(age ~ smoking_status, data=dados_reduzidos, var.equal=T)
#### post-hoc
model <- aov(age ~ smoking_status, data=dados_reduzidos)
TukeyHSD(model, conf.level=.95)

### Teste para 2 proporções
tabela = table(dados_reduzidos$hypertension,
dados_reduzidos$Residence_type)
n_hipertensos = c(41,42)
residencia = c(194,206) #41+153, 42+164

prop.test(n_hipertensos, residencia, alternative = "two.sided",correct =
FALSE)

barplot(tabela, beside = TRUE,
        legend = c("Pressão normal","Pressão alta"),
        ylim=c(0,200))

## Análise de Regressão
idade = c(21,19,27,23,20,31,30,37,36,40,39)
tempo = c(96,106,98,110,140,116,109,112,118,113,127)
### Análise de Regressão Linear

#### Gráfico de dispersão
plot(idade, tempo,
     ylim = c(85, 145),
     xlim = c(15, 45),
     main = "Tempo de reação ao estímulo",
     xlab="Idade (ano)",

```

```

        ylab = "Tempo (s)")

boxplot(tempo)

tempo2 = tempo[-5]
idade2 = idade[-5]
#tabela = tabela[-c(5,7)]

remover = (tempo == 140)&(idade==20)
tempo2 = tempo[!remover]
idade2 = idade[!remover]
#tabela = tabela[!remover]

plot(idade2, tempo2,
      ylim = c(85, 145),
      xlim = c(15, 45),
      main = "Tempo de reação ao estímulo",
      xlab = "Idade (ano)",
      ylab = "Tempo (s)")

#### Modelo
modelo = lm(tempo2~idade2) #Tempo em relação a idade
#modelo = lm(avg_glucose_level ~ age, data = dados_reduzidos)

plot(idade2, tempo2,
      ylim = c(85, 145),
      xlim = c(15, 45),
      main = "Tempo de reação ao estímulo",
      xlab = "Idade (ano)",
      ylab = "Tempo (s)")
abline(modelo) #Adiciona a reta do modelo de regressão linear

summary(modelo)

cor(idade2,tempo2) #Correlação de Pearson

### Análise de Regressão Logística

plot(dados_reduzidos$age, dados_reduzidos$heart_disease,
      xlab = "Idade (ano)",
      ylab = "Doença Cardíaca")

modelo_log = glm(heart_disease ~ age, data = dados_reduzidos, family =
binomial)
summary(modelo_log)

dados_plot = data.frame(age=seq(0, 80))
dados_plot$heart_disease = predict(modelo_log, dados_plot, type =
'response')

plot(dados_reduzidos$age, dados_reduzidos$heart_disease,
      xlab = "Idade (ano)",
      ylab = "Doença Cardíaca")
lines(dados_plot$age, dados_plot$heart_disease)

```