

```

# Eduardo Freire Mangabeira
# DRE 116137191

install.packages("faraway")
library(faraway)

# Dataset 1
data(diabetes)
head(diabetes)

# a) o grupo de pessoas com o corpo de tamanho mediano possui uma
# variação maior no índice de colesterol
# as médias são próximas da média populacional
plot(diabetes$frame, diabetes$chol, xlab="tamanho corporal", ylab="colesterol")
mean(diabetes$chol, na.rm = T)

# b) aqui parece haver uma relação entre tipo corporal e taxa HDL,
# quanto menor o corpo, maior o HDL.
plot(diabetes$frame, diabetes$hdl, xlab = "tamanho corporal", ylab= "hdl")

# c) as pessoas com corpos medios ou grandes possuem uma
# concentração maior de hemoglobinas glicolisadas, as pessoas
# com corpos pequenos possuem taxas muito próximas em média.
MaiorDeCinquenta <- diabetes[diabetes$age >= 50,]
MenorDeCinquenta <- diabetes[diabetes$age < 50,]
plot(MaiorDeCinquenta$glyhb, xlab= "hemoglobinas glicosiladas",
     ylab= "maiores de 50 anos")
plot(MaiorDeCinquenta$frame, MaiorDeCinquenta$glyhb,
     main="maiores de 50", xlab="tamanho corporal",
     ylab="hemoglobinas glicosiladas")
plot(MenorDeCinquenta$frame, MenorDeCinquenta$glyhb,
     main="menores de 50", xlab="tamanho corporal",
     ylab="hemoglobinas glicosiladas")

# d)
# fiz também um barplot para auxiliar na visualização.
# como o coeficiente de pearson é muito próximo de zero se
# pode afirmar que não há correlação entre as variáveis
# tamanho corporal e gênero na amostra observada.
gen_versus_size <- table(diabetes$gender, diabetes$frame)
barplot(gen_versus_size, beside=T, main="tamanho corporal",
        legend= T, ylim = c(0,150))
chisq.test(gen_versus_size, correct=T)

# e) pelo scatterplot, aparentemente existe alguma relação,
# vamos conferir se isso se confirma.
# Pelo coeficiente de correlação de pearson, existe uma correlação
# forte(entre 0.7 e 0.9) entre as variáveis peso e tamanho da cintura.
# Fazendo a análise de regressão, é possível observar que a cada
# unidade que representa o tamanho da cintura, são aumentadas
# unidades de peso a uma taxa de 0.1205
attach(diabetes)
plot(weight, waist, xlab= "peso", ylab="cintura")
cor(weight, waist, use = "complete.obs")
modelo <- lm(waist~weight, data = diabetes)
abline(modelo, col = "darkred")
# plotamos mais uma vez para ver os valores residuais
plot(modelo)
summary(modelo)
#

# f) pelo coeficiente de correlação de pearson, não há
# correlação entre o IMC e o colesterol, o que fica evidente
# também ao visualizar o gráfico da reta de regressão.
#
imc = weight / height * height
plot (chol, imc, xlab="colesterol", ylab= "IMC")
cor(imc, chol, use = "complete.obs")
modelo_2 <- lm(chol~imc, data = diabetes)
abline(modelo_2, col="darkred")
plot(modelo_2)
summary(modelo_2)

# g) a associação entre as variáveis é muito pequena, segundo
# o coeficiente de correlação.
# pela reta de regressão podemos ver que realmente não há
# correlação.
plot(bp.1s, chol, xlab=" 1ª pressão sistólica", ylab="colesterol")
cor(chol, bp.1s, use = "complete.obs")
modelo_3 <- lm(bp.1s~chol, data = diabetes)
abline(modelo_3, col="darkred")

#
plot(MaiorDeCinquenta$bp.1s, MaiorDeCinquenta$chol, main="maiores de 50",
     xlab="1ª pressão sistólica", ylab="colesterol")

```

```
cor(MaiorDeCinquenta$chol, MaiorDeCinquenta$bp.ls, use = "complete.obs")
modelo_4 <- lm(MaiorDeCinquenta$bp.ls~MaiorDeCinquenta$chol, data = diabetes)
abline(modelo_3, col="darkred")
```

```
plot(MenorDeCinquenta$bp.ls, MenorDeCinquenta$chol, main="menores de 50",
      xlab=" 1ª pressão sistólica", ylab="colesterol")
cor(MenorDeCinquenta$chol, MenorDeCinquenta$bp.ls, use="complete.obs")
modelo_4 <- lm(MenorDeCinquenta$bp.ls~MenorDeCinquenta$chol, data = diabetes)
abline(modelo_3, col="darkred")
```

```
# Os resultados mostram que mesmo fazendo recortes, o conjunto de
# dados continua mostrando que não há correlação entre as variáveis
```