Atividade 1 - Estimando os parâmetros

Disciplina: Regressão 1

Schênia Taynna Medeiros Silva - 20190156798

2023-07-24

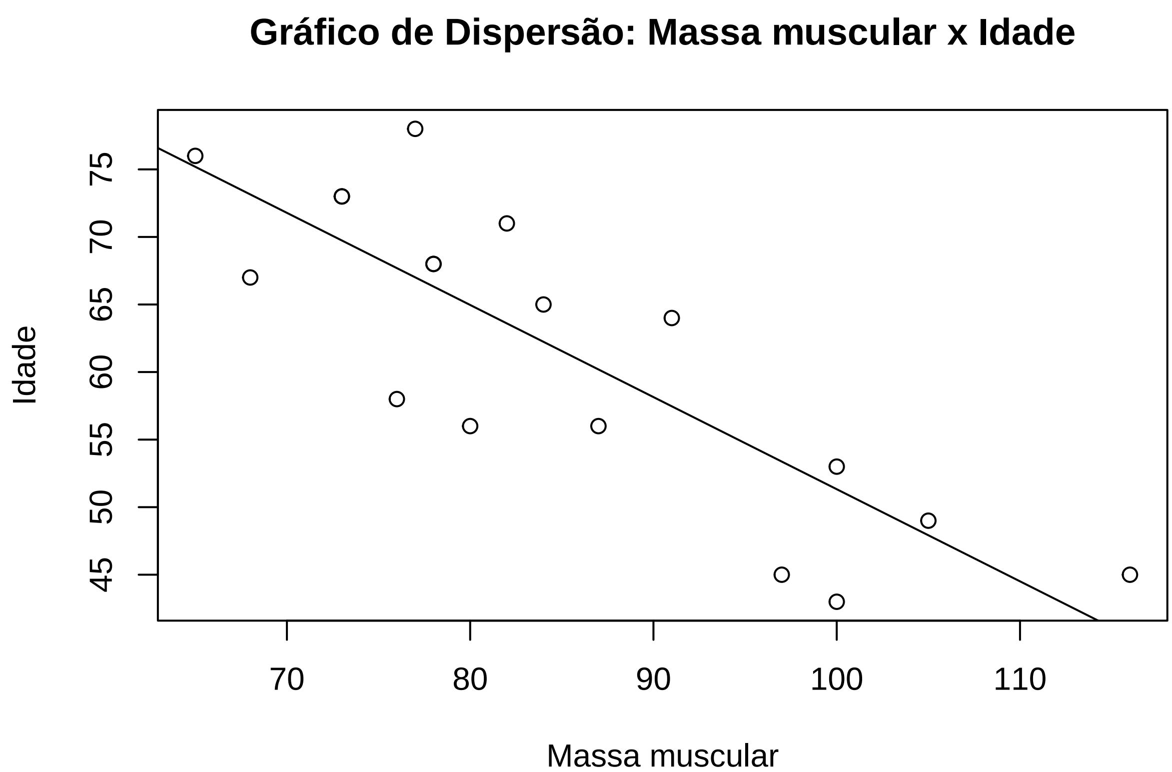
Questão 1

Considere os dados sobre massa muscular e da idade de mulheres adultas.

```
#Dados

idade<-c(71, 64, 43, 67, 56, 73, 68, 56, 76, 65, 45, 58, 45, 53, 49, 78, 73, 68)

massamuscular<-c(82, 91, 100, 68, 87, 73, 78, 80, 65, 84, 116, 76, 97, 100, 105,
                 77, 73, 78)

# Criando um dataframe
dados_Q1 <- data.frame(idade, massamuscular)
```

a) Ajuste um modelo de regressão linear simples para explicar a massa muscular em função da idade de mulheres adultas.

```
# grafico de dispersao
plot(idade~massamuscular,
     main = "Gráfico de Dispersão: Massa muscular x Idade",
     xlab = "Massa muscular", ylab = "Idade")

#ajuste do modelo
fit_Q1<-lm(idade~massamuscular)
abline(fit_Q1)
```



b) Quais foram os valores estimados para os coeficientes de regressão? Quem é o intercepto e a inclinação da reta?

```r
#"Resumo" do ajuste
summary(fit_Q1)
```

```
##
## Call:
## lm(formula = idade ~ massamuscular)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.692 -5.658  1.671  3.262 10.990
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   119.5125     9.6026  12.446 1.21e-09 ***
## massamuscular  -0.6818     0.1116  -6.111 1.50e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.361 on 16 degrees of freedom
## Multiple R-squared:    0.7,  Adjusted R-squared:  0.6813
## F-statistic: 37.34 on 1 and 16 DF,  p-value: 1.504e-05
```

```r
names(fit_Q1)
```

```
##  [1] "coefficients"  "residuals"     "effects"       "rank"
##  [5] "fitted.values" "assign"        "qr"            "df.residual"
##  [9] "xlevels"       "call"          "terms"         "model"
```

```r
names(summary(fit_Q1))
```

```
##  [1] "call"          "terms"         "residuals"     "coefficients"
##  [5] "aliased"       "sigma"         "df"            "r.squared"
##  [9] "adj.r.squared" "fstatistic"    "cov.unscaled"
```

os valores estimados para os coeficientes de regressão são: * Intercepto (coeficiente da constante): 119.5125 * Inclinação (coeficiente de "massamuscular"): -0.6818

O **intercepto** representa o valor esperado da variável "idade" quando a variável "massamuscular" é igual a zero. A **inclinação** representa a mudança esperada na variável "idade" para cada aumento de uma unidade na variável "massamuscular". O resultado do coeficiente é negativo, isso indica uma relação inversa entre "massamuscular" e "idade". Ou seja, à medida que a "massamuscular" aumenta, espera-se que a "idade" diminua.

c) Expresse a reta estimada. E interprete os parâmetros.

```r
#parametros \alpha e \beta
coef(fit_Q1)
```

```
##   (Intercept) massamuscular
##   119.5124786    -0.6818462
```

```r
fit_Q1$coefficients
```

```
##   (Intercept) massamuscular
##   119.5124786    -0.6818462
```

```r
fit_Q1
```

```
##
## Call:
## lm(formula = idade ~ massamuscular)
##
## Coefficients:
##   (Intercept)  massamuscular
##      119.5125        -0.6818
```

A reta estimada é dada pela equação:

massamuscular= 119.5125 - 0.6818 $\times$ idade

A equação da reta estimada nos fornece uma forma de prever a "massamuscular" com base na variável "idade" usando o modelo de regressão linear.

d) Qual a estimativa do erro padrão para o modelo de regressão ajustado

```r
## Estimador de sigma 2

summary(fit_Q1)$sigma
```

```
## [1] 6.361374
```

```r
summary(fit_Q1)$sigma^2
```

```
## [1] 40.46709
```

A estimativa do erro padrão para o modelo de regressão ajustado é de aproximadamente 6.361374.

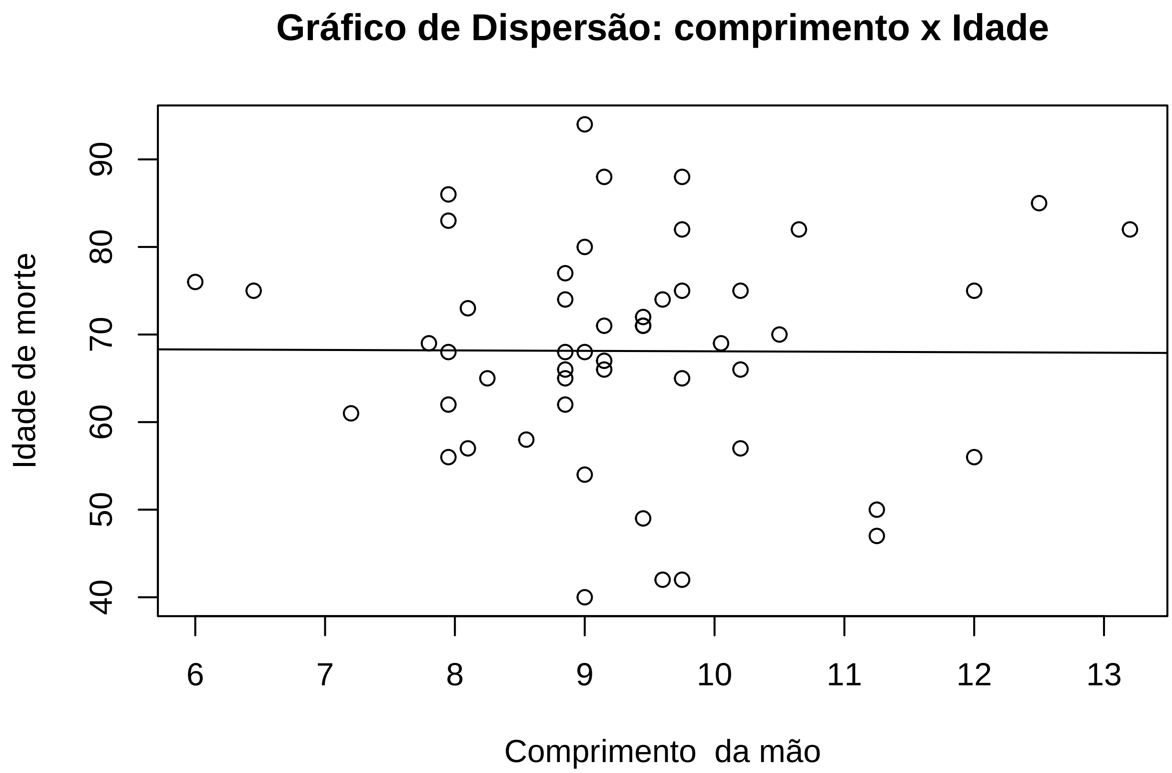
Questão 2

Considere os dados sobre idade de morte e do comprimento de linha da mão.

```r
# dados
idade_morte<-c(40, 42, 42, 47, 49, 50, 54, 56, 56, 57, 57, 58, 61, 62, 62, 65, 65, 65,
        66, 66, 66, 67, 68, 68, 68, 69, 69, 70, 71, 71, 71, 72, 73, 74, 74, 75,
        75, 75, 75, 76, 77, 80, 82, 82, 82, 83, 85, 86, 88, 88, 94)

comprimento<-c(9.00,9.60,9.75,11.25,9.45,11.25,9.00,7.95,12.00,8.10,10.20,8.55,
               7.20,7.95,8.85,8.25,8.85,9.75,8.85,9.15,10.20, 9.15,7.95,8.85,
               9.00,7.80,10.05,10.50,9.15,9.45,9.45,9.45,8.10,8.85,9.60,6.45,
               9.75,10.20,12.00, 6.00,8.85,9.00,9.75,10.65,13.20,7.95,12.50,
               7.95,9.15,9.75,9.00)

# Criando um dataframe
dados_Q2 <- data.frame(idade_morte, comprimento)
```

a) Ajuste um modelo de regressão linear simples para explicar idade de morte em função do comprimento de linha da mão.

```r
# grafico de dispersao
plot(idade_morte~comprimento,
     main = "Gráfico de Dispersão: comprimento x Idade",
     xlab = "Comprimento  da mão", ylab = "Idade de morte")

#ajuste do modelo
fit_Q2<-lm(idade_morte~comprimento)
abline(fit_Q2)
```



b) Quais foram os valores estimados para os coeficientes de regressão? Quem é o intercepto e a inclinação da reta?

```
summary(fit_Q2)
```

```
##
## Call:
## lm(formula = idade_morte ~ comprimento)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.1337  -6.7084   0.8034   6.9763  25.8663
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 68.60508   12.08350   5.678 7.32e-07 ***
## comprimento -0.05237    1.28442  -0.041    0.968
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.6 on 49 degrees of freedom
## Multiple R-squared:  3.393e-05,  Adjusted R-squared:  -0.02037
## F-statistic: 0.001663 on 1 and 49 DF,  p-value: 0.9676
```

```
names(fit_Q2)
```

```
##  [1] "coefficients"  "residuals"     "effects"       "rank"
##  [5] "fitted.values" "assign"        "qr"            "df.residual"
##  [9] "xlevels"       "call"          "terms"         "model"
```

```r
names(summary(fit_Q2))
```

```
##  [1] "call"          "terms"         "residuals"     "coefficients"
##  [5] "aliased"       "sigma"         "df"            "r.squared"
##  [9] "adj.r.squared" "fstatistic"    "cov.unscaled"
```

os valores estimados para os coeficientes de regressão são: * Intercepto (coeficiente da constante): 68.60508 * Inclinação (coeficiente de "comprimento"): -0.05237

O **intercepto** representa o valor esperado da variável "idade_morte" quando a variável "comprimento" é igual a zero. A **inclinação** representa a mudança esperada na variável "idade_morte" para cada aumento de uma unidade na variável "comprimento". O resultado do coeficiente é negativo, isso indica uma relação inversa entre "comprimento" e "idade_morte". Ou seja, à medida que a "comprimento" aumenta, espera-se que a "idade_morte" diminua.

c) Expresse a reta estimada. E interprete os parâmetros do modelo.

```r
fit_Q2
```

```
##
## Call:
## lm(formula = idade_morte ~ comprimento)
##
## Coefficients:
## (Intercept)  comprimento
##    68.60508     -0.05237
```

A reta estimada é dada pela equação:

comprimento = 68.60508 -0.05237 $\times$ idade de morte

A equação da reta estimada nos fornece uma forma de prever a "comprimento" com base na variável "idade_morte" usando o modelo de regressão linear.

d) Qual a estimativa do erro padrão para o modelo de regressão ajustado.

```r
summary(fit_Q2)$sigma
```

```
## [1] 12.60145
```

```r
summary(fit_Q2)$sigma^2
```

```
## [1] 158.7965
```

A estimativa do erro padrão para o modelo de regressão ajustado é de aproximadamente 12.60145

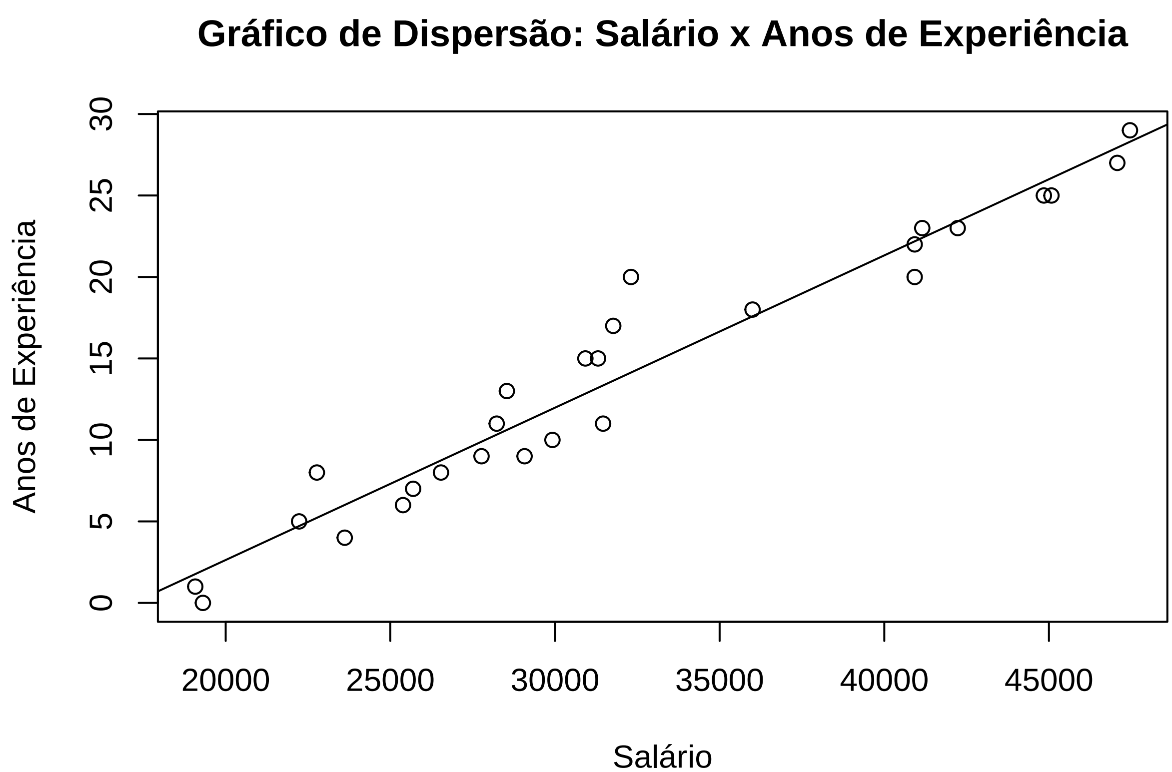
Questão 3.

Considere os dados sobre salário e anos de experiência de executivos.

```r
salario<-c(19307,31769,22769,31307,27769,30923,26538,22230,28538,32307,28230,
           19076,25384,25692,42230,40923,36000,47076,31461,29923,47461,41153,
           23615,40923,45076,29076,44846)

experiencia<-c(0, 17, 8, 15, 9, 15, 8, 5, 13, 20, 11, 1, 6, 7, 23, 20, 18, 27,
               11, 10, 29, 23, 4, 22, 25, 9, 25)

# Criando um dataframe
dados_Q3 <- data.frame(salario, experiencia)
```

a) Ajuste um modelo de regressão linear simples para explicar o salário dos executivos em função dos anos de experiência.

```r
# grafico de dispersao
plot(experiencia~salario,
     main = "Gráfico de Dispersão: Salário x Anos de Experiência",
     xlab = "Salário", ylab = "Anos de Experiência")

#ajuste do modelo
fit_Q3<-lm(experiencia~salario)
abline(fit_Q3)
```



b) Quais foram os valores estimados para os coeficientes de regressão? Quem é o intercepto e a inclinação da reta?

```r
#"Resumo" do ajuste
summary(fit_Q3)
```

```
##
## Call:
## lm(formula = experiencia ~ salario)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3442 -1.3667 -0.7431  0.6888  5.8651
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.606e+01  1.552e+00  -10.35 1.59e-10 ***
## salario      9.346e-04  4.651e-05   20.10  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.035 on 25 degrees of freedom
## Multiple R-squared:  0.9417, Adjusted R-squared:  0.9394
## F-statistic: 403.8 on 1 and 25 DF,  p-value: < 2.2e-16
```

```r
names(fit_Q3)
```

```
##  [1] "coefficients"  "residuals"     "effects"       "rank"
##  [5] "fitted.values" "assign"        "qr"            "df.residual"
##  [9] "xlevels"       "call"          "terms"         "model"
```

```r
names(summary(fit_Q3))
```

```
##  [1] "call"          "terms"         "residuals"     "coefficients"
##  [5] "aliased"       "sigma"         "df"            "r.squared"
##  [9] "adj.r.squared" "fstatistic"    "cov.unscaled"
```

os valores estimados para os coeficientes de regressão são: * Intercepto (coeficiente da constante): -1.606e+01 * Inclinação (coeficiente de "massamuscular"): 9.346e-04

O **intercepto** representa o valor esperado da variável "experiencia" quando a variável "salario" é igual a zero. A **inclinação** representa a mudança esperada na variável "experiencia" para cada aumento de uma unidade na variável "salario".

c) Expresse a reta estimada. E interprete os parâmetros do modelo.

```r
#parametros \alpha e \beta
fit_Q3
```

```
##
## Call:
## lm(formula = experiencia ~ salario)
##
## Coefficients:
## (Intercept)      salario
##  -1.606e+01    9.346e-04
```

A reta estimada é dada pela equação:

salario = -1.606e+01 + 9.346e-04 × anos de experiênica

A equação da reta estimada nos fornece uma forma de prever "salario" com base na variável "experiencia" usando o modelo de regressão linear.

d) Qual a estimativa do erro padrão para o modelo de regressão ajustado.

```
summary(fit_Q3)$sigma
```

```
## [1] 2.035433
```

```
summary(fit_Q3)$sigma^2
```

```
## [1] 4.142988
```

A estimativa do erro padrão para o modelo de regressão ajustado é de aproximadamente 2.035433