

Laboratório 08 - Estatística Aplicada - Manhã

Aluno: Marcus Vinícius Leite
Costa

Matrícula: 116110728

1)

a) O Código utilizado para gerar o gráfico de dispersão da letra A foi o seguinte:

```
##### QUESTAO 01 #####

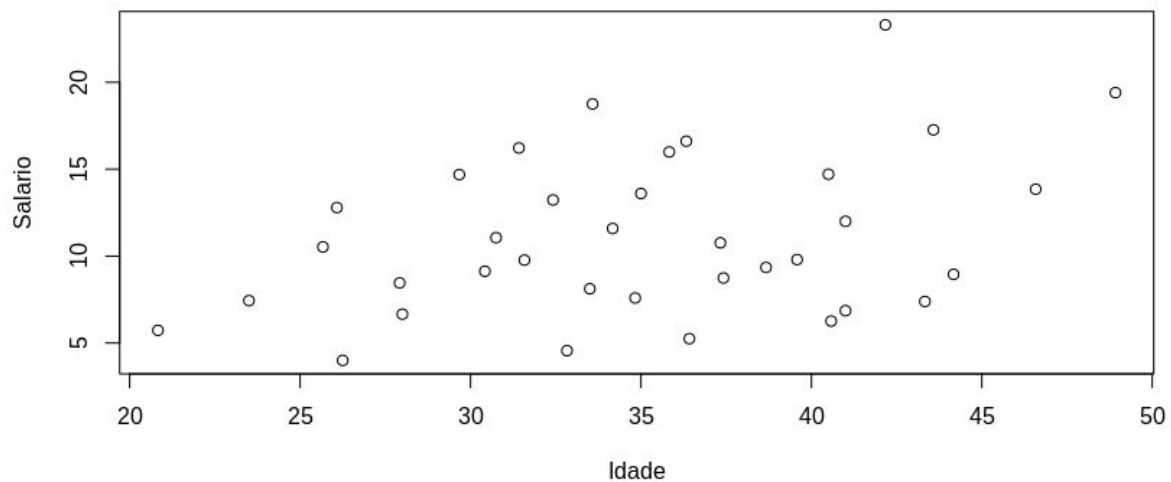
dados <- read.csv("dados.csv", sep = ";")
attach(dados)

## LETRA A ##

dados$salario <- as.numeric(gsub(",", "\\.", dados$salario))
dados$idade <- as.numeric(gsub(",", "\\.", dados$idade))

plot(dados$idade, dados$salario, xlab = "Idade", ylab = "Salario")
```

O gráfico de dispersão gerado foi o seguinte:



b) Olhando apenas o gráfico de dispersão, sem cálculos adicionais, é possível perceber uma leve relação entre as variáveis, já que é observável uma possível reta sendo construída que agrega vários pontos.

c) O código utilizado para calcular o coeficiente de correlação linear entre as variáveis salário e idade foi:

```
## LETRA C ##  
  
cor(dados$idade, dados$salario)
```

O valor obtido foi:

```
> cor(dados$idade, dados$salario)  
[1] 0.365186
```

Como esse coeficiente varia entre -1 e 1, e estar perto dessas extremidades significa uma alta correlação linear, podemos deduzir que o valor obtido indica a ocorrência de uma correlação fraca entre as variáveis idade e salário.

d) Código utilizado na letra D:

```
## LETRA D ##  
  
model <- lm(dados$idade ~ dados$salario)
```

Os coeficientes estimados foram:

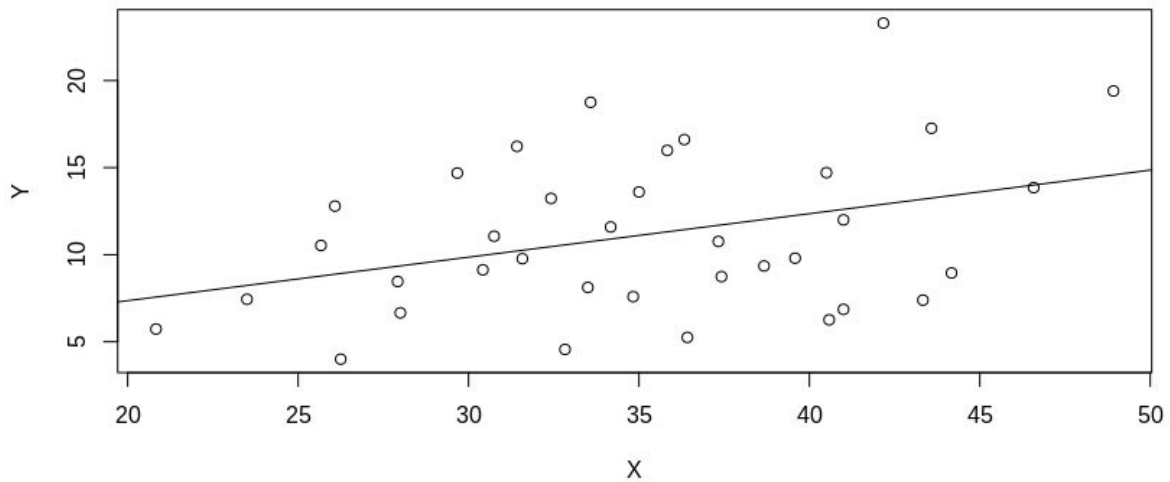
```
Call:
lm(formula = dados$idade ~ dados$salario)
```

```
Coefficients:
 (Intercept)  dados$salario
    29.1143      0.5338
```

e) Utilizando os seguintes comandos, obteve-se o gráfico de dispersão com a reta de regressão estimada:

```
## LETRA E ##

X< dados$idade
Y<-dados$salario
lm(Y~X)
plot(X,Y)
abline(lm(Y ~ X))
```



2) Na questão 02, os valores obtidos após calcular a correlação e a função `lm()` foi o seguinte:

```
> ##### QUESTAO 02 #####
```

```

>
> ## LETRA A ##
>
> cor(dados$idade, dados$salario)
[1] 0.365186
>
> a <- lm(dados$idade ~ dados$salario)
> summary(a)

Call:
lm(formula = dados$idade ~ dados$salario)

Residuals:
      Min       1Q   Median       3Q      Max
-11.3427  -4.7516  -0.5394   4.7326  10.2786

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    29.1143     2.8018  10.391 4.33e-12 ***
dados$salario   0.5338     0.2334   2.287  0.0285 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.333 on 34 degrees of freedom
Multiple R-squared:  0.1334,    Adjusted R-squared:  0.1079
F-statistic: 5.232 on 1 and 34 DF,  p-value: 0.02852

>
> ## LETRA B ##
>
> cor(dados$idade**2, dados$salario)
[1] 0.3648552
>
> b <- lm(dados$idade**2 ~ dados$salario)
> summary(b)

Call:
lm(formula = dados$idade^2 ~ dados$salario)

Residuals:
      Min       1Q   Median       3Q      Max

```

```
-654.62 -321.85 -79.85 303.11 810.61
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	855.37	196.97	4.343	0.00012	***
dados\$salario	37.48	16.40	2.285	0.02868	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 445.2 on 34 degrees of freedom

Multiple R-squared: 0.1331, Adjusted R-squared: 0.1076

F-statistic: 5.221 on 1 and 34 DF, p-value: 0.02868

```
>
```

```
> ## LETRA C ##
```

```
>
```

```
> cor(log(dados$idade), dados$salario)
```

```
[1] 0.364591
```

```
>
```

```
> c <- lm(log(dados$idade) ~ dados$salario)
```

```
> summary(c)
```

Call:

```
lm(formula = log(dados$idade) ~ dados$salario)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.41628	-0.13765	-0.00172	0.14630	0.28988

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3.361900	0.083310	40.354	<2e-16	***
dados\$salario	0.015841	0.006939	2.283	0.0288	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1883 on 34 degrees of freedom

Multiple R-squared: 0.1329, Adjusted R-squared: 0.1074

F-statistic: 5.212 on 1 and 34 DF, p-value: 0.0288

```
>
```

```

> ## LETRA D ##
>
> cor(sqrt(dados$idade), dados$salario)
[1] 0.3650559
>
> d <- lm(sqrt(dados$idade) ~ dados$salario)
> summary(d)

Call:
lm(formula = sqrt(dados$idade) ~ dados$salario)

Residuals:
      Min       1Q   Median       3Q      Max
-1.08260 -0.40921 -0.01852  0.41460  0.86005

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.38453    0.24016  22.420  <2e-16 ***
dados$salario  0.04573    0.02000   2.286  0.0286 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5429 on 34 degrees of freedom
Multiple R-squared:  0.1333,    Adjusted R-squared:  0.1078
F-statistic: 5.228 on 1 and 34 DF,  p-value: 0.02858

> ## LETRA E ##
>
> cor(1/dados$idade, dados$salario)
[1] -0.3622173
>
> e <- lm(1/dados$idade ~ dados$salario)
> summary(e)

Call:
lm(formula = 1/dados$idade ~ dados$salario)

Residuals:
      Min       1Q   Median       3Q      Max
-0.0084054 -0.0047047 -0.0000603  0.0038707  0.0157090

```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.0351105	0.0026003	13.503	3.21e-15 ***
dados\$salario	-0.0004907	0.0002166	-2.266	0.0299 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.005878 on 34 degrees of freedom

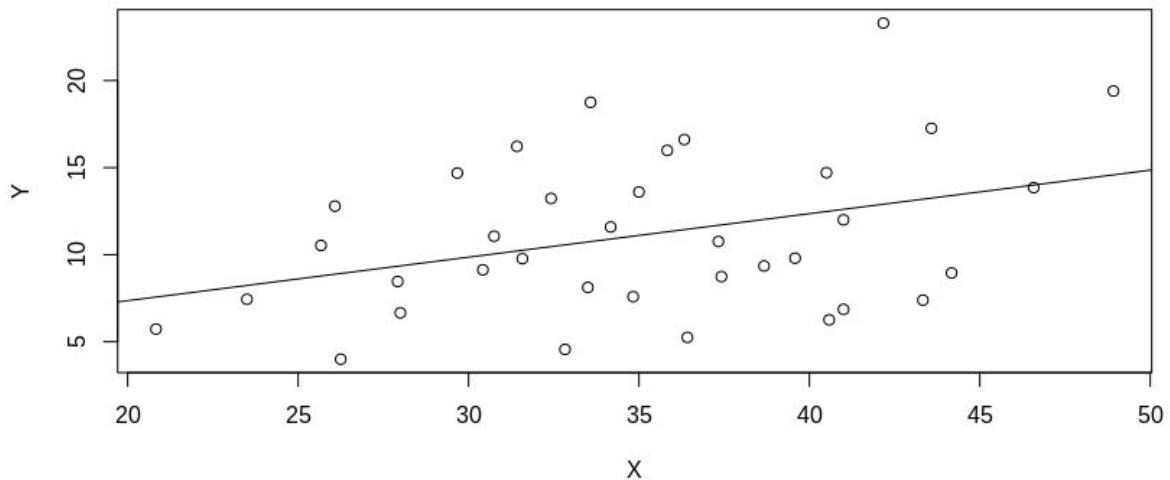
Multiple R-squared: 0.1312, Adjusted R-squared: 0.1056

F-statistic: 5.134 on 1 and 34 DF, p-value: 0.02994

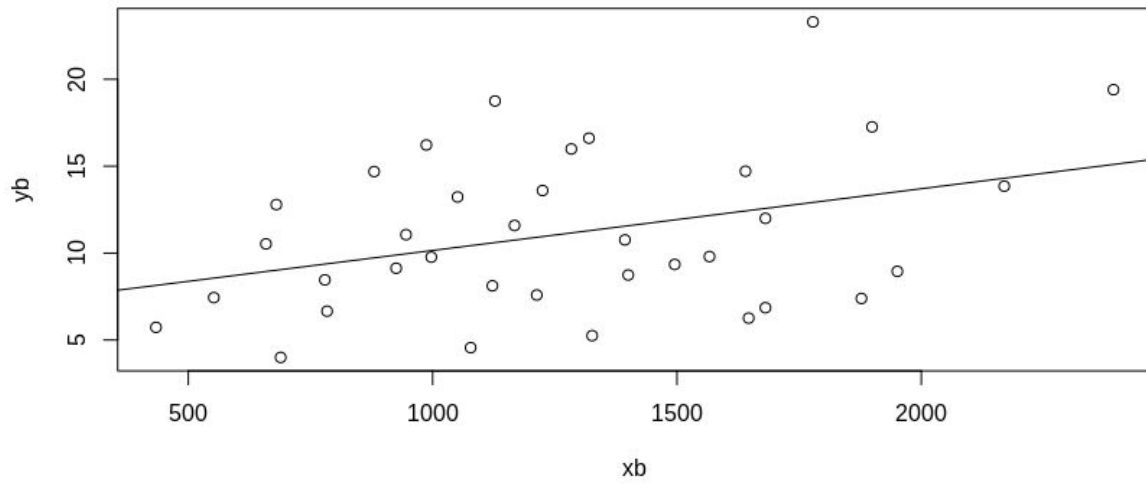
Como o p-valor da letra obtido na letra D foi o menor, podemos deduzir que ele é o que melhor se ajusta.

Gráficos de dispersão e suas retas de regressão da questão 02

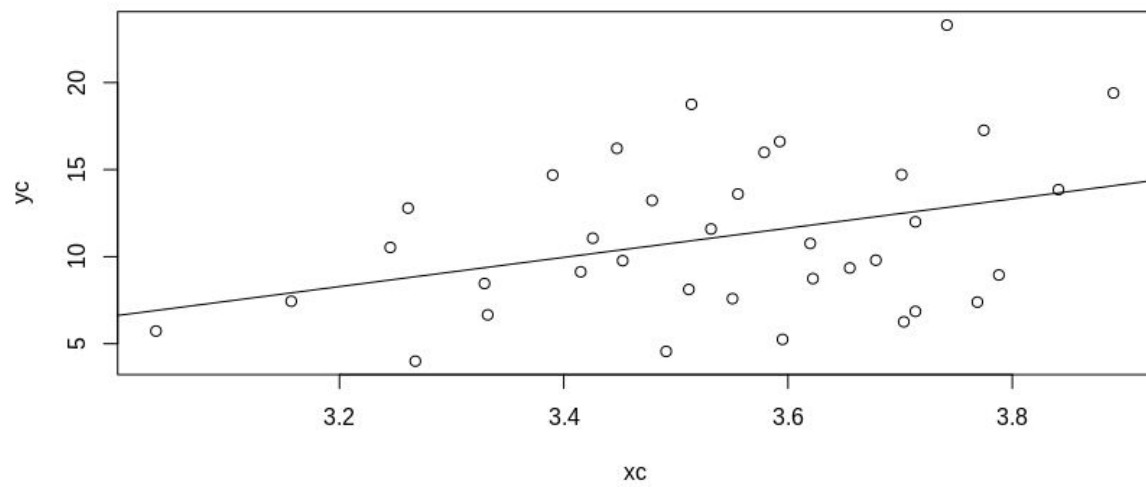
LETRA A:



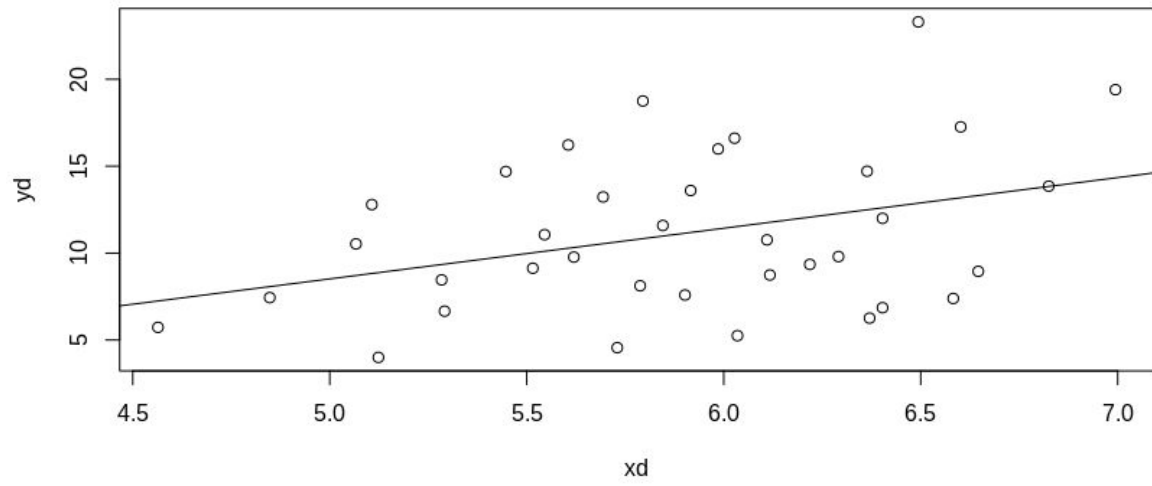
LETRA B:



LETRA C:



LETRA D:



LETRA E:

