

## EXERCÍCIOS 13ª aula de R – Regressão

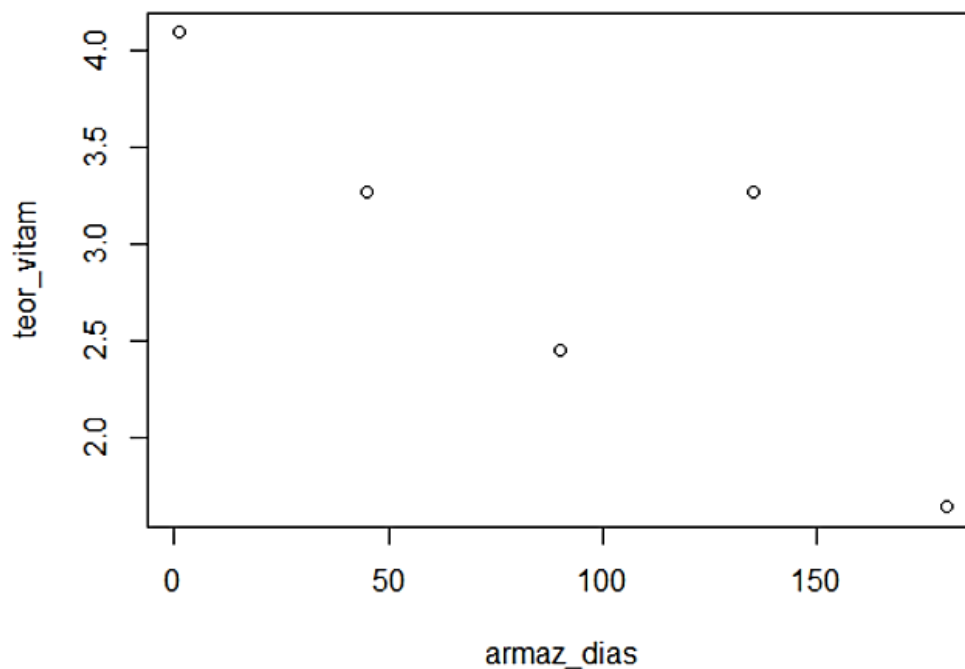
Todos os exercícios serão executados no RStudio e as operações e códigos utilizados devem ser escritos na sequência em que foram utilizados, para posterior correção.

1. Os dados apresentados na tabela abaixo relacionam o teor de vitamina C (mg de ácido ascórbico/100ml de suco de maçã) em função do período de armazenamento em dias.

Período de armazenamento (dias)	Teor de vitamina C
1	4,09
45	3,27
90	2,45
135	3,27
180	1,64

- a) Crie um data.frame e plote o gráfico.

```
armaz_dias<-c(1,45,90,135,180)
teor_vitam<-c(4.09,3.27,2.45,3.27,1.64)
dados<-data.frame(armaz_dias,teor_vitam)
dados
plot(dados)
```



- b) Ache a equação da reta que relaciona dos dados.

```
> reglin<-lm(teor_vitam~armaz_dias, dados)
> reglin
```

Call:

```
lm(formula = teor_vitam ~ armaz_dias, data = dados)
```

Coefficients:

```
(Intercept)    armaz_dias
   3.92980      -0.01093
```

$$teor_{vitam} = 3,92980 - 0,01093.armaz\_dias$$

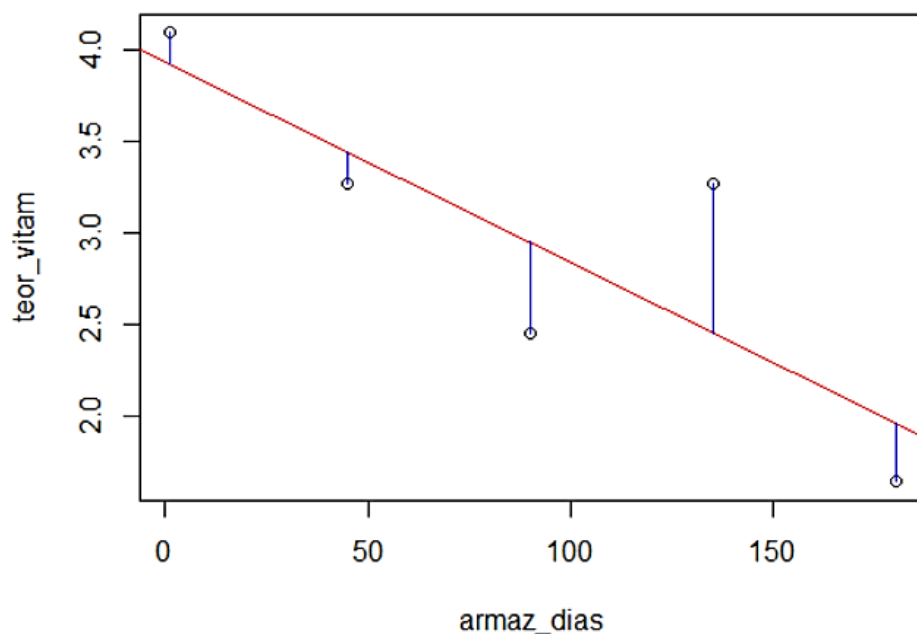
- c) Qual seria o teor de vitamina C se o suco ficar armazenado durante 20 dias?

```
> y<-(-0.01093*x+3.92980)
> x<-20
> y
[1] 3.7112
```

- d) Agora plote novamente os dados e acrescente ao gráfico, além da reta de regressão ajustada, segmentos de reta representando os resíduos, ou seja, segmentos de reta que vão dos valores observados (pontos) aos calculados (reta).

```
predict(reglin)
resid(reglin)
result<-data.frame(armaz_dias,
                    teor_vitam,
                    calculado=predict(reglin),
                    residuos=resid(reglin))

result
plot(armaz_dias,teor_vitam)
abline(reglin, col=2)
segments(
  result$armaz_dias,
  result$teor_vitam,
  result$armaz_dias,
  result$calculado,
  col=4
)
```



- e) Qual a conclusão dessa regressão?

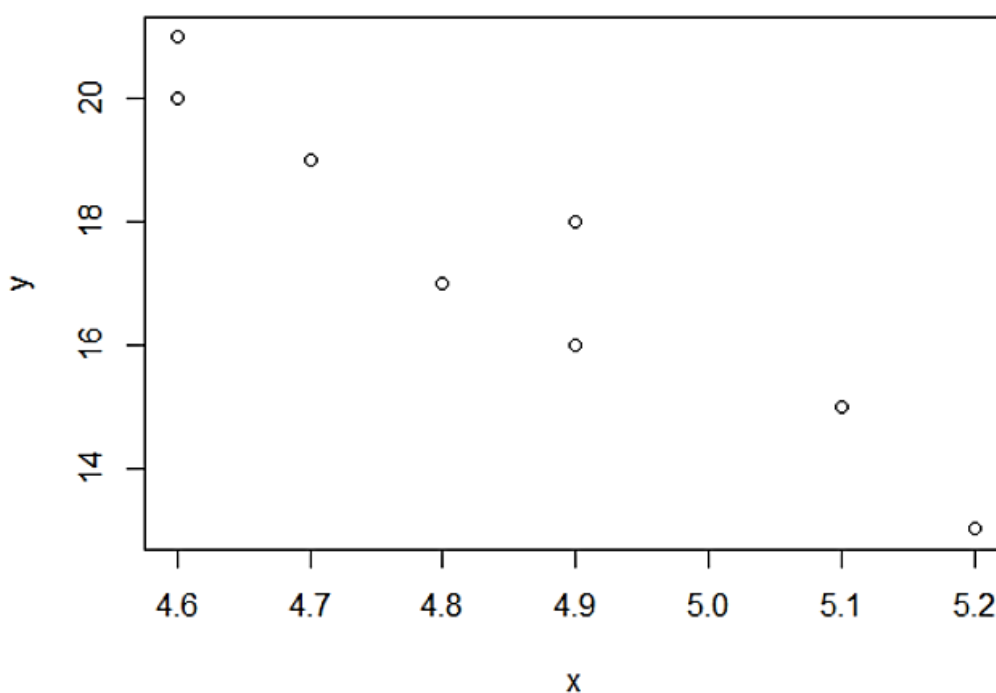
À medida em que o tempo passa, o teor de vitamina no suco diminui.

2) Para uma amostra de 8 operadores de máquina, foram coletados o número de horas de treinamento (x) e o tempo necessário para completar o trabalho (y). Os dados encontram-se na tabela abaixo:

X	Y
5,2	13
5,1	15
4,9	18
4,6	20
4,7	19
4,8	17
4,6	21
4,9	16

a) Faça o gráfico de dispersão para esses dados.

```
x<-c(5.2,5.1,4.9,4.6,4.7,4.8,4.6,4.9)
y<-c(13,15,18,20,19,17,21,16)
plot(x,y)
```



b) Determine a equação da reta.

```
> reglin<-lm(y~x)
> reglin
```

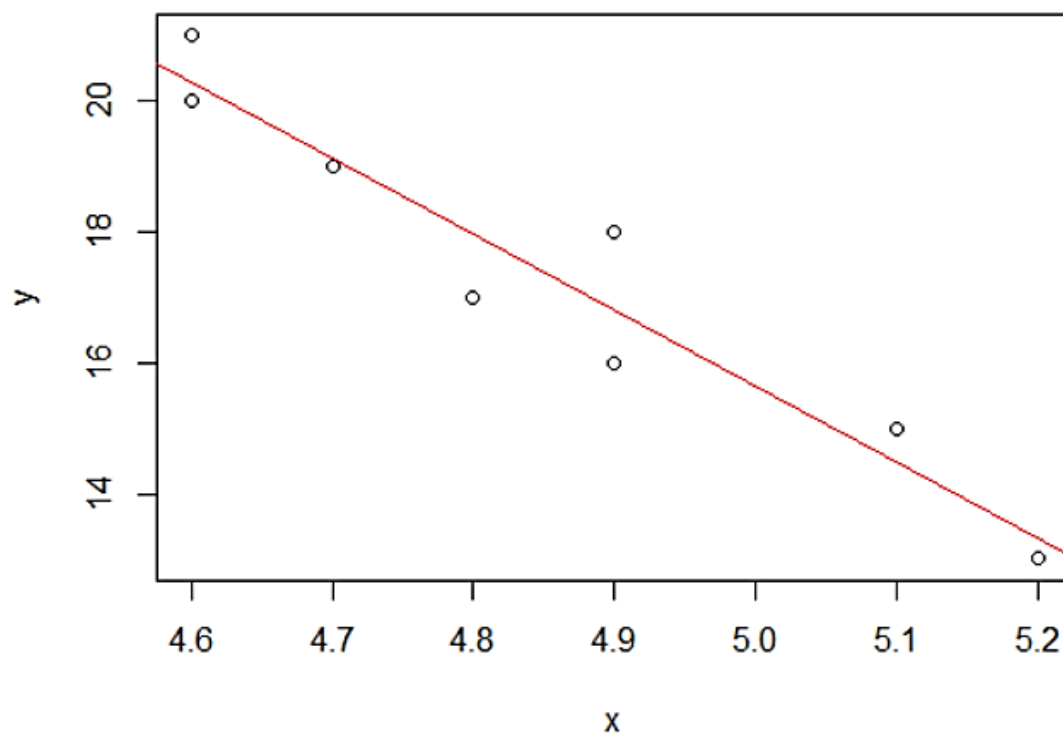
```
Call:
lm(formula = y ~ x)
```

```
Coefficients:
(Intercept)          x
      73.72      -11.62
```

$$y = 73,72 - 11,62.x$$

c) Trace no gráfico anterior, a reta de regressão.

```
abline(reglin, col=2)
```



d) Calcule e interprete o coeficiente de determinação.

```
> summary(reglin)
```

```
Call:
```

```
lm(formula = y ~ x)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-0.9559	-0.4301	-0.1985	0.5772	1.2059

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	73.721	6.785	10.865	3.6e-05	***
x	-11.618	1.398	-8.312	0.000164	***

```
---
```

```
Signif. codes:
```

```
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.815 on 6 degrees of freedom
```

```
Multiple R-squared: 0.9201, Adjusted R-squared: 0.9068
```

```
F-statistic: 69.09 on 1 and 6 DF, p-value: 0.0001645
```

o  $R^2$  é 0,9201, ou seja 92,01% de y é explicado pelos valores de x. Isso indica que a medida em que o tempo de treinamento aumenta, o tempo de realização do trabalho diminui.