

Curso Livre II

Aluno: Manuel Ferreira Junior

Matricula: 20180008601

Modulo II

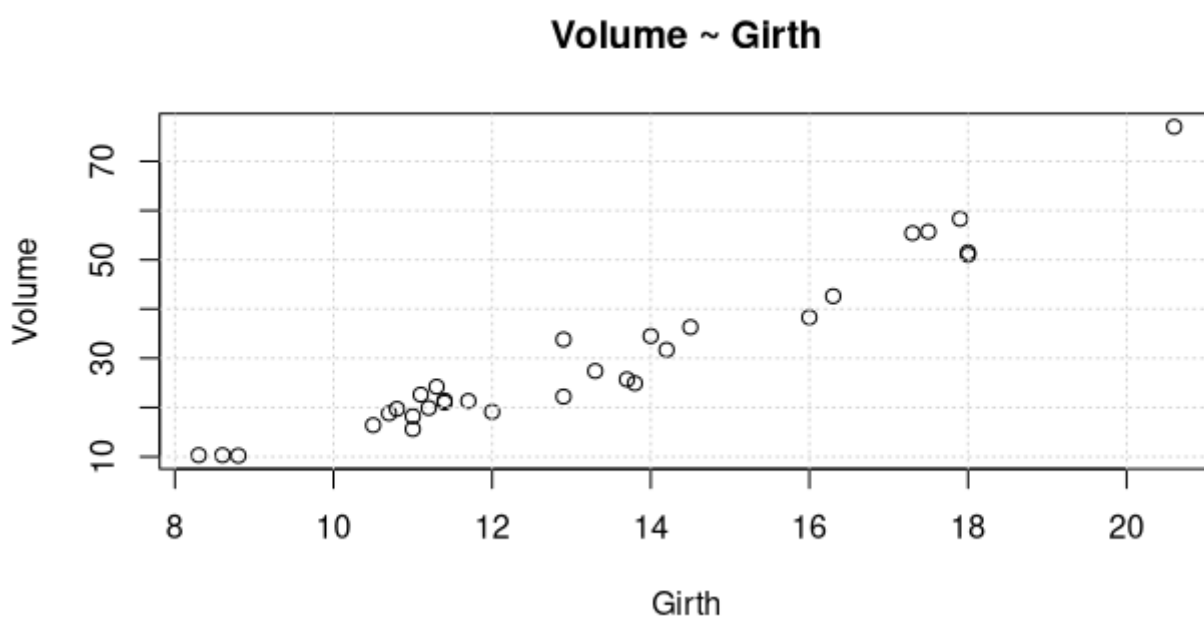
Tarefa 01 Acesse o banco de dados `trees` no R, utilizando a função `data()`. Estes dados fornecem as informações sobre o diâmetro (`Girth`), a altura (`Height`) e o volume (`Volume`) de 31 árvores cerejeiras. Faça uma análise sobre a relação linear entre as variáveis `Girth` e `Volume` e ajuste um modelo de regressão linear simples para explicar o volume da árvore em função do seu diâmetro. E depois responda as seguintes questões:

Carregando os dados

```
data(trees)
attach(trees)
```

A) Obtenha o gráfico de dispersão. Como você descreveria a relação entre as duas variáveis?

```
> plot(Girth,Volume)
> grid()
> title('Volume ~ Girth')
```



Aparentemente, segundo a relação das variáveis, a medida que o diâmetro aumenta, o volume cresce de forma linear também.

B) Qual o grau de correlação linear entre as variáveis?

```
> cor(Girth,Volume)
[1] 0.9671194
```

Note que, a correlação é bem próxima de 1, re-afirmando a relação linear entre as variáveis, sendo uma relação forte e positiva de 0.97, aproximadamente.

C) Qual a conclusão do teste para o coeficiente de correlação? Justifique utilizando o p-valor.

```
> cor.test(Volume, Girth)

Pearson's product-moment correlation

data:  Volume and Girth
t = 20.478, df = 29, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9322519 0.9841887
sample estimates:
      cor
0.9671194
```

Verifica-se que o p-value < 2.2e-16, então podemos rejeitar a hipótese nula de que a correlação entre as variáveis é igual a zero, ou seja, que existe uma relação linear entre as variáveis.

D) Ajuste um modelo de regressão linear simples.

```
> gvmmodel <- lm(Volume ~ Girth)
> gvmmodel

Call:
lm(formula = Volume ~ Girth)

Coefficients:
(Intercept)      Girth
   -36.943       5.066
```

E) Qual a variável resposta (interesse) e a regressora (explicativa)?

Para esse modelo, consideremos que a variável explicativa é o diâmetro da árvore e a variável resposta é o volume da árvore.

F) Quais foram os valores estimados dos coeficientes de regressão (o intercepto e a inclinação)? Como é possível interpretá-los?

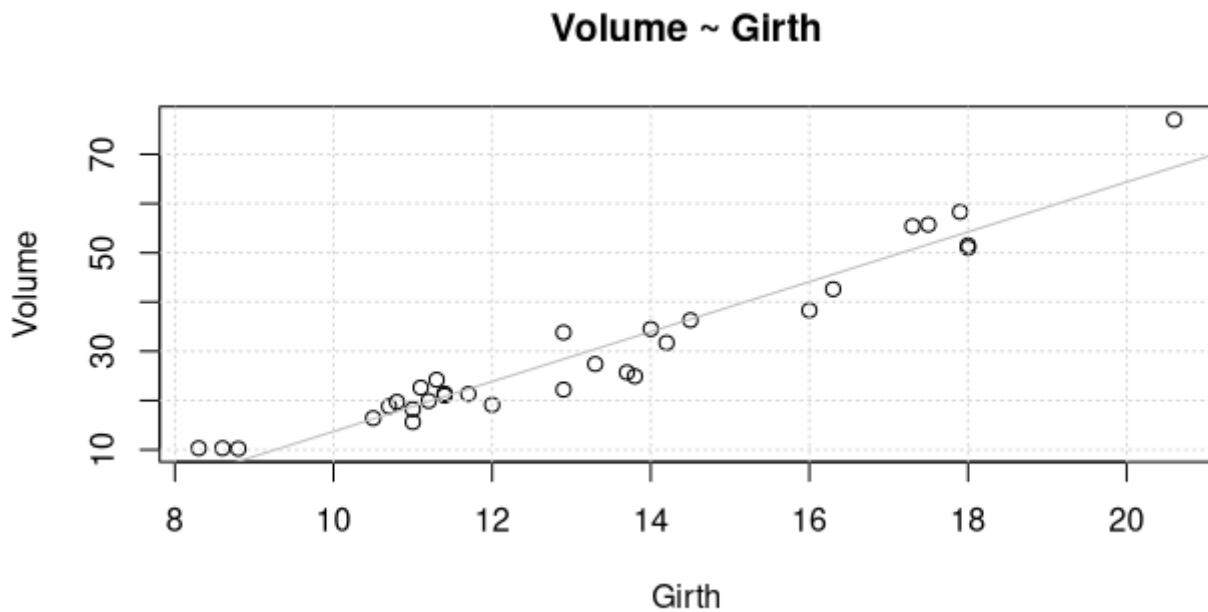
```
> coefs <- gvmmodel$coefficients
> coefs
(Intercept)      Girth
-36.943459      5.065856
```

Ao aumentarmos uma unidade o diâmetro, aumentamos em 5.06 unidades o volume da árvore, por exemplo.

G) Trace a reta de estimada no gráfico.

$$\hat{Y} = -36.943459 + 5.065856 \cdot X$$

```
> plot(Girth, Volume)
> abline(gvmmodel, col="gray")
> grid()
> title('Volume ~ Girth')
```



H) Qual o teste que utilizamos para verificar se a relação de regressão é significativa? Faça uma conclusão sobre o teste utilizando o respectivo p-valor.

Utilizamos o teste t-student. Concluimos que a relação entre as variáveis é estatisticamente significativa, pois o p-value é inferior a 2.2×10^{-16} , portanto existe uma correlação linear significativamente forte.

I) Qual a estatística utilizada para verificarmos se o modelo tem um alto poder explicativo para a variável resposta? Qual o valor desta estatística neste ajuste? Interprete este valor.

```
> summary(gvmodel)
```

Call:
lm(formula = Volume ~ Girth)

Residuals:

Min	1Q	Median	3Q	Max
-8.065	-3.107	0.152	3.495	9.587

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-36.9435	3.3651	-10.98	7.62e-12 ***
Girth	5.0659	0.2474	20.48	< 2e-16 ***

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

Residual standard error: 4.252 on 29 degrees of freedom
Multiple R-squared: 0.9353, Adjusted R-squared: 0.9331
F-statistic: 419.4 on 1 and 29 DF, p-value: < 2.2×10^{-16}

O coeficiente de determinação, ou R^2 . O valor da estatística neste ajuste é equivalente a 0.9353, ou seja, o modelo explica 93.53% da variabilidade dos dados de resposta ao redor de sua média.

J) Qual o volume médio previsto para uma árvore com o diâmetro de 13 polegadas?

```
> predict(gvmodel,newdata=data.frame(Girth=c(13)),  
+         interval="prediction")  
      fit      lwr      upr  
1 28.91267 20.07634 37.74901
```