

Exercícios

Marcus Nunes

25 de outubro de 2017

Relatório

1. Utilizando o menu do RStudio, crie um novo markdown, com output em pdf, chamado `exercicios.Rmd`. Altere o cabeçalho deste arquivo, de modo que ele fique similar ao do arquivo `relatorio.Rmd`.
2. Faça o resumo dos cinco números das variáveis quantitativas do conjunto de dados. É possível perceber alguma assimetria olhando estes resultados?

```
summary(iris[, 1:4])
```

```
##      Sepal.Length      Sepal.Width      Petal.Length      Petal.Width
##  Min.       :4.300    Min.       :2.000    Min.       :1.000    Min.       :0.100
##  1st Qu.:5.100    1st Qu.:2.800    1st Qu.:1.600    1st Qu.:0.300
##  Median :5.800    Median :3.000    Median :4.350    Median :1.300
##  Mean   :5.843    Mean   :3.057    Mean   :3.758    Mean   :1.199
##  3rd Qu.:6.400    3rd Qu.:3.300    3rd Qu.:5.100    3rd Qu.:1.800
##  Max.    :7.900    Max.    :4.400    Max.    :6.900    Max.    :2.500
```

3. Rode os comandos

```
library(GGally)
ggpairs(iris[, -5])
```

e interprete o resultado obtido.

4. Identifique as duas variáveis que se correlacionam com maior intensidade. Ajuste um modelo de regressão linear entre elas. Utilize a variável com maior média como variável preditora.

```
ajuste <- lm(Petal.Width ~ Petal.Length, data=iris)
```

5. Exiba os resultados do modelo de regressão ajustado no passo anterior em uma tabela. Interprete o resultado obtido.

```
summary(ajuste)
```

```
##
## Call:
## lm(formula = Petal.Width ~ Petal.Length, data = iris)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.56515 -0.12358 -0.01898  0.13288  0.64272
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.363076   0.039762  -9.131  4.7e-16 ***
## Petal.Length   0.415755   0.009582  43.387 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2065 on 148 degrees of freedom
```

```
## Multiple R-squared:  0.9271, Adjusted R-squared:  0.9266
## F-statistic: 1882 on 1 and 148 DF,  p-value: < 2.2e-16

lm_traducao <- function(ajuste){
  traducao <- summary(ajuste)$coefficients
  colnames(traducao) <- c("Estimativa", "Erro Padrão", "t", "p-valor")
  rownames(traducao) <- c("Intercepto", "Comprimento da Pétala")
  return(traducao)
}

library(knitr)
library(kableExtra)
kable(lm_traducao(ajuste), format="latex", booktabs=TRUE,
      caption="\\label{CarsLM_4} Resultado do ajuste de um modelo linear
aos dados \\texttt{iris}." , digits=4) %>%
  kable_styling(latex_options="hold_position")
```

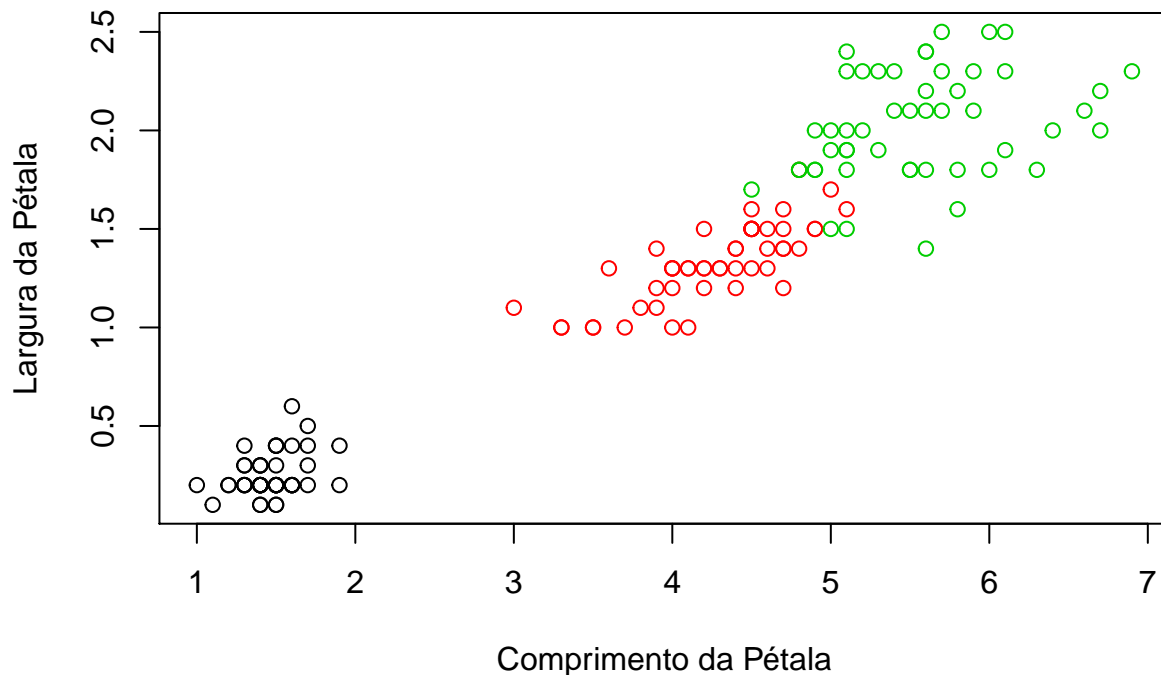
Tabela 1: Resultado do ajuste de um modelo linear aos dados iris.

	Estimativa	Erro Padrão	t	p-valor
Intercepto	-0.3631	0.0398	-9.1312	0
Comprimento da Pétala	0.4158	0.0096	43.3872	0

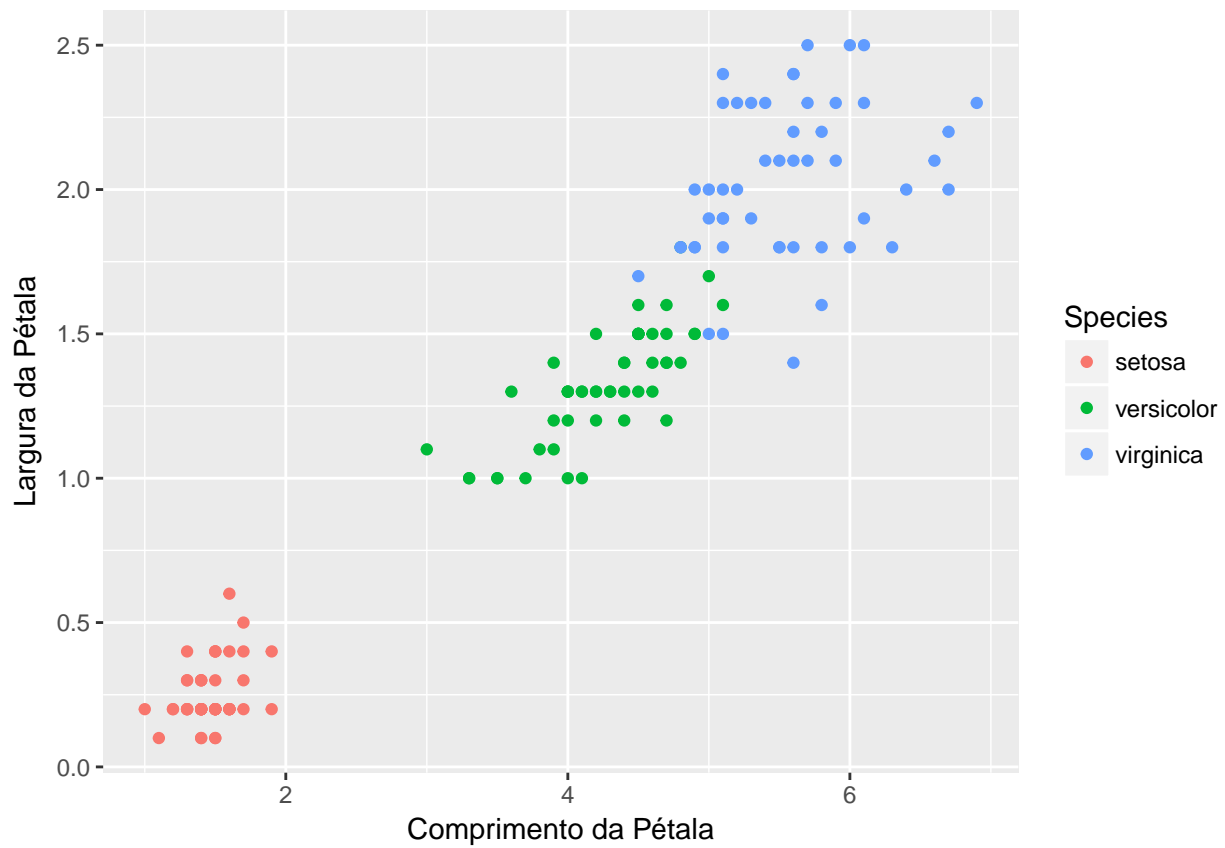
6. Crie um gráfico de dispersão relacionando estas duas variáveis. Pinte cada ponto com uma cor associada à sua espécie.

```
plot(Petal.Width ~ Petal.Length, data=iris, col=as.numeric(Species),
     xlab="Comprimento da Pétala", ylab="Largura da Pétala")

library(ggplot2)
```

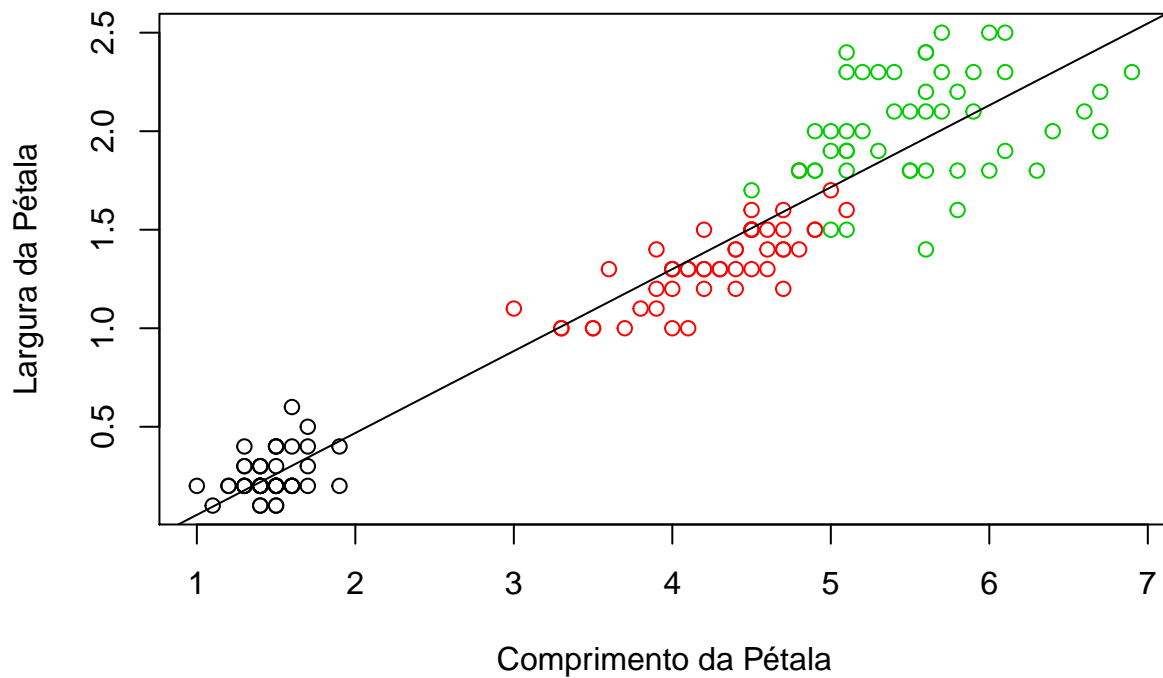


```
ggplot(iris, aes(x=Petal.Length, y=Petal.Width)) +  
  geom_point(aes(colour=Species)) +  
  labs(x="Comprimento da Pétala", y="Largura da Pétala")
```

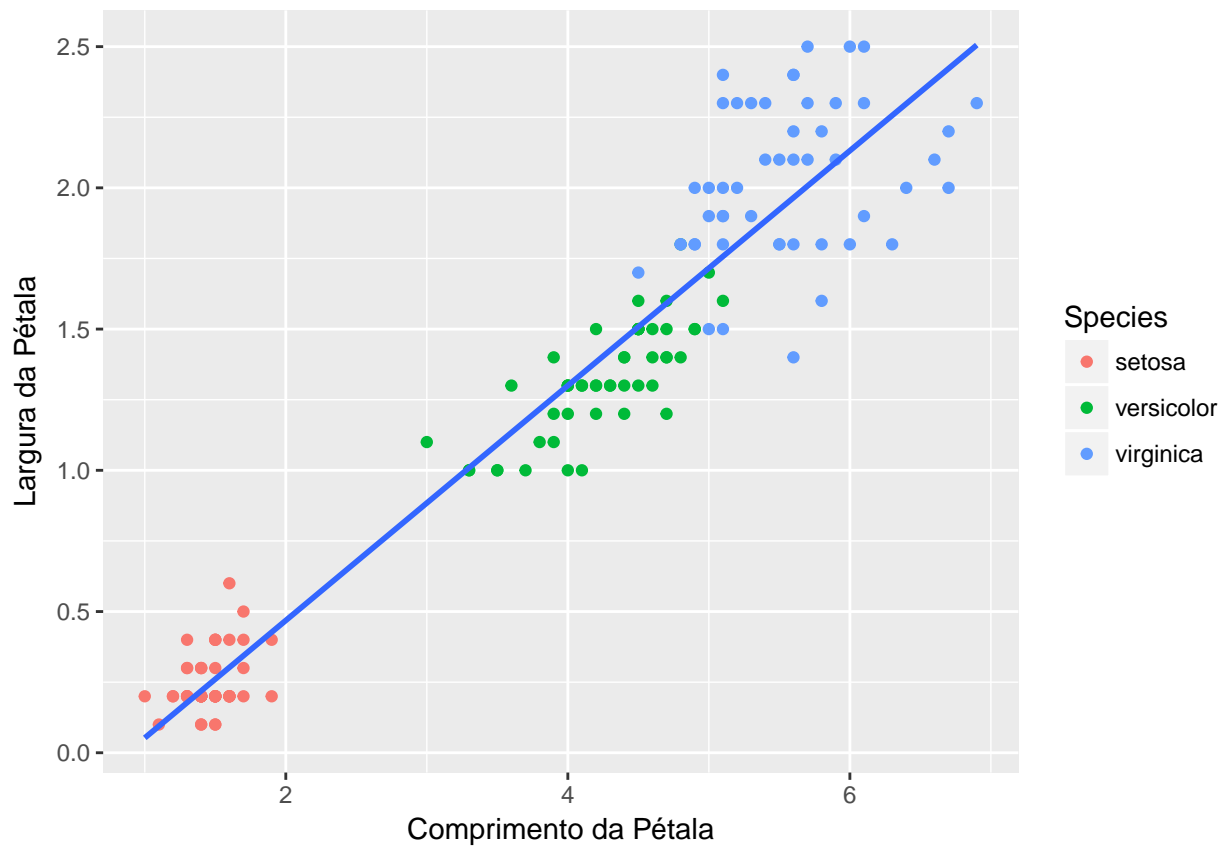


7. Adicione a reta de regressão ao gráfico criado no passo anterior.

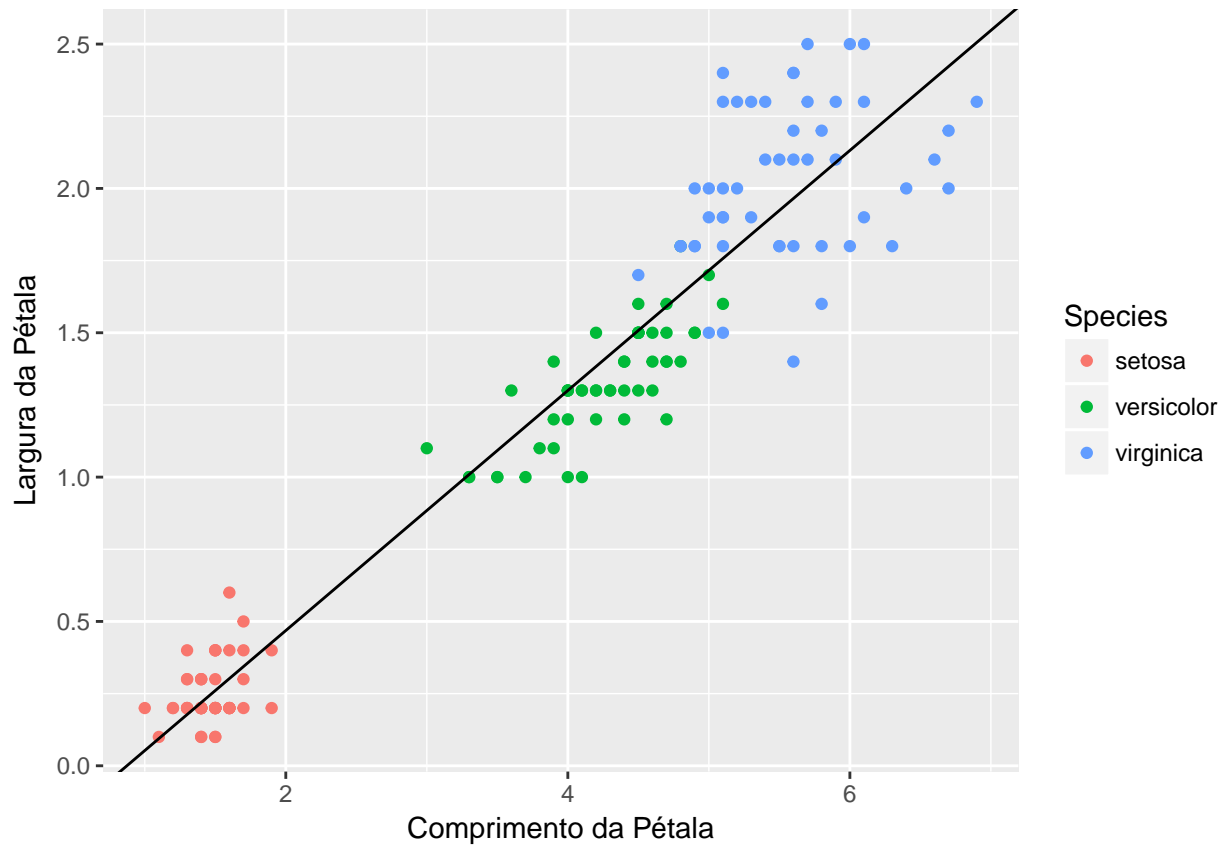
```
plot(Petal.Width ~ Petal.Length, data=iris, col=as.numeric(Species),  
      xlab="Comprimento da Pétala", ylab="Largura da Pétala")  
abline(ajuste)
```



```
ggplot(iris, aes(x=Petal.Length, y=Petal.Width)) +  
  geom_point(aes(colour=Species)) +  
  labs(x="Comprimento da Pétala", y="Largura da Pétala") +  
  stat_smooth(method="lm", se=FALSE)
```



```
ggplot(iris, aes(x=Petal.Length, y=Petal.Width)) +
  geom_point(aes(colour=Species)) +
  labs(x="Comprimento da Pétala", y="Largura da Pétala") +
  geom_abline(slope=coef(ajuste)[2], intercept=coef(ajuste)[1])
```

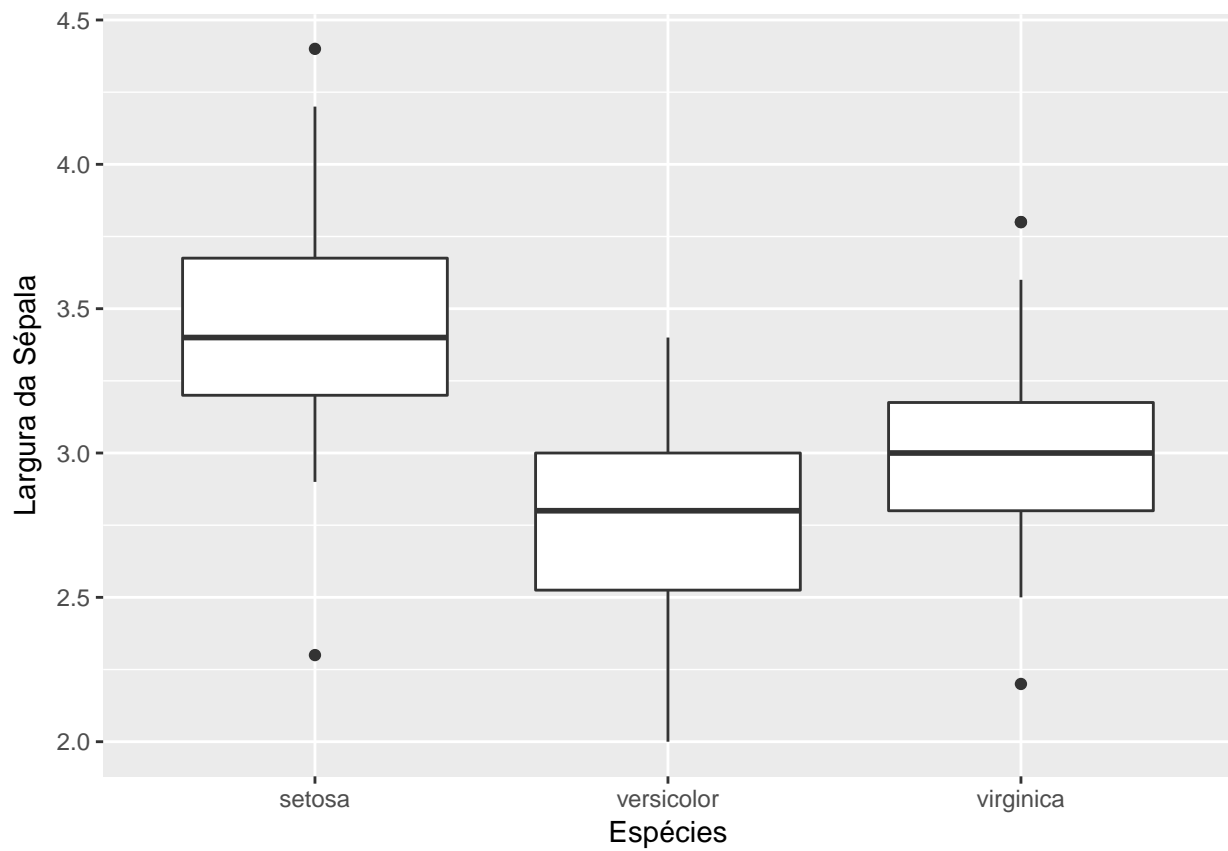


8. Faça um boxplot comparando as observações da variável `Sepal.Width` entre as três espécies de plantas.

```
boxplot(Sepal.Width ~ Species, data=iris, xlab="Espécies",
  ylab="Largura da Sépala")
```



```
ggplot(iris, aes(x=Species, y=Sepal.Width)) +  
  geom_boxplot() +  
  labs(x="Espécies", y="Largura da Sépala")
```



9. Teste as hipóteses

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_0 : \text{pelo menos um par } \mu_i \neq \mu_j, \text{ se } i \neq j$$

em que μ_i é a média da variável `Sepal.Width` para os grupos

- $i = 1$ (setosa)
- $i = 2$ (versicolor)
- $i = 3$ (virginica)

```
ajuste <- aov(Sepal.Width ~ Species, data=iris)
summary(ajuste)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Species        2  11.35    5.672   49.16 <2e-16 ***
## Residuals     147  16.96    0.115
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

10. Utilize a função `TukeyHSD` para verificar quais diferenças entre as médias são diferentes de zero de maneira significativa, com nível de confiança de 95%.

```
TukeyHSD(ajuste)
```

```
##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = Sepal.Width ~ Species, data = iris)
##
## $Species
##              diff              lwr              upr              p adj
## versicolor-setosa -0.658 -0.81885528 -0.4971447 0.0000000
## virginica-setosa   -0.454 -0.61485528 -0.2931447 0.0000000
## virginica-versicolor 0.204  0.04314472  0.3648553 0.0087802
```

```
plot(TukeyHSD(ajuste))
```

