



ME613 - Análise de Regressão

Parte 10

Samara F. Kiihl - IMECC - UNICAMP

file:///Users/amac/Documents/GitHub/ME613-UNICAMP/ME613-UNICAMP.github.io/nulas/slides/parte10/parte10.html#1

1/21

Propriedades

$$\hat{\beta}_{p \times 1} \sim \mathcal{N}_p(\beta, (\mathbf{X}^T \mathbf{X})^{-1} \sigma^2)$$

$$\frac{1}{\sigma} (\mathbf{X}^T \mathbf{X})^{1/2} (\beta - \hat{\beta}) \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I})$$

$$\frac{1}{\sigma^2} (\beta - \hat{\beta})^T (\mathbf{X}^T \mathbf{X}) (\beta - \hat{\beta}) \sim \chi^2(p)$$

$$(n - p) \frac{s^2}{\sigma^2} \sim \chi^2(n - p)$$

Portanto:

$$\frac{1}{ps^2} (\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\beta - \hat{\beta}) \sim F(p, n - p)$$

Região de Confiança

file:///Users/amac/Documents/GitHub/ME613-UNICAMP/ME613-UNICAMP.github.io/nulas/slides/parte10/parte10.html#1

2/21

Intervalo de Confiança para β_k

Um intervalo de $100(1 - \alpha)\%$ de confiança para β_k é dado por:

$$IC(\beta_k, 1 - \alpha) = \left[\hat{\beta}_k - t_{n-p, \alpha/2} \sqrt{\widehat{Var}(\hat{\beta}_k)}; \hat{\beta}_k + t_{n-p, \alpha/2} \sqrt{\widehat{Var}(\hat{\beta}_k)} \right]$$

Exemplo

Estudo sobre diversidade das espécies em Galápagos.

Conjunto de dados: 30 ilhas, 7 variáveis.

Species: the number of plant species found on the island

Endemics: the number of endemic species

Area: the area of the island (km²)

Elevation: the highest elevation of the island (m)

Nearest: the distance from the nearest island (km)

Scruz: the distance from Santa Cruz island (km)

Adjacent: the area of the adjacent island (square km)

5/21

file:///Users/ismac/Documents/GitHub/ME613-UNICAMP/ME613-UNICAMPgithub.io/nulas/slides/parte10/parte10.html#1

Exemplo

IC 95% para $\beta_{Adjacent}$:

```
confint(lmod)[6, ]
```

```
##          2.5 %          97.5 %
## -0.11133622 -0.03827344
```

7/21

file:///Users/ismac/Documents/GitHub/ME613-UNICAMP/ME613-UNICAMPgithub.io/nulas/slides/parte10/parte10.html#1

Exemplo

```
library(faraway)
lmod <- lm(Species ~ Area + Elevation + Nearest + Scruz + Adjacent, gala)
summary(lmod)
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.068221   19.154198   0.3690 0.7153508
## Area         -0.023938    0.022422  -1.0676 0.2963180
## Elevation     0.319465    0.053663   5.9532 3.823e-06
## Nearest       0.009144    1.054136   0.0087 0.9931506
## Scruz        -0.240524    0.215402  -1.1166 0.2752082
## Adjacent     -0.074805    0.017700  -4.2262 0.0002971
##
## n = 30, p = 6, Residual SE = 60.97519, R-Squared = 0.77
```

6/21

file:///Users/ismac/Documents/GitHub/ME613-UNICAMP/ME613-UNICAMPgithub.io/nulas/slides/parte10/parte10.html#1

Região de Confiança para β

Região de Confiança de $100 \times (1 - \alpha)\%$:

$$\frac{1}{ps^2}(\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\beta - \hat{\beta}) \leq F(p, n - p; 1 - \alpha)$$

8/21

file:///Users/ismac/Documents/GitHub/ME613-UNICAMP/ME613-UNICAMPgithub.io/nulas/slides/parte10/parte10.html#1

Exemplo

RC 95% para $\beta_{Adjacent}$ e β_{Area}

```
library(ellipse)
aa <- ellipse(lmod, which=c(2,6), level=0.95)
plot(aa, type="l", ylim=c(-0.13, 0), col="blue")
points(coef(lmod)[2], coef(lmod)[6], pch=19)
abline(v=confint(lmod)[2,], lty=2)
abline(h=confint(lmod)[6,], lty=2)
abline(h=c(max(aa[,2]), min(aa[,2])), lty=2, col="blue")
abline(v=c(max(aa[,1]), min(aa[,1])), lty=2, col="blue")
```

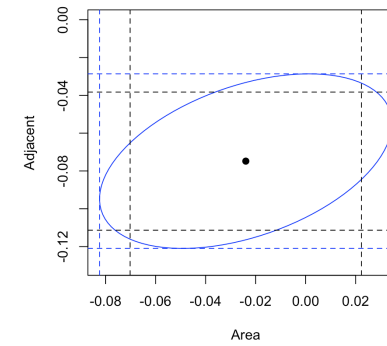
9/21

file:///Users/amac/Documents/GitHub/ME613-UNICAMP/ME613-UNICAMPgithub.io/nulas/slides/parte10/parte10.html#1

9/21

Exemplo

RC 95% para $\beta_{Adjacent}$ e β_{Area}



10/21

file:///Users/amac/Documents/GitHub/ME613-UNICAMP/ME613-UNICAMPgithub.io/nulas/slides/parte10/parte10.html#1

10/21

Teste de Hipótese Linear

Teste de Hipótese Linear

Teste de hipótese linear:

$$H_0: \mathbf{R}_{r \times p} \boldsymbol{\beta}_{p \times 1} = \mathbf{q}_{r \times 1}$$

$$H_1: \mathbf{R}_{r \times p} \boldsymbol{\beta}_{p \times 1} \neq \mathbf{q}_{r \times 1}$$

Para testar, começamos pensando no vetor de discrepância com relação à H_0 :

$$\mathbf{R}_{r \times p} \hat{\boldsymbol{\beta}}_{p \times 1} - \mathbf{q}_{r \times 1} = \mathbf{m}_{r \times 1}$$

queremos medir quão longe \mathbf{m} está de $\mathbf{0}$.

12/21

Teste de Wald para Hipótese Linear

Precisamos então conhecer a distribuição de \mathbf{m} , sob H_0 :

$$E(\mathbf{m}) = E(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}) = \mathbf{R}E(\hat{\boldsymbol{\beta}}) - \mathbf{q} = \mathbf{R}\boldsymbol{\beta} - \mathbf{q} \stackrel{H_0}{=} \mathbf{0}$$

$$Var(\mathbf{m}) = Var(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}) = \mathbf{R}Var(\hat{\boldsymbol{\beta}})\mathbf{R}^T = \sigma^2\mathbf{R}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{R}^T$$

Estatística do teste:

$$\begin{aligned} W &= \mathbf{m}^T [Var(\mathbf{m})]^{-1} \mathbf{m} \\ &= (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})^T [\sigma^2 \mathbf{R}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{R}^T]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}) \\ &\stackrel{H_0}{\sim} \chi^2(r) \end{aligned}$$

13/21

file:///Users/amac/Documents/GitHub/ME613-UNICAMP/ME613-UNICAMPgithub.io/nulas/slides/parte10/parte10.html#1

13/21

Exemplo

- Para $H_0: \beta_j = 0$, definimos

$$\mathbf{R} = [0 \quad 0 \quad \dots \quad 1 \quad 0 \quad \dots \quad 0] \text{ e } \mathbf{q} = 0.$$

- Para $H_0: \beta_k = \beta_j$, definimos

$$\mathbf{R} = [0 \quad 0 \quad 1 \quad \dots \quad -1 \quad 0 \quad \dots \quad 0] \text{ e } \mathbf{q} = 0.$$

- Para $H_0: \beta_1 + \beta_2 + \beta_3 = 1$, definimos

$$\mathbf{R} = [0 \quad 1 \quad 1 \quad 1 \quad 0 \quad \dots \quad 0] \text{ e } \mathbf{q} = 1.$$

15/21

file:///Users/amac/Documents/GitHub/ME613-UNICAMP/ME613-UNICAMPgithub.io/nulas/slides/parte10/parte10.html#1

15/21

Teste de Wald para Hipótese Linear

Problema: não conhecemos σ^2 . Temos que utilizar um estimador para σ^2 : s^2 .

Sabemos que $(n-p) \frac{s^2}{\sigma^2} \sim \chi^2(n-p)$.

A estatística do teste é:

$$\begin{aligned} F &= \frac{W}{r} \frac{\sigma^2}{s^2} \\ &= \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})^T [\mathbf{R}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{R}^T]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})}{r} \frac{\sigma^2}{s^2} \\ &= \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})^T [s^2 \mathbf{R}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{R}^T]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})}{r} \\ &\stackrel{H_0}{\sim} F_{r, n-p} \end{aligned}$$

14/21

file:///Users/amac/Documents/GitHub/ME613-UNICAMP/ME613-UNICAMPgithub.io/nulas/slides/parte10/parte10.html#1

14/21

Exemplo

- Para $H_0: \beta_0 = 0, \beta_1 = 0$ e $\beta_2 = 0$, definimos

$$\mathbf{R} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \end{pmatrix} \quad \mathbf{q} = \mathbf{0}$$

- Para $H_0: \beta_1 + \beta_2 = 1, \beta_3 + \beta_5 = 0$ e $\beta_4 + \beta_5 = 0$, definimos

$$\mathbf{R} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \quad \mathbf{q} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

16/21

file:///Users/amac/Documents/GitHub/ME613-UNICAMP/ME613-UNICAMPgithub.io/nulas/slides/parte10/parte10.html#1

16/21

Exemplo

```
library(systemfit)
data( "Kmenta" )
modelo <- lm(consump ~ price + income,data=Kmenta)
summary(modelo)$coef
```

```
##           Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) 99.8954229  7.51936214  13.285093 2.090605e-10
## price      -0.3162988  0.09067741  -3.488177 2.815290e-03
## income      0.3346356  0.04542183   7.367285 1.099860e-06
```

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

X_1 : price

X_2 : income

Exemplo

$H_0: \beta_1 = 0$

```
R <- matrix( c(0,1,0),ncol=length(coef(modelo)),byrow=TRUE)
linearHypothesis(modelo,R)
```

```
## Linear hypothesis test
##
## Hypothesis:
## price = 0
##
## Model 1: restricted model
## Model 2: consump ~ price + income
##
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      18 108.660
## 2      17  63.332  1    45.328 12.167 0.002815 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Exemplo

$H_0: \beta_1 = 2$

```
R <- matrix( c(0,1,0),ncol=length(coef(modelo)),byrow=TRUE)
q=2
linearHypothesis(modelo,R,q)
```

```
## Linear hypothesis test
##
## Hypothesis:
## price = 2
##
## Model 1: restricted model
## Model 2: consump ~ price + income
##
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      18 2494.21
## 2      17  63.33  1    2430.9 652.52 5.311e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Exemplo

$H_0: \beta_1 = 2 \text{ e } \beta_2 = 1.$

```
R <- matrix( c(0,1,0,
               0,0,1),ncol=length(coef(modelo)),byrow=TRUE)
q=matrix(c(2,1),ncol=1)
linearHypothesis(modelo,R,q)
```

```
## Linear hypothesis test
##
## Hypothesis:
## price = 2
## income = 1
##
## Model 1: restricted model
## Model 2: consump ~ price + income
##
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      19 7146.8
## 2      17  63.3  2    7083.4 950.7 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Leituras

- [The Matrix Cookbook](#)
- Faraway - [Linear Models with R](#): Seção 3.5.
- Draper & Smith - [Applied Regression Analysis](#): Seção 9.1.