Estatística Aplicada

Regressão linear múltipla

Francisco Cribari-Neto

Departamento de Estatística, Universidade Federal de Pernambuco

2025

O modelo de regressão múltipla

MOTIVAÇÃO: Mensurar o impacto sobre a média de y de mais de uma covariável.

O **modelo linear de regressão múltipla** é dado por

$$\boxed{y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} + e_t, \quad t = 1, \ldots, T}$$

Note que o modelo de regressão simples é obtido como um caso particular quando tomamos $K = 2$.

β_1: média de y_t quando $x_{t2} = \cdots = x_{tK} = 0$.

Para $j = 2, \ldots, K$, $\partial \mu_t / \partial x_{tj} = \beta_j$. Ou seja, β_j mede o impacto sobre a média da resposta quando o j-ésimo regressor aumenta em uma unidade e todos os demais regressores permanecem inalterados.

Exemplo motivador 1

Sejam:

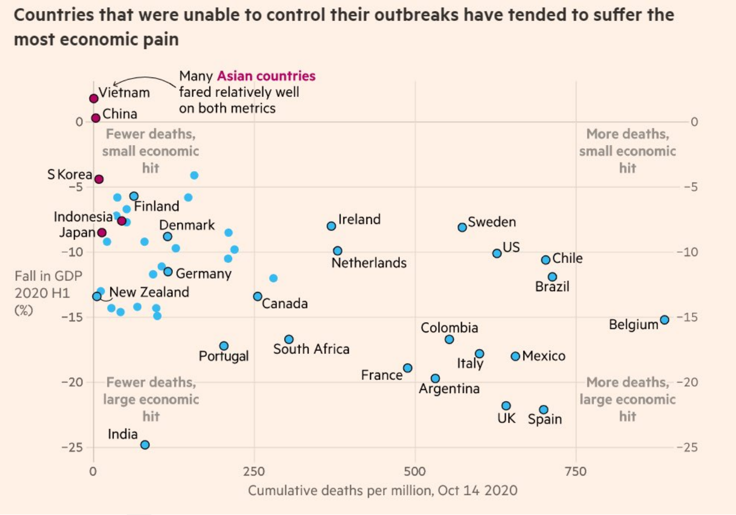
- y_t: Nota em um exame, cuja escala varia entre 0 e 100;
- x_{t2}: Número médio de horas de estudo semanal (aulas + estudos individuais) nos 12 meses anteriores ao exame;
- x_{t3}: Renda familiar;
- x_{t4}: Idade;
- x_{x5}: Sexo (masculino / feminino).

Nosso modelo:

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \beta_5 x_{t5} + e_t.$$

Aqui, β_2 é a pontuação <u>adicional</u> esperada decorrente de um aumento de uma hora semanal de estudo com os demais regressores mantidos constantes.

Exemplo motivador 2: Queda no PIB (%) na primeira metade de 2020 vs mortes cumulativas por Covid-19



Francisco Cribari

Exemplo motivador 2: Queda no PIB (%) na primeira metade de 2020 vs mortes cumulativas por Covid-19

Regressores adicionais:

- Montante relativo e extensão de auxílios individuais;
- Peso relativo do setor de serviços na economia;
- Grau de abertura da economia para o comércio internacional.

O modelo de regressão múltipla (cont.)

Podemos escrever o modelo, em forma matricial, como

$$\boxed{y = X\beta + e}$$

Aqui, $y = (y_1, \ldots, y_T)'$ é um vetor $T \times 1$ de respostas,
$\beta = (\beta_1, \ldots, \beta_K)'$ é o vetor de parâmetros (de dimensão $K \times 1$),
$e = (e_1, \ldots, e_T)'$ é um vetor $T \times 1$ contendo os erros e X é uma matriz $T \times K$ ($K < T$) com observações sobre os regressores.

Quando o modelo contém um intercepto, a primeira coluna de X é um vetor de uns.

O modelo de regressão múltipla (cont.)

Ou seja,

$$X = \begin{bmatrix} 1 & x_{12} & x_{13} & \dots & x_{1K} \\ 1 & x_{22} & x_{23} & \dots & x_{2K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{T2} & x_{T3} & \dots & x_{TK} \end{bmatrix}.$$

Podemos também escrever o modelo como

$$\boxed{y_t = x_t'\beta + e_t, \quad t = 1, \dots, T}$$

em que x_t é a t-ésima linha de X (como vetor coluna).

Suposições

Para o processo inferencial seja viabilizado, precisamos fazer algumas suposições.

SUPOSIÇÕES

[S0] O modelo em uso está correto, ou seja, ele fornece uma boa representação para a população / fenômeno de interesse.

[S1] O erro possui média zero, ou seja, $\mathbb{E}(e_t) = 0 \ \forall t$.

[S2] A variância do erro é constante, i.e.,

$$\mathrm{var}(e_t) = \sigma^2 \ \forall t$$

$(0 < \sigma^2 < \infty)$.

[S3] Os erros são não correlacionados, i.e.,

$$\mathrm{cov}(e_t, e_s) = 0 \ \forall t \neq s.$$

Suposições (cont.)

[S4] As colunas de X são linearmente independentes.

Em outras palavras, $\text{posto}(X) = K$.

Lembre que o **posto** de uma matriz é a dimensão da maior submatriz quadrada com determinante diferente de zero.

Em outras palavras, $\nexists\, c = (c_1, \ldots, c_K)' \neq 0$ tal que

$$c_1 + c_2 x_{t2} + \cdots + c_K x_{tK} = 0 \quad \forall\, t.$$

Para estimação intervalar e testes de hipóteses, adicionamos a seguinte suposição (que não é necessária para estimação pontual):

[S5] Os erros são normalmente distribuídos.

Revisão

DEF. Uma **forma quadrática** é uma função escalar definida como

$$Q(z) = z'Az,$$

em que z é um vetor de dimensão $K \times 1$ e A é uma matriz de dimensão $K \times K$.

Note que $Q(z)$ é um escalar.

A seguir, abordaremos a **derivada de uma função escalar em relação a um vetor**. Esse conceito representa um vetor de derivadas parciais, onde cada elemento corresponde à derivada da função em relação a uma das variáveis do vetor.

Revisão

Seja A uma matriz simétrica de dimensão $K \times K$ e sejam z e w vetores $K \times 1$. Então,

$$\frac{\partial z'w}{\partial z} = \frac{\partial(z_1 \times w_1 + \cdots + z_K \times w_K)}{\partial z} = w.$$

(Ou seja, derivamos o escalar $z'w$ em relação a cada elemento de z.)

Adicionalmente,

$$\frac{\partial z'Az}{\partial z} = 2Az.$$

Se A não for simétrica, então $\partial z'Az/\partial z = (A + A')z$.

Estimação pontual

OBJETIVO: Estimar β, o vetor de parâmetros do modelo $y = X\beta + e$.

MÉTODO DE ESTIMAÇÃO: Mínimos quadrados ordinários, ou seja, minimizaremos $S \equiv S(\beta) = \sum_{t=1}^{T} e_t^2 = e'e$.

Note que

$$
\begin{aligned}
S = e'e &= (y - X\beta)'(y - X\beta) \\
&= y'y - \beta'X'y - y'X\beta + \beta'X'X\beta \\
&= y'y - 2\beta'X'y + \beta'X'X\beta.
\end{aligned}
$$

FATO: $X'X$ é simétrica.

FATO: Sob [S4], $X'X$ é não singular.

Pausa para revisão

RESULTADO 1 Seja A uma matriz simétrica de ordem $T \times T$ e seja B uma matriz $T \times K$. Então, $B'AB$ é simétrica. Tomando $A = I_T$, temos que $B'B$ é simétrica.

RESULTADO 2 Seja B uma matriz de dimensão $T \times K$ com posto K. Então, $B'B$ é não singular.

Pausa para revisão (cont.)

Seja A uma matriz simétrica de dimensão $K \times K$ e seja z um vetor $K \times 1$.

- A é **positiva definida** se $z'Az > 0 \; \forall z \neq 0$.
- A é **positiva semi-definida** se $z'Az \geq 0 \; \forall z$.
- A é **negativa definida** se $z'Az < 0 \; \forall z \neq 0$.
- A é **negativa semi-definida** se $z'Az \leq 0 \; \forall z$.

Se A é positiva-definida (p.d.), então $|A| > 0$.

Se A é p.d., então A^{-1} é p.d.

Se A é p.d. e c é um escalar positivo, então cA é p.d.

Se A e B são p.d. (e de mesma dimensão), então $A + B$ é p.d.

Se A é p.d., então $a_{ii} > 0$ para $i = 1, \ldots, K$.

Estimação pontual (cont.)

A condição de primeira ordem é

$$\frac{\partial S}{\partial \beta} = -2X'y + 2(X'X)\beta = 0.$$

Segue que

$$X'X\,b = X'y,$$

em que $b = (b_1, \ldots, b_K)'$ é o **estimador de mínimos quadrados ordinários** de β.

Assim,

$$\boxed{b = (X'X)^{-1}X'y}$$

Mas e a condição de segunda ordem? Note que $\partial^2 S/\partial\beta\partial\beta' = 2\,X'X$, que é positiva definida. Ou seja, trata-se de ponto de mínimo.

Pausa para revisão

RESULTADO Seja B uma matriz de dimensão $T \times K$ com posto K. Então, $\text{posto}(B'B) = K$ e $B'B$ é positiva definida. Adicionalmente, $c\,B'B$ é positiva definida $\forall c > 0$.

CONSEQUÊNCIA 2 X'X é positiva definida. Assim, a condição de segunda ordem é satisfeita.

RESULTADO ADICIONAL 1 Seja C uma matriz de dimensão $K \times T$. Então, CC' é positiva semi-definida. Se $\text{posto}(C) = K$, então CC' é positiva definida.

RESULTADO ADICIONAL 2 (Decomposição de Choleski) Seja A uma matriz positiva definida de dimensão $K \times K$. Então, $A = LL'$, em que L é uma matriz triangular inferior de ordem $K \times K$ com elementos diagonais positivos.

OBS.: Alguns autores escrevem $A = L'L$. Nesse caso, L é triangular superior.

Pausa para revisão (cont.)

EXEMPLO SIMPLES DE DECOMPOSIÇÃO DE CHOLESKI Considere a matriz

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 13 \end{bmatrix}.$$

Temos que $A = LL'$, em que

$$L = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}.$$

Note que

$$LL' = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 13 \end{bmatrix} = A.$$

OBS.: Se a decomposição for definida como $L'L$, então L é triangular superior, ou seja,

$$L = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}.$$

Pausa para revisão (cont.)

O R utiliza a segunda notação, ou seja, $L'L$.

```
> A = matrix(c(1, 2, 2, 13), nrow = 2, ncol = 2)
> A
     [,1] [,2]
[1,]    1    2
[2,]    2   13
> A.chol = chol(A)
> A.chol
     [,1] [,2]
[1,]    1    2
[2,]    0    3
```

Estimação pontual no R

Para estimar regressões lineares usando o software R, use a função `lm` ('linear model'). Por exemplo,

```
> ajuste = lm(y ~ x2 + x3 + x4)
> summary(ajuste)
```

Para remover o intercepto do modelo, adicione `-1`:
`lm(y ~ -1 + x2 + x3 + x4)`.

Para mais detalhes,

```
> help(lm)
```

Estimação pontual no R I

Façamos um exemplo.

```
> # Carregar o conjunto de dados
> data(mtcars)
> # Detalhes sobre os dados
> help(mtcars)
> # Ver as primeiras linhas
> head(mtcars)
                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
> # sumário dos dados
> with(mtcars, summary(mpg)) |> round(2)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  10.40   15.43   19.20   20.09   22.80   33.90 
> with(mtcars, summary(wt)) |> round(2)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
```

Estimação pontual no R II

```
   1.51    2.58    3.33    3.22    3.61    5.42
> with(mtcars, summary(hp)) |> round(2)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  52.00   96.50  123.00  146.69  180.00  335.00
> # correlações
> with(mtcars, cor(mpg, wt)) |> round(2)
[1] -0.87
> with(mtcars, cor(mpg, hp)) |> round(2)
[1] -0.78
> with(mtcars, cor(wt, hp)) |> round(2)
[1] 0.66
> # Ajustar o modelo de regressão linear
> modelo = lm(mpg ~ wt + hp, data = mtcars)
> summary(modelo)

Call:
lm(formula = mpg ~ wt + hp, data = mtcars)

Residuals:
         Min           1Q       Median           3Q          Max
-3.940979469 -1.600221937 -0.182013641  1.049855518  5.853790850
```

Estimação pontual no R III

```
Coefficients:
                Estimate     Std. Error  t value      Pr(>|t|)
(Intercept) 37.22727011645  1.59878753800 23.28469   < 2.22e-16
wt          -3.87783074240  0.63273349438 -6.12870 0.0000011196
hp          -0.03177294698  0.00902970968 -3.51871    0.0014512

Residual standard error: 2.59341178 on 29 degrees of freedom
Multiple R-squared:  0.826785452,       Adjusted R-squared:
0.814839621
F-statistic: 69.2112134 on 2 and 29 DF,  p-value: 9.10905439e-12
```

Estimação pontual no R

A estimativa de β_2 obtida usando o modelo de regressão simples foi -5.344. Aqui, ela passa a ser -3.8778. Concluímos que para cada aumento de 1000 libras no peso do carro a eficiência (medida em MPG) diminui em 3.8778 quando a potência do motor é mantida constante.

A estimativa de β_3 é -0.0318. Assim, um aumento de uma hp ("horsepower") de potência do motor conduz a uma redução de eficiência de 0.0318 (medida em MPG) quando o peso do carro é mantido constante. Ou seja, para um dado peso do carro, um aumento de 10 hp na potência conduz a uma redução de eficiência de aproximadamente 0.3 MPG.

Resíduos

Os erros representam discrepâncias entre as respostas $(y_1, \ldots, y_T)$ e a reta de regressão populacional $(x_t'\beta)$:

$$e_t = y_t - x_t'\beta = \beta_1 - \beta_2 x_{t2} - \cdots - \beta_K x_{tK}.$$

Podemos 'estimar' tais erros tomando as discrepâncias entre as respostas $(y_1, \ldots, y_T)$ e a reta de regressão amostral $(x_t'b)$. Essas medidas são chamadas de **resíduos**:

$$\hat{e}_t = y_t - x_t'b = y_t - b_1 - b_2 x_{t2} - \cdots - b_K x_{tK}.$$

Denotaremos por $\hat{e}$ o vetor de resíduos de mínimos quadrados ordinários, ou seja, $\hat{e} = (\hat{e}_1, \ldots, \hat{e}_T)'$.

Estimação de σ^2

OBJETIVO: Obter um estimador para σ^2.

Lembre que $\sigma^2 = \text{var}(e_t) = \mathbb{E}(e_t^2)$ (dado que e_t possui media zero).

IDEIA: Podemos usar a variância amostral dos resíduos como um estimador para a variância populacional dos erros.

MELHORAMENTO: Ao fazer isso, incorporaremos uma 'correção de graus de liberdade'.

Temos, assim, o seguinte estimador da variância dos erros:

$$\boxed{\hat{\sigma}^2 = \frac{1}{T-K}\sum_{t=1}^{T} \hat{e}_t^2}$$

Ou seja, $\hat{\sigma}^2 = \hat{e}'\hat{e}/(T-K)$.

Pausa para revisão

Seja $Z = (Z_1, \ldots, Z_m)'$ um **vetor aleatório** de dimensão $m \times 1$ com média $\xi = (\xi_1, \ldots, \xi_m)'$.

A **matrix de covariância** de Z é

$$\text{cov}(Z) = \mathbb{E}\left[(Z-\xi)(Z-\xi)'\right].$$

Os elementos diagonais são as variâncias: $\text{var}(Z_1), \ldots, \text{var}(Z_m)$.

Os demais elementos são covariâncias: $\text{cov}(Z_1, Z_2), \text{cov}(Z_1, Z_3), \ldots$.

Uma vez que $\text{cov}(Z_i, Z_j) = \text{cov}(Z_j, Z_i)$ para todo $i \neq j$, $\text{cov}(Z)$ é **simétrica**.

ALHOS E BUGALHOS: Nas derivações a seguir, não confundir $\text{cov}(b)$ com $\text{cov}(e)$. Note também que as dimensões são distintas: $K \times K$ e $T \times T$ (respec.).

OBS.: Dado que $\mathbb{E}(e) = 0$ (por [S1]), segue que $\text{cov}(e) = \mathbb{E}(ee')$.

Propriedades de b

OBJETIVO: Estudar as propriedades de b, o EMQO de β. Assumiremos que [S4] vale. Note que [S0] garante que $X\beta + e$ fornece uma boa representação para y.

PROPRIEDADE 1 Sob [S0] e [S1], b é não viesado.

Note que

$$
\begin{aligned}
b = (X'X)^{-1}X'y &\overset{[S0]}{=} (X'X)^{-1}X'(X\beta + e) \\
&= \beta + (X'X)^{-1}X'e.
\end{aligned}
$$

Segue que

$$
\begin{aligned}
\mathbb{E}(b) &= \mathbb{E}(\beta) + \mathbb{E}[(X'X)^{-1}X'e] \\
&= \beta + (X'X)^{-1}X'\mathbb{E}(e) \\
&\overset{[S1]}{=} \beta + 0 \\
&= \beta.
\end{aligned}
$$

Ou seja, $\mathbb{E}(b) = \beta \ \forall \beta \in \mathbb{R}^K$. Note que só precisamos de [S0] e [S1].

Propriedades de b (cont.)

PROPRIEDADE 2 Sob [S0], [S1], [S2] e [S3], $\text{cov}(b) = \sigma^2(X'X)^{-1}$.

Note que

$$\text{cov}(b) = \mathbb{E}\{[b - \mathbb{E}(b)][b - \mathbb{E}(b)]'\} \overset{\text{[S0] e [S1]}}{=} \mathbb{E}[(b-\beta)(b-\beta)'].$$

Mas $b - \beta = (X'X)^{-1}X'e$. Segue que

$$\begin{aligned}
\text{cov}(b) &= \mathbb{E}[(X'X)^{-1}X'ee'X(X'X)^{-1}] \\
&= (X'X)^{-1}X'\mathbb{E}(ee')X(X'X)^{-1} \\
&\overset{\text{[S1] a [S3]}}{=} (X'X)^{-1}X'\sigma^2 I_T X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1},
\end{aligned}$$

em que I_T é a matriz identidade de ordem T.

Note que $\text{cov}(e) = \sigma^2 I_T$.

Propriedades de b (cont.)

Como notamos,

$$\text{cov}(e) = \mathbb{E}(ee') = \begin{bmatrix} \mathbb{E}(e_1^2) & \mathbb{E}(e_1e_2) & \cdots & \mathbb{E}(e_1e_T) \\ \mathbb{E}(e_2e_1) & \mathbb{E}(e_2^2) & \cdots & \mathbb{E}(e_2e_T) \\ \vdots & \vdots & \vdots & \vdots \\ \mathbb{E}(e_Te_1) & \mathbb{E}(e_Te_2) & \cdots & \mathbb{E}(e_T^2) \end{bmatrix}.$$

Segue que

$$\text{cov}(e) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 I_T.$$

Para esse resultado, usamos [S1], [S2] e [S3].

Propriedades de b (cont.)

Note que podemos facilmente estimar essa matriz de variâncias e covariâncias:

$$\widehat{\text{cov}}(b) = \hat{\sigma}^2(X'X)^{-1}.$$

Os **erros-padrão** de $b_1, \ldots, b_K$ são obtidos tomando as raízes quadradas dos elementos diagonais de $\widehat{\text{cov}}(b)$.

PROPRIEDADE 3 O estimador de máxima verossimilhança de β obtido sob a suposição de normalidade coincide com b.

A <u>função de verossimilhança</u> construída assumindo que $y_1, \ldots, y_T$ são variáveis aleatórias independentes tais que $y_t \sim \mathcal{N}(x_t'\beta, \sigma^2)$, $t = 1, \ldots, T$, é dada por

$$L \equiv L(\beta, \sigma^2 \,|\, y, X) = (2\pi\sigma^2)^{-T/2} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right).$$

A <u>função de log-verossimilhança</u> é

$$\ell \equiv \ell(\beta, \sigma^2 \,|\, y, X) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta).$$

Propriedades de b (cont.)

As <u>condições de primeira ordem</u> são

$$\frac{\partial \ell}{\partial \beta} = -\frac{1}{2\sigma^2}(-2X'y + 2X'X\beta) = 0,$$
$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{T}{2} \times \frac{1}{\sigma^2} + \frac{1}{2\sigma^4}e'e = 0.$$

Da primeira equação, obtemos

$$\boxed{\hat{\beta}_{\text{MV}} = (X'X)^{-1}X'y}$$

que é igual a b.

Da segunda equação, obtemos $\hat{e}'\hat{e}/\hat{\sigma}^4 = T/\hat{\sigma}^2$. Ou seja,

$$\boxed{\hat{\sigma}^2_{\text{MV}} = \frac{\hat{e}'\hat{e}}{T}}$$

que é diferente de $\hat{\sigma}^2$.

Propriedades de b (cont.)

PROPRIEDADE 4 b é um **estimador linear**, pois pode ser escrito como um múltiplo de y: $b = Ay$, em que $A = (X'X)^{-1}X'$.

PROPRIEDADE 5 Se $y \sim \mathcal{N}$, então $b \sim \mathcal{N}$.

PROPRIEDADE 6 (Teorema de Gauss-Markov) Sob [S0], [S1], [S2] e [S3], b é o melhor estimador linear não viesado de β.

Ou seja, $\forall a \in \mathbb{R}^K$ e para qualquer estimador b^* de β que seja linear e não viesado, segue que

$$\text{var}(a'b^*) \geq \text{var}(a'b) \quad \Rightarrow \quad \text{var}(a'b^*) - \text{var}(a'b) \geq 0.$$

Ou seja,

$$a'[\text{cov}(b^*)]a - a'[\text{cov}(b)]a \geq 0 \quad \Rightarrow \quad a'[\text{cov}(b^*) - \text{cov}(b)]a \geq 0.$$

Ou seja, a diferença entre $\text{cov}(b^*)$ e $\text{cov}(b)$ deve ser positiva semi-definida.

Propriedades de b (cont.)

PROVA. Defina a classe de estimadores lineares de β como

$$b^* = [(X'X)^{-1}X' + C]y,$$

em que C é uma matriz (não-aleatória) de dimensão $K \times T$.

Note que o estimador de MQO b é um membro desta classe quando C é uma matriz de zeros.

Note ainda que

$$\begin{aligned} b^* &= [(X'X)^{-1}X' + C]y \\ &\overset{[S0]}{=} [(X'X)^{-1}X' + C](X\beta + e) \\ &= \beta + Ce + CX\beta + (X'X)^{-1}X'e. \end{aligned}$$

Dado que $\mathbb{E}(e) = 0$ (por [S1]), então $\mathbb{E}(b^*) = \beta + CX\beta$.

Assim, para que b^* seja não viesado, requeremos $CX = 0$.

Propriedades de b (cont.)

Temos então

$$b^* - \beta = Ce + (X'X)^{-1}X'e = [C + (X'X)^{-1}X']e.$$

Portanto,

$$(b^* - \beta)(b^* - \beta)' = [C + (X'X)^{-1}X']ee'[C' + X(X'X)^{-1}].$$

A matriz de covariância de b^* é

$$\begin{aligned}
\text{cov}(b^*) &= \mathbb{E}[(b^* - \beta)(b^* - \beta)'] \\
&= [C + (X'X)^{-1}X']\mathbb{E}(ee')[C' + X(X'X)^{-1}] \\
&\overset{\text{[S1] a [S3]}}{=} [C + (X'X)^{-1}X']\sigma^2 I_T[C' + X(X'X)^{-1}] \\
&= \sigma^2[CC' + (X'X)^{-1}X'C' + CX(X'X)^{-1} + (X'X)^{-1}X'X(X'X)^{-1}] \\
&= \sigma^2[CC' + (X'X)^{-1}].
\end{aligned}$$

(Note que usamos acima a propriedade de não viés, i.e., $CX = 0$ e $X'C' = 0$, e também [S0], [S1], [S2] e [S3].)

Propriedades de b (cont.)

Assim,

$$\text{cov}(b^*) = \sigma^2(X'X)^{-1} + \sigma^2 CC' = \text{cov}(b) + \sigma^2 CC'.$$

Ou seja,

$$\text{cov}(b^*) - \text{cov}(b) = \sigma^2 CC',$$

uma matriz positiva semi-definida.

$\square$

OBS. Se $C = 0$, então $b^* = b$ e, claro, $\text{cov}(b^*) - \text{cov}(b) = 0$.

OBS. Se $\text{posto}(C) = K$ (que implica $b^* \neq b$), então $\sigma^2 CC'$ é positiva definida.

OBS. Lembre desse resultado que revisamos: *Seja C uma matriz de dimensão $K \times T$. Então, CC' é positiva semi-definida. Se $\text{posto}(C) = K$, então CC' é positiva definida.*

Duas propriedades de uma matriz importante

Considere a seguinte matriz (de ordem $T \times T$):

$$M = I_T - X(X'X)^{-1}X'.$$

[PROP. 1] A matriz M é **simétrica**:

$$M' = (I_T - X(X'X)^{-1}X')' = I_T - X(X'X)^{-1}X' = M.$$

[PROP. 2] A matriz M é **idempotente**:

$$\begin{aligned} MM &= (I_T - X(X'X)^{-1}X')(I_T - X(X'X)^{-1}X') \\ &= I_T - X(X'X)^{-1}X' - X(X'X)^{-1}X' + X(X'X)^{-1}X'X(X'X)^{-1}X' \\ &= I_T - X(X'X)^{-1}X' - X(X'X)^{-1}X' + X(X'X)^{-1}X' \\ &= I_T - X(X'X)^{-1}X' = M. \end{aligned}$$

Mais adiante, denotaremos a matriz $X(X'X)^{-1}X'$ por H e a chamaremos de **matriz chapéu**.

Revisão

Vimos que se $Z_1, Z_2, \ldots, Z_k \overset{\text{indep}}{\sim} \mathcal{N}(0,1)$, então $\sum_{i=1}^{k} Z_i^2 \sim \chi_k^2$.

Agora, sejam $Z_1, Z_2, \ldots, Z_k$ variáveis aleatórias independentes tal que $Z_i \sim \mathcal{N}(\mu_i, 1)$, para $i = 1, \ldots, k$. Então $\sum_{i=1}^{k} Z_i^2 \sim \chi_{k,\lambda}^2$, em que $\lambda = \sum_{i=1}^{k} \mu_i^2$. Essa é a **distribuição qui-quadrado não central**.

Dois resultados importantes

LEMA 1. Seja $s \sim \mathcal{N}(\psi, I_T)$ e seja A uma matriz não estocástica de dimensão $T \times T$ simétrica e idempotente. Então,

$$s'As \sim \chi^2_{a,\lambda},$$

em que $a = \text{posto}(A)$ e $\lambda = \psi' A \psi / 2$.

LEMA 2. Seja $s \sim \mathcal{N}(\psi, I_T)$, seja A uma matriz não estocástica simétrica de dimensão $T \times T$ e seja B uma matriz não estocástica de dimensão $K \times T$. Se

$$BA = 0,$$

então

$$s'As \quad \text{e} \quad Bs$$

são independentes.

Propriedades de b (cont.)

PROPRIEDADE 7 Sob normalidade (e [S0] a [S3]), b e $\hat{\sigma}^2$ são independentes.

PROVA. Dado que $e \sim \mathcal{N}(0, \sigma^2 I_T)$, segue que $y \sim \mathcal{N}(\mu, \sigma^2 I_T)$ e

$$b = (X'X)^{-1}X'y \sim \mathcal{N}(\beta, \sigma^2(X'X)^{-1}).$$

Seja $z = y/\sigma$. Segue que

$$z \sim \mathcal{N}(\sigma^{-1}X\beta, I_T).$$

Seja $M = I_T - X(X'X)^{-1}X'$. Temos que

$$\frac{T-K}{\sigma^2}\hat{\sigma}^2 = z'Mz.$$

Mostremos que isso é verdade.

Propriedades de b (cont.)

Temos

$$\begin{aligned}
\frac{T-K}{\sigma^2}\hat{\sigma}^2 &= \frac{T-K}{\sigma^2} \times \overbrace{\frac{\hat{e}'\hat{e}}{T-K}}^{\hat{\sigma}^2} = \frac{1}{\sigma^2}(y-Xb)'(y-Xb) \\
&= \frac{1}{\sigma^2}[y - X(X'X)^{-1}X'y]'[y - X(X'X)^{-1}X'y] \\
&= \frac{1}{\sigma^2}y'[I_T - X(X'X)^{-1}X']'[I_T - X(X'X)^{-1}X']y \\
&= \frac{1}{\sigma^2}y'M'My.
\end{aligned}$$

Lembre que M é simétrica (i.e., $M = M'$) e idempotente (i.e., $MM = M$). Isso implica que $M'M = M$ e, assim,

$$\frac{T-K}{\sigma^2}\hat{\sigma}^2 = \frac{1}{\sigma^2}y'MMy = \left(\frac{y}{\sigma}\right)' M \left(\frac{y}{\sigma}\right) = z'Mz.$$

Sigamos em frente...

Propriedades de b (cont.)

Pelo Lema 1,

$$z'Mz \sim \chi^2_{a,\lambda}.$$

Aqui, $a = \text{posto}(M)$. Note que, dado que M é idempotente, segue que $\text{posto}(M) = \text{tr}(M)$ e

$$\begin{aligned}
\text{tr}(M) &= \text{tr}(I_T - X(X'X)^{-1}X') \\
&= \text{tr}(I_T) - \text{tr}(X(X'X)^{-1}X') \\
&= T - \text{tr}(X'X(X'X)^{-1}) \\
&= T - \text{tr}(I_K) \\
&= T - K.
\end{aligned}$$

Ou seja, $a = T - K$. Agora precisamos determinar o valor de λ.

Propriedades de b (cont.)

Dado que $z \sim \mathcal{N}(\sigma^{-1}X\beta, I_T)$, segue que

$$\lambda = \frac{\beta' X' M X \beta}{2\sigma^2}.$$

Note que

$$\begin{aligned}
X'MX &= X'[I_T - X(X'X)^{-1}X']X \\
&= X'X - X'X(X'X)^{-1}X'X \\
&= X'X - X'X \\
&= 0.
\end{aligned}$$

Assim, $\lambda = 0$.

Temos, assim, que $z'Mz \sim \chi^2_{T-K}$. Ou seja, $(T-K)/\sigma^2 \times \hat{\sigma}^2 \sim \chi^2_{T-K}$.

Propriedades de b (cont.)

O passo final de nossa prova será aplicar o Lema 2 com $A = M$ e $B = (X'X)^{-1}X'$. Note que

$$\begin{aligned} BA &= (X'X)^{-1}X'[I_T - X(X'X)^{-1}X'] \\ &= (X'X)^{-1}X' - (X'X)^{-1}X'X(X'X)^{-1}X' \\ &= (X'X)^{-1}X' - (X'X)^{-1}X' \\ &= 0. \end{aligned}$$

Portanto, $z'Mz$ e $(X'X)^{-1}X'z$ são independentes. Temos, assim, que

$$\hat{\sigma}^2 \quad \text{e} \quad b$$

são independentes.

$\square$

PROPRIEDADE 8 b e $\hat{\sigma}^2$ são consistentes. (Provaremos mais adiante.)

Propriedades de $\hat{\sigma}^2$

PROPRIEDADE 1 Sob normalidade (e [S0] a [S3]),

$$\boxed{\frac{T-K}{\sigma^2}\hat{\sigma}^2 \sim \chi^2_{T-K}}$$

Note que esse resultado foi provado ao longo da prova de que $\hat{\sigma}^2$ e b são independentes. Ele segue do nosso Lema 1.

PROPRIEDADE 2 Sob normalidade (e [S0] a [S3]), $\hat{\sigma}^2$ é não viesado.

Note que

$$\mathbb{E}\left(\frac{T-K}{\sigma^2}\hat{\sigma}^2\right) = T-K \;\; \text{(média da dist. } \chi^2_{T-K}\text{)}$$

Assim,

$$\mathbb{E}(\hat{\sigma}^2) = \sigma^2.$$

□

Propriedades de $\hat{\sigma}^2$ (cont.)

Note ainda que

$$\text{var}\left(\frac{T-K}{\sigma^2}\hat{\sigma}^2\right) = 2(T-K) \ \text{(variância da dist. } \chi^2_{T-K})$$

Ou seja,

$$\frac{(T-K)^2}{\sigma^4}\text{var}(\hat{\sigma}^2) = 2(T-K).$$

Ou seja,

$$\text{var}(\hat{\sigma}^2) = \frac{2\sigma^4}{T-K} \to 0$$

quando $T \to \infty$.

Propriedades de $\hat{\sigma}^2$ (cont.)

PROPRIEDADE 3 Mesmo sem normalidade (mas sob [S0] a [S3]), $\hat{\sigma}^2$ é não viesado.

PROVA. Note, de início, que

$$\begin{aligned}
\hat{e} &= y - Xb \\
&= y - X(X'X)^{-1}X'y \\
&= [I_T - X(X'X)^{-1}X']y \\
&= My.
\end{aligned}$$

Sob [S0],

$$\begin{aligned}
\hat{e} &= [I_T - X(X'X)^{-1}X'](X\beta + e) \\
&= X\beta - X(X'X)^{-1}X'X\beta + Me \\
&= Me.
\end{aligned}$$

Precisamos obter $\mathbb{E}(\hat{\sigma}^2) = \mathbb{E}[\hat{e}'\hat{e}/(T-K)]$.

Propriedades de $\hat{\sigma}^2$ (cont.)

Note que

$$
\begin{aligned}
\mathbb{E}(\hat{\sigma}^2) = \mathbb{E}\left(\frac{\hat{e}'\hat{e}}{T-K}\right) &= \frac{1}{T-K}\mathbb{E}[e'\overbrace{M'M}^{M}e] \\
&= \frac{1}{T-K}\mathbb{E}[\mathrm{tr}(e'Me)] \\
&= \frac{1}{T-K}\mathbb{E}[\mathrm{tr}(Mee')] \\
&= \frac{1}{T-K}\,\mathrm{tr}(M\,\overbrace{\mathbb{E}(ee')}^{\sigma^2 I_T}) \\
&= \frac{\sigma^2}{T-K}\,\mathrm{tr}(\overbrace{I_T - X(X'X)^{-1}X'}^{M}) \\
&= \frac{\sigma^2}{T-K}\times(T-K) \\
&= \sigma^2.
\end{aligned}
$$

Revisão

O **quantil** $p \in (0, 1)$ da variável aleatória W é

$$w_p = \inf \{w \in \mathbb{R} \mid \Pr(W \le w) \ge p\}.$$

O quantil 0.5 é a mediana e os quantis 0.25 e 0.75 são conhecidos como primeiro e terceiro quartis, respectivamente.

Por exemplo, os quantis de nível 0.90, 0.95, 0.975, 0.99, 0.995 da distribuição normal padrão são

```
> qnorm(c(0.90, 0.95, 0.975, 0.99, 0.995)) |> round(4)
[1] 1.2816 1.6449 1.9600 2.3263 2.5758
```

Para a distribuição t_5, obtemos

```
> qt(c(0.90, 0.95, 0.975, 0.99, 0.995), df = 5) |> round(4)
[1] 1.4759 2.0150 2.5706 3.3649 4.0321
```

Estimação intervalar

OBJETO DE DESEJO: Realizar inferência intervalar sobre os parâmetros do modelo de regressão linear.

Em particular, para um $\alpha \in (0, 1)$ selecionado, desejamos encontrar um intervalo aleatório cujos limites sejam estatísticas e cuja probabilidade de conter o valor verdadeiro do parâmetro seja $1 - \alpha$. No que segue, para $\delta \in (0, 1)$, z_δ denota o quantil δ da distribuição normal padrão.

SUPOSIÇÃO IMPORTANTE: Assumiremos **normalidade**, i.e., nossa suposição [S5].

CASO IRREALISTA Suponha que σ^2 é conhecido. Para $j = 1, \ldots, K$,

$$b_j \sim \mathcal{N}(\beta_j, \mathrm{var}(b_j)),$$

em que $\mathrm{var}(b_j)$ é o elemento (j, j) da matriz $\sigma^2(X'X)^{-1}$.

Note que

$$\frac{b_j - \beta_j}{\sqrt{\mathrm{var}(b_j)}} \sim \mathcal{N}(0, 1).$$

Estimação intervalar (cont.)

Assim,

$$\Pr\left(z_{\alpha/2} \leq \frac{b_j - \beta_j}{\sqrt{\mathrm{var}(b_j)}} \leq z_{1-\alpha/2}\right) = 1 - \alpha.$$

Dada a simetria da distribuição normal, $z_{\alpha/2} = -z_{1-\alpha/2}$. Segue que

$$\Pr\left(b_j - z_{1-\alpha/2}\sqrt{\mathrm{var}(b_j)} \leq \beta_j \leq b_j + z_{1-\alpha/2}\sqrt{\mathrm{var}(b_j)}\right) = 1 - \alpha.$$

Chegamos ao seguinte intervalo de confiança de nível $(1-\alpha) \times 100\%$ para β_j:

$$\boxed{\left[b_j - z_{1-\alpha/2}\sqrt{\mathrm{var}(b_j)}, b_j + z_{1-\alpha/2}\sqrt{\mathrm{var}(b_j)}\right]}$$

Estimação intervalar (cont.)

CASO REALISTA Suponha que σ^2 é desconhecido.

Seja c_{jj} o j-ésimo elemento diagonal de $(X'X)^{-1}, j = 1, \ldots, K$. Trabalharemos com os seguintes resultados:

RESULTADO 1: $(b_j - \beta_j)/\sqrt{\sigma^2 c_{jj}} \sim \mathcal{N}(0, 1)$.

RESULTADO 2: $(T - K)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{T-K}$.

RESULTADO 3: b_j e $\hat{\sigma}^2$ são independentes.

Segue desses três resultados combinados que

$$\frac{b_j - \beta_j}{\sqrt{\widehat{\text{var}}(b_j)}} = \frac{b_j - \beta_j}{\sqrt{\hat{\sigma}^2 c_{jj}}} \sim t_{T-K}.$$

Estimação intervalar (cont.)

EXPLICAÇÃO: Temos

$$\frac{\frac{b_j-\beta_j}{\sqrt{\sigma^2 c_{jj}}}}{\sqrt{\frac{(T-K)\hat{\sigma}^2/\sigma^2}{T-K}}} = \frac{b_j - \beta_j}{\sqrt{\hat{\sigma}^2 c_{jj}}}.$$

Note que: (i) O numerador (lado esquerdo da igualdade) possui distribuição $\mathcal{N}(0,1)$, (ii) O denominador (lado esquerdo da igualdade) é a raiz quadrada de uma variável aleatória χ^2_{T-K} dividida por $T-K$ (número de graus de liberdade), (iii) O numerador e o denominador são independentes.

OBS. A <u>independência</u> entre o numerador e o denominador segue da independência entre b e $\hat{\sigma}^2$.

Estimação intervalar (cont.)

Fazendo desenvolvimento análogo ao anterior, chegamos ao seguinte intervalo de confiança de nível $(1-\alpha)\times 100\%$ para β_j:

$$\boxed{\left[b_j - t_{T-K;1-\alpha/2}\sqrt{\widehat{\text{var}}(b_j)}, b_j + t_{T-K;1-\alpha/2}\sqrt{\widehat{\text{var}}(b_j)}\,\right]}$$

em que $t_{m;\delta}$ denota o quantil δ da distribuição t_m.

Ou seja,

$$\boxed{\left[b_j - t_{T-K;1-\alpha/2}\sqrt{\hat{\sigma}^2 c_{jj}}, b_j + t_{T-K;1-\alpha/2}\sqrt{\hat{\sigma}^2 c_{jj}}\,\right]}$$

PERGUNTA PARA REFLEXÃO: Qual dos dois intervalos (caso irrealista e caso realista) tende a ser mais amplo e por quê?

Estimação intervalar (cont.)

EXERCÍCIO: Seguindo os desenvolvimentos anteriores, mas fazendo as devidas modificações, construa um intervalo de confiança de nível $(1-\alpha)\times 100\%$ para σ^2.

Estimação intervalar (cont.)

No R, use a função `confint`. Por exemplo,

```
> data(cars)
> help(cars)
> ajuste = lm(dist ~ speed, data = cars)
> intervalos de confiança de 95%
> confint(ajuste)
                     2.5 %          97.5 %
(Intercept) -31.16784960239 -3.99034017863
speed         3.09696432814  4.76785319011
> # intervalos de confiança de 90%
> confint(ajuste, level = 0.90)
                       5 %            95 %
(Intercept) -28.91451427065 -6.24367551037
speed         3.23550067632  4.62931684193
> # estimativas pontuais (b_1 e b_2)
> coef(ajuste)
    (Intercept)           speed
-17.57909489051   3.93240875912
```

Testes de hipóteses

OBJETO DE DESEJO: Realizar inferência por testes de hipóteses sobre os parâmetros do modelo de regressão linear.

SUPOSIÇÃO IMPORTANTE: Assumiremos normalidade, i.e., nossa suposição [S5].

TESTE 1 Suponha que desejamos testar $\mathcal{H}_0 : \beta_j = \beta_j^{(0)}$ vs.

$\mathcal{H}_1 : \beta_j \neq \beta_j^{(0)}$, para algum $j \in \{1, \ldots, K\}$.

Seja $\operatorname{ep}(b_j)$ o erro-padrão de b_j, i.e., $\operatorname{ep}(b_j) = \sqrt{\widehat{\operatorname{var}}(b_j)}$.

Lembre que $(b_j - \beta_j)/\operatorname{ep}(b_j) \sim t_{T-K}$. Note que

$$t = \frac{b_j - \beta_j^{(0)}}{\operatorname{ep}(b_j)}$$

é uma <u>estatística de teste</u> que podemos usar. Ela mede a evidência amostral contra $\mathcal{H}_0$.

Note que, sob $\mathcal{H}_0$, $\beta_j = \beta_j^{(0)}$ e, assim, sob a hipótese nula

$$t \sim t_{T-K}.$$

Testes de hipóteses (cont.)

Desta forma, rejeitamos $\mathcal{H}_0$ ao nível de significância $\alpha \in (0, 1)$ se

$$|t| > t_{T-K;1-\alpha/2}.$$

Esse é teste é chamado de **teste t**.

O teste apresentado é um teste bilateral ou bicaudal. Podemos realizar testes unilaterais ou unicaudais.

CASO 1. Suponha que desejamos testar $\mathcal{H}_0 : \beta_j = \beta_j^{(0)}$ vs. $\mathcal{H}_1 : \beta_j > \beta_j^{(0)}$.

Rigorosamente, testamos $\mathcal{H}_0 : \beta_j \leq \beta_j^{(0)}$ vs. $\mathcal{H}_1 : \beta_j > \beta_j^{(0)}$. Ao escrever $\mathcal{H}_0 : \beta_j = \beta_j^{(0)}$, desconsideramos a possibilidade que β_j seja menor que $\beta_j^{(0)}$.

Usamos a mesma estatística de teste e rejeitamos a hipótese nula se

$$t > t_{T-K;1-\alpha}.$$

Testes de hipóteses (cont.)

EXEMPLO:

$$\text{consumo}_t = \beta_1 + \beta_2\, \text{renda}_t + \beta_3\, \text{riqueza}_t + e_t,$$

$\mathcal{H}_0 : \beta_3 = 0$ vs $\mathcal{H}_1 : \beta_3 > 0$ (i.e., riqueza não tem impacto ou tem impacto positivo).

CASO 2. Suponha que desejamos testar $\mathcal{H}_0 : \beta_j = \beta_j^{(0)}$ vs. $\mathcal{H}_1 : \beta_j < \beta_j^{(0)}$.

Rigorosamente, testamos $\mathcal{H}_0 : \beta_j \geq \beta_j^{(0)}$ vs. $\mathcal{H}_1 : \beta_j < \beta_j^{(0)}$. Ao escrever $\mathcal{H}_0 : \beta_j = \beta_j^{(0)}$, desconsideramos a possibilidade que β_j seja maior que $\beta_j^{(0)}$.

Usamos a mesma estatística de teste e rejeitamos a hipótese nula se

$$t < -t_{T-K;1-\alpha}.$$

Testes de hipóteses (cont.)

EXEMPLO:

$$\text{inflação}_t = \beta_1 + \beta_2 \text{ taxa de juros}_t + e_t,$$

$\mathcal{H}_0 : \beta_2 = 0$ vs $\mathcal{H}_1 : \beta_2 < 0$ (i.e., taxa de juros não tem impacto ou tem impacto negativo).

TERMINOLOGIA. Quando $\beta_j^{(0)} = 0$, a estatística t é chamada de razão t.

Note que podemos lidar com casos mais complexos. Por exemplo, suponha que desejamos testar $\mathcal{H}_0 : \beta_2 = \beta_3$ vs. $\mathcal{H}_1 : \beta_2 \neq \beta_3$. Equivalentemente, $\mathcal{H}_0 : \beta_2 - \beta_3 = 0$ vs. $\mathcal{H}_1 : \beta_2 - \beta_3 \neq 0$.

EXEMPLO:

$$\text{consumo}_t = \beta_1 + \beta_2 \text{ renda do trabalho}_t + \beta_3 \text{ renda de outras fontes}_t + e_t.$$

Testes hipóteses (cont.)

Nossa estatística de teste:

$$t = \frac{b_2 - b_3}{\sqrt{\widehat{\text{var}}(b_2 - b_3)}} = \frac{b_2 - b_3}{\sqrt{\widehat{\text{var}}(b_2) + \widehat{\text{var}}(b_3) - 2\widehat{\text{cov}}(b_2, b_3)}}.$$

Note que $\widehat{\text{var}}(b_2)$, $\widehat{\text{var}}(b_3)$ e $\widehat{\text{cov}}(b_2, b_3)$ são os elementos $(2, 2)$, $(3, 3)$ e $(2, 3)$ de $\hat{\sigma}^2(X'X)^{-1}$.

Rejeitamos $\mathcal{H}_0$ ao nível de significância α se

$$|t| > t_{T-K;1-\alpha/2}.$$

Testes de hipóteses (cont.)

PERGUNTA: Possuímos alguma informação, relevante ou não, sobre a distribuição (da estatística de teste) sob a hipótese alternativa?

Suponha que $\mathcal{H}_0 : \beta_j = \beta_j^{(0)}$ é falsa. Seja $\beta_j^{(1)}$ o valor verdadeiro do parâmetro.

Lembre que se $Z_1 \sim \mathcal{N}(0, 1)$ e $Z_2 \sim \chi^2_\nu$ são variáveis aleatórias independentes, então

$$\frac{Z_1}{\sqrt{Z_2/\nu}} \sim t_\nu.$$

FATO: Se $Z_1 \sim \mathcal{N}(\varphi, 1)$, então

$$\frac{Z_1}{\sqrt{Z_2/\nu}} \sim t_{\nu,\varphi}$$

(t de Student não central, i.e., t de Student com ν graus de liberdade e parâmetro de não centralidade φ; para detalhes, ver, e.g., `https://en.wikipedia.org/wiki/Noncentral_t-distribution`).

Testes de hipóteses (cont.)

Se $W \sim t_{\nu,\varphi}$, então

$$\mathbb{E}(W) = \varphi\sqrt{\frac{\nu}{2}}\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \quad \text{para} \quad \nu > 1.$$

Também,

$$\mathrm{var}(W) = \frac{\nu(1+\varphi^2)}{\nu-2} - \varphi^2\left(\sqrt{\frac{\nu}{2}}\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}\right)^2 \quad \text{para} \quad \nu > 2.$$

FATO: A seguinte aproximação é válida:

$$\sqrt{\frac{\nu}{2}}\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \approx \frac{1}{1-\frac{3}{4\nu-1}}.$$

FATO: Se $\varphi \neq 0$, a distribuição é assimétrica. A assimetria da distribuição $t_{\nu,\varphi}$ diminui à medida que ν aumenta.

Testes de hipóteses (cont.)

Temos que $b_j \sim \mathcal{N}(\beta_j^{(1)}, \sigma^2 c_{jj})$. Assim,

$$b_j - \beta_j^{(0)} \sim \mathcal{N}(\beta_j^{(1)} - \beta_j^{(0)}, \sigma^2 c_{jj}).$$

Ou seja,

$$\frac{b_j - \beta_j^{(0)}}{\sqrt{\sigma^2 c_{jj}}} \sim \mathcal{N}\left(\frac{\beta_j^{(1)} - \beta_j^{(0)}}{\sqrt{\sigma^2 c_{jj}}}, 1\right).$$

Segue que, sob $\mathcal{H}_1$,

$$t \sim t_{T-K,\varphi},$$

em que

$$\varphi = \frac{\beta_j^{(1)} - \beta_j^{(0)}}{\sqrt{\sigma^2 c_{jj}}}.$$

Francisco Cribari

Testes de hipóteses (cont.)

Ou seja, a estatística t pode ser expressa como

$$\frac{\mathcal{N}(\varphi, 1)}{\sqrt{\chi^2_{T-K}/(T-K)}},$$

as variáveis aleatórias no numerador e denominador sendo independentes.

Quando $\mathcal{H}_0$ é verdadeira, $\varphi = 0$ e $t \sim t_{T-K}$ (central). Quando $\mathcal{H}_0$ é falsa, $\varphi \neq 0$ e $t \sim t_{T-K;\varphi}$ (não central).

Note que quanto maior $|\varphi|$, mais distintas serão as distribuições nula e não nula de t.

Assim, quanto maior $|\varphi|$, maior deverá ser o poder do teste.

Note que $|\varphi|$ aumenta quando: (i) $|\beta_j^{(1)} - \beta_j^{(0)}|$ aumenta, ou (ii) T aumenta (o que causa redução no erro-padrão de b_j).

OBS. A consistência do teste segue de (ii).

Testes de hipóteses (cont.)

PRÓXIMO OBJETIVO: Testar várias restrições simultaneamente.

TESTE 2 Suponha que desejamos testar a hipótese nula de que não há relação de regressão (linear) entre a resposta e os regressores. Ou seja, queremos testar $\mathcal{H}_0 : \beta_2 = \beta_3 = \cdots = \beta_K = 0$ vs. $\mathcal{H}_1$: não $\mathcal{H}_0$ (i.e., pelo menos uma das inclinações é diferente de zero).

Note que, sob a hipótese nula, $y_t = \beta_1 + e_t$.

Sejam $\beta_s = (\beta_2, \ldots, \beta_K)'$, $b_s = (b_2, \ldots, b_K)'$ e $\widehat{\text{cov}}(b_s)$ a matriz de covariância estimada de b_s.

Note que queremos testar $\mathcal{H}_0 : \beta_s = 0$ vs. $\mathcal{H}_1 : \beta_s \neq 0$.

Temos que

$$\frac{(b_s - \beta_s)'[\widehat{\text{cov}}(b_s)]^{-1}(b_s - \beta_s)}{K-1} \sim F_{K-1,T-K}.$$

Testes de hipóteses (cont.)

Note que, sob a hipótese nula, $\beta_s = 0$. Desta forma, podemos usar a seguinte estatística de teste:

$$\boxed{F = \frac{b_s'[\widehat{\text{cov}}(b_s)]^{-1} b_s}{K-1}}$$

Sob $\mathcal{H}_0$, $F \sim F_{K-1,T-K}$. Assim, rejeitamos $\mathcal{H}_0$ ao nível de significância α se

$$F > F_{K-1,T-K;1-\alpha}.$$

Esse teste é conhecido como **teste F**.

Estimação pontual no R I

Revisitemos uma regressão que já estimamos.

```
> # Carregar o conjunto de dados
> data(mtcars)
> # Detalhes sobre os dados
> help(mtcars)
> # Ajustar o modelo de regressão linear
> modelo = lm(mpg ~ wt + hp, data = mtcars)
> summary(modelo)

Call:
lm(formula = mpg ~ wt + hp, data = mtcars)

Residuals:
         Min           1Q       Median           3Q          Max
-3.940979469 -1.600221937 -0.182013641  1.049855518  5.853790850

Coefficients:
                Estimate     Std. Error  t value     Pr(>|t|)
(Intercept) 37.22727011645  1.59878753800 23.28469   < 2.22e-16
wt          -3.87783074240  0.63273349438 -6.12870 0.0000011196
```

Estimação pontual no R II

```
hp           -0.03177294698  0.00902970968 -3.51871    0.0014512

Residual standard error: 2.59341178 on 29 degrees of freedom
Multiple R-squared:  0.826785452,        Adjusted R-squared:
0.814839621
F-statistic: 69.2112134 on 2 and 29 DF,  p-value: 9.10905439e-12
```

O valor da estatística F para o teste de $\mathcal{H}_0 : \beta_2 = \beta_3 = 0$ é 69.2112 e o p-valor do teste é muito pequeno.

Testes de hipóteses (cont.)

Há uma relação entre os testes F e t no modelo de regressão simples ($K = 2$). Testamos aqui $\mathcal{H}_0 : \beta_2 = 0$ vs. $\mathcal{H}_1 : \beta_2 \neq 0$.

Note que, nesse caso,

$$F = \frac{b_2^2}{\widehat{\text{var}}(b_2)} = t^2.$$

Lembre que se $Z \sim t_{T-K}$, então $Z^2 \sim F_{1,T-K}$.

Nesse caso, o teste t possui uma vantagem. PERGUNTA: Qual? RESPOSTA: É possível fazer testes unilaterais com o teste t.

REFLEXÃO: Considere o modelo $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + e_t$. Testar $\mathcal{H}_0 : \beta_2 = 0$ e $\mathcal{H}_0 : \beta_3 = 0$ usando dois testes t é equivalente a testar $\mathcal{H}_0 : \beta_2 = \beta_3 = 0$ usando um teste F? Por quê?

RESPOSTA: Não. Os testes t e F usam informações distintas. A estatística F inclui $\widehat{\text{cov}}(b_2, b_3)$.

Pausa para revisão

Suponha que Z é um vetor aleatório de dimensão $J \times 1$ e que

$$Z \sim \mathcal{N}(\mu, \Sigma),$$

em que μ é um vetor de médias de dimensão $J \times 1$ e Σ é uma matriz de covariância positiva definida de dimensão $J \times J$. Então,

$$(Z - \mu)'\Sigma^{-1}(Z - \mu)$$

tem distribuição χ^2_J.

Testes de hipóteses (cont.)

TESTE 3 Ambicionamos generalidade e agora desejamos testar

$$\mathcal{H}_0 : R\beta = r \text{ vs. } \mathcal{H}_1 : R\beta \neq r$$

R: matriz $J \times K$ de posto J que impõe J restrições sobre β

r: vetor $J \times 1$ dado

OBJETO DE DESEJO: Uma estatística de teste.

Comecemos notando que

$$b \sim \mathcal{N}(\beta, \sigma^2(X'X)^{-1}) \;\Rightarrow\; Rb - r \sim \mathcal{N}(R\beta - r, R\sigma^2(X'X)^{-1}R').$$

Note que, sob a hipótese nula, a média da distribuição é zero ($R\beta - r = 0$).

Testes de hipóteses (cont.)

Sabemos que

$$\frac{T-K}{\sigma^2}\hat{\sigma}^2 \sim \chi^2_{T-K}.$$

Sabemos também que, sob $\mathcal{H}_0$,

$$(Rb-r)'[\sigma^2 R(X'X)^{-1}R']^{-1}(Rb-r) \sim \chi^2_J.$$

Por fim, sabemos que b e $\hat{\sigma}^2$ são independentes.

Dividindo a segunda variável aleatória χ^2 padronizada pela primeira variável aleatória χ^2 padronizada chegamos à estatística de teste

$$\boxed{F = \frac{(Rb-r)'[R(X'X)^{-1}R']^{-1}(Rb-r)}{J\hat{\sigma}^2}}$$

Sob $\mathcal{H}_0$, $F \sim F_{J,T-K}$. Assim, rejeitamos $\mathcal{H}_0$ ao nível de significância α se

$$F > F_{J,T-K;1-\alpha}.$$

Testes de hipóteses (cont.)

EXEMPLO: Considere o seguinte modelo

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + e_t,$$

em que

- y_t: gasto familiar com alimentação;
- x_{t2}: renda familiar do trabalho;
- x_{t3}: renda familiar de outras fontes;
- x_{t4}: riqueza.

Suponha que desejamos testar $\mathcal{H}_0 : \beta_2 = \beta_3$ e $\beta_4 = 0$. Ou seja,

$$\mathcal{H}_0 : \beta_2 - \beta_3 = 0 \text{ e } \beta_4 = 0$$

e $J = 2$ (duas restrições).

Testes de hipóteses (cont.)

Ou seja, desejamos testar a igualdade entre as duas propensões marginais a consumir e, simultaneamente, que a riqueza não impacta o gasto familiar com alimentação médio.

Aqui,

$$r = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

A matriz R, de dimensão 2×4, é

$$R = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Podemos testar nossa hipótese nula usando o teste F.

Estimação intervalar

Voltemos a estimação intervalar. Nosso interesse agora residirá na construção de **regiões de confiança** para dois ou mais coeficientes de regressão.

CASO 1 Nosso interesse reside em construir uma **região de confiança** de nível $(1-\alpha)\times 100\%$ para β. Sob normalidade (suposição [S5]),

$$b \sim \mathcal{N}(\beta, \sigma^2(X'X)^{-1}).$$

Assim,

$$(b-\beta)'\sigma^{-2}(X'X)(b-\beta) \sim \chi^2_K.$$

Lembre que, sob [S5], $[(T-K)/\sigma^2]\hat{\sigma}^2 \sim \chi^2_{T-K}$. Assim,

$$\begin{aligned} F_\beta &= \frac{(b-\beta)'\sigma^{-2}(X'X)(b-\beta)/K}{[(T-K)/\sigma^2]\hat{\sigma}^2/(T-K)} \\ &= \frac{(b-\beta)'(X'X)(b-\beta)}{\hat{\sigma}^2 K} \sim F_{K,T-K}. \end{aligned}$$

Estimação intervalar (cont.)

A **região de confiança** de nível $(1-\alpha)\times 100\%$ para β é dada pelo conjunto de valores de β tais que

$$\Pr(F_\beta \leq F_{K,T-K;1-\alpha}) = 1-\alpha.$$

CASO 2 Particionemos β como $\beta = (\beta_1', \beta_2')'$, em que β_1 é um vetor $q \times 1$ e β_2 é um vetor $(K-q)\times 1$. Suponha que desejamos obter uma **região de confiança** de nível $(1-\alpha)\times 100\%$ para β_1. Nosso modelo é

$$y = X_1\beta_1 + X_2\beta_2 + e,$$

em que $X = [X_1\ X_2]$, X_1 sendo de dimensão $T\times q$ e X_2 tendo dimensão $T \times (K-q)$.

Estimação intervalar (cont.)

Sejam $b = (b_1', b_2')'$ e

$$F_{\beta_1} = \frac{(b_1 - \beta_1)'(X_1'X_1)(b_1 - \beta_1)}{\hat{\sigma}^2 q} \sim F_{q,T-K}.$$

A **região de confiança** de nível $(1 - \alpha) \times 100\%$ para β_1 é dada pelo conjunto de valores de β_1 tais que

$$\Pr(F_{\beta_1} \leq F_{q,T-K;1-\alpha}) = 1 - \alpha.$$

Estimação intervalar (cont.) I

EXEMPLO: Considere a regressão $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + e_t$ e suponha que desejamos obter uma região de confiança de nível 95% para (β_2, β_3). Usemos o conjunto de dados `mtcars`.

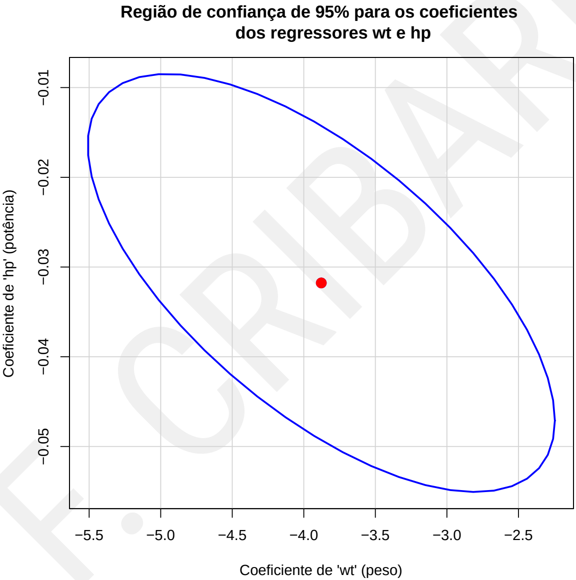
Nosso código:

```
# 1. Carregar o pacote car
library(car)

# 2. Carregar o conjunto de dados
data(mtcars)

# 3. Ajustar o modelo de regressão linear
modelo = lm(mpg ~ wt + hp, data = mtcars)

# 4. Gerar o gráfico da região de confiança de 95%
# Use confidenceEllipse() para coeficientes de regressão
confidenceEllipse(modelo, which.coef = c(2, 3), levels = 0.95,
     xlab = "Coeficiente de 'wt' (peso)",
```

Estimação intervalar (cont.) II

```
      ylab = "Coeficiente de 'hp' (potência)",
      main = "Região de confiança de 95% para os coeficientes
      dos regressores wt e hp")

# Adicionar ponto correspondente aos coeficientes estimados
points(coef(modelo)[2], coef(modelo)[3], pch = 19,
col = "red", cex = 1.5)
```

Estimação intervalar (cont.)



Uma pergunta relevante sobre testes de hipóteses

PERGUNTA: Como realizar estimação intervalar e testes de hipóteses sem assumir normalidade?

Note que sem normalidade as estatísticas t e F não possuem, sob a hipótese nula distribuições t de Student e F de Snedecor, respectivamente. De onde, então, obteríamos valores críticos para esses testes?

Teste z

Suponha que desejamos testar $\mathcal{H}_0 : \beta_j = \beta_j^{(0)}$ vs $\mathcal{H}_1 : \beta_j \neq \beta_j^{(0)}$ sem assumir normalidade.

A estatística de teste é

$$t = \frac{b_j - \beta_j^{(0)}}{\text{ep}(b_j)}.$$

Sem normalidade, a distribuição nula de t não é t_{T-K}. Assim, não podemos usar valores críticos obtidos da distribuição t de Student.

RESULTADO: Sob $\mathcal{H}_0$, $t \xrightarrow{d} \mathcal{N}(0,1)$.

Assim, um teste (aproximado) é: Rejeite $\mathcal{H}_0$ ao nível de significância α se $|t| > z_{1-\alpha/2}$. Esse teste é conhecido como teste z.

Não temos mais $\Pr(\text{rejeitar } \mathcal{H}_0 \,|\, \mathcal{H}_0 \text{ é verdadeira}) = \alpha$. Sabemos apenas que $\Pr(\text{rejeitar } \mathcal{H}_0 \,|\, \mathcal{H}_0 \text{ é verdadeira}) \to \alpha$ quando $T \to \infty$.

Revisão

RESULTADO: Sejam $\{Z_T; T \in \mathbb{N}\}$ e $\{V_T; T \in \mathbb{N}\}$ sequências de variáveis aleatórias, Z variável aleatória e $c \in \mathbb{R}$ (constante real). Se $V_T \xrightarrow{p} c$ e $Z_T \xrightarrow{d} Z$, então $V_T \times Z_T \xrightarrow{d} cZ$.

CASO PARTICULAR ($Z_T = Z \; \forall T$): $V_T \times Z \xrightarrow{d} cZ$.

CASO AINDA MAIS PARTICULAR ($Z_T = Z \; \forall T$ e $c = 1$): $V_T \times Z \xrightarrow{d} Z$.

EXEMPLO: Seja $V_T = 1$ com probabilidade $1 - 1/T$ e $V_T = 0$ com probabilidade $1/T$ e seja $Z \sim \mathcal{N}(0, 1)$. Dado que $V_T \xrightarrow{p} 1$, segue que $V_T \times Z \xrightarrow{d} \mathcal{N}(0, 1)$.

Teste Wald

Suponha que desejamos testar $\mathcal{H}_0 : R\beta = r$ vs $\mathcal{H}_1 : R\beta \neq r$ (sem [S5]).

Sob $\mathcal{H}_0$ e com normalidade,

$$(Rb - r)'[\sigma^2 R(X'X)^{-1}R']^{-1}(Rb - r) \sim \chi^2_J.$$

Sob $\mathcal{H}_0$ e sem normalidade,

$$\frac{\sigma^2}{\hat{\sigma}^2} \times (Rb - r)'[\sigma^2 R(X'X)^{-1}R']^{-1}(Rb - r) \xrightarrow{d} \chi^2_J.$$

Dado que $\hat{\sigma}^2$ é consistente para σ^2, temos que $\sigma^2/\hat{\sigma}^2 \xrightarrow{p} 1$. Segue que, sob $\mathcal{H}_0$,

$$W = (Rb - r)'[\hat{\sigma}^2 R(X'X)^{-1}R']^{-1}(Rb - r) \xrightarrow{d} \chi^2_J.$$

Teste Wald (cont.)

Use valores críticos da distribuição nula assintótica (aproximação). Rejeite $\mathcal{H}_0$ se $W > \chi^2_{J;1-\alpha}$.

Não temos mais $\Pr(\text{rejeitar } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ é verdadeira}) = \alpha$. Sabemos apenas que $\Pr(\text{rejeitar } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ é verdadeira}) \to \alpha$ quando $T \to \infty$.

Estimação intervalar sem normalidade

Suponha que desejamos obter um intervalo de confiança de nível $(1-\alpha)\times 100\%$ para β_j, $j = 1, \ldots, K$, para algum $\alpha \in (0,1)$.

Com normalidade, nosso intervalo é

$$\text{IC}_{1-\alpha}(\beta_j) = b_j \pm t_{T-K;1-\alpha/2}\, \text{ep}(b_j)$$

e $\Pr(\text{IC}_{1-\alpha}(\beta_j) \text{ contém } \beta_j) = 1-\alpha$.

Sem normalidade, nosso intervalo é

$$\text{IC}_{1-\alpha}(\beta_j) = b_j \pm z_{1-\alpha/2}\, \text{ep}(b_j)$$

e $\Pr(\text{IC}_{1-\alpha}(\beta_j) \text{ contém } \beta_j) \to 1-\alpha$.

Consequências de especificação incorreta

Considere os seguintes modelos:

$$\begin{aligned}
&[\text{M1}]\;\; y = X_1\beta_1 + X_2\beta_2 + e_1 = X\beta + e_1,\\
&[\text{M2}]\;\; y = X_1\beta_1 + e_2.
\end{aligned}$$

Ou seja, o modelo M1 usa os regressores utilizados em M2 e mais algum(ns) regressor(es) adicional(is). O modelo M2 pode ser obtido do modelo M1 impondo $\beta_2 = 0$.

CASO 1 O modelo M1 é o correto, mas estimamos M2.

O efeito da(s) variável(is) omitida(s) é absorvido pelo erro, que deixa de ter média zero:

$$\begin{aligned}
b_1 &= (X_1'X_1)^{-1}X_1'y\\
&= (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + e_1)\\
&= \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'e_1.
\end{aligned}$$

Assim,

$$\mathbb{E}(b_1) = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 \neq \beta_1.$$

Consequências de especificação incorreta (cont.)

Ou seja, o EMQO se torna **viesado**.

MENSAGEM: A omissão de regressores importantes resulta em estimadores viesados.

CASO 2 O modelo M2 é o verdadeiro, mas estimamos M1.

Ou seja, incluímos na regressão variáveis explicativas que não são relevantes.

Aqui, não violamos a suposição de que os erros têm média zero ([S1]) e, assim, as consequências não são tão severas.

O EMQO $b = (X'X)^{-1}X'y$ permanece não viesado:

$$\mathbb{E}(b) = \beta = \begin{bmatrix} \beta_1 \\ 0 \end{bmatrix}.$$

A consequência indesejável é que o EMQO se torna **ineficiente**.

Exemplo empírico I

```
cars                  package:datasets                  R Documentation

Speed and Stopping Distances of Cars

Description:

     The data give the speed of cars and the distances taken to stop.
     Note that the data were recorded in the 1920s.

Usage:

     cars

Format:

     A data frame with 50 observations on 2 variables.

       [,1]  speed  numeric  Speed (mph)
       [,2]  dist   numeric  Stopping distance (ft)
```

Exemplo empírico I

```
> # dados
> data(cars); attach(cars)
> # informação sobre os dados
> help(cars)
> # número de observações
> nobs = length(dist)
> # sumário dos dados
> summary(cars)
     speed           dist
 Min.   : 4.0   Min.   :  2.00
 1st Qu.:12.0   1st Qu.: 26.00
 Median :15.0   Median : 36.00
 Mean   :15.4   Mean   : 42.98
 3rd Qu.:19.0   3rd Qu.: 56.00
 Max.   :25.0   Max.   :120.00
> # desvios-padrão
> round(sd(speed), 4); round(sd(dist), 4)
[1] 5.2876
[1] 25.7694
> # ajuste de regressão
> ajuste = lm(dist~speed)
> # sumário do ajuste
> summary(ajuste)

Call:
lm(formula = dist ~ speed)

Residuals:
         Min           1Q       Median           3Q          Max
-29.06908029  -9.52532117  -2.27185401   9.21471533  43.20128467

Coefficients:
```

Exemplo empírico II

```
                 Estimate     Std. Error  t value   Pr(>|t|)
(Intercept) -17.579094891   6.758440169 -2.60106   0.012319
speed         3.932408759   0.415512777  9.46399 1.4898e-12

Residual standard error: 15.3795867 on 48 degrees of freedom
Multiple R-squared:  0.651079381, Adjusted R-squared:  0.643810201
F-statistic: 89.5671065 on 1 and 48 DF,  p-value: 1.4898365e-12

> # calculemos as principais quantidades do ajuste de regressão
> # usando as fórmulas apresentadas em aula
> # matriz X
> X = cbind(1, speed)
> # vetor b
> b = solve(t(X) %*% X) %*% t(X) %*% dist
> # resíduos
> echapeu = dist - X %*% b
> # estimativa de sigma^2
> sigma2chapeu = crossprod(echapeu)/(nobs-ncol(X))
> # desvio padrão
> sqrt(sigma2chapeu)
              [,1]
[1,] 15.3795867488
> sqrt(sigma2chapeu) |> as.double()
[1] 15.3795867488
> # erros-padrão de b_1 e b_2
> ep = sqrt(as.double(sigma2chapeu)*diag(solve(t(X) %*% X)))
> ep
                        speed
6.758440169379 0.415512776657
> # estimativas de MQO
> b
                 [,1]
```

Exemplo empírico III

```
      -17.57909489051
speed   3.93240875912
> # vejamos as estimativas obtidas via função lm
> coef(ajuste)
    (Intercept)           speed
-17.57909489051   3.93240875912
> # razões t
> t = b / ep
> t
                [,1]
      -2.60105800302
speed  9.46398999030
> # p-valor do teste de H_0: \beta_1 = 0 vs H_1: \beta_1 \not= 0
> pv1 = 2*(1-pt(abs(t[1]), df=nobs-ncol(X)))
> pv1
[1] 0.0123188161538
> # p-valor do teste de H_0: \beta_2 = 0 vs H_1: \beta_2 \not= 0
> pv2 = 2*(1-pt(abs(t[2]), df=nobs-ncol(X)))
> pv2
[1] 1.48991929905e-12
```

Nosso próximo tópico

Na saída do R do ajuste da nossa regressão, falta entendermos essas informações:

```
Multiple R-squared:  0.651079381, Adjusted R-squared:  0.643810201
```

Esse é nosso próximo tópico.