

ECONOMETRIA CLÁSSICA E BAYESIANA

Ralph S. Silva

Departamento de Métodos Estatísticos
Instituto de Matemática
Universidade Federal do Rio de Janeiro

Sumário

Tópicos

Referências

Periódicos

Introdução

Modelo de Regressão Linear Múltipla

O Estimador de Mínimos Quadrados

Possíveis Tópicos

- ▶ Introdução: Econometria.
- ▶ Modelo de regressão linear.
- ▶ Estimador de mínimos quadrados.
- ▶ Teste de hipóteses e seleção de modelos.
- ▶ Forma funcional e mudança de estrutura.
- ▶ Modelos de regressão não-linear, semi-paramétrico e não-paramétrico.
- ▶ Endogeneidade e estimação por variável instrumental.
- ▶ Modelo de regressão generalizado e heterocedasticidade.
- ▶ Sistema de equações.
- ▶ Modelos para dados de painel.
- ▶ Métodos de estimação em econometria.
- ▶ Estimação da distância mínima e método generalizado dos momentos.
- ▶ Estimação por máxima verossimilhança.
- ▶ Estimação e inferência utilizando métodos de simulação.
- ▶ Estimação e inferência bayesiana.

Outros Tópicos

- ▶ Escolha discreta.
- ▶ Contagem de eventos.
- ▶ Variáveis de dependência limitada: truncamento, censura e seleção de amostra.
- ▶ Correlação serial.
- ▶ Dados não estacionários.

Referências: Livros

- ▶ Bauwens, L., Lubrano, M. e Richard, J-F. (2000). *Bayesian Inference in Dynamic Econometric Models*.
- ▶ Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*.
- ▶ Greene, W. H. (2011). *Econometric Analysis*, 7th ed.
- ▶ Gujarati, D. (2002). *Basic Econometrics*, 4th ed.
- ▶ Hayashi, F. (2000). *Econometrics*.
- ▶ Hendry, D. F. e Nielsen, B. (2007). *Econometric Modelling: A Likelihood Approach*.
- ▶ Koop, G. (2003). *Bayesian Econometrics*.
- ▶ Lancaster, T. (2004). *Introduction to Modern Bayesian Econometrics*.
- ▶ Stock, J. H. e Watson, M. W. (2006). *Introduction to Econometrics*.
- ▶ Wooldridge, J. M. (2006). *Introductory to Econometrics: A Modern Approach*.
- ▶ Petersen, K. B. e Pedersen, M. S. (2012). *The Matrix Cookbook*.
http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/3274/pdf/imm3274.pdf

Periódicos

- ▶ *Econometric Reviews*
- ▶ *Econometric Theory*
- ▶ *Econometrica*
- ▶ *Empirical Economics*
- ▶ *Foundations and Trends in Econometrics*
- ▶ *Journal of Business and Economic Statistics*
- ▶ *The Econometrics Journal*
- ▶ *The Journal of Applied Econometrics*
- ▶ *The Journal of Econometrics*
- ▶ *The Review of Economics and Statistics*

Introdução

Econometria é a área que foca em aplicações de estatística matemática e as ferramentas da inferência estatística na mensuração empírica de relações postuladas pela teoria econômica (Greene).

Exemplo 1: Função de Consumo de Keynes.

$$C = f(X) \quad \left\{ \begin{array}{l} C : \text{consumo} \\ X : \text{renda} \end{array} \right.$$

$$0 < \frac{dC}{dX} < 1.$$

a propensão marginal a consumir.

A propensão média a consumir (PMC), C/X , decai quando a renda sobe

$$\frac{d(C/X)}{dX} = \left(\frac{dC}{dX} - \frac{C}{X} \right) \frac{1}{X} < 0 \quad \Rightarrow \quad \frac{dC}{dX} < \frac{C}{X}.$$

O modelo de função consumo mais simples é o linear (com restrições)

$$C = \alpha + X\beta, \quad \alpha > 0 \quad \text{e} \quad 0 < \beta < 1.$$

- ▶ Podemos estudar alguns conjuntos de dados.
- ▶ Será verdade que $\alpha > 0$ e $0 < \beta < 1$?
- ▶ A relação é estável através do tempo?
- ▶ Os parâmetros mudam se considerarmos instantes de tempo diferentes? (Uma mudança na propensidade média a economizar.)
- ▶ Existem diferenças sistemáticas nas relações para diferentes países? Como poderíamos explicar estas diferenças?
- ▶ Existem outros fatores que poderiam melhorar a habilidade do modelo em explicar a relação entre consumo e renda?

Exemplo 2: (Detalhes no arquivo R: exemplo_02.r pag. 9)

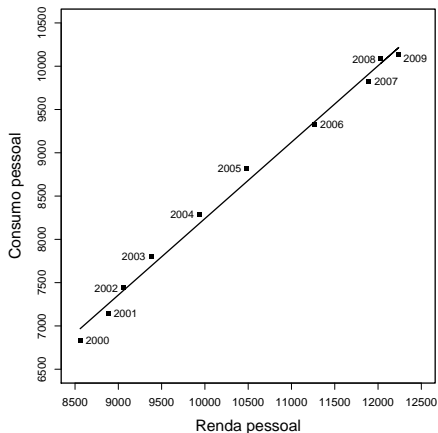


Figura : Consumo e renda agregados dos EUA entre 2000 e 2009.

O ajuste por mínimos quadrados resulta em $\hat{\alpha} = -587,19$ e $\hat{\beta} = 0,8827$.

Para os dados reais, temos uma estimativa negativa para a constante do modelo.

- ▶ A teoria está errada?
- ▶ Isto se deve a variabilidade dos dados?
- ▶ Será que o estimador de mínimos quadrados é adequado nesta situação?
- ▶ Como ficaria a estimação por máxima verossimilhança?
- ▶ Como ficaria a estimação bayesiana? E a distribuição a priori?

Elasticidade

Em economia, o termo elasticidade refere-se a como uma variável econômica muda em função de outra variável econômica.

A elasticidade é medida em termos percentuais ao invés de absoluto.

Isto significa que medimos uma mudança na variável como uma porcentagem da quantidade original da variável.

A mudança em porcentagem da variável x é definida por

$$\text{Mudança percentual em } x = \frac{\text{Mudança em } x}{\text{Valor original de } x} = \frac{\Delta x}{x} \quad (\text{de } x \text{ para } x + \Delta x).$$

A elasticidade de y com respeito a x é dada por

$$\epsilon \triangleq \text{Elasticidade de } y \text{ com respeito a } x = \frac{\text{Mudança percentual em } y}{\text{Mudança percentual em } x} = \frac{\Delta y/y}{\Delta x/x}.$$

Em termos infinitesimais, temos que

$$\epsilon = \text{Elasticidade de } y \text{ com respeito a } x = \frac{\partial y/y}{\partial x/x} = \frac{x}{y} \frac{\partial y}{\partial x}.$$

Modelo de Regressão Linear Múltipla

Notação:

$$\begin{aligned}y &= f(x_1, x_2, \dots, x_p) + \varepsilon \\ &= x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p + \varepsilon,\end{aligned}$$

sendo que

- ▶ y é a variável dependente,
- ▶ $(x_1, x_2, \dots, x_p)'$ é o vetor de variáveis independentes (regressoras, exploratórias ou explicativas),
- ▶ $(\beta_1, \beta_2, \dots, \beta_p)'$ é o vetor de coeficientes; e
- ▶ ε é um distúrbio aleatório.

Para o modelo de regressão acima, a **elasticidade de y com respeito a k -ésima variável x_k** é dada por

$$\epsilon = \frac{x_k}{y} \frac{\partial y}{\partial x_k} = \frac{x_k}{y} \beta_k.$$

Equação da demanda:

$$\text{quantidade} = \beta_1 + \text{preço} \times \beta_2 + \text{renda} \times \beta_3 + \varepsilon.$$

Equação da demanda inversa:

$$\text{preço} = \gamma_1 + \text{quantidade} \times \gamma_2 + \text{renda} \times \beta_3 + \xi.$$

Estas são relações válidas de um mercado.

Ganhos e Educação:

$$\text{ganhos} = \beta_1 + \text{educação} \times \beta_2 + \text{idade} \times \beta_3 + \varepsilon,$$

$$\text{ganhos} = \beta_1 + \text{educação} \times \beta_2 + \text{idade} \times \beta_3 + \text{idade}^2 \times \beta_4 + \varepsilon.$$

Hipóteses do Modelo de Regressão Linear

HP.1: Linearidade: $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p + \varepsilon_i$;

HP.2: Posto completo: não existe nenhuma relação exata entre as variáveis independentes do modelo;

HP.3: Exogeneidade das variáveis independentes:

$$E(\varepsilon_i | x_{i1}, x_{i2}, \dots, x_{ip}) = 0 \quad \Rightarrow \quad \text{Cov}(\mathbf{x}_i, \varepsilon_i) = 0$$

com $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$.

HP.4: Homocedasticidade e autocorrelação nula: cada distúrbio, ε_i , tem a mesma variância, σ^2 , e é não correlacionado com qualquer outro distúrbio ε_j ;

HP.5: Geração dos dados: os dados em $(x_{i1}, x_{i2}, \dots, x_{ip})$ podem ser qualquer mistura de constantes e variáveis aleatórias; e

HP.6: Distribuição normal: os distúrbios são normalmente distribuídos.

Notação

Temos que

$$\begin{aligned}y_i &= x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p + \varepsilon_i \\ &= \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \quad \text{para } i = 1, 2, \dots, n.\end{aligned}$$

- ▶ n é o número de observações;
- ▶ $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$ é o vetor da i -ésima observação das variáveis independentes;
- ▶ $\mathbf{x}_k = (x_{k1}, x_{k2}, \dots, x_{kn})'$ é o vetor da k -ésima variável independente;
- ▶ $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ é o vetor da variável dependente;
- ▶ $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$ é o vetor dos distúrbios;
- ▶ $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$ é o vetor de coeficientes; e
- ▶ $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$ é a matriz de variáveis independentes.

Assim, o modelo de regressão linear múltipla pode ser reescrito por

$$\begin{aligned}\mathbf{y} &= \mathbf{x}_1\beta_1 + \mathbf{x}_2\beta_2 + \cdots + \mathbf{x}_p\beta_p + \boldsymbol{\varepsilon} \quad \text{ou} \\ \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.\end{aligned}$$

Exemplos

- ▶ $y = \gamma x^\beta e^\varepsilon \Rightarrow \ln y = \ln \gamma + \beta \ln x + \varepsilon$ é linear, mas
- ▶ $y = \gamma x^\beta + \varepsilon$ é não linear.

Os seguintes modelos são lineares:

$$y = \alpha + \beta x + \varepsilon$$

$$y = \alpha + \beta \cos x + \varepsilon$$

$$y = \alpha + \beta \frac{1}{x} + \varepsilon$$

$$y = \alpha + \beta \ln x + \varepsilon$$

$$\ln y = \alpha + \beta \ln x + \varepsilon$$

$$z = \ln \gamma + \beta \ln x + \varepsilon = \alpha + \beta w + \varepsilon.$$

O Modelo Log-linear

$$\ln y = \beta_1 + \beta_2 \ln x_2 + \cdots + \beta_p \ln x_p + \varepsilon$$

(Elasticidade constante)

A elasticidade de y com respeito a mudanças em x_k é dado por

$$\epsilon = \frac{\partial y / y}{\partial x / x} = \frac{\partial y}{\partial x} \times \frac{\partial \ln y / \partial y}{\partial \ln x / \partial x} = \frac{\partial \ln y}{\partial \ln x_k} = \beta_k.$$

Exemplo: Mercado de Gasolina dos EUA - 1953-2004

Considere o seguinte modelo para o consumo de gasolina per capita:

$$\ln(G/\text{Pop}) = \beta_1 + \beta_2 \ln(\text{Renda}/\text{Pop}) + \beta_3 \ln \text{Preço} + \beta_4 \ln(\text{PCN}) + \beta_5 \ln(\text{PCU}) + \varepsilon,$$

sendo PCN o preço dos carros novos e PCU o preço dos carros usados.

- ▶ Elasticidade renda e preço para a gasolina; e
- ▶ Elasticidade para demanda com respeito a preço de carros novos e usados.

Em geral, utiliza-se um modelo **semi-log** para modelar taxas de crescimento: $\ln y_t = \mathbf{x}'_t \boldsymbol{\beta} + \gamma t + \varepsilon_t$. A taxa de crescimento por período é dada por $\partial \ln y_t / \partial t = \gamma$.

O modelo **translog** permite modelar uma variedade de estruturas.

- ▶ Primeiro, temos $y = g(x_1, x_2, \dots, x_k)$. Podemos transformar para $\ln y = \ln g(x_1, x_2, \dots, x_k) = h(x_1, x_2, \dots, x_k) = f(\ln x_1, \ln x_2, \dots, \ln x_k)$ pois obviamente $x_k = \exp(\ln x_k)$.
- ▶ Daí, aproximar por uma expansão de Taylor de segunda ordem em torno de $\mathbf{x} = \mathbf{1}$. Definimos $\ln \mathbf{x} = (\ln x_1, \ln x_2, \dots, \ln x_k)'$. Logo,

$$\begin{aligned} \ln y &= f(\mathbf{0}) + \sum_{k=1}^p \frac{\partial f(\ln \mathbf{x})}{\partial \ln x_k} \Big|_{\ln \mathbf{x}=\mathbf{0}} \ln x_k \\ &+ \frac{1}{2} \sum_{k=1}^p \sum_{\ell=1}^p \frac{\partial^2 f(\ln \mathbf{x})}{\partial \ln x_k \partial \ln x_\ell} \Big|_{\ln \mathbf{x}=\mathbf{0}} \ln x_k \ln x_\ell + \varepsilon. \end{aligned}$$

- ▶ Segue-se que

$$\ln y = \beta_0 + \sum_{k=1}^p \beta_k \ln x_k + \frac{1}{2} \sum_{k=1}^p \sum_{\ell=1}^p \gamma_{k\ell} \ln x_k \ln x_\ell + \varepsilon.$$

Distúrbios do Modelo de Regressão

Temos:

- ▶ $E(\varepsilon_i|\mathbf{X}) = 0$ por hipótese, ou seja,

$$E(\varepsilon|\mathbf{X}) = \begin{bmatrix} E(\varepsilon_1|\mathbf{X}) \\ E(\varepsilon_2|\mathbf{X}) \\ \vdots \\ E(\varepsilon_n|\mathbf{X}) \end{bmatrix} = \mathbf{0};$$

- ▶ $E(\varepsilon_i|\varepsilon_1, \dots, \varepsilon_{i-1}, \varepsilon_{i+1}, \dots, \varepsilon_n) = 0$;
- ▶ Por hipótese, todos os distúrbios são simplesmente uma amostra aleatória de alguma população;
- ▶ $E(\varepsilon_i) = E_{\mathbf{X}}(E(\varepsilon_i|\mathbf{X})) = E_{\mathbf{X}}(0) = 0$;
- ▶ Para cada ε_i , $\text{Cov}(\varepsilon_i, \mathbf{X}) = \text{Cov}(E(\varepsilon_i|\mathbf{X}), \mathbf{X}) = \text{Cov}(0, \mathbf{X}) = \mathbf{0}$;

Note que $E(\varepsilon_i) = 0$ **não implica** que $E(\varepsilon_i|\mathbf{x}_i) = 0$. Isto é, $E(\varepsilon_i) = 0$ **não implica** que $\text{Cov}(0, \mathbf{X}) = \mathbf{0}$.

- ▶ $\mathbf{y} = \mathbf{X}\beta + \varepsilon \Rightarrow E(\mathbf{y}|\mathbf{X}) = E(\mathbf{X}\beta|\mathbf{X}) + E(\varepsilon|\mathbf{X}) = \mathbf{X}\beta + \mathbf{0} = \mathbf{X}\beta$.

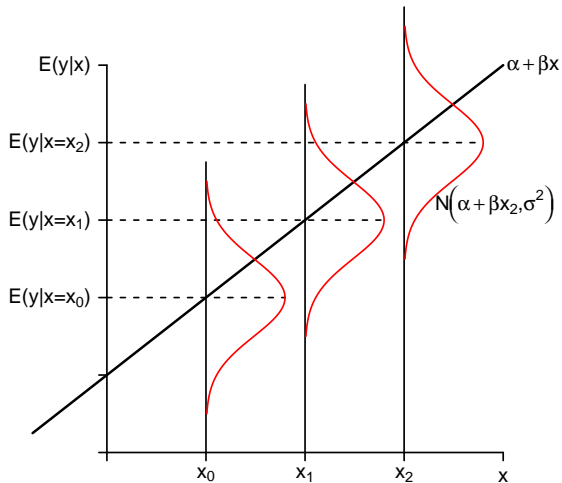
Distúrbios Esféricos

Temos:

- ▶ $\text{Var}(\varepsilon_t|\mathbf{X}) = \sigma^2$, para todo $i = 1, 2, \dots, n$;
- ▶ $\text{Cov}(\varepsilon_i, \varepsilon_j|\mathbf{X}) = 0$, para todo $i \neq j$ (correlação nula), ou seja,

$$\begin{aligned} E(\varepsilon\varepsilon'|\mathbf{X}) &= \begin{bmatrix} E(\varepsilon_1\varepsilon_1|\mathbf{X}) & E(\varepsilon_1\varepsilon_2|\mathbf{X}) & \cdots & E(\varepsilon_1\varepsilon_n|\mathbf{X}) \\ E(\varepsilon_2\varepsilon_1|\mathbf{X}) & E(\varepsilon_2\varepsilon_2|\mathbf{X}) & \cdots & E(\varepsilon_2\varepsilon_n|\mathbf{X}) \\ \vdots & \vdots & \ddots & \vdots \\ E(\varepsilon_n\varepsilon_1|\mathbf{X}) & E(\varepsilon_n\varepsilon_2|\mathbf{X}) & \cdots & E(\varepsilon_n\varepsilon_n|\mathbf{X}) \end{bmatrix} \\ &= \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}; \end{aligned}$$

- ▶ $\text{Var}(\varepsilon) = E_{\mathbf{X}}(\text{Var}(\varepsilon|\mathbf{X})) + \text{Var}_{\mathbf{X}}(E(\varepsilon|\mathbf{X})) = \sigma^2 \mathbf{I}$;
- ▶ Por hipótese, $(\varepsilon|\mathbf{X}) \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$.

Exemplo 3: (Detalhes no arquivo R: exemplo_03.r pag. 21)**Figura :** O Modelo Clássico de Regressão.

Mínimos Quadrados

Temos que

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i.$$

- ▶ $\boldsymbol{\beta}$ e ε_i são quantidades da população;
- ▶ A regressão populacional é $E(y_i | \mathbf{x}_i) = \mathbf{x}_i' \boldsymbol{\beta}$;
- ▶ \mathbf{b} e e_i são estimativas amostrais;
- ▶ A estimativa de $E(y_i | \mathbf{x}_i)$ é $\hat{y}_i = \mathbf{x}_i' \mathbf{b}$;
- ▶ Seja $\varepsilon_i = y_i - \mathbf{x}_i' \boldsymbol{\beta}$ o distúrbio da i -ésima observação;
- ▶ Para qualquer valor de \mathbf{b} , estimamos ε_i com o resíduo

$$e_i = y_i - \mathbf{x}_i' \mathbf{b};$$

- ▶ Das definições, $y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i = \mathbf{x}_i' \mathbf{b} + e_i$;
- ▶ Podemos estimar $\boldsymbol{\beta}$ com a amostra de (y_i, \mathbf{x}_i) , $i = 1, 2, \dots, n$;
- ▶ Na estimação de $\boldsymbol{\beta}$ por \mathbf{b} , queremos que os pontos observados fiquem perto da reta ajustada;
- ▶ Perto neste caso é definido por algum critério de ajuste;
- ▶ **Critério de Mínimos Quadrados.**

Temos

$$S(\mathbf{b}_0) \triangleq \sum_{i=1}^n \mathbf{e}_{i0}^2 = \sum_{i=1}^n (y_i - \mathbf{x}'_i \mathbf{b}_0)^2,$$

com \mathbf{b}_0 denotando a escolha do vetor de coeficientes.

Queremos minimizar $S(\mathbf{b}_0)$, ou seja,

$$\min_{\mathbf{b}_0} S(\mathbf{b}_0) = \min_{\mathbf{b}_0} \mathbf{e}'_0 \mathbf{e}_0 = \min_{\mathbf{b}_0} (\mathbf{y} - \mathbf{X}\mathbf{b}_0)'(\mathbf{y} - \mathbf{X}\mathbf{b}_0).$$

Mas

$$\begin{aligned} S(\mathbf{b}_0) &= \mathbf{y}'\mathbf{y} - \mathbf{b}'_0 \mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b}_0 + \mathbf{b}'_0 \mathbf{X}'\mathbf{X}\mathbf{b}_0 \\ &= \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\mathbf{b}_0 + \mathbf{b}'_0 \mathbf{X}'\mathbf{X}\mathbf{b}_0. \end{aligned}$$

A condição necessária para um mínimo é

$$\frac{\partial S(\mathbf{b}_0)}{\partial \mathbf{b}_0} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b}_0 = \mathbf{0}.$$

Consulte o “The Matrix Cookbook” para mais detalhes sobre as derivadas.

Seja \mathbf{b} a solução. Então, \mathbf{b} satisfaz as equações normais de mínimos quadrados:

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}.$$

Se a inversa de $\mathbf{X}'\mathbf{X}$ existir, que segue pela hipótese de posto completo de \mathbf{X} , então a solução é

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

Para esta solução minimizar a soma de quadrados,

$$\frac{\partial^2 S(\mathbf{b}_0)}{\partial \mathbf{b}_0 \partial \mathbf{b}_0'} = 2\mathbf{X}'\mathbf{X}$$

deve ser uma matriz positiva definida.

Seja $q = \mathbf{c}'\mathbf{X}'\mathbf{X}\mathbf{c}$ para algum vetor arbitrário \mathbf{c} diferente de zero. Então

$$q = \mathbf{v}'\mathbf{v} = \sum_{i=1}^n v_i^2 \quad \text{sendo} \quad \mathbf{v} = \mathbf{X}\mathbf{c}.$$

A menos que cada elemento de \mathbf{v} seja zero, q é positivo.

Exemplo 4: $y_i = \beta_1 + \beta_2 T_i + \beta_3 G_i + \beta_3 R_i + \beta_4 P_i + \varepsilon_i$

Investimento Real (Y)	Constante (1)	Tendência (T)	PIB Real (G)	Taxa de Juros (R)	Taxa de Inflação (P)
0,161	1	1	1,058	5,16	4,40
0,172	1	2	1,088	5,87	5,15
0,158	1	3	1,086	5,95	5,37
0,173	1	4	1,122	4,88	4,99
0,195	1	5	1,186	4,50	4,16
0,217	1	6	1,254	6,44	5,75
0,199	1	7	1,246	7,83	8,82
0,163	1	8	1,232	6,25	9,31
0,195	1	9	1,298	5,50	5,21
0,231	1	10	1,370	5,46	5,83
0,257	1	11	1,439	7,46	7,40
0,259	1	12	1,479	10,28	8,64
0,225	1	13	1,474	11,77	9,31
0,241	1	14	1,503	13,42	9,44
0,204	1	15	1,475	11,02	5,99

Observações anuais de 1968 a 1982.

(Mostrar exemplo no R: exemplo_04.r pag. 25)

Teorema - Regressão Ortogonal: *Se as variáveis em um modelo de regressão são não correlacionadas, isto é, são ortogonais, então os coeficientes angulares da regressão são os mesmos que os coeficientes na regressão individual simples.*

Aspectos Algébricos da Solução de Mínimos Quadrados

- ▶ As equações normais são

$$\mathbf{X}'\mathbf{X}\mathbf{b} - \mathbf{X}'\mathbf{y} = -\mathbf{X}'(\mathbf{y} - \mathbf{X}\mathbf{b}) = -\mathbf{X}'\mathbf{e} = \mathbf{0}.$$

- ▶ Para cada coluna \mathbf{x}_k de \mathbf{X} , temos $\mathbf{x}_k'\mathbf{e} = 0$.
- ▶ Se a primeira coluna de \mathbf{X} for de 1's, então
 1. Os resíduos de mínimos quadrados somam zero, $\mathbf{x}_1'\mathbf{e} = \mathbf{1}'\mathbf{e} = \sum_{i=1}^n e_i = 0$.
 2. A regressão no hiperplano passa pelo ponto das médias dos dados. A primeira equação normal implica que $\bar{y} = \bar{\mathbf{x}}'\mathbf{b}$.
 3. A média dos valores ajustados da regressão é igual a média dos valores atuais.

Isto só vale se a regressão tiver uma constante.

Projeção

O valor dos resíduos de mínimos quadrados é dado por

$$\begin{aligned}\mathbf{e} &= \mathbf{y} - \mathbf{X}\mathbf{b} \\ &= \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{y} - \mathbf{P}\mathbf{y} \\ &= \mathbf{I}\mathbf{y} - \mathbf{P}\mathbf{y} \\ &= (\mathbf{I} - \mathbf{P})\mathbf{y} \\ &= \mathbf{M}\mathbf{y},\end{aligned}$$

sendo $\mathbf{P} \triangleq \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ e $\mathbf{M} \triangleq \mathbf{I} - \mathbf{P}$.

Temos que \mathbf{P} é a matriz de projeção (leva \mathbf{y} em $\hat{\mathbf{y}} = \mathbf{X}\mathbf{b}$):

$$\mathbf{P} = \mathbf{P}', \quad \text{e} \quad \mathbf{P} = \mathbf{P}^2.$$

E \mathbf{M} é uma matriz importante pois leva \mathbf{y} em \mathbf{e} :

$$\mathbf{M} = \mathbf{M}', \quad \mathbf{M} = \mathbf{M}^2, \quad \text{e} \quad \mathbf{M}\mathbf{X} = \mathbf{0}.$$

Toda matriz simétrica e idempotente é positiva definida, em particular \mathbf{P} e \mathbf{M} .

A Matriz de Projeção

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\mathbf{b} + \mathbf{e} \quad \Leftrightarrow \quad \mathbf{y} = \hat{\mathbf{y}} + \mathbf{e} \quad \Rightarrow \\ \hat{\mathbf{y}} &= \mathbf{y} - \mathbf{e} = (\mathbf{I} - \mathbf{M})\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{P}\mathbf{y}. \end{aligned}$$

$$\mathbf{P}\mathbf{M} = \mathbf{M}\mathbf{P} = \mathbf{0}, \quad \text{e} \quad \mathbf{P}\mathbf{X} = \mathbf{X}.$$

$$\mathbf{y} = \mathbf{P}\mathbf{y} + \mathbf{M}\mathbf{y} = \text{projeção} + \text{resíduos}.$$

Podemos reescrever a soma de quadrados da seguinte forma:

$$\begin{aligned} \mathbf{y}'\mathbf{y} &= \mathbf{y}'\mathbf{P}'\mathbf{P}\mathbf{y} + \mathbf{y}'\mathbf{M}'\mathbf{M}\mathbf{y} \\ &= \hat{\mathbf{y}}'\hat{\mathbf{y}} + \mathbf{e}'\mathbf{e}. \end{aligned}$$

Outras relações úteis:

$$\begin{aligned} \mathbf{e}'\mathbf{e} &= \mathbf{y}'\mathbf{M}'\mathbf{M}\mathbf{y} = \mathbf{y}'\mathbf{M}\mathbf{y} = \mathbf{y}'\mathbf{e} = \mathbf{e}'\mathbf{y} \\ \mathbf{e}'\mathbf{e} &= \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b}. \end{aligned}$$

Regressão Particionada e Regressão Parcial

Suponha que a regressão envolva dois conjuntos de variáveis \mathbf{X}_1 e \mathbf{X}_2 . Logo,

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon.$$

Qual é a solução algébrica para \mathbf{b}_2 ?

$$\begin{bmatrix} \mathbf{X}'_1\mathbf{X}_1 & \mathbf{X}'_1\mathbf{X}_2 \\ \mathbf{X}'_2\mathbf{X}_1 & \mathbf{X}'_2\mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_1\mathbf{y} \\ \mathbf{X}'_2\mathbf{y} \end{bmatrix}$$

Resolvendo para \mathbf{b}_1 (primeira equação):

$$\mathbf{X}'_1\mathbf{X}_1\mathbf{b}_1 + \mathbf{X}'_1\mathbf{X}_2\mathbf{b}_2 = \mathbf{X}'_1\mathbf{y}$$

$$(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{X}_1\mathbf{b}_1 + (\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{X}_2\mathbf{b}_2 = (\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{y}$$

$$\mathbf{b}_1 = (\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{y} - (\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{X}_2\mathbf{b}_2$$

$$= (\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1(\mathbf{y} - \mathbf{X}_2\mathbf{b}_2).$$

\mathbf{b}_1 é o conjunto de coeficientes na regressão de \mathbf{y} sobre \mathbf{X}_1 , menos um vetor de correção. Suponha que $\mathbf{X}'_1\mathbf{X}_2 = \mathbf{0}$, então $\mathbf{b}_1 = (\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{y}$.

Na solução de \mathbf{b}_2 temos

$$(\mathbf{X}_2' \mathbf{X}_1) \mathbf{b}_1 + (\mathbf{X}_2' \mathbf{X}_2) \mathbf{b}_2 = \mathbf{X}_2' \mathbf{y}$$

$$(\mathbf{X}_2' \mathbf{X}_1)(\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' (\mathbf{y} - \mathbf{X}_2 \mathbf{b}_2) + (\mathbf{X}_2' \mathbf{X}_2) \mathbf{b}_2 = \mathbf{X}_2' \mathbf{y}$$

$$\mathbf{X}_2' \mathbf{P}_1 \mathbf{y} - \mathbf{X}_2' \mathbf{P}_1 \mathbf{X}_2 \mathbf{b}_2 + (\mathbf{X}_2' \mathbf{X}_2) \mathbf{b}_2 = \mathbf{X}_2' \mathbf{y}$$

$$\mathbf{X}_2' (\mathbf{I} - \mathbf{P}_1) \mathbf{X}_2 \mathbf{b}_2 = \mathbf{X}_2' (\mathbf{I} - \mathbf{P}_1) \mathbf{y}$$

$$\mathbf{X}_2' \mathbf{M}_1 \mathbf{X}_2 \mathbf{b}_2 = \mathbf{X}_2' \mathbf{M}_1 \mathbf{y}$$

$$\mathbf{b}_2 = (\mathbf{X}_2' \mathbf{M}_1 \mathbf{X}_2)^{-1} \mathbf{X}_2' \mathbf{M}_1 \mathbf{y} = (\mathbf{X}_2^{*'} \mathbf{X}_2^*)^{-1} \mathbf{X}_2^{*'} \mathbf{y}^*,$$

sendo $\mathbf{X}_2^* = \mathbf{M}_1 \mathbf{X}_2$ e $\mathbf{y}^* = \mathbf{M}_1 \mathbf{y}$.

\mathbf{M}_1 é a matriz construtora de resíduos definida para a regressão sobre as colunas de \mathbf{X}_1 .

Logo, $\mathbf{M}_1 \mathbf{X}_2$ é uma matriz de resíduos.

Cada coluna de $\mathbf{M}_1 \mathbf{X}_2$ é um vetor de resíduos da regressão correspondente da coluna \mathbf{X}_2 sobre as variáveis em \mathbf{X}_1 .

Teorema (Frisch-Waugh (1933) e Lovell (1963)): *Na regressão linear de mínimos quadrados do vetor \mathbf{y} sobre dois conjuntos de variáveis \mathbf{X}_1 e \mathbf{X}_2 , o subvetor \mathbf{b}_2 é o conjunto de coeficientes obtidos quando os resíduos da regressão de \mathbf{y} sobre \mathbf{X}_1 sozinho são regredidos sobre o conjunto de resíduos obtidos quando cada coluna de \mathbf{X}_2 é regredida sobre \mathbf{X}_1 .*

Considere uma regressão de \mathbf{y} sobre um conjunto de variáveis \mathbf{X} e uma variável adicional \mathbf{z} , e seus respectivos coeficientes por \mathbf{b} e c .

Corolário: *O coeficiente sobre \mathbf{z} em uma regressão múltipla de \mathbf{y} sobre $\mathbf{W} = [\mathbf{X}, \mathbf{z}]$ é calculada como*

$$c = (\mathbf{z}' \mathbf{M} \mathbf{z})^{-1} \mathbf{z}' \mathbf{M} \mathbf{y} = (\mathbf{z}^{*'} \mathbf{z}^*)^{-1} \mathbf{z}^* \mathbf{y}^*,$$

sendo $\mathbf{z}^* = \mathbf{M} \mathbf{z}$ e $\mathbf{y}^* = \mathbf{M} \mathbf{y}$ os vetores de resíduos da regressão de mínimos quadrados de \mathbf{z} e \mathbf{y} sobre \mathbf{X} . Lembrando que $\mathbf{M} = \mathbf{I} - \mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'$.

Coeficiente de Correlação Parcial

Seja a regressão $y = \beta_1 + \beta_2 z + \beta_3 x + \varepsilon$. Além disso,

(a) y_i^* os resíduos da regressão de y sobre x , isto é, $y_i^* = y_i - a - bx_i$; e

(b) z_i^* os resíduos da regressão de z sobre x , isto é, $z_i^* = z_i - c - dx_i$,

para $i = 1, 2, \dots, n$.

Então, a correlação parcial r_{yz}^* é a correlação simples entre y^* e z^* , isto é,

$$r_{yz}^* = \frac{\mathbf{z}^{*'} \mathbf{y}^*}{\sqrt{(\mathbf{z}^{*'} \mathbf{z}^*)(\mathbf{y}^{*'} \mathbf{y}^*)}}.$$

pois o modelo de regressão tem a constante β_1 . Daí, $\mathbf{1}' \mathbf{z}^* = \mathbf{1}' \mathbf{y}^* = 0$.

Notemos que $\mathbf{y}^* = \mathbf{M}\mathbf{y}$ e $\mathbf{z}^* = \mathbf{M}\mathbf{z}$ com $\mathbf{M} = \mathbf{I} - \mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Soma de Quadrados

Considere agora os resíduos \mathbf{u} da regressão: $\mathbf{y} = \mathbf{X}\mathbf{d} + \mathbf{z}\mathbf{c} + \mathbf{u}$.

A menos que $\mathbf{X}'\mathbf{z} = \mathbf{0}$, \mathbf{d} não será igual a $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

A menos que $\mathbf{c} = 0$, \mathbf{u} não será igual a $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$.

Notemos que

$$\begin{aligned} \mathbf{c} &= (\mathbf{z}^{*'} \mathbf{z}^*)^{-1} \mathbf{z}^{*'} \mathbf{y}^*, \quad \text{e} \\ \mathbf{d} &= (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' (\mathbf{y} - \mathbf{z} \mathbf{c}) = \mathbf{b} - (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{z} \mathbf{c}. \end{aligned}$$

Segue-se que

$$\mathbf{u} = \mathbf{y} - \mathbf{X} \mathbf{b} + \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{z} \mathbf{c} - \mathbf{z} \mathbf{c} = \mathbf{e} - \mathbf{M} \mathbf{z} \mathbf{c} = \mathbf{e} - \mathbf{z}^* \mathbf{c}.$$

Portanto,

$$\mathbf{u}' \mathbf{u} = \mathbf{e}' \mathbf{e} + \mathbf{c}^2 (\mathbf{z}^{*'} \mathbf{z}^*) - 2 \mathbf{c} \mathbf{z}^{*'} \mathbf{e}.$$

Mas

$$\begin{aligned} \mathbf{e} &= \mathbf{M} \mathbf{y} = \mathbf{y}^* \\ \mathbf{z}^{*'} \mathbf{e} &= \mathbf{z}^{*'} \mathbf{M} \mathbf{y} = \mathbf{z}' \mathbf{M}' \mathbf{M} \mathbf{y} = \mathbf{z}' \mathbf{M}' \mathbf{y} = \mathbf{z}^{*'} \mathbf{y}^* \\ \mathbf{z}^{*'} \mathbf{e} &= \mathbf{z}^{*'} \mathbf{y}^* = \mathbf{c} (\mathbf{z}^{*'} \mathbf{z}^*), \end{aligned}$$

pois pela equação normal $(\mathbf{z}^{*'} \mathbf{z}^*) \mathbf{c} = \mathbf{z}^{*'} \mathbf{y}^*$.

Assim, $\mathbf{u}' \mathbf{u} = \mathbf{e}' \mathbf{e} - \mathbf{c}^2 (\mathbf{z}^{*'} \mathbf{z}^*)$.

Teorema (Mudança na soma de quadrados quando uma variável é adicionada ao modelo): Se $\mathbf{e}'\mathbf{e}$ é a soma de quadrados dos resíduos quando \mathbf{y} é regredido sobre \mathbf{X} e $\mathbf{u}'\mathbf{u}$ é a soma de quadrados dos resíduos quando \mathbf{y} é regredido sobre \mathbf{X} e \mathbf{z} , então

$$\mathbf{u}'\mathbf{u} = \mathbf{e}'\mathbf{e} - c^2(\mathbf{z}^{*\prime}\mathbf{z}^*) \leq \mathbf{e}'\mathbf{e},$$

sendo c o coeficiente de \mathbf{z} na regressão completa e $\mathbf{z}^* = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{z}$ o vetor de resíduos quando \mathbf{z} é regredido sobre \mathbf{X} .

Temos que $\mathbf{e}'\mathbf{e} = \mathbf{y}^{*\prime}\mathbf{y}^*$ e pela equação normal

$$c(\mathbf{z}^{*\prime}\mathbf{z}^*) = \mathbf{z}^{*\prime}\mathbf{y}^* \Rightarrow c^2(\mathbf{z}^{*\prime}\mathbf{z}^*) = (\mathbf{z}^{*\prime}\mathbf{y}^*)^2/(\mathbf{z}^{*\prime}\mathbf{z}^*).$$

Portanto,

$$\frac{\mathbf{u}'\mathbf{u}}{\mathbf{y}^{*\prime}\mathbf{y}^*} = \frac{\mathbf{y}^{*\prime}\mathbf{y}^* - (\mathbf{z}^{*\prime}\mathbf{y}^*)^2/\mathbf{z}^{*\prime}\mathbf{z}^*}{\mathbf{y}^{*\prime}\mathbf{y}^*} = 1 - r_{yz}^{*2}.$$

Exemplo 4 (continuação): $y_i = \beta_1 + \beta_2 T_i + \beta_3 G_i + \beta_3 R_i + \beta_4 P_i + \varepsilon_i$

Seguimos a análise dos dados da equação de investimento.

Calculamos a correlação (simples) entre a variável investimento e as variáveis de tendência temporal (T), produto interno bruto (G), taxa de juros (R) e taxa de inflação (P).

Calculamos também a correlação parcial entre investimento e as quatro regressoras dado as demais regressoras.

Tabela : Correlação entre investimento e as demais variáveis.

	Correlação Simples	Correlação Parcial
Tendência	0,7496	-0,9360
PIB	0,8632	0,9680
Taxa de juros	0,5871	-0,5167
Taxa de inflação	0,4777	-0,0221

(Mostrar exemplo no R: exemplo_04.r pag. 35)

Qualidade de Ajuste e Análise de Variância

Podemos perguntar se a variação em \mathbf{x} é um bom preditor da variação em y .

A variação total é dada por:

$$SQTot = \sum_{i=1}^n (y_i - \bar{y})^2.$$

Já sabemos que $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e} = \hat{\mathbf{y}} + \mathbf{e}$. Logo, para uma única observação, temos que

$$\begin{aligned} y_i &= \hat{y}_i + e_i = \mathbf{x}_i' \mathbf{b} + e_i \\ y_i - \bar{y} &= \hat{y}_i - \bar{y} + e_i = (\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{b} + e_i. \end{aligned}$$

Exemplo 5: (Detalhes no arquivo R: exemplo_05.r pag. 37)

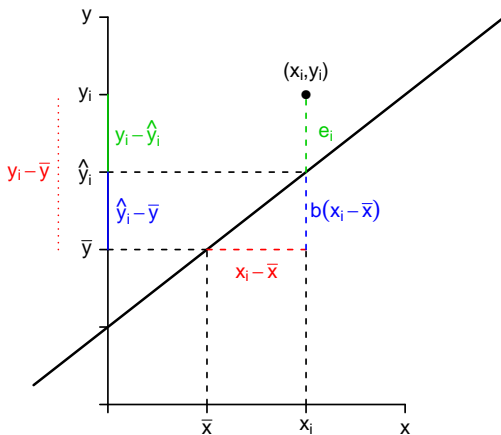


Figura : Decomposição de y .

Algumas Propriedades, Notações e Somas de Quadrados

$$\mathbf{1}\bar{X} = \mathbf{1}\frac{1}{n}\mathbf{1}'\mathbf{x} = \begin{bmatrix} \bar{X} \\ \bar{X} \\ \vdots \\ \bar{X} \end{bmatrix} = \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{x}$$

$$\begin{bmatrix} x_1 - \bar{X} \\ x_2 - \bar{X} \\ \vdots \\ x_n - \bar{X} \end{bmatrix} = [\mathbf{x} - \mathbf{1}\bar{X}] = [\mathbf{x} - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{x}] = [\mathbf{I}\mathbf{x} - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{x}] = [\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}']\mathbf{x} = \tilde{\mathbf{M}}\mathbf{x}.$$

Definimos $\tilde{\mathbf{M}} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'$. Esta também é uma matriz idempotente.

Temos que $\tilde{\mathbf{M}}\mathbf{1} = \mathbf{0} \Leftrightarrow \mathbf{1}'\tilde{\mathbf{M}} = \mathbf{0}'$. Logo, $\sum_{i=1}^n (x_i - \bar{X}) = \mathbf{1}'[\tilde{\mathbf{M}}\mathbf{x}] = \mathbf{0}'\mathbf{x} = 0$. Além disso,

$$\sum_{i=1}^n (x_i - \bar{X})^2 = (\mathbf{x} - \mathbf{1}\bar{X})'(\mathbf{x} - \mathbf{1}\bar{X}) = (\tilde{\mathbf{M}}\mathbf{x})'(\tilde{\mathbf{M}}\mathbf{x}) = \mathbf{x}'\tilde{\mathbf{M}}'\tilde{\mathbf{M}}\mathbf{x} = \mathbf{x}'\tilde{\mathbf{M}}\mathbf{x},$$

$$\begin{aligned}\tilde{\mathbf{M}}\mathbf{y} &= \tilde{\mathbf{M}}\mathbf{X}\mathbf{b} + \tilde{\mathbf{M}}\mathbf{e} \\ \tilde{\mathbf{M}}\mathbf{e} &= \mathbf{e} \quad \text{e} \quad \mathbf{e}'\tilde{\mathbf{M}}\mathbf{X} = \mathbf{e}'\mathbf{X} = \mathbf{0}'.\end{aligned}$$

Finalmente,

$$\underbrace{\mathbf{y}'\tilde{\mathbf{M}}\mathbf{y}}_{SQ_{Tot}} = \underbrace{\mathbf{b}'\mathbf{X}'\tilde{\mathbf{M}}\mathbf{X}\mathbf{b}}_{SQ_{Reg}} + \underbrace{\mathbf{e}'\mathbf{e}}_{SQ_{Res}}.$$

- ▶ SQ_{Tot} é a soma de quadrados total;
- ▶ SQ_{Reg} é a soma de quadrados da regressão; e
- ▶ SQ_{Res} é a soma de quadrados dos resíduos.

Coeficiente de Determinação

$$R^2 = \frac{SQ_{Reg}}{SQ_{Tot}} = \frac{\mathbf{b}'\mathbf{X}'\tilde{\mathbf{M}}\mathbf{X}\mathbf{b}}{\mathbf{y}'\tilde{\mathbf{M}}\mathbf{y}} = 1 - \frac{\mathbf{e}'\mathbf{e}}{\mathbf{y}'\tilde{\mathbf{M}}\mathbf{y}}.$$

Tabela : Análise de variância.

Soma de Quadrados	Fonte	Graus de Liberdade	Média Quadrática
Regressão	$\mathbf{b}'\mathbf{X}'\mathbf{y} - n\bar{y}^2$	$p - 1$	
Resíduo	$\mathbf{e}'\mathbf{e}$	$n - p$	s^2
Total	$\mathbf{y}'\mathbf{y} - n\bar{y}^2$	$n - 1$	$S_{yy}/(n - 1) = s_y^2$

Temos

$$\begin{aligned}
 \mathbf{b}'\mathbf{X}'\tilde{\mathbf{M}}\mathbf{X}\mathbf{b} &= \mathbf{b}'\mathbf{X}'\tilde{\mathbf{M}}[\mathbf{y} - \mathbf{e}] = \mathbf{b}'\mathbf{X}'\tilde{\mathbf{M}}\mathbf{y} - \mathbf{b}'\mathbf{X}'\tilde{\mathbf{M}}\mathbf{e} \\
 &= \mathbf{b}'\mathbf{X}'\left[\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right]\mathbf{y} \quad (\text{porque } \mathbf{X}'\tilde{\mathbf{M}}\mathbf{e} = \mathbf{0}) \\
 &= \mathbf{b}'\mathbf{X}'\mathbf{y} - \frac{1}{n}\mathbf{b}'\mathbf{X}'\mathbf{1}\mathbf{1}'\mathbf{y} = \mathbf{b}'\mathbf{X}'\mathbf{y} - \frac{1}{n}\hat{\mathbf{y}}'\mathbf{1}\mathbf{1}'\mathbf{y} \\
 &= \mathbf{b}'\mathbf{X}'\mathbf{y} - \frac{1}{n}(\mathbf{y} - \mathbf{e})\mathbf{1}\mathbf{1}'\mathbf{y} = \mathbf{b}'\mathbf{X}'\mathbf{y} - \frac{1}{n}\left[\sum_{i=1}^n y_i - \sum_{i=1}^n e_i\right]\sum_{i=1}^n y_i \\
 &= \mathbf{b}'\mathbf{X}'\mathbf{y} - n\bar{y}^2.
 \end{aligned}$$

Exemplo 4 (continuação): $y_i = \beta_1 + \beta_2 T_i + \beta_3 G_i + \beta_3 R_i + \beta_4 P_i + \varepsilon_i$

(Mostrar exemplo no R: exemplo_04.r pag. 41)

Tabela : Análise de variância: equação de investimento.

Soma de Quadrados	Fonte	Graus de Liberdade	Média Quadrática	R^2
Regressão	0,0159025	4	0,00397563	0,9724
Resíduo	0,0004508	10	0,00004508	
Total	0,0163533	14	0,00116810	

Teorema (Mudança no R^2 quando uma variável é adicionada ao modelo): Seja R_{Xz}^2 o coeficiente de determinação na regressão de \mathbf{y} sobre \mathbf{X} e uma variável adicional \mathbf{z} (**u os resíduos**), seja R_X^2 o mesmo para a regressão de \mathbf{y} sobre \mathbf{X} sozinha (**e os resíduos**), e seja r_{yz}^{*2} o coeficiente de correlação parcial entre y e z . Então,

$$R_{Xz}^2 = R_X^2 + (1 - R_X^2)r_{yz}^{*2} \geq R_X^2.$$

Prova: Vimos que $\mathbf{z}^* = \mathbf{Mz}$, $\mathbf{y}^* = \mathbf{My} = \mathbf{e} \Rightarrow \mathbf{e}'\mathbf{e} = \mathbf{y}^{*'}\mathbf{y}^*$, e

$$\mathbf{u}'\mathbf{u} = \mathbf{e}'\mathbf{e} - c^2(\mathbf{z}^{*'}\mathbf{z}^*) = \mathbf{y}^{*'}\mathbf{y}^* - \frac{(\mathbf{z}^{*'}\mathbf{y}^*)^2}{\mathbf{z}^{*'}\mathbf{z}^*} = \mathbf{e}'\mathbf{e}(1 - r_{yz}^{*2}).$$

Logo,

$$R_{Xz}^2 = 1 - \frac{\mathbf{u}'\mathbf{u}}{\mathbf{y}'\tilde{\mathbf{M}}\mathbf{y}} = 1 - \frac{\mathbf{e}'\mathbf{e}}{\mathbf{y}'\tilde{\mathbf{M}}\mathbf{y}} + \frac{\mathbf{e}'\mathbf{e}}{\mathbf{y}'\tilde{\mathbf{M}}\mathbf{y}}r_{yz}^{*2} = R_X^2 + (1 - R_X^2)r_{yz}^{*2} \geq R_X^2.$$

Exemplo 6: Ajuste de uma Função Consumo

(Mostrar exemplo no R: `exemplo_06.r` pag. 43-44)

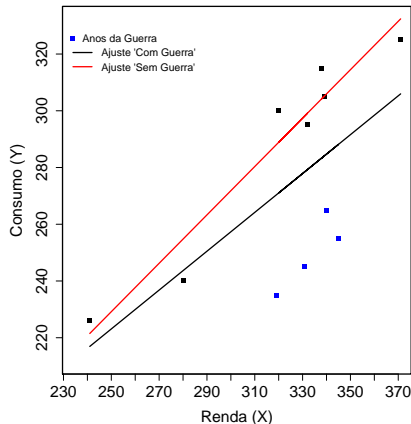


Figura : Dados de consumo e renda de 1940-1950 dos EUA. Fonte: *Economic Report of the President, U.S. Government Printing Office, Washington, D.C., 1983.*

Exemplo 6 (continuação): Ajuste de uma Função Consumo

Tabela : Análise de variância: consumo e renda.

Soma de Quadrados	Fonte	Graus de Liberdade	Média Quadrática	R^2
Regressão	5768,2068	1	5768,2068	0,4571
Resíduo	6849,9751	9	761,1083	
Total	12618,1818	10	1261,8182	

Tabela : Análise de variância: consumo, renda e dicotômica para anos de guerra.

Soma de Quadrados	Fonte	Graus de Liberdade	Média Quadrática	R^2
Regressão x	5768,2068	1	5768,2068	0,9464
Regressão w	6173,5189	1	6173,5189	
Regressão	11941,7257	2	5970,8629	
Resíduo	676,4562	8	84,5570	
Total	12618,1818	10	1261,8182	

O R^2 Ajustado para os Graus de Liberdade

R^2 nunca decresce quando uma nova variável regressora é adicionado a equação de regressão. Portanto, é tentador adicionar diversas variáveis regressoras no modelo para que R^2 se aproxime de 1.

O R^2 ajustado para os graus de liberdade é dado por

$$\bar{R}^2 = 1 - \frac{\mathbf{e}'\mathbf{e}/(n-p)}{\mathbf{y}'\tilde{\mathbf{M}}\mathbf{y}/(n-1)}.$$

Para efeitos computacionais, temos

$$\bar{R}^2 = 1 - \frac{n-1}{n-p} \times R^2.$$

- ▶ O \bar{R}^2 **pode decrescer** quando uma nova variável é adicionada ao conjunto de variáveis independentes.
- ▶ Na verdade, \bar{R}^2 **pode até ser negativo**.

- Uma alternativa para comparar modelos com uma maior penalização pela perda dos graus de liberdade é dada por

$$\overline{R}^2 = 1 - \frac{n+p}{n-p} \times R^2.$$

O modelo que tem \overline{R}^2 máximo é o mesmo que tem o valor mínimo do critério de predição (CP)

$$CP = \frac{\mathbf{e}'\mathbf{e}}{n-p} \left(1 + \frac{p}{n}\right) = s^2 \left(1 + \frac{p}{n}\right).$$

- Veremos outros critérios de comparação de modelos baseados no máximo da função de verossimilhança com penalização pelo número de parâmetros (AIC, BIC, HQIC, etc).

R^2 e o Termo Constante do Modelo

A prova que $0 \leq R^2 \leq 1$ requer que \mathbf{X} contenha uma coluna de 1's.

Se isto não ocorrer, então $\tilde{\mathbf{M}}\mathbf{e} \neq \mathbf{e}$, $\mathbf{e}'\tilde{\mathbf{M}}\mathbf{X} \neq \mathbf{0}$ e o termo $2\mathbf{e}'\tilde{\mathbf{M}}\mathbf{X}\mathbf{b}$ em

$$\mathbf{y}'\tilde{\mathbf{M}}\mathbf{y} = (\tilde{\mathbf{M}}\mathbf{X}\mathbf{b} + \tilde{\mathbf{M}}\mathbf{e})'(\tilde{\mathbf{M}}\mathbf{X}\mathbf{b} + \tilde{\mathbf{M}}\mathbf{e})$$

não se cancela.

Daí,

$$R^2 = 1 - \frac{\mathbf{e}'\mathbf{e}}{\mathbf{y}'\tilde{\mathbf{M}}\mathbf{y}}$$

pode resultar em qualquer valor.

Portanto, devemos ter cuidado ao analisar o R^2 de um modelo de regressão que não contenha o termo constante.

Este é um dos motivos que, em geral, deixa-se o termo constante no modelo regressão mesmo que testes de hipóteses indiquem o contrário.

Regressão Transformada Linearmente

Na regressão de \mathbf{y} sobre \mathbf{X} , suponha que as colunas de \mathbf{X} são transformadas linearmente.

Exemplo: (Precificação de arte) A teoria I da determinação de preços de leilão em pinturas de Monet diz que o preço é determinado pelas dimensões (largura W e altura H) da pintura,

$$\begin{aligned}\ln \text{preço} &= \beta_1 + \beta_2 \ln W + \beta_3 \ln H + \varepsilon \\ &= \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon.\end{aligned}$$

A teoria II diz que os compradores estão interessados na área da pintura e a razão do aspecto,

$$\begin{aligned}\ln \text{preço} &= \gamma_1 + \gamma_2 \ln(WH) + \gamma_3 \ln(W/H) + \xi \\ &= \gamma_1 z_1 + \gamma_2 z_2 + \gamma_3 z_3 + \xi.\end{aligned}$$

É evidente que $z_1 = x_1$, $z_2 = x_2 + x_3$ e $z_3 = x_2 - x_3$.

Em termos matriciais, temos $\mathbf{Z} = \mathbf{XP}$ sendo

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

Teorema: Na regressão linear de \mathbf{y} sobre $\mathbf{Z} = \mathbf{XP}$ sendo \mathbf{P} uma matriz não singular que transforma as colunas de \mathbf{X} , os coeficientes são iguais a $\mathbf{P}^{-1}\mathbf{b}$ sendo \mathbf{b} o vetor de coeficientes da regressão linear de \mathbf{y} sobre \mathbf{X} . Além disto, os R^2 das duas regressões são idênticos.

Prova: Os coeficientes são

$$\begin{aligned} \mathbf{d} &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = [(\mathbf{XP})'(\mathbf{XP})]^{-1}(\mathbf{XP})'\mathbf{y} = (\mathbf{P}'\mathbf{X}'\mathbf{XP})^{-1}\mathbf{P}'\mathbf{X}'\mathbf{y} \\ &= \mathbf{P}^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{P}^{-1}\mathbf{P}'\mathbf{X}'\mathbf{y} = \mathbf{P}^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{P}^{-1}\mathbf{b}. \end{aligned}$$

Temos também que

$$\mathbf{u} = \mathbf{y} - \mathbf{Z}\mathbf{d} = \mathbf{y} - \mathbf{Z}\mathbf{P}^{-1}\mathbf{b} = \mathbf{y} - \mathbf{XPP}^{-1}\mathbf{b} = \mathbf{y} - \mathbf{Xb} = \mathbf{e}.$$

Logo, $\mathbf{u}'\mathbf{u} = \mathbf{e}'\mathbf{e} \Rightarrow R^2_{y|z} = R^2_{y|x}$.

Hipóteses do Modelo de Regressão Linear

- HP.1: Linearidade: $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{ip}\beta_p + \varepsilon_i$;
- HP.2: Posto completo: a matriz \mathbf{X} $n \times p$ é de posto completo;
- HP.3: Exogeneidade das variáveis independentes: $E(\varepsilon_i | \mathbf{x}_i) = 0$;
- HP.4: Homocedasticidade e autocorrelação nula: cada distúrbio, ε_i , tem a mesma variância, σ^2 , e é não correlacionado com qualquer outro distúrbio ε_j ;
- HP.5: Geração dos dados: os dados em $(x_{i1}, x_{i2}, \dots, x_{ip})$ podem ser qualquer mistura de constantes e variáveis aleatórias; e
- HP.6: Distribuição normal: os distúrbios são normalmente distribuídos.

O Estimador de Mínimos Quadrados

Por hipótese,

$$E(\varepsilon|\mathbf{x}) = 0 \Rightarrow \text{Cov}(\mathbf{x}, \varepsilon) = \text{Cov}(E(\varepsilon|\mathbf{x}), \mathbf{x}) = \mathbf{0} \quad \text{e} \quad E(\varepsilon) = E_{\mathbf{x}}(E(\varepsilon|\mathbf{x})) = 0.$$

Mas,

$$\begin{aligned} \text{Cov}(\mathbf{x}, \varepsilon) &= E_{\mathbf{x}, \varepsilon}(\mathbf{x}\varepsilon) = E_{\mathbf{x}, y}(\mathbf{x}(y - \mathbf{x}'\beta)) = \mathbf{0} \quad \Rightarrow \\ E_{\mathbf{x}, y}(\mathbf{x}y) &= E_{\mathbf{x}}(\mathbf{x}\mathbf{x}')\beta \quad (\text{população}). \end{aligned}$$

Tomemos a equação normal de mínimos quadrados $\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\mathbf{b}$ que dividida por n , podemos reescrever como

$$\left(\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i y_i \right) = \left(\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' \right) \mathbf{b} \quad (\text{amostra}).$$

Assim, o estimador de mínimos está parecido com sua respectiva quantidade populacional. Podemos aplicar a lei dos grandes números sob certas condições.

Preditor Linear Ótimo

Considere o problema de encontrar um preditor linear ótimo para y .

Critério: o erro quadrático médio (EQM) do preditor linear y .

Vamos ignorar a [HP.6](#) e desconsiderar [HP.1](#) que $E(y|\mathbf{x})$ é linear.

Queremos minimizar

$$\text{EQM} = E_{\mathbf{x},y}([y - \mathbf{x}'\boldsymbol{\gamma}]^2).$$

Isto pode ser reescrito como

$$\text{EQM} = E_{\mathbf{x},y}([y - E(y|\mathbf{x})]^2) + E_{\mathbf{x},y}([E(y|\mathbf{x}) - \mathbf{x}'\boldsymbol{\gamma}]^2).$$

A condição necessária é

$$\begin{aligned} \frac{\partial E_{\mathbf{x},y}([E(y|\mathbf{x}) - \mathbf{x}'\boldsymbol{\gamma}]^2)}{\partial \boldsymbol{\gamma}} &= E_{\mathbf{x},y} \left(\frac{\partial [E(y|\mathbf{x}) - \mathbf{x}'\boldsymbol{\gamma}]^2}{\partial \boldsymbol{\gamma}} \right) \\ &= -2E_{\mathbf{x},y}(\mathbf{x}[E(y|\mathbf{x}) - \mathbf{x}'\boldsymbol{\gamma}]) = \mathbf{0} \Rightarrow \\ E_{\mathbf{x},y}(\mathbf{x}E(y|\mathbf{x})) &= E_{\mathbf{x},y}(\mathbf{x}\mathbf{x}')\boldsymbol{\gamma}. \end{aligned}$$

Mas

$$\begin{aligned} E_{\mathbf{x},y}(\mathbf{x}E(y|\mathbf{x})) &= \text{Cov}(\mathbf{x}, E(y|\mathbf{x})) + E_{\mathbf{x}}(\mathbf{x})E_{\mathbf{x}}(E(y|\mathbf{x})) \\ &= \text{Cov}(\mathbf{x}, y) + E_{\mathbf{x}}(\mathbf{x})E_y(y) = E_{\mathbf{x},y}(\mathbf{x}y). \end{aligned}$$

Portanto, a condição necessária para minimizar o preditor linear, EQM, é

$$E_{\mathbf{x},y}(\mathbf{x}y) = E_{\mathbf{x},y}(\mathbf{x}\mathbf{x}')\gamma.$$

Teorema: *Se o processo de geração dos dados (y_i, \mathbf{x}_i) , $i = 1, 2, \dots, n$, é tal que a lei dos grandes números se aplica ao estimador*

$$\left(\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i y_i \right) = \left(\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' \right) \mathbf{b}$$

das matrizes

$$E_{\mathbf{x},y}(\mathbf{x}y) = E_{\mathbf{x}}(\mathbf{x}\mathbf{x}')\beta,$$

então o mínimo do preditor linear do erro quadrático esperado de y_i é estimado pela regressão de mínimos quadrados.

Estimação Não Tendenciosa

O estimador de mínimos quadrados é não tendencioso para cada amostra.

Temos

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} \quad \text{e}$$

$$E(\mathbf{b}|\mathbf{X}) = \boldsymbol{\beta} + E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}|\mathbf{X}) = \boldsymbol{\beta} \quad (\text{pela HP.3}).$$

Portanto,

$$E(\mathbf{b}) = E_X(E(\mathbf{b}|\mathbf{X})) = E_X(\boldsymbol{\beta}) = \boldsymbol{\beta}.$$

Viés: Omissão de Variáveis Relevantes

Suponha que o modelo correto seja

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon$$

sendo p_1 e p_2 os números de colunas em \mathbf{X}_1 e \mathbf{X}_2 respectivamente.

A regressão de \mathbf{y} sobre \mathbf{X}_1 (sem incluir \mathbf{X}_2) resulta no estimador

$$\mathbf{b}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y} = \beta_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\beta_2 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\varepsilon.$$

Logo,

$$E(\mathbf{b}_1|\mathbf{X}) = \beta_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\beta_2 = \beta_1 + \mathbf{P}_{1.2}\beta_2,$$

sendo $\mathbf{P}_{1.2} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$.

Portanto, \mathbf{b}_1 é um estimador tendencioso de β_1 , exceto para os casos que $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$ ou $\beta_2 = \mathbf{0}$.

A estimação do modelo pode ser vista ao impor **incorretamente** a restrição **$\beta_2 = \mathbf{0}$** . Isto introduz um viés na estimação de \mathbf{b}_1 .

Inclusão de Variáveis Irrelevantes

Suponha que o modelo de regressão correto seja

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon},$$

mas que estimamos como a modelo de regressão (variáveis extras)

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}.$$

Aqui, a inclusão de \mathbf{X}_2 no modelo de regressão pode ser vista como uma falha em impor que $\boldsymbol{\beta}_2 = \mathbf{0}$. Portanto, não a nada a se provar aqui.

Temos que

$$E(\mathbf{b}|\mathbf{X}) = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \mathbf{0} \end{bmatrix}.$$

É possível provar que ao aumentar o número de variáveis desnecessárias no modelo faz com que a precisão das estimativas diminua (a menos que $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$).

Exemplo 7: Distribuição Amostral do Estimador de Mínimos Quadrados

(Detalhes no arquivo R: exemplo_07.r pag. 57)

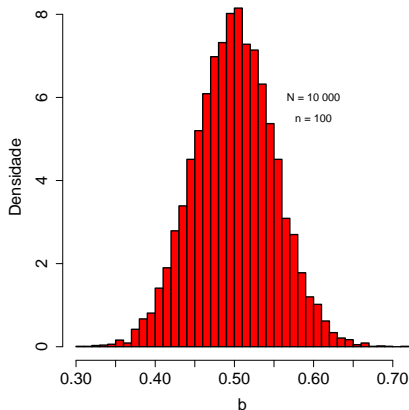


Figura : Distribuição amostral do estimador b de β na regressão $y = \alpha + \beta x + \varepsilon$.

Variância Amostral do Estimador de Mínimos Quadrados

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}.$$

Logo,

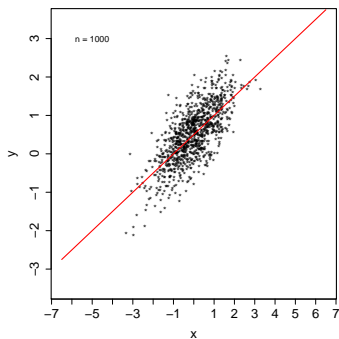
$$\begin{aligned}\text{Var}(\mathbf{b}|\mathbf{X}) &= \text{E}[(\mathbf{b} - \boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta})'| \mathbf{X}] \\ &= \text{E}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} | \mathbf{X}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' | \mathbf{X})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \quad (\text{pela HP.4}) \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.\end{aligned}$$

Suponha uma regressão linear simples da forma $y = \beta_1 + \beta_2 x + \varepsilon$. Assim, temos

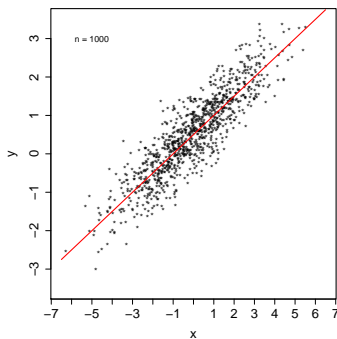
$$\text{Var}(b_2|\mathbf{X}) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Exemplo 8: Variância do Estimador de Mínimos Quadrados

(Detalhes no arquivo R: exemplo_08.r pag. 59-60)



(a) Menor variância em x .



(b) Maior variância em x .

Figura : Variância do estimador b de β na regressão $y = \alpha + \beta x + \varepsilon$.

Estimativas para os dados com menor dispersão em x:

```

                Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.52368     0.01579   33.16  <2e-16 ***
x1              0.51920     0.01566   33.15  <2e-16 ***

```

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

Residual standard error: 0.4992 on 998 degrees of freedom

Multiple R-squared: 0.524, Adjusted R-squared: 0.5235

Estimativas para os dados com maior dispersão em x:

```

                Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.52294     0.01580   33.09  <2e-16 ***
x2              0.49683     0.00826   60.15  <2e-16 ***

```

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

Residual standard error: 0.4996 on 998 degrees of freedom

Multiple R-squared: 0.7838, Adjusted R-squared: 0.7836

Agora, seja $\mathbf{b}_0 = \mathbf{C}\mathbf{y}$ um outro estimador não tendencioso e linear de β , sendo que \mathbf{C} é uma matriz $p \times n$. Então,

$$\beta = E(\mathbf{b}_0|\mathbf{X}) = E(\mathbf{C}\mathbf{y}|\mathbf{X}) = E(\mathbf{C}\mathbf{X}\beta + \mathbf{C}\varepsilon|\mathbf{X}) \Rightarrow \mathbf{C}\mathbf{X} = \mathbf{I}.$$

(existem vários candidatos para \mathbf{C} que satisfazem a igualdade) e

$$\text{Var}(\mathbf{b}_0|\mathbf{X}) = \sigma^2 \mathbf{C}\mathbf{C}'.$$

Agora, seja $\mathbf{D} = \mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ tal que $\mathbf{D}\mathbf{y} = \mathbf{b}_0 - \mathbf{b}$. Então, $\mathbf{b}_0 = \mathbf{D}\mathbf{y} + \mathbf{b} = (\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}$ e

$$\text{Var}(\mathbf{b}_0|\mathbf{X}) = \sigma^2 \left[(\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')' \right].$$

Sabemos que $\mathbf{C}\mathbf{X} = \mathbf{I} = \mathbf{D}\mathbf{X} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} \Rightarrow \mathbf{D}\mathbf{X} = \mathbf{0}$. Portanto,

$$\text{Var}(\mathbf{b}_0|\mathbf{X}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} + \sigma^2 \mathbf{D}\mathbf{D}' = \text{Var}(\mathbf{b}|\mathbf{X}) + \sigma^2 \mathbf{D}\mathbf{D}'.$$

e $\mathbf{D}\mathbf{D}'$ é uma forma quadrática ($\mathbf{a}'\mathbf{D}\mathbf{D}'\mathbf{a} \geq 0$).

Logo, $\text{Var}(\mathbf{b}|\mathbf{X}) \leq \text{Var}(\mathbf{b}_0|\mathbf{X})$ para qualquer $\mathbf{b}_0 \neq \mathbf{b}$ e \mathbf{X} fixo.

Teorema de Gauss-Markov

No modelo de regressão linear clássico com matriz de variáveis regressoras \mathbf{X} , o estimador de mínimos quadrados \mathbf{b} é o estimador não tendencioso e linear de variância mínima de β . Para qualquer vetor de constantes \mathbf{v} , o estimador não tendencioso e linear de variância mínima de $\mathbf{v}'\beta$ no modelo de regressão clássico é $\mathbf{v}'\mathbf{b}$, sendo \mathbf{b} o estimador de mínimos quadrados.

O teorema também implica que o estimador de mínimos quadrados para cada coeficiente também tem variância mínima. Basta tomar um vetor $\mathbf{v} = (0, \dots, 0, 1, 0, \dots, 0)$ com o valor 1 no coeficiente a ser estimado.

A variância incondicional é dada por

$$\begin{aligned}\text{Var}(\mathbf{b}) &= \text{E}_{\mathbf{X}}(\text{Var}(\mathbf{b}|\mathbf{X})) + \text{Var}_{\mathbf{X}}(\text{E}(\mathbf{b}|\mathbf{X})) \\ &= \sigma^2 \text{E}_{\mathbf{X}}((\mathbf{X}'\mathbf{X})^{-1}),\end{aligned}$$

pois $\text{E}(\mathbf{b}|\mathbf{X}) = \beta \Rightarrow \text{Var}_{\mathbf{X}}(\text{E}(\mathbf{b}|\mathbf{X})) = \mathbf{0}$.

Estimação da Variância

Sabemos que $\text{Var}(\mathbf{b}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. Precisamos estimar σ^2 .

Temos que $\text{Var}(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2$ e e_i é uma estimativa de ε_i . Então,

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n e_i^2$$

parece um estimador natural de σ^2 .

Os resíduos de mínimos quadrados são

$$\mathbf{e} = \mathbf{M}\mathbf{y} = \mathbf{M}[\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}] = \mathbf{M}\boldsymbol{\varepsilon}, \quad \text{pois} \quad \mathbf{M}\mathbf{X} = \mathbf{0}.$$

Este estimador de σ^2 é baseado em $\mathbf{e}'\mathbf{e} = \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}$.

Note que os resíduos são correlacionados:

$$\text{Var}(\mathbf{e}|\mathbf{X}) = E(\mathbf{e}\mathbf{e}'|\mathbf{X}) = E(\mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{M}|\mathbf{X}) = \mathbf{M}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X})\mathbf{M} = \mathbf{M}\sigma^2\mathbf{I}\mathbf{M} = \sigma^2\mathbf{M}.$$

Note que \mathbf{M} é uma matriz positiva definida.

Agora, estamos interessados em

$$E(\mathbf{e}'\mathbf{e}|\mathbf{X}) = E(\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}|\mathbf{X}).$$

A quantidade escalar $\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}$ é uma matriz 1×1 , logo é igual ao seu traço.

Segue-se que

$$\begin{aligned} E(\text{tr}(\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon})|\mathbf{X}) &= E(\text{tr}(\mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')|\mathbf{X}) = \text{tr}(E(\mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X})) = \text{tr}(\mathbf{M}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X})) \\ &= \text{tr}(\mathbf{M}\sigma^2\mathbf{I}) = \sigma^2\text{tr}(\mathbf{M}). \end{aligned}$$

O traço de \mathbf{M} é

$$\text{tr}(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \text{tr}(\mathbf{I}_n) - \text{tr}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}) = \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{I}_p) = n - p.$$

Portanto, $E(\mathbf{e}'\mathbf{e}|\mathbf{X}) = (n - p)\sigma^2$.

Logo, o estimador natural de σ^2 é tendencioso. Um estimador não tendencioso de σ^2 é

$$s^2 = \frac{\mathbf{e}'\mathbf{e}}{n - p}.$$

Este estimador é não tendencioso incondicionalmente também,

$$E(s^2) = E_X(E(s^2|\mathbf{X})) = E_X(\sigma^2) = \sigma^2.$$

Podemos estimar $\text{Var}(\mathbf{b}|\mathbf{X})$ por $\widehat{\text{Var}}(\mathbf{b}|\mathbf{X}) = s^2(\mathbf{X}'\mathbf{X})^{-1}$.

A raiz quadrada do k -ésimo elemento da diagonal de $s^2(\mathbf{X}'\mathbf{X})^{-1}$ é o erro padrão do estimador b_k .

Normalidade (HP.6)

Se tivermos a hipótese $\varepsilon \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{I})$, então

$$\begin{aligned}(\mathbf{b}|\mathbf{X}) &\sim \mathcal{N}_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}), \quad \text{e} \\(b_k|\mathbf{X}) &\sim \mathcal{N}(\beta_k, \sigma^2(\mathbf{X}'\mathbf{X})_{kk}^{-1}).\end{aligned}$$

Lembre-se que $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + \mathbf{A}\boldsymbol{\varepsilon}$ com $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, ou seja, sob HP.6 \mathbf{b} é uma combinação linear de normais. Portanto, \mathbf{b} (condicional a \mathbf{X}) tem distribuição normal.

Propriedades dos Estimadores de Mínimos Quadrados para Amostras Grandes

Utilizaremos somente **HP.1** até **HP.5** (sem normalidade).

Consistência de \mathbf{b}

Suponha que **HP.5a** $(\mathbf{x}_i, \varepsilon_i)$, $i = 1, \dots, n$ seja uma sequência de observações independentes e que

$$\text{plim} \frac{\mathbf{X}'\mathbf{X}}{n} = \mathbf{Q}, \quad \text{uma matriz positiva definida.}$$

Temos que

$$\mathbf{b} = \beta + \left(\frac{\mathbf{X}'\mathbf{X}}{n} \right)^{-1} \left(\frac{\mathbf{X}'\varepsilon}{n} \right).$$

Se \mathbf{Q}^{-1} existir, então

$$\text{plim} \mathbf{b} = \beta + \mathbf{Q}^{-1} \text{plim} \left(\frac{\mathbf{X}'\varepsilon}{n} \right).$$

pois a inversa é uma função contínua da matriz original.

Segue-se

$$\frac{1}{n} \mathbf{X}' \boldsymbol{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \varepsilon_i = \frac{1}{n} \sum_{i=1}^n \mathbf{w}_i = \bar{\mathbf{w}}.$$

Portanto,

$$\text{plim} \mathbf{b} = \boldsymbol{\beta} + \mathbf{Q}^{-1} \text{plim} \bar{\mathbf{w}}.$$

Temos que

$$E(\mathbf{w}_i) = E_{\mathbf{X}}(E(\mathbf{w}_i | \mathbf{X})) = E_{\mathbf{X}}(\mathbf{x}_i E(\varepsilon_i | \mathbf{X})) = \mathbf{0} \Rightarrow E(\bar{\mathbf{w}}) = \mathbf{0}.$$

Além disso,

$$\text{Var}(\bar{\mathbf{w}} | \mathbf{X}) = E(\bar{\mathbf{w}} \bar{\mathbf{w}}' | \mathbf{X}) = \frac{1}{n} \mathbf{X}' E(\varepsilon \varepsilon' | \mathbf{X}) \mathbf{X} \frac{1}{n} = \left(\frac{\sigma^2}{n} \right) \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right) \Rightarrow$$

$$\text{Var}(\bar{\mathbf{w}}) = E_{\mathbf{X}}(\text{Var}(\bar{\mathbf{w}} | \mathbf{X})) + \text{Var}_{\mathbf{X}}(E(\bar{\mathbf{w}} | \mathbf{X})) = \left(\frac{\sigma^2}{n} \right) E_{\mathbf{X}} \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right).$$

Se $\lim_{n \rightarrow \infty} E_{\mathbf{X}} \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right) = \mathbf{Q}$, então $\lim_{n \rightarrow \infty} \text{Var}(\bar{\mathbf{w}}) = \mathbf{0}$ e $\text{plim} \frac{\mathbf{X}' \boldsymbol{\varepsilon}}{n} = \mathbf{0} \Rightarrow$

$$\text{plim} \mathbf{b} = \boldsymbol{\beta} + \mathbf{Q}^{-1} \mathbf{0} = \boldsymbol{\beta}.$$

Assim, \mathbf{b} é um estimador consistente de $\boldsymbol{\beta}$ no modelo de regressão linear.

Condições de Grenander

- HG.1:** Para cada coluna de \mathbf{X} , \mathbf{x}_k , se $d_{nk}^2 = \mathbf{x}_k' \mathbf{x}_k$, então $\lim_{n \rightarrow \infty} d_{nk}^2 = +\infty$. Assim, \mathbf{x}_k não degenera para uma sequência de zeros. A soma de quadrados continuará a crescer com o tamanho da amostra.
- HG.2:** $\lim_{n \rightarrow \infty} x_{ik}^2 / d_{nk}^2 = 0$ para todo $i = 1, \dots, n$. Esta condição implica que nenhuma observação dominará $\mathbf{x}_k' \mathbf{x}_k$ e quando $n \rightarrow \infty$, observações individuais serão menos importantes.
- HG.3:** Seja \mathbf{R}_n a correlação amostral das colunas de \mathbf{X} , excluindo a variável constante. Então, $\lim_{n \rightarrow \infty} \mathbf{R}_n = \mathbf{C}$, uma matriz positiva definida. Isto implica que a condição de posto completo sempre será satisfeita. Nós já assumimos que \mathbf{X} tem posto completo para uma amostra finita, então esta hipótese assegura que a condição nunca será violada.

Estas condições garantem que a matriz de dados são “bem comportadas” em amostras grandes mesmo nos casos envolvendo séries temporais com tendência temporal, séries polinômiais e variáveis de tendência (as hipóteses anteriores eram muito restritivas).

Normalidade Assintótica de \mathbf{b}

Utilizaremos a [HP.3](#) (exogeneidade) e adicionaremos a hipótese que as observações são independentes.

Sabemos que

$$\begin{aligned}\sqrt{n}(\mathbf{b} - \boldsymbol{\beta}) &= \left(\frac{\mathbf{X}'\mathbf{X}}{n} \right)^{-1} \left(\frac{1}{\sqrt{n}} \right) \mathbf{X}'\boldsymbol{\varepsilon} \Rightarrow \\ \left[\text{plim} \left(\frac{\mathbf{X}'\mathbf{X}}{n} \right)^{-1} \right] \left(\frac{1}{\sqrt{n}} \right) \mathbf{X}'\boldsymbol{\varepsilon} &= \mathbf{Q}^{-1} \left(\frac{1}{\sqrt{n}} \right) \mathbf{X}'\boldsymbol{\varepsilon}.\end{aligned}$$

Mas

$$\left(\frac{1}{\sqrt{n}} \right) \mathbf{X}'\boldsymbol{\varepsilon} = \sqrt{n}(\bar{\mathbf{w}} - \mathbf{E}(\bar{\mathbf{w}})) \quad \text{sendo} \quad \mathbf{E}(\bar{\mathbf{w}}) = \mathbf{0}.$$

Podemos utilizar a versão multivariada do teorema central do limite de Lindeberg-Feller para obter a distribuição limite de $\sqrt{n}\bar{\mathbf{w}}$.

Mas $\bar{\mathbf{w}}$ é a média de n vetores aleatórios independentes $\mathbf{w}_i = \mathbf{x}_i\varepsilon_i$ com média $\mathbf{0}$ e variância $\text{Var}(\mathbf{w}_i) = \text{Var}(\mathbf{x}_i\varepsilon_i) = \sigma^2\text{E}(\mathbf{x}_i\mathbf{x}_i') = \sigma^2\mathbf{Q}_i$. Segue-se que a variância de $\sqrt{n}\bar{\mathbf{w}}$ é

$$\sigma^2\bar{\mathbf{Q}}_n = \sigma^2 \left(\frac{1}{n} \right) [\mathbf{Q}_1 + \mathbf{Q}_2 + \cdots + \mathbf{Q}_n].$$

Se a soma não for dominada por nenhum termo e se os regressores forem bem comportados,

$$\lim_{n \rightarrow \infty} \sigma^2\bar{\mathbf{Q}}_n = \sigma^2\mathbf{Q}.$$

Portanto, se $[\mathbf{x}_i\varepsilon_i]$, $i = 1, \dots, n$, forem vetores independentes distribuídos com média $\mathbf{0}$ e variância $\sigma^2\mathbf{Q}_i$ e se $\text{plim} \frac{\mathbf{X}'\mathbf{X}}{n} = \mathbf{Q}$, então

$$\frac{1}{\sqrt{n}}\mathbf{X}'\varepsilon \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{Q}).$$

Segue-se que

$$\mathbf{Q}^{-1} \left(\frac{1}{\sqrt{n}} \right) \mathbf{X}' \varepsilon \xrightarrow{d} \mathcal{N}(\mathbf{Q}^{-1} \mathbf{0}, \mathbf{Q}^{-1} \sigma^2 \mathbf{Q} \mathbf{Q}^{-1}) \Rightarrow \sqrt{n}(\mathbf{b} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}^{-1}).$$

Teorema:

Se $\{\varepsilon_i\}$ são independentemente distribuídos com média 0 e variância σ^2 e x_{ik} é tal que as condições de Grenander são satisfeitas, então

$$\mathbf{b} \stackrel{a}{\sim} \mathcal{N} \left(\boldsymbol{\beta}, \frac{\sigma^2}{n} \mathbf{Q}^{-1} \right).$$

Na prática, é necessário estimar $(1/n)\mathbf{Q}^{-1}$ com $(\mathbf{X}'\mathbf{X})^{-1}$ e σ^2 com $\mathbf{e}'\mathbf{e}/(n-p)$.

Consistência de s^2

Temos a variância assintótica de \mathbf{b} : $(\sigma^2/n)\mathbf{Q}^{-1}$.

Podemos restringir a atenção a s^2 como estimador consistente de σ^2 .

$$\begin{aligned}s^2 &= \frac{1}{n-p} \boldsymbol{\varepsilon}' \mathbf{M} \boldsymbol{\varepsilon} = \frac{1}{n-p} \left[\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\varepsilon} \right] \\ &= \frac{n}{n-p} \left[\frac{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{n} - \left(\frac{\boldsymbol{\varepsilon}' \mathbf{X}}{n} \right) \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right)^{-1} \left(\frac{\mathbf{X}' \boldsymbol{\varepsilon}}{n} \right) \right].\end{aligned}$$

Mas

$$\text{plim} \left(\frac{\boldsymbol{\varepsilon}' \mathbf{X}}{n} \right) \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right)^{-1} \left(\frac{\mathbf{X}' \boldsymbol{\varepsilon}}{n} \right) = 0, \quad \text{pois}$$

$$\text{plim} \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right)^{-1} = \mathbf{Q}^{-1} \quad \text{e} \quad \text{plim} \left(\frac{\mathbf{X}' \boldsymbol{\varepsilon}}{n} \right) = 0.$$

Agora, vamos analisar

$$\overline{\boldsymbol{\varepsilon}^2} = \frac{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{n} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2.$$

Neste caso, $E(\varepsilon_i^2) = \sigma^2$. Pelo teorema de Markov, precisamos de $E|\varepsilon_i^2|^{1+\delta} < \infty$, $\delta > 0$, tal que a hipótese mínima até agora é que ε_i tenha momentos finitos um pouco maior que 2.

Além disso, se assumirmos que cada ε_i tem a mesma distribuição, então pelo teorema de Khinchine, momentos finitos até 2 é suficiente.

Convergência em média quadrática exigiria que $E(\varepsilon_i^4) = \phi_\varepsilon < \infty$.

Então, os termos da soma são independentes, com média σ^2 e variância $\phi_\varepsilon - \sigma^4$. Logo, sob condições fracas,

$$\text{plim} \frac{\mathbf{e}'\mathbf{e}}{n-p} = \text{plim} \frac{\varepsilon' \mathbf{M} \varepsilon}{n-p} = \text{plim} s_n^2 = \sigma^2 \Rightarrow \text{plim} s_n^2 \left(\frac{\mathbf{X}'\mathbf{X}}{n} \right)^{-1} = \sigma^2 \mathbf{Q}^{-1}.$$

O estimador apropriado da matriz de covariância assintótica de \mathbf{b} é

$$\mathbf{S}_b = s^2 (\mathbf{X}'\mathbf{X})^{-1}.$$

Distribuição Assintótica de uma Função de \mathbf{b}

Seja $\mathbf{f}(\mathbf{b})$ um conjunto de J funções contínuas, e continuamente diferenciáveis do estimador de mínimos quadrados, e seja

$$\mathbf{C}(\mathbf{b}) = \frac{\partial \mathbf{f}(\mathbf{b})}{\partial \mathbf{b}'},$$

sendo \mathbf{C} uma matriz $J \times p$ cuja j -ésima linha é o vetor de derivadas da j -ésima função com respeito a \mathbf{b}' .

Pelo teorema de Slutsky,

$$\text{plim} \mathbf{f}(\mathbf{b}) = \mathbf{f}(\boldsymbol{\beta}) \quad \text{e} \quad \text{plim} \mathbf{C}(\mathbf{b}) = \frac{\partial \mathbf{f}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} = \boldsymbol{\Gamma}.$$

Teorema

Se $\mathbf{f}(\mathbf{b})$ é um conjunto de funções contínuas e continuamente diferenciável de \mathbf{b} tal que $\boldsymbol{\Gamma} = \frac{\partial \mathbf{f}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}$, e se $\mathbf{b} \stackrel{a}{\sim} \mathcal{N}\left(\boldsymbol{\beta}, \frac{\sigma^2}{n} \mathbf{Q}^{-1}\right)$, então

$$\mathbf{f}(\mathbf{b}) \stackrel{a}{\sim} \mathcal{N}\left(\mathbf{f}(\boldsymbol{\beta}), \boldsymbol{\Gamma} \left(\frac{\sigma^2}{n} \mathbf{Q}^{-1}\right) \boldsymbol{\Gamma}'\right).$$

A prova do teorema utiliza a aproximação de Taylor:

$$\mathbf{f}(\mathbf{b}) = \mathbf{f}(\boldsymbol{\beta}) + \boldsymbol{\Gamma}(\mathbf{b} - \boldsymbol{\beta}) + \text{outros termos.}$$

Na prática, estimaremos a variância de $\mathbf{f}(\mathbf{b})$ por $\mathbf{S}_{f(\mathbf{b})} = s^2 \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{C}'$.

Exemplo 9: Demanda por Gasolina

$$\begin{aligned} \ln(\text{G/Pop})_t &= \beta_1 + \gamma \ln(\text{G/Pop})_{t-1} + \beta_2 \ln \text{Preço}_t + \beta_3 \ln(\text{Renda/Pop})_t \\ &+ \beta_4 \ln(\text{PCN})_t + \beta_5 \ln(\text{PCU})_t + \varepsilon_t, \end{aligned}$$

sendo PCN o preço dos carros novos, PCU o preço dos carros usados e G o consumo de gasolina (gasto total (GasExp) dividido pelo preço (GasP)).

- ▶ A elasticidade a longo prazo do preço é $\epsilon_2 = \frac{\beta_2}{1 - \gamma}$.
- ▶ A elasticidade a longo prazo da renda é $\epsilon_3 = \frac{\beta_3}{1 - \gamma}$.

Utilizaremos mínimos quadrados para estimar os parâmetros do modelo.

(Mostrar exemplo no R: exemplo_09.r pag. 75-77)

Sejam $\mathbf{x}_t = (1, \ln \text{Preço}_t, \ln(\text{Renda/Pop})_t, \ln(\text{PCN})_t, \ln(\text{PCU})_t)'$,
 $y_t = \ln(\text{G/Pop})_t$ e $\beta = (\beta_1, \dots, \beta_5)'$.

Então podemos reescrever o modelo como:

$$\begin{aligned} y_t &= \mathbf{x}_t' \beta + \gamma y_{t-1} + \varepsilon_t \\ y_t - \frac{1}{1-\gamma} \mathbf{x}_t' \beta &= \gamma \left(y_{t-1} - \frac{1}{1-\gamma} \mathbf{x}_{t-1}' \beta \right) + \varepsilon_t \\ y_t - \mu_t &= \gamma (y_{t-1} - \mu_{t-1}) + \varepsilon_t \\ z_t &= \gamma z_{t-1} + \varepsilon_t, \end{aligned}$$

sendo $z_t = y_t - \mu_t$, $z_{t-1} = y_{t-1} - \mu_{t-1}$ e $\mu_t = \frac{1}{1-\gamma} \mathbf{x}_t' \beta$.

Temos então um modelo autoregressivo de ordem 1 (AR(1)) estacionário (por hipótese pois estaremos removendo todas as tendências com os regressores em \mathbf{x}_t).

Da estacionariedade do modelo AR(1), temos que

$E(z_t) = \gamma E(z_{t-1}) + E(\varepsilon) \Rightarrow E(z_t) = 0 \Rightarrow E(y_t) = \mu_t$ (média de longo prazo).

Agora, podemos calcular as elasticidade de longo prazo do consumo em relação ao preço ou a renda.

A elasticidade a longo prazo do consumo em relação ao preço:

$$\epsilon_2 = \frac{\partial \ln(G/Pop)_t}{\partial \ln \text{Preço}_t} = \frac{\partial \mu_t}{\partial \ln \text{Preço}_t} = \frac{\beta_2}{1 - \gamma}.$$

A elasticidade a longo prazo do consumo em relação a renda:

$$\epsilon_3 = \frac{\partial \ln(G/Pop)_t}{\partial \ln(\text{Renda}/Pop)_t} = \frac{\partial \mu_t}{\partial \ln(\text{Renda}/Pop)_t} = \frac{\beta_3}{1 - \gamma}.$$

Seja $\theta = (\beta_1, \gamma, \beta_2, \dots, \beta_5)'$. Então,

$$\begin{aligned} \frac{\partial \epsilon_2(\theta)}{\partial \theta} &= \left(0, \frac{\beta_2}{(1 - \gamma)^2}, \frac{1}{1 - \gamma}, 0, 0, 0 \right)' \\ \frac{\partial \epsilon_3(\theta)}{\partial \theta} &= \left(0, \frac{\beta_3}{(1 - \gamma)^2}, 0, \frac{1}{1 - \gamma}, 0, 0 \right)'. \end{aligned}$$

Estimação por Máxima Verossimilhança

Suponha **HP.6** seja verdade. Então, o estimador dos coeficientes da regressão por mínimos quadrados coincide com o estimador dos coeficientes por máxima verossimilhança.

O logaritmo da função de verossimilhança baseada em uma amostra aleatória de tamanho n com erros normais é dada por

$$\ell(\beta, \sigma^2) \triangleq \ln L(\beta, \sigma^2) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta).$$

O estimador de máxima verossimilhança de (β, σ^2) é dado por

$$\begin{aligned} \mathbf{b}_{EMV} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}, \quad \text{e} \\ \hat{\sigma}_{EMV}^2 &= \frac{1}{n} (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) = \frac{1}{n} \mathbf{e}'\mathbf{e}. \end{aligned}$$

Por ser um estimador de máxima verossimilhança, \mathbf{b}_{EMV} é assintoticamente eficiente entre os estimadores consistentes, e assintoticamente normal.

Veremos Estimadores de Máxima Verossimilhança mais adiante.

Estimação por Intervalo

Vamos começar utilizando todas as hipóteses do modelo de regressão linear, inclusive [HP.6](#).

Dado $\varepsilon \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{I})$, temos que

$$\begin{aligned}(\mathbf{b}|\mathbf{X}) &\sim \mathcal{N}_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}), \quad \text{e} \\(b_k|\mathbf{X}) &\sim \mathcal{N}(\beta_k, \sigma^2(\mathbf{X}'\mathbf{X})_{kk}^{-1}).\end{aligned}$$

Definamos s_{kk} o k -ésimo elemento da diagonal de $(\mathbf{X}'\mathbf{X})^{-1}$.

Daí, temos

$$Z_k = \frac{b_k - \beta_k}{\sqrt{\sigma^2 s_{kk}}} \sim \mathcal{N}(0, 1).$$

Mas σ^2 é desconhecido.

Resultado 1: O posto de uma matriz simétrica idempotente é igual ao seu traço.

Resultado 2: Se $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ e \mathbf{C} é uma matriz tal que $\mathbf{C}'\mathbf{C} = \mathbf{I}$, então $\mathbf{C}'\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

Resultado 3: Se $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ e \mathbf{A} é uma matriz idempotente, então $\mathbf{x}'\mathbf{A}\mathbf{x} \sim \chi_m^2$ sendo m o número de raízes unitárias de \mathbf{A} que é igual ao posto de \mathbf{A} .

Resultado 4: Se $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\mathbf{L}\mathbf{x}$ uma função linear e $\mathbf{x}'\mathbf{A}\mathbf{x}$ uma forma quadrática idempotente, então $\mathbf{L}\mathbf{x}$ e $\mathbf{x}'\mathbf{A}\mathbf{x}$ são independentes se $\mathbf{L}\mathbf{A} = \mathbf{0}$.

A quantidade

$$\frac{(n-p)S^2}{\sigma^2} = \frac{\mathbf{e}'\mathbf{e}}{\sigma^2} = \left(\frac{\varepsilon}{\sigma}\right)' \mathbf{M} \left(\frac{\varepsilon}{\sigma}\right)$$

é uma forma quadrática idempotente do vetor normal (ε/σ) com $\text{tr}(\mathbf{M}) = n - p$ e a forma linear dada por

$$\left(\frac{\mathbf{b} - \boldsymbol{\beta}}{\sigma}\right) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\frac{\varepsilon}{\sigma}\right).$$

Temos que $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{M} = \mathbf{0}$ pois $\mathbf{M}\mathbf{X} = \mathbf{0}$. Daí, temos que a independência.

Teorema - Independência de \mathbf{b} e s^2 : *Se ε é um vetor com distribuição normal, então o estimador \mathbf{b} de mínimos quadrados dos coeficientes é estatisticamente independente do vetor de resíduos \mathbf{e} e portanto, todas as funções de \mathbf{e} , incluindo s^2 .*

Destes resultados, temos que

$$\frac{b_k - \beta_k}{\sqrt{s^2 s_{kk}}} \sim t_{n-p}.$$

Consequentemente,

$$IC_{100(1-\alpha)\%}(\beta_k) = (b_k - t_{n-p, 1-\alpha/2} \sqrt{s^2 s_{kk}}, b_k + t_{n-p, 1-\alpha/2} \sqrt{s^2 s_{kk}}).$$

Exemplo 9 (continuação): Demanda por Gasolina

$$\ln(G/Pop)_t = \beta_1 + \beta_2 \ln \text{Preço}_t + \beta_3 \ln(\text{Renda}/Pop)_t + \beta_4 \ln(PCN)_t + \beta_5 \ln(PCU)_t + \varepsilon_t.$$

(Mostrar exemplo no R: exemplo_09.r pag. 82)

Tabela : Intervalos de Confiança de 95%.

Variável	2,5%	97,5%
Constante	-22,7264	-19,6958
ln Preço	-0,1093	0,0668
ln Renda/Pop	0,9395	1,2522
ln PCN	-0,6896	-0,0576
ln PCU	-0,1878	0,2279

Os intervalos de confiança para ϵ_2 e ϵ_3 baseados em teoria assintótica:

$$IC_{95\%}(\epsilon_2) = (-0,7099; -0,1129) \quad \text{e} \quad IC_{95\%}(\epsilon_3) = (0,6523; 1,2888).$$

Combinação Linear dos Coeficientes

- ▶ Seja \mathbf{w} um vetor $p \times 1$ de constantes.
- ▶ Estamos interessados na combinação linear $\alpha \triangleq \mathbf{w}'\boldsymbol{\beta}$.
- ▶ Estimamos α com $a \triangleq \mathbf{w}'\mathbf{b}$ sendo \mathbf{b} o estimador de mínimos quadrados.
- ▶ Como \mathbf{b} tem distribuição normal, temos

$$a \sim \mathcal{N}(\mathbf{w}'\boldsymbol{\beta}, \sigma^2 \mathbf{w}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{w}).$$

- ▶ Ainda estimamos σ^2 com s^2 .
- ▶ Estimamos a variância de a com $s_a^2 \triangleq s^2 \mathbf{w}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{w}$.
- ▶ Segue-se que $(a - \alpha)/s_a \sim t_{n-p}$ (pois a e s_a^2 são independentes).
- ▶ Daí podemos construir intervalos de confiança para α .
- ▶ Podemos testar somas ou diferenças entre os coeficientes.

Predição

- ▶ Suponha que desejamos prever o valor de y^0 associado ao vetor de regressores \mathbf{x}^0 .
- ▶ Este valor seria $y^0 = \mathbf{x}^{0'}\boldsymbol{\beta} + \varepsilon^0$.
- ▶ Segue do teorema de Gauss-Markov que $\hat{y}^0 = \mathbf{x}^{0'}\mathbf{b}$ é o estimador não tendencioso linear de variância mínima de $E(y^0|\mathbf{x}^0)$.
- ▶ O erro de previsão é $e^0 = y^0 - \hat{y}^0 = (\boldsymbol{\beta} - \mathbf{b})'\mathbf{x}^0 + \varepsilon^0$.
- ▶ A variância de predição para este estimador é

$$\text{Var}(e^0|\mathbf{X}, \mathbf{x}^0) = \sigma^2 + \text{Var}((\boldsymbol{\beta} - \mathbf{b})'\mathbf{x}^0|\mathbf{X}, \mathbf{x}^0) = \sigma^2(1 + \mathbf{x}^{0'}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^0).$$

- ▶ Podemos estimar esta variância ao substituir σ^2 pelo estimador s^2 .
- ▶ Assim, um intervalo de predição de y^0 é dada por

$$\text{IC}_{100(1-\alpha)\%}(y^0) = (\hat{y}^0 - t_{\alpha/2; n-p} \times ep(e^0), \hat{y}^0 + t_{\alpha/2; n-p} \times ep(e^0)).$$

Exemplo 10: (Mostrar exemplo no R: exemplo_10.r pag. 85)

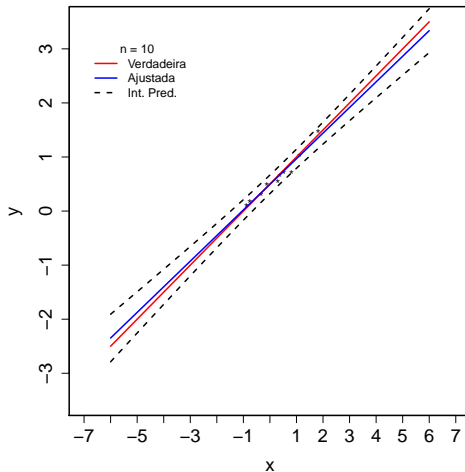


Figura : Intervalo de predição.

Multicolinearidade

Quando os regressores são altamente correlacionados, temos que

- ▶ pequenas mudanças nos dados podem produzir grandes variações na estimativas dos parâmetros;
- ▶ coeficientes podem ter erros padrões grandes e baixos níveis de significância mesmo que eles sejam conjuntamente significantes e o R^2 para a regressão seja elevado; e
- ▶ coeficientes podem ter um sinal equivocado ou magnitudes implausíveis.

Seja \mathbf{X} a matriz de variáveis regressoras com uma constante e $p - 1$ variáveis medidas como desvios de suas médias, $\mathbf{x}_k = \mathbf{z}_k - \bar{z}_k \mathbf{1}$.

Seja $\mathbf{X}_{(k)}$ a matriz com todas as variáveis exceto \mathbf{x}_k . Então, o k -ésimo elemento de $(\mathbf{X}'\mathbf{X})^{-1}$ é dado por

$$\begin{aligned} (\mathbf{x}'_k \mathbf{M}_{(k)} \mathbf{x}_k)^{-1} &= \left[\mathbf{x}'_k \mathbf{x}_k - \mathbf{x}'_k \mathbf{X}_{(k)} (\mathbf{X}'_{(k)} \mathbf{X}_{(k)})^{-1} \mathbf{X}'_{(k)} \mathbf{x}_k \right]^{-1} \\ &= \left[\mathbf{x}'_k \mathbf{x}_k \left(1 - \frac{\mathbf{x}'_k \mathbf{X}_{(k)} (\mathbf{X}'_{(k)} \mathbf{X}_{(k)})^{-1} \mathbf{X}'_{(k)} \mathbf{x}_k}{\mathbf{x}'_k \mathbf{x}_k} \right) \right]^{-1} = \frac{1}{(1 - R_k^2) s_{kk}}, \end{aligned}$$

sendo R_k^2 o coeficiente de determinação da regressão de x_k sobre todas as outras variáveis do modelo e $s_{kk} = \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2$.

A variância do estimador de mínimos quadrados do k -ésimo coeficiente é σ^2 vezes a razão acima,

$$\text{Var}(b_k | \mathbf{X}) = \frac{\sigma^2}{(1 - R_k^2) \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}.$$

Com as outras coisas iguais,

- ▶ quanto maior a correlação de x_k com as outras variáveis, maior será a variância de b_k , devido a multicolinearidade;
- ▶ quanto maior a variância de x_k , menor será a variância de b_k ; e
- ▶ quanto melhor o ajuste geral da regressão, menor será a variância devido a σ^2 .

Dados reais nunca serão ortogonais ($R_k^2 = 0$). Então, a multicolinearidade estará sempre presente.

Quando a multicolinearidade é um problema?

Podemos utilizar o **número de condição** de $\mathbf{X}'\mathbf{X}$: a raiz quadrada da razão da maior raiz característica de $\mathbf{X}'\mathbf{X}$ (sendo as colunas padronizadas para tamanho unitário) sobre a menor. Valores maiores que 20 sugerem uma indicação de problema.

Possíveis caminhos são

- ▶ aumentar o número de observações;
- ▶ retirar as variáveis mais problemáticas do modelo; e
- ▶ utilizar componentes principais. (Dificuldade de interpretação.)

Exemplo 11: Dados de Longley.

(Mostrar exemplo no R: exemplo_11.r pag. 89-90)

Tabela : Correlação entre as variáveis.

	Emprego	Preço	PIB	Militar	Ano
Emprego	1,0000	0,9709	0,9836	0,4573	0,9713
Preço	0,9709	1,0000	0,9916	0,4647	0,9911
PIB	0,9836	0,9916	1,0000	0,4464	0,9953
Militar	0,4573	0,4647	0,4464	1,0000	0,4172
Ano	0,9713	0,9911	0,9953	0,4172	1,0000

Tabela : Ajuste do modelo com 15 observações.

	Coeficiente	Erro padrão	Estat. t	Valor p
Constante	1459415,1	714182,9	2,04348	0,06825
Ano	-721,756	369,985	-1,95077	0,07965
Preço	-181,12296	135,5251	-1,33646	0,21101
PIB	0,09107	0,02026	4,49478	0,00115
Militar	-0,07494	0,26113	-0,28698	0,77999

Tabela : Ajuste do modelo com 16 observações.

	Coeficiente	Erro padrão	Estat. t	Valor p
Constante	1169087,5	835902,4	1,39859	0,18949
Ano	-576,464	433,487	-1,32983	0,21049
Preço	-19,76807	138,893	-0,14233	0,88940
PIB	0,06439	0,01995	3,22746	0,00805
Militar	-0,01015	0,30857	-0,03288	0,97436