Departamento de Ciências Florestais
ESALQ - USP

Análise de Regressão Aplicada

João L. F. Batista

Piracicaba
– 2004 –

“Twice two equals four: ’tis true,
But too empty, too trite.
What I look for is a clue
To some matters not so light.”

W. Busch, 1909

SUMÁRIO

CAPÍTULO 1
MODELOS LINEARES SIMPLES

"We must be careful not to confuse data with the abstractions we use to analyze them."

William James (1842-1910)

1.1 Relação entre Duas Variáveis

A questão de interesse envolve a relação entre duas ou mais variáveis quantitativas.

Estamos sempre interessados em que?

"descrições"

"explicações"

"predições"

"controle"

Que formas de relação são aceitas na nossa área de pesquisa?

Relação Funcional: é uma relação matemática.

- Área do Círculo = $\pi \times [\text{raio}]^2$
- Salário de Trabalhador temporário = = Custo por hora $\times$ Número de Horas Trabalhadas
- Energia = massa $\times$ [velocidade da luz]^2

Relação Estatística: é uma relação "*estocástica*".

- Peso do peixe = $f(\text{comprimento do peixe})$
- Altura da árvore = $f(\text{diâmetro da árvore})$
- Volume de Run-off = $f(\text{Volume de água na chuva})$

A relação estatística:

- *Não é "perfeita".*
- *É "variável" ao redor de uma "relação média" (relação funcional).*

1.2 Estrutura Geral dos Modelos Lineares Simples

Trataremos de uma classe de modelos estatísticos que chamamos de "*Modelo Linear Simples*".

Os modelos lineares simples têm algumas características importantes:

1. A *variável resposta* (Y = "dependent variable") varia de acordo com a *variável preditora* (X = "independent variable") de **modo sistemático**.

   $\Longrightarrow$ FUNÇÃO DE REGRESSÃO

2. Existe um efeito estocástico que causa uma **variação** ou **dispersão** das observações ao redor da porção sistemática da modelo.

   $\Longrightarrow$ ERRO ALEATÓRIO (ε)

   O que gera esse *efeito estocástico*?

3. A relação entre variável resposta (Y), variável preditora (X) e erro aleatório (ε) é sempre **aditiva**.

   $\Longrightarrow$ MODELO LINEAR NOS PARÂMETROS

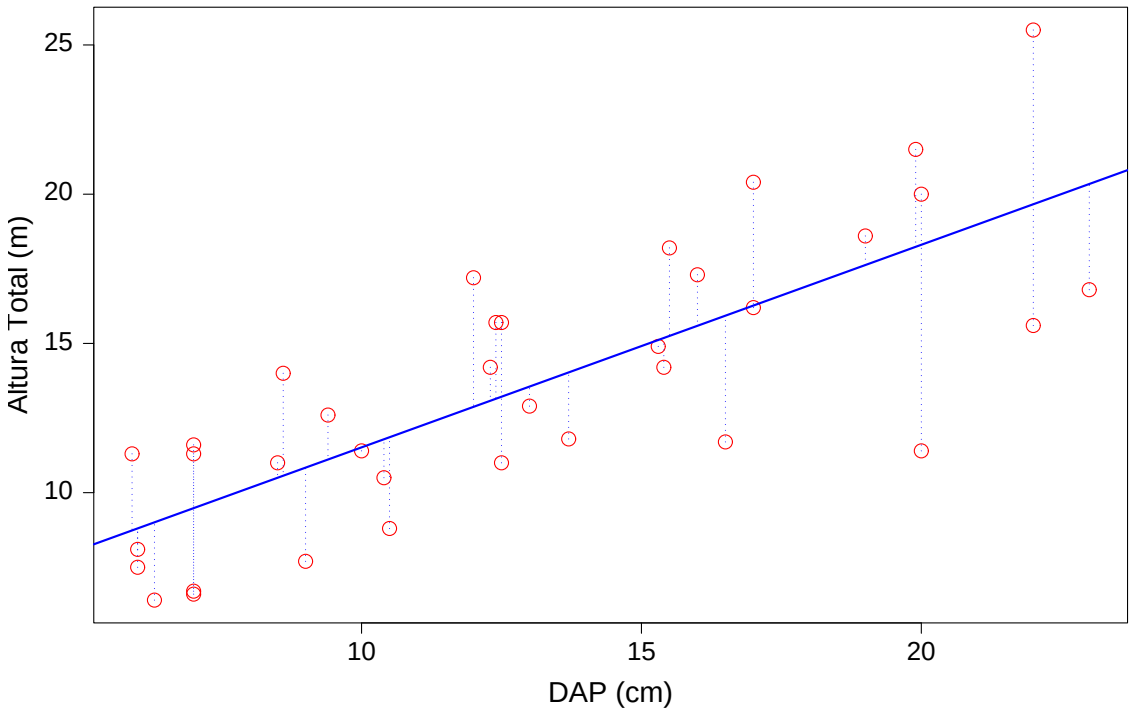


Figura 1.1: Relação entre DAP (diâmetro à altura do peito) e altura total de árvores de *Eucalyptus saligna*. Floresta com idade desconhecida que sofreu várias colheitas sem manejo adequado, região de Itatinga, SP.

1.2.1 Estrutura Formal dos Modelos Lineares Simples

Utilizaremos com freqüência definições formais dos modelos nesse curso, pois as características dos modelos lineraes e as distinção entre os modelos são mais claramente compreendidas através da definição formal.

No caso da estrutura geral dos modelos lineares a definição é:

1. A relação de aditividade é definida pela expressão da **reta**:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

$\Longrightarrow$ Y é a variável resposta que se assume como **variável estocástica** ou **variável aleatória**;

$\Longrightarrow$ X é a variável preditora que se assume como **variável matemática**;

$\Longrightarrow$ ε é o erro aleatório, sendo, portanto, também uma **variável aleatória**;

$\Longrightarrow$ β_0 e β_1 são **constantes matemáticas**, geralmente chamadas de parâmetros ou coeficientes de regressão.

Essa expressão implica que a variável aleatória Y é uma função linear da variável aleatória ε.

2. Para cada nível de X, Y segue uma **distribuição estatística**, pois é uma variável aleatória.

   Ainda mais formalmente: **condicional** a $X = x$, Y segue uma distribuição estatística.

3. Os **valores esperados** dessas distribuições variam sistematicamente com X.

CONCEITOS BÁSICOS DE ESTATÍSTICA

Para entendermos a definição acima é necessária que estejam claros alguns conceitos básicos de estatística:

- variável matemática versus variável aleatória;
- variável aleatória e distribuição estatística;
- transofrmação linear de variável aleatória;
- valor esperado de variável aleatória.

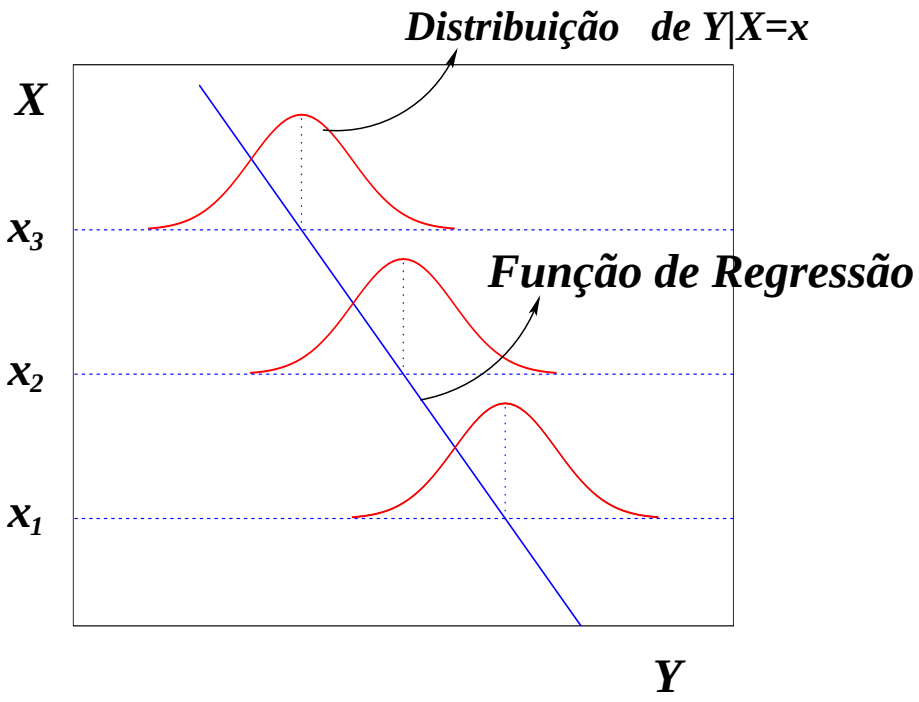


Figura 1.2: Esquema da definição formal dos modelos lineares simples. Note que a posição dos eixos X e Y está invertida para enfatizar o aspecto de Y ser uma variável aleatória.

1.2.2 Implicações da Estrutura dos Modelos Lineares Simples

A definição formal da estrtura geral dos modelos lineares simples deixa clara algumas implicações desses modelos.

O fato desses modelos serem representados geometricamente por uma reta tem algumas implicações matemáticas:

1. Implica que os modelos são **lineares nos parâmetros**, i.e., a alteração de um coeficiente de regressão não afeta o outro.

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad \Rightarrow \quad \begin{cases} \partial Y/\partial \beta_0 = 1 \\ \\ \partial Y/\partial \beta_1 = X \end{cases}$$

   Quando a alteração num coeficiente afeta outros coeficientes de um modelo, dissemos que o modelo é **não-linear**.

Os modelos matemáticos com fundamentação teórica na nárea de recursos naturais são, em geral, não-lineraes.

- Modelo de Potência: relação entre variáveis biométricas de organismos individuais como, por exemplo, massa e volume:
$$Y = \beta_0 \, X^{\beta_1} + \varepsilon$$
- Modelo Chapman-Richards: relação entre tamanho e idade, tanto para organismos (árvores), como para populações (florestas):
$$Y = \beta_0 \left[1 - \exp(\beta_1 X^{\beta_2})\right]^{\beta_3} + \varepsilon$$

2. Implica que os modelos são **lineares na variável preditora**, i.e., são *modelos de primeira ordem*.
$$Y = \beta_0 + \beta_1 X + \varepsilon \quad \Rightarrow \quad \frac{\partial Y}{\partial X} = \beta_1$$
Os modelos polinomiais são lineares nos parâmetros, mas não são lineares na variável preditora:

Segunda Ordem: $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon \quad \Rightarrow \quad \frac{\partial Y}{\partial X} = \beta_1 + 2\beta_2 X$

Terceira Ordem: $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon \quad \Rightarrow$
$$\Rightarrow \quad \frac{\partial Y}{\partial X} = \beta_1 + 2\beta_2 X + 3\beta_3 X^2$$

3. Os coeficientes de regressão têm interpretação matemática e teórica clara:
   - β_0 é o **intercepto** da reta, i.e., o ponto em que a reta cruza o eixo das ordenadas (Y).
     Pode ter interpretação teórica ou não.
   - β_1 é o **coeficiente de inclinação** da reta.
     Ele significa que uma alteração unitária em X, resultará numa alteração de β_1 em Y.

O componente aleatório desses modelos gera implicações estatísticas, da qual a principal é:

Relação entre Y e ε: a distribuição da variável resposta e do erro aleatório é a mesma, para em cada nível de variável preditora ($X = x$),
$$\begin{aligned}(Y|X = x) &= \beta_0 + \beta_1 x + \varepsilon \\ (Y|X = x) &= k + \varepsilon\end{aligned}$$

1.3 O Modelo Linear Simples de Erros Aleatórios (MLEA)

Uma vez que a estrutura geral dos modelos lineares simples foi apresentada, podemos apresentar uma série de modelos de modo a explorar progressivamente a estrutura, ajuste e aplicação dos modelos lineares.

1.3.1 Estrutura do MLEA

A definição formal do MLEA é:

$$\begin{aligned} \text{Na população:} \quad & Y = \beta_0 + \beta_1 X + \varepsilon \\ \text{Na amostra:} \quad & y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \end{aligned}$$

onde

- i - índice das obseravações numa amostra de tamanho n: $i = 1, 2, \ldots, n$;
- y_i - valores observados da variável resposta (Y - variável aleatória);
- x_i - valores observados da variável preditora (X - variável matemática);
- β_0 e β_1 - coeficientes de regressão (constantes) a serem estimados;
- ε_i - erro aleatório (variável aleatória) associado à $i^{\text{ésima}}$ observação, do qual assumiremos (inicialmente) apenas uma propriedade: $\boldsymbol{E}\{\varepsilon_i | X = x_i\} = 0$.

Como essa definição formal será repetida muitas vêzes (com acréscimos), utilizaremos uma notação mais sintética para apresentá-la:

(1.1) MLEA: $Y = \beta_0 + \beta_1 X + \varepsilon,$ onde $\boldsymbol{E}\{\varepsilon | X = x\} = 0.$

1.3.2 Pressuposição

O MLEA tem **uma única pressuposição**: que, para cada nível da variável preditora ($X = x$), o valor esperado do erro aleatório é zero (média zero para o erro aleatório):

$$\boldsymbol{E}\{\varepsilon | X = x\} = 0.$$

Essa pressuposição define exatamente onde passa a função de regressão (reta de regressão).

Observando os valores esperados para Y em cada nível da X vemos que:

$$\begin{aligned} \boldsymbol{E}\{Y|X=x_i\} &= \boldsymbol{E}\{\beta_0+\beta_1 x_i+\varepsilon_i\} \\ &= \beta_0+\beta_1 x_i+\boldsymbol{E}\{\varepsilon_i|X=x_i\} \\ &= \beta_0+\beta_1 x_i \end{aligned}$$

O MLEA assume que as médias da variável resposta em cada nível da variável preditora estão alinhadas sobre a reta de regressão.

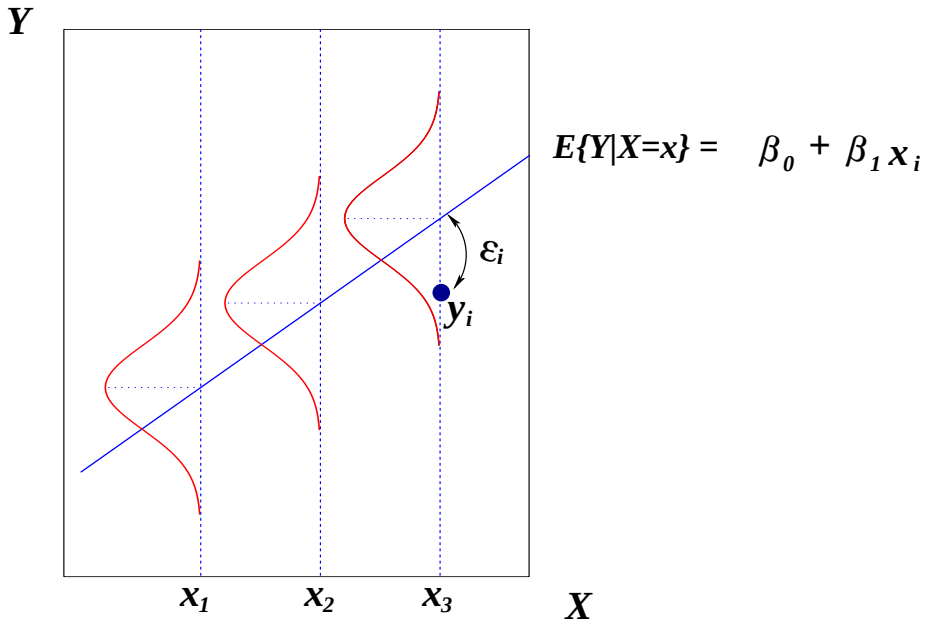


Figura 1.3: Gráfico demonstrando onde passa a reta de regressão do MLEA como decorrência da pressuposição $\boldsymbol{E}\{\varepsilon|X=x\}=0$.

1.3.3 Métodos de Ajuste

Valores de β_0 e β_1, sendo parâmetros populacionais, serão sempre desconhecidos.

A partir de uma amostra podemos **estimá-los**.

Quais critérios podemos utilizar para estimar os coeficientes de regressão?

- Algum conceito de **optimização** deve ser utilizado para ajustar a linha de regressão aos dados.
- Podemos definir "**perda**" de diversas formas:

  $\sum$[desvios]^2

  $\sum$[desvios absolutos]

  [desvio absoluto máximo]
- O MÉTODO DOS QUADRADOS MÍNIMOS define $\sum$[desvios]^2 como o critério para estimar β_0 e β_1.
- A utilização de $\sum$[desvios]^2 como a **função de perda** não é tão arbitrária como parece.
  - As estimativas obtidas por este critério têm propriedades estatísticas desejáveis.
  - O método faz uma poderação de modo a dar maior ênfase aos grandes desvios. A perda é maior para um grande desvio do que para uma série de pequenos desvios.

1.3.4 Estimadores de Quadrados Mínimos (EQM)

Para uma amostra de observações pareadas (y_i, x_i), $i = 1, 2, \ldots, n$, podemos definir os desvios como

$$e_i \quad = \quad y_i - (b_0 + b_1 x_i) \quad = \quad y_i - b_0 - b_1 x_i,$$

onde b_0 e b_1 são as estimativas de β_0 e β_1, respectivamente, que pretendemos encontrar.

O Método de Quadrados Mínimos minimiza, portanto, a função

$$Q = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

Q é uma função quadrática com relação às estimativas dos coeficientes de regressão (b_0 e b_1), assim podemos encontrar o seu **máximo** ou **mínimo** através das derivadas parciais:

$$\begin{aligned} \frac{\partial Q}{\partial b_0} &= \sum_{i=1}^{n} 2\,(y_i - b_0 - b_1 x_i)\,(-1) = -2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i) \\ \frac{\partial Q}{\partial b_1} &= \sum_{i=1}^{n} 2\,(y_i - b_0 - b_1 x_i)\,(-x_i) = -2 \sum_{i=1}^{n} x_i\,(y_i - b_0 - b_1 x_i) \end{aligned}$$

Igualando as derivadas parciais a **zero** e solucionado o sistema de equações resultante encontramos as estimativas:

$$\begin{array}{rl} -2\sum (y_i - b_0 - b_1 x_i) & = 0 \\ -2\sum x_i\,(y_i - b_0 - b_1 x_i) & = 0 \end{array}$$
$$\Downarrow$$
$$\begin{array}{rl} \sum (y_i - b_0 - b_1 x_i) & = 0 \\ \sum x_i\,(y_i - b_0 - b_1 x_i) & = 0 \end{array}$$
$$\Downarrow$$
$$\begin{array}{rl} \sum y_i - nb_0 - b_1 \sum x_i & = 0 \\ \sum x_i y_i - b_0 \sum x_i - b_1 \sum x_i^2 & = 0 \end{array}$$
$$\Downarrow$$

(1.2)

EQUAÇÕES NORMAIS

$$\begin{array}{lll} \sum y_i & = nb_0 & +b_1 \sum x_i \\ \sum x_i y_i & = b_0 \sum x_i & +b_1 \sum x_i^2 \end{array}$$

A soluação do Sistema de Equações Normais resulta em:

$$b_1 = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2} = \frac{S_{xy}}{S_{xx}} \quad (1.3)$$

$$b_0 = \frac{1}{n}\left(\sum y_i - b_1 \sum x_i\right) = \overline{y} - b_1 \overline{x} \quad (1.4)$$

onde:

S_{xy} é chamada de **Soma de Produtos de** X **por** Y

S_{xx} é chamada de **Soma de Quadrados de** X.

1.3.5 Propriedades dos EQM

CONCEITOS BÁSICOS DE ESTATÍSTICA

Para prosseguirmos necessitamos rever um conceito básico de estatística:

- propriedade de estimadores estatísticos: estimador não viciado.

EQM são funções lineares dos valores observados de Y.

Prova:

$$\begin{aligned} b_1 &= \frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{\sum(x_i-\overline{x})^2} = \sum \frac{(x_i-\overline{x})}{\sum(x_i-\overline{x})^2} y_i = \sum k_i y_i \\ b_0 &= \overline{y} - b_1\overline{x} = \sum y_i/n - \sum k_i y_i \overline{x} = \sum \left[1/n - \overline{x}k_i\right] y_i \\ &= \sum \left[\frac{1}{n} - \frac{\overline{x}(x_i-\overline{x})}{\sum(x_i-\overline{x})^2}\right] y_i = \sum k_i^* y_i \end{aligned}$$

EQM são estimadores não viciados dos coeficientes de regressão:

$$\boldsymbol{E}\{b_0\} = \beta_0 \quad \text{e} \quad \boldsymbol{E}\{b_1\} = \beta_1.$$

Prova:

$$\begin{aligned} \boldsymbol{E}\{b_1\} &= \boldsymbol{E}\left\{\sum\left[\frac{(x_i-\overline{x})}{\sum(x_i-\overline{x})^2}\right] y_i\right\} \\ &= \left(1/\sum(x_i-\overline{x})^2\right) \sum\left[(x_i-\overline{x})\right] \boldsymbol{E}\{y_i\} \\ &= \left(1/\sum(x_i-\overline{x})^2\right) \sum\left[(x_i-\overline{x})\right] (\beta_0+\beta_1 x_i) \\ &= \left(1/\sum(x_i-\overline{x})^2\right) \left[\sum x_i(\beta_0+\beta_1 x_i) - \sum \overline{x}(\beta_0+\beta_1 x_i)\right] \\ &= \left(1/\sum(x_i-\overline{x})^2\right) \left[\beta_0 n\overline{x} + \beta_1 \sum x_i^2 - \beta_0 n\overline{x} - \beta_1 \left(\sum x_i\right)^2/n\right] \\ &= \beta_1 \left(1/\sum(x_i-\overline{x})^2\right) \left[\sum x_i^2 - \left(\sum x_i\right)^2/n\right] \\ &= \beta_1 \end{aligned}$$

$$\begin{aligned}
\boldsymbol{E}\{b_0\} &= \boldsymbol{E}\{\overline{y} - b_1\overline{x}\} \\
&= \boldsymbol{E}\left\{\sum_{i=1}^{n} y_i/n\right\} - \overline{x}\beta_1 \\
&= (1/n)\boldsymbol{E}\left\{\sum_{i=1}^{n}(\beta_0 + \beta_1 x_i)\right\} - \beta_1\overline{x} \\
&= (1/n)\sum_{i=1}^{n}\left(\boldsymbol{E}\{\beta_0 + \beta_1 x_i\}\right) - \beta_1\overline{x} \\
&= (1/n)\left(n\beta_0 + \beta_1\sum_{i=1}^{n} x_i\right) - \beta_1\overline{x} \\
&= \beta_0 + \beta_1\overline{x} - \beta_1\overline{x} \\
&= \beta_0
\end{aligned}$$

1.3.6 Propriedades do Modelo Ajustado

Inicialmente devemos definir alguns termos novos:

Valor Observado: é o valor da variável resposta para uma dada obervação: y_i.

Valor Ajustado: é o valor que o modelo ajustado retorna para um dado nível da variável preditora:

$$\widehat{y}_i = b_0 + b_1 x_i$$

Resíduo: é a diferença do valor observado e o valor ajustado (**nessa ordem**):

$$e_i \quad = \quad y_i - \widehat{y}_i \quad = \quad y_i - b_0 - b_1 x_i$$

NOTA IMPORTANTE:

É importante fazer distinção entre o **resíduo** (e_i) e o **erro aleatório** (ε_i).

O resíduo se refere **sempre** ao modelo ajustado aplicado à sua amostra de ajuste.

O erro aleatório se refere ao modelo no âmbito da população:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad \Rightarrow \quad \varepsilon = Y - (\beta_0 + \beta_1 X) \quad \Rightarrow \quad \varepsilon = Y - \boldsymbol{E}\{Y\}$$

Vejamos agora as propriedades do modelos ajustado pelo método dos quadrados mínimos:

1. O valor ajustado é um estimador não-viciado do valor esperado para variável resposta:

$$\begin{aligned} \boldsymbol{E}\{\widehat{y}_i\} &= \boldsymbol{E}\{b_0 + b_1 x_i\} \\ &= \beta_0 + \beta_1 x_i \\ &= \boldsymbol{E}\{Y|X = x_i\} \end{aligned}$$

2. A soma dos resíduos é nula (igual a zero).

$$\sum e_i = 0$$

   Não leve a sérios softwares de regressão que apresentem a soma dos resíduos como critério de qualidade de ajuste de um modelo linear.

3. A soma dos resíduos ao quadrado é mínima.

   Os estimadores b_0 e b_1 foram encontrados de modo a minimizar:

$$Q = \sum (y_i - b_0 - b_1 x_i)^2 = \sum e_i^2.$$

4. A soma dos valores ajustados é igual a soma dos valores observados na amostra:

$$\sum \widehat{y}_i = \sum y_i.$$

5. Os resíduos são **independentes** dos valores da variável preditora:

$$\sum e_i \, x_i = 0.$$

6. Os resíduos são **independentes** dos valores ajustados:

$$\sum e_i \, \widehat{y}_i = 0.$$

7. A função de regressão ajusada **sempre passa pelo ponto** $(\overline{y}, \overline{x})$:

$$b_0 = \overline{y} - b_1 \overline{x} \quad \Rightarrow \quad \overline{y} = b_0 + b_1 \overline{x}$$

FORMA ALTERNATIVA DO MODELO LINEAR SIMPLES

Essa propriedade nos permite definir um versão alternativa para apresentação do MLEA:

$$\begin{aligned}\widehat{y}_i &= b_0 + b_1 x_i \\ &= b_0 + b_1 x_i + (b_1\overline{x} - b_1\overline{x}) \\ &= (b_0 + b_1\overline{x}) + b_1(x_i - \overline{x})\end{aligned}$$

$$\widehat{y}_i = \overline{y} + b_1(x_i - \overline{x}) \tag{1.5}$$

Note que essa forma alternativa também pode ser aplicada à versão populacional do modelo:

$$\boldsymbol{E}\{Y|X = x\} = \boldsymbol{E}\{Y\} + \beta_1(x - \overline{x}) \tag{1.6}$$

A versão alternativa deixa claro que

$$x_i = \overline{x} \quad \Rightarrow \quad \widehat{y}_i = \overline{y},$$

portanto, a função de regressão ajustada sempre passará pelo ponto $(\overline{y}, \overline{x})$.

1.3.7 Exemplo de Aplicação: Relação altura-DAP em *E.saligna*

Como exemplo de aplicação utilizaremos dados de altura total (variável resposta) e DAP (variável preditora) de 36 árvores de *E. saligna* (tabela 1.1).

O objetivo é obter um modelo que permita saber a altura total das árvores com base apenas na medida do DAP (relação hipsométrica).

Estimativas de Quadrados Mínimos

- Valores de somas de quadrados e produtos:

  $S_{xx} = 944.72$ $S_{yy} = 725.6075$ $S_{xy} = 641.25$

- EQM:

  $$b_1 = \frac{725.6075}{944.72} = 0.6787725$$

Tabela 1.1: Dados de altura total (variável resposta) e DAP (variável preditora) de 36 árvores de *E. saligna*.

ÁRVORE	ALTURA TOTAL (y_i)	DAP (x_i)	y_i^2	x_i^2	$x_i y_i$
1	21.5	19.9	462.25	396.01	427.85
2	15.7	12.4	246.49	153.76	194.68
3	11.7	16.5	136.89	272.25	193.05
4	7.7	9.0	59.29	81.00	69.30
5	6.6	7.0	43.56	49.00	46.20
6	8.8	10.5	77.44	110.25	92.40
7	12.9	13.0	166.41	169.00	167.70
8	20.0	20.0	400.00	400.00	400.00
9	11.6	7.0	134.56	49.00	81.20
10	6.4	6.3	40.96	39.69	40.32
11	18.2	15.5	331.24	240.25	282.10
12	11.0	8.5	121.00	72.25	93.50
13	11.3	7.0	127.69	49.00	79.10
14	15.7	12.5	246.49	156.25	196.25
15	10.5	10.4	110.25	108.16	109.20
16	7.5	6.0	56.25	36.00	45.00
17	14.2	12.3	201.64	151.29	174.66
18	11.4	10.0	129.96	100.00	114.00
19	11.3	5.9	127.69	34.81	66.67
20	11.4	20.0	129.96	400.00	228.00
21	6.7	7.0	44.89	49.00	46.90
22	14.9	15.3	222.01	234.09	227.97
23	8.1	6.0	65.61	36.00	48.60
24	14.0	8.6	196.00	73.96	120.40
25	11.0	12.5	121.00	156.25	137.50
26	25.5	22.0	650.25	484.00	561.00
27	12.6	9.4	158.76	88.36	118.44
28	20.4	17.0	416.16	289.00	346.80
29	17.3	16.0	299.29	256.00	276.80
30	17.2	12.0	295.84	144.00	206.40
31	16.2	17.0	262.44	289.00	275.40
32	11.8	13.7	139.24	187.69	161.66
33	16.8	23.0	282.24	529.00	386.40
34	15.6	22.0	243.36	484.00	343.20
35	18.6	19.0	345.96	361.00	353.40
36	14.2	15.4	201.64	237.16	218.68
TOTAL	486.30	465.60	7294.71	6966.48	6930.73

$$b_0 = \left(\frac{486.30}{36}\right) - [0.6787725]\left(\frac{465.60}{36}\right) = 13.50833 - [0.6787725]12.93333$$
$$b_0 = 4.729542$$

Verificando as Propriedades do Modelo Ajustado

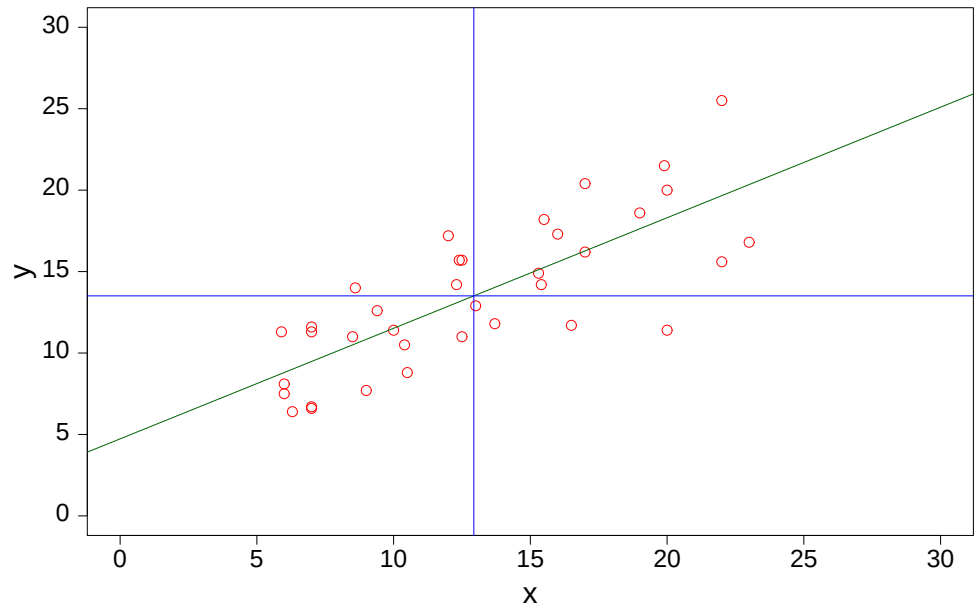
- Valores ajustados:

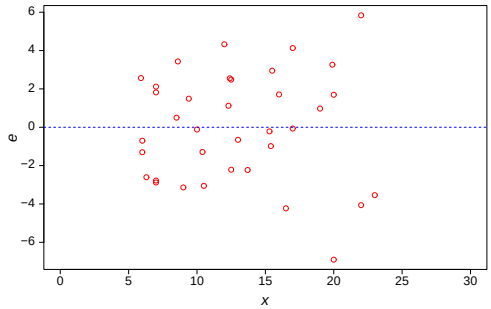
ÁRVORE (i)	ALTURA (y_i)	DAP (x_i)	VAL. AJUSTADOS ($\widehat{y_i}$)	RESÍDUOS (e_i)
1	21.5	19.9	18.237115	3.2628846
2	15.7	12.4	13.146321	2.5536787
3	11.7	16.5	15.929289	-4.2292887
4	7.7	9.0	10.838495	-3.1384947
5	6.6	7.0	9.480950	-2.8809496
6	8.8	10.5	11.856653	-3.0566535
7	12.9	13.0	13.553585	-0.6535848
8	20.0	20.0	18.304993	1.6950073
9	11.6	7.0	9.480950	2.1190504
10	6.4	6.3	9.005809	-2.6058088
11	18.2	15.5	15.250516	2.9494838
12	11.0	8.5	10.499108	0.5008916
13	11.3	7.0	9.480950	1.8190504
14	15.7	12.5	13.214199	2.4858014
15	10.5	10.4	11.788776	-1.2887762
16	7.5	6.0	8.802177	-1.3021770
17	14.2	12.3	13.078444	1.1215559
18	11.4	10.0	11.517267	-0.1172672
19	11.3	5.9	8.734300	2.5657002
20	11.4	20.0	18.304993	-6.9049927
21	6.7	7.0	9.480950	-2.7809496
22	14.9	15.3	15.114762	-0.2147617
23	8.1	6.0	8.802177	-0.7021770
24	14.0	8.6	10.566986	3.4330144
25	11.0	12.5	13.214199	-2.2141986
26	25.5	22.0	19.662538	5.8374622
27	12.6	9.4	11.110004	1.4899963
28	20.4	17.0	16.268675	4.1313250
29	17.3	16.0	15.589902	1.7100975
30	17.2	12.0	12.874812	4.3251877
31	16.2	17.0	16.268675	-0.0686750
32	11.8	13.7	14.028726	-2.2287256
33	16.8	23.0	20.341310	-3.5413103
34	15.6	22.0	19.662538	-4.0625378
35	18.6	19.0	17.626220	0.9737799
36	14.2	15.4	15.182639	-0.9826390
TOTAL	486.3000	—	486.3000	$-5.412337 \times 10^{-15}$

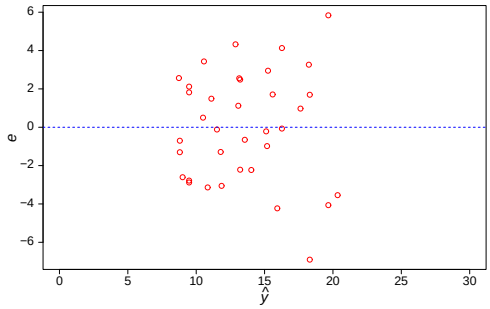
  - Note a validade das propriedades $\sum e_i = 0$ e $\sum y_i = \sum \widehat{y}_i$.

  - É Ainda possível verificar que:
    $\sum e_i x_i = -1.043610 \times 10^{-14}$ e $\sum e_i \widehat{y}_i = -1.909584 \times 10^{-14}$.
  - Mas é interessante verificar que $\sum e_i y_i = 290.3446 \neq 0$.

- Algumas das propriedades do MLEA podem ser melhor visualizadas graficamente:







1.3.8 Cuidados na Aplicação dos Modelos Lineares

- O fato da função de regressão representar uma variação sistemática de Y em função de X **não implica necessariamente** numa relação de causa e efeito.

  Qual a relação causal entre altura e diâmetro em árvores?

- Frequentemente a relação teórica (biológica) entre as variáveis é não-linear, mas o modelo linear pode ser uma **boa aproximação dentro da amplitude de interesse**.

  O escopo do modelo depende de ambas: *amplitude dos dados* e *escolha da forma do modelo.*

- **Interpolação** versus **Extrapolação**.

  Todas as propriedades estatísticas ligadas aos valores ajustados pelo modelo assumem que o modelo é utilizado **dentro** da amplitude original dos dados, i.e., o modelo é usado para **interpolar**.

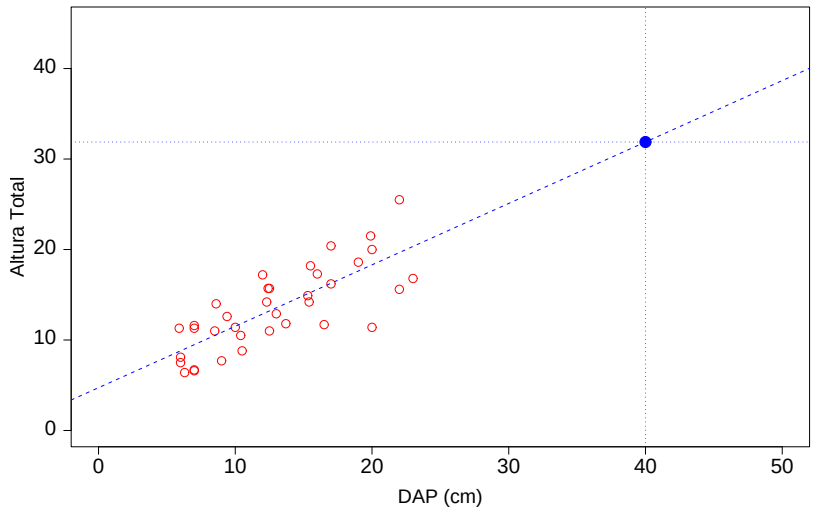
  Exemplos:

  – O valor ajustado para $X = 0$ pode não ser possível, pois é impossível observá-lo.

  – No exemplo de altura-DAP de *E. saligna*, alguém poderia (**erradamente**) tentar encontrar o valor ajustado para DAP = 40 cm:

  $$\widehat{y}_i = b_0 + b_1(40) = 31.9$$

  Esse valor ajustado é realista?



1.3.9 Exercícios

1.3-1. Analise os modelos abaixo e responda se eles são modelos lineares de primeira ordem e se o MLEA pode ser aplicado no ajuste deles.

(a) $Y = \beta_0 + \beta_1 X^3 + \varepsilon$;

(b) $Y = \beta_0 + \beta_1 \log(X) + \varepsilon$;

(c) $(1/Y) = \beta_0 + \beta_1 \log(X) + \varepsilon$;

1.3-2. Faça uma interpretação teórica dos coeficientes de regressão dos modelos abaixo:

(a) [Altura] $= \beta_0 + \beta_1$ [DAP] $+ \varepsilon$ (para árvores);

(b) [Vazão] $= \beta_0 + \beta_1$ [Precipitação] $+ \varepsilon$ (numa bacia hidrográfica);

(c) [Número de ovas] $= \beta_0 + \beta_1$ [Comprimento] $+ \varepsilon$ (peixe esturjão);

(d) [Produção] $= \beta_0 + \beta_1$ [Idade] $+ \varepsilon$ (para uma floresta plantada);

1.3-3. O MLEA assume que a variável preditora é uma variável matemática.

Que atitude prática pode ser tomada para "*amenizar*" essa pressuposição no caso da variável preditora ser aleatória?

Exemplos: DAP, precipitação, comprimento do peixe, etc.

1.3-4. Que argumentos teóricos podem ser levantados para afirmar que a pressuposição fundamental do MLEA ($\boldsymbol{E}\{\varepsilon|X = x\} = 0$) é realista?

1.3-5. Demostre as seguintes propriedades do modelo ajustado:

(a) $\sum e_i = 0$;

(b) $\sum \widehat{y}_i = \sum y_i$;

(c) $\sum e_i x_i = 0$;

(d) $\sum e_i \widehat{y}_i = 0$.

1.3-6. A tabela abaixo apresenta dados de comprimento de peixes esturjões (cm, x_i) e produção de ovas (milhares, y_i).

i	x_i	y_i	x_iy_i	y_i^2	x_i^2
1	105	45	4725	2025	11025
2	113	63	7119	3969	12769
3	125	86	10750	7396	15625
4	137	118	16166	13924	18769
5	141	112	15792	12544	19881
6	153	169	25857	28561	23409
7	165	201	33165	40401	27225
8	177	237	41949	56169	31329
9	188	263	49444	69169	35344
10	198	268	53064	71824	39204
$\sum$	1502	1562	258031	305982	234580

(a) Ajuste o MLEA aos dados.

(b) Verifique empiricamente, por cálculos e gráficos, as propriedades do MLEA.

1.3-7. No exemplo do esturjão, utilize o modelo ajustado.

(a) Encontrar os valores ajustados para os comprimentos: 50, 100, 150, 200, 250 cm;

(b) Comente o "*realismo*" dos valores ajustados encontrados.

1.4 O Modelo Linear Simples de Erros Aleatórios Homocedásticos e Independentes (MLEAHI)

O segundo modelo linear simples que veremos, acrescenta duas pressuposições novas ao MLEA. Além dos erros serem aleatórios com valor esperado nulo para cada nível da variável preditora ($\boldsymbol{E}\{\varepsilon|X=x\}=0$), assumiremos também

1. Os erros são homocedásticos, i.e., têm variância homogênea. A variância dos erros é constante para todos os níveis da variável preditora. Utilizaremos a seguinte notação matemática:
$$\boldsymbol{Var}\{\varepsilon|X=x\} = \boldsymbol{Var}\{\varepsilon\} = \sigma^2$$
2. Os erros são independentes, i.e., a correlação entre dois erros é sempre nula:
$$\boldsymbol{Cov}\{\varepsilon_i, \varepsilon_j\} = 0$$

1.4.1 Estrutura e Pressuposições do MLEAHI

A definição formal do MLEAHI (na notação sintética) é

$$\text{MLEAHI:} \qquad Y = \beta_0 + \beta_1 X + \varepsilon, \tag{1.7}$$

onde:

$$\begin{aligned} X &- \text{variável não estocástica} \\ \boldsymbol{E}\{\varepsilon|X=x\} &= 0; \\ \boldsymbol{Var}\{\varepsilon\} &= \sigma^2; \\ \boldsymbol{Cov}\{\varepsilon_i, \varepsilon_j\} &= 0. \end{aligned}$$

CONCEITOS BÁSICOS DE ESTATÍSTICA

Para prosseguirmos necessitamos rever alguns conceitos básicos de estatística:

- variância de uma variável aleatória;
- covariância de duas variáveis aleatórias.

CONSIDERAÇÕES SOBRE AS PRESSUPOSIÇÕES

Variável preditora não estocástica: valem as considerações já apresentadas para o MLEA.

Valor esperado nulo para o erro ($\boldsymbol{E}\{\varepsilon|X = x\} = 0$): valem as considerações e deduções já apresentadas para o MLEA.

Homocedasticidade: podemos considerar a pressuposição dos erros homocedásticos por dois aspectos:

- Se considerarmos o termo do erro como sendo gerado por erros aleatórios de mensuração, assumir homocedasticidade implica em assumir que a dispersão dos erros de mensuração é a mesma, não variando com a magnitude do que se mede, i.e, com o valor da variável preditora ou da variável resposta.

  Em processos de mensuração, a homocedasticidade é uma propriedade muito desejável e na maioria dos processos controlados ela deve ocorrer.

- Se considerarmos o termo do erro como resultando a dispersão da variável resposta ao redor de seu valor esperado para cada nível da variável preditora ($Y|X = x$ é variável estocásitca), então a questão se torna mais problemáticas.

  Em geral, as variáveis que medimos na natureza possuem **dispersão proporcional**, i.e., a magnitude da dispersão é proporcional ao valor da variável.

  Por exemplo: é comum que a dispersão do volume de árvores em florestas plantadas fique ao redor de 15%.

Veremos, mais adiante, como tratar os casos em que a homocedasticidade não pode ser assumida.

Independência: A pressuposição da independência é fundamental na maioria dos modelos estatísticos. Podemos considerar que:

- Se assumirmos que as observações de campo são realizadas de modo independente, essa pressuposição é bastante razoável.
- A aleatorização realizada na amostragem e na instalação de experimento é um procedimento que visa garantir essa independência.
- Entretanto, numa série de situações a independência pode ficar comprometida devido a influências de dependências temporais e espaciais.

1.4.2 Estimadores de Quadrados Mínimos e suas Propriedades

Os EQM são os mesmos do MLEA (equações 1.3 e 1.4), que podem ser resumidamente apresentadas como

$$b_1 = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sum(x_i - \overline{x})^2} = \frac{S_{xy}}{S_{xx}} \qquad \text{e} \qquad b_0 = \overline{y} - b_1\overline{x}.$$

Os EQM têm as seguintes propriedades no MLEAHI:

1. EQM são funções lineares dos valores observados de Y. Com foi visto no MLEA:

$$\begin{aligned} b_1 = & \quad \sum \tfrac{(x_i-\overline{x})}{\sum(x_i-\overline{x})^2} y_i & = \sum k_i y_i \\ b_0 = & \quad \sum \left[\tfrac{1}{n} - \tfrac{\overline{x}(x_i-\overline{x})}{\sum(x_i-\overline{x})^2}\right] y_i & = \sum k_i^* y_i \end{aligned}$$

2. EQM são estimadores não viciados dos coeficientes de regressão. Com foi visto no MLEA:

$$\boldsymbol{E}\{b_0\} = \beta_0 \quad \text{e} \quad \boldsymbol{E}\{b_1\} = \beta_1.$$

3. As variâncias dos EQM são proporcionais à variância do erro (σ^2):

$$\boldsymbol{Var}\{b_1\} = \sigma^2 \frac{1}{\sum(x_i - \overline{x})^2} \tag{1.8}$$

$$\boldsymbol{Var}\{b_0\} = \sigma^2 \left[\frac{1}{n} + \frac{\overline{x}^2}{\sum(x_i - \overline{x})^2}\right] \tag{1.9}$$

Vejamos a variância do estimador da inclinação (b_1):

$$\boldsymbol{Var}\{b_1\} = \boldsymbol{Var}\{\sum k_i y_i\} = \sum k_i^2 \boldsymbol{Var}\{y_i\} + \sum k_i k_j \boldsymbol{Cov}\{y_i, y_j\}.$$

Note que

$$\boldsymbol{Cov}\{y_i, y_j\} = \boldsymbol{Cov}\{\beta_0 + \beta_1 x_i + \varepsilon_i, \beta_0 + \beta_1 x_j + \varepsilon_j\} = \boldsymbol{Cov}\{\varepsilon_i, \varepsilon_j\} = 0.$$

Ou seja, independência dos erros implica em independência das observações da variável resposta.

Voltando à variância de b_1:

$$\begin{aligned} \boldsymbol{Var}\{b_1\} &= \sum k_i^2 \boldsymbol{Var}\{y_i\} = \sum k_i^2 \boldsymbol{Var}\{\varepsilon_i\} = \sum k_i^2 \sigma^2 = \sigma^2 \sum k_i^2 \\ &= \sigma^2 \sum \left[\frac{(x_i - \overline{x})}{\sum(x_i - \overline{x})^2}\right]^2 = \sigma^2 \frac{\sum x_i - \overline{x})^2}{[\sum(x_i - \overline{x})^2]^2} \\ \boldsymbol{Var}\{b_1\} &= \sigma^2 \frac{1}{\sum(x_i - \overline{x})^2} \end{aligned}$$

A variância do estimador do intercepto (b_0):

$$\textbf{\textit{Var}}\{b_0\} \quad = \quad \textbf{\textit{Var}}\{\sum k_i^* y_i\} = \sum k_i^{*2} \textbf{\textit{Var}}\{y_i\} + \sum k_i^* k_j^* \textbf{\textit{Cov}}\{y_i, y_j\}.$$

Novamente temos que

$$\textbf{\textit{Cov}}\{y_i, y_j\} = \textbf{\textit{Cov}}\{\varepsilon_i, \varepsilon_j\} = 0 \qquad \text{e} \qquad \textbf{\textit{Var}}\{y_i\} = \textbf{\textit{Var}}\{\varepsilon_i\} = \sigma^2,$$

portanto,

$$\begin{aligned}
\textbf{\textit{Var}}\{b_0\} &= \sum k_i^{*2} \textbf{\textit{Var}}\{y_i\} = \sum \left[\frac{1}{n} - \frac{\overline{x}(x_i - \overline{x})}{\sum (x_i - \overline{x})^2}\right]^2 \sigma^2 \\
&= \sigma^2 \sum \left[\frac{1}{n^2} - \frac{2\,\overline{x}(x_i - \overline{x})}{\sum (x_i - \overline{x})^2} + \frac{\overline{x}^2 (x_i - \overline{x})^2}{\left[\sum (x_i - \overline{x})^2\right]^2}\right] \\
&= \sigma^2 \left[\frac{n}{n^2} - \frac{2\,\overline{x}}{\sum (x_i - \overline{x})^2} \sum (x_i - \overline{x}) + \frac{\overline{x}^2}{\left[\sum (x_i - \overline{x})^2\right]^2} \sum (x_i - \overline{x})^2\right] \\
\textbf{\textit{Var}}\{b_0\} &= \sigma^2 \left[\frac{1}{n} + \frac{\overline{x}^2}{\sum (x_i - \overline{x})^2}\right]
\end{aligned}$$

Comentário: as variâncias dos EQM dos coeficientes de regressão inversamente proporcionais à soma de quadrados da variável preditora ($\sum (x_i - \overline{x})^2$), logo:

- o **aumento do tamanho da amostra** (n) tem um efeito indireto (via soma de quadrados de X) reduzindo a variância dos estimadores, mas no caso do intercepto (b_0) terá efeito direto também;
- o **aumento da amplitude de amostragem** ($x_{\max} - x_{\min}$) tem um efeito indireto (via soma de quadrados de X) reduzindo a variância dos estimadores.

  O efeito do aumento da amplitude de amostragem pode ser "*intuitivamente*" captado olhando-se o gráfico de dispersão da variável resposta pela variável preditora:

Questão: *Quais as implicações disso no planejamento de um experimento ou estudo baseado em amostragem?*

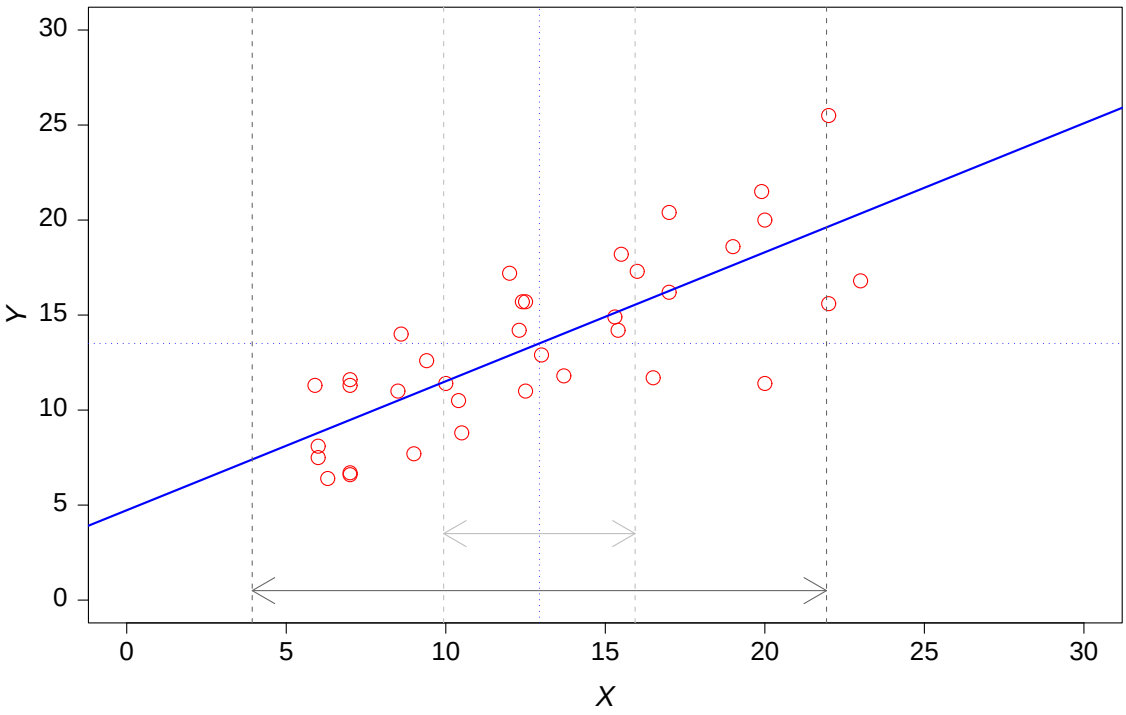


Figura 1.4: Gráfico exemplificando a amplitude de amostragem da variável preditora (X) no ajuste da reta de regressão.

4. Os EQM **não são independentes**.

   Prova: lembrando que as observações da variável resposta são independentes (***Cov***$\{y_i, y_j\} =$ ***Cov***$\{\varepsilon_i, \varepsilon_j\} = 0$), temos:

$$
\begin{aligned}
\textbf{\textit{Cov}}\{b_0, b_1\} &= \textbf{\textit{Cov}}\{\sum k_i^* y_i, \sum k_i y_i\} \\
&= \sum k_i^* k_i \textbf{\textit{Var}}\{y_i\} \\
&= \sigma^2 \sum \left(\frac{1}{n} - \frac{\overline{x}(x_i - \overline{x})}{\sum (x_i - \overline{x})^2} \right) \left(\frac{(x_i - \overline{x})}{\sum (x_i - \overline{x})^2} \right) \\
&= \sigma^2 \sum \left(\frac{(x_i - \overline{x})}{n \sum (x_i - \overline{x})^2} - \frac{\overline{x}(x_i - \overline{x})^2}{\left[\sum (x_i - \overline{x})^2\right]^2} \right) \\
&= \sigma^2 \left(\frac{\sum (x_i - \overline{x})}{n \sum (x_i - \overline{x})^2} - \frac{\overline{x}}{\left[\sum (x_i - \overline{x})^2\right]^2} \sum (x_i - \overline{x})^2 \right)
\end{aligned}
$$

$$\boldsymbol{Cov}\{b_0, b_1\} = \frac{-\sigma^2 \overline{x}}{\sum (x_i - \overline{x})^2}$$

Comentário: é importante notar que:

- A covariância entre os EQM também é inversamente proporcional soma de quadrados da variável preditora, sendo reduzida à medida que se aumenta a amplitude de amostragem da variável preditora.
- A covariância entre os EQM é **negativa**, ou seja, para um dado problema, aumento no valor de b_1 implicam em redução no valor de b_0, e vice-versa.

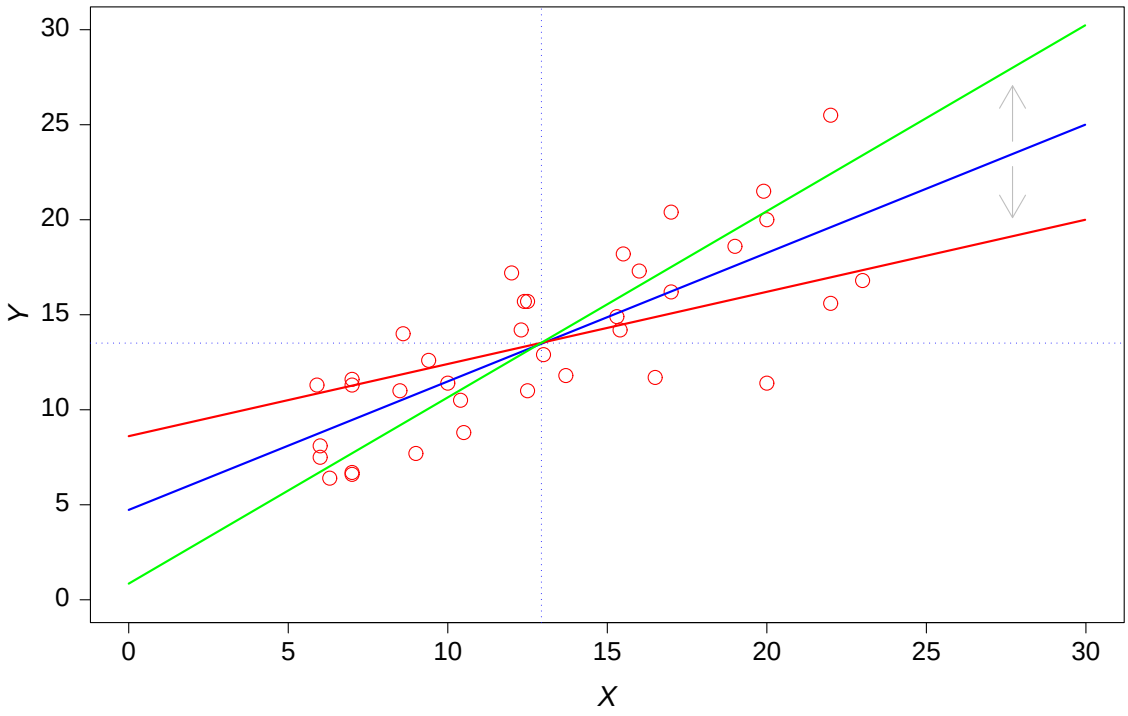


Figura 1.5: Gráfico mostrando o efeito da covariância entre os EQM sobre a reta de regressão.

- Na versão alternativa do modelo linear simples:

$$\widehat{y}_i = \overline{y} + b_1(x_i - \overline{x})$$

há independência entre coeficientes b_1 e $\overline{y}$:

$$\begin{aligned} \textbf{\textit{Cov}}\{\overline{y}, b_1\} &= \textbf{\textit{Cov}}\{\sum \frac{1}{n} y_i, \sum k_i y_j\} = \sum \frac{1}{n} k_i \textbf{\textit{Var}}\{y_i\} \\ &= \sigma^2 \sum \frac{1}{n} \frac{x_i - \overline{x}}{\sum (x_i - \overline{x})^2} = \frac{\sigma^2}{n \sum (x_i - \overline{x})^2} \sum (x_i - \overline{x}) = 0 \end{aligned}$$

5. Estimadores de Variância Mínima (**Teorema de Gauss**): Dentre todos estimadores lineares não-viciados, os EQM (no modelo MLEAHI) são estimadores de variância mínima.

   b.l.u.e. - *"best linear unbiased estimators"*.

> CONCEITOS BÁSICOS DE ESTATÍSTICA
>
> Para prosseguirmos necessitamos rever um conceito básico de estatística:
>
> - propriedade de estimadores estatísticos: estimador de variância mínima.

Comentário:

- O **Teorema de Gauss** é apresentado frequentemente nos livros texto de regressão com o nome de Gauss-Markov.
- Esse teorema prova a **generalidade** e **robustez** dos estimadores de quadrados mínimos.

1.4.3 Propriedades do Modelo Ajustado

Os valores ajustados são combinações lineares dos valores observados

Podemos acrescentar mais essa propriedade (oitava) àquelas já demonstradas para os modelos lineares simples ajustados por quadrados mínimos (modelo MLEA).

Prova: partindo da versão alternativa do modelo linear simples:

$$\begin{aligned} \widehat{y}_i &= \overline{y} + b_1 (x_i - \overline{x}) \\ &= \sum_j \frac{1}{n} y_j + (x_i - \overline{x}) \sum_j \frac{(x_j - \overline{x})}{\sum_j (x_j - \overline{x})^2} y_j \end{aligned}$$

$$
\begin{aligned}
&= \sum_j \left[\frac{1}{n} + \frac{(x_i - \overline{x})(x_j - \overline{x})}{\sum_j (x_j - \overline{x})^2}\right] y_j \\
\widehat{y}_i &= \sum_j k_{ij} y_j
\end{aligned}
$$

Variância dos valores ajustados:

Novamente partindo da versão alternativa do modelo linear simples:

$$
\begin{aligned}
\boldsymbol{Var}\{\widehat{y}_i\} &= \boldsymbol{Var}\{\overline{y} + b_1(x_i - \overline{x})\} \\
&= \boldsymbol{Var}\{\overline{y}\} + (x_i - \overline{x})^2 \boldsymbol{Var}\{b_1\} + (x_i - \overline{x}) \boldsymbol{Cov}\{\overline{y}, b_1\} \\
&= \frac{\sigma^2}{n} + (x_i - \overline{x})^2 \sigma^2 \frac{1}{\sum (x_i - \overline{x})^2}
\end{aligned}
$$

$$
\boldsymbol{Var}\{\widehat{y}_i\} = \sigma^2 \left[\frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum (x_i - \overline{x})^2}\right] \tag{1.10}
$$

Comentário: note que a variância dos valores ajustados é diretamente proporcional ao quadrado da distância o nível da variável preditora e do valor médio dessa variável.

Quando o nível da variável preditora considerado é o valor médio temos:

$$
\boldsymbol{Var}\{\widehat{y}_i | x_i = \overline{x}\} = \frac{\sigma^2}{n}
$$

que é o valor mínimo da variância dos valores esperados, sendo igual à variância da média para variável resposta.

Assim, à medida que nos afastamos da região central dos dados $(\overline{x}, \overline{y})$, a variância dos valores esperados aumenta.

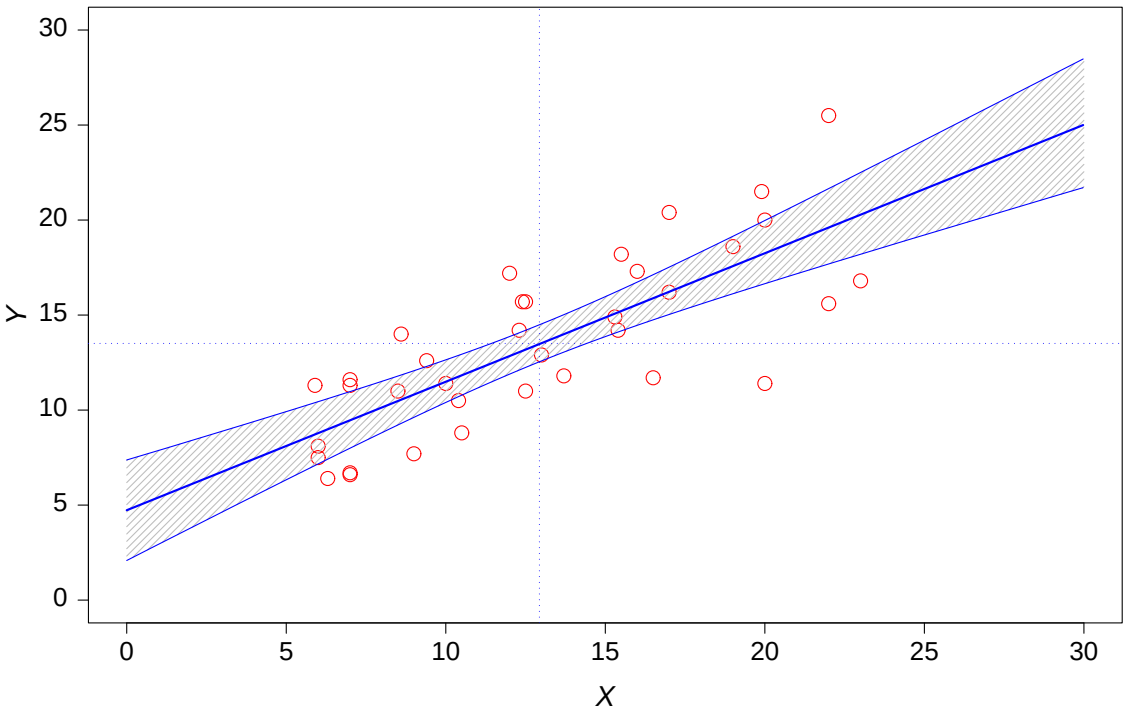


Figura 1.6: Gráfico exemplificando a variação na variância do valor ajustado em função da distância do ponto central $(\overline{y}, \overline{x})$. As linhas acima e abaixo da reta de regressão representam o valor esperado mais ou menos, respectivamente, duas vêzes o erro padrão do valor ajustado ($\sqrt{\boldsymbol{Var}\{\widehat{y}_i\}}$).

1.4.4 Estimador da Variância do Erro

Listemos as fórmulas de variância e covariância encontradas até agora no MLEAHI:

$$\begin{aligned}
\boldsymbol{Var}\{b_1\} &= \sigma^2 \frac{1}{\sum(x_i - \overline{x})^2}, \\
\boldsymbol{Var}\{b_0\} &= \sigma^2 \left[\frac{1}{n} + \frac{\overline{x}^2}{\sum(x_i - \overline{x})^2}\right], \\
\boldsymbol{Cov}\{b_0, b_1\} &= \frac{-\sigma^2 \overline{x}}{\sum(x_i - \overline{x})^2}, \\
\boldsymbol{Var}\{\widehat{y}_i\} &= \sigma^2 \left[\frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum(x_i - \overline{x})^2}\right].
\end{aligned}$$

Em todas as fórmulas o único parâmetro é a variância do erro ($\boldsymbol{Var}\{\varepsilon_i|x = x_i\} = \sigma^2$), para o qual devemos encontrar um estimador de modo que as variâncias acima possam ser efetivamente estimadas.

Uma vez que os resíduos (e_i) são os representantes dos erros (ε_i) quando ajustamos o modelo a um conjunto de dados, a soma de quadrados do resíduos pode ser um bom estimador da variância do erro (σ^2).

$$\begin{aligned}\boldsymbol{E}\{SQR\} &= \boldsymbol{E}\{Q\} = \boldsymbol{E}\left\{\sum e_i^2\right\} = \boldsymbol{E}\left\{\sum (y_i - \widehat{y}_i)^2\right\} \\ &= \sum \boldsymbol{E}\left\{y_i^2 - 2y_i\widehat{y}_i + \widehat{y}_i^2\right\} \\ &= \sum \left[\boldsymbol{E}\left\{y_i^2\right\} - 2\boldsymbol{E}\{y_i\widehat{y}_i\} + \boldsymbol{E}\left\{\widehat{y}_i^2\right\}\right]\end{aligned}$$

Nesse ponto, necessitamos de uma das propriedades da variância e covariância de variáveis aleatórias:

$$\begin{aligned}\boldsymbol{Var}\{Z\} = \boldsymbol{E}\left\{Z^2\right\} - [\boldsymbol{E}\{Z\}]^2 \quad &\Rightarrow \quad \boldsymbol{E}\left\{Z^2\right\} = \boldsymbol{Var}\{Z\} + [\boldsymbol{E}\{Z\}]^2 . \\ \boldsymbol{Cov}\{Z, W\} = \boldsymbol{E}\{ZW\} - \boldsymbol{E}\{Z\}\,\boldsymbol{E}\{W\} \quad &\Rightarrow \quad \boldsymbol{E}\{ZW\} = \boldsymbol{Cov}\{Z, W\} + \boldsymbol{E}\{Z\}\,\boldsymbol{E}\{W\}\end{aligned}$$

Voltando ao valor esperado da soma de quadrados do resíduo

$$\begin{aligned}\boldsymbol{E}\{SQR\} &= \sum \left[\boldsymbol{Var}\{y_i\} + (\boldsymbol{E}\{y_i\})^2 - 2(\boldsymbol{Cov}\{y_i, \widehat{y}_i\} + \boldsymbol{E}\{y_i\}\,\boldsymbol{E}\{\widehat{y}_i\}) + \boldsymbol{Var}\{\widehat{y}_i\} + (\boldsymbol{E}\{\widehat{y}_i\})^2\right] \\ &= \sum \left[\boldsymbol{Var}\{y_i\} + (\boldsymbol{E}\{Y\})^2 - 2\boldsymbol{Var}\{\widehat{y}_i\} - 2\boldsymbol{E}\{Y\}\,\boldsymbol{E}\{Y\} + \boldsymbol{Var}\{\widehat{y}_i\} + (\boldsymbol{E}\{Y\})^2\right] \\ &= \sum [\boldsymbol{Var}\{y_i\} - \boldsymbol{Var}\{\widehat{y}_i\}] \\ &= \sum \left[\sigma^2 - \sigma^2\left(\frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum (x_i - \overline{x})^2}\right)\right] \\ &= \sigma^2 \sum \left[1 - \frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum (x_i - \overline{x})^2}\right] \\ \boldsymbol{E}\{SQR\} &= \sigma^2(n-2)\end{aligned}$$

Portanto, a soma de quadrados do resíduo pode compor um estimador não-viciado para a variância do erro. Esse estimador é chamado de Quadrado Médio do Resíduo (QMR):

$$\widehat{\sigma^2} = QMR = \frac{SQR}{n-2} = \frac{\sum e_i^2}{n-2} = \frac{\sum (y_i - \widehat{y}_i)^2}{n-2} = \frac{\sum (y_i - b_0 - b_1 x_i)^2}{n-2} \tag{1.11}$$

O QMR é composto dividindo-se a SQR pelos *graus de liberdade*, i.e., o número de observações (n) menos o número de coeficientes de regressão (2) estimados pelos quadrados mínimos.

A SQR pode ser obtida por dois caminhos:

1. Diretamente, a partir dos próprios resíduos:

   (1.12) $$SQR = \sum e_i^2 = \sum (y_i - \widehat{y}_i)^2.$$

2. Indiretamente, através das somas de quadrados e soma de produtos:

   (1.13) $$SQR = S_{yy} - b_1 S_{xy} = \sum (y_i - \overline{y})^2 - b_1 \sum (x_i - \overline{x})(y_i - \overline{y})$$

   Nesse segundo caso, é necessário manter o valor de b_1 com um grande número de algarismos para que erros de aproximação não comprometam o cálculo da SQR.

1.4.5 Desigualdade de Gauss

Para realizarmos inferências a partir de um modelo estatístico é necessário que se assuma explicitamente uma distribuição estatística ligada ao modelo.

Gauss, trabalhando com erros de mensurações astronômicas no século XIX, estabeleceu uma desigualdade que é muito útil para inferências aproximadas que podem ser aplicadas ao MLEAHI.

DESIGUALDADE DE GAUSS

Para qualquer variável aleatória contínua Z, com desvio padrão σ e cuja densidade é simétrica em relação à sua moda μ, aplica-se a desigualdade:

(1.14) $$P(|Z - \mu| \geq \lambda\sigma) \leq \begin{cases} 1 - \lambda/\sqrt{3} & (\lambda \leq 2/\sqrt{3}) \\ 4/(9\lambda^2) & (\lambda \geq 2/\sqrt{3}) \end{cases}$$

Por exemplo se tomarmos $\lambda = 2$ teremos:

$$P(|Z - \mu| \geq 2\sigma) \leq 0.1111$$

ou seja, para a variável aleatória Z existe uma probabilidade de **no máximo** 11% de que uma observação assuma valores que se distanciem 2 desvios padrão (2σ) da moda.

De modo reverso, há uma probabilidade de **no mínimo** 89% de que os valores de Z fiquem na faixa de mais ou menos 2 desvios padrão ao redor da moda.

A desigualdade de Gauss permite a construção de **Intervalos de Confiança Aproximados** para qualquer variável aleatória que satisfaça a duas condições:

1. seja variável contínua; e
2. tenha função de densidade simétrica em relação à moda.

1.4.6 Inferência sobre os Coeficientes de Regressão

Utilizando a Desigualdade de Gauss e o estimador da variância do erro podemos construir intervalos de confiança aproximados para os coeficientes de regressão.

Intercepto: a variância da estimativa do intercepto (b_0) é:

$$\boldsymbol{Var}\{b_0\} = \sigma^2 \left[\frac{1}{n} + \frac{\overline{x}^2}{\sum(x_i - \overline{x})^2}\right],$$

a qual pode ser estimada por

$$\widehat{\boldsymbol{Var}}\{b_0\} = QMR \left[\frac{1}{n} + \frac{\overline{x}^2}{\sum(x_i - \overline{x})^2}\right].$$

Portanto, o **Intervalo de Confiança Aproximado** de 89% ($\approx 90\%$) pode ser construído por

$$b_0 \pm 2\sqrt{QMR \left[\frac{1}{n} + \frac{\overline{x}^2}{\sum(x_i - \overline{x})^2}\right]}.$$

Inclinação: a variância da estimativa do inclinação (b_1) é:

$$\boldsymbol{Var}\{b_1\} = \frac{\sigma^2}{\sum(x_i - \overline{x})^2},$$

a qual pode ser estimada por

$$\widehat{\boldsymbol{Var}}\{b_1\} = \frac{QMR}{\sum(x_i - \overline{x})^2}.$$

Portanto, o **Intervalo de Confiança Aproximado** de 89% ($\approx 90\%$) pode ser construído por

$$b_1 \pm 2\sqrt{\frac{QMR}{\sum(x_i - \overline{x})^2}}.$$

Interpretação dos Intervalos de Confiança Aproximados (ICA)

- Os ICA são **condicionais** aos níveis da variável preditora, pois as estimativas de quadrados mínimos foram encontradas com base nesses níveis.
- O coeficiente de confiança não é exato, mas sim **mínimo** (intervalo conservador).
- No contexto do MLEAHI, um ICA de 90% deveria ser interpretado como:

  Dentre amostras repeditas de tamanho n, nos mesmos níveis da variável preditora X, pelo menos 90% delas resultarão em intervalos que includem o valor do coeficiente de regressão.

1.4.7 Inferências sobre os Valores Ajustados

Os valores ajustados por um modelo de regressão linear podem ter interpretações bastante diversas durante a aplicação do modelo.

Estimação versus Predição

É importante diferenciar dois conceitos chaves da aplicação de modelos estatísticos:

Estimação: processo de encontar valores *plausíveis para um parâmetro*, como por exemplo: média, mediana, variância, desvio padrão, **coeficiente de regressão**, etc.

Predição: processo de encontrar valores *plausíveis para uma variável aleatória*.

Ao ajustarmos um modelo linear simples, o que representa os valores ajustados que se alinham ao longo da reta de regressão?

1. Eles representam a melhor **estimativa** para o valor esperado da variável resposta Y, para cada nível da variável preditora X, pois já foi demonstrado que eles são estimadores não-viciados dela:

$$\boldsymbol{E}\{\widehat{y}_i\} = \boldsymbol{E}\{Y|X = x_i\}\,.$$

2. Se for assumido que a variável resposta (Y) é simétrica em relação ao seu valor esperado ($\boldsymbol{E}\{Y|X = x_i\}$), que é também a sua moda, então podemos dizer que os valores ajustados são:

a melhor predição para qualquer valor que a variável reposta Y pode assumir, para cada nível da variável preditora X.

Revisitemos o gráfico esquemático dos modelos lineares simples:

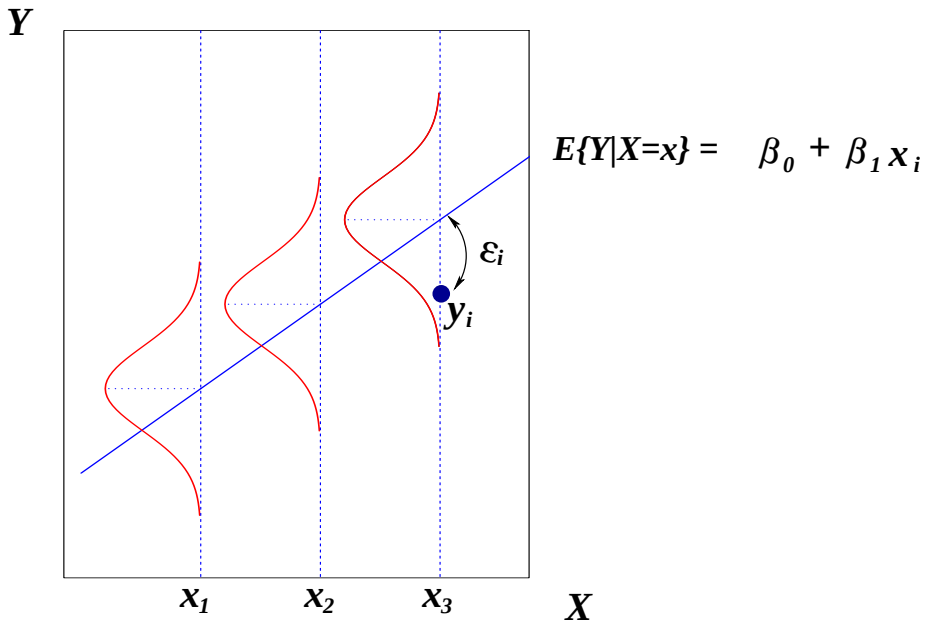


Figura 1.7: Gráfico esquemático dos modelos lineares simples.

Utilizaremos a seguinte terminologira:

Estimação da Resposta Média ($\widehat{y}$): procedimento de inferência relacionado ao *valor esperado* da variável resposta.

Estimação resulta em **Intervalos de Confiança**.

Predição de Uma Nova Observação ($\widehat{y}_h$): procedimento de inferência relacionado a um *valor qualquer* da variável resposta.

Predição resulta em **Intervalos de Predição**.

Estimação da Resposta Média

A reposta média é estimada pelo modelo ajustado:

$$\widehat{y}_i = b_0 + b_1 x_i$$

com variância igual a

$$\boldsymbol{Var}\{\widehat{y}_i\} = \sigma^2 \left[\frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum(x_i - \overline{x})^2}\right].$$

e variância estimada por

$$\widehat{\boldsymbol{Var}}\{\widehat{y}_i\} = QMR \left[\frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum(x_i - \overline{x})^2}\right].$$

O ICA de 90% é construído como

$$\widehat{y}_i \pm 2\sqrt{QMR \left[\frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum(x_i - \overline{x})^2}\right]}.$$

Predição de uma Nova Observação

A predição de uma nova observação também parte do modelo ajustado, aplicado a um novo valor da variável preditora ($X = x_h$):

$$\widehat{y}_h = b_0 + b_1 x_h.$$

Não é um parâmetro que estamos procurando, mas uma observação amostrada da distribuição de $Y|X = x_h$.

O que sabemos da distribuição de $Y|X = x_h$?

- Que tem valor esperado igual a $\boldsymbol{E}\{Y|X = x_h\} = \beta_0 + \beta_1 x_h$.
- Que tem variânica equal a $\boldsymbol{Var}\{Y|X = x_h\} = \sigma^2$.
- Que é uma variável aleatória contínua com densidade simétrica em relação à sua moda ($\boldsymbol{E}\{Y|X = x_h\}$): Desigualdade de Gauss se aplica.

Entretanto, nós não conhecemos os valores dos parâmetros, mas somente temos estimativas deles.

Portanto, a variabilidade de uma nova observação terá duas fontes de incerteza:

1. A incerteza na estimativa do valor esperado de $Y|X = x_h$ ($\boldsymbol{E}\{Y|X = x_h\}$).
   (Onde passa a linha de regressão?)
   $\Rightarrow$ variância da distribuição amostral de $\widehat{y}$.

2. A incerteza das observações da variável aleatória $Y|X = x_h$ em relação ao seu valor esperado $\boldsymbol{E}\{Y|X = x_h\}$.

   (Onde ocorrerá uma nova observação?)

   $\Rightarrow$ variância da distribuição de $Y|X = x_h$.

Assim, a variância de uma nova observação será:

$$\begin{aligned}
\boldsymbol{Var}\{\widehat{y}_h\} &= \boldsymbol{Var}\{\widehat{y}\} + \boldsymbol{Var}\{Y|X = x_h\} \\
&= \sigma^2 \left[\frac{1}{n} + \frac{(x_h - \overline{x})^2}{\sum(x_i - \overline{x})^2}\right] + \sigma^2 \\
&= \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_h - \overline{x})^2}{\sum(x_i - \overline{x})^2}\right],
\end{aligned}$$

que pode ser estimada por

$$\widehat{\boldsymbol{Var}}\{\widehat{y}_h\} = QMR \left[1 + \frac{1}{n} + \frac{(x_h - \overline{x})^2}{\sum(x_i - \overline{x})^2}\right].$$

O **Intervalo de Predição Aproximado (IPA)** de 90% para uma nova observação é dado por

$$\widehat{y}_h \pm 2\sqrt{QMR \left[1 + \frac{1}{n} + \frac{(x_h - \overline{x})^2}{\sum(x_i - \overline{x})^2}\right]}.$$

Predição da Média Amostral de m Novas Observações

Em algumas situações especiais o modelo linear pode ser utilizado para inferir sobre o valor de uma média amostral de um conjunto de m novas observações.

É importante lembrarmos que a média amostral é uma *variável aleatória* com algumas propriedades importantes em relação à distribuição da variável original.

CONCEITOS BÁSICOS DE ESTATÍSTICA

Precisamos rever um conceito básico que é central na teoria estatística:

- Teorema Central do Limite.

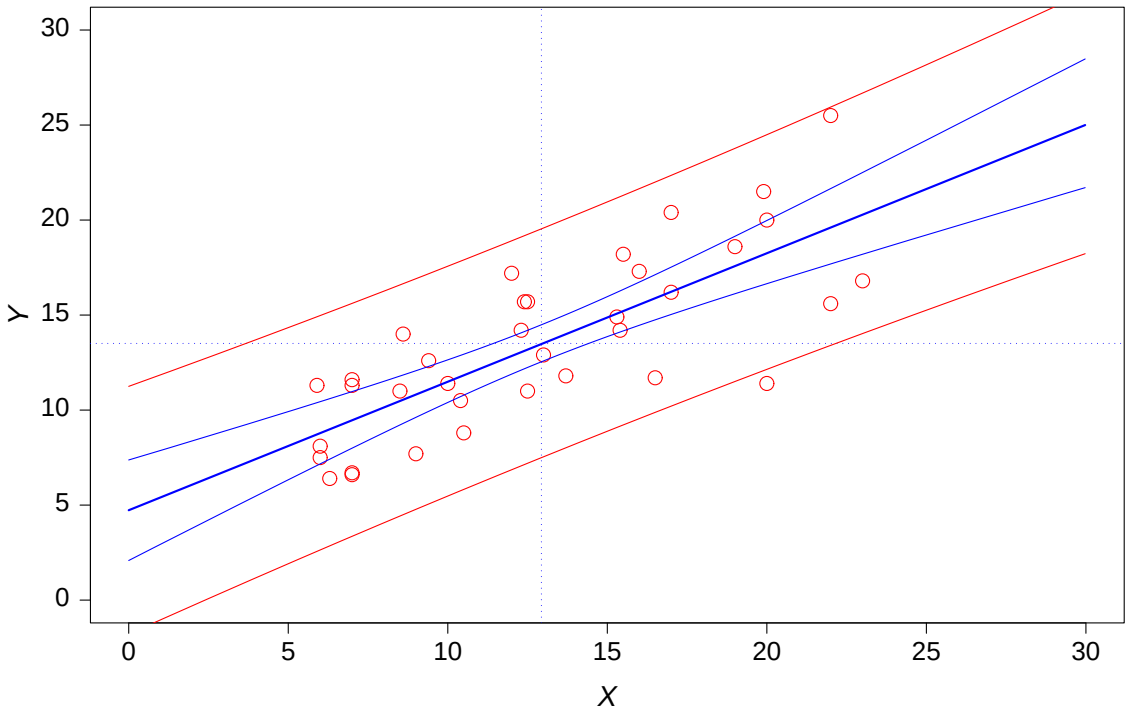


Figura 1.8: Gráfico exemplificando intervalo de confiança (linha azul) e intervalo de predição (linha vermelha).

De forma análogoa à questão de uma nova observação, a média amostral de m novas observações será obtida pela linha de regressão:

$$\widehat{\overline{y}}_m = b_0 + b_1 x_h$$

e terá variância igual a:

$$\begin{aligned} \boldsymbol{Var}\{\widehat{\overline{y}}_m\} &= \boldsymbol{Var}\{\widehat{y}\} + \frac{\boldsymbol{Var}\{Y|X=x_h\}}{m} \\ &= \sigma^2\left[\frac{1}{n} + \frac{(x_h-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right] + \frac{\sigma^2}{m} \\ &= \sigma^2\left[\frac{1}{m} + \frac{1}{n} + \frac{(x_h-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right], \end{aligned}$$

que pode ser estimada por

$$\widehat{\boldsymbol{Var}}\{\widehat{\overline{y}}_m\} = QMR\left[\frac{1}{m} + \frac{1}{n} + \frac{(x_h-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right].$$

O IPA de 90% para média de m novas observações é dado por

$$\widehat{\overline{y}}_m \pm 2\sqrt{QMR\left[\frac{1}{m}+\frac{1}{n}+\frac{(x_h-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right]}.$$

É interessante ressaltar que o Teorema Central do Limite também se aplica a esse caso, pois se fizermos o tamanho de amostra m das novas observações crescer, teremos no limite a seguintes situação:

$$\lim_{m\to\infty} \textbf{\textit{Var}}\{\widehat{\overline{y}}_m\} \;=\; \lim_{m\to\infty}\left[\textbf{\textit{Var}}\{\widehat{y}\}+\frac{\textbf{\textit{Var}}\{Y|X=x_h\}}{m}\right] = \textbf{\textit{Var}}\{\widehat{y}\}$$

Ou seja, a variância da média de m novas observações tende à variância da resposta média, à medida que o tamanho de amostra tende ao infinito ($m \to \infty$).

1.4.8 Exemplo de Aplicação: Vazão no rio Tinga

Conjunto de dados: dados quinzenais de vazão e precipitação do rio Tinga na Estação Experimental de Itatinga.

Modelo: a vazão (l/s) do rio como variável resposta e precipitação (mm) como variável preditora.

Modelo ajustado: estimativas

$b_0 = 11.759777$;	$QMR = 23.3$;	$\overline{y} = 16.63134$;
$b_1 = 0.126446$;	$\overline{x} = 38.52675$;	$S_{xx} = 236739.1$.

Interpretação dos coeficientes de regressão: b_0 é a vazão do rio quando não ocorre precipitação = 11.8 l/s.

b_1 é a vazão do rio por mm de precipitação = 0.13 $(l/s)/mm$.

ICA de 90% para os coeficientes de regressão:

$$\begin{aligned} b_0 &\Rightarrow 11.8 \pm 2(0.535745) \Rightarrow 11.8 \pm 1.0(l/s) \\ b_1 &\Rightarrow 0.13 \pm 2(0.009931) \Rightarrow 0.13 \pm 0.018\frac{l/s}{mm} \end{aligned}$$

ICA de 90% para vazão média do rio no caso de precipitações de 50 mm:

$$\begin{aligned} \widehat{y} &= 18.08209 \\ &\Rightarrow 18.1 \pm 2(0.3919499) \Rightarrow 18.1 \pm 0.78(l/s) \end{aligned}$$

IPA de 90% para vazão do rio no caso de uma chuva (precipitação) de 50 mm:

$$\begin{aligned} \widehat{y}_h &= 18.08209 \\ &\Rightarrow 18.1 \pm 2(4.842894) \Rightarrow 18.1 \pm 9.7(l/s) \end{aligned}$$

IPA de 90% para vazão média do rio num mês que ocorram 3 chuvas (precipitações) de 50 mm:

$$\begin{aligned} \widehat{\overline{y}}_m &= 18.08209 \\ &\Rightarrow 18.1 \pm 2(2.814301) \Rightarrow 18.1 \pm 5.6(l/s) \end{aligned}$$

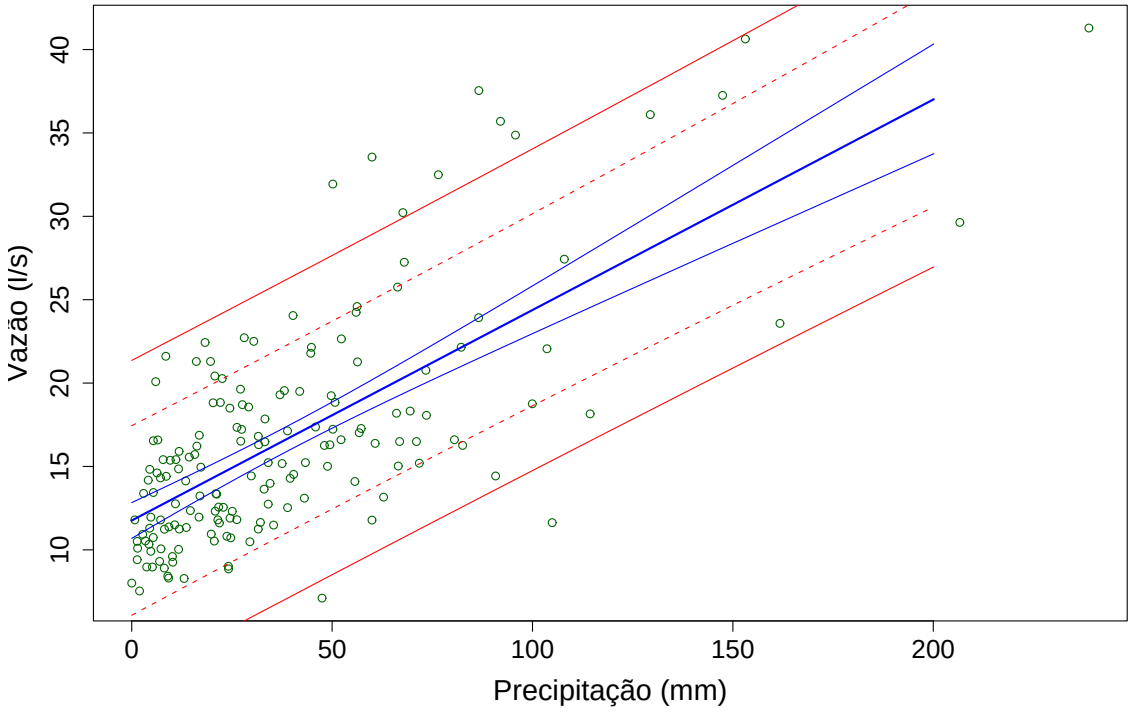


Figura 1.9: Gráfico ilustrando os ICAs e IPAs calculados acima.

1.4.9 Exercícios

1.4-1. Utilizando os dados do peixe esturjão (exercício 1.3-6, na página 1-19) realize as seguintes inferências com base no MLEAHI:

(a) ICA de 90% para os coeficientes de regressão;

(b) ICA de 90% para a resposta média, quando o comprimento dos peixes é de 140 cm.

(c) IPA de 90% para a produção de ovas de um peixe com comprimento de 140 cm.

(d) IPA de 90% para a produção média de ovas numa amostra de 5 peixes com comprimento de 140 cm.

1.4-2. Utilizando os dados do exemplo de relação de altura-DAP para *E. saligna*, realize as seguintes inferências com base no MLEAHI:

(a) ICA de 90% para os coeficientes de regressão;

(b) ICA de 90% para a altura média, das árvores com DAP de 17 cm.

(c) IPA de 90% para a altura de uma árvore com DAP de 17 cm.

(d) IPA de 90% para a altura média de uma amostra de 10 árvore com DAP de 17 cm.

1.4-3. Considere os resultados apresentados sobre a vazão do rio Tinga e discuta as situações abaixo. Estabeleça as suas conclusões.

Suponha que foi realizada uma colheita florestal na bacia do rio Tinga e, logo após a colheita, observou-se:

(a) uma vazão de 30 l/s relacionada a uma chuva de 50 mm;

(b) uma vazão média de 25 l/s relacionada a 3 ocorrências de precipitação de 50 mm.

Três anos após a colheita florestal foi observado:

(c) uma vazão de 7 l/s relacionada a uma chuva de 50 mm;

(d) uma vazão média de 14 l/s relacionada a 3 ocorrências de precipitação de 50 mm.

A colheita florestal influenciou a vazão do rio?

1.4-4. Uma pesquisadora florestal coordena um estudo da relação entre o logaritmo da riqueza de espécies de aves de subbosque em fragmentos florestais (variável resposta) e o tamanho dos fragmentos (variável preditora). Assumindo que a relação linear simples é apropriada para essas variáveis, deseja-se realizar inferências sobre o coeficiente de inclinação (β_1) do modelo. Foram selecionados originalmente fragmentos cujos tamanhos variam de 0.5 ha a 10 ha. Mas surgiu uma discussão no projeto:

- alguns pesquisadores argumentam que o estudo deve se deter nesses tamanhos de fragmentos e fazer obter um grande tamanho de amostra;
- outros pesquisadores argumentam que seria mais apropriado ampliar a faixa de tamanho de fragmentos (de 0.5 a 200 ha), mesmo que isso, devido ao maior esforço amostral, resultasse numa ligeira redução do tamanho da amostra.

Discuta a situação.

1.5 O Modelo Linear Simples de Erros Gaussianos (MLEG)

Os modelos lineares simples vistos até o momento tiveram um conjunto crescente de pressuposições que permitiram um aprofundamento da aplicação da análise de regressão aos dados. Acrescentaremos agora, a pressuposição que define uma distribuição estatística aos erros de modo a construir o "*modelo linear simples completo*".

1.5.1 Estrutura e Pressuposições do MLEG

A pressuposição a ser acrescentada é que os erros possuem distribuição Gaussiana (distribuição Normal).

A definição formal do MLEG (na notação sintética) é

$$\text{MLEG:} \qquad Y = \beta_0 + \beta_1 X + \varepsilon, \tag{1.15}$$

onde:

$$\begin{aligned} X &- \text{variável não estocástica} \\ \boldsymbol{E}\{\varepsilon|X=x\} &= 0; \\ \boldsymbol{Var}\{\varepsilon\} &= \sigma^2; \\ \boldsymbol{Cov}\{\varepsilon_i,\varepsilon_j\} &= 0; \\ (\varepsilon|X=x) &\sim N. \end{aligned}$$

Para tornar a notação ainda mais sintética utilizaremos a seguinte notação:

$$(\varepsilon|X=x) \overset{\text{iid}}{\sim} N\left(0,\sigma^2\right)$$

que significa:

> erros (para cada nível de X) **i**ndependentes **i**denticamente **d**istribuidos (iid) segundo distribuição Gaussiana (Normal - N) com média 0 e variância σ^2.

Considerações sobre a Pressuposição de Distribuição Gaussiana

Variável preditora não estocástica: valem as considerações já apresentadas para o MLEA.

Valor esperado nulo para o erro ($\boldsymbol{E}\{\varepsilon|X = x\} = 0$): valem as considerações e deduções já apresentadas para o MLEA.

Homocedasticidade: valem as considerações e deduções já apresentadas para o MLEAHI.

Independência: valem as considerações e deduções já apresentadas para o MLEAHI.

Distribuição Gaussiana: novamente podemos olhar os erros como resultado de erros do processo de mensuração ou como resultados da aleatoriedade da variável resposta.

- Do ponto de vista de erros de mensuração, já vimos que num bom processo de mensuração espera-se que os erros tenham média zero (ausência de erros sistemáticos), com variância constante e com distribuição simétrica à média. A distribuição Gaussiana é o "modelo perfeito" para uma distribuição dos erros.
- Sob o aspecto da aleatoriedade da variável resposta, pode-se argumentar que quando a variável resposta é o resultado de uma série muito grande de efeitos num processo natural complexo, a distribuição Gaussiana é um bom modelo. Muitas variáveis de carácter biológico tendem a ter distribuição Gaussiana, embora não seja possível generalizá-la para todas variáveis.
- Quando a variável resposta é resultado do total de observações individuais, como por exemplo o volume de madeira numa parcela, ou é resultado da média de observações individuais, como por exemplo a altura média das árvores, podemos evocar o Teorema Central do Limite para justificar a pressuposição de distribuição Gaussiana.

CONCEITOS BÁSICOS DE ESTATÍSTICA

Para prosseguirmos necessitamos dois conceitos fundamentais da estatística:

Distribuição Gaussiana e
Teorema Central do Limite

1.5.2 Propriedades dos Estimadores de Quadrados Mínimos e Inferência

Os EQM são os mesmos do MLEA e do MLEAHI (equações 1.3 e 1.4), que podem ser resumidamente apresentadas como

$$b_1 = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sum(x_i - \overline{x})^2} = \frac{S_{xy}}{S_{xx}} \qquad \text{e} \qquad b_0 = \overline{y} - b_1\overline{x}.$$

Já vimos que esses estimadores têm as seguintes propriedades:

1. EQM são funções lineares dos valores observados da variável resposta (MLEA).
2. EQM são estimadores não viciados dos coeficientes de regressão (MLEA).
3. As variâncias dos EQM são proporcionais à variância do erro (σ^2) (MLEAHI).
4. Os EQM são correlacionados (MLEAHI).
5. Estimadores de Variância Mínima (**Teorema de Gauss**, MLEAHI).

Ao assumirmos que os erros têm distribuição Gaussiana podemos acrescentar a essas propriedade:

6. Os EQM dos coeficientes de regressão têm distribuição Gaussiana.

CONCEITOS BÁSICOS DE ESTATÍSTICA

Para prosseguirmos necessitamos de um conceito de estatística:

Funções Lineares de Variáveis Gaussianas

Inferência sobre o Coeficiente de Inclinação

Sendo b_1 um EQM de β_1 ele terá distribuição Gaussiana com:

$$\begin{aligned} \text{média: } \boldsymbol{E}\{b_1\} &= \beta_1 \\ \text{variância: } \boldsymbol{Var}\{b_1\} &= \frac{\sigma^2}{\sum(x_i - \overline{x})^2}. \end{aligned}$$

Utilizando uma notação sintética:

$$b_1 \sim N\left(\beta_1; \frac{\sigma^2}{\sum(x_i - \overline{x})^2}\right). \tag{1.16}$$

Para estimarmos a variância de b_1 utilizamos o QMR no lugar da variância do erro (σ^2):

$$\widehat{\boldsymbol{Var}}\{b_1\} = \frac{QMR}{\sum(x_i - \overline{x})^2}$$

sendo o seu *erro padrão* dado por:

$$\boldsymbol{S}\{b_1\} = \sqrt{\frac{QMR}{\sum(x_i - \overline{x})^2}}.$$

A distribuição Gaussiana de b_1 implica em:

$$\begin{aligned} \frac{b_1 - \beta_1}{\sqrt{\boldsymbol{Var}\{b_1\}}} &\sim N(0;1) \\ \frac{b_1 - \beta_1}{\sqrt{\widehat{\boldsymbol{Var}}\{b_1\}}} &\sim t(n-2), \end{aligned}$$

onde $t(n-2)$ é da distribuição t de Student com parâmetro "*graus de liberdade*" $\nu = n - 2$.

Intervalos de Confinaça Exatos: para construirmos I.C. **exatos**, com coeficiente de confiança de $(1-\alpha)100\%$, podemos utilizar a distribuição t de Student. De acordo com ela temos:

$$P\left[b_1 - t(1-\alpha/2; n-2)\,\boldsymbol{S}\{b_1\} \leq \beta_1 \leq b_1 + t(\alpha/2; n-2)\,\boldsymbol{S}\{b_1\}\right] = 1 - \alpha,$$

sendo α o nível de significância adotado.

Assim, o I.C. de $(1-\alpha)100\%$ é

$$b_1 \pm t(1-\alpha/2; n-2)\,\boldsymbol{S}\{b_1\}$$

onde $t(1-\alpha/2;\ n-2)$ é o quantil $(1-\alpha/2)$ da distribuição t com $(n-2)$ graus de liberade.

Testes de Hipóteses: utilizando a mesma distribuição t de Student, podemos construir testes de hipósteses bilaterais do tipo:

$$\begin{aligned} \text{Hipótese Nula} \qquad & H_0 : \beta_1 = \theta_0 \\ \text{Hipótese Alternativa} \qquad & H_\alpha : \beta_1 \neq \theta_0 \end{aligned}$$

utilizando a estatística

$$t^* = \frac{b_1 - \theta_0}{\boldsymbol{S}\{b_1\}}$$

e aplicando, sob nível de significância (ou tamanho do teste) de α, a regressa de decisão:

- se $|t^*| \geq t(1-\alpha/2;\, n-2)$, rejeita H_0;
- se $|t^*| < t(1-\alpha/2;\, n-2)$, não rejeita H_0.

Hipótese unilaterais também podem ser testadas aplicando de forma análoga o teste t unilateral, embora sejam raramente utilizadas na análise de regressão.

CONCEITOS BÁSICOS DE ESTATÍSTICA

A clareza do apresentado acima depende da compreensão de conceitos básicos sobre:

- Testes Estatísticos de Hipótese;
- Teste t para uma amostra.

Inferência sobre o Coeficiente do Intercepto

Podemos desenvolver a inferência sobre o coeficiente do intercepto de modo análogo ao desenvolvido para o coeficiente da inclinação da reta de regressão.

Sendo b_0 um EQM de β_0 ele terá distribuição Gaussiana com:

$$\begin{aligned} \text{média: } \boldsymbol{E}\{b_0\} &= \beta_0 \\ \text{variância: } \boldsymbol{Var}\{b_0\} &= \sigma^2 \left[\frac{1}{n} + \frac{\overline{x}^2}{\sum(x_i - \overline{x})^2}\right]. \end{aligned}$$

Utilizando uma notação sintética:

$$b_0 \sim N\left(\beta_0; \sigma^2 \left[\frac{1}{n} + \frac{\overline{x}^2}{\sum(x_i - \overline{x})^2}\right]\right). \tag{1.17}$$

Para estimarmos a variância de b_0 utilizamos o QMR no lugar da variância do erro (σ^2):

$$\widehat{\boldsymbol{Var}}\{b_0\} = QMR\left[\frac{1}{n} + \frac{\overline{x}^2}{\sum(x_i - \overline{x})^2}\right].$$

sendo o seu *erro padrão* dado por:

$$\boldsymbol{S}\{b_0\} = \sqrt{QMR\left[\frac{1}{n} + \frac{\overline{x}^2}{\sum(x_i - \overline{x})^2}\right]}.$$

A distribuição Gaussiana de b_0 implica em:

$$\begin{aligned}\frac{b_0 - \beta_0}{\sqrt{\boldsymbol{Var}\{b_0\}}} &\sim N(0;1)\\ \frac{b_0 - \beta_0}{\sqrt{\widehat{\boldsymbol{Var}}\{b_0\}}} &\sim t(n-2),\end{aligned}$$

onde $t(n-2)$ é da distribuição t de Student com parâmetro "*graus de liberdade*" $\nu = n - 2$.

Intervalos de Confinaça Exatos: I.C. **exatos**, com coeficiente de confiança de $(1-\alpha)100\%$:

$$b_1 \pm t(1-\alpha/2; n-2)\,\boldsymbol{S}\{b_1\}.$$

Testes de Hipóteses: para testar hipósteses bilaterais:

Hipótese Nula	$H_0 : \beta_0 = \theta_0$
Hipótese Alternativa	$H_\alpha : \beta_0 \neq \theta_0$

utilizando a estatística

$$t^* = \frac{b_0 - \theta_0}{\boldsymbol{S}\{b_0\}}$$

com regressa de decisão (nível de significância de α):

- se $|t^*| \geq t(1-\alpha/2;\, n-2)$, rejeita H_0;
- se $|t^*| < t(1-\alpha/2;\, n-2)$, não rejeita H_0.

1.5.3 Valor Ajustado e Intervalos de Confiança e Predição

Assim como os EQM dos coeficientes de regressão, também foi demonstrado que o valor ajustado é uma função linear das observações da variável resposta.

Se assumirmos que o erro tem distribuição Gaussiana, isso resulta que os valores ajustados ($\widehat{y}_i$), como estimativas do valor esperado da variável resposta ($\boldsymbol{E}\{Y|X=x_i\}$), também terão distribuição Gaussiana com média igual a

$$\boldsymbol{E}\{\widehat{y}_i\} = \boldsymbol{E}\{Y|X=x_i\}\,.$$

Intervalo de Confiança para a Reposta Média

No caso da resposta média (linha de regressão), visto que a variância da estimativa era

$$\boldsymbol{Var}\{\widehat{y}_i\} = \sigma^2\left[\frac{1}{n} + \frac{(x_i-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right],$$

assim a distribuição da estimativa da resposta média é dada por:

$$\widehat{y}_i \sim N\left(\boldsymbol{E}\{Y|X=x_i\}\,;\sigma^2\left[\frac{1}{n} + \frac{(x_i-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right]\right). \tag{1.18}$$

O **erro padrão da estimativa** será obtido por

$$\boldsymbol{S}\{\widehat{y}_i\} = \sqrt{QMR\left[\frac{1}{n} + \frac{(x_i-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right]}$$

e Intervalos de Confiança **exatos**, com coeficiente de confiança de $(1-\alpha)100\%$, são obtidos por:

$$\widehat{y}_i \pm t(1-\alpha/2; n-2)\,\boldsymbol{S}\{\widehat{y}_i\}$$

Intervalo de Predição para uma Nova Observação

No caso de uma nova observação (predição), visto que a variância é obtida por

$$\boldsymbol{Var}\{\widehat{y}_h\} = \sigma^2\left[1 + \frac{1}{n} + \frac{(x_h-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right],$$

assim a distribuição de uma nova observação é dada por:

$$\widehat{y}_h \sim N\left(\boldsymbol{E}\left\{Y|X=x_h\right\}; \sigma^2\left[1+\frac{1}{n}+\frac{(x_h-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right]\right). \tag{1.19}$$

O **erro padrão de predição** é obtido por

$$\boldsymbol{S}\{\widehat{y}_h\} = \sqrt{QMR\left[1+\frac{1}{n}+\frac{(x_h-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right]}$$

e **Intervalos de Predição exatos**, com **probabilidade** de $(1-\alpha)100\%$, são obtidos por:

$$\widehat{y}_h \pm t(1-\alpha/2; n-2)\, \boldsymbol{S}\{\widehat{y}_h\}.$$

Intervalo de Predição para a Média Amostral de m Novas Observações

No caso da média de m novas observações, vimos que a variância esperada é

$$\boldsymbol{Var}\{\widehat{\overline{y}}_m\} = \sigma^2\left[\frac{1}{m}+\frac{1}{n}+\frac{(x_h-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right].$$

Logo, ela tem distribuição

$$\widehat{\overline{y}}_m \sim N\left(\boldsymbol{E}\left\{Y|X=x_h\right\}; \sigma^2\left[\frac{1}{m}+\frac{1}{n}+\frac{(x_h-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right]\right).$$

Com o **erro padrão de predição** pode ser calculado por

$$\boldsymbol{S}\{\widehat{\overline{y}}_m\} = \sqrt{QMR\left[\frac{1}{m}+\frac{1}{n}+\frac{(x_h-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right]},$$

Intervalos de Predição exatos, com **probabilidade** de $(1-\alpha)100\%$, são dados pela expressão

$$\widehat{\overline{y}}_m \pm t(1-\alpha/2; n-2)\, \boldsymbol{S}\{\widehat{\overline{y}}_m\}.$$

1.5.4 Intervalos Simultâneos e Bandas de Confiança

Todos intervalos relacionados com o valor esperado até o momento se referem a intervalos **individuais**, i.e., somente um único nível da variável preditora é considerado.

Para podermos considerar vários níveis da variável preditora simultaneamente devemos construir:

- *Intervalos Simulatâneos* quando o número de níveis é conhecido;
- *Bandas de Confiança* quando o número de níveis não é conhecido ou é muito grande.

Inferência Simultânea para Resposta Média

No caso da resposta média veremos dois métodos:

Método de Bonferroni: aplica a desiguladade de Bonferroni na construção de intervalos simultâneos *aproximados*.

DESIGUALDADE DE BONFERRONI

Suponha g eventos ($A_1, A_2, \ldots, A_g$), cada qual com probabilidade $P(A_i) = \gamma$, $i = 1, 2, \ldots, g$. A probabilidade da intersecção dos eventos complementares ($\overline{A}_1, \overline{A}_1, \ldots, \overline{A}_g$) segue a desigualdade:

$$P\left(\cap_{i=1}^{g} \overline{A}_i\right) \geq 1 - g\,\gamma,$$

com igualdade apenas quandos os eventos forem independentes.

No caso de estarmos interessados em g intervalos de confiança, mas desejamos manter o nível de significância dos intervalos simultâneos em α, i.e., o coeficiente de confiança em $(1 - \alpha)$, teremos

$$1 - \alpha \;=\; 1 - g\gamma \qquad \Rightarrow \qquad \gamma = \frac{\alpha}{g}.$$

Para construirmos g intervalos simultâneos com nível de significância α, cada intervalo individual terá que ter nível de significância α/g.

Os g I.C. simultâneos de Bonferroni, portanto, são definidos como:

$$\widehat{y}_i \pm B_g\, \boldsymbol{S}\{\widehat{y}_i\}$$

onde:

$$B_g = t\left(1 - \frac{\alpha}{2g}; n - 2\right)$$

O método de Bonferroni é **conservador**, i.e., o coeficiente de confiança verdadeiro é sempre maior que o valor nominal.

À medida que o número de inferências simultâneas aumenta, os intervalos de Bonferroni rapidamente se tornam extremamente amplos e deixam de ser úteis na prática.

Número de Intervalos Simultâneos (g)	Coeficiente de Confiança em cada Intervalo	B_g $t(1-\frac{\alpha}{2g};8)$	Tamanho Relativo de B_g (%)
1	0.9500	2.306	100
5	0.9900	3.355	146
10	0.9950	3.833	166
20	0.9975	4.334	188
50	0.9990	5.041	219
100	0.9995	5.617	244

Método de Working-Hotelling: constroe uma *Banda de Confiança* para a linha de regressão (resposta média), considerando *todos os níveis possíveis* da variável preditora:

$$\widehat{y}_i \pm W \, \boldsymbol{S}\{\widehat{y}_i\}$$

onde

$$W = \sqrt{2\, F(1-\alpha; 2; n-2)}$$

sendo utilizado o quantil $(1-\alpha)$ da distribuição F com 2 graus de liberdade no numerador e $n-2$ graus de liberdade no denominador.

Para um número pequeno de intervalos simultâneos o método de Bonferroni gera intervalos menores, mas à medida que o número de intervalos aumenta, a B.C. de Working-Hotelling tende a ser menor que os I.C. simultâneos pelo método de Bonferroni.

Número de Níveis de X g	$\alpha = 0.05; n-2 = 8$ Bonferroni B	Working-Hotteling W	B/W
1	2.306	2.986	0.77
2	2.752	2.986	0.92
3	3.016	2.986	1.01
4	3.206	2.986	1.07
5	3.355	2.986	1.12
10	3.833	2.986	1.28
20	4.334	2.986	1.45

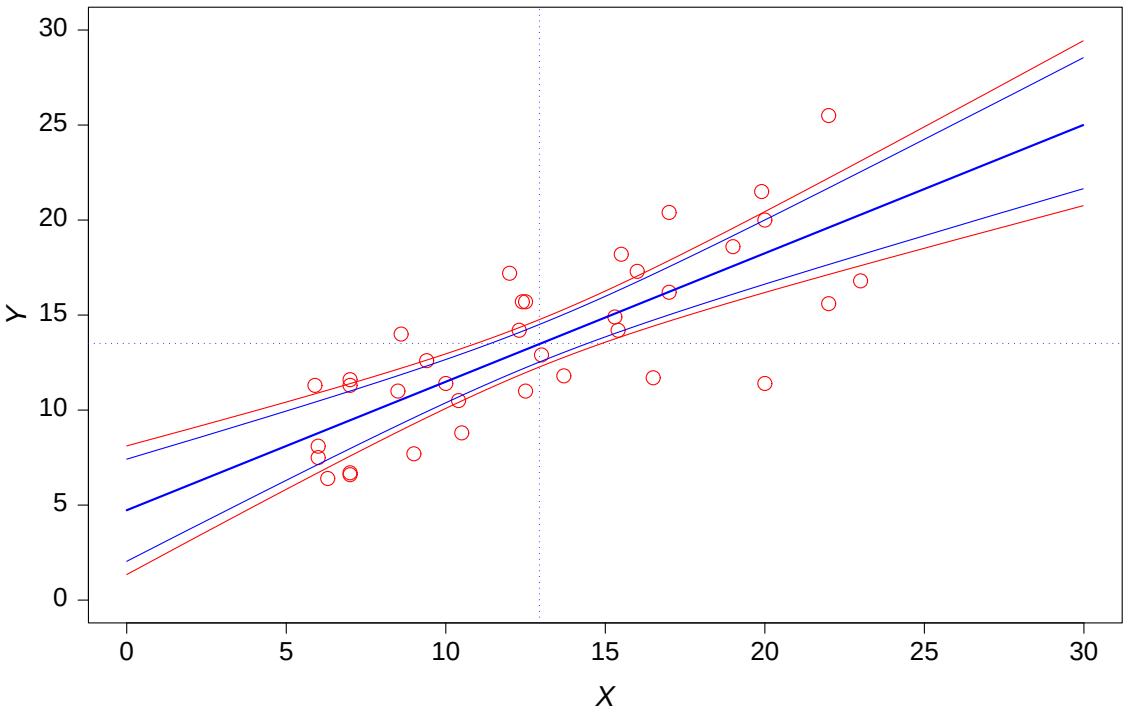


Figura 1.10: Gráfico comparando a Banda de Confiança de 95% (em vermelho) para a resposta média (construida pelo método Working-Hotelling) e Intervalos de Confiança individuais (em azul).

Em comparação a I.C. individuais, a Banda de Confiança tende sempre a ser um pouco mais larga.

Inferência Simultânea para uma Nova Observação

Método de Bonferroni: também pode ser utilizado nos Intervalos de Predição Simultâneos. Para g I.P. simultâneos temos:

$$\widehat{y}_h \pm B_g \, \boldsymbol{S}\{\widehat{y}_h\}$$

onde:

$$B_g = t\left(1 - \frac{\alpha}{2g}; n-2\right).$$

Permanece o problema de serem muito conservadores.

1.5.5 Verificando a Qualidade do Ajuste do Modelo

Embora a pressuposição de erros Gaussianos não seja essencial para estudar a qualidade do ajuste dos modelos lineares, ela permite a realização de testes formais.

A variabilidade da variável resposta ao redor da sua média pode ser vista como formada por dois componentes independentes:

$$\begin{aligned} S_{yy} &= \sum(y_i-\overline{y})^2 = \sum\left[y_i-\overline{y}+\widehat{y}_i-\widehat{y}_i\right]^2 = \sum\left[(\widehat{y}_i-\overline{y})+(y_i-\widehat{y}_i)\right]^2 \\ &= \sum\left[(\widehat{y}_i-\overline{y})^2+2(\widehat{y}_i-\overline{y})(y_i-\widehat{y}_i)+(y_i-\widehat{y}_i)^2\right] \\ &= \sum(\widehat{y}_i-\overline{y})^2+2\sum(\widehat{y}_i-\overline{y})(y_i-\widehat{y}_i)+\sum(y_i-\widehat{y}_i)^2 \\ &= \sum(\widehat{y}_i-\overline{y})^2+2\left[\sum\widehat{y}_i(y_i-\widehat{y}_i)+\sum\overline{y}(y_i-\widehat{y}_i)\right]+\sum(y_i-\widehat{y}_i)^2 \\ &= \sum(\widehat{y}_i-\overline{y})^2+2\left[\sum\widehat{y}_ie_i+\overline{y}\sum e_i\right]+\sum(y_i-\widehat{y}_i)^2 \\ &= \sum(\widehat{y}_i-\overline{y})^2+\sum(y_i-\widehat{y}_i)^2 \end{aligned}$$

$$SQT = SQM + SQR$$

onde:

SQT - Soma de Quadrados Total: variabiliade dos valores observados y_i em relação à média amostral, variabilidade "*total*" da variável resposta.

Fórmula facilitada de cálculo:

$$S_{yy} = \sum(y_i-\overline{y})^2 = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}$$

SQM - Soma de Quadrados do Modelo: variabilidade dos valores ajustados ao redor da média amostral, variabilidade "*explicada*" pelo modelo.

Essa variabilidade explicada é a parte *sistemática* explicada pela variável preditora: $\beta_0+\beta_1x_i$.

Fórmulas facilitadas de cálculo:

$$\begin{aligned} SQM = \sum(\widehat{y}_i-\overline{y})^2 &= b_1^2\sum(x_i-\overline{x})^2 = b_1^2\left[\sum x_i^2-\frac{\left(\sum x_i\right)^2}{n}\right] \\ &= b_1\sum(x_i-\overline{x})(y_i-\overline{y}) = b_1\left[\sum x_iy_i-\frac{\sum x_i\ \sum y_i}{n}\right] \end{aligned}$$

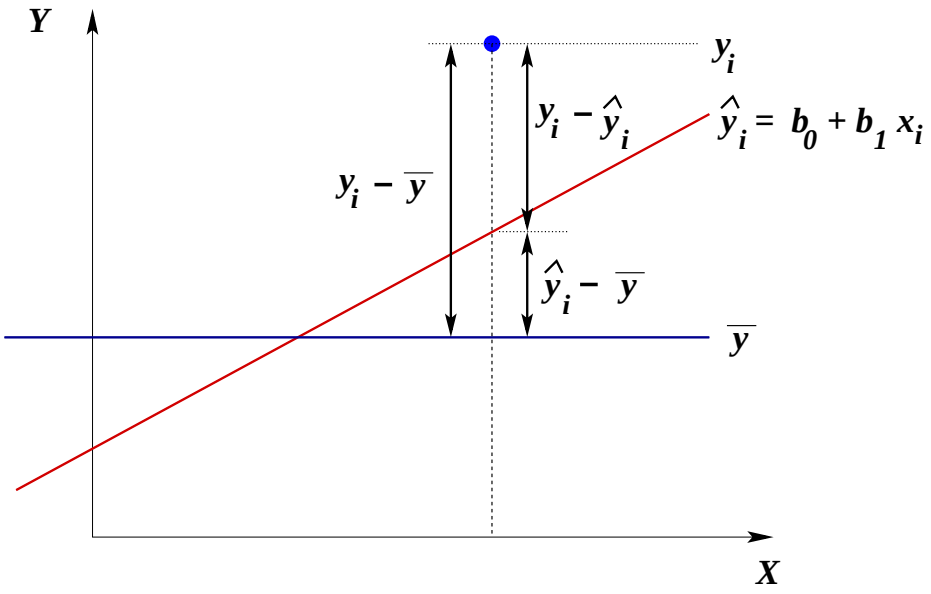


Figura 1.11: Gráfico esquemático mostrando como a Soma de Quadrados Total (SQT) é particionada em Soma de Quadrados do Modelo (SQM) e Soma de Quadrados do Resíduo (SQR).

SQR - Soma de Quadrados do Resíduo: variabilidade dos valores observados y_i em relação aos valores ajustados $\widehat{y}_i$, variabilidade "*não explicada*" pelo modelo.

Essa variabilidade é a parte *aleatória* que resulta do termo do erro: ε_i.

Fórmula facilitada de cálculo:

$$SQR = SQT - SQM$$

Coeficiente de Determinação

Uma forma natural de se quantificar a qualidade do ajuste do modelo é verificando o tamanho relativo da SQM em relação à SQT. A razão dessas somas de quadrado é chamada de *Coeficiente de Determinação*:

$$r^2 = \frac{SQM}{SQT} = 1 - \frac{SQR}{SQT}, \tag{1.20}$$

embora seja freqüentemente mais prático calculá-lo por

$$r^2 = b_1^2 \frac{\sum (x_i - \overline{x})^2}{\sum (y_i - \overline{y})^2}.$$

- Note que $0 \leq r^2 \leq 1$.
- O r^2 é interpretado como:

  a proporção da variação da variável resposta (Y) explicada pelo modelo linear, i.e., pela variável preditora (X).

- Não existe uma escala absoluta para analisar o r^2, os limites para os quais se considera o ajuste "*bom*" dependem muito do tipo de dados e situação em que se aplica a regressão.

  Por exemplo, para equações de volume de árvores individuais em florestas plantadas espera-se r^2 superiores a 0.90. Já para modelos de produção de florestas plantadas é comum se considerar bons valores de r^2 acima de 0.70 ou 0.60.

Coeficiente de Correlação

Uma outra medida comumente utilizada em modelos lineares simples é o *Coeficiente de Correlação*:

$$r = \sqrt{r^2} = \sqrt{b_1^2 \frac{\sum(x_i - \overline{x})^2}{\sum(y_i - \overline{y})^2}} = b_1 \sqrt{\frac{\sum(x_i - \overline{x})^2}{\sum(y_i - \overline{y})^2}}.$$

- O r quantifica o grau de associação *linear* entre variável resposta e preditora.
- O r possui amplitude $-1 \leq r \leq 1$, tal que:

  $r \approx -1$: alta correlação negativa;

  $r \approx 0$: ausência de correlação;

  $r \approx 1$: alta correlação positiva.

- O sinal r que define correlação positiva ou negativa é definido pelo sinal de b_1.

Teste F do Modelo

Uma outra forma de estudarmos a qualidade do ajuste é verificando o que as somas de quadrado *estimam*.

Já vimos que

$$\boldsymbol{E}\{QMR\} = \boldsymbol{E}\left\{\frac{SQR}{n-2}\right\} = \sigma^2,$$

logo a SQR corrigida para os graus de liberdade ($n-2$) é um estimador da variância do erro.

No caso da SQM temos:

$$\begin{aligned}
\boldsymbol{E}\{SQM\} &= \boldsymbol{E}\left\{b_1^2 \sum (x_i-\overline{x})^2\right\} = \left(\sum (x_i-\overline{x})^2\right) \boldsymbol{E}\left\{b_1^2\right\} \\
&= \left(\sum (x_i-\overline{x})^2\right) \left[\boldsymbol{Var}\{b_1\} + (\boldsymbol{E}\{b_1\})^2\right] \\
&= \left(\sum (x_i-\overline{x})^2\right) \left[\frac{\sigma^2}{\sum (x_i-\overline{x})^2} + \beta_1^2\right] \\
&= \sigma^2 + \beta_1^2 \sum (x_i-\overline{x})^2
\end{aligned}$$

Note que se ao associarmos um *Quadrado Médio do Modelo* à variabilidade explicada pelo modelo linear simples temos:

$$QMM = \frac{SQM}{\text{número de parâmetros no modelo} - 1} = \frac{SQM}{1} = SQM,$$

e, portanto,

$$\boldsymbol{E}\{QMM\} = \boldsymbol{E}\{SQM\} = \sigma^2 + \beta_1^2 \sum (x_i-\overline{x})^2.$$

Note ainda que

- σ^2 é a variância do erro (não explicada pelo modelo); e
- $\beta_1^2 \sum (x_i-\overline{x})^2$ é a variabilidade explicada pelo modelo.

A razão desses quadrados médios é uma medida da relevância da variabilidade explicada pelo modelo comparada a variabilidade do erro:

$$\begin{aligned}
F^* &= \frac{QMM}{QMR} = \frac{SQM/1}{SQR/(n-2)} \\
F^* &= \frac{\frac{SQM/\sigma^2}{1}}{\frac{SQR/\sigma^2}{n-2}} = \frac{\chi^2(1)}{1} \div \frac{\chi^2(n-2}{n-2}.
\end{aligned}$$

Sendo essa estatística uma razão de variáveis Qui-Quadrado, ele remete à distribuição F.

Ela pode ser utilizada para testar as hipóteses:

$$\begin{aligned}
H_0: &\quad \beta_1 = 0 \\
H_\alpha: &\quad \beta_1 \neq 0.
\end{aligned}$$

Sob H_0 teríamos

$$F^* \approx \frac{\sigma^2 + (0)\sum(x_i - \overline{x})^2}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1,$$

e, portanto,

$$F^* \sim \text{Distribuição } F\left(\nu_1 = 1; \nu_2 = n-2\right).$$

A regra de decisão para se testar as hipóteses acima, no nível de significância α, ficam:

- se $F^* \geq F(1-\alpha; 1; n-2)$ rejeita-se H_0; e
- se $F^* < F(1-\alpha; 1; n-2)$ não se rejeita H_0.

Algumas considerações sobre o teste F do modelo:

- *No modelo linear simples*, o teste F do modelo é equivalente ao teste t para as hipóteses:

  $$\begin{aligned} H_0: &\quad \beta_1 = 0 \\ H_\alpha: &\quad \beta_1 \neq 0. \end{aligned}$$

  De fato, pode se demonstrar que

  $$t(n-2) = \sqrt{F(1; n-2)}.$$

- Existe uma relação direta em o R^2 e o teste F do modelo:

  $$\begin{aligned} R^2 = \frac{SQM}{SQT} \quad &\Rightarrow \quad \frac{1}{R^2} = \frac{SQM + SQR}{SQM} = 1 + \frac{SQR}{SQM} \\ &\Rightarrow \quad \frac{1}{R^2} - 1 = \frac{SQR}{SQM} \Rightarrow \frac{R^2}{1-R^2} = \frac{SQM}{SQR} \\ &\Rightarrow \quad \left(\frac{n-p}{p-1}\right)\frac{R^2}{1-R^2} = \left(\frac{n-p}{p-1}\right)\frac{SQM}{SQR} = \frac{SQM/(p-1)}{SQR/(n-p)} = F \\ &\Rightarrow \quad F = \left(\frac{n-p}{p-1}\right)\frac{R^2}{1-R^2} \end{aligned}$$

  onde $p = 2$ é o número de coeficientes de regressão do modelo linear simples.

  Essa relação mostra que o teste F do modelo pondera a proporção da variabilidade explicada pelo modelo (R^2) e a não explicada ($1 - R^2$), pelos respectivos graus de liberdade ($p - 1$ e $n - p$).

- A interpretação do teste F é distinta da do R^2:

  > *ao se rejeitar H_0, conclui-se que exite uma relação linear entre variável resposta (Y) e variável preditora (X).*

1.5.6 Considerações sobre a Aplicação do MLEG

- O grau de associação *aceitável* entre variável resposta e variável preditora depende:
  - da situação em que se aplica o modelo,
  - do que se espera do modelo.
- O R^2 é enfatizado no caso de se utilizar o modelo para *predição*, pois é um índice descritivo da qualidade das predições que serão realizadas.
- O teste F do modelo é enfatizado nos caso em que se busca o modelo como uma *explicação* para relação entre Y e X. Teste F significativo indica a existência de relação *linear* entre Y e X.
- No caso de amostras de pequeno tamanho (n pequeno) é comum se observar:
  - teste F significativo;
  - valor de R^2 muito baixo.

  Geralmente, amostras pequenas podem ser suficientes para comprovar a existência de relação linear entre Y e X, mas são insuficientes para gerar um modelo com boa capacidade preditiva.
- Em situações onde se utiliza uma amostra **muito** pequena, às vezes se observa
  - teste F não significativo;
  - valor de R^2 alto.

  Quando isso ocorre, pode se concluir que o tamanho da amostra é insuficiente para qualquer aplicação, pois:
  - teste F não significativo em amostras pequenas indica *incapacidade* para detectar a relação linear (caso ela exista);
  - valor de R^2 alto em amostras pequenas indica que a SQT foi mal estimada (subestimada), pois a amplitude da variável resposta está mal representada na amostra.
- Em algumas áreas é como se utilizar uma estatística descritiva da qualidade do ajuste chamada *erro padrão da estimativa* ou *erro padrão do resíduo* que é simplesmente:

  $$\boldsymbol{S}\{\widehat{y}\} = \sqrt{QMR}.$$

  Trata-se de uma medida *geral* da incerteza associada a linha de regressão.

1.5.7 Exemplo de Aplicação 1: Relação Vazão-Precipitação no rio Tinga

- O modelo ajustado fica:

$$\widehat{y}_i = 11.759777 + 0.126446\, x_i.$$

- Os erros padrões das estimativas dos coeficientes de regressão:

$$\boldsymbol{S}\{b_0\} = 0.535745 \qquad \boldsymbol{S}\{b_1\} = 0.009931$$

- Os I.C. de 95% para as estimativas dos coeficientes de regressão:

  t **tabelado:** $t(0.975, 164) = 1.974535$;

  b_0**:** $11.759777 \pm 1.974535(0.535745) = 11.76 \pm 1.06$;

  b_1**:** $0.126446 \pm 1.974535(0.009931) = 0.13 \pm 0.02$.

- Teste t para as hipóteses:

$$H_0: \quad \beta_k = 0 \qquad \text{contra} \qquad H_\alpha: \beta_k \neq 0.$$

$$b_0: \quad t^* = \frac{11.759777}{0.535745} = 21.95, \qquad |21.95| \geq 1.974535 \Rightarrow \text{ rejeita } H_0.$$

$$b_1: \quad t^* = \frac{0.126446}{0.009931} = 12.73, \qquad |12.733| \geq 1.974535 \Rightarrow \text{ rejeita } H_0.$$

- I.C. individuais de 95% para resposta média nos níveis de 50 e 100 mm de precipitação:

  $\widehat{y}_{50}$**:** $18.08209 \pm 1.974535(0.3919499) = 18.08 \pm 0.77$;

  $\widehat{y}_{100}$**:** $24.40441 \pm 1.974535(0.7164615) = 24.40 \pm 1.42$.

- I.C. simultâneos de 95% para resposta média nos níveis de 50 e 100 mm de precipitação:

  Bonferroni: $t(1 - 0.05/(2 * 2); 164) = t(0.9875; 164) = 2.262168$

  $\widehat{y}_{50}$**:** $18.08209 \pm 2.262168(0.3919499) = 18.08 \pm 0.89$;

  $\widehat{y}_{100}$**:** $24.40441 \pm 2.262168(0.7164615) = 24.40 \pm 1.62$.

- I.P. individuais de 95% para uma nova observação nos níveis de 50 e 100 mm de precipitação:

$\widehat{y}_{50}$**:** $18.08209 \pm 1.974535(4.847356) = 18.08 \pm 9.57$;

$\widehat{y}_{100}$**:** $24.40441 \pm 1.974535(4.884317) = 24.40 \pm 9.64$.

- I.P. simultâneos de 95% para resposta média nos níveis de 50 e 100 mm de precipitação:

  Bonferroni: $t(1 - 0.05/(2 * 2); 164) = t(0.9875; 164) = 2.262168$

  $\widehat{y}_{50}$**:** $18.08209 \pm 2.262168(4.847356) = 18.08 \pm 10.97$;

  $\widehat{y}_{100}$**:** $24.40441 \pm 2.262168(4.884317) = 24.40 \pm 11.05$.

- Coeficiente de Determinação: 0.4971, embora não muito alto é razoável para esse tipo de relação.

- Teste F do modelo:

  $$\left.\begin{array}{l} F^* = 162.1 \\ F(0.95; 1; 164) = 3.898787 \end{array}\right\} \Rightarrow 162.1 > 3.8988 \Rightarrow \text{ rejeita-se } H_0$$

  Note que:

  $$\left[\begin{array}{ccc} t(0.975, 164) & = & \sqrt{F(0.95; 1; 164)} \\ 1.974535 & = & \sqrt{3.898787} \end{array}\right] \qquad \left[\begin{array}{ccc} t^*(H_0 : \beta_1 = 0) & = & \sqrt{F^*} \\ 12.73 & = & \sqrt{162.13} \end{array}\right]$$

- Erro padrão da estimativa: 4.832 l/s.

  Considerando a vazão média nos dados (16.63 l/s), verifica-se que o modelo não está bom para se fazer de predições, como os intervalos de predição já mostraram.

1.5.8 Exercícios

1.5-1. Utilizando os dados do peixe esturjão (exercício 1.3-6, na página 1-19) realize as seguintes inferências com base no MLEG:

(a) IC de 90% para os coeficientes de regressão;

(b) IC de 90% para a resposta média, quando o comprimento dos peixes é de 140 cm.

(c) IP de 90% para a produção de ovas de um peixe com comprimento de 140 cm.

(d) IP de 90% para a produção média de ovas numa amostra de 5 peixes com comprimento de 140 cm.

1.5-2. Compare os intervalos *exatos* do exercício anterior como os intervalos *aproximados* do exercício 1.4-1, na página 1-40.

1.5-3. Utilizando os dados do exemplo de relação de altura-DAP para *E. saligna*, realize as seguintes inferências com base no MLEG:

(a) IC de 95% para os coeficientes de regressão;

(b) IC de 95% para a altura média, das árvores com DAP de 17 cm.

(c) IP de 95% para a altura de uma árvore com DAP de 17 cm.

(d) IP de 95% para a altura média de uma amostra de 10 árvore com DAP de 17 cm.

(e) Banda de Confiança de 95% para a resposta média.

(f) Intervalos de Predição Simultâneos de 95% para altura de árvores com DAP 10, 17, 20 e 23 cm.

1.5-4. Discuta a qualidade do ajuste do MLEG, em termos de Coeficiente de Determinação e teste F do modelo, para os seguintes dados:

(a) Relação altura DAP para árvores de *E. saligna*.

(b) Produção de ovas do esturjão.

(c) Vazão do rio Tinga, considere as estimativas:

$b_0 = 11.759777$;	$QMR = 23.3$;	$S_{xx} = 236739.1$;
$b_1 = 0.126446$;	$\overline{x} = 38.52675$;	$S_{yy} = 68325.07$;
$n = 168$;	$\overline{y} = 16.63134$;	$S_{xy} = 37489.2$.

CAPÍTULO 2

A ABORDAGEM DE VEROSSIMILHANÇA NO MODELO LINEARE SIMPLE

- Até o momento apresentamos os estimadores de quadrados mínimos como forma de ajuste dos modelos lineares simples.
- Apresentaremos agora, para o MLEG, a abordagem de verossimilhança que permite um entendimento mais amplo dos modelos lineares e permitirá a visualização de possíveis formas mais gerais desses modelos.
- Faremos, inicialmente, uma revisão do conceito de verossimilhança e dos estimadores de máxima verossimilhança.
- Depois, desenvolveremos os conceitos no caso do modelo linear simples (MLEG).

2.1 O Conceito de Verossimilhança

- O termo "*likelihood*" no inglês, que tecnicamente no português é traduzido por "*verossimilhança*", foi cunhado por Ronald Fisher para distinguir esse conceito do conceito de probabilidade.
- A diferença pode parecer sutil, mas tem sérias implicações na aplicação dos modelos estatísticos.
- Tomemos como exemplo uma variável aleatória Y com distribuição Poisson. A função de probabilidade dessa variável é

$$P(Y = y) = \frac{\mathrm{e}^{-\mu}\mu^{y}}{y!}$$

onde μ é o parâmetro da distribuição, representando o valor esperado de Y ($\boldsymbol{E}\{Y\} = \mu$).

- Se assumirmos que o valor do parâmetro é conhecido, por exemplo $\mu = 23$, fica simples calcular a probabilidade de ocorrência de um valor de Y, por exemplo $Y = 24$:

$$P(Y = 24) = \frac{\mathrm{e}^{-\mu}\mu^{24}}{24!} = \frac{\mathrm{e}^{-23}23^{24}}{24!} = 0.00919136$$

- Na aplicação desse modelo, no entanto, a ocorrência de $Y = 24$ é obtida através de observações, mas o valor do parâmetro será sempre desconhecido.
- Na verdade, nosso interesse será estimar o parâmetro a partir da observação.
- Nessa situação, a função de probabilidade passa a depender dos valores plausíveis para o parâmetro, uma vez que o valor da observação ($Y = 24$) é conhecido.
- A função de probabilidade se torna *função de verossimilhança*:

$$\mathcal{L}\{\mu|X = 24\} = \frac{\mathrm{e}^{-\mu}\mu^{24}}{24!}$$

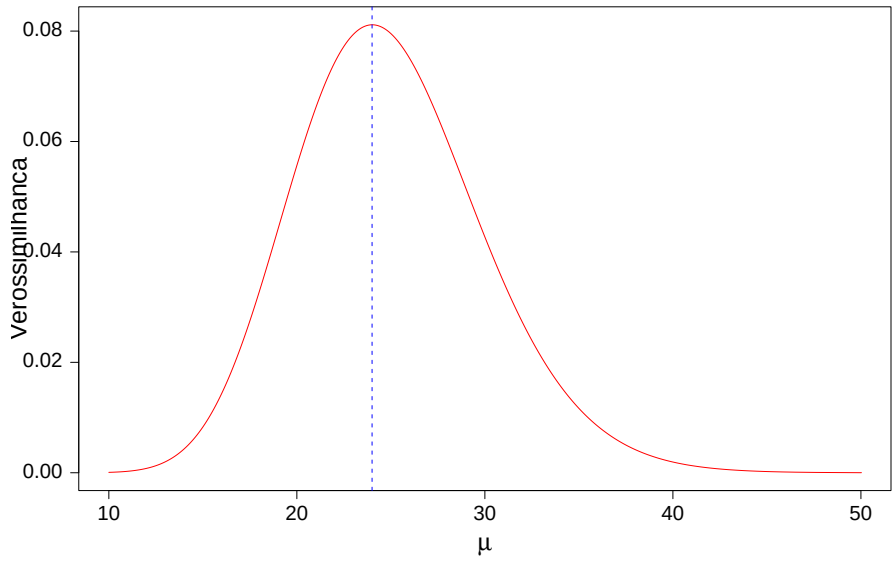


Figura 2.1: Função de verossimilhança para uma variável Poisson com observação $Y = 24$.

2.1.1 A Função de Verossimilhança na Distribuição Gaussiana

- O conceito de função de verossimilhança se aplica a qualquer distribuição estatística.
- No caso de variáveis contínuas, a função de verossimilhança é obtida a partir da *função de densidade probabilística* .
- Vejamos o caso da distribuição Gaussiana, com apenas uma observação $Y = y$:

$$\mathcal{L}\{\mu, \sigma | Y = y\} = \left(2\pi\sigma^2\right)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}\left(y - \mu\right)^2\right]$$

- Note que nesse caso temos dois parâmetros, de modo que o gráfico da verossimilhança tem que ser estudado em três dimensões:

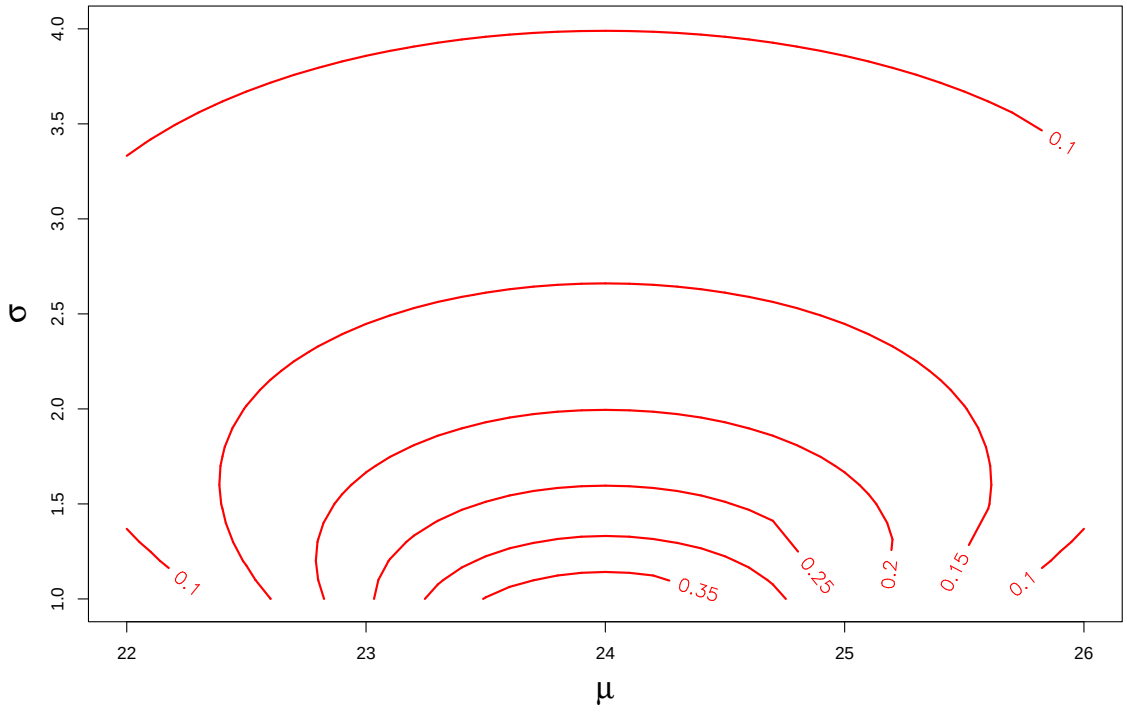


Figura 2.2: Função de verossimilhança para uma variável Gaussiana com uma única observação $Y = 24$.

- Esse gráfico gera dois comentários:

1. À medida que o número de parâmetros no modelo aumenta, aumentam as dimensões da função de verossimilhança.

   No caso da distribuição Gaussiana ela é bivariada: média e variância.

2. Como é possível se inferir sobre dois parâmetros tendo apenas uma observação?

   Não é possível!!! Na prática isso não faz sentido e o gráfico acima é totalmente equivocado.

2.1.2 A Função de Verossimilhança com Múltiplas Observações Independentes

- Voltemos ao exemplo da distribuição Poisson, e consideremos que temos agora em mão uma amostra com 3 observações independentes: $\mathbf{Y}_n = y_1, y_2, \ldots, y_n$.
- Assumindo que $\mu = 24$, qual a probabilidade de observarmos a amostra que temos:

$$P(\mathbf{Y}_n|\mu) = P(Y = y_1|\mu)\, P(Y = y_2|\mu)\, \ldots\, P(Y = y_n|\mu) = \prod_{i=1}^{n} P(Y = y_i|\mu)$$

- De forma análoga, a função de verossimilhança assume a forma de um produtório no caso de múltiplas observações independentes:

$$\mathcal{L}\{\mu|\mathbf{Y}_n\} = \mathcal{L}\{\mu|Y = y_1\}\, \mathcal{L}\{\mu|Y = y_2\}\ \ldots\, \mathcal{L}\{\mu|Y = y_n\} = \prod_{i=1}^{n} \mathcal{L}\{\mu|Y = y_i\}$$

- No caso de variável contínua, a função de verossimilhança gerada por n observações independentes é o produtório das funções de densidade:

$$\begin{aligned} \mathcal{L}\{\mu, \sigma|\mathbf{Y}_n = y_1, y_2, \ldots, y_n\} &= \prod_{i=1}^{n} \left(2\pi\sigma^2\right)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}\left(y_i - \mu\right)^2\right] \\ \mathcal{L}\{\mu, \sigma\} &= \left(2\pi\sigma^2\right)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \mu\right)^2\right] \end{aligned}$$

Função de Log-Verossimilhança Negativa

- A função de verossimilhança para múltiplas observações é de tratamento matemático complicado pois envolve o produtório da função de densidade para cada observação.

Uma vez que os valores da função de densidade são sempre menores que um, a função de verossimilhança rapidamente se aproxima do zero, o que é problemático no uso de computadores.

- Para facilitar o tratamento matemático e as operações de cálculos, trabalhamos com a *função de log-verossimilhança negativa* :

$$\mathbf{L}\{\mu, \sigma|\mathbf{Y}_n\} = -\ln\left(\mathcal{L}\{\mu, \sigma|\mathbf{Y}_n\}\right)$$

- No caso da distribuição Gaussiana temos:

$$\begin{aligned}
\mathbf{L}\{\mu, \sigma|\mathbf{Y}_n\} &= -\ln\left\{\left(2\pi\sigma^2\right)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right]\right\} \\
&= -\left\{-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right\} \\
\mathbf{L}\{\mu, \sigma|\mathbf{Y}_n\} &= \frac{n}{2}\ln(2\pi) + \frac{n}{2}\ln(\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2
\end{aligned}$$

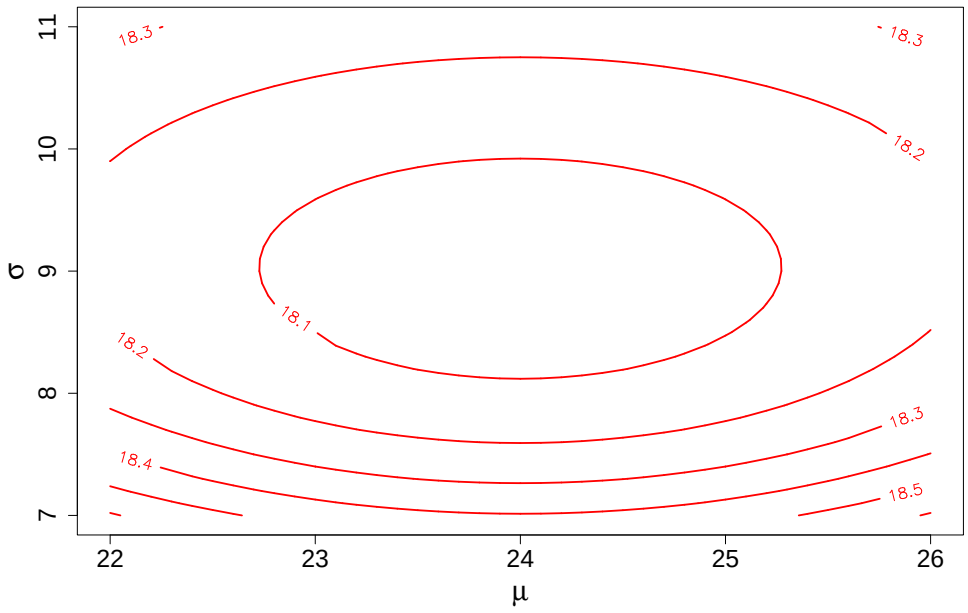


Figura 2.3: Função de log-verossimilhança negativa para uma variável Gaussiana com múltiplas observações.

2.2 Estimadores de Máxima Verossimilhança (EMV)

- Os Estimadores de Máxima Verossimilhança são aqueles que:
  - maximizam a função de verossimilhança, ou
  - minimizam a função de log-verossimilhança negativa.
- Como trabalharemos com a função de log-verossimilhança negativa, vejamos os passos para se encontrar o ponto de mínimo de uma função:
  1. encontrar a primeira derivada em relação ao parâmetro;
  2. igualar a derivada a zero e solucionar para o parâmetro desejado;
  3. verificar se a segunda derivada é positiva.

2.2.1 Exemplo: Distribuição Exponencial

- Como primeiro exemplo, analisaremos uma distribuição com único parâmetro.
- Y é uma variável aleatória contínua com distribuição Exponencial e, portanto, com função de densidade probabilística

$$f_Y(y|\lambda) = \lambda \exp(-\lambda y) \qquad \text{onde} \quad y \geq 0, \lambda > 0.$$

- A função de verossimilhança fica

$$\mathcal{L}\{\lambda|\mathbf{Y}_n\} = \prod_{i=1}^{n} \lambda \exp(-\lambda y_i) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} y_i\right).$$

- A função de log-verossimilhança negativa é

$$\mathbf{L}\{\lambda\} = -n \ln \lambda + \lambda \sum_{i=1}^{n} y_i.$$

- Igualando a primeira derivada a zero:

$$\frac{d\mathbf{L}\{\lambda\}}{d\lambda} = -\frac{n}{\lambda} + \sum_{i=1}^{n} y_i = 0.$$

- E, portanto, o EMV de λ é:

$$\widehat{\lambda} = \frac{n}{\sum_{i=1}^{n} y_i}.$$

- Como a segunda derivada é positiva:

$$\frac{d^2\mathbf{L}\{\lambda\}}{d\lambda^2} = \frac{n}{\lambda^2},$$

concluí-se que $\widehat{\lambda}$ é de fato o EMV, pois minimiza $\mathbf{L}\{\lambda\}$.

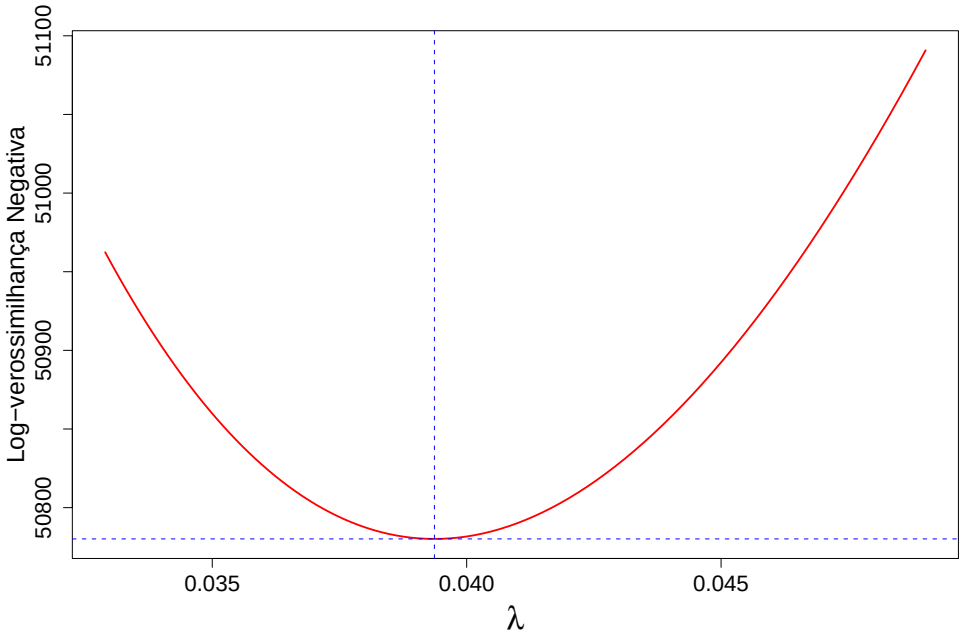


Figura 2.4: Função de verossimilhança para o parâmetro λ da distribuição Exponencial, aplicada a dados de DAP de 11.991 árvores de floresta nativa no Maranhão.

2.2.2 Exemplo: Distribuição Gaussiana

- Y é uma variável aleatória contínua com distribuição Gaussiana e, portanto, com função de densidade probabilística

$$f_Y(y|\mu,\sigma^2) = \left(2\pi\sigma^2\right)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2\right]$$

onde $-\infty < y < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0.$

- A função de verossimilhança fica

$$\begin{aligned}\mathcal{L}\{\mu,\sigma^2|\mathbf{Y}_n\} &= \prod_{i=1}^{n}\left(2\pi\sigma^2\right)^{-1/2}\exp\left[-\frac{1}{2\sigma^2}(y_i-\mu)^2\right] \\ &= \left(2\pi\sigma^2\right)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right]\end{aligned}$$

- A função de log-verossimilhança negativa é

$$\mathbf{L}\{\mu,\sigma^2\} = \frac{n}{2}\ln(2\pi) + \frac{n}{2}\ln(\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2$$

- Como temos agora dois parâmetros, igualaremos as primeiras derivadas parciais de cada parâmetro a zero:

$$\begin{aligned}\frac{\partial\mathbf{L}\{\mu,\sigma^2\}}{\partial\mu} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\mu)=0 \qquad \Rightarrow \qquad \sum_{i=1}^{n}(y_i-\mu)=0 \\ \frac{\partial\mathbf{L}\{\mu,\sigma^2\}}{\partial\sigma^2} &= \frac{n}{2\sigma^2}-\frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i-\mu)^2=0 \qquad \Rightarrow \qquad n-\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2=0\end{aligned}$$

- A solução para o sistema de duas equações gera os EMV:

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n}y_i \qquad \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(y_i-\widehat{\mu})^2$$

- É importante notar que o EMV da variância não divide a soma de quadrados pelos graus de liberdade ($n-1$), mas pelo tamanho da amostra (n).

  Logo, ele **não** é um estimador não-viciado da variância populacional:

$$\begin{aligned}\boldsymbol{E}\left\{\widehat{\sigma^2}\right\} &= \boldsymbol{E}\left\{\frac{1}{n}\sum_{i=1}^{n}(y_i-\widehat{\mu})^2\right\} = \boldsymbol{E}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[y_i^2-2y_i\widehat{\mu}+\widehat{\mu}^2\right]\right\} \\ &= \frac{1}{n}\boldsymbol{E}\left\{\left[\sum y_i^2-2\frac{(\sum y_i)^2}{n}+\frac{(\sum y_i)^2}{n}\right]\right\} \\ &= \frac{1}{n}\boldsymbol{E}\left\{\left[\sum y_i^2-\frac{(\sum y_i)^2}{n}\right]\right\} = \frac{1}{n}\left[\sum\boldsymbol{E}\left\{y_i^2\right\}-\frac{1}{n}\boldsymbol{E}\left\{(\sum y_i)^2\right\}\right] \\ &= \frac{1}{n}\left[\sum\left(\boldsymbol{Var}\{y_i\}+(\boldsymbol{E}\left\{y_i\right\})^2\right)-\frac{1}{n}\left(\boldsymbol{Var}\{\sum y_i\}+(\boldsymbol{E}\left\{\sum y_i\right\})^2\right)\right] \\ &= \frac{1}{n}\left[\left(n\sigma^2+n\mu^2\right)-\frac{1}{n}\left(n\sigma^2+n^2\mu^2\right)\right] \\ &= \frac{1}{n}\left[n\sigma^2+n\mu^2-\sigma^2-n\mu^2\right] \\ &= \frac{n-1}{n}\sigma^2\end{aligned}$$

- Mas note que a medida que n cresce, $\widehat{\sigma^2}$ tende ao valor do parâmetro:

$$\lim_{n\to\infty} \boldsymbol{E}\left\{\widehat{\sigma^2}\right\} = \lim_{n\to\infty} \frac{n-1}{n}\sigma^2 = \sigma^2$$

2.2.3 Propriedades dos Estimadores de Máxima Verossimilhança

Como foi visto os EMV podem ser viciados ou não-viciados, mas eles possuem uma série de propriedades que os torna "*altamente confiáveis*".

Greene (2000) apresenta as propriedades dos EMV da seguinte maneira:

Consistência: os EMV são consistentes, i.e., eles **convergem em probabilidade** para o valor do parâmetro.

Se $\widehat{\theta}_{\text{MV}}$ é o EMV de θ então:

$$\lim_{n\to\infty} P\left(|\widehat{\theta}_{\text{MV}} - \theta| > \epsilon\right) = 0$$

onde ϵ é uma constante real (arbitrariamente pequena).

A consistência garante que em grandes amostras ($n \to \infty$) os EMV, para efeitos práticos, são não-viciados.

Eficiência Assintótica: O Teorema do Limite Inferior de Cramer-Rao afirma que, para um dado parâmetro θ qualquer, existe um limite inferior para variância de qualquer estimador não-viciado:

$$\mathbf{I}(\theta)^{-1} = \left[-\boldsymbol{E}\left\{\frac{\partial^2 \ln \mathcal{L}\{\theta\}}{\partial\theta^2}\right\}\right]^{-1} = \left[\boldsymbol{E}\left\{\frac{\partial^2 \mathbf{L}\{\theta\}}{\partial\theta^2}\right\}\right]^{-1}.$$

Os EMV são *assimptoticamente* eficientes, i.e.,

$$\lim_{n\to\infty} \boldsymbol{Var}\{\widehat{\theta}_{\text{MV}}\} = \mathbf{I}(\theta)^{-1},$$

ou seja, para grandes amostras ($n \to \infty$), os EMV atingem o Limite Inferior de Cramer-Rao.

Normalidade Assimptótica: Os EMV convergem em distribuição para distribuição Gaussiana.

Se $F_{n;\widehat{\theta}_{\text{MV}}}(x)$ é a função de distribuição do EMV de θ, então

$$\lim_{n\to\infty} |F_{n;\widehat{\theta}_{\text{MV}}}(x) - F_G(x)| = 0$$

onde $F_G(x)$ é a função de distribuição Gaussiana.

Em outras palavras: para grandes amostras ($n \to \infty$), os EMV têm distribuição Gaussiana:

$$\widehat{\theta}_{\text{MV}} \overset{a}{\sim} N\left(\theta, \mathbf{I}(\theta)^{-1}\right).$$

Invariância: os EMV são invariantes sob transformações monotônicas.

Se estamos interessados em estimar $\gamma = g(\theta)$, onde $g(\cdot)$ é uma função monotônica, então

$$\widehat{\gamma}_{\text{MV}} = g(\widehat{\theta}_{\text{MV}}).$$

- As três primeiras propriedades mostram que, para grandes amostras ($n \to \infty$), os EMV são estimadores não viciados, de variância mínima e com distribuição Gaussiana.
- A quarta propriedade têm uma aplicação interessante no caso da relação entre variância e erro padrão:
  - O EQM da variância é não-viciado, mas isso não garante que o EQM do erro padrão também seja.
  - Já no caso dos EMV, temos certeza que:

$$\widehat{\sigma}_{\text{MV}} = \sqrt{\widehat{\sigma^2}_{\text{MV}}}$$

Família Exponencial de Distribuições

É importante lembrar que tais propriedades são válidas se a função de densidade da variável aleatória em questão ($f_Y(x, \theta)$) obedecer a algumas *condições de regularidade*:

1. As três primeiras derivadas de $f_Y(x, \theta)$ com relação a θ são finitas para todos valores de x.
2. As condições para se obter a primeira e segunda derivada de $\ln f_Y(x, \theta)$ são alcançadas.
3. Para todos valores de θ, os valores de

$$\left| \frac{\partial^3 \ln f_Y(x, \theta)}{\partial \theta^3} \right|$$

é menor que a função que tem valor esperado finito.

Existe uma classe de distribuições estatísticas para as quais as condições acima sempre são verdadeiras.

A *Família Exponencial de Distribuições* são distribuições estatísticas cuja log-verossimilhança pode ser fatorada da seguinte forma:

$$\mathbf{L}\{\theta|\mathbf{Y}\} = a(\mathbf{Y}) + b(\theta) + \sum_{j=1}^{k} c_j(\mathbf{Y}) s_j(\theta)$$

onde $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ e $s(\cdot)$ são funções.

A distribuição Gaussiana faz parte da Família Exponencial.

2.3 Analisando a Função de Verossimilhança

- Frequentemente quando trabalhamos com distribuições estatísticas com vários parâmetros, apenas um deles é de interessendo, tornando os demais parâmetros inconvenientes, mas necessários.
- Por exemplo, quando se usa a distribuição Normal o parâmetro de interesse é, em geral, a média (μ), sendo o desvio padrão (σ) um parâmetro inconveniente.
- Nestas circunstâncias não é conveniente se analisar as regiões de verossimilhança, pois os parâmetro inconvenientes não possuem interpretação prática.
- Existem várias técnicas para lidar com essa situação, mas abordaremos apenas duas.

2.3.1 Verossimilhança Estimada

- A Verossimilhnaça Estimada consiste em substituir os parâmetros inconvenientes pelas suas estimativas de máxima verossimilhança.
- Suponhamos que, no caso da distribuição Gaussiana, estejamos interessados apenas na média, sendo a variância um parâmetro inconveniente, a função de log-verossimilhança negativa estimada fica:

$$\mathbf{L}_E\{\mu\} = \mathbf{L}\{\mu, \widehat{\sigma^2}_{\text{MV}}\} = \frac{n}{2}\ln(2\pi) + \frac{n}{2}\ln(\widehat{\sigma^2}_{\text{MV}}) + \frac{1}{2\widehat{\sigma^2}_{\text{MV}}}\sum_{i=1}^{n}(y_i - \mu)^2$$

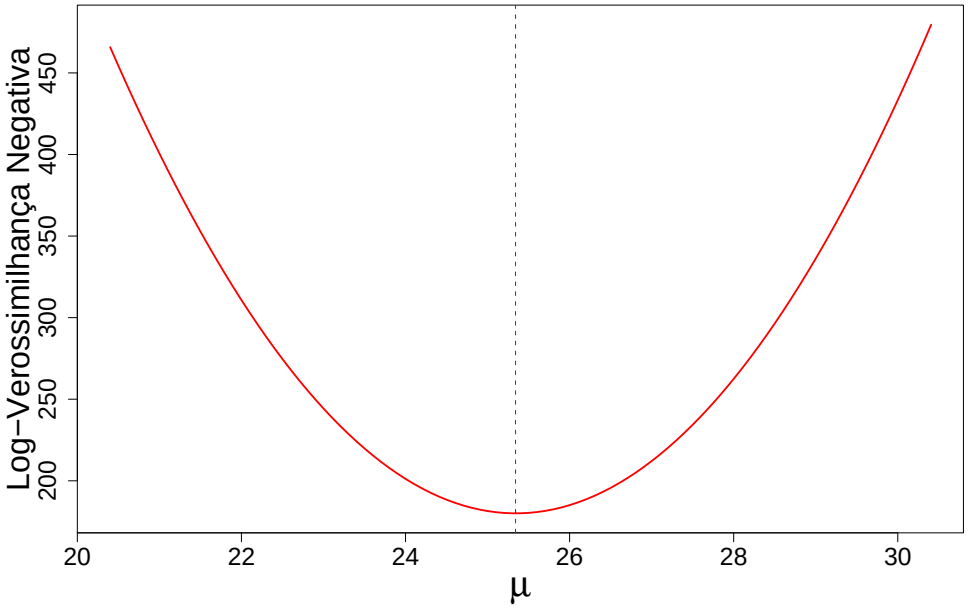


Figura 2.5: Função de log-verossimilhança negativa **estimada** para a média de uma distribuição Gaussiana, applicada a dados do DAP médio por parcela em floresta nativa do Maranhão.

2.3.2 Verossimilhança Perfilhada

- A Verossimilhnaça Perfilhada consiste em encontar as estimativas de máxima verossimilhança dos parâmetros inconvenientes **para cada valor do parâmetro de interesse**.
- Suponhamos que, no caso da distribuição Gaussiana, estejamos interessados apenas na média, sendo a variância um parâmetro inconveniente, a função de log-verossimilhança negativa perfilada fica:

$$\mathbf{L}_P\{\mu\} = \mathbf{L}\{\mu, \sigma^2 = s(\mu)\} = \frac{n}{2}\ln(2\pi) + \frac{n}{2}\ln(s(\mu)) + \frac{1}{2s(\mu)}\sum_{i=1}^{n}(y_i - \mu)^2$$

$$\text{sendo} \quad s(\mu) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2$$

$$\mathbf{L}_P\{\mu\} = \frac{n}{2}\ln(2\pi) + \frac{n}{2}\ln\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2\right) + \frac{n}{2}$$

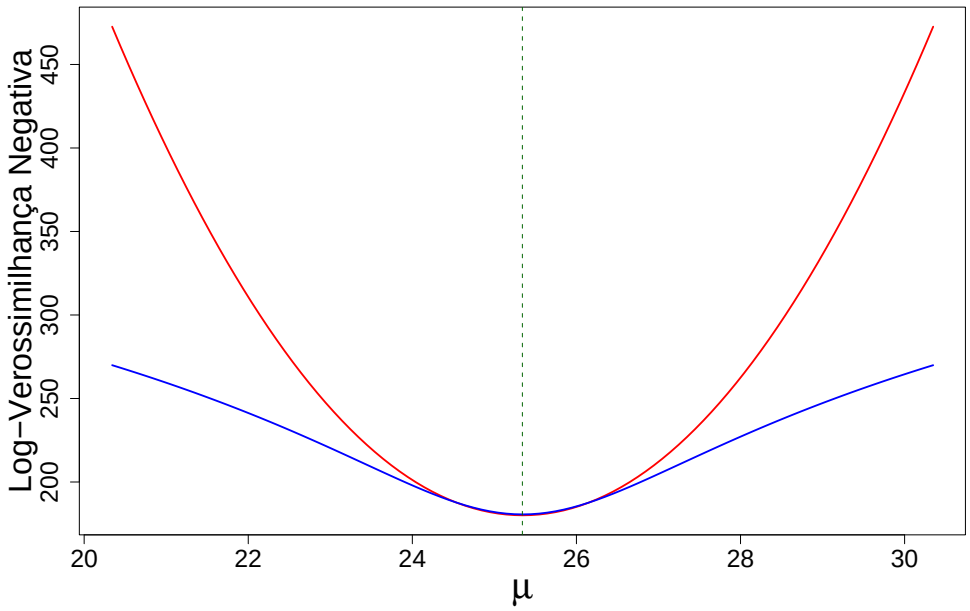


Figura 2.6: Função de log-verossimilhança negativa **perfilada** (em azul) e log-verossimilhança negativa **estimada** (em vermelho) para a média de uma distribuição Gaussiana, applicada a dados do DAP médio por parcela em floresta nativa do Maranhão.

- Note que a verossimilhança perfilada e verossimilhança estimada têm o mesmo ponto de mínimo (ponto das estimativas de máxima verossimilhança) e que na vizinhança desse ponto as curvas são coincidentes.
- Entretanto, à medida que se afasta do ponto de mínimo a curva da verossimilhança perfilada é mais aberta, o que indica um maior grau de incerteza sobre a estimativa da média.
- No entanto, a verossimilhança perfilada é considerada mais coerente e realista, uma vez que para cada valor de μ no gráfico, ela busca o EMV da variância assumindo aquele valor de média.

2.4 Razão de Verossimilhança

- Suponha que tenhamos duas hipóteses concorrentes a respeito do valor que um dado parâmetro θ assume:

$$H_A : \theta = \theta_A \qquad \text{e} \qquad H_B : \theta = \theta_B.$$

- A **plausibilidade** de cada hipótese, face uma amostra $\mathbf{Y}_n$ observada, é indicada pela respectiva função de verossimilhança:

$$H_A \Rightarrow \mathcal{L}\{\theta_A|\mathbf{Y}_n\} \qquad \text{e} \qquad H_B \Rightarrow \mathcal{L}\{\theta_B|\mathbf{Y}_n\}.$$

- A função de verossimilhança de maior valor indica a *hipótese mais plausível* das duas. Digamos que $\mathcal{L}\{\theta_A\} > \mathcal{L}\{\theta_B\}$.

- A **Razão de Verossimilhança**

$$\frac{\mathcal{L}\{\theta_A\}}{\mathcal{L}\{\theta_B\}}$$

  pode ser interpretada como a *força de evidência* com que os dados favorecem a hipótese A *vis-a-vis* a hipótese B.

- O logartimo da razão de verossimilhança é equivalente à **diferença** das log-verossimilhanças negativas:

$$\ln\left[\frac{\mathcal{L}\{\theta_A\}}{\mathcal{L}\{\theta_B\}}\right] = \ln[\mathcal{L}\{\theta_A\}] - \ln[\mathcal{L}\{\theta_B\}] = \mathbf{L}\{\theta_B\}] - \mathbf{L}\{\theta_A\}$$

2.4.1 Construindo Intervalos de Verossimilhança

- Os *Intervalos de Verossimilhança* (IV) são intervalos de *plausibilidade*.

- Valores dentro do IV são considerados igualmente plausíveis.

- Como os EMV maximizam a verosimilhança, eles são considerados os valores *mais plausíveis*.

- Para definirmos os Intervalos de Verossimilhança, precisamos definir um limite mínimo para a razão de verossimilhança a partir do qual os valores deixam de serem considerados igualmente plausíveis.

- Royall (1997) propõe dois níveis:

– Razão de Verossimilhança = **8** (diferença da log-verossimilhança negativa = $\ln(8)$) seria análoga ao nível de significância de 5%.

– Razão de Verossimilhança = **32** (diferença da log-verossimilhança negativa = $\ln(32)$) seria análoga ao nível de significância de 1%.

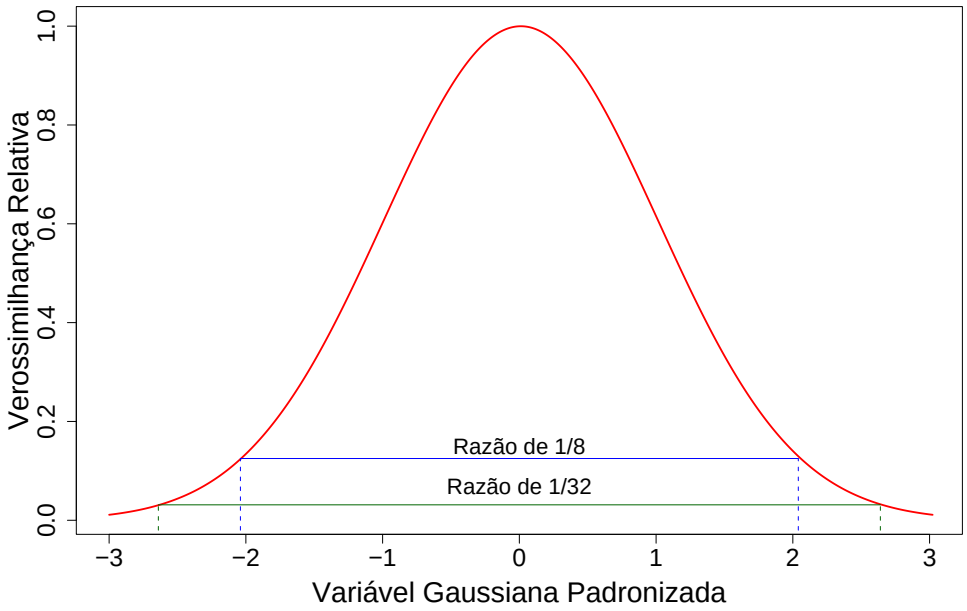


Figura 2.7: Intervalos de Verossimilhança para razões de 8 e 32 na distribuição Gaussiana Padronizada, mostrados na **função de verossimilhança** em escala relativa ao ponto de máximo.

- Utilizando as razões de 8 e 32 podemos construir os Intervalos de Verossimilhanç, mas essa idéia atualmente não é amplamente aceita entre os estatísticos.

2.5 O Modelo Linear Simples

- O modelo linear simples de erros gaussianos (MLEG) já foi apresentado na perspectiva dos estimadores de quadrados mínimos.
- Sua estrutura formal pode ser apresentada como:

$$\text{MLEG:} \qquad Y = \beta_0 + \beta_1 X + \varepsilon,$$

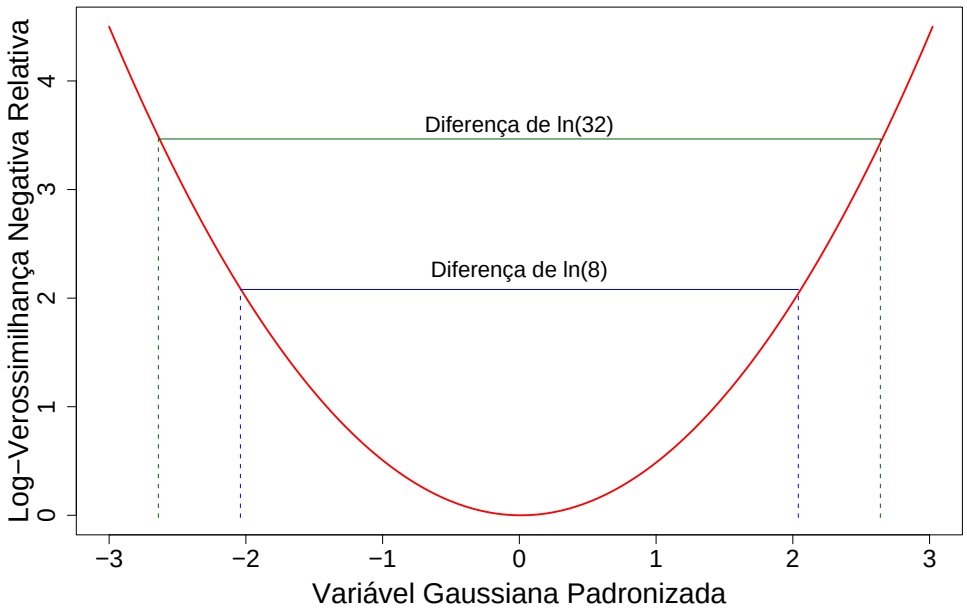


Figura 2.8: Intervalos de Verossimilhança para razões de 8 e 32 na distribuição Gaussiana Padronizada, mostrados na **função de log-verossimilhança** negativa em escala relativa ao ponto de mínimo.

onde:

$$\begin{aligned} X &- \text{variável não estocástica;} \\ (\varepsilon|X=x) &\overset{\text{iid}}{\sim} N\left(0,\sigma^2\right). \end{aligned}$$

- Na abordagem de verossimilhança o modelo pode ser apresentado a partir da função de verossimilhança:

$$\mathcal{L}\{\beta_0,\beta_1,\sigma^2|\mathbf{Y}_n\} = \prod_{i=1}^{n}\left(2\pi\sigma^2\right)^{-1/2}\exp\left[-\frac{1}{2\sigma^2}\left[y_i-(\beta_0+\beta_1 x_i)\right]^2\right]. \quad (2.1)$$

- Mas para fins de manipulação matemática utilizaremos a função de log-verossimilhança negativa:

$$\mathbf{L}\{\beta_0,\beta_1,\sigma^2|\mathbf{Y}_n\} = \frac{n}{2}\ln(2\pi)+\frac{n}{2}\ln(\sigma^2)+\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2 \quad (2.2)$$

2.5.1 EMV no Modelo Linear Simples

- Para encontrarmos os EMV do MLEG basta aplicarmos o procedimento de máxima verossimilhança nas equações 2.2.
- Primeiras derivadas em relação aos parâmetros do modelo:

$$\begin{aligned}
\frac{\partial \mathbf{L}\{\beta_0,\beta_1,\sigma^2\}}{\partial \beta_0} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)=0 \\
&\Rightarrow \sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)=0 \\
&\Rightarrow \sum y_i = n\beta_0 + \left(\sum x_i\right)\beta_1 \\
\\
\frac{\partial \mathbf{L}\{\beta_0,\beta_1,\sigma^2\}}{\partial \beta_1} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)x_i=0 \\
&\Rightarrow \sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)x_i=0 \\
&\Rightarrow \sum y_i x_i = \left(\sum x_i\right)\beta_0 + \left(\sum x_i^2\right)\beta_1 \\
\\
\frac{\partial \mathbf{L}\{\beta_0,\beta_1,\sigma^2\}}{\partial \sigma^2} &= \frac{n}{\sigma^2}-\frac{1}{\sigma^4}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2=0 \\
&\Rightarrow +n-\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2=0
\end{aligned}$$

- Das duas primeiras derivadas parciais extraímos o *Sistema de Equações Normais* :

$$\begin{aligned}
\sum y_i &= n\beta_0 + \left(\sum x_i\right)\beta_1 \\
\sum y_i x_i &= \left(\sum x_i\right)\beta_0 + \left(\sum x_i^2\right)\beta_1
\end{aligned}$$

  cuja solução gera os EMV para o modelo:

$$\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum y_i x_i - \left(\sum y_i \sum x_i\right)/n}{\sum x_i^2 - \left(\sum x_i\right)^2/n} = \frac{\sum (x_i-\overline{x})(y_i-\overline{y})}{\sum (x_i-\overline{x})^2} = \frac{S_{xy}}{S_{xx}} \\
\widehat{\beta}_0 &= \frac{\sum y_i}{n} - \widehat{\beta}_1 \frac{\sum x_i}{n} = \overline{y} - \widehat{\beta}_1 \overline{x}
\end{aligned}$$

- Note no caso desse modelo (MLEG), os EMV ($\widehat{\beta}_0$, $\widehat{\beta}_1$) são idênticos aos EQM (b_0, b_1).

- Aplicando os EMV dos coeficientes de regressão na terceira derivada, obtemos o EMV para variância do erro e erro padrão da estimativa:

$$\begin{aligned} \widehat{\sigma}^2 &= \frac{1}{n}\sum(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i)^2 \\ \widehat{\sigma} &= \sqrt{\frac{1}{n}\sum(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i)^2} \end{aligned}$$

2.5.2 Inferência sobre os Parâmetros: Coeficientes de Regressão e Variância do Erro

- A função de log-verossimilhança negativa do MLEG (equação 2.2) pode ser utilizada para fazer inferência sobre os parâmetros.
- Utilizando a verossimilhança perfilada para variância do erro obtemos:

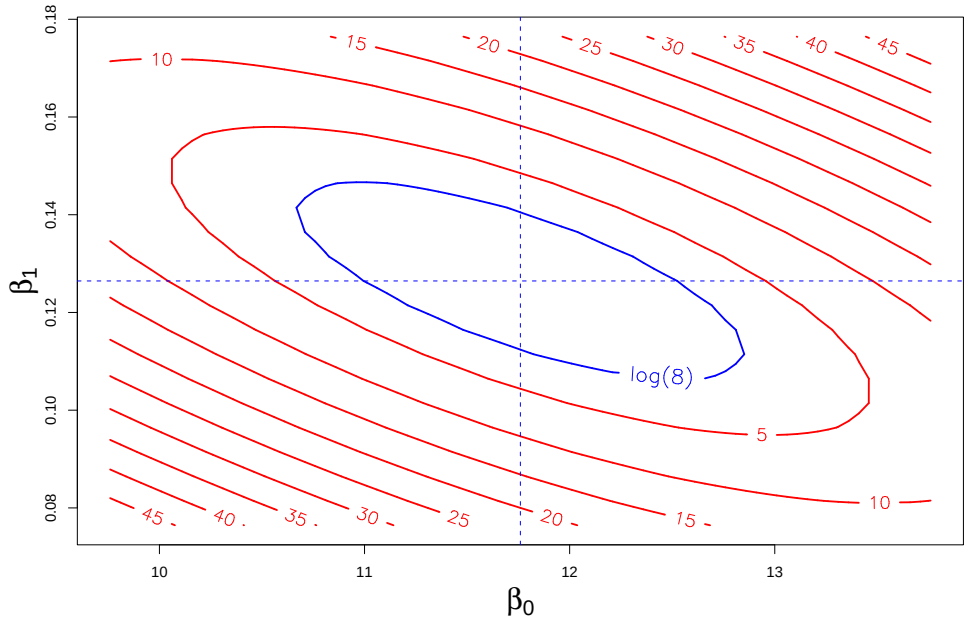


Figura 2.9: Gráfico de contorno para a função de log-verossimilhança negativa perfilada (escala relativa) para as EMV dos coeficientes de regessão no exemplo do rio Tinga (vazão em função de precipitação). A linha sólida em azul define a *Região de Verossimilhança* de razão 8 para os coeficientes de regressão.

- Utilizando a verossimilhança perfilada para construirmos um gráfico para cada coeficiente de regressão temos:

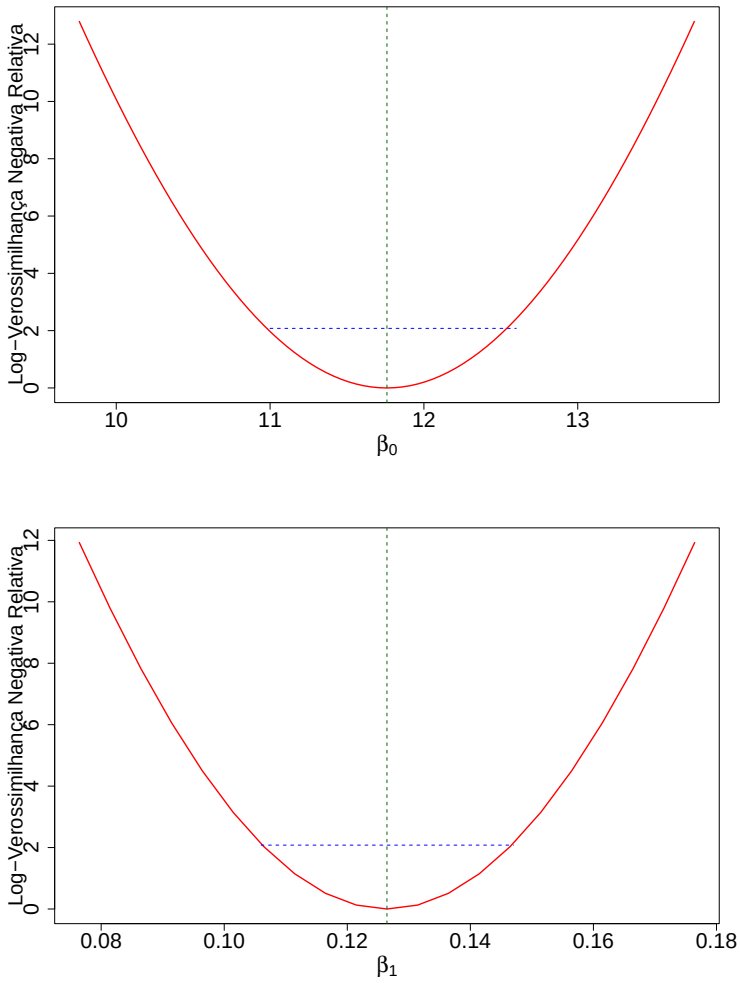


Figura 2.10: Gráficos da função de log-verossimilhança negativa perfilada (escala relativa) para a EMV do coeficiente do **intercepto** ($\widehat{\beta}_0$) e **inclinação** ($\widehat{\beta}_1$) no exemplo do rio Tinga (vazão em função de precipitação). A linha tracejada em azul define o *Intervalo de Verossimilhança* de razão 8 para o coeficiente de regressão.

- Um análise cuidadosa do gráfico de região de verossimilhança e dos gráficos de intervalo de confiança revela *não coincidência* que existe entre regiões e intervalos de verossimilhança (e de confiança).
- Uma das vantagens da abordagem da verossimilhança é a aplicação da mesma

metodologia para qualquer parâmetro.

Por exemplo, podemos construir um gráfico para avaliar a qualidade com que o erro padrão da estimativa está sendo estimado:

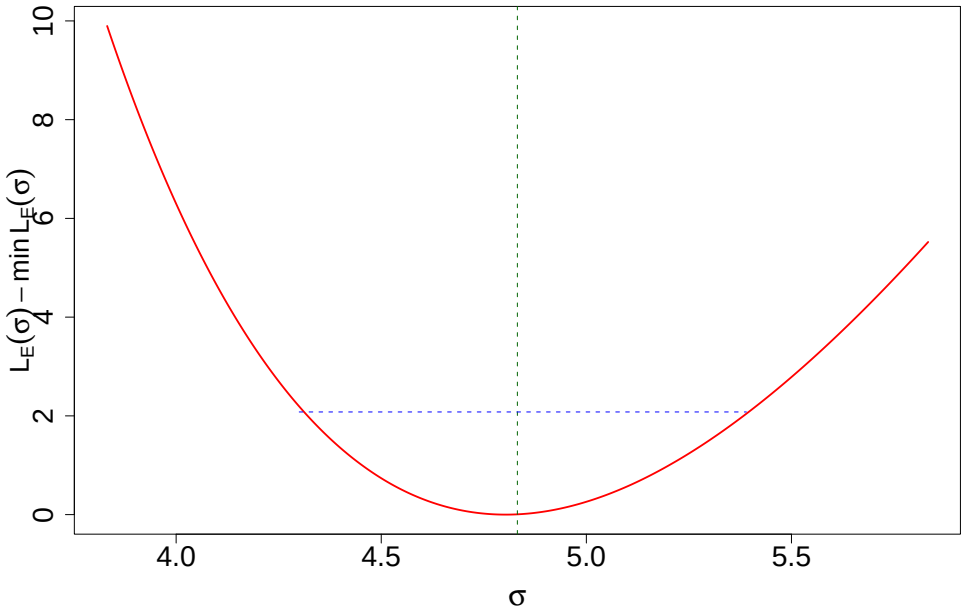


Figura 2.11: Gráficos da função de log-verossimilhança negativa estimada (escala relativa) para a EMV do erro padrão da estimativa (σ). no exemplo do rio Tinga (vazão em função de precipitação). A linha tracejada em azul define o *Intervalo de Verossimilhança* de razão 8 para o coeficiente de regressão.

- Note a assimetria do intervalo de verossimilhança de $\widehat{\sigma}$.

2.6 Aplicando o Modelo Linear Simples

2.6.1 Utilizando o Modelo para Inferência

- A inferência baseada em verossimilhança também pode ser utilizada na aplicação dos modelos ajustados.
- Toda inferência sobre parâmetros baseada na verossimilhança parte da função de verossimilhança, o que pode ser diferente da abordagem tradicional.
- Já no caso de predição, inferência sobre variáveis aleatórias, a abordagem de verossimilhança não traz novidades.

2.6.2 Inferência sobre a Resposta Média

- Para utilizarmos a verossimilhança para fazer inferência sobre a resposta média, temos que lembrar que a resposta média é um **parâmetro** estimado (sem viés) pelo valor ajustado:
$$\boldsymbol{E}\{\widehat{y}_k|X = x_k\} = \boldsymbol{E}\{Y|X = x_k\}.$$
- O estimador da resposta média pode ser visto como uma função dos estimadores dos coeficientes de regressão:
$$\widehat{y}_k = \widehat{\beta}_0 + \widehat{\beta}_1 x_k,$$
sendo, portanto, um EML e, consequentemente, têm distribuição Gaussiana com média $\boldsymbol{E}\{Y|X = xk\}$ e variância $\boldsymbol{Var}\{\widehat{y}_k\}$.
- Assim, a função de verossimilhança para a resposta média é:
$$\mathbf{L}\{\widehat{y}_k|\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\sigma}^2, X = x_k\} = \frac{n}{2}\ln(2\pi) + \frac{n}{2}\ln(\widehat{\boldsymbol{Var}}\{\widehat{y}_k\}) + \frac{1}{2\widehat{\boldsymbol{Var}}\{\widehat{y}_k\}}(y_k - \widehat{y}_k)^2$$
onde:
$$\widehat{\boldsymbol{Var}}\{\widehat{y}_k\} = \widehat{\sigma}^2\left[\frac{1}{n} + \frac{(x_k - \overline{x})^2}{\sum(x_i - \overline{x})^2}\right].$$
- Note que os únicos elementos que variam nessa função são
  - x_k: valores condicionantes da variável preditora, que geram alteração em $\widehat{y}_k$; e
  - y_k: valores da variável resposta sobre os quais se deseja obter $\mathbf{L}\{\widehat{y}_k\}$ (*não são valores observados*).

- Para obtermos a função de log-verossimilhança *relativa* necessitamos subtair o valor de mínimo.

  Se utilizarmos o ponto de mínimo para todas observações, através da expressão

  $$\mathbf{L}\{\widehat{y}_k\}_R = \mathbf{L}\{\widehat{y}_k\} - \min\left\{\mathbf{L}\{\widehat{y}_k\}\right\},$$

  obteremos o gráfico 2.12, que mostra o ponto de mínimo no encontro das médias amostrais.

  Esse gráfico enfatiza que a verossimilhança não é a mesma ao longo da reta de regressão pois estamos trabalhando com uma amostra e, consequentemente, valores próximos às médias amostrais são mais plausíveis.

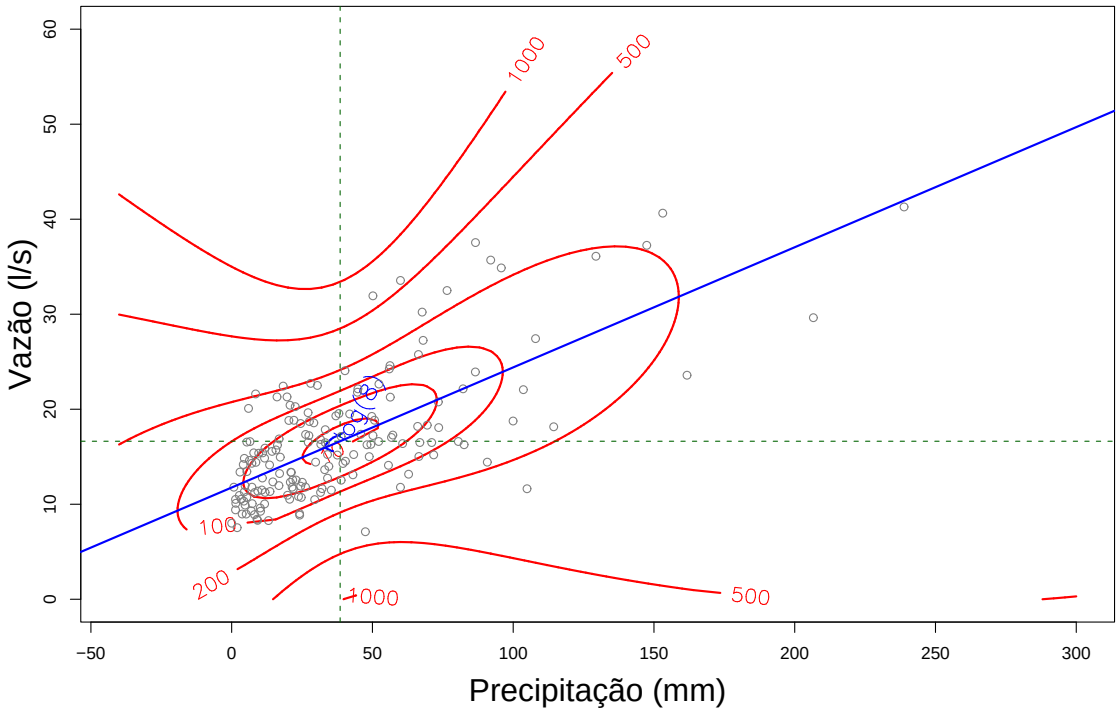


Figura 2.12: Log-Verossimilhança para o MLEG fazendo o cálculo não condicional, no exemplo do rio Tinga.

- Entretanto, não é isso que geralmente queremos ao fazer inferência sobre a resposta média.

O que desejamos é uma inferência condicional à $X = x_k$, assim a verossimilhança relativa deve ser tomada com o ponto de mínimo para cada valor x_k:

$$\mathbf{L}\{\widehat{y}_k\}_R = \mathbf{L}\{\widehat{y}_k\} - \min\left\{\mathbf{L}\{\widehat{y}_k\}|X = x_k\right\}.$$

O resultao é o gráfico 2.13 em que a verossimilhança diminiu *a partir da reta de regressão*

As *isolinhas* sugerem a forma de "um vale", sendo que no fundo desse vale está a reta de regressão, mas o vale fica mais estreito próximo ao ponto de encontro das médias amostrais e se alarga à medida que se afasta desse ponto.

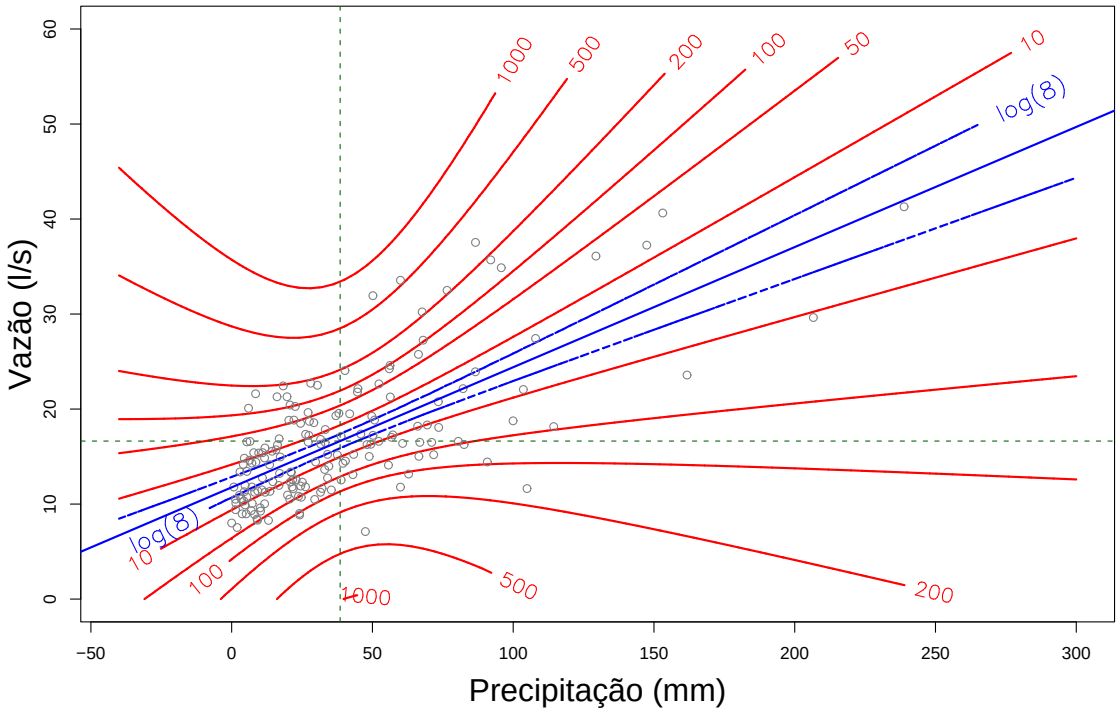


Figura 2.13: Log-Verossimilhança para o MLEG fazendo o cálculo condicional a $X = x$, no exemplo do rio Tinga.

- Para se obter o Intervalos de Verossimilhança para um nível de X em particular, basta aplicar a função de verossimilhança para um nível fixo de $X = x_k$ (gráfico 2.14).

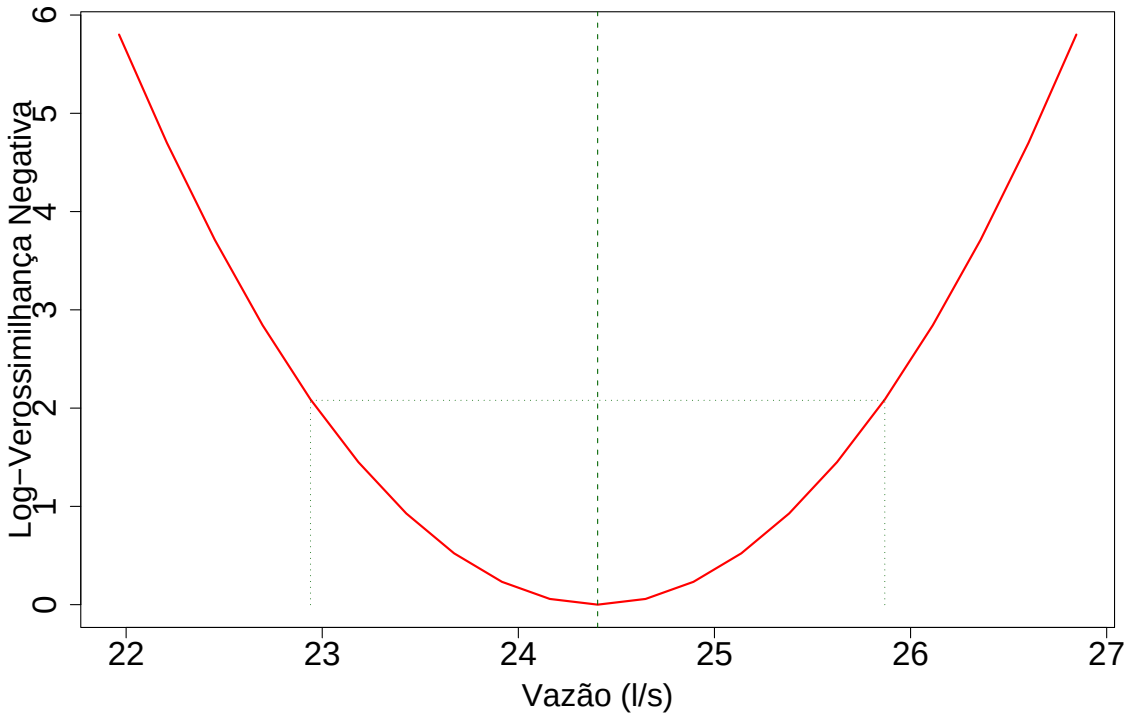


Figura 2.14: Gráfico da log-verossimilhança negativa para o MLEG fazendo o cálculo condicional a $X = 100$, no exemplo do rio Tinga. O Intervalo de Verossimilhança apresentado é para a razão de 1:8 (diferença de $\log(8)$).

2.6.3 Intervalos de Predição para uma Nova Observação

- No caso da predição de uma nova observação, a abordagem é identica a já vista.
- A predição de uma nova observação pelo modelo implica em observar um nova valor de uma variável aleatória Gaussiana com média

$$\widehat{y}_h = b_0 + b_1 x_h$$

e variância

$$\textbf{\textit{Var}}\{\widehat{y}_h\} = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_h - \overline{x})^2}{\sum (x_i - \overline{x})^2}\right].$$

- O intervalo de predição, no caso de um único nível da variável preditora, será gerado a partir da distribuição Gaussiana (figura 2.15).

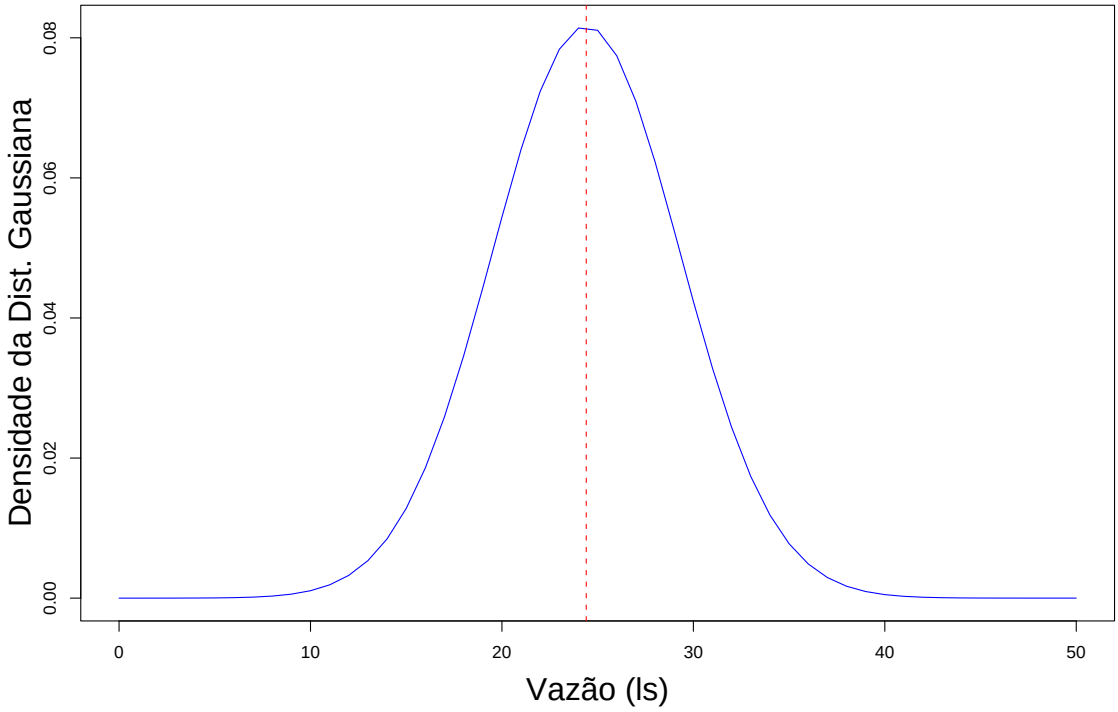


Figura 2.15: Gráfico da distribuição Gaussiana para a predição de uma nova observação pelo MLEG para o nível da variável preditora $X = 100$.

- No caso de observar a reta de regressão, os intervalos de predição *individuais* gerarão linhas praticamente paralelas à reta de regressão (figura 2.16).
- A mesma abordagem é válida para construir intervalos de predição para a média amostral de m novas predições.

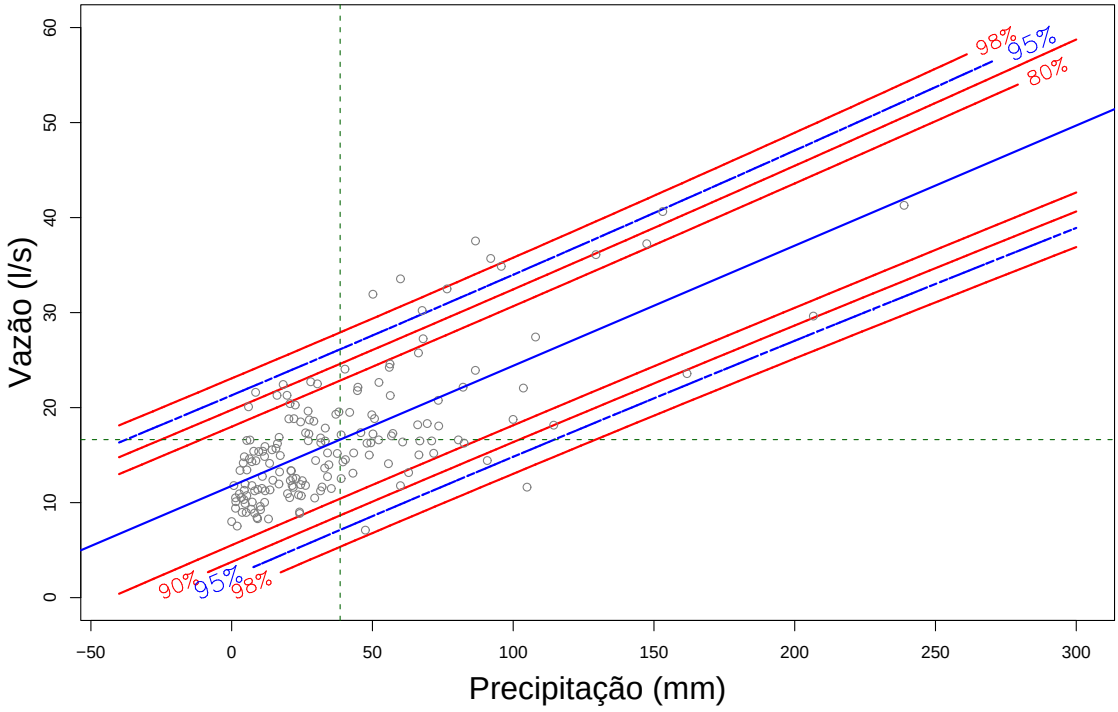


Figura 2.16: Gráfico das linhas de intervalo de predição individuais, com base na distribuição Gaussiana para a predição de uma nova observação pelo MLEG. Exemplo: rio Tinga.

2.7 Verificando a Qualidade do Ajuste

2.7.1 Verossimilhança do Modelo Ajustado

- Ao utilizarmos a abordagem da verossimilhança no MLEG sempre utilizaremos os EML para todos os parâmetros.
- OS EML implicam no ponto de máximo da função de verossimilhança (mínimo da log-verossimilhança negativa).
- Para um dado modelo ajustado (digamos MLEG.1), a função de

log-verossimilhança negativa assume a forma:

$$\begin{aligned}\mathbf{L}\{\text{MLEG.1}\} &= \max\left\{\mathbf{L}\{\beta_0,\beta_1,\sigma^2|\mathbf{Y}_n\}\right\} = \mathbf{L}\{\widehat{\beta}_0,\widehat{\beta}_1,\widehat{\sigma}^2|\mathbf{Y}_n\}\\ \mathbf{L}\{\text{MLEG.1}\} &= \frac{n}{2}\ln(2\pi)+\frac{n}{2}\ln(\widehat{\sigma}^2)+\frac{1}{2\widehat{\sigma}^2}\sum_{i=1}^{n}(y_i-\widehat{\beta}_0-\widehat{\beta}_1x_i)^2\\ &= \frac{n}{2}\ln(2\pi)+\frac{n}{2}\ln(\widehat{\sigma}^2)+\frac{1}{2\widehat{\sigma}^2}SQR\end{aligned}$$

Mas como o EML para variância do erro envolve SQR

$$\widehat{\sigma}^2_{MV}=\frac{1}{n}\sum_{i=1}^{n}(y_i-\widehat{\beta}_0-\widehat{\beta}_1x_i)^2=\frac{SQR}{n}$$

a expressão simplifica para:

$$\begin{aligned}\mathbf{L}\{\text{MLEG.1}\} &= \frac{n}{2}\ln(2\pi)+\frac{n}{2}\ln\left(\frac{SQR}{n}\right)+\frac{n}{2}\\ &= \frac{n}{2}\ln(2\pi)+\frac{n}{2}+\frac{n}{2}\ln\left(\frac{SQR}{n}\right)\\ \mathbf{L}\{\text{MLEG.1}\} &= \frac{n}{2}\ln(2\pi)+\frac{n}{2}-\frac{n}{2}\ln(n)+\frac{n}{2}\ln\left(SQR\right).\end{aligned}$$

- Note que para uma dada amostra o único termo que pode variar se diferentes modelos forem aplicados à amostra é o termo que envolve a SQR, pois todos os demais dependem do tamanho da amostra n.

2.7.2 Qualidade do Ajuste e Razão de Verossimilhança

- Na abordagem tradicional vimos dois critérios para verificar a qualidade do ajuste de um modelo linear:

  Teste F: que utiliza a estatística:

  $$F^*=\frac{QMM}{QMR}=\frac{SQM/1}{SQR/(n-2)}$$

  comparando-a contra a distribuição F com $\nu_1=1$ e $\nu_2=n-2$ graus de liberdade.

Coeficiente de Determinação: que quantifica a proporção da variabilidade da variável resposta é explicada pelo modelo

$$R^2 = \frac{SQM}{SQR} = 1 - \frac{SQT}{SQR}.$$

- Na abordagem da verossimilhança, a comparação se dá através da razão de verossimilhança, na qual comparamos duas hipóteses *vis-a-vis*.
- Como vimos, no modelo linear simples, o teste F é equivalente a testar as hipóteses:

$$H_0 : \beta_1 = 0 \qquad \text{contra} \qquad H_\alpha : \beta_1 \neq 0.$$

- Essas duas hipóteses podem ser traduzidas em dois modelos.

  Hipótese Alternativa: é o próprio MLEG, onde temos:

$$M_A : \boldsymbol{E}\{Y|X = x_k\} = \beta_0 + \beta_1 x_k,$$

  no qual temos 3 parâmetros: β_0, β_1 e σ^2.

  Hipótese Nula: é um modelo reduzido em relação ao MLEG:

$$M_B : \boldsymbol{E}\{Y|X = x_k\} = \beta_0 = \boldsymbol{E}\{Y\} = \mu,$$

  no qual temos apenas dois parâmetros: μ e σ^2.
  Nesse modelo reduzido, os EMV são:

$$\begin{aligned} \widehat{\mu}_{MV} &= \frac{\sum_{i=1}^n y_i}{n} = \overline{y} \\ \widehat{\sigma}_{MV} &= \frac{1}{n}\sum_{i=1}^n (y_i - \widehat{\mu}_{MV})^2 \end{aligned}$$

- Podemos comparar esses modelos pela razão de verossimilhança (ou pela diferença da log-verossimilhança negativa):

$$\begin{aligned} \frac{\mathcal{L}\{M_A\}}{\mathcal{L}\{M_B\}} &\Rightarrow \mathbf{L}\{M_B\} - \mathbf{L}\{M_A\} \\ \mathbf{L}\{M_A\} &= \frac{n}{2}\ln(2\pi) + \frac{n}{2} - \frac{n}{2}\ln(n) + \frac{n}{2}\ln\left(SQR_A\right) \\ &= \frac{n}{2}\ln(2\pi) + \frac{n}{2} - \frac{n}{2}\ln(n) + \frac{n}{2}\ln\left(\sum(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i)^2\right) \\ &= \frac{n}{2}\ln(2\pi) + \frac{n}{2} - \frac{n}{2}\ln(n) + \frac{n}{2}\ln\left(SQR_{MLEG}\right) \end{aligned}$$

$$\begin{aligned}
\mathbf{L}\{M_B\} &= \frac{n}{2}\ln(2\pi) + \frac{n}{2} - \frac{n}{2}\ln(n) + \frac{n}{2}\ln\left(SQR_B\right) \\
&= \frac{n}{2}\ln(2\pi) + \frac{n}{2} - \frac{n}{2}\ln(n) + \frac{n}{2}\ln\left(\sum(y_i - \overline{y})^2\right) \\
&= \frac{n}{2}\ln(2\pi) + \frac{n}{2} - \frac{n}{2}\ln(n) + \frac{n}{2}\ln\left(SQT_{MLEG}\right)
\end{aligned}$$

- Ou seja:
  - a SQR do modelo M_A é a própria soma de quadrados do resíduo do MLEG;
  - a SQR do modelo reduzio M_B é na verdade a SQT do MLEG.
- A diferença da log-verossimilhança negativa fica

$$\mathbf{L}\{M_B\} - \mathbf{L}\{M_A\} = \frac{n}{2}\ln\left(SQT_{MLEG}\right) - \frac{n}{2}\ln\left(SQR_{MLEG}\right) = \frac{n}{2}\ln\left[\frac{SQT}{SQR}\right].$$

- Essa diferença pode ser expressa em termos de estatística F^*:

$$\begin{aligned}
\mathbf{L}\{M_B\} - \mathbf{L}\{M_A\} &= \frac{n}{2}\ln\left[\frac{SQT}{SQR}\right] = \frac{n}{2}\ln\left[\frac{SQM + SQR}{SQR}\right] = \frac{n}{2}\ln\left[\frac{SQM}{SQR} + 1\right] \\
&= \frac{n}{2}\ln\left[\frac{1}{n-2}F^* + 1\right].
\end{aligned}$$

- Também pode ser expressa em termos de R^2:

$$\begin{aligned}
\mathbf{L}\{M_B\} - \mathbf{L}\{M_A\} &= \frac{n}{2}\ln\left[\frac{SQT}{SQR}\right] = -\frac{n}{2}\ln\left[\frac{SQR}{SQT}\right] \\
&= -\frac{n}{2}\ln\left[1 - R^2\right].
\end{aligned}$$

- **Conclusão:** na perspectiva da verossimilhança,
  - as duas estatísticas para verificar a qualidade do ajuste do modelo linear são análogas à razão de verossimilhança (ou diferença da log-verossimilhança negativa);
  - essas estatísticas comparam 2 modelos hierárquicos:

    Modelo Completo: onde $\beta_1 \neq 0$, e
    Modelo Reduzido: onde $\beta_1 = 0$.

2.7.3 O Critério de Akaike na Comparação de Modelos

- A utilização da razão da verossimilhança não considera que os dois modelos sendo comparados podem diferir bastante no número de parâmetros ajustados.
- Geralmente, espera-se que um modelo com mais parâmetros tende a apresentar um ajuste melhor que um com menos parâmetros, pois os parâmetros são os elementos que nos permitem explicar o comportamento dos dados.
- Pelo *Princípio da Parsimônia*, ou *Princípio de Okham*, entre dois modelos com igual poder explicativo opta-se pelo modelo mais simples, i.e., com menor número de parâmetros.
- Um critério baseado da log-verossimilhança que considera o número de parâmetros é o **Critério de Informação de Akaike** (AIC), que penaliza a log-verossimilhança com duas vêzes o número de parâmetros:

(2.3) $$AIC = -2\ln\left[\mathcal{L}\{\theta\}\right] + 2p = 2\mathbf{L}\{\theta\} + 2p$$

- No caso de um modelo linear ajustado, o AIC se torna:

$$\begin{aligned} AIC &= 2\mathbf{L}\{MLEG\} + 2p \\ &= n\ln(2\pi) + n - n\ln(n) + n\ln\left(SQR\right) + 2p. \end{aligned}$$

- Se utilizarmos o AIC para comparar modelos hierárquicos, todos os termos relativos ao tamanho da amostra (n) podem ser eliminados, ficando apenas o termo dependente da soma de quadrados do resíduo:

(2.4) $$AIC = n\ln\left(SQR\right) + 2p.$$

2.7.4 AIC e Qualidade de Ajuste

- No caso do teste da qualidade de ajuste do modelo linear simples, o AIC seria aplicado aos dois modelos:

$$\begin{aligned} AIC(M_{\text{COMPLETO}}) &= n\ln\left(SQR\right) + 6 \\ AIC(M_{\text{REDUZIDO}}) &= n\ln\left(SQT\right) + 4 \end{aligned}$$

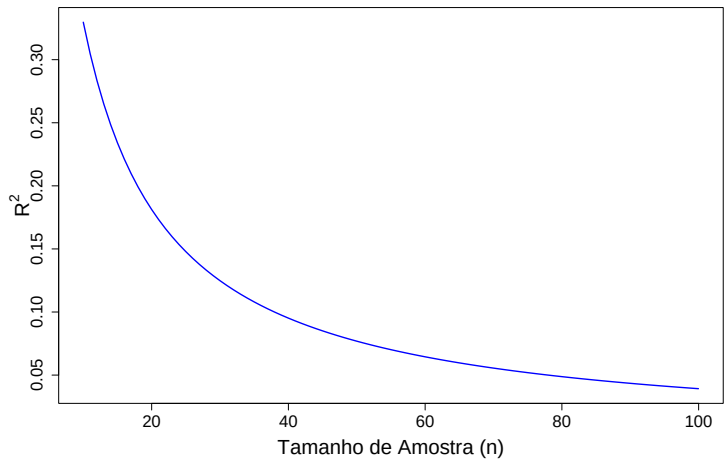
- Note que no modelo completo (modelo linear simples) temos o número de parâmetros $p = 3$, pois os parâmetros são os coeficientes de regressão (β_0 e β_1) e a variância do erro (σ^2).

- Já no modelo reduzido tempos $p = 2$: μ e σ^2.
- A qualidade do ajuste deve ser analisada pela diferença dos AICs:

$$\begin{aligned}\Delta_{\text{AJUSTE}} &= AIC(M_{\text{REDUZIDO}}) - AIC(M_{\text{COMPLETO}}) \\ &= n\left[\ln\left(SQT\right) - \ln\left(SQR\right)\right] - 2 \\ &= n\ln\left[\frac{SQT}{SQR}\right] - 2\end{aligned}$$

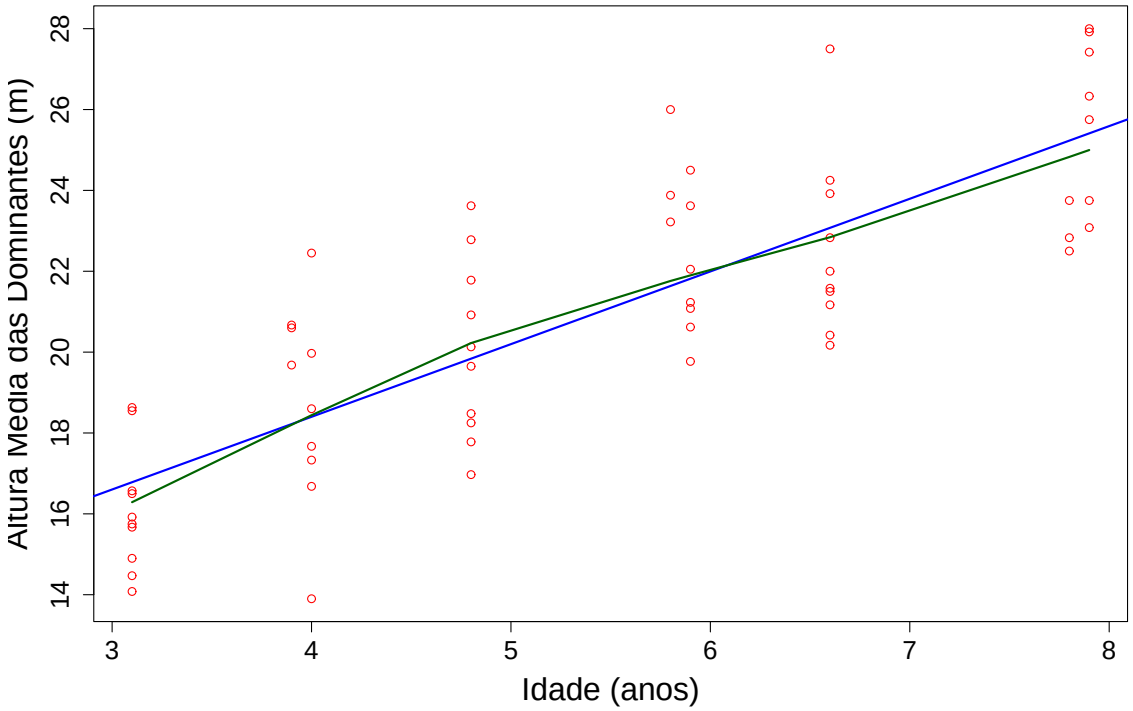
- A *interpretação direta* da diferença de AIC é geralmente utilizando um nível arbitrário para diferença da log-verossimilhança ($\ln(8)$):
  - Para $\Delta_{\text{AJUSTE}} < 2$ ($\approx \ln(8)$) os modelos completo e reduzidos podem ser considerados igualmente plausíveis, i.e., o MLEG falhou em se ajustar.
  - Para $\Delta_{\text{AJUSTE}} > 2$ ($\approx \ln(8)$) o modelo completo pode ser considerado superior ao reduzido, i.e., o MLEG tem boa qualidade de ajuste.
- Como a abordagem da verossimilhança é *condicional ao dados e aos modelos sendo comparados*, ao se interpretar os AIC ou suas diferenças nenhuma referência é feita ao tamanho da amostra (n), ao número de parâmetros (p) ou a graus de liberdade.
- É interessante notar que é possível estabelecer uma relação direta entre Δ_{AJUSTE} e o valor de R^2 mínimo necessário para se concluir que houve um bom ajuste:

$$\begin{aligned}\Delta_{\text{AJUSTE}} &= -n\ln\left[1 - R^2\right] - 2 \\ R^2 &= 1 - \exp\left[-\frac{4}{n}\right]\end{aligned}$$



2.8 Exemplo: Crescimento das Árvores Dominantes em *E. grandis*

Relação entre a altura média das árvores dominantes (variável resposta) e a idade (variável preditora) de povoamentos de *E. grandis* no Estado de São Paulo.

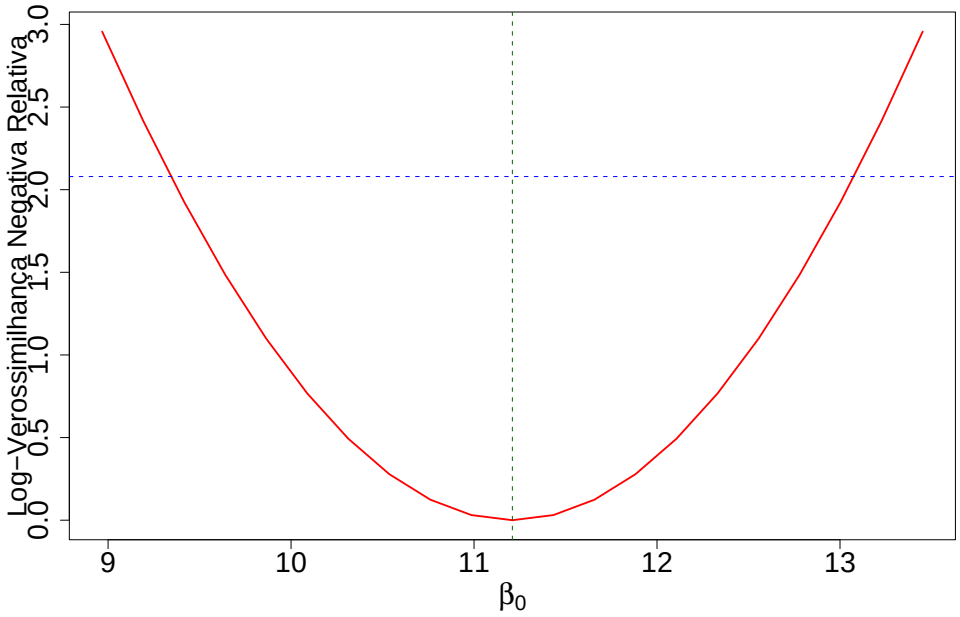


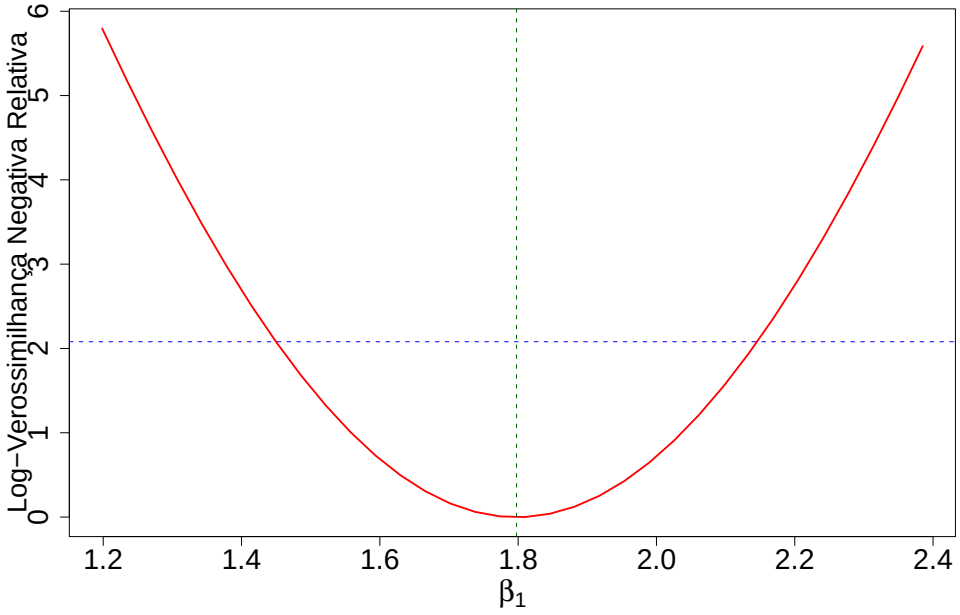
- Algumas estatísticas:

  $n = 60$; $\sum(x_i - \overline{x})^2 = 154.5698$;
  $\overline{x} = 5.368333$; $\sum(y_i - \overline{y})^2 = 759.7539$;
  $\overline{y} = 20.85983$; $\sum(x_i - \overline{x})(y_i - \overline{y}) = 277.8607$.

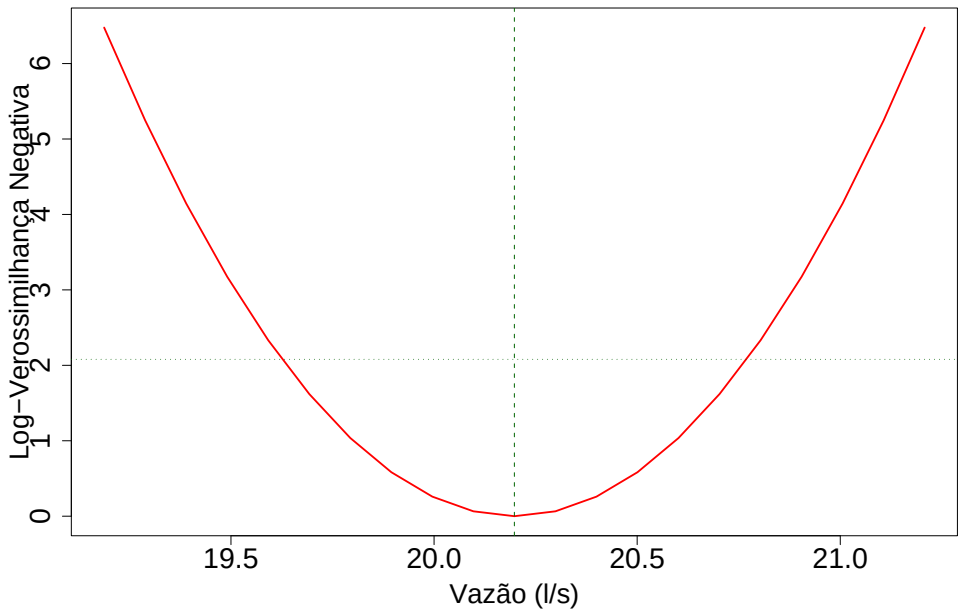
- EMV do modelo linear simples: $\widehat{\beta}_0 = 11.209511$, $\widehat{\beta}_1 = 1.797638$ e $\widehat{\sigma}^2 = 4.337681$.

- I.V. para os EMV:





- Inferência sobre a resposta média para idade de 5 anos:



- Qualidade do ajuste:$\Delta_{\text{AJUSTE}} = 71.80741$, ou seja, a qualidade de ajuste é muito boa.

2.9 Exercícios

Para as situações abaixo, ajuste o MLEG encontrando:

- As EMV para os parâmetros.
- Os Intervalos de Verossimilhança (razão 8) para os coeficientes de regressão (utilize a Verossimilhança Estimada).
- Os Intervalos de Confiança de 95% para os coeficientes de regressão e compara com os I.V..
- O Intervalo de Verossimilhança (razão 8) para a resposta média para x_k especificado (utilize a Verossimilhança Estimada).
- O Intervalo de Confiança de 95% para a resposta média para x_k especificado e compare com o I.V..

2.9-1. Relação entre o diâmetro médio das árvores (variável resposta, em cm) e a idade (variável preditora, em $anos$) em povoamentos de *Eucalyptus grandis*:

$n = 60$; $\sum(x_i - \overline{x})^2 = 154.5698$;
$\overline{x} = 5.368333$; $\sum(y_i - \overline{y})^2 = 101.9314$;
$\overline{y} = 9.576167$; $\sum(x_i - \overline{x})(y_i - \overline{y}) = 88.58172$.

2.9-2. Relação entre o incremnto médio anual (variável resposta, em $m^3 ha^{-1}$) e a idade (variável preditora, em $anos$) em povoamentos de *Eucalyptus grandis*:

$n = 60$; $\sum(x_i - \overline{x})^2 = 154.5698$;
$\overline{x} = 5.368333$; $\sum(y_i - \overline{y})^2 = 1996.933$;
$\overline{y} = 22.46667$; $\sum(x_i - \overline{x})(y_i - \overline{y}) = 14.08667$.

2.9-3. Relação entre a turbidez (variável resposta, em FTU) e a precipitação (variável preditora, em mm) no rio tinga.

$n = 149$; $\sum(x_i - \overline{x})^2 = 222707.8$;
$\overline{x} = 40.43273$; $\sum(y_i - \overline{y})^2 = 1139.351$;
$\overline{y} = 3.087248$; $\sum(x_i - \overline{x})(y_i - \overline{y}) = 415.3505$.

CAPÍTULO 3
MISCELÂNEA SOBRE MODELOS LINEARES

Nesse capítulo trataremos de dois temas não diretamente ligados entre si mas que são elementos importantes da utilização do Modelo Linear simples:

- o teste da *Falta de Ajuste*;
- a questão de quando a variável preditora é estocástica; e
- a questão de erros de mensuração nas variáveis.

3.1 Teste da Falta de Ajuste

- O teste da falta de ajuste é uma maneira de verificar a pressuposição de que o modelo linear é verdadeiramente o modelo adequado para os dados.
- Ele será definido em termos de comparação de dois modelos:
  1. o modelo *reduzido* e
  2. o modelo *completo*.

3.1.1 Modelo Reduzido

- Para aplicarmos o teste da falta de ajuste, assumiremos o MLEG deixando explícito que em cada nível da preditora ($X = x_i$) temos diversas repetições (n_i):
  - Modelo Linear: $y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij}$;
  - Comportamento do erro: $(\varepsilon_{ij}|X = x_i) \overset{\text{iid}}{\sim} N(0, \sigma^2)$;

  - Variável preditora (X) é não-estocástica com níveis estabelecidos $i = 1, 2, \ldots, m$, e em cada nível há um número n_i repetido de observações, sendo o número total de observações $n = \sum_{i=1}^{m} n_i$.
  - Nesse modelo, o número de parâmetros (excluindo a variância do erro) é $p = 2$.

- O modelo assim concebido será considerado como modelo reduzido, pois os valores esperados da variável resposta em cada nível da variável preditora se alinham de acordo com a reta de regressão:

$$\boldsymbol{E}\{Y|X = x_i\} = \mu_i = \beta_0 + \beta_1 x_i.$$

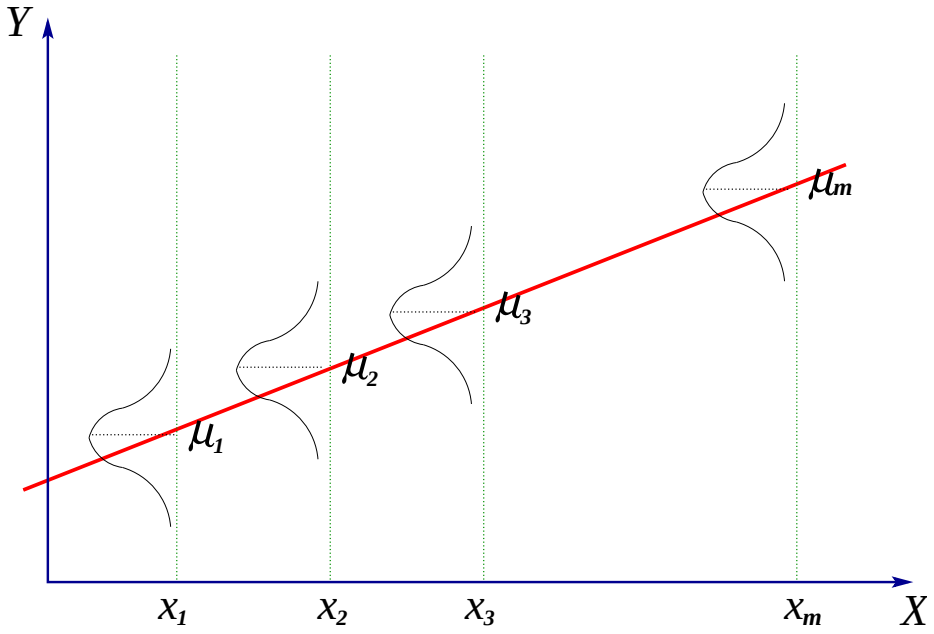


Figura 3.1: Modelo Linear Simples (MLEG) como um modelo reduzido, onde as médias se alinha ao longo da reta de regressão.

- Ao ajustar o modelo pelo método de quadrados mínimos (ou MMV) obtermos como modelo ajustado:

$$\widehat{y}_{ij} = b_0 + b_1 x_i,$$

  onde b_0 e b_1 são as EQM dos coeficientes de regressão.

- A variância do erro nesse modelo será estimada pela soma de quadrado de resíduos:

$$SQR_{\text{M.REDUZIDO}} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \widehat{y}_{ij})^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - b_0 - b_1 x_i)^2,$$

que é a própria SQR do MLEG. Associados essa SQR temos $n-2$ graus de liberdade.

3.1.2 Modelo Completo

- O modelo completo é obtido retirando-se a restrição de relação linear entre os valores esperados da variável resposta em cada nível da variável preditora.
- Em comparação com o MLEG, o modelo completo nesse caso é
  - Modelo (não necessariamente linear): $y_{ij} = g(x_i) + \varepsilon_{ij}$;
  - Comportamento do erro: $(\varepsilon_{ij}|X = x_i) \overset{\text{iid}}{\sim} N(0, \sigma^2)$;
  - Variável preditora (X) é não-estocástica com níveis estabelecidos $i = 1, 2, \ldots, m$.
  - A função $g(\cdot)$ é desconhecida, podendo ser qualquer função, de modo que a única afirmação que podemos fazer sobre o modelo é:

$$\boldsymbol{E}\{Y|X = x_i\} = \mu_i.$$

  - Nesse modelo, o número de parâmetros (excluindo a variância do erro) é igual ao número de níveis de X que forem utilizados: m.
- Ajustando esse modelo pelo método de quadrados mínimos (ou MMV) obtemos o seguinte modelo ajustado:

$$\widehat{y}_{ij} = \overline{y}_i,$$

  onde $\overline{y}_i$ é a média amostral de Y para cada um dos níveis de X.
- A variância do erro nesse modelo será estimada pela soma de quadrado de resíduos:

$$SQR_{\text{M.COMPLETO}} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \widehat{y}_{ij})^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \overline{y}_i)^2.$$

  Associados essa SQR temos como graus de liberdade

$$\sum_{i=1}^{m}(n_i - 1) = \sum_{i=1}^{m} n_i - m = n - m.$$

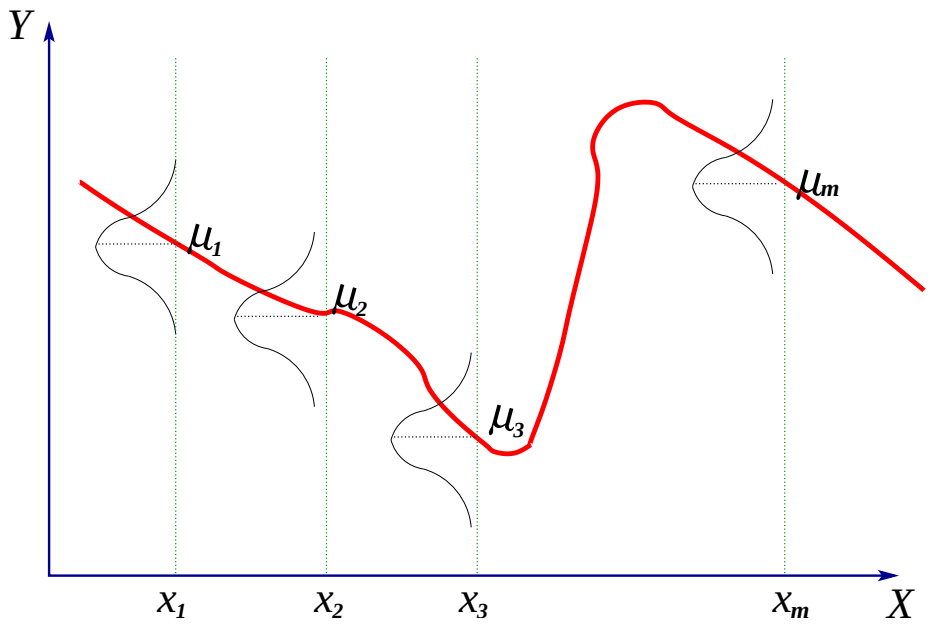


Figura 3.2: Modelo completo no teste da falta de ajuste, onde as médias seguem uma função desconhecida.

3.1.3 Relação entre Modelo Completo e Modelo Reduzido

- Ao analisarmos a SQR do modelo reduzido em relação ao modelo completo encontramos uma relação importante entre esses modelos:

$$\begin{aligned} SQR_{\text{M.REDUZIDO}} &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left[y_{ij}-\widehat{y}_i\right]^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left[y_{ij}-\widehat{y}_i+\overline{y}_i-\overline{y}_i\right]^2 \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left[\left(y_{ij}-\overline{y}_i\right)+\left(\overline{y}_i-\widehat{y}_i\right)\right]^2 \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(y_{ij}-\overline{y}_i\right)^2+\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\overline{y}_i-\widehat{y}_i\right)^2+2\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(y_{ij}-\overline{y}_i\right)\left(\overline{y}_i-\widehat{y}_i\right). \\ SQR_{\text{M.REDUZIDO}} &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(y_{ij}-\overline{y}_i\right)^2+\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\overline{y}_i-\widehat{y}_i\right)^2 \end{aligned}$$

Note que

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(y_{ij}-\overline{y}_i\right)\left(\overline{y}_i-\widehat{y}_i\right) = \sum_{i=1}^{m}\left[\left(\overline{y}_i-\widehat{y}_i\right)\sum_{j=1}^{n_i}\left(y_{ij}-\overline{y}_i\right)\right] = 0$$

pois $\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_i)=0$ para todo j.

- Assim a SQR no modelo reduzido (MLEG) pode ser interpretada como tendo dois componentes:

$$SQR_{\text{M.REDUZIDO}} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_i)^2 + \sum_{i=1}^{m}\sum_{j=1}^{n_i}(\overline{y}_i-\widehat{y}_i)^2$$

  - $\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_i)^2$ é a soma de quadrados do resíduo do **modelo completo**, que é chamada de *Soma de Quadrados do Resíduo Puro* ($SQRP$).
  - $\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\overline{y}_i-\widehat{y}_i)^2$ é a soma de quadrados dos desvios do modelo reduzido **em relação ao** modelo completo, que é chamada de *Soma de Quadrados da Falta de Ajuste* ($SQFA$).

- No contexto do modelo linear simples (MLEG) temos então que

$$SQR = SQRP + SQFA \qquad \Rightarrow \qquad SQFA = SQR - SQRP.$$

- É importante notar a influência a número de observações em cada nível de X sobre a $SQFA$:

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\overline{y}_i-\widehat{y}_i)^2 = \sum_{i=1}^{m} n_i(\overline{y}_i-\widehat{y}_i)^2,$$

  ou seja, em cada nível de X, a "*falta de ajuste*" é poderada pelo número de observações.

3.1.4 Teste F para Falta de Ajuste

- Na abordagem estatística do teste de hipóteses, a falta de ajuste é verificada construindo-se um teste para as hipóteses:

$$\begin{aligned} H_0: &\quad \boldsymbol{E}\{Y|X=x_i\} = \mu_i = \beta_0+\beta_1 x_i \\ H_\alpha: &\quad \boldsymbol{E}\{Y|X=x_i\} = \mu_i \neq \beta_0+\beta_1 x_i \end{aligned}$$

- No caso de H_0 ser verdadeira, a $SQFA$ será bem pequena, mas no caso de H_0 falsa ela será grande.

- Para construirmos um teste, devemos associar graus de liberdade (gl) com as somas de quadrados:

$$\begin{aligned} SQR_{\text{M.REDUZDIO}} = SQR \quad &\Rightarrow \quad gl_{\text{M.REDUZIDO}} = n-2 \\ SQR_{\text{M.COMPLETO}} = SQRP \quad &\Rightarrow \quad gl_{\text{M.COMPLETO}} = n-m \\ SQFA = SQR - SQRP \quad &\Rightarrow \quad gl_{\text{FA}} = (n-2)-(n-m) = m-2 \end{aligned}$$

- No contexto do MLEG, para cada soma de quadrados teremos uma estimativa de variância (quadrado médio):

$$\begin{aligned} QMR &= \frac{SQR}{n-2} \\ QMRP &= \frac{SQRP}{n-m} \\ QMRFA &= \frac{SQFA}{m-2}. \end{aligned}$$

- O valor esperado dessas variâncias são:

$$\begin{aligned} \boldsymbol{E}\{QMRP\} &= \sigma^2 \\ \boldsymbol{E}\{QMFA\} &= \sigma^2 + \left[\sum_{i=1}^{n} n_i \left(\mu_i - (\beta_0 + \beta_1 x_i)\right)^2\right] / (m-2) \end{aligned}$$

- O $QMRP$ é o melhor estimador para a variância do erro, independentemente de qual modelo de regressão é verdadeiro.
- No caso de H_0 verdadeira, o $QMFA$ também é um bom estimador da variância do erro, pois $\mu_i = \beta_0 + \beta_1 x_i$.
- Para testarmos H_0, podemos comparar esses dois estimadores da variância do erro através da estatística

$$F^* = \frac{QMFA}{QMRP} = \frac{SQFA/(m-2)}{SQRP/(m-n)}.$$

- Sob H_0, essa estatística terá distribuição F com $(m-2)$ graus de liberdade no numerador e $(m-n)$ graus de liberdade no denominador, o que sugere a regra de decisão:
  - Se $F^* \geq F(1-\alpha; m-2; m-n) \Rightarrow$ rejeita-se H_0;
  - Se $F^* < F(1-\alpha; m-2; m-n) \Rightarrow$ não se rejeita H_0.

3.1.5 Falta de Ajuste pelo AIC

- Utilizando a abordagem da verossimilhança os modelos completos e reduzidos seriam comparados utilizando o AIC:

$$\begin{aligned} \text{AIC}\{\text{M.REDUZIDO}\} &= n \ln(SQR) + 2p \\ \text{AIC}\{\text{M.COMPLETO}\} &= n \ln(SQRP) + 2m \end{aligned}$$

- O teste da falta de ajuste é

$$\begin{aligned}\Delta_{\text{FALTA AJUSTE}} &= n\ln(SQR) + 2p - n\ln(SQRP) - 2m \\ &= n\ln\left(\frac{SQR}{SQRP}\right) + 2(p-m) \\ &= n\ln\left(1 + \frac{SQFA}{SQRP}\right) + 2(p-m)\end{aligned}$$

- Se $\Delta_{\text{FALTA AJUSTE}} = \ln(8) \approx 2$ for fixado como nível a partir do qual as diferenças passam a ser consideradas marcantes, temos a seguinte relação

$$\begin{aligned}2 &> n\ln\left(1 + \frac{SQFA}{SQRP}\right) + 2(p-m) \\ \frac{SQFA}{SQRP} &> 1 - \exp\left[\frac{2 + 2(p-m)}{n}\right]\end{aligned}$$

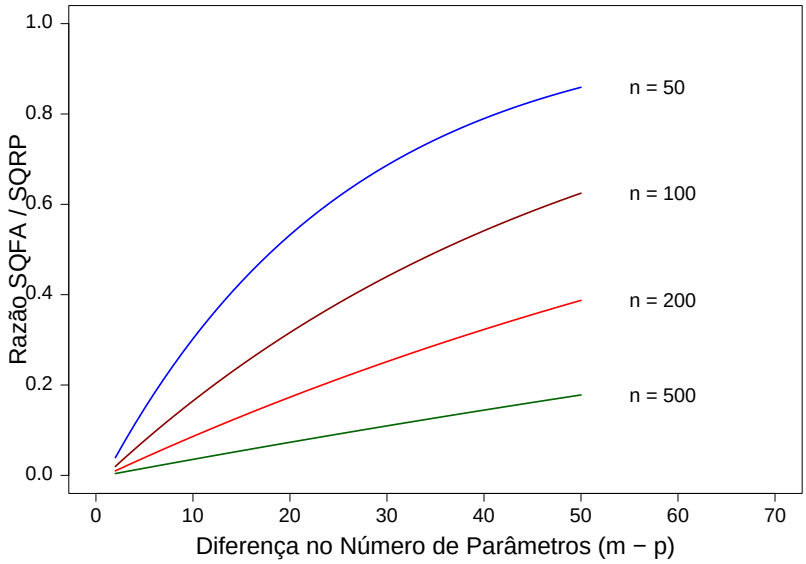


Figura 3.3: Gráfico mostrando a como a razão $\frac{SQFA}{SQRP}$ deve mudar em função da diferença do número de parâmetros entre o modelo completo e reduzido para razão de verossimilhança de $\ln(8)$.

3.1.6 Exemplo: Altura Média das Árvores Dominantes em *E. grandis*

- Relação entre a altura média das árvores dominantes (m) (numa parcela) e a idade do povoamento (anos) para povoamentos de *Eucalyptus grandis* na região central do Estado de São Paulo.
- A relação se mostra aparentemtemente linear:

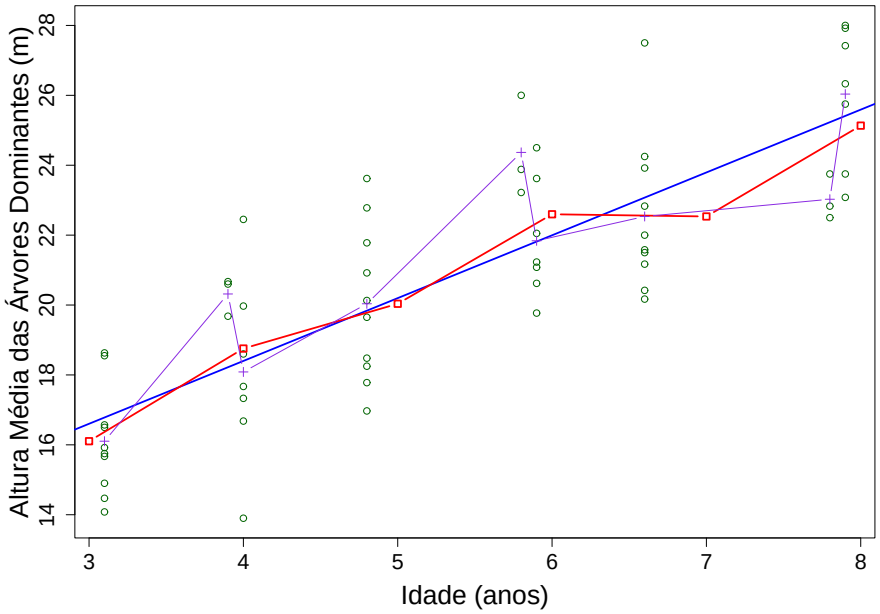


Figura 3.4: Relação entre a altura média das árvores dominantes e idade em povoamentos de *Eucalyptus grandis* na região central do Estado de São Paulo. Linha azul: modelo linear simples; linha vermelha: médias por classes de idade arredonda para ano; linha violeta: médias por classes de idade arredonda para décimo de ano.

- Teste F da falta de ajuste:

FONTE DE VARIAÇÃO	gl	Soma de Quadrados	Quadrado Médio	F	Prob. $F^* > F$
Modelo de Regressão	1	499.49	499.49	111.31	4.103×10^{-15}
Resíduos	58	260.26	4.49		
Falta de Ajuste	4	18.582	4.6455	1.038	0.3962
Resíduo Puro	54	241.68	4.48		

- **Conclusão:** não se rejeita H_0 do teste da falta de ajuste, i.e., não há evidências suficientes para dizer que o modelo linear não é apropriado.
- Comparação dos modelos pelo AIC:

$$\begin{aligned} \text{AIC}\{\text{M.COMPLETO}\} &= 331.2566 \\ \text{AIC}\{\text{M.REDUZIDO}\} &= 335.7011 \\ \Delta_{\text{FALTA AJUSTE}} &= 4.444489 \end{aligned}$$

  Conclusão: o modelo completo é superior ao modelo reduzido.
- Mas vejamos uma comparação de como foram definidas as classes de idade:

$$\begin{aligned} \text{AIC}\{\text{ARREDONDAMENTO } 1\} &= 319.5352 \\ \text{AIC}\{\text{ARREDONDAMENTO } 0\} &= 331.2566 \\ \Delta_{\text{ARREDONDAMENTO}} &= 11.72135 \end{aligned}$$

  Conclusão: o modelo com arredondamento em décimo de ano é superior ao modelo com arredondamento em ano !!!!!!!!
- **Qual o problema do AIC?**

  Falta robustez! Basta uma observação equivocada para o modelo completo ser superior ao modelo linear!

3.2 Variável Preditora Estocástica

- Os modelos lineares simples discutidos até o momento sempre assumiam que a variável preditora é não estocástica, de modo que o único efeito aleatório no modelo provem do erro.
- Ao fazer a re-definição do MLEG sem a pressuposição de X ser fixo, podemos tomar o modelo anterior fazendo cada propriedade dos erros condicional ao valores observados da variável preditora ($\mathbf{X} = \mathbf{x}$) temos:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

onde

$$\begin{aligned} &\boldsymbol{E}\{\varepsilon|\mathbf{X} = \mathbf{x}\} = 0 \\ &\boldsymbol{Var}\{\varepsilon|\mathbf{X} = \mathbf{x}\} = \sigma^2 \\ &\boldsymbol{Cov}\{\varepsilon_i, \varepsilon_i\} = 0 \\ &(\varepsilon|\mathbf{X} = \mathbf{x}) \overset{\text{iid}}{\sim} N(0, \sigma^2) \end{aligned}$$

- Ao assumirmos que os modelos definidos até aqui são condicionais aos valores observados da variável preditora ($\mathbf{X} = \mathbf{x}$), não ocorre nenhuma mudança na forma de obter os EQM (e também os EMV).

$$\begin{aligned} b_0 &\Rightarrow (b_0|\mathbf{X} = \mathbf{x}) \\ b_1 &\Rightarrow (b_1|\mathbf{X} = \mathbf{x}) \\ \boldsymbol{Var}\{b_0\} &\Rightarrow \boldsymbol{Var}\{b_0|\mathbf{X} = \mathbf{x}\} \\ \boldsymbol{Var}\{b_1\} &\Rightarrow \boldsymbol{Var}\{b_1|\mathbf{X} = \mathbf{x}\} \end{aligned}$$

- Vejamos como ficam as propriedades desses estimadores:

  Estimadores Não Viciados: para verificarmos isso basta tomarmos os valores esperados em relação à variação aleatória da variável preditora ($\mathbf{X}$).
  O valor esperado incondicional será o valor esperado do valor esperado condicional:

$$\begin{aligned} \boldsymbol{E}_{\mathbf{X}}\{b_0\} &= \boldsymbol{E}\{\boldsymbol{E}\{b_0|\mathbf{X} = \mathbf{x}\}\} = \boldsymbol{E}\{\beta_0\} = \beta_0 \\ \boldsymbol{E}_{\mathbf{X}}\{b_1\} &= \boldsymbol{E}\{\boldsymbol{E}\{b_1|\mathbf{X} = \mathbf{x}\}\} = \boldsymbol{E}\{\beta_1\} = \beta_1 \end{aligned}$$

  Variância dos Estimadores: para encontrarmos a variância é necessário considerar a variabilidade da variável preditora:

$$\begin{aligned} \boldsymbol{Var}_{\mathbf{X}}\{b_0\} &= \boldsymbol{E}_{\mathbf{X}}\{\boldsymbol{Var}\{b_0|\mathbf{X} = \mathbf{x}\}\} + \boldsymbol{Var}_{\mathbf{X}}\{\boldsymbol{E}\{b_0|\mathbf{X} = \mathbf{x}\}\} \\ &= \boldsymbol{E}_{\mathbf{X}}\left\{\sigma^2\left[\frac{1}{n} + \frac{\overline{x}^2}{\sum(x_i - \overline{x})^2}\right]\right\} + \boldsymbol{Var}_{\mathbf{X}}\{\beta_0\} \\ &= \sigma^2\left[\frac{1}{n} + \boldsymbol{E}_{\mathbf{X}}\left\{\frac{\overline{x}^2}{\sum(x_i - \overline{x})^2}\right\}\right] \end{aligned}$$

$$\begin{aligned} \boldsymbol{Var}_{\mathbf{X}}\{b_1\} &= \boldsymbol{E}_{\mathbf{X}}\{\boldsymbol{Var}\{b_1|\mathbf{X} = \mathbf{x}\}\} + \boldsymbol{Var}_{\mathbf{X}}\{\boldsymbol{E}\{b_1|\mathbf{X} = \mathbf{x}\}\} \\ &= \boldsymbol{E}_{\mathbf{X}}\left\{\frac{\sigma^2}{\sum(x_i - \overline{x})^2}\right\} + \boldsymbol{Var}_{\mathbf{X}}\{\beta_1\} \\ &= \sigma^2\,\boldsymbol{E}_{\mathbf{X}}\left\{\frac{1}{\sum(x_i - \overline{x})^2}\right\} \end{aligned}$$

  Ou seja, a variância depende do comportamento da variável preditora, o que só pode ser definido com pressuposições adicionais.

  Intervalos de Confiança: como calcular intervalos de confiança se a variância dos estimadores depende das propriedades de $\mathbf{X}$?

A idéia de valore esperado considerando a variação de $\mathbf{X}$ como sendo o valor esperado do valor esperado condicional pode ser aplicada a intervalos de confiança (Rice, 1995):

$$\begin{aligned} \boldsymbol{E}_{\mathbf{X}}\{IC(\beta_k)\} &= \boldsymbol{E}\{\boldsymbol{E}\{IC(\beta_k|\mathbf{X}=\mathbf{x})\}\} \\ &= \boldsymbol{E}\{(1-\alpha)\} \\ &= 1-\alpha. \end{aligned}$$

Conclusão: mesmo não conhecendo exatamente a variância dos estimadores, a coeficiente de confiança dos intervalos calculados com base na variância condicional continuam válidos.

Qualidade dos Estimadores: foi mostrado que para variável preditora não estocástica, os EMQ são os melhores estimadores para o MLEG.

Isso significa que condicionalmente aos valores observados ($\mathbf{X}=\mathbf{x}$) os EQM são os melhores estimadores.

No caso da variável preditora estocástica, os EQM são os melhores para cada um dos conjuntos observáveis valores de $\mathbf{X}$ e, consequentemente, são os melhores estimadores quando variável preditora é estocástica.

CONCLUSÃO FINAL: o Teorema de Gauss continua válido no caso de variável preditora estocástica.

3.3 Erro de Mensuração nas Variáveis

- Um outro aspecto que ainda não foi abordado é a questão de erros aleatórios de mensuração nas variáveis preditora e resposta.
- Na área florestal, os erros de mensuração são considerados negligenciáveis e são geralmente ignorados no ajuste de modelos de regressão. Entretanto é importante conhecer as conseqüências da presença de tais erros.
- Vejamos o que ocorre se as variáveis resposta e preditora forem medidas com erro aleatório de mensuração:

$$\begin{aligned} y &= y^* + v \qquad v \overset{\text{iid}}{\sim} N(0,\sigma_v^2) \\ x &= x^* + u \qquad u \overset{\text{iid}}{\sim} N(0,\sigma_u^2) \end{aligned}$$

onde:

y^* é a medida verdadeira da variável resposta;

v é o erro aleatório de mensuração da variável resposta;

x^* é a medida verdadeira da variável preditora; e

u é o erro aleatório de mensuração da variável preditora.

- O MLEG nesse caso fica:

$$\begin{aligned} y &= \beta_0 + \beta_1 x^* + \varepsilon + v \\ &= \beta_0 + \beta_1 x^* + \varepsilon^* \end{aligned}$$

ou seja, qualquer erro de mensuração na variável resposta pode ser absorvido pelo termo do erro aleatório, sendo válido a teoria desenvolvida até o momento.

- No caso da variável preditora a situação se complica:

$$\begin{aligned} y &= \beta_0 + \beta_1 x^* + \varepsilon \\ &= \beta_0 + \beta_1 (x - u) + \varepsilon \\ &= \beta_0 + \beta_1 x - \beta_1 u + \varepsilon \\ &= \beta_0 + \beta_1 x + [\varepsilon - \beta_1 u] \\ &= \beta_0 + \beta_1 x + w \end{aligned}$$

- O problema nesse caso é que a variável preditora (x) e o erro aleatório (w) não são independentes:

$$\boldsymbol{Cov}\{x, w\} = \boldsymbol{Cov}\{x^* + u, \varepsilon - \beta_1 u\} = -\beta_1 \sigma_u^2.$$

- Isso viola uma pressuposição fundamental do modelo linear que é a independência entre a variável preditora e o comportamento do erro.

- O resultado é que os EQM se tornam viciados (e os de EMV se tornam inconsistentes) com viés proporcional à magnitude do erro de mensuração (σ_u^2).

- A solução é utilizar um modelo que considere os erros de mensuração como parte integrante do modelo, o qual terá a forma:

$$y = \beta_0 + \beta_1 x + \theta u + \varepsilon.$$

3.4 Exercícios

Dados de povoamentos de *Eucalyptus grandis* na região central do Estado de São Paulo.

CLASSE DE IDADE	IDADE (ANOS)	DIÂMETRO MÉDIO (CM)	CLASSE DE IDADE	IDADE (ANOS)	DIÂMETRO MÉDIO (CM)
3.0	3.1	6.89	6.0	5.9	8.43
3.0	3.1	8.48	6.0	5.9	10.37
3.0	3.1	7.74	6.0	5.9	9.54
3.0	3.1	7.26	6.0	5.9	9.08
3.0	3.1	7.81	6.0	5.9	9.71
3.0	3.1	7.44	6.0	5.9	8.92
3.0	3.1	9.03	6.0	5.9	11.01
3.0	3.1	8.62	6.0	5.8	11.53
3.0	3.1	7.54	6.0	5.8	10.36
3.0	3.1	7.70	6.0	5.8	10.61
4.0	4.0	7.84	7.0	6.6	8.89
4.0	4.0	9.59	7.0	6.6	10.70
4.0	4.0	8.62	7.0	6.6	9.85
4.0	4.0	8.16	7.0	6.6	9.40
4.0	4.0	8.90	7.0	6.6	10.04
4.0	4.0	8.21	7.0	6.6	9.23
4.0	4.0	10.15	7.0	6.6	11.33
4.0	3.9	10.66	7.0	6.6	11.96
4.0	3.9	9.47	7.0	6.6	10.96
4.0	3.9	9.81	7.0	6.6	11.23
5.0	4.8	7.98	8.0	7.9	9.39
5.0	4.8	9.73	8.0	7.9	11.36
5.0	4.8	8.91	8.0	7.8	10.53
5.0	4.8	8.42	8.0	7.8	9.91
5.0	4.8	9.16	8.0	7.8	10.45
5.0	4.8	8.26	8.0	7.9	9.57
5.0	4.8	10.18	8.0	7.9	11.64
5.0	4.8	10.94	8.0	7.9	12.26
5.0	4.8	9.71	8.0	7.9	11.38
5.0	4.8	10.12	8.0	7.9	11.60

1. Ajuste o MLEG para os dados acima.
2. Realize o teste F da Qualidade do ajuste.
3. Realize o teste F da Falta de ajuste.
4. Construa um gráfico com os dados e adicione o modelo reduzido (reta de regressão) e o modelo completo do teste da falta de ajuste.

5. Encontre os valores de AIC relacionados ao teste da falta de ajuste.

CAPÍTULO 4
DIAGNÓSTICO E MEDIDAS REMEDIADORAS

Toda a teoria apresentada assume que as pressuposições dos modelos lineares simples são realistas para o conjunto de observações nos quais a análise de regressão será aplicada. Embora tenham sido mostrados modelos com vários graus de exigência em termos de pressuposições, o Modelo Linear de Erros Gaussianos é o que nos permite maior gama de inferências. Ao aplicar esse modelo, entretanto, é prudente verificar se as pressuposições assumidas são de fato realistas para os dados em mãos. Para isso discutiremos várias técnicas de *diagnóstico*.

Frequentemente, algumas das pressuposições não são realistas, tornando o modelo inapropriado, mas algumas medidas podem ser tomadas para solucionar o problema. Chamaremos tais medidas de *medidas remediadoras*.

4.1 Diagnósticos de Problemas nos Modelos Lineares

- A abordagem de diagnósticos priveligia a utilização de gráficos sobre testes e análises numéricas.
- Embora as pressuposições sempre se refiram ao erro (ε), é importante familiarizar-se com o comportamento da variável resposta (Y).

4.1.1 Diagnósticos envolvendo a Variável Resposta

- O diagnóstico gráfico-visual do comportamento da variável resposta envolve geralmente a utilização dos gráficos:

  gráfico de densidade: permite visualizar a forma da distribuição de Y;

gráfico de caixa (“box-plot”): investiga a forma da distribuição enfatizando os quantis da distribuição (1o., mediana, 3o.), a distância inter-quartil e observações discrepantes.

gráfico cronológico: busca detectar prováveis dependências temporais;

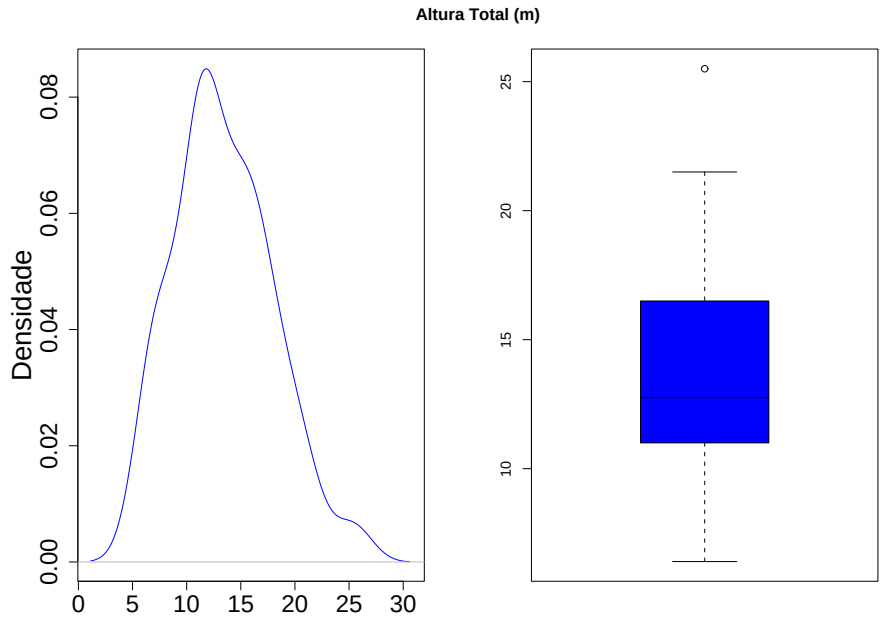
gráfico ramo-folha(“stem-leaf”): gráfico *numérico* que permite analisar a forma da distribuição e a presença de observações discrepantes.

- O objetivo principal é detectar tendências relativas a
  - assimetria;
  - curtose; e
  - observações discrepantes.

4.1.2 Exemplos de Diagnóstico da Variável Resposta

Exemplo 1: Relação Altura-Diâmetro em Árvores de *E. saligna*

A variável altura apresenta certa assimetria à esquerda e o padrão geral se distancia da dist. Gaussiana.



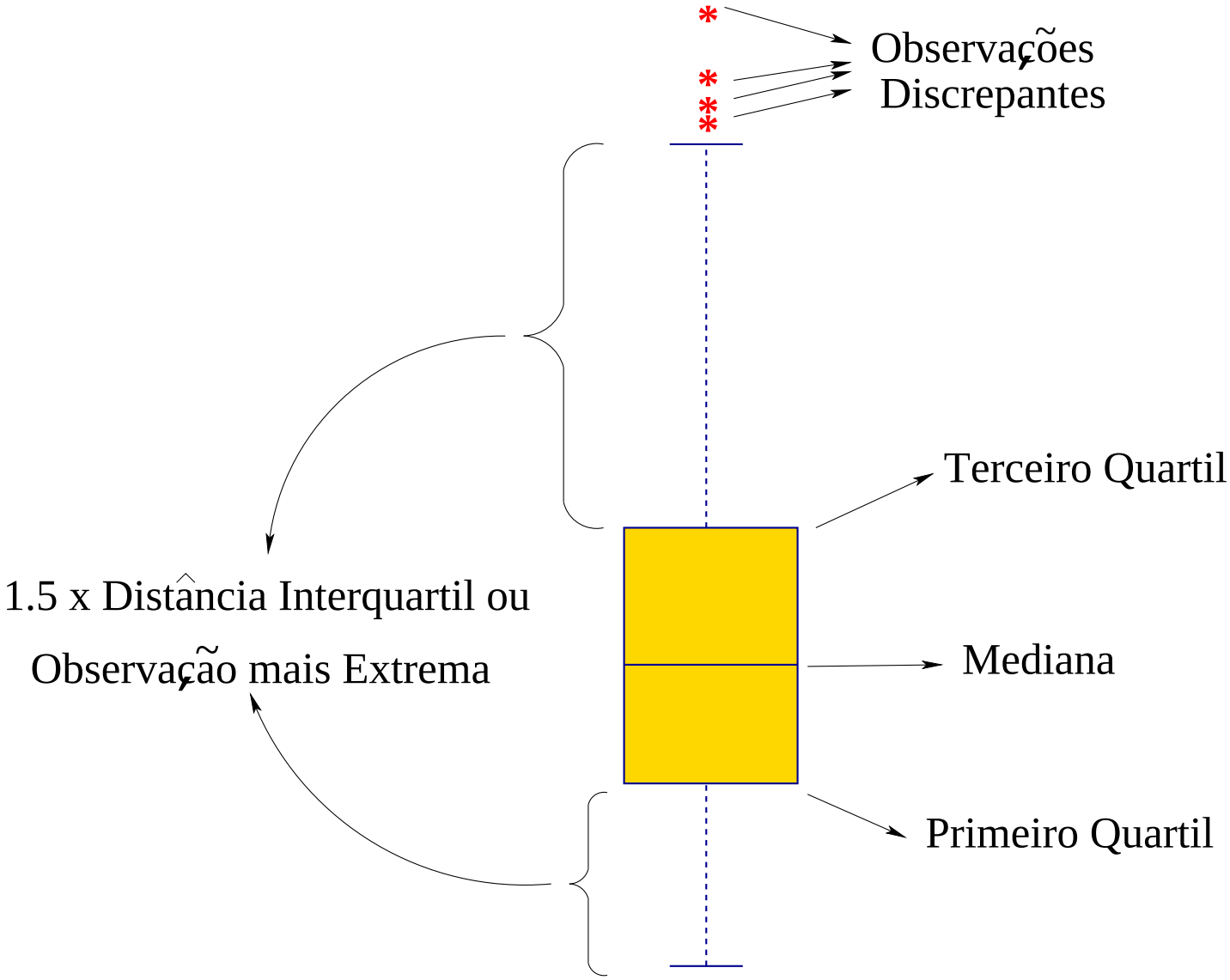
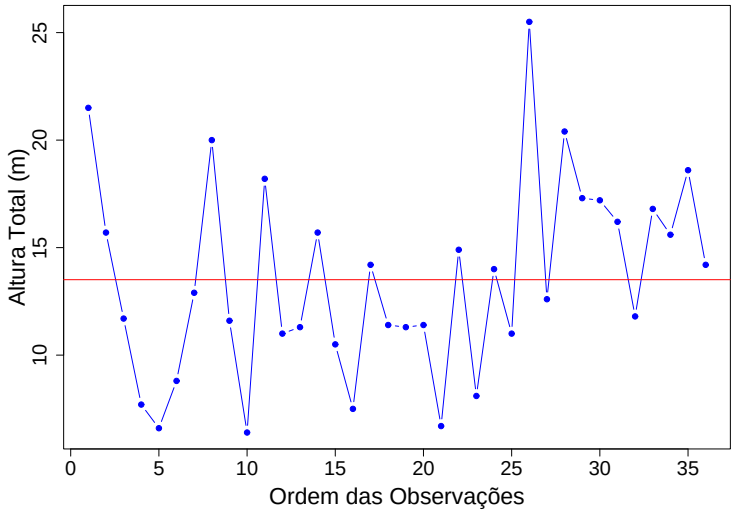


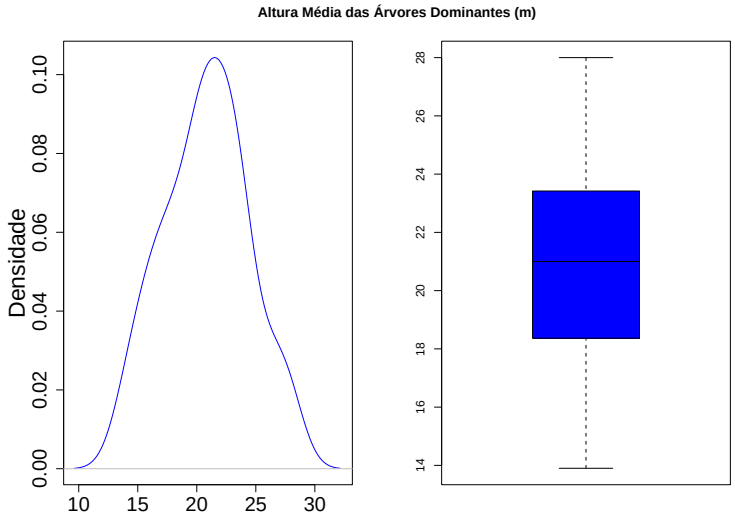
Figura 4.1: Figura que mostra a estrutura do gráfico de caixa ("*box-plot*"). A caixa é delimitada pelo primeiro e terceiro quartil, tendo como a mediana como uma linha interior. As linhas que saem da caixa se expandem até $1,5$ vêzes a distância interquartil ou até a observação mais extrema. A partir dessa distância, as observações são marcadas individualmente como observações discrepantes. O conceito de observação discrepante pode ser mudado alterando-se a distância até onde a linhas se expandem. Para se representar também o tamanho da amostra, a largura da caixa pode ser proporcional ao número de observações.

O gráfico cronológico não mostra nenhuma tendência marcante nos dados, indicando ausência de dependência temporal.

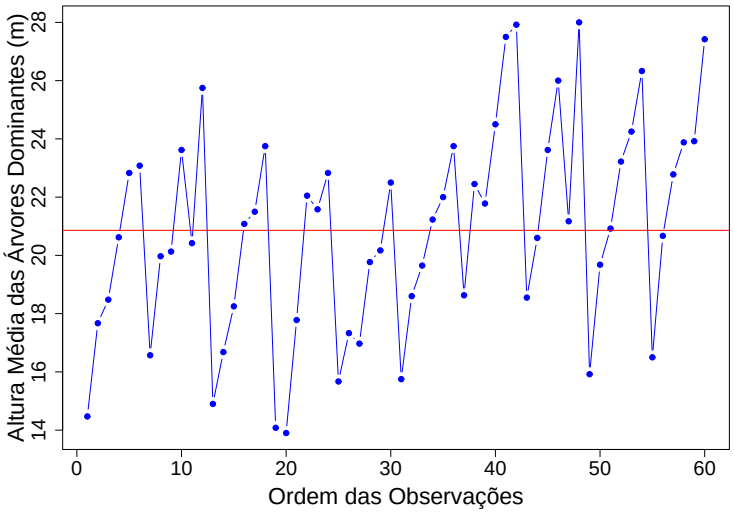


Exemplo 2: Altura das Árvores Dominantes em Povoamentos de *E. grandis*

A altura média das árvores dominantes se mostra bastante simétrica e com curtose adequada.

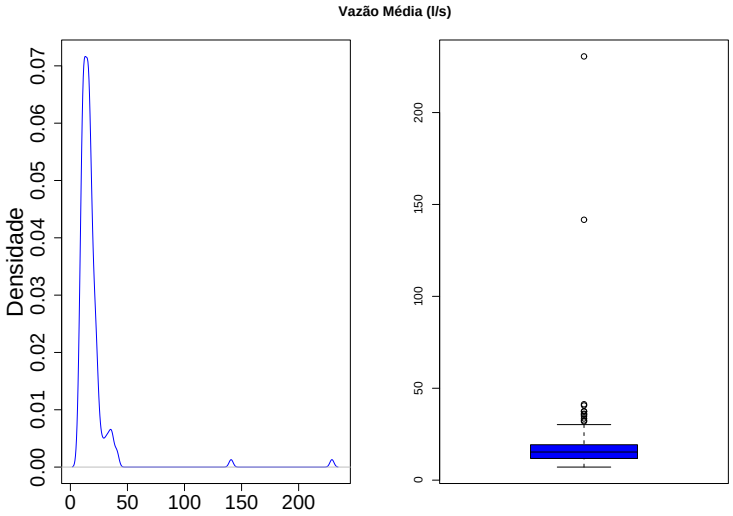


O gráfico revela a existência de tendência tempora fruto do fato dos dados resultarem de parcelas que foram re-medidas ao longo do tempo, denotando assim o padrão de crescimento dos povoamentos com a idade.

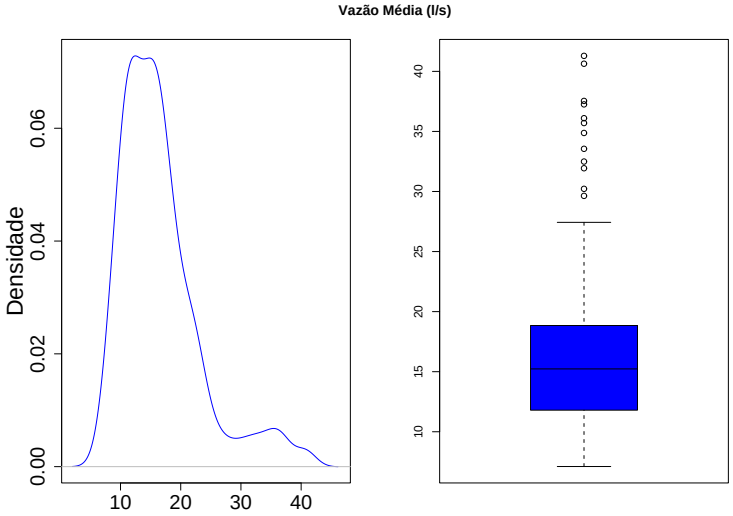


Exemplo 3: Vazão Média (Quinzenal) do Rio Tinga (E.E. de Itatinga)

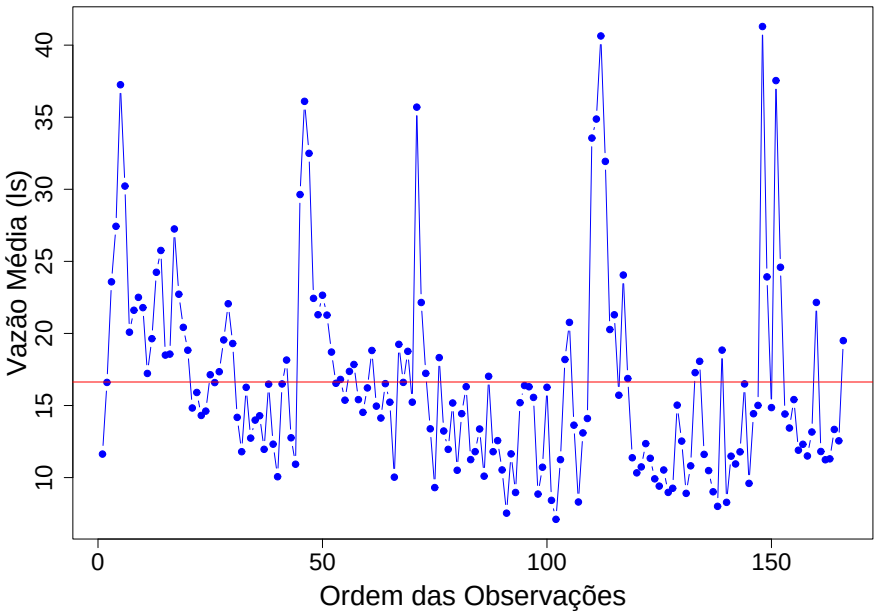
Os gráficos mostram a presença de duas observações discrepantes que impedem uma análise mais criteriosa do comportamento da variável vazão.



Elimando-se as duas observações discrepantes, nota-se que a variável vazão tem forma bastante assimétrica à esquerda, com tendência a platicurtose na região central dos valores.



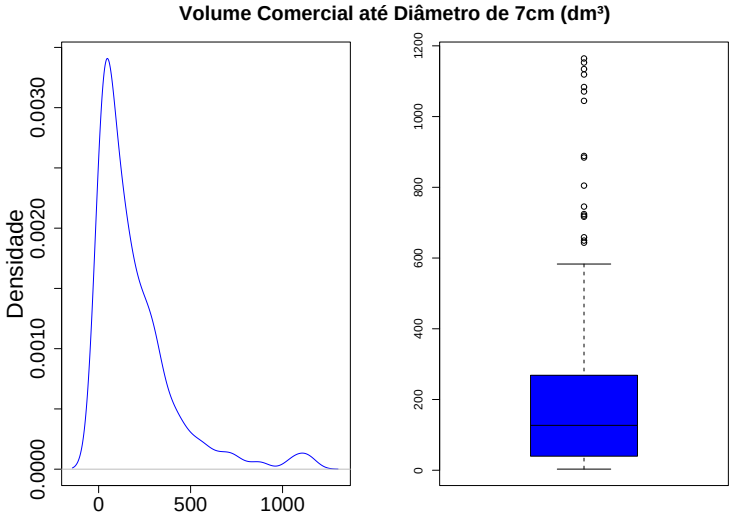
Como os dados de vazão são de quinzenas sucessivas, a ordem dos dados representa uma ordem cronológica e os dados mostram claramente uma dependência temporal.



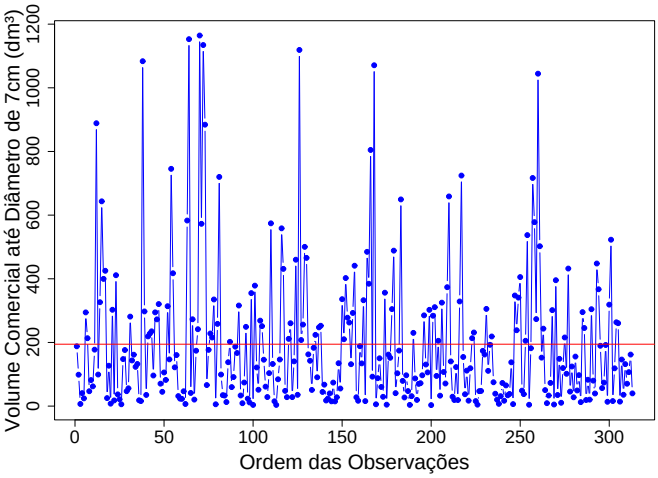
Exemplo 4: Volume de Árvores de Caixeta (*Tabebuia cassinoides*)

Os dados se referem ao volume comercial até o diâmetro de utilização de 7 cm de árvores de caixeta.

A variável resposta tem forte simetria à esquerda, mas não parece apresentar problema de curtose.



O gráfico do volume na ordem de observação revela certa alternância de valores grandes e pequenos, mas não demonstra nenhum padrão de dependência temporal.



4.1.3 O Resíduo

- Para diagnóstico de problemas com as pressuposições no modelo linear, o resíduo é o elemento fundamental pois ele representa a estimativa do termo de erro do modelo, que é o elemento sobre qual as pressuposições são colocadas.
- É importante, no entanto, manter bem clara a diferênça entre **resíduo**

$$e_i = y_i - \widehat{y}_i$$

e erro aleatório

$$\varepsilon_i = y_i - \boldsymbol{E}\{Y|X = x_i\}.$$

- Note que a maneira de definir o resíduo implica em:

resíduo positivo: valor ajustado ($\widehat{y}_i$) **subestima** o valor observado (y_i); e

resíduo negativo: valor ajustado ($\widehat{y}_i$) **superestima** o valor observado (y_i).

PROPRIEDADES DO RESÍDUO

- É importante lembrar que o processo de ajuste do modelo linear (Método de Quadrados Mínimos) impõe ao resíduo algumas propriedades.
- Isso deve ser considerado para não confundirmos tais propriedades com aquilo que esperamos observar no resíduo por ele ser um estimador do erro do modelo.

Média: como $\sum e_i = 0$, temos que

$$\overline{e} = \frac{\sum e_i}{n} = 0.$$

Qualquer diferença em relação ao valor nulo na média do resíduo resulta de erros de arredondamento.

Note que a pressuposição $\boldsymbol{E}\{\varepsilon_i\} = 0$, não pode ser verificada por essa via.

Não independência: os resíduos não são independentes entre si, pois:

$$e_i = y_i - \widehat{y}_i = y_i - (b_0 + b_1 x_i),$$

ou seja, todos os resíduos dependem das EQM para os coeficientes de regressão.

Associados ao resíduo temos **apenas** $n-2$ graus de liberdade e sabemos que:

$$\begin{aligned}\sum e_i &= 0\\ \sum e_i x_i &= 0\\ \sum e_i \widehat{y}_i &= 0\end{aligned}$$

Entretanto, os graus de liberdade ($n-2$) em grandes amostras é praticamente igual ao tamanho da amostra (n), de modo que, na prática, a dependência entre resíduos pode ser ignorada no diagnóstico do modelo linear, pois exerce muito pouca influência sobre.

Variância: a variância do resíduo é obtida através da soma de quadrados do resíduo:

$$\widehat{\boldsymbol{Var}}\{e_i\} = \frac{\sum(e_i-\overline{e})^2}{n-2} = \frac{\sum e_i^2}{n-2} = \frac{SQR}{n-2} = QMR.$$

Já vimos que o QMR é um estimador não viciado da variância do erro ($\boldsymbol{E}\{QMR\} = \sigma^2$).

Um outro aspecto que devemos notar é em cada nível de X, a variância do resíduo continua sendo um bom estimador da variância do erro, mesmo no caso da variância do erro não ser constante:

$$\begin{aligned}\boldsymbol{Var}\{e_{ij}|X=x_i\} &= \boldsymbol{E}\left\{e_{ij}^2|X=x_i\right\} - \left[\boldsymbol{E}\{e_{ij}|X=x_i\}\right]^2\\ &= \boldsymbol{E}\left\{e_{ij}^2|X=x_i\right\}\\ &= \boldsymbol{E}\left\{\varepsilon_{ij}^2|X=x_i\right\}\\ &= \boldsymbol{Var}\{\varepsilon_{ij}|X=x_i\}\\ &= \sigma_i^2\end{aligned}$$

RESÍDUO PADRONIZADO

- As pressuposições fundamentais do MLEG podem ser apresentadas na forma

$$\varepsilon_i \overset{\text{iid}}{\sim} N(0,\sigma^2),$$

logo, podemos definir um *erro padronizado* por

$$\frac{\varepsilon_i - \overline{\varepsilon}}{\sigma} = \frac{\varepsilon_i}{\sigma} \sim N(0,1).$$

- De modo análogo, podemos definir o **resíduo padronizado** por

$$\frac{e_i - \overline{e}}{\sqrt{QMR}} = \frac{e_i}{\sqrt{QMR}} \sim t(n-2),$$

  ou seja, no caso das pressuposições do MLEG serem realistas, o resíduo padronizado segue a distribuição t de Student com $n-2$ graus de liberdade.

- No caso de grandes amostras, podemos lançar mão da aproximação

$$\frac{e_i}{\sqrt{QMR}} \sim N(0,1),$$

  que implica que 95% das observações devem ter resíduo entre **2** e **-2**, aproximadamente (exatamente seria 1.96 e -1.96).

- Essa aproximação estabelece um limite para encontarmos observações com resíduo discrepante.

4.1.4 Diagnóstico através da Análise do Resíduo

Uma série de problemas envolvendo o modelo linear, inclusive violações das pressuposições, podem ser diagnosticados através da análise do resíduo:

1. A relação entre variável resposta e variável preditora não é linear.
2. A variância do erro não é constante (*heteroscedasticidade*).
3. Os erros não são independentes.
4. O modelo se ajusta bem à maioria das observações, com exceção de algumas observações discrepantes.
5. O erro não tem distribuição Gaussiana.
6. Uma ou mais variáveis preditoras importantes foram omitidas do modelo (modelo linear múltiplo).

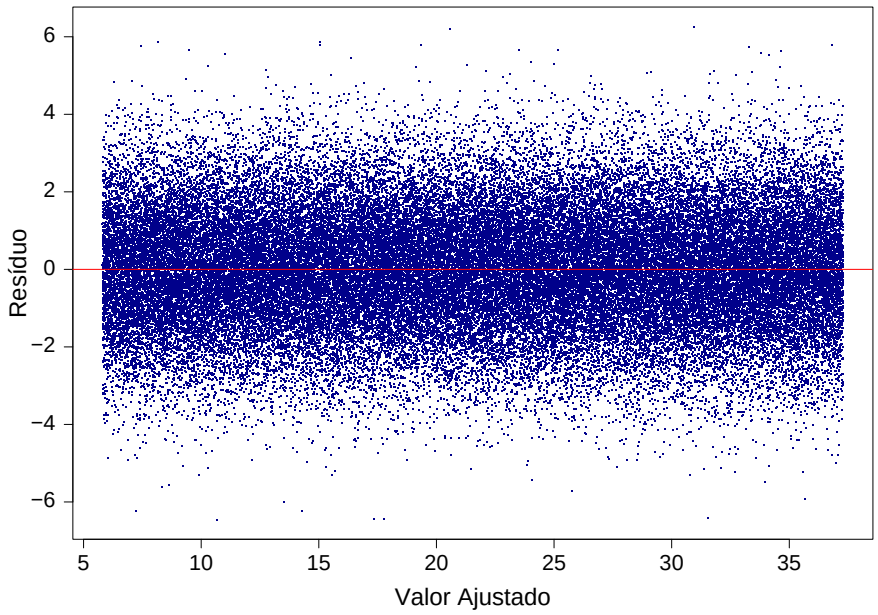
O diagnóstico do resíduos é realizado através de uma série de gráficos:

- Gráficos de dispersão do resíduo contra os valores ajustados: $e \times \widehat{y}$.
- Gráfico do resíduo em ordem cronológica de observação: $e \times t$ ou $e \times (1, 2, \ldots, n)$.

- Gráficos de dispersão do resíduo contra variáveis preditoras que foram omitidas do modelo.
- Gráfico de caixa (*box plot*) do resíduo, particularmente quando comparamos o comportamento do resíduo em diferentes *sub-grupos*.
- Gráfico Quantil-quantil: quantis empíricos do resíduo contra os quantis teóricos assumindo a distribuição Gaussiana.

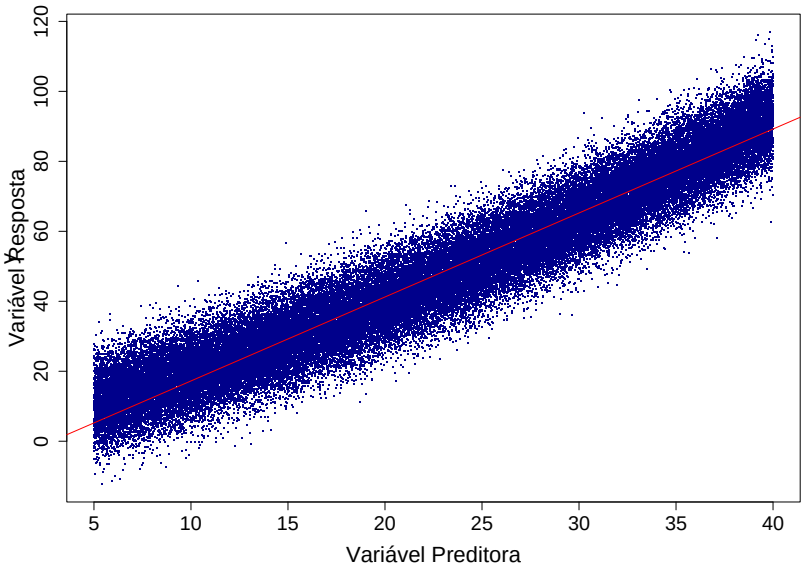
GRÁFICO DE DISPERSÃO DO RESÍDUO CONTRA OS VALORES AJUSTADOS

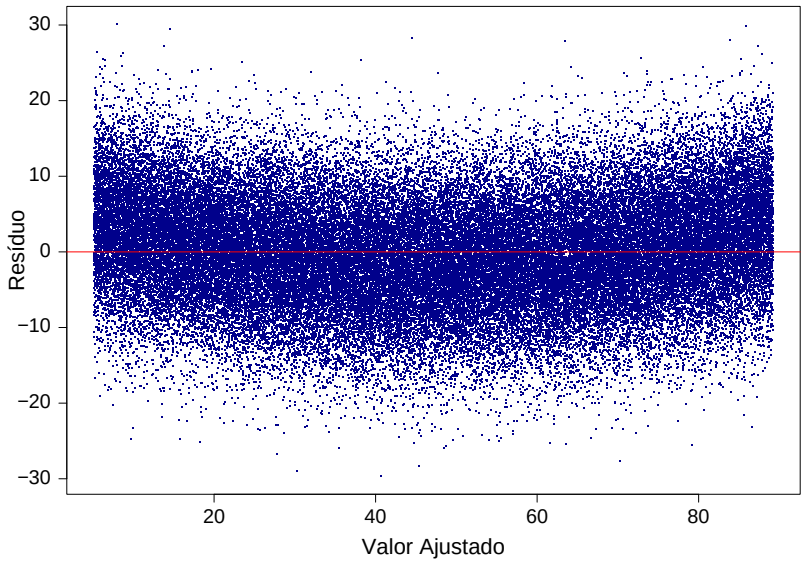
- O gráfico de dispersão do resíduo contra os valores ajustados permite avaliar os seguintes aspectos do modelo linear:
  - a relação entre variável resposta e variável preditora não é linear;
  - heteroscedasticidade;
  - magnitude do erro de predição em observações indiiduais.
- Exemplo do padrão experado para o gráfico de dispersão do resíduo:
  - faixa retangular de largura constante centrada no eixo das abscissas (eixo-x);
  - a densidade de pontos se reduz à medida que se distancia do eixo-x.



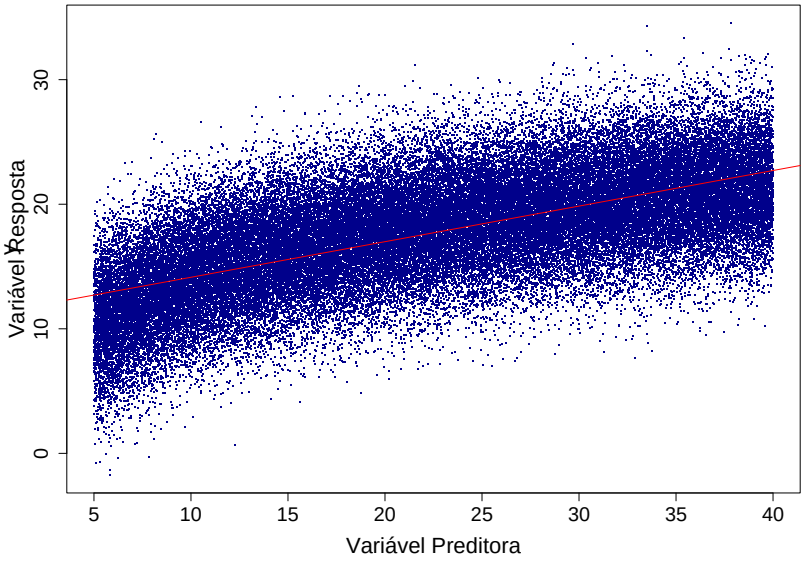
- Exemplos do que pode acontecer quando a relação entre variável resposta e variável preditora não é linear:

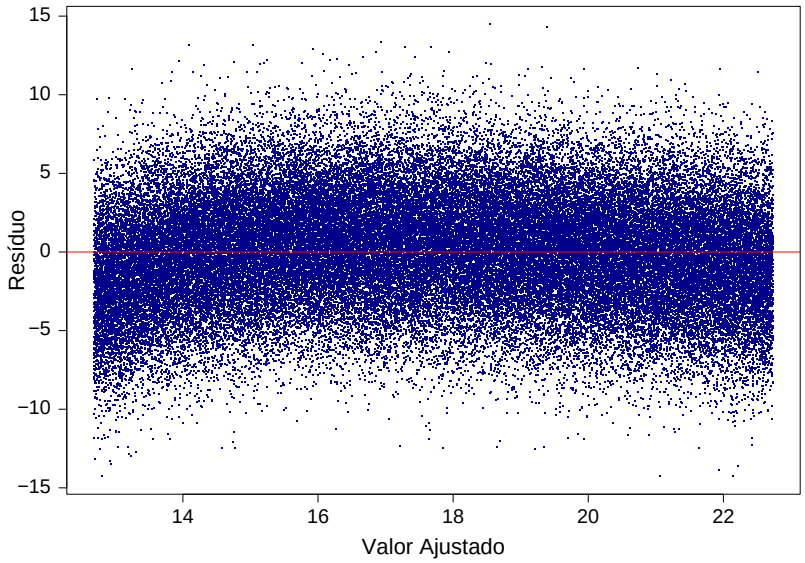
$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$$



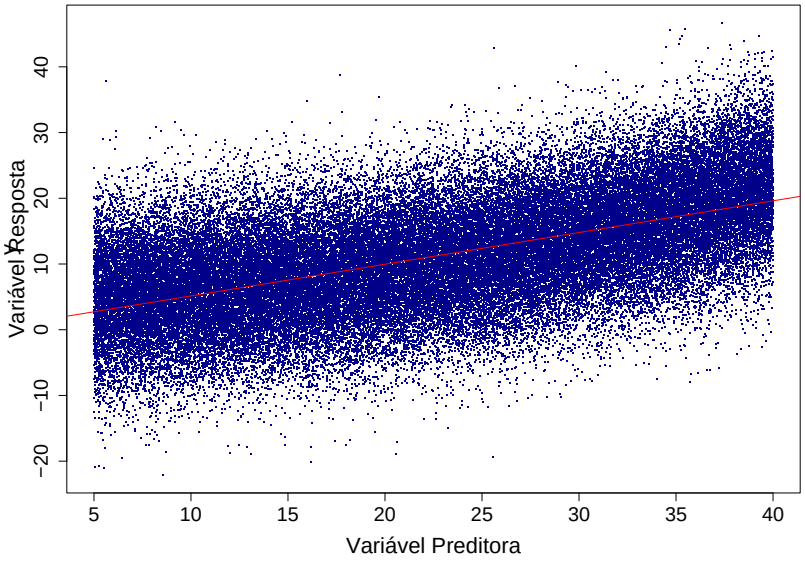


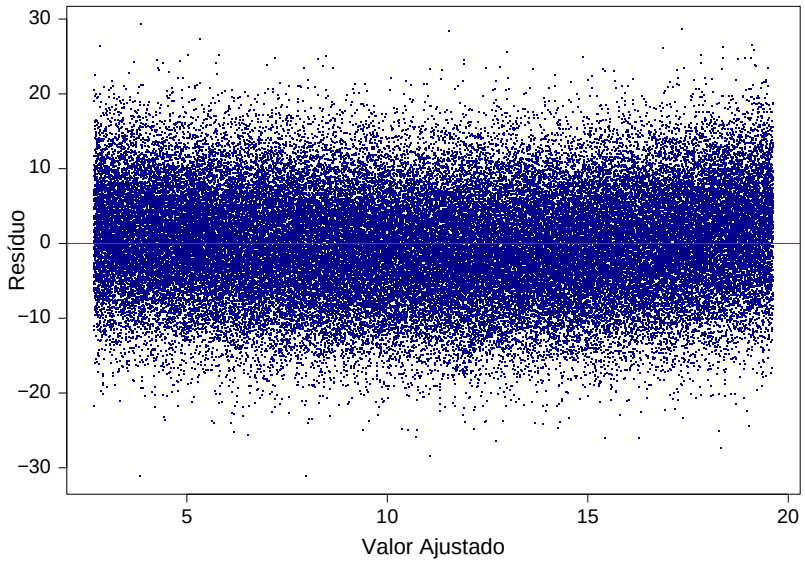
$$y_i = \beta_0 + \beta_1 \ln(x_i) + \varepsilon_i$$



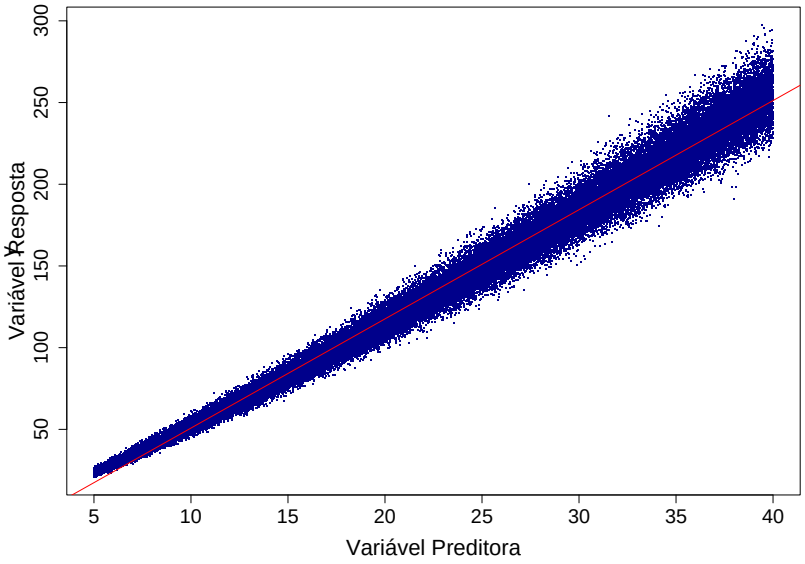


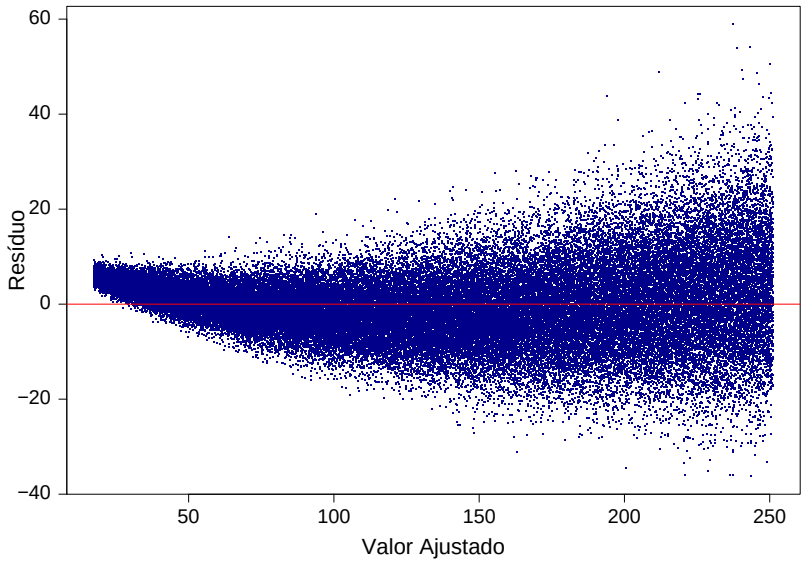
$$\ln(y_i) = \beta_0 + \beta_1 x_i + \varepsilon_i$$





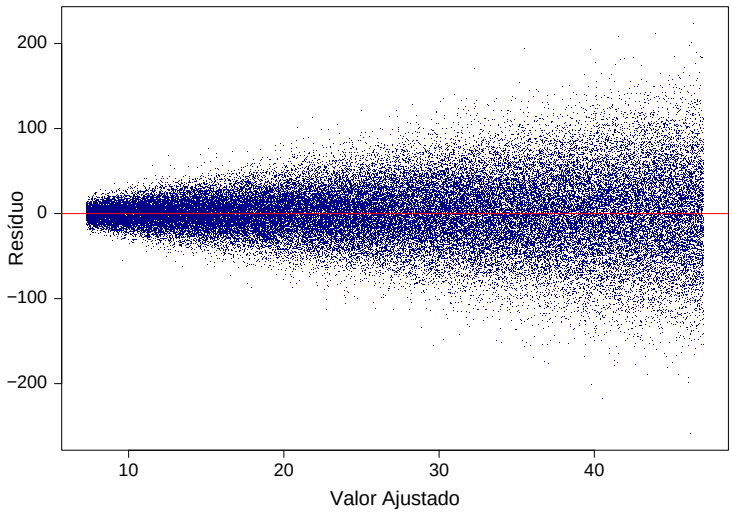
$$\ln(y_i) = \beta_0 + \beta_1 \ln(x_i) + \varepsilon_i$$





- Exemplos do que pode acontecer quando a relação entre variável resposta e variável preditora não linear, mas a variância não é constante:

$$\boldsymbol{Var}\{\varepsilon_i\} = k\, x_i$$



$$\boldsymbol{Var}\{\varepsilon_i\} = k\,(1/x_i)$$

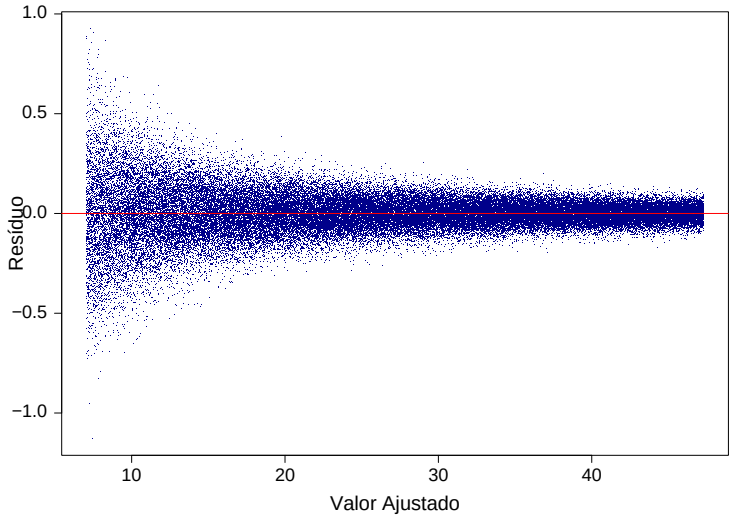


GRÁFICO DE DISPERSÃO DO PESÍDUO PADRONIZADO

Aspectos a Estudar: o gráfico de dispersão do resíduo padronizado contra a variável respostas permite observar os seguintes aspectos do modelo linear:

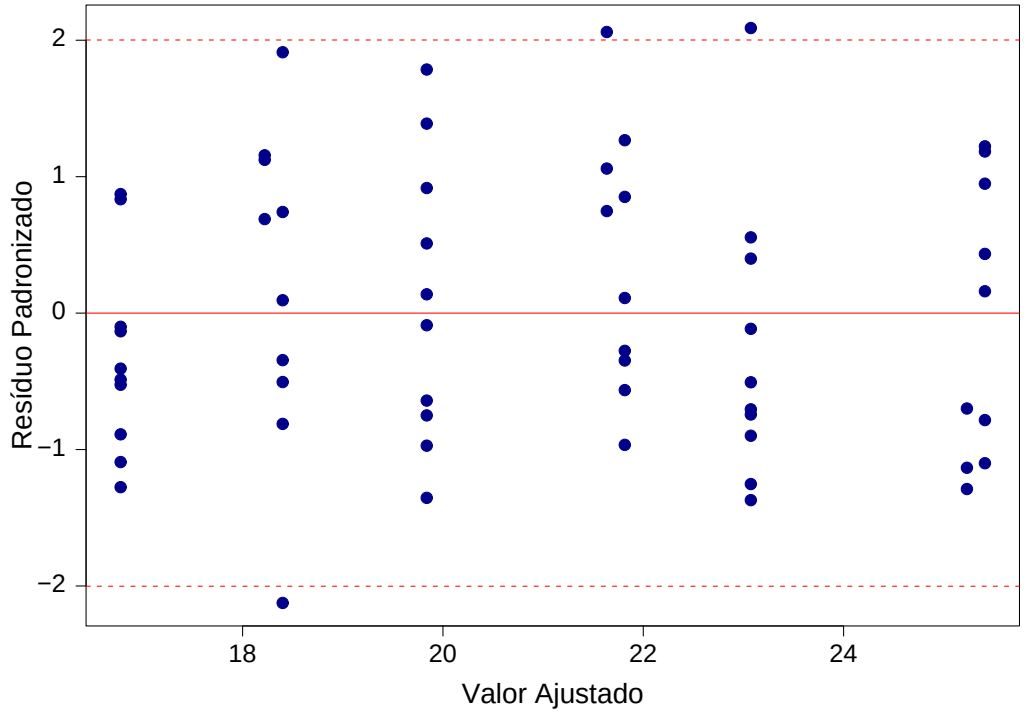
- a relação entre variável resposta e variável preditora não é linear;
- heteroscedasticidade;
- detectar observações discrepantes.

Exemplo: Altura das Árvores Dominantes em *E. grandis*

- Como o tamanho da amostra é $n = 60$, espera-se que o número de observações com resíduo padronizado acima de $|2|$ seja:

$$P(|Z| \geq 2) \times 60 \approx 0.05 \times 60 = 3.$$

- Observa-se no gráfico 3 observações além do limite $|2|$, o que sugere que tais observações não são discrepantes.



Exemplo: Relação entre Volume Comercial e Volume Cilíndrico em árvores de caixeta (*Tabebuia cassinoides*).

- Como o tamanho da amostra é $n = 313$, espera-se que o número de observações com resíduo padronizado acima de $|2|$ seja:

$$0.05 \times 313 = 15.65 \approx 16.$$

- Observa-se no gráfico 16 observações além do limite $|2|$.
- No entanto observa-se mais 3 observações além do limite $|4|$. O número de observações que se espera acima desse limite é:

$$P(|Z| \geq 4) \times 313 = (6.334248 \times 10^{-5}) \times 313 = 0.01982620 \approx 0.$$

Isso indica que de fato as observações são discrepantes!

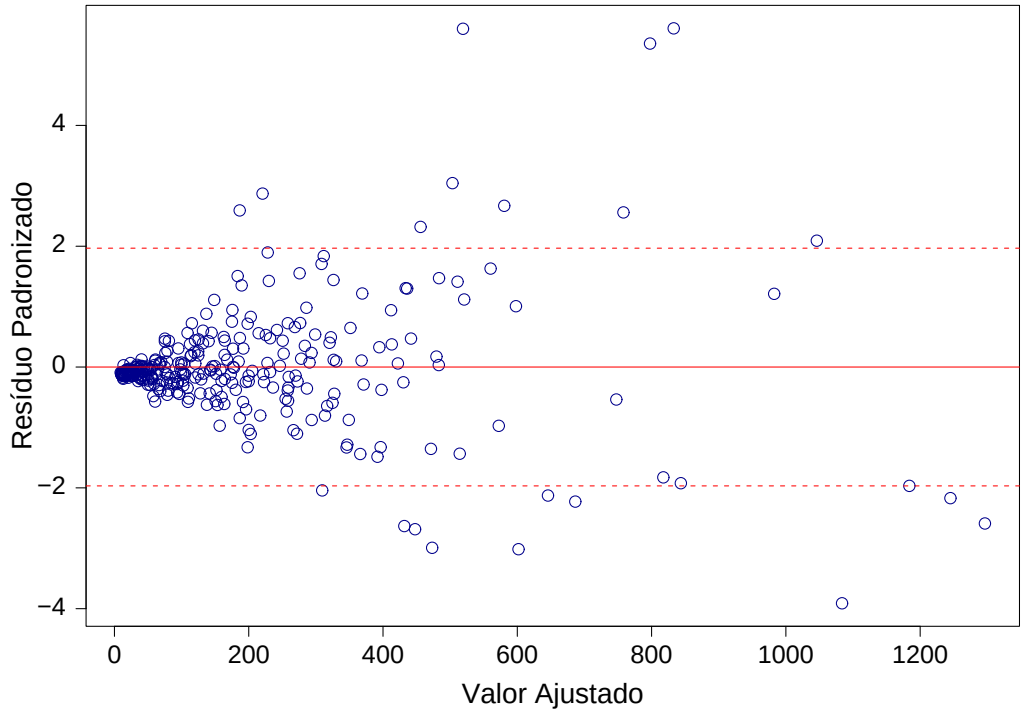


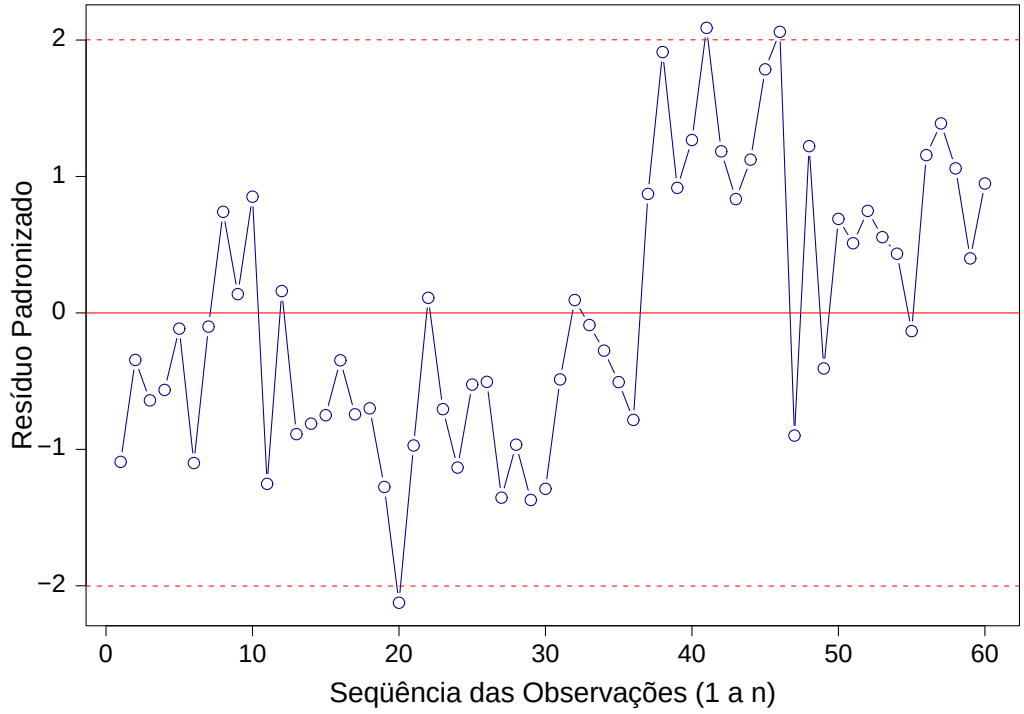
GRÁFICO CRONOLÓGICO

Objetivo: o objetivo de grafar o resíduo contra a ordem cronológica dos dados ou contar a seqüência das observações $(1, 2, \ldots, n)$ é tentar detectar dependências temporais

que possam violar a pressuposição de independência do erro.

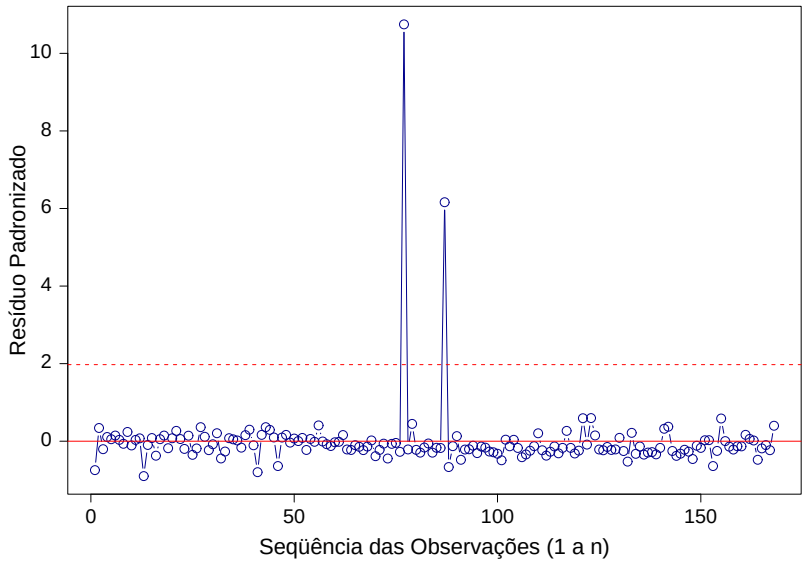
Exemplo: Altura Média das Árvores Dominantes em *E. grandis*.

- No caso da altura média foi detectado um padrão temporal na variável resposta.
- Mas ao se ajustar o modelo, esse padrão não se apresenta no resíduo padronizado.



Exemplo: Vazão média quinzenal do rio Tinga em função da precipitação na E.E. de Itatinga.

- Foi detectado um padrão temporal bem claro na variável resposta: por tratar-se de quinzenas sucessivas.
- Detecta-se no resíduo padronizado duas observações extremamente discrepantes.



- Ignorando-se as observações discrepantes, nota-se que o resíduo padronizado mantem um padrão de peridiocidade nos seus valores positivos e negativos revelando a presença de dependência temporal nos dados.

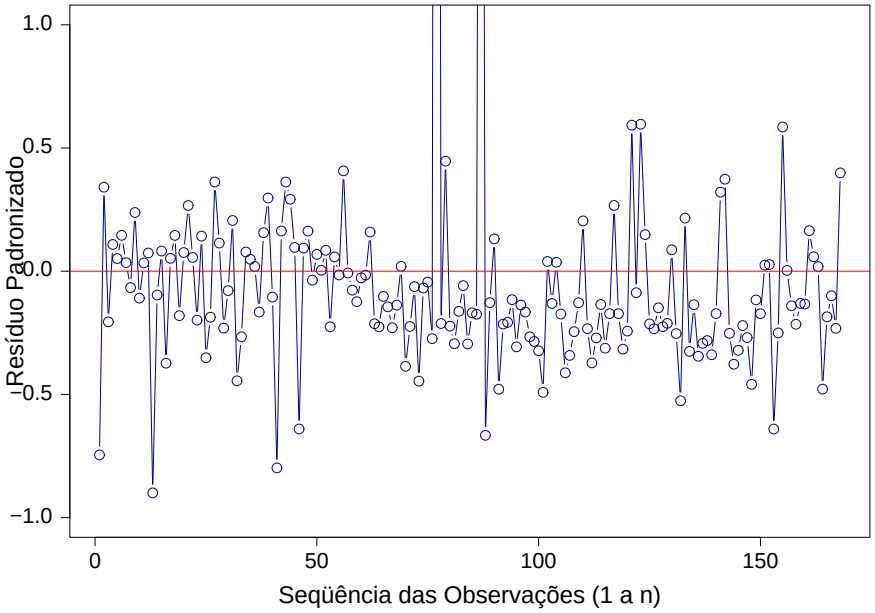


GRÁFICO QUANTIL-QUANTIL PARA DISTRIBUIÇÃO DO RESÍDUO

Gráfico Quantil-Quantil para Distribuição Gaussiana: compara os quantis empíricos do resíduo contra os quantis teóricos que se espera assumindo que o erro tem distribuição Gassiana.

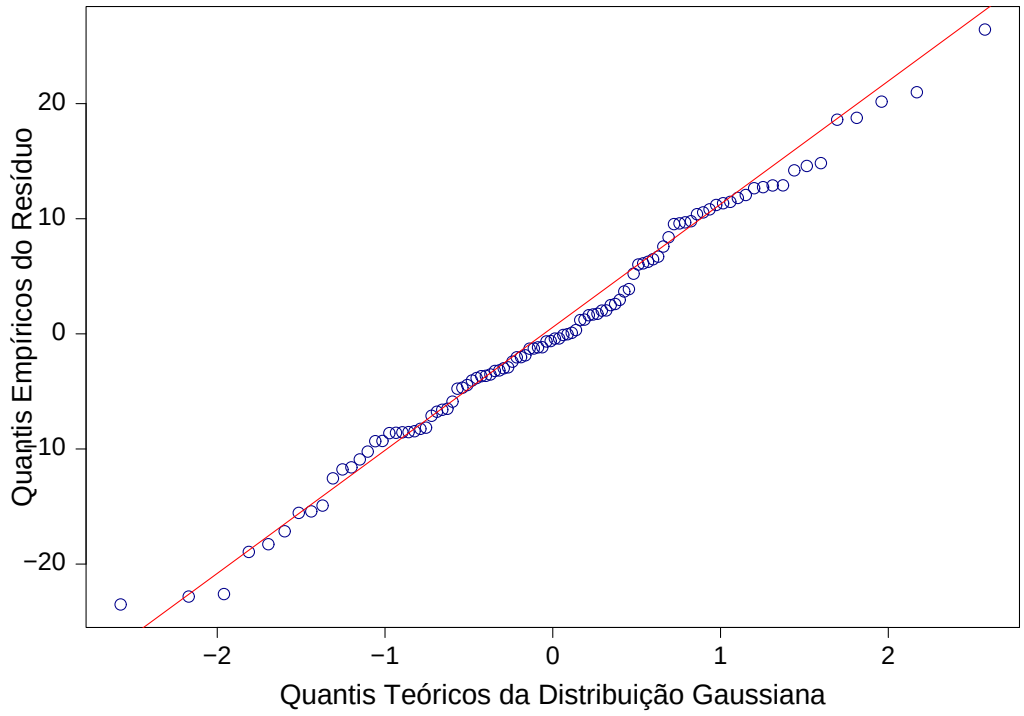
Quantis empíricos do resíduo: ordena-se o resíduo em ordem crescente, sendo que cada valor observado ordenado $e_{[i]}$ é definido como sendo o quantil o $i^{\text{ésimo}}$ quantil.

Quantis teóricos: utilizando a função de distribuição Gaussiana ($\Phi(\cdot)$), obtem-se o $i^{\text{ésimo}}$ quantil teórico pela expressão

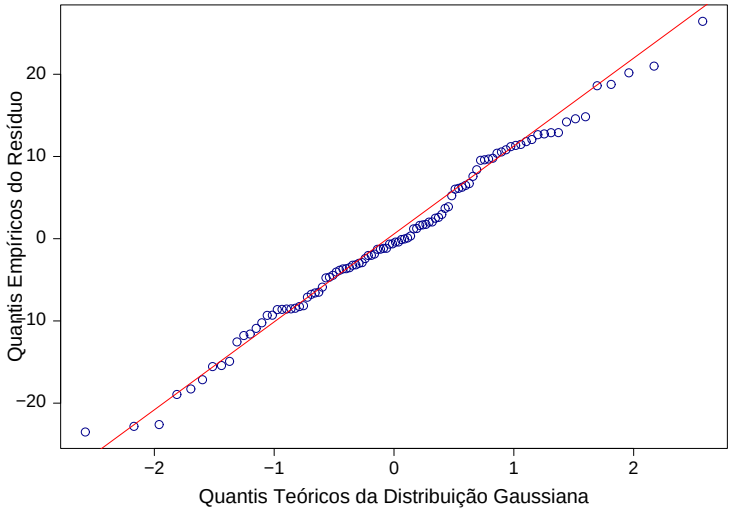
$$Z_{[i]} = \sqrt{QMR}\left[\Phi^{-1}\left(\frac{i-0.375}{n+0.25}\right)\right]$$

Gráfico: grafa-se $Z_{[i]}$ (nas abscissas) contra $e_{[i]}$ (nas ordenadas). Se o erro segue a distribuição Gaussiana, os quantis devem se alinhar ao longo de uma reta.

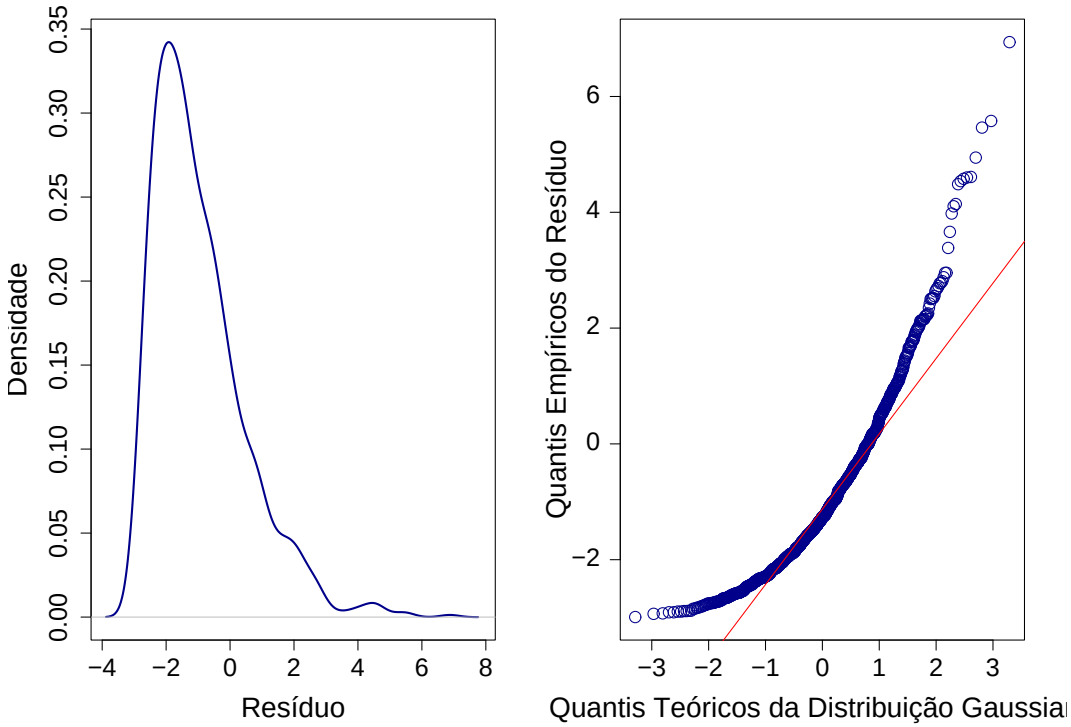
Exemplo Teórico: Distribuição Gaussiana com tamanho de amostra $n = 100$.



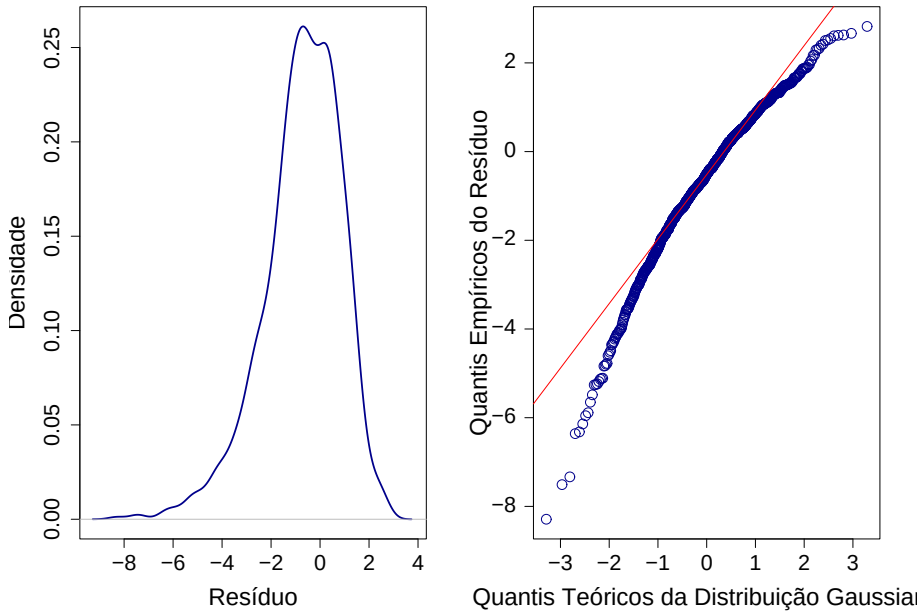
Exemplo Teórico: Distribuição Gaussiana com tamanho de amostra $n = 1000$.



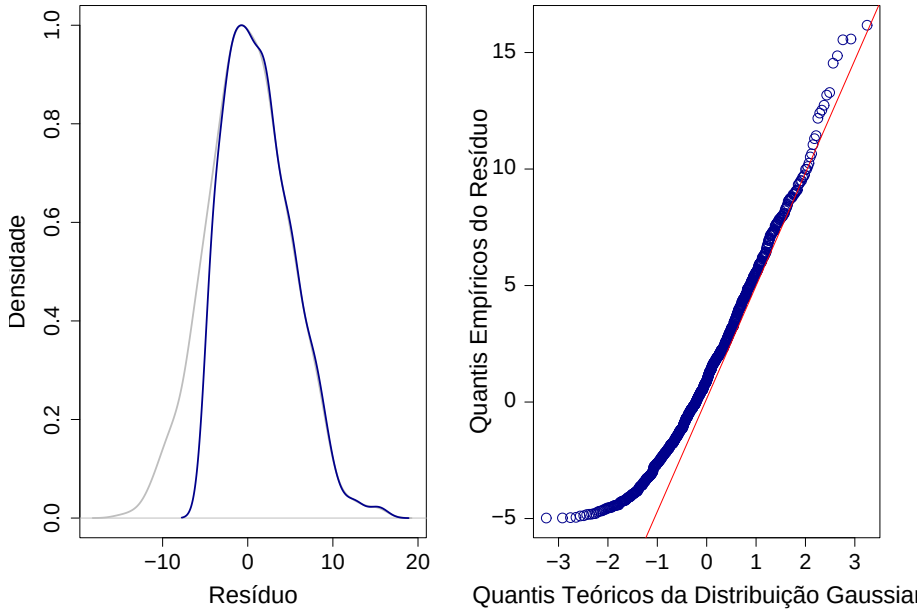
Exemplo Teórico: distribuição assimétrica à direita ($n = 1000$).



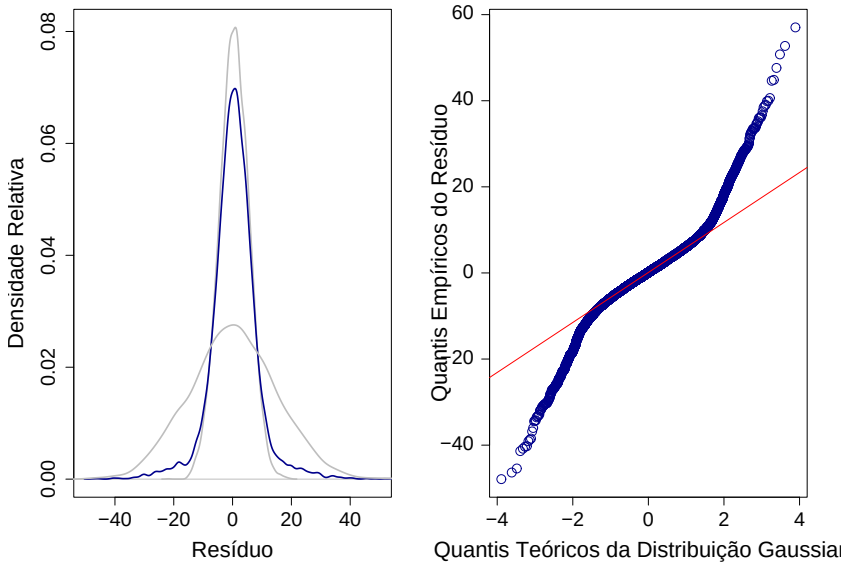
Exemplo Teórico: distribuição assimétrica à esquerda ($n = 1000$).



Exemplo Teórico: distribuição truncada à esquerda ($n = 1000$).



Exemplo Teórico: distribuição leptocúrtica ($n = 10000$) produzida pela *contaminação* de distribuição Gaussiana: $80\%G(0, 5) + 20\%G(0, 15)$.



Exemplo Teórico: distribuição platocúrtica ($n = 10000$).

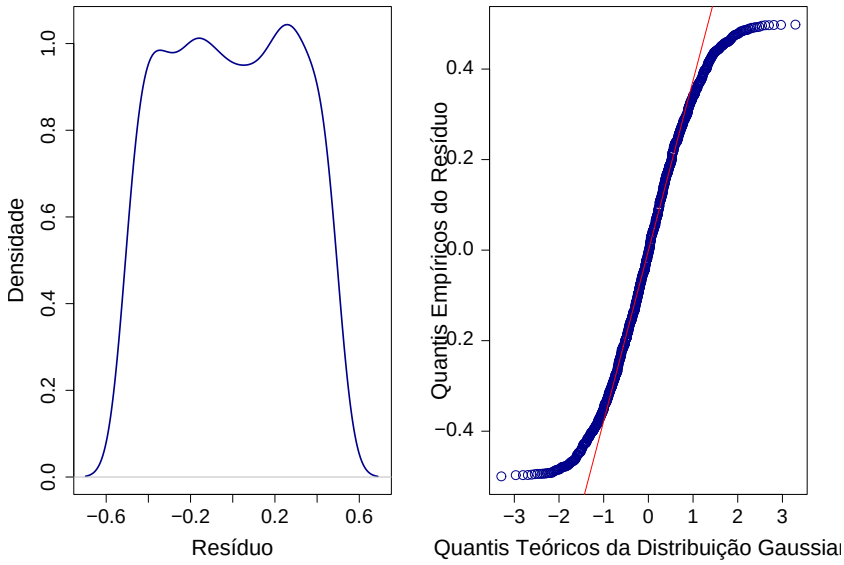


GRÁFICO DA RAIZ QUADRADA DO RESÍDUO ABSOLUTO

- Embora o gráfico de dispersão do resíduo contra valor ajustado permita verificar o problema de heteroscedasticidade, esse problema é mais eficientemente investigado fazendo o gráfico da raiz do resíduo absoluto ($\sqrt{|e_i|}$) contra o valor ajustado.

Exemplo: Volume de Árvores de Caixeta (*Tabebuia cassinoides*). Variância crescente com o valor ajustado.

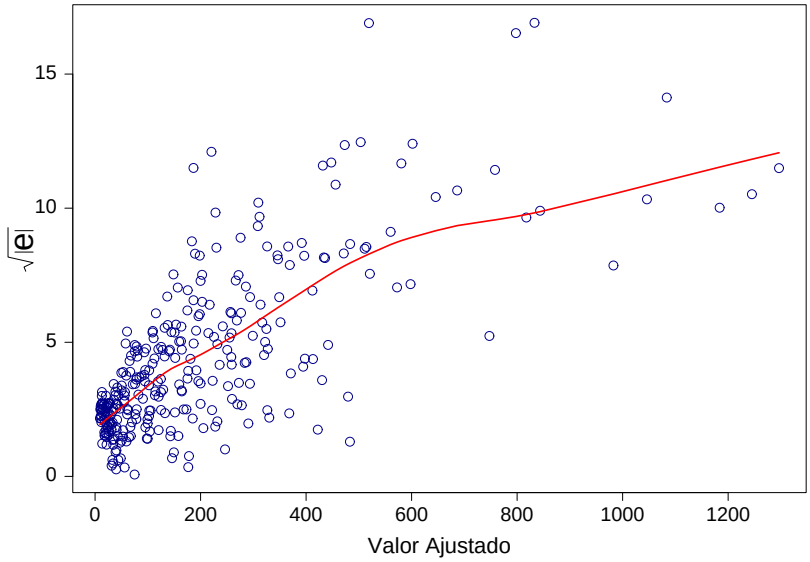
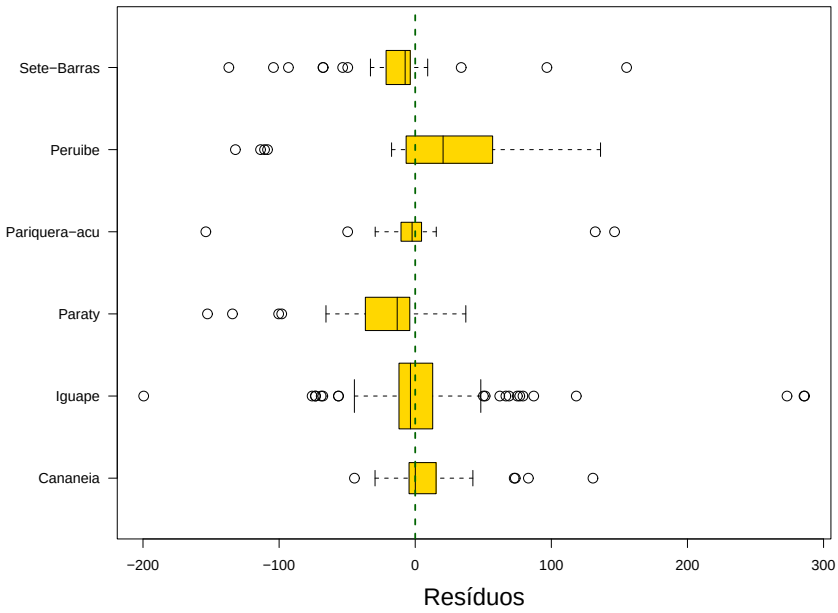


GRÁFICO DE CAIXA DO RESÍDUO

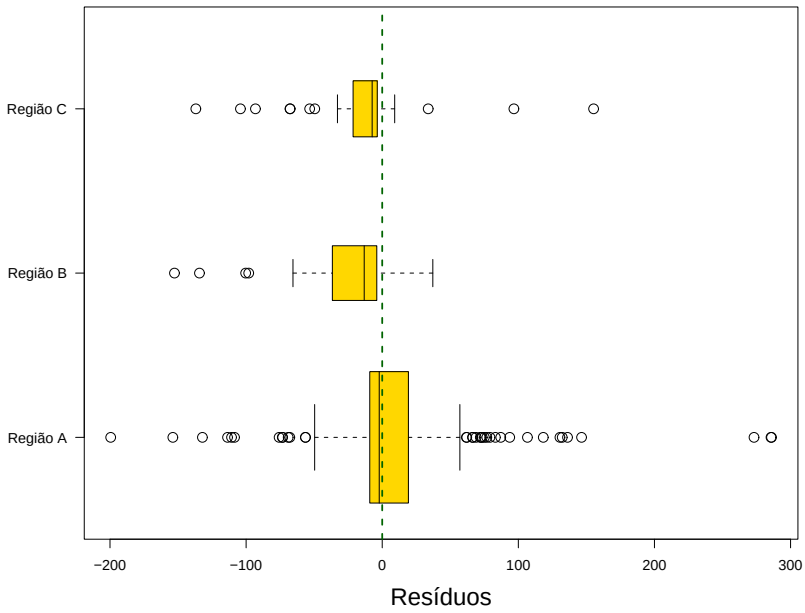
- O gráfico de caixa (*boxplot*) nos permite verificar a distribuição de uma variável aleatória.
- Isso é particularmente útil quando queremos estudar o comportamento do resíduo em diferentes *sub-grupos* do conjunto de dados que estamos analisando.
- Nesse caso interessa comparar o comportamento do resíduo em sub-grupos do conjunto de dados, mostrando como o modelo ajustado se comporta em diferentes situações de aplicação.

Exemplo: Volume de Árvores de Caixeta (*Tabebuia cassinoides*).

Comportamento do resíduo do modelo por Município da área de estudo.



Comportamento do resíduo do modelo por regiões da área de estudo.



4.2 Medidas Remediadoras nos Modelos Lineares

Quando realizamos uma análise de regressão e detectamos problemas no modelo linear simples, em última análise gostaríamos de tomar medidas que corrigissem estes problemas (medidas remediadoras).

4.2.1 Visão Geral de Problemas e Medidas Remediadoras

Relação funcional entre variável resposta e preditora não é linear. Nesse caso podemos:

- Procurar transformações da variável resposta e/ou a variável preditora, para verificar se a relação linear se estabelece numa outra escala, como na escala logarítmica, por exemplo.
- Entrar com novo termo no modelo: a variável preditora na forma quadrática ou cúbica.

  Isso corresponde a introduzir novas variáveis preditoras no modelo, tornando-o um modelo linear múltiplo.
- Abandonar o modelo linear simples e buscar um modelo não-linear.

Os erros não são independentes. Não há medida remediadora para esse problema.

A única solução é aplicar modelos que incorporem a dependência na estrutura dos erros, utilizando o **Método de Quadrados Mínimos Generalizados**.

Heteroscedasticidade. Duas medidas são possíveis:

- Se a variância *varia de modo sistemático* de acordo com a variável preditora, pode se utilizar o **Método dos Quadrados Mínimos Ponderados** (Regressão Ponderada).
- É possível também buscar uma transformação dos dados (variável resposta) para estabilizar a variância.

Os erros não seguem a distribuição Gaussiana. Como já vimos, esse é um problema menor, pois podemos construir muitas inferências úteis sem assumir essa pressuposição.

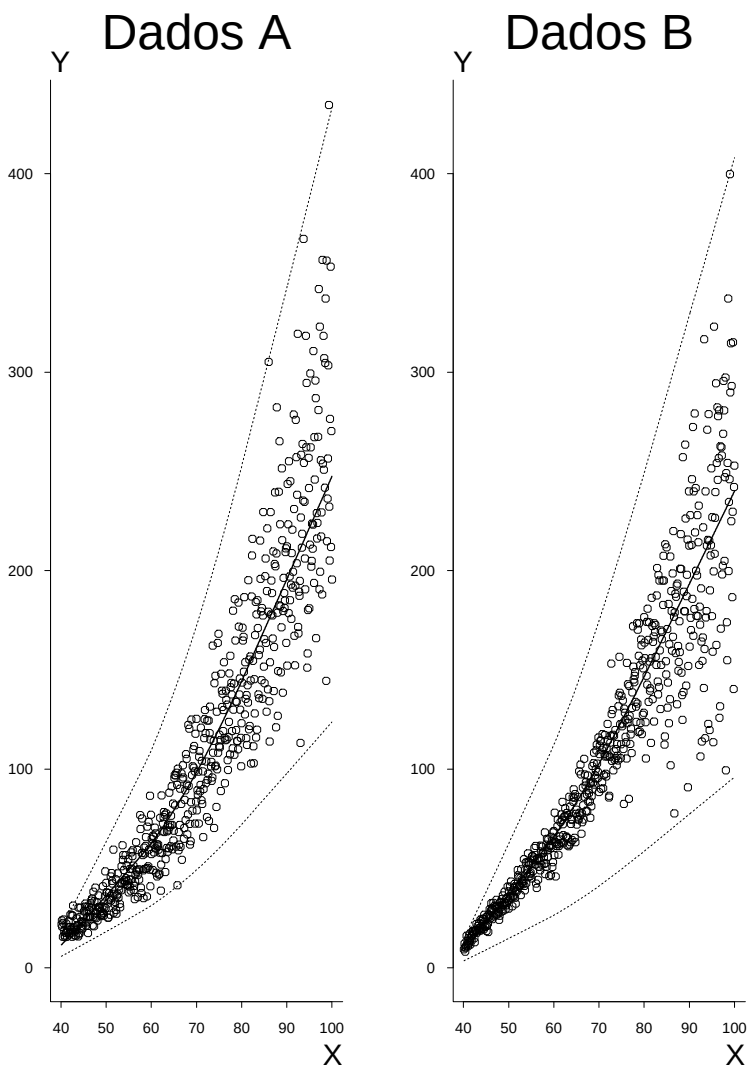
Se a distribuição Gaussiana for fundamental para a análise, a única medida é buscar transformações dos dados (variável resposta) para estabilizar a variância **e** gerar erros Gaussianos.

Freqüentemente isso não é possível! A solução nesse caso é utilizar modelos não Gaussianos, ou seja, **Modelos Lineares Generalizados**.

EXEMPLO CLÁSSICO ENVOLVENDO HETEROCEDASTICIDADE E DIST. GAUSSIANA

- Os dois conjuntos de dados abaixo sugerem uma relação funcional semelhante entre X e Y:

$$\boldsymbol{E}\{y_i\} \;=\; \beta_0\, x_i^{\beta_1}.$$



- Transformando esse modelo para escala logarítmica temos:

$$\begin{aligned} \boldsymbol{E}\{\ln(y)\} &= \ln(\beta_0) + \beta_1 \ln(x_i) \\ \boldsymbol{E}\{y^*\} &= \beta_0^* + \beta_1 x_i^* \end{aligned}$$

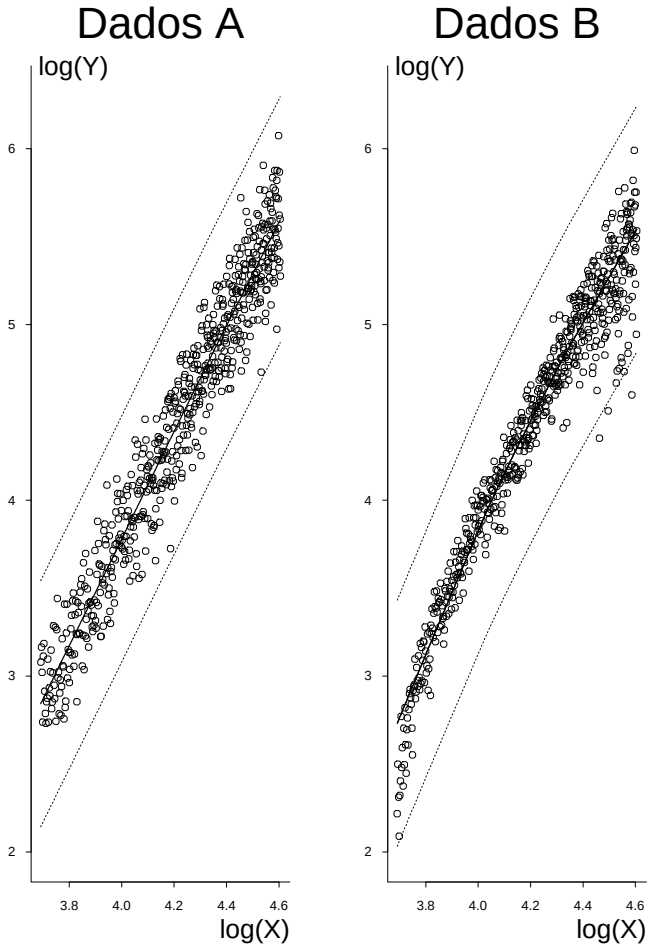
- Mas o que ocorre com a estrutura dos erros quando ajustamos o modelo na escala logarítmica ?

$$y_i^* = \beta_0^* + \beta_1 x_i^* + \varepsilon_i^* \Longrightarrow y_i = \beta_0\, x_i^{\beta_1}\, \varepsilon_i$$

onde

$$\begin{aligned} \varepsilon_i^* = \ln(\varepsilon_i) &\Rightarrow \varepsilon_i = \exp(\varepsilon_i^*) \\ \varepsilon_i^* \sim \text{Gaussiana}(0, \sigma^2) &\Rightarrow \varepsilon_i \sim \text{Lognormal} \end{aligned}$$

- Como ficam os dados com a transformação logarítimica?



Conclusão

- Se os erros têm distribuição **Lognormal** e são **multiplicativos**, então o modelo é

$$y_i = \beta_0 \, x_i^{\beta_1} \, \varepsilon_i$$

A transformação logarítimica

(a) homogeniza a variância dos erros e

(b) normaliza a distribuição dos erros.

Esse modelo deve ser ajustado na forma do modelo linear simples

$$\begin{aligned} \ln(y_i) &= \beta_0^* + \beta_1 \ln(x_i) + \ln(\varepsilon_i) \\ y_i^* &= \beta_0^* + \beta_1 x_i^* + \varepsilon_i^* \end{aligned}$$

- Se o modelo verdadeiro for:

$$y_i = \beta_0 \, x_i^{\beta_1} + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2)$$

a solução é utilizar **regressão não-linear**.

- Se o modelo verdadeiro for:

$$y_i = \beta_0 \, x_i^{\beta_1} + \varepsilon_i \qquad \varepsilon_i \sim N(0, \underline{\underline{x_i^m \sigma^2}})$$

a solução é **regressão não-linear ponderada**.

4.2.2 Transformações em Modelos Lineares

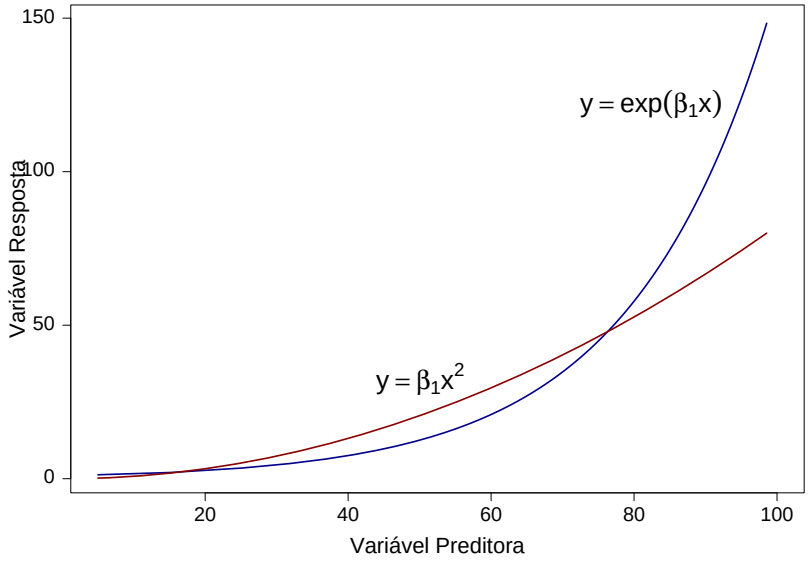
- As transformações são medidas remediadoras que *podem* solucionar vários problemas dos modelos lineares.

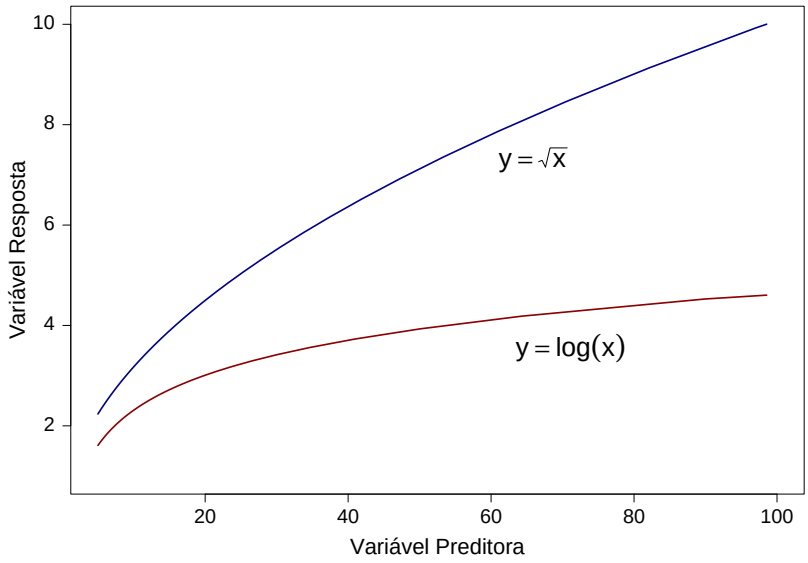
Transformações da Variável Preditora

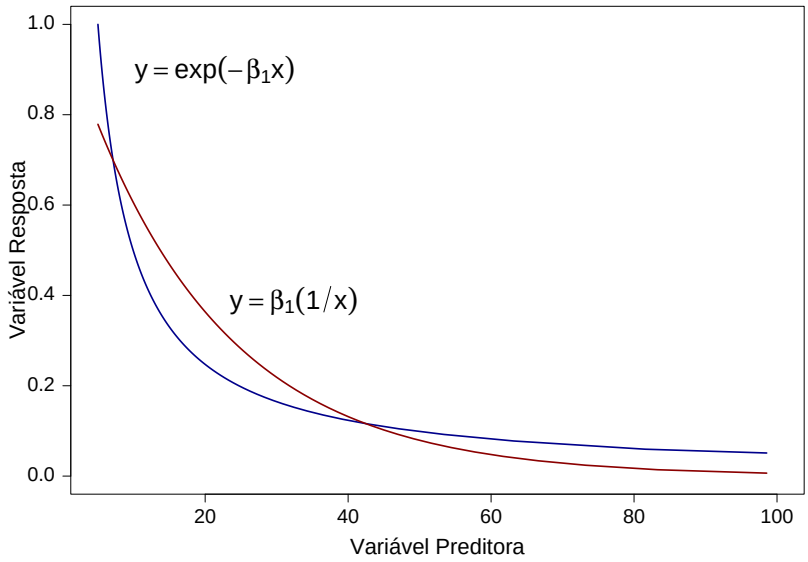
- Se o único problema no modelo é que a relação funcional não se apresenta linear **mas** o erro apresenta homocedasticidade e distribuição Gaussiana,

então a solução é *tranformar apenas a variável preditora*.

- Se a estrutura do erro é adequada, qualquer transformação na variável resposta alterará a estrutura do erro.
- Algumas transformações comumente realizadas na variável preditora:







Transformações da Variável Resposta

- Quando o modelo apresenta problema de heteroscedasticidade e os erros não seguem a distribuição Gaussiana, **então** a variável resposta **tem que ser transformada**.
- Alguns aspectos a considerar ao transformar a variável resposta são:

  1. As transformações mais comuns são:

$$\begin{aligned} Y^* &= \sqrt{Y} \\ Y^* &= \log_e(Y) \\ Y^* &= \frac{1}{Y} \end{aligned}$$

  2. A transformação para outras escala pode ser devida a considerações teóricas. Por exemplo: a transformação logarítimica de Y implica que a taxa de mudança é proporcional ao tamanho de Y:

$$\frac{dY}{dt} = kY$$

$$\begin{aligned}
\frac{dY}{Y} &= kdt \\
\int \frac{1}{Y} dY &= \int kdt \\
\log(Y) &= kt + c \\
\log(Y) &= b_0 + b_1 t \\
Y &= \mathrm{e}^{b_0}\mathrm{e}^{b_1 t} \\
Y &= \alpha \mathrm{e}^{b_1 t}
\end{aligned}$$

3. **Após** a transformação, devemos voltar aos gráficos de diagnóstico do resíduo e verificar se o remédio foi **efetivo**.
4. As transformações que estabilizam a variância quando esta muda de acordo com $\boldsymbol{E}\{Y\}$ são:

$$\begin{aligned}
\sigma_i^2 \propto \boldsymbol{E}\{Y_i\} &\Rightarrow Y^* = \sqrt{Y} \\
\sigma_i \propto \boldsymbol{E}\{Y_i\} &\Rightarrow Y^* = \log(Y) \\
\sqrt{\sigma_i} \propto \boldsymbol{E}\{Y_i\} &\Rightarrow Y^* = \frac{1}{Y}
\end{aligned}$$

5. Quando a variável resposta é transformada, **todos** os testes e inferências são realizados na **escala transformada**.

   Para os I.C. e I.P. serem interpretáveis, eles devem ser convertidos de volta à escala original.

4.3 Regressão Linear Ponderada

A regressão linear ponderada é a técnica mais utilizada para se contornar o problema de heteroscedasticidade nos dados.

4.3.1 Quadrados Mínimos Ponderados

- Já vimos que as EQM são obtidas minimizando a função da soma de quadrados do resíduo

$$Q = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2.$$

- No Método de Quadrados Mínimos (*Ordinários*) todas as observações têm igual importância na minimização de Q.
- Podemos generalizar a idéia do método de quadrados mínimos considerando a possibilidade de que a importância das observações individuais não é a mesma.
- Assim podemos crirar uma função de **Soma de Quadrados Ponderada do Resíduo**:

$$Q_w = \sum_{i=1}^{n} w_i(y_i - \beta_0 - \beta_1 x_i)^2.$$

onde w_i é o peso dado a cada observação individual ($i = 1, 2, \ldots, n$).

- A minimização de Q_w em relação aos coeficientes de regressão (β_0 e β_1) resulta num **Sistema de Equações Normais Ponderado**:

$$\begin{aligned} \sum w_i y_i &= b_0 \sum w_i + b_1 \sum w_i x_i \\ \sum w_i x_i y_i &= b_0 \sum w_i x_i + b_1 \sum w_i x_i^2 \end{aligned}$$

cuja a solução gera as **Estimativas de Quadrados Mínimos Ponderados** (EQMP):

$$\begin{aligned} b_1 &= \frac{\sum w_i x_i y_i - [(\sum w_i x_i)(\sum w_i y_i)/n]}{\sum w_i x_i^2 - [(\sum w_i x_i)^2/n]} \\ b_0 &= \frac{\sum w_i y_i}{\sum w_i} - b_1 \frac{\sum w_i x_i}{\sum w_i} \end{aligned}$$

4.3.2 Ponderação como Transformação do Modelo

- A ponderação da Soma de Quadrados do Resíduo implica numa transformação do modelo original.
- O modelo linear simples apropriado para ponderação é o MLEG, mas com erro heteroscedático:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad (\varepsilon_i | X = x_i) \sim N(0,\, \sigma_i^2)$$

onde a variância do erro é proporcional a uma potência da variável resposta

$$\textbf{\textit{Var}}\{\varepsilon_i | X = x_i\} = \sigma_i^2 = \sigma^2(x_i^m),$$

sendo σ^2 o parâmetro da variância e m a potência adequadara para se estabelecer a proporcionalidade.

- Nesse caso, os pesos adequados para os o método de quadrados mínimos ponderados deve ser o inverso de uma potência da variável reposta:

$$\sigma_i^2 = \sigma^2(x_i^m) \qquad \Rightarrow \qquad w_i = \frac{1}{x_i^m}.$$

- A função da soma de quadrados ponderada se torna:

$$\begin{aligned} Q_w &= \sum_{i=1}^{n} w_i(y_i - \beta_0 - \beta_1 x_i)^2 \\ &= \sum_{i=1}^{n} \frac{1}{x_i^m}(y_i - \beta_0 - \beta_1 x_i)^2 \\ &= \sum_{i=1}^{n} \left(\frac{y_i}{x_i^{m/2}} - \beta_0 \frac{1}{x_i^{m/2}} - \beta_1 \frac{x_i}{x_i^{m/2}} \right)^2 \end{aligned}$$

- Portanto, minimizar Q_w acima é equivalente a aplicar o método de quadrados mínimos ordinários ao modelo (*linear múltiplo*)

$$\begin{aligned} \frac{y_i}{x_i^{m/2}} &= \beta_0 \frac{1}{x_i^{m/2}} + \beta_1 \frac{x_i}{x_i^{m/2}} + \frac{\varepsilon_i}{x_i^{m/2}} \\ y_i' &= \beta_0^* x_{1i}' + \beta_1^* x_{2i}' + \varepsilon_i' \end{aligned}$$

- Se o erro no modelo original (ε_i) era heteroscedástico, o erro no modelo transformado (ε_i') tem variância constante

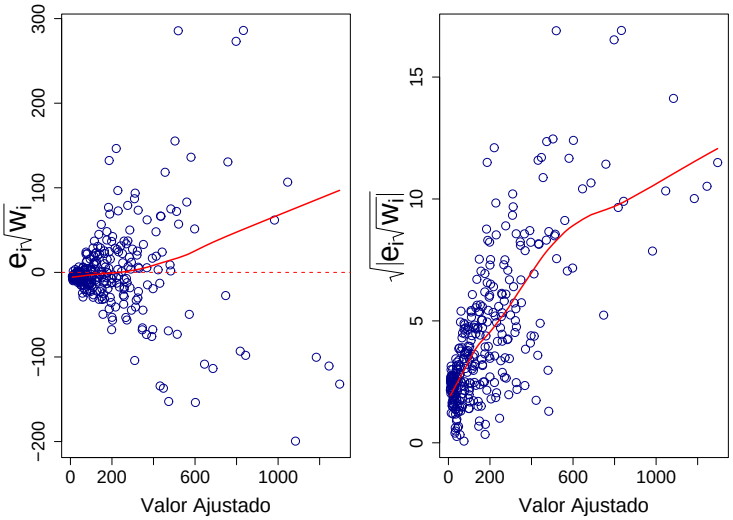
$$\varepsilon_i \sim N(0, \sigma^2 x_i^m) \implies \varepsilon_i' = \frac{\varepsilon_i}{x_i^{m/2}} \sim N(0, \sigma^2).$$

- **Conclusão:**
  - Quadrados mínimos ponderados implica numa transformação da escala da variável resposta.
  - Para se corrigir heteroscedasticidade é freqüentemente necessário testar diversos valores de m ($w_i = x_i^{-m}$), para se encontrar o peso que de fato homogeniza as variâncias.
  - Em geral testa-se valores na amplitude de $m = -4, \ldots, 0, \ldots, 4$, sendo $m = 0$ é o modelo não ponderado.
  - O teste é realizado vizualmente analisando-se
    - o gráfico de dispersão do resíduo ou
    - o gráfico de dispersão da raiz do resíduo absoluto.

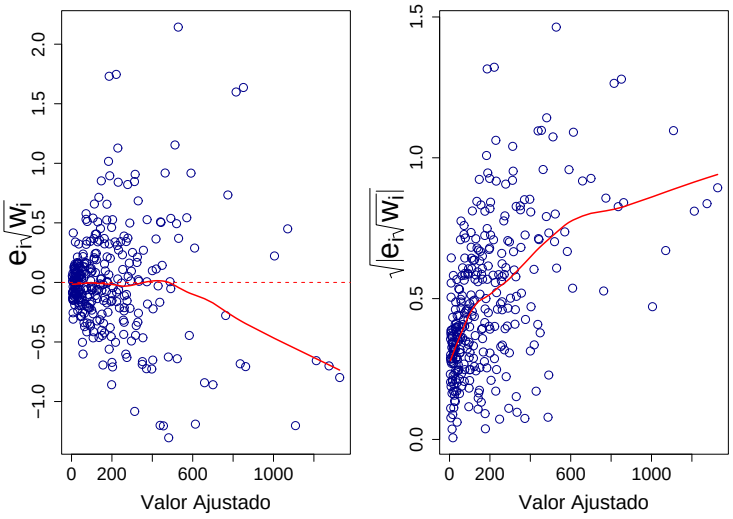
4.3.3 Exemplo de Regressão Ponderada: Volume de Árvores de Caixeta *Tabebuia Cassionoides*

Efeito da ponderação no ajuste do volume comercial em função do volume cilíndrico (modelo de Spurr).

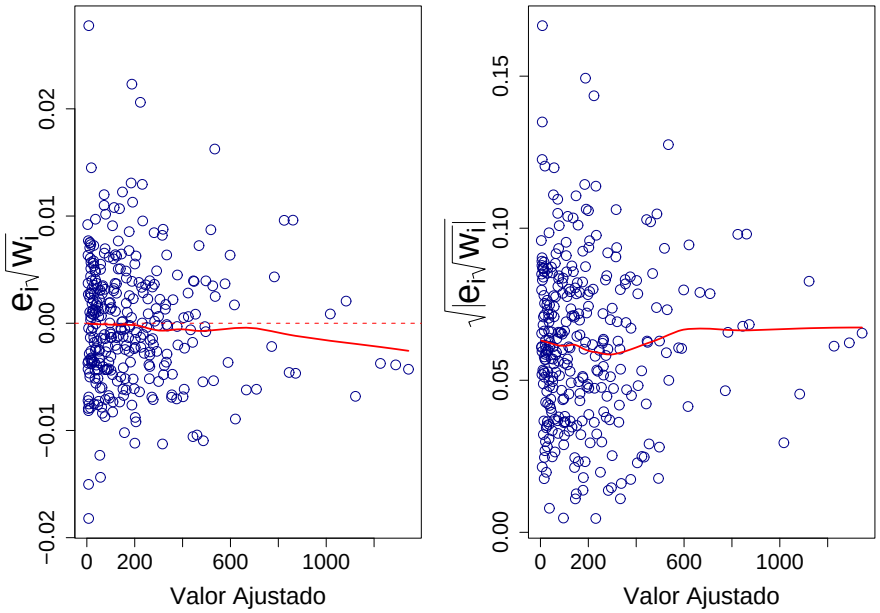
m = 0



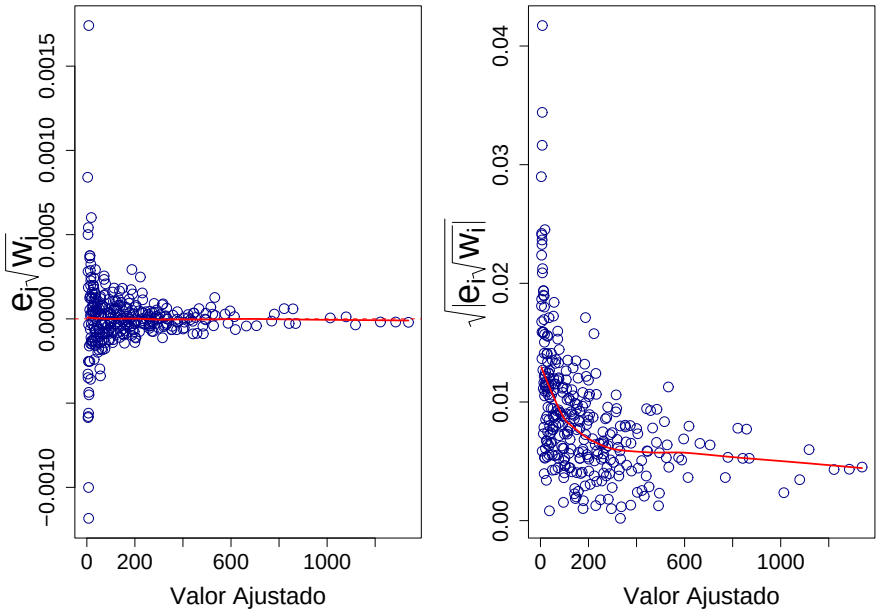
m = 1



m = 2


$e_i\sqrt{w_i}$
−0.02
−0.01
0.00
0.01
0.02
0
200
600
1000
Valor Ajustado
$\sqrt{|e_i\sqrt{w_i}|}$
0.00
0.05
0.10
0.15
0
200
600
1000
Valor Ajustado

m = 3


$e_i\sqrt{w_i}$
−0.0010
0.0000
0.0005
0.0010
0.0015
0
200
600
1000
Valor Ajustado
$\sqrt{|e_i\sqrt{w_i}|}$
0.00
0.01
0.02
0.03
0.04
0
200
600
1000
Valor Ajustado

CAPÍTULO 5
REGRESSÃO LINEAR POR MATRIZES

5.1 O Modelo Linear Simples

5.1.1 Representação do Sistema de Equações

- O modelo de regressão linear simples de erros Gaussianos (MLEG) tem a forma de uma equação indexada

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

onde $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$.

- A indexação da equação representa na verdade um *sistema de equações*:

$$\left.\begin{array}{c} y_1 = \beta_0 + \beta_1 x_1 + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_2 + \varepsilon_2 \\ \cdots \\ y_n = \beta_0 + \beta_1 x_n + \varepsilon_n \end{array}\right\} \Rightarrow \left[\begin{array}{c} y_1 \\ y_2 \\ \vdots \\ y_n \end{array}\right] = \left[\begin{array}{cc} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{array}\right] \left[\begin{array}{c} \beta_0 \\ \beta_1 \end{array}\right] + \left[\begin{array}{c} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{array}\right]$$

- Em notação matricial, o modelo de regressão linear simples de erros Gaussianos (MLEG) pode ser expresso como

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

onde:

$\mathbf{y}$ é o vetor da variável resposta;

$\boldsymbol{\beta}$ é o vetor dos coefcientes de regressão;

$\boldsymbol{\varepsilon}$ é o vetor dos erros aleatórios;

$\mathbf{X}$ é a matrix que contém a variávl preditora, sendo chamada de **matrix de delineamento**.

Note que na matrix de delineamento, a primeira coluna é *um vetor unitário*, assim ela pode ser vista como a junção de dois vetores:

$$\mathbf{X} = [\mathbf{i} \quad \mathbf{x}_1]$$

5.1.2 Estimativa de Quadrados Mínimos

- Os estimadores de quadrados mínimos são encontrados solucionado o sitema de *Equações Normais:*

$$\left.\begin{array}{lll} b_0 n & +b_1 \sum x_i & = \sum y_i \\ b_0 \sum x_i & +b_1 \sum x_i^2 & = \sum x_i y_i \end{array}\right\} \Rightarrow \left[\begin{array}{cc} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{array}\right] \left[\begin{array}{c} b_0 \\ b_1 \end{array}\right] = \left[\begin{array}{c} \sum y_i \\ \sum x_i y_i \end{array}\right]$$

  o qual em notação matricial fica

$$\mathbf{X'Xb} = \mathbf{X'y}$$

- A solução do sistema é obtida pré-multiplicando o sistema por $[\mathbf{X'X}]^{-1}$:

$$\begin{aligned} \mathbf{X'Xb} &= \mathbf{X'y} \\ [\mathbf{X'X}]^{-1}\mathbf{X'Xb} &= [\mathbf{X'X}]^{-1}\mathbf{X'y} \\ \mathbf{Ib} &= [\mathbf{X'X}]^{-1}\mathbf{X'y} \\ \mathbf{b} &= [\mathbf{X'X}]^{-1}\mathbf{X'y} \end{aligned}$$

- Esta solução é a mesma obtida pelos métodos da algebra convencional.

$$\begin{aligned} \mathbf{X'X} &= \left[\begin{array}{cc} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{array}\right] \\ |\mathbf{X'X}| &= n\sum x_i^2 - \left(\sum x_i\right)^2 = n\sum (x_i - \overline{x})^2 \\ [\mathbf{X'X}]^{-1} &= \frac{1}{n\sum (x_i - \overline{x})^2} \left[\begin{array}{cc} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{array}\right] \\ [\mathbf{X'X}]^{-1}\mathbf{X'y} &= \frac{1}{n\sum (x_i - \overline{x})^2} \left[\begin{array}{cc} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{array}\right] \left[\begin{array}{c} \sum y_i \\ \sum x_i y_i \end{array}\right] \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n\sum(x_i-\overline{x})^2}\begin{bmatrix} \sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i \\ -\sum x_i \sum y_i + n\sum x_i y_i \end{bmatrix} \\
&= \frac{1}{n\sum(x_i-\overline{x})^2}\begin{bmatrix} \sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i \\ n\sum x_i y_i - \sum x_i \sum y_i \end{bmatrix} \\
\Rightarrow\ & b_1 = \frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{\sum(x_i-\overline{x})^2} \\
\Rightarrow\ & b_0 = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i + (\sum x_i)^2 \sum y_i - (\sum x_i)^2 \sum y_i}{n\sum(x_i-\overline{x})^2} \\
\Rightarrow\ & b_0 = \frac{(\sum x_i^2 \sum y_i - (\sum x_i)^2 \sum y_i) - (\sum x_i \sum x_i y_i - (\sum x_i)^2 \sum y_i)}{n\sum(x_i-\overline{x})^2} \\
\Rightarrow\ & b_0 = \frac{(\sum x_i^2 - (\sum x_i)^2)}{\sum(x_i-\overline{x})^2}\frac{\sum y_i}{n} - \frac{(\sum x_i y_i - \sum x_i \sum y_i)}{\sum(x_i-\overline{x})^2}\frac{\sum x_i}{n} \\
\Rightarrow\ & b_0 = \overline{y} - b_1\overline{x}
\end{aligned}
$$

5.2 O Modelo Linear Múltiplo

5.2.1 Sistema de Equações

- Os modelos lineares múltiplos compõem sistemas de equações com diversas variáveis preditoras

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i.$$

- Todos os modelos de regressão linear, seja simples ou múltipla, podem ser representados na forma matricial por

$$
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix} + \beta_2 \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{bmatrix} + \cdots + \beta_p \begin{bmatrix} x_{1p} \\ x_{2p} \\ \vdots \\ x_{np} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}
$$

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

- A forma matricial do modelo linear mostra que cada observação da variável resposta é modelada como uma **combinação linear** das variáveis preditoras (colunas da matrix de delineamento) **mais** o erro aleatório.
- Os vetores e matrizes que compõem o modelo linear são:

Vetor da Variável Resposta $\Rightarrow \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$

Vetor dos Coeficientes de Regressão $\Rightarrow \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$

Matrix das Variáveis Preditoras $\Rightarrow \mathbf{X} = [\mathbf{i}\ \mathbf{x}_1\, \mathbf{x}_2\, \cdots\, \mathbf{x}_p]$

(**Matrix de Delineamento**) $\Rightarrow \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & x_{23} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} & \dots & x_{np} \end{bmatrix}$

Vetor dos Erros Aleatórios $\Rightarrow \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$

5.2.2 Valores Ajustado e Resíduos

- De forma análoga ao modelo linear simples, no modelo linear múltiplo, o valor esperado da variável resposta é dado pela combinação linear das variáveis preditoras para cada nível específico delas:

$$\begin{aligned} \boldsymbol{E}\{Y|X_1 = x_{k1}, X_2 = x_{k2}, \dots, X_p = x_{kp}\} &= \\ = \boldsymbol{E}\{y_k\} &= \beta_0 + \beta_1 x_{k1} + \beta_2 x_{k2} + \cdots + \beta_p x_{kp}. \end{aligned}$$

- Em notação matricial, diremos que o vetor de valores esperados é um produto matricial:

$$\boldsymbol{E}\{\mathbf{Y}|\mathbf{X}\} = \mathbf{X}\boldsymbol{\beta}.$$

- Se encontrarmos estimativas para os coeficientes de regressão

$$\mathbf{b}' = [b_0\, b_1\, b_2\, \cdots\, b_p],$$

então o *valor ajustado* é definido como

$$\widehat{y}_i \;=\; b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_p x_{ip}.$$

Que em notação matricial é traduzido pelo *vetor de valores ajustados*

$$\begin{bmatrix} \widehat{y}_1 \\ \widehat{y}_2 \\ \vdots \\ \widehat{y}_n \end{bmatrix} \;=\; b_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + b_1 \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix} + b_2 \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{bmatrix} + \cdots + b_p \begin{bmatrix} x_{1p} \\ x_{2p} \\ \vdots \\ x_{np} \end{bmatrix}$$

$$\widehat{\mathbf{y}} \;=\; \mathbf{X}\mathbf{b}.$$

- O resíduo é definido como diferença entre valor observado e valor ajustado:

$$e_i = y_i - \widehat{y}_i \quad \Rightarrow \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} y_1 - \widehat{y}_1 \\ y_2 - \widehat{y}_2 \\ \vdots \\ y_n - \widehat{y}_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} \widehat{y}_1 \\ \widehat{y}_2 \\ \vdots \\ \widehat{y}_n \end{bmatrix} = \mathbf{y} - \widehat{\mathbf{y}}$$

5.2.3 Estimativas de Quadrados Mínimos

- As estimativas de Quadrados Mínimos são encontradas a partir da função da *Soma de Quadrado do Resíduo*:

$$\begin{aligned} SQR &= \sum e_i^2 = \mathbf{e}'\mathbf{e} = (\mathbf{y} - \widehat{\mathbf{y}})'\mathbf{e} = \mathbf{y}'\mathbf{e} - \widehat{\mathbf{y}}'\mathbf{e} \\ &= \mathbf{y}'(\mathbf{y} - \widehat{\mathbf{y}}) - \widehat{\mathbf{y}}'(\mathbf{y} - \widehat{\mathbf{y}}) \\ &= \mathbf{y}'\mathbf{y} - 2\widehat{\mathbf{y}}'\mathbf{y} + \widehat{\mathbf{y}}'\widehat{\mathbf{y}} \\ &= \mathbf{y}'\mathbf{y} - 2(\mathbf{X}\mathbf{b})'\mathbf{y} + (\mathbf{X}\mathbf{b})'(\mathbf{X}\mathbf{b}) \\ &= \mathbf{y}'\mathbf{y} - 2\mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} \end{aligned}$$

- Para encontrar as estimativas é necessário encontrar a primeira derivada em relação aos parâmetros, igualar a zero e ressolver o sistema:

$$\begin{aligned} \frac{\partial SQR}{\partial \mathbf{b}} &= -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b} = 0 \\ \Rightarrow \quad & \mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y} \end{aligned}$$

- Esse é o próprio *Sistema de Equações Normais* para um modelo linear múltiplo em notação matricial.

- Num modelo linear com p variáveis preditoras o sistema fica:

$$\begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} & \sum x_{i3} & \cdots & \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \sum x_{i1}x_{i3} & \cdots & \sum x_{i1}x_{ip} \\ \sum x_{i2} & \sum x_{i2}x_{i1} & \sum x_{i2}^2 & \sum x_{i2}x_{i3} & \cdots & \sum x_{i2}x_{ip} \\ \sum x_{i3} & \sum x_{i3}x_{i1} & \sum x_{i3}x_{i2} & \sum x_{i3}^2 & \cdots & \sum x_{i3}x_{ip} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{ip} & \sum x_{ip}x_{i1} & \sum x_{ip}x_{i2} & \sum x_{ip}x_{i3} & \cdots & \sum x_{ip}^2 \end{bmatrix} \times \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_p \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \sum x_{i2}y_i \\ \sum x_{i3}y_i \\ \vdots \\ \sum x_{ip}y_i \end{bmatrix}$$

- Note que a matrix $(\mathbf{X}'\mathbf{X})$ é uma matrix **simétrica**.

- A solução do Sistema de Equações Normais é

$$\mathbf{b} = [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{y}.$$

  Ou seja: só haverá solução se:

  - A matrix inversa $[\mathbf{X}'\mathbf{X}]^{-1}$ existir.
  - O que implica que a matrix $(\mathbf{X}'\mathbf{X})$ deve ter posto completo.
  - O que implica que a matrix $\mathbf{X}$ deve ter posto completo.
  - O que implica que as colunas de $\mathbf{X}$ devem ser linearmente independentes.
  - O que implica que *as variáveis preditoras devem ser linearmente independentes*.

- Para que tenhamos certeza que a solução resulta num ponto de mínimo da função da Soma de Quadrados do Resíduo devemos olhar a segunda derivada:

$$\frac{\partial^2 SQR}{\partial \mathbf{b}^2} = 2\mathbf{X}'\mathbf{X}.$$

  A solução é de mínimo se:

- $(\mathbf{X}'\mathbf{X}) > 0$.
- O que implica que a matrix $(\mathbf{X}'\mathbf{X})$ deve ser *definida positiva*.
- O que implica que todos auto-valores de $(\mathbf{X}'\mathbf{X})$ devem ser positivos.

5.2.4 Matrix de Projeção

A apresentação das EQM nos permite redefinir, em termos matriciais, o vetor de valores ajustados e o vetor de resíduos:

Valor Ajustado: vimos que é dado por um produto matricial

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{b} = \mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{y}.$$

- Definiremos como **Matrix de Projeção** ("*hat matrix*") a matrix

$$\mathbf{P} = \mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}',$$

de forma que o vetor de valores ajustados pode ser definida pelo produto matricial

$$\hat{\mathbf{y}} = \mathbf{P}\,\mathbf{y}.$$

- A matrix de projeção é uma matrix
  - de ordem $(n \times n)$

$$\begin{array}{ccc} \mathbf{X} & [\mathbf{X}'\mathbf{X}]^{-1} & \mathbf{X}' \\ (n \times p) & (p \times p) & (p \times n) \end{array}$$

  - simétrica

$$\mathbf{P}' = (\mathbf{X}[\mathbf{X}'\mathbf{X}]^{-\mathbf{1}}\mathbf{X}')' = \mathbf{X}[\mathbf{X}'\mathbf{X}]^{-\mathbf{1}}\mathbf{X}' = \mathbf{P}$$

  - **idempotente**.

$$\begin{aligned} \mathbf{P}\mathbf{P} &= (\mathbf{X}[\mathbf{X}'\mathbf{X}]^{-\mathbf{1}}\mathbf{X}')\,(\mathbf{X}[\mathbf{X}'\mathbf{X}]^{-\mathbf{1}}\mathbf{X}') \\ &= \mathbf{X}[\mathbf{X}'\mathbf{X}]^{-\mathbf{1}}\,(\mathbf{X}'\mathbf{X})\,[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}' \\ &= \mathbf{X}[\mathbf{X}'\mathbf{X}]^{-\mathbf{1}}\mathbf{X}' \\ &= \mathbf{P} \end{aligned}$$

- A matrix de projeção nos será muito útil na análise de regressão e diagnóstico dos resíduos.

Resíduo: o vetor dos resíduos foi definido como o vetor das diferenças entre valor observado e valor ajustado:

$$\begin{aligned}\mathbf{e} &= \mathbf{y} - \widehat{\mathbf{y}} \\ &= \mathbf{y} - \mathbf{P}\mathbf{y} \\ &= (\mathbf{I} - \mathbf{P})\mathbf{y} \\ &= \mathbf{M}\mathbf{y}\end{aligned}$$

- A matrix $\mathbf{M}$ também é uma matrix especial, pois ela é
  - de ordem $(n \times n)$;
  - simétrica

$$\mathbf{M}' = (\mathbf{I} - \mathbf{P})' = \mathbf{I}' - \mathbf{P}' = \mathbf{I} - \mathbf{P} = \mathbf{M};$$

  - **idempotente**

$$\begin{aligned}\mathbf{M}\mathbf{M} &= (\mathbf{I} - \mathbf{P})(\mathbf{I} - \mathbf{P}) \\ &= (\mathbf{I} - \mathbf{P})\mathbf{I} - (\mathbf{I} - \mathbf{P})\mathbf{P} \\ &= (\mathbf{I} - \mathbf{P}) - (\mathbf{I}\mathbf{P} - \mathbf{P}\mathbf{P}) \\ &= (\mathbf{I} - \mathbf{P}) - (\mathbf{P} - \mathbf{P}) \\ &= \mathbf{I} - \mathbf{P} \\ &= \mathbf{M}\end{aligned}$$

5.3 A Regressão como Projeção no Espaço Vetorial

5.3.1 Ortogonalidade das Matrizes $\mathbf{P}$ e $\mathbf{M}$

- As matrizes $\mathbf{P}$ e $\mathbf{M}$ são ortogonais:

$$\mathbf{M}\mathbf{P} = (\mathbf{I} - \mathbf{P})\,\mathbf{P} = \mathbf{I}\mathbf{P} - \mathbf{P}\mathbf{P} = \mathbf{P} - \mathbf{P} = \mathbf{0}.$$

- A ortogonalidade implica que o espaço vetorial gerado pelas colunas de $\mathbf{P}$ é *"perpendicular"* ao espaço vetorial gerado pelas colunas de $\mathbf{M}$.

5.3.2 Lei do Cosseno

- A **Lei do Cosseno** define o ângulo entre dois vetores.

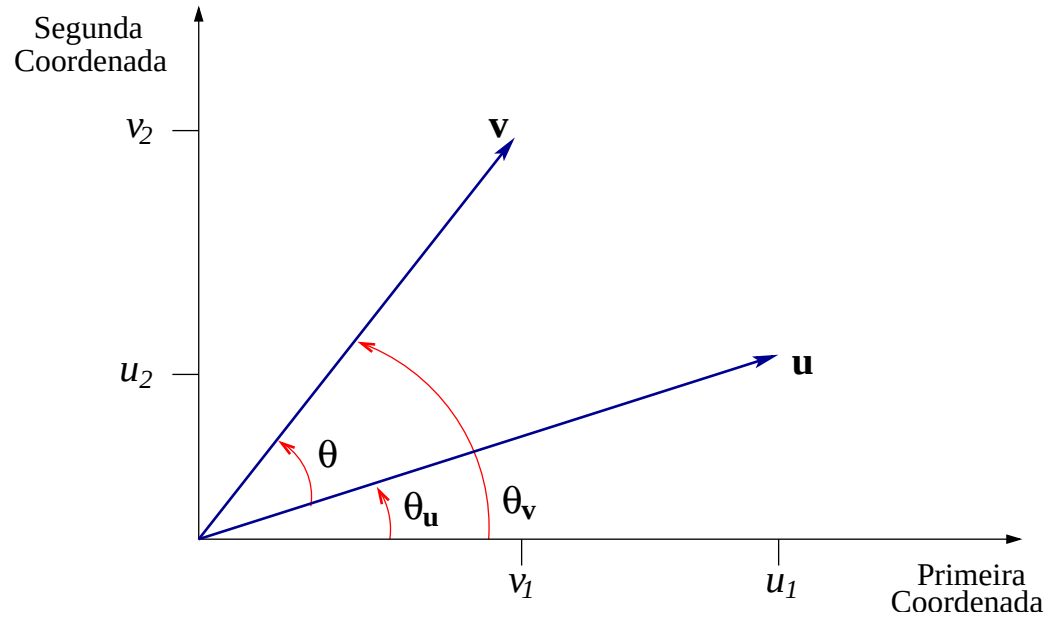


Figura 5.1: Ilustração da lei do cosseno que define o ângulo entre dois vetores: $\mathbf{v}$ e $\mathbf{u}$.

- No espaço bi-dimensional a figura 5.1 ilustra a relação entre as coordenadas de dois vetores e o ângulo formado entre eles.
- No caso do espaço bi-dimensional, a Lei do Cosseno se fundamenta nas seguintes definições trigonométricas:

$$\cos(\theta_{\mathbf{v}}) = \frac{v_1}{L_{\mathbf{v}}} \qquad \cos(\theta_{\mathbf{u}}) = \frac{u_1}{L_{\mathbf{u}}}$$
$$\sin(\theta_{\mathbf{v}}) = \frac{v_2}{L_{\mathbf{v}}} \qquad \sin(\theta_{\mathbf{u}}) = \frac{u_2}{L_{\mathbf{u}}}$$

  onde $L_{\mathbf{u}}$ e $L_{\mathbf{v}}$ são os comprimentos dos vetores $\mathbf{u}$ e $\mathbf{v}$, respectivamente, e na relação

$$\cos(\theta) = \cos(\theta_{\mathbf{v}} - \theta_{\mathbf{u}}) = \cos(\theta_{\mathbf{v}})\cos(\theta_{\mathbf{u}}) + \sin(\theta_{\mathbf{v}})\sin(\theta_{\mathbf{u}}).$$

- Assim, o cosseno do ângulo formado pelos vetores é obtido por

$$\cos(\theta) = \left(\frac{v_1}{L_{\mathbf{v}}}\right)\left(\frac{u_1}{L_{\mathbf{u}}}\right) + \left(\frac{v_2}{L_{\mathbf{v}}}\right)\left(\frac{u_2}{L_{\mathbf{u}}}\right) = \frac{v_1u_1 + v_2u_2}{L_{\mathbf{u}}L_{\mathbf{v}}}$$

- Expressando o cosseno do ângulo em termos vetoriais temos:

$$\cos(\theta) = \frac{\mathbf{v}'\mathbf{u}}{||\mathbf{v}||\,||\mathbf{u}||} = \frac{\mathbf{v}'\mathbf{u}}{\sqrt{\mathbf{v}'\mathbf{v}}\,\sqrt{\mathbf{u}'\mathbf{u}}}.$$

  que é a Lei do Cosseno, sendo válida para vetores com k-dimensões.

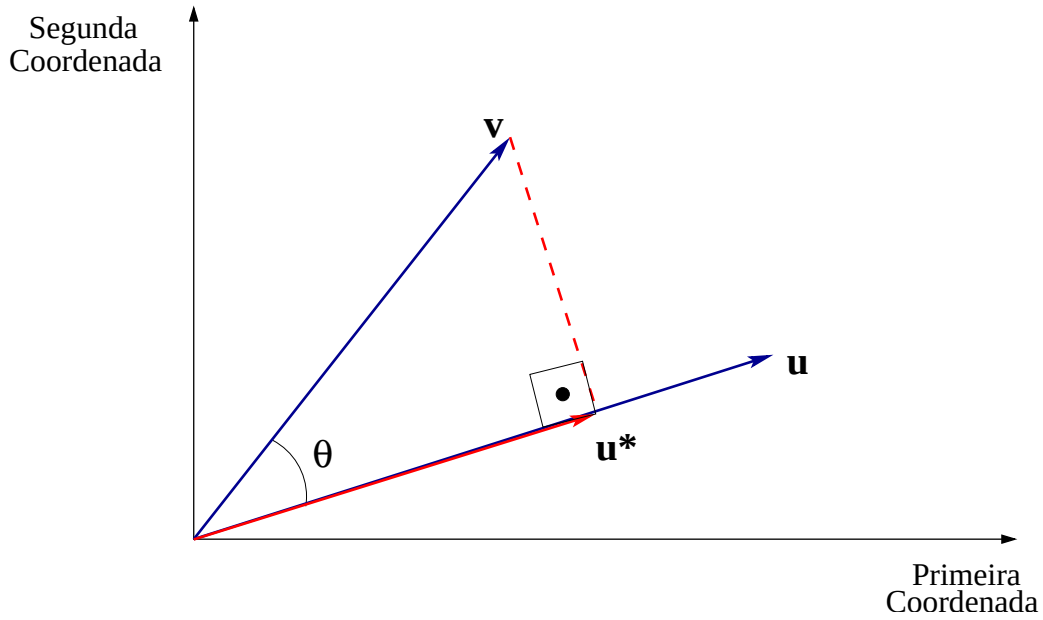


Figura 5.2: Ilustração da projeção do vetor $\mathbf{v}$ sobre o vetor $\mathbf{u}$.

5.3.3 Projeção de Vetores

- A figura 5.2 ilustra a projeção do vetor $\mathbf{v}$ sobre o vetor $\mathbf{u}$ num espaço vetorial bi-dimensional.
- Pela Lei do Cosseno, o comprimento da projeção $\mathbf{u}^*$ é

$$||\mathbf{u}^*|| = L_{\mathbf{v}} \cos(\theta) = L_{\mathbf{v}} \left(\frac{v_1 u_1 + v_2 u_2}{L_{\mathbf{u}} L_{\mathbf{v}}} \right) = \frac{\mathbf{v}'\mathbf{u}}{\sqrt{\mathbf{u}'\mathbf{u}}}$$

- Como a projeção $\mathbf{u}^*$ e o vetor $\mathbf{u}$ estão alinhados, a projeção é um múltiplo escalar desse:

$$\mathbf{u}^* = k\,\mathbf{u} = \left(\frac{||\mathbf{u}^*||}{||\mathbf{u}||} \right) \mathbf{u} = \left(\frac{\mathbf{v}'\mathbf{u}/\sqrt{\mathbf{u}'\mathbf{u}}}{\sqrt{\mathbf{u}'\mathbf{u}}} \right) \mathbf{u} = \left(\frac{\mathbf{v}'\mathbf{u}}{\mathbf{u}'\mathbf{u}} \right) \mathbf{u}$$

5.3.4 Valor Ajustado e Resíduo como Projeções Vetoriais

- O modelo linear também pode ser entendido como uma projeção vetorial.
- A figura 5.3 ilustra essa projeção vetorial num espaço tridimensional (apenas três observações), com duas variáveis preditoras: $\mathbf{x}_0$ (vetor unitário) e $\mathbf{x}_1$.

- O vetor dos valores ajustados

$$\widehat{\mathbf{y}} = \mathbf{P}\,\mathbf{y} = \mathbf{X}[\mathbf{X}'\mathbf{X}^{-1}]\mathbf{X}'\,\mathbf{y}$$

é interpretado como a *projeção vetorial* do vetor da variável resposta sobre o *espaço vetorial gerado pelas colunas da matrix de delineamento* ($\mathbf{X}$).

- O vetor dos resíduos

$$\mathbf{e} = \mathbf{M}\,\mathbf{y} = [\mathbf{I} - \mathbf{P}]\,\mathbf{y} = [\mathbf{I} - \mathbf{X}[\mathbf{X}'\mathbf{X}^{-1}]\mathbf{X}']\,\mathbf{y}$$

é a projeção do vetor da variável resposta sobre o *espaço vetorial perpendicular* ao epaço vetorial gerado pelas colunas da matrix $\mathbf{X}$.

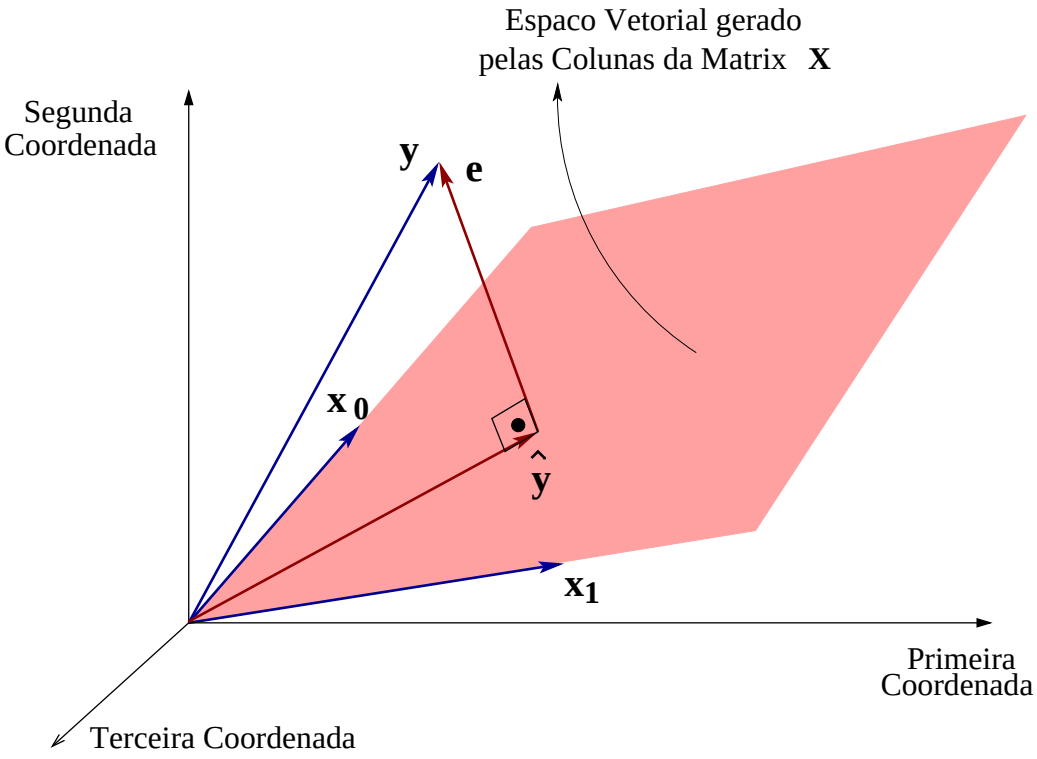


Figura 5.3: Ilustração da projeção vetorial do vetor de observações ($\mathbf{y}$) sobre o espaço gerado pelas colunas da matrix de delineamento ($\mathbf{X}$).

5.3.5 Propriedades Vetoriais do Modelo Linear

Decomposição da Soma de Quadrados: a abordagem vetorial ilustra bem a decomposição da *Soma de Quadrados Total*, i.e., a variabilidade da variável resposta:

$$\begin{aligned}
\mathbf{e} = \mathbf{y} - \widehat{\mathbf{y}} \quad \Rightarrow \quad \mathbf{y} &= \widehat{\mathbf{y}} + \mathbf{e} \\
||\mathbf{y}||^2 &= ||\widehat{\mathbf{y}}||^2 + ||\mathbf{e}||^2 \\
\mathbf{y}'\mathbf{y} &= \widehat{\mathbf{y}}'\widehat{\mathbf{y}} + \mathbf{e}'\mathbf{e} \\
\mathbf{y}'\mathbf{y} - n(\overline{y})^2 &= \widehat{\mathbf{y}}'\widehat{\mathbf{y}} - n(\overline{y})^2 + \mathbf{e}'\mathbf{e} \\
\mathbf{y}'\mathbf{y} - n(\frac{1}{n}\mathbf{y}'\mathbf{i})^2 &= \widehat{\mathbf{y}}'\widehat{\mathbf{y}} - n(\frac{1}{n}\widehat{\mathbf{y}}'\mathbf{i})^2 + \mathbf{e}'\mathbf{e} \\
\mathbf{y}'\mathbf{I}\mathbf{y} - \frac{1}{n}\mathbf{y}'\mathbf{i}\mathbf{i}'\mathbf{y} &= \widehat{\mathbf{y}}'\mathbf{I}\widehat{\mathbf{y}} - \frac{1}{n}\widehat{\mathbf{y}}'\mathbf{i}\mathbf{i}'\widehat{\mathbf{y}} + \mathbf{e}'\mathbf{e} \\
\mathbf{y}'[\mathbf{I} - \frac{1}{n}\mathbf{i}\mathbf{i}']\mathbf{y} &= \widehat{\mathbf{y}}'[\mathbf{I} - \frac{1}{n}\mathbf{i}\mathbf{i}']\widehat{\mathbf{y}} + \mathbf{e}'\mathbf{e} \\
\mathbf{y}'\mathbf{M}^0\mathbf{y} &= \widehat{\mathbf{y}}'\mathbf{M}^0\widehat{\mathbf{y}} + \mathbf{e}'\mathbf{e} \\
\mathbf{y}'\mathbf{M}^0\mathbf{y} &= \mathbf{b}'\mathbf{X}'\mathbf{M}^0\mathbf{X}\mathbf{b} + \mathbf{e}'\mathbf{e}
\end{aligned}$$

Nessa expressão temos:

- $\mathbf{y}'\mathbf{M}^0\mathbf{y} = \sum(y_i - \overline{y})^2$ é a Soma de Quadrados Total (SQT);
- $\mathbf{e}'\mathbf{e} = \sum e_i^2$ é a Soma de Quadrado do Resíduo (SQR), i.e., a variabilidade **não explicada** pelo modelo;
- $\mathbf{b}'\mathbf{X}'\mathbf{M}^0\mathbf{X}\mathbf{b}$ é a Soma de Quadrados do Modelo (SQM), i.e., a variabilidade **explicada** pelo modelo;

$$SQT = SQM + SQR$$

Vetor do Resíduo é perpendicular ao espaço gerado pelas colunas da matrix $\mathbf{X}$:

$$\begin{aligned}
\mathbf{X}'\mathbf{e} &= \mathbf{X}'(\mathbf{y} - \widehat{\mathbf{y}}) \\
&= \mathbf{X}'\mathbf{y} - \mathbf{X}'\widehat{\mathbf{y}} \\
&= \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\mathbf{b} \\
&= \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X}^{-1}\mathbf{X}'\mathbf{y}) \\
&= \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{y} \\
&= \mathbf{0}
\end{aligned}$$

Vetor do Resíduo é perpendicular ao vetor do valor ajutado:

$$\begin{aligned}
\mathbf{e}'\widehat{\mathbf{y}} &= (\mathbf{y}-\widehat{\mathbf{y}})'\,\widehat{\mathbf{y}} \\
&= \mathbf{y}'\widehat{\mathbf{y}} - \widehat{\mathbf{y}}'\widehat{\mathbf{y}} \\
&= \mathbf{y}'\mathbf{P}\mathbf{y} - (\mathbf{P}\mathbf{y})'\mathbf{P}\mathbf{y} \\
&= \mathbf{y}'\mathbf{P}\mathbf{y} - \mathbf{y}\mathbf{P}'\mathbf{P}\mathbf{y} \\
&= \mathbf{y}'\mathbf{P}\mathbf{y} - \mathbf{y}\mathbf{P}\mathbf{y} \\
&= \mathbf{0}
\end{aligned}$$

5.4 O Modelo Linear Múltiplo Gaussiano

- O modelo linear múltiplo foi visto até do ponto de vista matricial apenas, mas acrescentando-se o erro aleatório com distribuição Gaussiano veremos que o modelo linear múltiplo tem as mesmas propriedades estatísticas já vistas para o modelo linear simples.

5.4.1 Erros Esféricos

- O componente do erro no Modelo Linear Múltiplo Gaussiano é freqüentemente chamado de *erros esféricos* pois as propriedades de independência e homoscedasticidade é apresentada de forma matricial.
- Já foi visto que:
  - Os erros de valor esperado nulo

    $$\boldsymbol{E}\{\varepsilon_i|X=x_i\}=0 \quad \text{para todo } i=1,2,\ldots,n.$$

    No caso do modelo múltiplo generalizaremos a condição para os níveis de todas variáveis preditoras

    $$\boldsymbol{E}\{\varepsilon_i|\mathbf{X}\}=0 \quad \text{para todo } i=1,2,\ldots,n.$$

  - Os erros são homoscedáticos e idependentes

    $$\begin{aligned}
    \boldsymbol{Var}\{\varepsilon_i|X=x_i\} &= \sigma^2 \quad \text{para todo } i=1,2,\ldots,n; \\
    \boldsymbol{Cov}\{\varepsilon_i,\varepsilon_j|X=x_i\} &= 0 \quad \text{para todo } i\neq j.
    \end{aligned}$$

- Assim, a matrix de covariância do erro terá forma:

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

- Como cada termo do erro tem distribuição Gaussiana

$$(\varepsilon_i|\mathbf{X}) \sim N(0,\sigma^2) \quad \text{para todo } i = 1, 2, \ldots, n,$$

  o termo do erro será composta composto por uma distribuição Gaussiana multivariada

$$(\boldsymbol{\varepsilon}|\mathbf{X}) \sim N(\mathbf{0}, \sigma^2\,\mathbf{I}).$$

CAPÍTULO 6

INTERPRETAÇÃO E INFERÊNCIA NO MODELO LINEAR MÚLTIPLO

6.1 Modelo Linear Mútliplo de Erros Esféricos

O modelo linear múltiplo modela a variável resposta em função de diversas variáveis preditoras.

Vejamos em detalhes a sua estrutura:

Modelo na Forma Indexada:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \qquad \text{para } i = 1, 2, \ldots, n.$$

Modelo na Forma Matricial:

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

onde

- $\mathbf{y}$ é o vetor da variável resposta;
- $\mathbf{X}$ é a matrix das variáveis preditoras, matrix de delineamento;
- $\boldsymbol{\beta}$ é o vetor dos coeficientes de regressão a serem estimados; e
- $\boldsymbol{\varepsilon}$ é o vetor do erro aleatório.

Pressuposição 1: a relação funcional linear é *"verdadeira"*, i.e., a relação

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

representa o fenômeno sendo modelado de modo realista.

Pressuposição 2: a matrix de delineamento

$$\mathbf{X} = [\mathbf{i}\ \mathbf{x}_1\, \mathbf{x}_2\, \cdots \mathbf{x}_p]$$

é uma matrix de ordem $n \times (p+1)$ sendo que:

a) $n > (p+1)$, i.e., o número de observações é maior que o número de variáveis preditoras. De preferência bem maior!

b) posto $(X) = (p+1)$, i.e., a matrix de delineamento é uma matrix de posto completo.

Essa pressuposição é necessária para que o Sistema de Equações Normais tenha solução única.

Pressuposição 3: dada a matrix $\mathbf{X}$, o erro tem valor esperado nulo:

$$\boldsymbol{E}\{\varepsilon_i|\mathbf{X}\} = 0 \quad \Rightarrow \quad \boldsymbol{E}\{\boldsymbol{\varepsilon}|\mathbf{X}\} = \mathbf{0}.$$

Isso implica que

- cada observação de ε_i é independente de todas as variáveis preditoras:

$$\boldsymbol{Cov}\{\varepsilon_i, \mathbf{X}\} = 0;$$

- mesmo para matrix de delineamento ($\mathbf{X}$) estocástica, o valor esperado do erro é nulo:

$$\boldsymbol{E}\{\boldsymbol{\varepsilon}\} = \boldsymbol{E}_{\mathbf{X}}\{\boldsymbol{E}\{\boldsymbol{\varepsilon}|\mathbf{X}\}\} = \boldsymbol{E}_{\mathbf{X}}\{\mathbf{0}\} = \mathbf{0};$$

- o valor esperado da variável resposta é

$$\boldsymbol{E}\{\mathbf{y}|\mathbf{X}\} = \mathbf{X}\boldsymbol{\beta}.$$

Pressuposição 4: os erros são **esféricos**, i.e., homoscedáticos e independentes (não-correlacionados):

$$\boldsymbol{Var}\{\varepsilon_i|\mathbf{X}\} = \sigma^2, \qquad i = 1, 2, \ldots, n$$

$$\boldsymbol{Cov}\{\varepsilon_i, \varepsilon_j\} = 0, \qquad i \neq j.$$

- Na notação matricial temos

$$\boldsymbol{Var}\{\boldsymbol{\varepsilon}|\mathbf{X}\} = \boldsymbol{E}\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}\} = \sigma^2\mathbf{I}$$

- Erros esféricos implica a independência **entre erros**, i.e., entre observações, o que equivale a afirmar a ausência de qualquer forma de autocorrelação (temporal, espacial, etc.).

Pressuposição 5: as variáveis preditoras **não são estocástica**.

- O que foi assumido para o Modelo Linear Simples é generalizado para todas variáveis preditoras no Modelo Linear Múltiplo.
- Essa pressuposição é essencialmente uma conveniência para simplificar as demonstrações, uma vez que para vários resultados elas pode ser relaxada.
- Somente em situações *experimentais* é que é possível garantir totalmente que as variáveis preditoras serão de fato não-estocásticas.

Pressuposição 6: os erros têm distribuição Gaussiana Multivariada, com *vetor de médias* nulo e matrix de variância-covariância igual a $\sigma^2\mathbf{I}$:

$$\boldsymbol{\varepsilon}|\mathbf{X} \sim N\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$$

- Essa pressuposição é necessária para garantir que as inferências realizadas (teste de hipóteses e intervalos de confiança) sejam **exatas**.

Resumindo as pressuposições em notação matemática sintética temos:

1. $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$;
2. $\mathbf{X}$ é uma matrix $n \times p$ de posto p;
3. $\boldsymbol{E}\{\boldsymbol{\varepsilon}|\mathbf{X}\} = \mathbf{0}$;
4. $\boldsymbol{E}\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}\} = \sigma^2\mathbf{I}$;
5. $\mathbf{X}$ é uma matrix não-estocástica;
6. $\boldsymbol{\varepsilon}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.

6.2 Interpretação do Modelo Linear Múltiplo (MLM)

O MLM por envolver diversas variáveis preditoras, permite diversas formas de aplicação relativas à relação funcional.

6.2.1 Modelo com p Variáveis Preditoras Distintas

- O MLM **Geral** assume que cada variável preditora é distinta das demais.
- Tomando $x_{i0} = 1$, o MLM pode ser expresso por

$$y_i = \beta_0 x_{i0} + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

- Nesse modelo, a relação funcional é

$$\boldsymbol{E}\{Y|X_0 = 1, X_1 = x_{k1}, \ldots, X_p = x_{kp}\} = \beta_0 x_{k0} + \beta_1 x_{k1} + \cdots + \beta_p x_{kp}$$

- O número de variáveis preditoras define casos particulares.
- Se $p = 1$, temos

$$\boldsymbol{E}\{Y|X_0 = 1, X_1 = x_{k1}\} = \beta_0 + \beta_1 x_{k1}$$

  que é o modelo linear simples, onde a interpretação geométrica da relacional funcional é de uma reta no espaço bidimensional (Y, X_1).

- Se $p = 2$, temos

$$\boldsymbol{E}\{Y|X_0 = 1, X_1 = x_{k1}, X_2 = x_{k2}\} = \beta_0 + \beta_1 x_{k1} + \beta_2 x_{k2}$$

  – Nesse caso, a relação funcional representa um **plano** no espaço tridimensional (Y, X_1, X_2).
  – Esse plano é chamado de **superfície resposta** ou **superfície de regressão**.
  – A interpetação dos coeficientes de regressão fica:

  β_0**:** intercepto, i.e., ponto no eixo-y que intercepta a superfície resposta (plano).

$$\boldsymbol{E}\{Y|X_0 = 1, X_1 = 0, X_2 = 0\} = \beta_0.$$

  β_1**:** alteração na variável resposta ocasionada pela **alteração de uma unidade** na variável X_1, quando X_2 permanece constante:

$$\begin{aligned}\boldsymbol{E}\{Y|X_0 = 1, X_1 = k, X_2 = x_{k2}\} - \\ -\boldsymbol{E}\{Y|X_0 = 1, X_1 = (k+1), X_2 = x_{k2}\} = \\ = (\beta_0 + \beta_1(k) + \beta_2 x_{k2}) - (\beta_0 + \beta_1(k+1) + \beta_2 x_{k2}) \\ = \beta_1\end{aligned}$$

  É a "**inclinação**" das projeções das retas da superfície resposta sobre o plano (Y, X_1).

β_2**:** alteração na variável resposta ocasionada pela **alteração de uma unidade** na variável X_2, quando X_1 permanece constante.
É a "**inclinação**" das projeções das retas da superfície resposta sobre o plano (Y, X_2).

– Essa interpretação é limitada e deve ser utilizada com cautela, pois na prática geralmente temos

$$\boldsymbol{Cov}\{X_1, X_2\} \neq 0,$$

logo, não é possível alterar X_1 e manter X_2 constante.

- Se tivermos p variáveis preditoras, temos a superfície resposta formada pelo **hiper-plano** defindo por

$$\boldsymbol{E}\{Y|X_0 = 1, X_1 = x_{k1}, \ldots, X_p = x_{kp}\} = \beta_0 + \beta_1 x_{k1} + \cdots + \beta_p x_{kp}$$

no espaço $(p+1)$-dimensional $(Y, X_1, X_2, \ldots, X_p)$.

– A interpretação dos coeficientes é análoga e também dever ser realizada com cautela.

– β_0 é o intercepto.

– β_j é alteração na variável resposta ocasionada pela **alteração de uma unidade** na variável X_j, quando todas as demais variáveis preditoras X_k ($k \neq j$) permanecem constantes.

6.2.2 Modelos Polinomias

- Modelo polinomiais são baseados **numa única** variável preditora, mas expressa a relação funcional na forma de um polinômio:

$$\begin{aligned} p &= \text{é a ordem do polinômio;} \\ X_k &= X^k \\ \boldsymbol{E}\{Y|X = x_i\} &= \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \cdots + \beta_p x_i^p \end{aligned}$$

- A interpretação dos coeficientes de regressão se baseia na interpretação das constantes de um polinômio.

- Exemplo: $p = 2$ implica numa parábola:

$$\begin{aligned} \boldsymbol{E}\{Y|X = x_i\} &= \beta_0 + \beta_1 x_i + \beta_2 x_i^2 \\ \beta_0 &\Rightarrow \text{intercepto;} \end{aligned}$$

$$\begin{aligned} \beta_2 &\Rightarrow \text{ concavidade da parábola :} \\ &\Rightarrow \beta_2 > 0 \text{ parab. convexa,} \\ &\Rightarrow \beta_2 < 0 \text{ parab. côncova;} \\ x = \frac{\beta_1}{2\beta_2} &\Rightarrow \text{ vértice da parábola;} \\ y = \frac{3\beta_1^2 + 4\beta_2\beta_0}{4\beta_2} &\Rightarrow \text{ máximo ou mínimo da parábola.} \end{aligned}$$

6.2.3 Modelos de Interação

- A variável resposta é modeada em função de algumas variáveis preditoras e da **interação** entre elas.
- Exemplo de duas variáveis preditoras:

$$\boldsymbol{E}\{Y|X_1 = x_{i1}, X_2 = x_{i2}\} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 (x_{i1} x_{i2}).$$

- Para se realizar o ajuste do modelo, a interação entra no modelo como uma terceira variável preditora $X_3 = X_1 X_2$:

$$\boldsymbol{E}\{Y|X_1 = x_{i1}, X_2 = x_{i2}, X_3 = x_{i3} = x_{i1}x_{i2}\} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}.$$

- Em termos da variáveis preditoras originais temos um espaço tridimensional (Y, X_1, X_2), mas a superfície resposta não é mais um plano devido à interação.
- Para deixar mais claro o **conceito de interação** vejamos alguns casos dos modelos de interação com duas variáveis:

  Modelo de Efeitos Aditivos Sem Interação:

$$\boldsymbol{E}\{Y|X_1 = x_{i1}, X_2 = x_{i2}\} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}.$$

  A superfície resposta é um plano, cuja projeção no plano (Y, X_1) resulta em **retas paralelas**:

$$\boldsymbol{E}\{Y|X_1 = x_{i1}, X_2 = x_{i2}\} = (\beta_0 + \beta_2 x_{i2}) + \beta_1 x_{i1},$$

  Para $X_2 = k$

$$\begin{aligned} \boldsymbol{E}\{Y|X_1 = x_{i1}, X_2 = k\} &= (\beta_0 + \beta_2 k) + \beta_1 x_{i1} \\ &= \beta_0' + \beta_1 x_{i1}. \end{aligned}$$

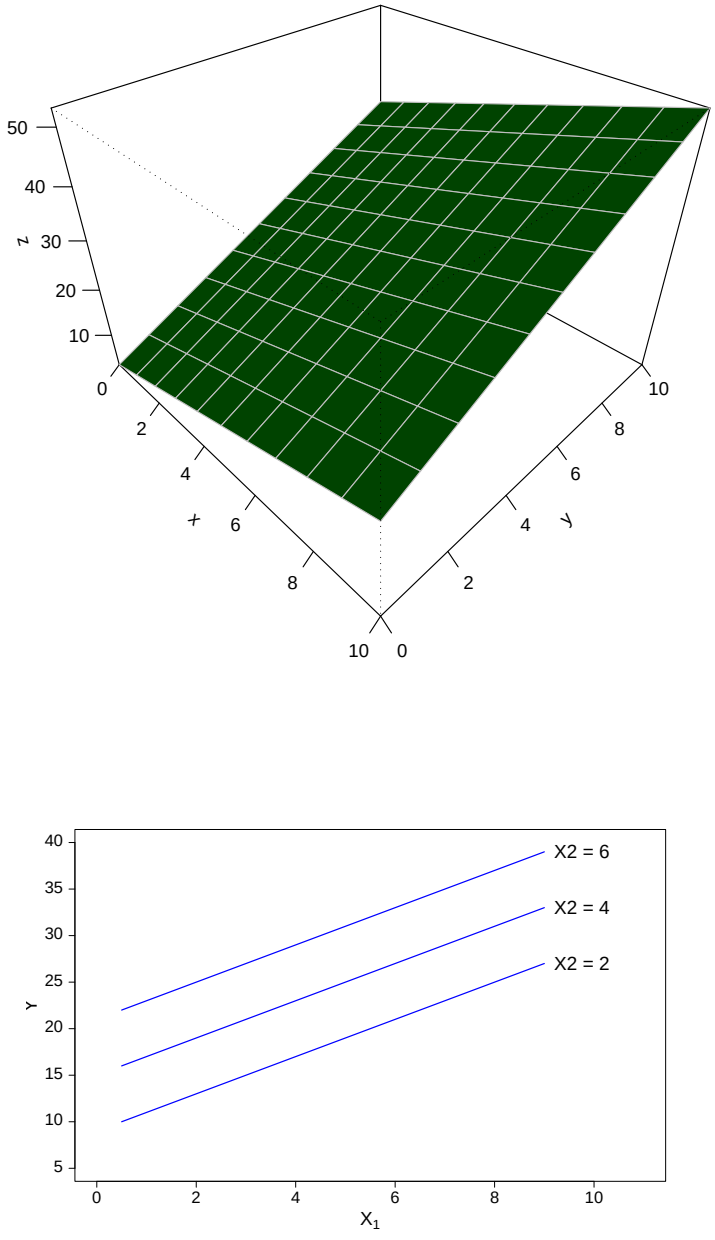


Figura 6.1: Interpretação geométrica do modelo de efeitos aditivos **sem interação**.

Modelo de Efeitos Aditivos Com Interação:

$$\boldsymbol{E}\{Y|X_1 = x_{i1}, X_2 = x_{i2}\} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 (x_{i1} x_{i2}).$$

A superfície resposta **não** é um plano, pois a superfície se "**curva**" devido à interação.

As projeções de retas da superfície resposta no plano (Y, X_1) resulta em **retas não-paralelas**:

$$\boldsymbol{E}\{Y|X_1 = x_{i1}, X_2 = x_{i2}\} = (\beta_0 + \beta_2 x_{i2}) + (\beta_1 + \beta_3 x_{i2})\, x_{i1},$$

Para $X_2 = k$

$$\begin{aligned} \boldsymbol{E}\{Y|X_1 = x_{i1}, X_2 = k\} &= (\beta_0 + \beta_2 k) + (\beta_1 + \beta_3 k)\, x_{i1} \\ &= \beta_0' + \beta_1' x_{i1}. \end{aligned}$$

A inclinação de cada reta projetada depende dos valores de X_2.

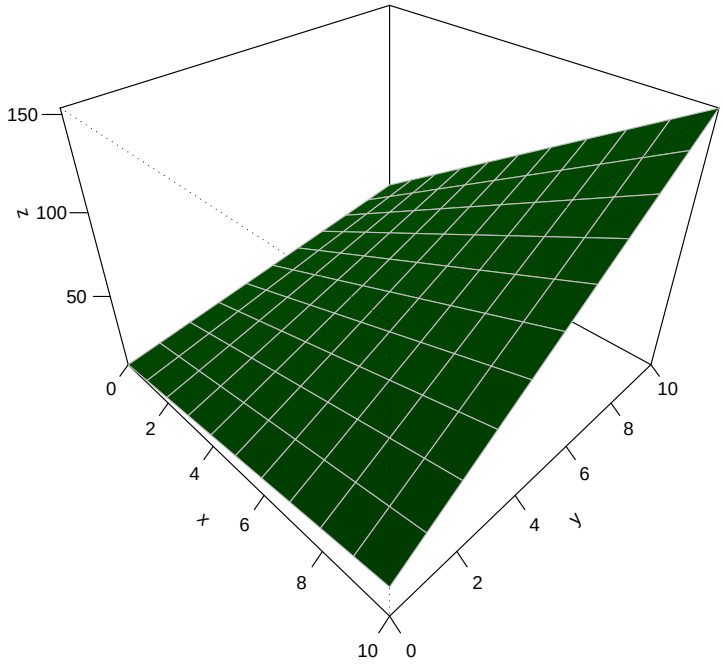
- Os modelos de interação com k variáveis preditoras implicam em diferentes ordens de interação o que torna a sua interpretação muito complexa.

  Exemplo: efeitos aditivos de 3 variáveis preditoras e todas as interações:

$$\begin{aligned} \boldsymbol{E}\{Y\} &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + && \text{(efeitos principais)} \\ &\quad + \beta_4 (x_{i1} x_{i2}) + \beta_5 (x_{i1} x_{i3}) + \beta_6 (x_{i2} x_{i3}) + && \text{(interações duplas)} \\ &\quad + \beta_7 (x_{i1} x_{i2} x_{i3}) && \text{(interação tripla)} \end{aligned}$$

- Os modelos de interação podem incluir efeitos polinomias, aumentando ainda mais a sua complexidade:

$$\begin{aligned} \boldsymbol{E}\{Y\} &= \beta_0 + && \text{(intercepto)} \\ &\quad + \beta_1 x_{i1} + \beta_2 x_{i1}^2 + && \text{(efeito principal 1)} \\ &\quad + \beta_3 x_{i2} + \beta_4 x_{i2}^2 + && \text{(efeito principal 2)} \\ &\quad + \beta_5 (x_{i1} x_{i2}) && \text{(interação de primeira ordem)} \\ &\quad + \beta_6 (x_{i1}^2 x_{i2}) + \beta_7 (x_{i1} x_{i2}^2) && \text{(interações de primeira por segunda ordem)} \\ &\quad + \beta_8 (x_{i1}^2 x_{i2}^2) && \text{(interações de segunda ordem)} \end{aligned}$$



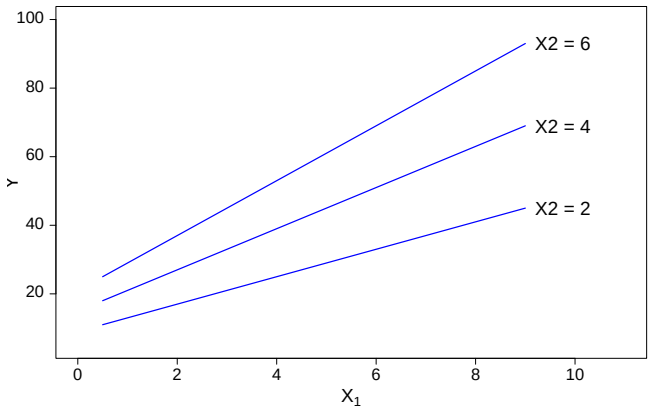


Figura 6.2: Interpretação geométrica do modelo de efeitos aditivos **com interação**.

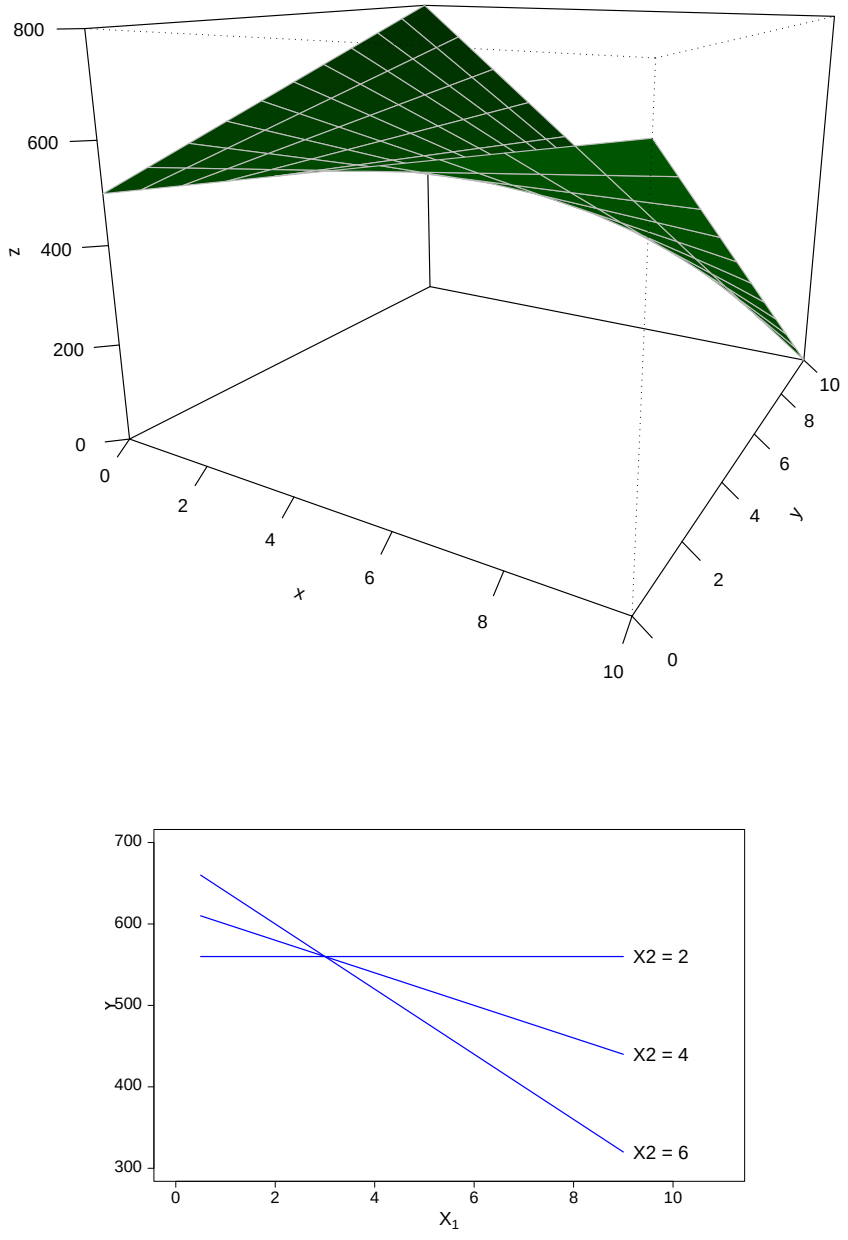


Figura 6.3: Interpretação geométrica do modelo de efeitos aditivos **com interação**.

- O aumento da complexidade dos modelos polinomiais e polinomiais de interação freqüentemente resulta em problemas sérios de ajuste, devido à existência de **correlação** entre as variáveis preditoras (*multicolinearidade*).
- A única forma de contornar completamente esse problema é através de delineamentos experimentais que gerem uma matrix de delineamento *ortogonal*, i.e., as variáveis preditoras não são correlacionadas.

6.2.4 Modelos de Variáveis Transformadas

Freqüentemente o MLM é uma simples aproximação para relações não-lineares.

O Modelo *Log-Linear* é um exemplo de relação multiplicativa que, após a transformação logarítmica, se torna linear:

$$\begin{aligned} y &= \mathrm{e}^{\beta_0}\, x_1^{\beta_1}\, x_2^{\beta_2} \cdots x_p^{\beta_p}\, \mathrm{e}^{\varepsilon} = \mathrm{e}^{\beta_0} \prod_{k=1}^{p} x_k^{\beta_k} \mathrm{e}^{\varepsilon} \\ \ln(y) &= \beta_0 + \beta_1 \ln(x_1) + \beta_2 \ln(x_2) + \cdots + \beta_p \ln(x_p) + \varepsilon \end{aligned}$$

O Modelo Exponencial Multiplicativo é um modelo que também pode ser aproximada pelo MLM através de uma transformação logarítmica:

$$\begin{aligned} y &= \mathrm{e}^{\beta_0}\, \mathrm{e}^{\beta_1 x_1}\, \mathrm{e}^{\beta_1 x_1}\, \mathrm{e}^{\varepsilon} = \mathrm{e}^{\beta_0} \prod_{k=1}^{p} x_k^{\beta_k} \mathrm{e}^{\varepsilon} \\ \ln(y) &= \beta_0 + \beta_1 \ln(x_1) + \beta_2 \ln(x_2) + \cdots + \beta_p \ln(x_p) + \varepsilon \end{aligned}$$

O Modelo Inverso é uma relação não-linear que pode ser aproximado pela transformação da variável resposta

$$\begin{aligned} y &= \frac{1}{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon} \\ \frac{1}{y} &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \end{aligned}$$

6.3 Inferência no Modelo Linear Mútliplo Gaussiano

6.3.1 Estimativas de Máxima Verossimilhança

- Como assumimos que $\boldsymbol{\varepsilon}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, tempos que

$$\begin{aligned} \boldsymbol{E}\{\mathbf{y}\} &= \boldsymbol{E}\{\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\} = \mathbf{X}\boldsymbol{\beta} \\ \text{e} \\ \boldsymbol{Cov}\{\mathbf{y}\} &= \boldsymbol{Cov}\{\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\} = \boldsymbol{Cov}\{\boldsymbol{\varepsilon}\} = \sigma^2\mathbf{I}. \end{aligned}$$

  Logo, temos que

$$\mathbf{y}|\mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}),$$

  ou seja, a variável resposta tem distribuição Gaussiana Multivariada.

- A função de verossimilhança da Gaussiana Multivariada nesse caso é a própria função de densidade da distribuição Guassiana Multivariada, pois a nossa amostra consiste de um único vetor de observações:

$$\mathcal{L}\{\boldsymbol{\beta}, \sigma^2\} = \left(2\pi\sigma^2\right)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right].$$

  Note que

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_p x_{ip})^2.$$

- Para encontrar as EMV dos coeficientes de regressão é melhor utilizar a função de log-verossimilhança que é

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

  cuja derivada parcial em relação aos coeficientes é igual a

$$\begin{aligned} \frac{\partial \mathbf{L}\{\boldsymbol{\beta}, \sigma^2\}}{\partial \boldsymbol{\beta}} = -\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) &= \mathbf{0} \\ \Rightarrow \quad \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} &= \mathbf{0} \\ \mathbf{X}\widehat{\boldsymbol{\beta}} &= \mathbf{y} \end{aligned}$$

$$\text{(Sistema de Equações Normais)} \qquad \mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

$$\text{(Estimativas de Máxima Verossimilhança)} \qquad \widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

  Note que os EMV são idênticos aos EQM!

- Para estimar a variância do erro temos

$$\begin{aligned}\frac{\partial \mathbf{L}\{\boldsymbol{\beta}, \sigma^2\}}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = 0 \\ &\Rightarrow \quad \frac{1}{2\sigma^2}\left[-n + \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})\right] = 0 \\ \widehat{\sigma}^2 &= \frac{1}{n}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) \\ \widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{i1} - \widehat{\beta}_2 x_{i2} - \cdots - \widehat{\beta}_p x_{ip})^2\end{aligned}$$

6.3.2 Inferência sobre os Coeficientes de Regressão

- As demonstrações acima apresentam as EMV (e as EQM) como **combinações lineares** do vetor da variável resposta.

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

- Portanto, se $\mathbf{y}$ tem distribuição Gaussiana Multivariada o mesmo acontece com $\mathbf{b}$.
- Mas qual o valor esperado e a variância-covariância de $\mathbf{b}$?
- Primeiramente vejamos uma expressão alternativa para $\mathbf{b}$:

$$\begin{aligned}\mathbf{b} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\end{aligned}$$

- O valor esperado, portanto, é

$$\begin{aligned}\boldsymbol{E}\{\mathbf{b}\} &= \boldsymbol{E}\left\{\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right\} \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{E}\{\boldsymbol{\varepsilon}\} \\ &= \boldsymbol{\beta}.\end{aligned}$$

- A variância fica

$$\begin{aligned}\boldsymbol{Var}\{\mathbf{b}\} &= \boldsymbol{Var}\{\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\} \\ &= \boldsymbol{Var}\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\}\end{aligned}$$

$$\begin{aligned} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{Var}\{\boldsymbol{\varepsilon}\}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\sigma^2\mathbf{I}\right)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

- Logo, as EMV (e as EQM) são

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

- Para estimarmos a variância de $\mathbf{b}$ necessitamos apenas de uma estimativa para a variância do erro. Como ela é estimada pelo QMR temos

$$\widehat{\boldsymbol{Var}}\{\mathbf{b}\} = QMR\,(\mathbf{X}'\mathbf{X})^{-1}.$$

- Para obtermos a variância da estimativa de um coeficiente de regressão particular b_k, basta estrair o k$^{\text{ésimo}}$ elemento da diagonal principal da matrix de variância-covariância:

$$\widehat{\boldsymbol{Var}}\{b_k\} = QMR\,\left[(\mathbf{X}'\mathbf{X})^{-1}\right]_{kk}.$$

- Com base nesses resultados, podemos construir Intervalos de Confiança de $(1-\alpha)100\%$ para cada uma das estimativas dos coeficientes de regressão:

$$b_k \pm t(1-\alpha/2; n-p-1)\sqrt{\widehat{\boldsymbol{Var}}\{b_k\}},$$

  onde $t(1-\alpha/2; n-p)-1$ é o quantil $(1-\alpha/2)$ da distribuição t de Student com $(n-p-1)$ graus de liberdade.

- Note que o modelo múltiplo possui p variável preditoras, mas são $(p+1)$ coeficientes de regressão, pois include o intercepto (β_0).

- Para testarmos a hipótese nula

$$H_0 : \beta_k = \beta_{k0},$$

  utilizamos a estatística

$$t^* = \frac{b_k - \beta_{k0}}{\sqrt{\widehat{\boldsymbol{Var}}\{b_k\}}},$$

  que sob H_0 segue a distribuição t de Student com $(n-p-1)$ graus de liberade.

- No caso de inferência simultânea, o fato de $\mathbf{b}$ possuir distribuição Gaussiana Multivariada, nos permite definir uma Região de Confiança de $(1-\alpha)100\%$ através da expressão

$$\frac{(\mathbf{b}-\boldsymbol{\beta})'(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{b}-\boldsymbol{\beta})}{2QMR} \leq F(1-\alpha;p;n-p-1)$$

onde $F(1-\alpha;p;n-p-1)$ é o quantil $(1-\alpha)$ da distribuição F com p graus de liberdade no numerador e $(n-p-1)$ graus de liberdade no denominador.

- Intervalos de Confiança Simultâneos de $(1-\alpha)100\%$ podem ser construídos utilizando o Método de Bonferoni:

$$b_k \pm B\sqrt{\widehat{\boldsymbol{Var}}\{b_k\}},$$

onde $B = t(1-\alpha/(2g), n-p-1)$ e g é número de intervalos simultâneos.

6.3.3 Inferência sobre Resposta Média e Predição

- Uma **única** resposta média para uma nova observação das variáveis preditoras pode ser definida como um produto vetorial interno:

Vetor da nova observação: $\mathbf{x}_h = \begin{bmatrix} 1 \\ x_{h1} \\ x_{h2} \\ \vdots \\ x_{hp} \end{bmatrix}$

Resposta Média: $\widehat{y}_h = \mathbf{x}_h'\mathbf{b} = \begin{bmatrix} 1 & x_{h1} & x_{h2} & \cdots & x_{hp} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix}$

- A notação matricial nos permite encontrar um **vetor** de respostas média, basta que as novas observações sejam organizadas nas linhas de uma matrix de novas observações:

Matrix de m novas observações: $\mathbf{X}_h = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & x_{m2} & \cdots & x_{mp} \end{bmatrix}$

Vetor de Respostas Médias: $\widehat{\mathbf{y}}_\mathbf{h} = \mathbf{X}_h\mathbf{b}$

- Uma vez que a resposta média é uma estimativa, seu vetor de valores esperados será

$$\begin{aligned} \boldsymbol{E}\{\widehat{\mathbf{y}}_\mathbf{h}\} &= \boldsymbol{E}\{\mathbf{X}_h\mathbf{b}\} \\ &= \mathbf{X}_h\boldsymbol{E}\{\mathbf{b}\} \\ &= \mathbf{X}_h\boldsymbol{\beta}, \end{aligned}$$

  sendo portanto uma estimativa não-viciada da superfície de regressão.

- A variância do vetor de respostas médias é obtida por

$$\begin{aligned} \boldsymbol{Var}\{\widehat{\mathbf{y}}_\mathbf{h}\} &= \boldsymbol{Var}\{\mathbf{X}_h\mathbf{b}\} \\ &= \mathbf{X}_h\,\boldsymbol{Var}\{\mathbf{b}\}\,\mathbf{X}_h{}' \\ &= \mathbf{X}_h\left(\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)\mathbf{X}_h{}' \\ &= \sigma^2\left(\mathbf{X}_h(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h{}'\right) \end{aligned}$$

- Note que $\boldsymbol{Var}\{\widehat{\mathbf{y}}_\mathbf{h}\}$ é uma matrix de variância-covariância de ordem $(m \times m)$.

- Essa matrix é estimada por

$$\widehat{\boldsymbol{Var}}\{\widehat{\mathbf{y}}_\mathbf{h}\} = QMR\left(\mathbf{X}_h(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h{}'\right),$$

  e a variância de uma resposta média em particular ($\widehat{y}_k$) é obtida extraindo-se o k$^{\text{ésimo}}$ elemento da diagonal da matrix acima

$$\widehat{\boldsymbol{Var}}\{\widehat{y}_k\} = QMR\left[\mathbf{X}_h(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h{}'\right]_{kk}.$$

- O vetor de respostas médias é também uma combinação linear do vetor da variável reposta

$$\widehat{\mathbf{y}}_\mathbf{h} = \mathbf{X}_h\,\mathbf{b} = \mathbf{X}_h(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathbf{y}$$

  portanto

$$\widehat{\mathbf{y}}_\mathbf{h} \sim N\left(\mathbf{X}_h\boldsymbol{\beta}\,,\,\sigma^2\left[\mathbf{X}_h(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h{}'\right]\right).$$

- Esses resultados permitem a construção dos Intervalos de Confiança na forma

$$\widehat{y}_k \pm t(1-\alpha/2; n-p-1)\sqrt{\boldsymbol{Var}\{\widehat{y}_k\}}.$$

- No caso de inferência simultâneas, podemos tanto utilizar o Método de Bonferoni quanto utilizarmos a Banda de Confiança pelo Método Working-Hotteling:

$$\widehat{y}_k \pm W \sqrt{\boldsymbol{Var}\{\widehat{y}_k\}}$$

onde

$$W = \sqrt{(p+1)\, F(1-\alpha; p; n-p-1)}.$$

- No caso de **predições**, sabemos que o vetor de predição será igual ao vetor das respostas médias, mas a variância deve incluir a incerteza sobre onde está a superfície de regressão e a incerteza intrinseca à variável resposta que possui variância igual a $\sigma^2\mathbf{I}$.

Assim, a variância do vetor de predições é

$$\begin{aligned} \widehat{\mathbf{y}}_\mathbf{h} &= \sigma^2\mathbf{I} + \sigma^2 \left[\mathbf{X}_h(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h{}'\right] \\ &= \sigma^2 \left[\mathbf{I} + \mathbf{X}_h(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h{}'\right] \end{aligned}$$

que pode ser estimada por

$$\widehat{\boldsymbol{Var}}\{\widehat{\mathbf{y}}_\mathbf{h}\} = QMR \left[\mathbf{I} + \mathbf{X}_h(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h{}'\right],$$

sendo que a variância de uma predição em particular ($\widehat{y}_k$) será dada pelo k$^{\text{ésimo}}$ elemento da diagonal dessa matrix

$$\widehat{\boldsymbol{Var}}\{\widehat{y}_k\} = QMR \left[\mathbf{I} + \mathbf{X}_h(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h{}'\right]_{kk}.$$

- Portanto, os **Intervalos de Predição** de $(1-\alpha)100\%$ de probabilidade são obtidos por

$$\widehat{y}_k \pm t(1-\alpha/2; n-p-1) \sqrt{\widehat{\boldsymbol{Var}}\{\widehat{y}_k\}}.$$

6.4 Qualidade do Ajuste no Modelo Linear Múltiplo

6.4.1 Teste F do Modelo

- Já foi visto que a variabilidade da variável resposta, media na forma de soma de quadrados, pode ser desdobrada em dois compontentes:

$$\begin{array}{rcl} SQT = & SQM & +SQR \\ \sum_{i=1}^{n}(y_i-\overline{y})^2 = & \sum_{i=1}^{n}(\widehat{y}_i-\overline{y})^2 & +\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2 \end{array}$$

- Em notação matricial esse desdobramento é obtido por

$$\begin{array}{rcl} SQT = & SQM & +SQR \\ \mathbf{y}'\mathbf{M}^0\mathbf{y} = & \mathbf{y}'\mathbf{P}\mathbf{y} & +\mathbf{y}'\mathbf{M}\mathbf{y} \\ \mathbf{y}'\mathbf{M}^0\mathbf{y} = & \mathbf{y}'\mathbf{P}\mathbf{y} & +\mathbf{y}'(\mathbf{I}-\mathbf{P})\mathbf{y} \end{array}$$

- Associados a cada soma de quadrados temos os seus graus de liberdade, o que nos permite obter os quadrados médios:

$$\begin{array}{lrcl} \text{Quadrado Médio do Modelo:} & QMM & = & \dfrac{SQM}{p} \\ \text{Quadrado Médio do Resíduo:} & QMM & = & \dfrac{SQR}{n-p-1}. \end{array}$$

- Esses quadrados médios são estimativas da variância do erro

$$\begin{array}{rcl} \boldsymbol{E}\{QMR\} & = & \sigma^2 \\ \boldsymbol{E}\{QMM\} & = & \sigma^2+\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}. \end{array}$$

- Assim, podemos testar a hipótese nula

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

contra a hipótese alternativa

$$H_\alpha : \text{para pelo menos um } j \quad \beta_j \neq 0, \qquad j = 1, 2, \ldots, p.$$

utilizando a estatística

$$F^* = \frac{QMM}{QMR}.$$

- Sob a hipótese nula, essa estatítisca tem distribuição F com p graus de liberdade no numerador e $(n-p-1)$ graus de liberdade no denominador.

6.4.2 Coeficiente de Determinação

- De modo análogo ao modelo linear simples, o coeficiente de determinação indica a proporção da variabilidade da variável resposta que é explicada pelo modelo de regressão, sendo calculado da mesma forma:

$$\begin{aligned} R^2 &= \frac{SQM}{SQT} = \frac{\mathbf{y}'\mathbf{P}\mathbf{y}}{\mathbf{y}'\mathbf{M}^0\mathbf{y}} \\ R^2 &= 1 - \frac{SQR}{SQT} = 1 - \frac{\mathbf{y}'\mathbf{M}\mathbf{y}}{\mathbf{y}'\mathbf{M}^0\mathbf{y}} \end{aligned}$$

- No modelo linear múltiplo, o número de coeficientes sendo ajustados pode ser bastante grande, de modo que é comum ajustar o coeficiente de determinação para considerar a relação entre tamanho de amostra (n) e o número de coeficientes de regressão ($p+1$):

$$R^2 = 1 - \frac{SQR/(n-p-1)}{SQT/(n-1)} = 1 - \left(\frac{n-1}{n-p-1}\right)\frac{SQR}{SQT}$$

- O resultado é que coef. de determinação ajustado é sempre inferior ao coeficiente de determinação sem o ajuste.

6.5 Exercícios

6.5-1. Um florestal ajustou o seguinte modelo a dados de 60 parcelas de *E. grandis* em primeira rotação:

$$\ln(y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon$$

onde:

$\ln(y_i)$ é o logaritmo da produção ($m^3 ha^{-1}$);

x_{i1} é o inverso da idade da floresta (em anos);

x_{i2} é a altura média das árvores dominantes (m); e

x_{i3} é a área basal da floresta ($m^2 ha^{-1}$).

Ele encontrou $SQR = 0.3467$ e as seguintes matrizes:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 60.0000 & 12.3179 & 1251.5900 & 997.5800 \\ 12.3179 & 2.7907 & 245.5503 & 195.6235 \\ 1251.5900 & 245.5503 & 26867.7127 & 21411.1413 \\ 997.5800 & 195.6235 & 21411.1413 & 17356.2476 \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 3.5508 & -5.7135 & -0.1106 & -0.0033 \\ -5.7135 & 11.0231 & 0.1609 & 0.0056 \\ -0.1106 & 0.1609 & 0.0058 & -0.0026 \\ -0.0033 & 0.0056 & -0.0026 & 0.0034 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 282.87197 \\ 56.76911 \\ 5979.62121 \\ 4785.02693 \end{bmatrix}$$

Pede-se:

(a) Encontrre as estimativas de quadrados mínimos dos coeficientes de regressão.

(b) Encontre os erros padrão dessas estimativas.

(c) Encontre os Intervalos de Confiança de 95% para essas estimativas.

(d) Encontre os valores ajustados para as seguintes novas observações:

$$\mathbf{X}_h = \begin{bmatrix} 1 & 1/3.5 & 16 & 12 \\ 1 & 1/4.5 & 20 & 14 \\ 1 & 1/5.5 & 24 & 16 \\ 1 & 1/6.5 & 28 & 18 \end{bmatrix}$$

(e) Assumindo que os valores ajustados representam estimativas para a resposta média, encontre os *Intervalos de Confiança* de 95% para essas estimativas.

(f) Converta os intervalos de confiança para a escala original da variável resposta.

(g) Assumindo que os valores ajustados representam predições encontre os *Intervalos de Predição* de 95% para essas estimativas da resposta média.

(h) Converta os predição de confiança para a escala original da variável resposta.

CAPÍTULO 7

TESTE DE HIPÓTESES NO MODELO LINEAR MÚLTIPLO

7.1 Desdobramento da Soma de Quadrados do Modelo

- A soma de quadrados do modelo (SQM) num modelo linear múltiplo pode ser desdobrada em efeitos aditivos relativos a cada variável preditora.
- Examinemos o caso de duas variáveis preditoras:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon$$

- Utilizaremos a notação $SQM(X_1, X_2)$ para indicarmos a soma de quadrados do modelo para o modelo com as variáveis preditoras X_1 e X_2, o mesmo sendo válido para a soma de quadrados do resíduo.

$$SQT = SQM(X_1, X_2) + SQR(X_1, X_2).$$

- Se considerarmos uma única variável preditora para a **mesma** variável resposta temos:

$$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon$$

$$SQT = SQM(X_1) + SQR(X_1).$$

- Como em ambos os modelos se trata da mesma variável resposta, a soma de quadrados totais (SQT) é a mesma, logo

$$SQM(X_1, X_2) + SQR(X_1, X_2) = SQM(X_1) + SQR(X_1)$$

$$SQM(X_1, X_2) - SQM(X_1) = SQR(X_1) - SQR(X_1, X_2)$$

- Essa diferença entre a soma de quadrados do modelo, ou entre a soma de quadrados do resíduo, entre os dois modelos é chamada de **Soma de Quadrados Extra** de X_2 dado que X_1 está no modelo.

  Ela será designada por: $SQM(X_2|X_1)$.

- A $SQM(X_2|X_1)$ é:

  – o **acréscimo** na SQM

  $$SQM(X_2|X_1) = SQM(X_1, X_2) - SQM(X_1);$$

  – a **redução** na SQR

  $$SQM(X_2|X_1) = SQR(X_1) - SQR(X_1, X_2)$$

  devido à introdução da variável preditora X_2 no modelo que continha a variável X_1.

- Lembremos que:

  – a SQT é interpretada como a variabilidade *total* da variável resposta;
  – a SQM é a variabilidade *explicada* pelo modelo; e
  – a SQR é a variabilidade *não explicada* pelo modelo.

  Dessa forma, a $SQM(X_2|X_1)$ é ou aumento na *"capacidade explicativa"* ao se introduzir a variável X_2 no modelo que já continha a variável X_1.

7.1.1 Interpretação Vetorial do Desdobramento da SQM

- Já foi mostrado que no modelo

  $$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon$$

  o desdobramento da SQT pode ser apresentada na forma matricial

  $$\begin{array}{rcl} SQT = & SQM & + SQR \\ \mathbf{y'M^0y} = & \mathbf{b'X'M^0Xb} & + \mathbf{e'e}. \end{array}$$

- Para facilitar a interpretação vetorial, tomemos o mesmo modelo subtraindo de cada variável a respectiva média:

  $$(y_i - \overline{y}) = \beta_1(x_{i1} - \overline{x}_1) + \beta_2(x_{i2} - \overline{x}_2) + \varepsilon$$

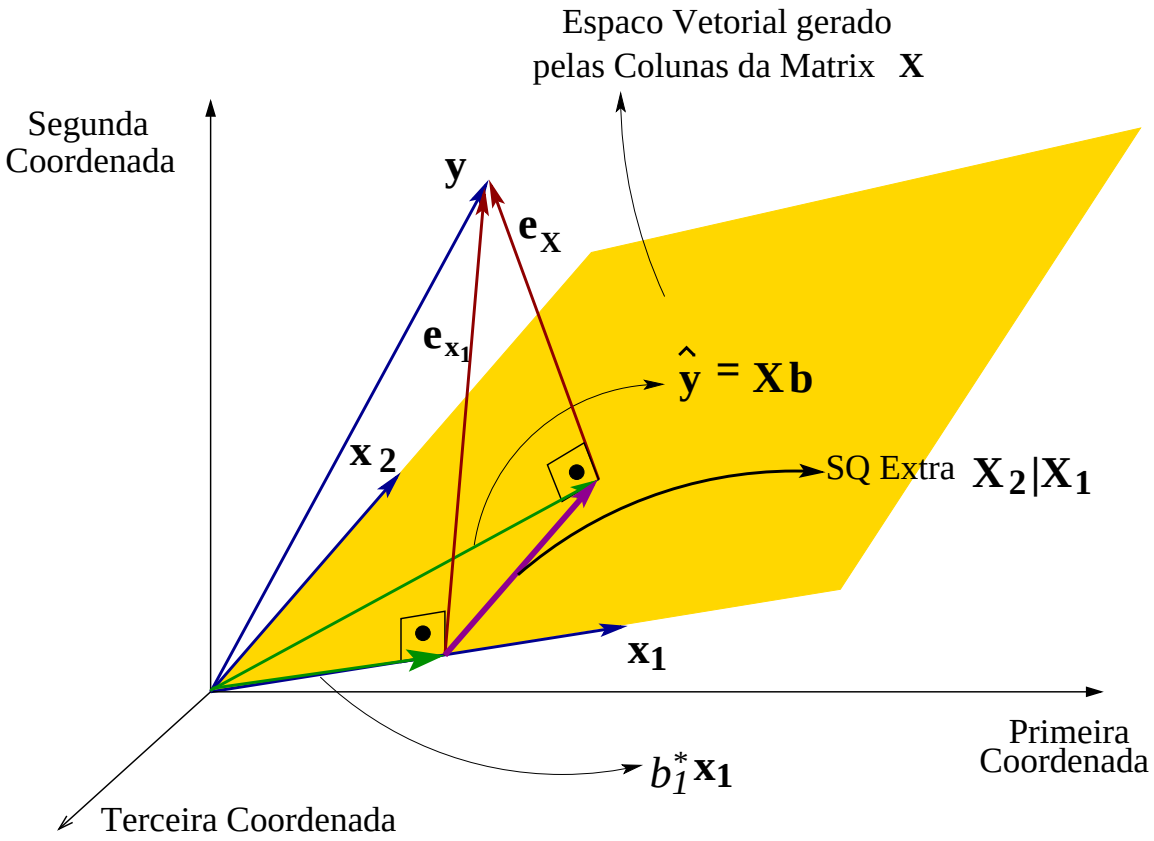


Figura 7.1: Interpretação vetorial da **Soma de Quadrados Extra** $X_2|X_1$, num espaço vetorial tridimensional.

de forma que o desdobramento da SQT se torna

$$\begin{aligned} SQT &= SQM + SQR \\ \mathbf{y}'\mathbf{y} &= \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} + \mathbf{e}'\mathbf{e}. \end{aligned}$$

- O que significa que

  $SQT = ||\mathbf{y}||^2$ é o quadrado do comprimento do vetor da variável resposta;

  $SQM = ||\mathbf{X}\mathbf{b}||^2$ é o quadrado do comprimento do vetor dos valores ajustados, i.e., a projeção de $\mathbf{y}$ no espaço gerado pelas colunas da matrix de delineamento;

  $SQR = ||\mathbf{e}||^2$ é o quadrado do comprimento do vetor dos resíduos.

- A figura 7.1 apresenta uma ilustração da interpretação vetorial da SQ Extra.

- Pela figura podemos perceber que o vetor da SQ Extra pode ($\mathbf{v}_{X_2|X_1}$) ser representado por *duas diferenças de projeções*:

  Projeções de $\mathbf{y}$ sobre $\mathbf{X}$ e $\mathbf{x}_1$:

  - $\mathbf{Xb}$ é a projeção de $\mathbf{y}$ sobre o plano gerado pelas colunas de $\mathbf{X} = [\mathbf{x}_1\, \mathbf{x}_2]$.
  - $b_1^*\, \mathbf{x}_1$ é a projeção de $\mathbf{y}$ sobre $\mathbf{x}_1$.
  - A SQ Extra é a diferença dessas projeções:

  $$\mathbf{v}_{X_2|X_1} = \mathbf{Xb} - b_1^*\mathbf{x}_1$$

  Nessa diferença, $\mathbf{v}_{X_2|X_1}$ é ortogonal a $\mathbf{x}_1$:

  $$\begin{aligned} \mathbf{v}_{X_2|X_1}{}'\mathbf{x}_1 &= (\mathbf{e}_{\mathbf{x}_1} - \mathbf{e}_{\mathbf{X}})'\,\mathbf{x}_1 \\ &= \mathbf{e}_{\mathbf{x}_1}{}'\mathbf{x}_1 - \mathbf{e}_{\mathbf{X}}{}'\mathbf{x}_1 \\ &= 0 - 0 = 0 \end{aligned}$$

  Isso acontece porque:

  - $\mathbf{e}_{\mathbf{x}_1}$ é o resíduo associado a $\mathbf{x}_1$, portanto $\mathbf{e}_{\mathbf{x}_1}{}'\mathbf{x}_1 = 0$.
  - $\mathbf{e}_{\mathbf{X}}$ é o resíduo associado à matrix de delineamento $\mathbf{X}$, sendo portanto ortogonal ao plano gerado pelas suas colunas e, consequentemente, ortogonal a cada uma das variáveis preditoras:

  $$\begin{aligned} \mathbf{0} &= \mathbf{e}_{\mathbf{X}}{}'\mathbf{X} \\ &= \mathbf{e}_{\mathbf{X}}{}'[\mathbf{x}_1 \quad \mathbf{x}_2] \\ &= [\mathbf{e}_{\mathbf{X}}{}'\mathbf{x}_1 \quad \mathbf{e}_{\mathbf{X}}{}'\mathbf{x}_2] \\ &= [\mathbf{0} \quad \mathbf{0}] \end{aligned}$$

  Como resultado forma-se um triângulo retângulo na diferença dessas projeções e aplicando-se o Teorema de Pitágoras obtemos:

  $$\begin{aligned} ||\mathbf{v}_{X_2|X_1}||^2 &= ||\mathbf{Xb}||^2 - ||b_1^*\mathbf{x}_1||^2 \\ \mathbf{v}_{X_2|X_1}{}'\mathbf{v}_{X_2|X_1} &= \mathbf{b}'\mathbf{X}'\mathbf{Xb} - b_1^{*2}\mathbf{x}_1{}'\mathbf{x}_1 \end{aligned}$$

  $$SQM(X_2|X_1) = SQM(X_1, X_2) - SQM(X_1)$$

  Projeções de $\mathbf{y}$ sobre os planos ortogonais a $\mathbf{X}$ e a $\mathbf{x}_1$:

  - $\mathbf{e}_{\mathbf{X}}$ é a projeção de $\mathbf{y}$ sobre o plano *ortogonal* ao plano gerado pelas colunas de $\mathbf{X} = [\mathbf{x}_1\, \mathbf{x}_2]$.

- $\mathbf{e}_{\mathbf{x}_1}$ é *ortogonal* à projeção de $\mathbf{y}$ sobre $\mathbf{x}_1$.
- A SQ Extra é a diferença dessas projeções:

$$\mathbf{v}_{X_2|X_1} = \mathbf{e}_{\mathbf{x}_1} - \mathbf{e}_{\mathbf{X}}$$

Nessa diferença, $\mathbf{v}_{X_2|X_1}$ é ortogonal a $\mathbf{e}_{\mathbf{X}}$:

$$\begin{aligned} \mathbf{e'_X v}_{X_2|X_1} &= \mathbf{e'_X}(\mathbf{Xb} - b_1^*\mathbf{x}_1) \\ &= \mathbf{e'_X Xb} - b_1^*\mathbf{e'_X x}_1 \\ &= 0 - b_1^*(0) = 0. \end{aligned}$$

Como resultado forma-se um triângulo retângulo na diferença dessas projeções e aplicando-se o Teorema de Pitágoras obtemos:

$$\begin{aligned} ||\mathbf{v}_{X_2|X_1}||^2 &= ||\mathbf{e}_{\mathbf{x}_1}||^2 - ||\mathbf{e}_{\mathbf{X}}||^2 \\ \mathbf{v}_{X_2|X_1}{}'\mathbf{v}_{X_2|X_1} &= \mathbf{e}_{\mathbf{x}_1}{}'\mathbf{e}_{\mathbf{x}_1} - \mathbf{e}_{\mathbf{X}}{}'\mathbf{e}_{\mathbf{X}} \end{aligned}$$

$$SQM(X_2|X_1) = SQR(X_1) - SQR(X_1, X_2)$$

- A SQ Extra também pode ser vista ela própria como uma projeção:
  - Inicialmente ajustamos a variável resposta em função apenas da variável preditora X_1:
  $$(y_i - \overline{y}) = \beta_1^*(x_{i1} - \overline{x}_1) + \varepsilon^*$$
  que gera os resíduos
  $$e_{i\mathbf{x}_1} = (y_i - \overline{y}) - b_1^*(x_{i1} - \overline{x}_1) \quad \Rightarrow \quad \mathbf{e}_{\mathbf{x}_1} = \mathbf{y} - b_1^*\mathbf{x}_1.$$
  - Esses **resíduos** são ajustados em função da variável preditora X_2
  $$e_{i\mathbf{x}_1} = \beta_2^*(x_{i2} - \overline{x}_2) + \varepsilon$$
  que resulta num modelo onde podemos fazer o desdobramento
  $$\begin{aligned} SQT = &\quad SQM &&+SQR \\ \mathbf{e}'_{\mathbf{x}_1}\mathbf{e}_{\mathbf{x}_1} = &\quad b_2^*\mathbf{x}'_2\mathbf{x}_2 &&+\mathbf{e'_X e_X} \\ SQR(X_1) = &\quad SQM(X_2|X_1) &&+SQR(X_1, X_2). \end{aligned}$$
  - Portanto, a $SQM(X_2|X_1)$ é a **projeção do resíduo do modelo de $Y|X_1$ sobre o vetor da variável preditora X_2.**

7.1.2 Desdobramento da SQM para Variáveis Ortogonais

- Pela interpretação da figura 7.1 podemos ver também que

$$SQM(X_2|X_1) \neq SQM(X_1|X_2),$$

pois trata-se de duas situações diferentes.

- Também interpretando a figura 7.1, vemos que para duas variáveis preditoras X_1 e X_2 **não ortogonais** ocorrerá que

$$\begin{aligned} SQM(X_2|X_1) &\neq SQM(X_2) \\ SQM(X_1|X_2) &\neq SQM(X_1). \end{aligned}$$

- Por outro lado, se as variáveis preditoras X_1 e X_2 são **não correlacionadas**, então:

$$\begin{aligned} \textbf{\textit{Cov}}\{X_1, X_2\} = \frac{\sum_{i=1}^{n}(x_{i1}-\overline{x}_1)(x_{i2}-\overline{x}_2)}{\sqrt{\sum_{i=1}^{n}(x_{i1}-\overline{x}_1)^2 \sum_{i=1}^{n}(x_{i2}-\overline{x}_2)^2}} &= 0 \\ \frac{\mathbf{x'}_1\mathbf{x}_2}{\sqrt{(\mathbf{x'}_1\mathbf{x}_1)(\mathbf{x'}_2\mathbf{x}_2)}} &= 0 \quad \Rightarrow \quad \mathbf{x'}_1\mathbf{x}_2 = 0 \end{aligned}$$

ou seja, elas são **ortogonais**.

- Num modelo linear onde as variáveis preditoras são ortogonais temos a soma de quadrados do modelo igual a

$$\begin{aligned} SQM(X_1, X_2) &= \mathbf{b'X'Xb} = \begin{bmatrix} b_1 & b_2 \end{bmatrix} [\mathbf{x'}_1\, \mathbf{x'}_2][\mathbf{x}_1\, \mathbf{x}_2] \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \\ &= \begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} \mathbf{x'}_1\mathbf{x}_1 & \mathbf{x'}_1\mathbf{x}_2 \\ \mathbf{x'}_2\mathbf{x}_1 & \mathbf{x'}_2\mathbf{x}_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \\ &= \begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} \mathbf{x'}_1\mathbf{x}_1 & 0 \\ 0 & \mathbf{x'}_2\mathbf{x}_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \\ &= b_1^2(\mathbf{x'}_1\mathbf{x}_1) + b_2^2(\mathbf{x'}_2\mathbf{x}_2) \\ &= SQM(X_1) + SQM(X_2). \end{aligned}$$

- Conseqüentemente, a ortogonalidade resulta em

$$\begin{aligned} SQM(X_1|X_2) &= SQM(X_1) \\ SQM(X_2|X_1) &= SQM(X_2). \end{aligned}$$

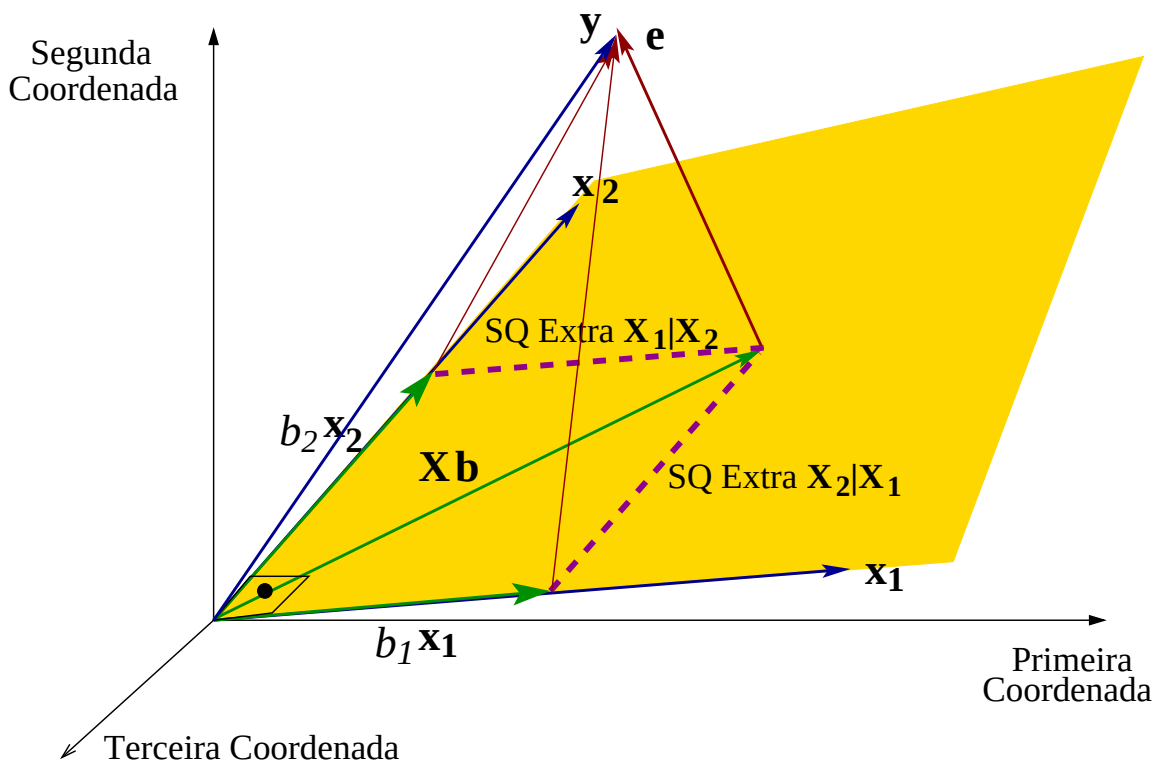


Figura 7.2: Interpretação vetorial da Soma de Quadrados Extra para um modelo com duas **variáveis preditoras ortogonais**, num espaço vetorial tridimensional.

7.1.3 Desdobramento da SQM para Diversas Variáveis

- Vejamos a decomposição para 3 variáveis preditoras, a partir dos modelos

Modelo 1:	$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i$
Modelo 2:	$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$
Modelo 3:	$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$

- Tomando o modelo 3 e 2 temos:

$$\begin{aligned} SQT &= SQT \\ SQM(X_1, X_2, X_3) + SQR(X_1, X_2, X_3) &= SQM(X_1, X_2) + SQR(X_1, X_2) \\ SQM(X_1, X_2, X_3) - SQM(X_1, X_2) &= SQR(X_1, X_2) - SQR(X_1, X_2, X_3) \end{aligned}$$

o que define a SQ Extra:

$$SQM(X_3|X_1, X_2) = SQR(X_1, X_2) - SQR(X_1, X_2, X_3)$$

$$= SQM(X_1, X_2, X_3) - SQM(X_1, X_2)$$

e, portanto,

$$SQM(X_1, X_2, X_3) = SQM(X_1, X_2) + SQM(X_3|X_1, X_2).$$

- Tomando o modelo 2 e 1 temos:

$$\begin{aligned} SQT &= SQT \\ SQM(X_1, X_2) + SQR(X_1, X_2) &= SQM(X_1) + SQR(X_1) \\ SQM(X_1, X_2) - SQM(X_1) &= SQR(X_1) - SQR(X_1, X_2) \end{aligned}$$

o que define a SQ Extra:

$$\begin{aligned} SQM(X_2|X_1) &= SQR(X_1) - SQR(X_1, X_2) \\ &= SQM(X_1, X_2) - SQM(X_1) \end{aligned}$$

e, portanto,

$$SQM(X_1, X_2) = SQM(X_1) + SQM(X_2|X_1).$$

- Substituindo esse segundo resultado no primeiro obtemos

$$SQM(X_1, X_2, X_3) = SQM(X_1) + SQM(X_2|X_1) + SQM(X_3|X_1, X_2).$$

- É importante notar que essa não é a única decomposição possível. Se alteramos a **seqüência** das variáveis preditoras no modelo podemos ter diferentes decomposições:

$$\begin{aligned} SQM(X_1, X_2, X_3) &= SQM(X_3) + SQM(X_2|X_3) + SQM(X_1|X_3, X_2) \\ &= SQM(X_2) + SQM(X_3|X_2) + SQM(X_1|X_2, X_3) \\ &= SQM(X_1) + SQM(X_1|X_3) + SQM(X_2|X_1, X_3) \end{aligned}$$

- A idéia da decomposição seqüencial é facilmente generalizável para um modelo múltiplo com p variáveis preditoras:

$$\begin{aligned} SQM(X_1, X_2, X_3, \ldots, X_p) = {} & SQM(X_1) + SQM(X_2|X_1) + SQM(X_3|X_1, X_2) + \\ & + \cdots + SQM(X_p|X_1, X_2, X_3, \ldots, X_{p-1}). \end{aligned}$$

- **Conclusão 1:** a SQM de um Modelo Linear Múltiplo pode ser decomposta numa **seqüência** de Soma de Quadrados Extra ou **Soma de Quadrados Marginais**, cada uma associada a **um grau de liberdade**, relativas a introdução seqüencial das variáveis preditoras no modelo múltiplo.

- **Conclusão 2:** a decomposição seqüencial não é única, mas todas elas totalizam a mesma SQM.

7.2 Teste F Parcial

- Já vimos que o teste F da qualidade do ajuste verifica o ajuste geral do modelo linear múltiplo aos dados.
- O desdobramento da Soma de Quadrados do Modelo abre, no entanto, a possibilidade de se testar a importância de diferentes variáveis preditoras no modelo linear múltiplo.

7.2.1 Teste F Seqüencial

- A decomposição seqüencial da SQM nos permite realizar uma comparação estatística seqüencial de modelos progressivamente complexos.
- Suponha a série de modelos:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \varepsilon_i \\ y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \\ y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \\ &\vdots \quad\quad \vdots \\ y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i. \end{aligned}$$

- A decomposição seqüencial da SQM do último modelo gera uma série de somas de quadrados marginais, cada uma com um grau de liberdade, nos permitindo construir uma tabela de Análise de Variância:

Fonte de Variação		**Graus de Lib.**	**S.Q.**
Modelo		p	SQM
	X_1	1	$SQM(X_1)$
	$X_2 \vert X_1$	1	$SQM(X_2 \vert X_1)$
	$X_3 \vert X_1, X_2$	1	$SQM(X_3 \vert X_1, X_2)$
	$\cdots$	$\cdots$	$\cdots$
	$X_p \vert X_1, X_2, \ldots, X_{p-1}$	1	$SQM(X_p \vert X_1, X_2, \ldots, X_{p-1})$
Resíduo		$n-p-1$	SQR
Total		$n-1$	SQT

- Observando cada soma de quadrados marginal, temos:

$$\begin{aligned} SQM(X_k|X_1, X_2, \ldots, X_{k-1}) &= SQM(X_1, X_2, \ldots, X_{k-1}, X_k) \\ &\quad -SQM(X_1, X_2, \ldots, X_{k-1}) \\ &= \mathbf{b}_k^{'}\mathbf{X}_k^{'}\mathbf{X}_k\mathbf{b}_k - \mathbf{b}_{k-1}^{'}\mathbf{X}_{k-1}^{'}\mathbf{X}_{k-1}\mathbf{b}_{k-1} \end{aligned}$$

  onde $\mathbf{X}_k$ e $\mathbf{X}_{k-1}$ são as matrizes de delineamento com k e $k-1$ variáveis preditoras, respectivamente, e $\mathbf{b}_k$ e $\mathbf{b}_{k-1}$ são os vetores de coeficientes de regressão correspondentes.

- Associada a essa soma de quadrados, temos apenas um grau de liberdade, uma vez que ela se refere a uma única variável preditora (e um único coeficiente de regressão).

- O valor esperado do quadrado médio associado a essa soma de quadrados marginal é

$$\boldsymbol{E}\left\{\frac{SQM(X_k|X_1, X_2, \ldots, X_{k-1})}{1}\right\} = \sigma^2 + \left(\boldsymbol{\beta}_k^{'}\mathbf{X}_k^{'}\mathbf{X}_k\boldsymbol{\beta}_k - \boldsymbol{\beta}_{k-1}^{'}\mathbf{X}_{k-1}^{'}\mathbf{X}_{k-1}\boldsymbol{\beta}_{k-1}\right)$$

  Note que no caso de $\beta_k = 0$

$$\boldsymbol{\beta}_k^{'}\mathbf{X}_k^{'}\mathbf{X}_k\boldsymbol{\beta}_k = \boldsymbol{\beta}_{k-1}^{'}\mathbf{X}_{k-1}^{'}\mathbf{X}_{k-1}\boldsymbol{\beta}_{k-1}$$

  e o valor esperado se reduz a

$$\boldsymbol{E}\left\{\frac{SQM(X_k|X_1, X_2, \ldots, X_{k-1})}{1}\right\} = \sigma^2.$$

- Tomando agora o modelo com todas variáveis preditoras, o valor esperado do quadrado médio do resíduo é

$$\boldsymbol{E}\{QMR\} = \boldsymbol{E}\left\{\frac{SQR(X_1, X_2, \ldots, X_{k-1}, X_k)}{n-p-1}\right\} = \sigma^2.$$

- Assim, cada SQM marginal permite testar uma hipótese *marginal*

$$\begin{aligned} H_0 : \beta_k &= 0|(\beta_1, \beta_2, \ldots, \beta_{k-1}) \\ H_\alpha : \beta_k &\neq 0|(\beta_1, \beta_2, \ldots, \beta_{k-1}) \\ &\qquad \text{onde } k = 1, 2, \ldots, p; \end{aligned}$$

  com a estatística

$$F^* = \frac{SQM(X_k|X_1, X_2, \ldots, X_{k-1})/1}{SQR/(n-p-1)},$$

  que sob H_0 tem distribuição F com 1 grau de liberdade do numerador e $n-p-1$ graus de liberdade do denominador.

- Uma vez que a SQM foi decomposta **ortogonalmente** nas soma de quadrados marginais, os testes F seqüenciais **não são redundantes**.

7.2.2 Exemplo: Modelagem da Produção em Floresta de *E. grandis*

- Os modelos de produção de florestas geralmente modelam a produção, medida em termos de volume de madeira por unidade de área (m^3/ha), em função de três variáveis principais:

  Idade da floresta com variável temporal;

  Sítio: uma variável que totaliza os efeitos ambientais de um dado local sobre a "*capacidade produtiva*" da floresta.

  A altura média das árvores dominantes numa certa idade de referência é tomada como representativa dessa capacidade produtiva.

  Densidade da floresta medida índice de ocupação do "*espaço de crescimento*" disponível ao crescimento da floresta.

  Geralmente utiliza-se a área basal da floresta (m^2/ha) como representativa dessa densidade.

- A figura 7.3 mostra gráficos de dispersão das variáveis assumindo o modelo

$$\log(\text{Produção}) = \beta_0 + \beta_1 \frac{1}{\text{Idade}} + \beta_2(\text{Sítio}) + \beta_3(\text{Área Basal}) + \varepsilon$$

  numa floresta de *Eucalyptus grandis* em primeira rotação na região central do Estado de São Paulo.

- A ordem que as variáveis preditoras são apresentadas acima tem um fundamento prático de modelagem.

  – Sendo um modelo de produção, a idade é uma variável fundamental para que se possa fazer as predições ao longo do tempo.

  Se não houver relação entre produção e idade, os dados não permitem o desenvolvimento do modelo de produção.

  – A variável sítio representa a variação da capacidade produtiva devendo, portanto, ser considerada imediamente após à idade, pois modelaria a variação espacial da produção resultante da heterogeneidade da capacidade produtiva dos diferentes ambientes.

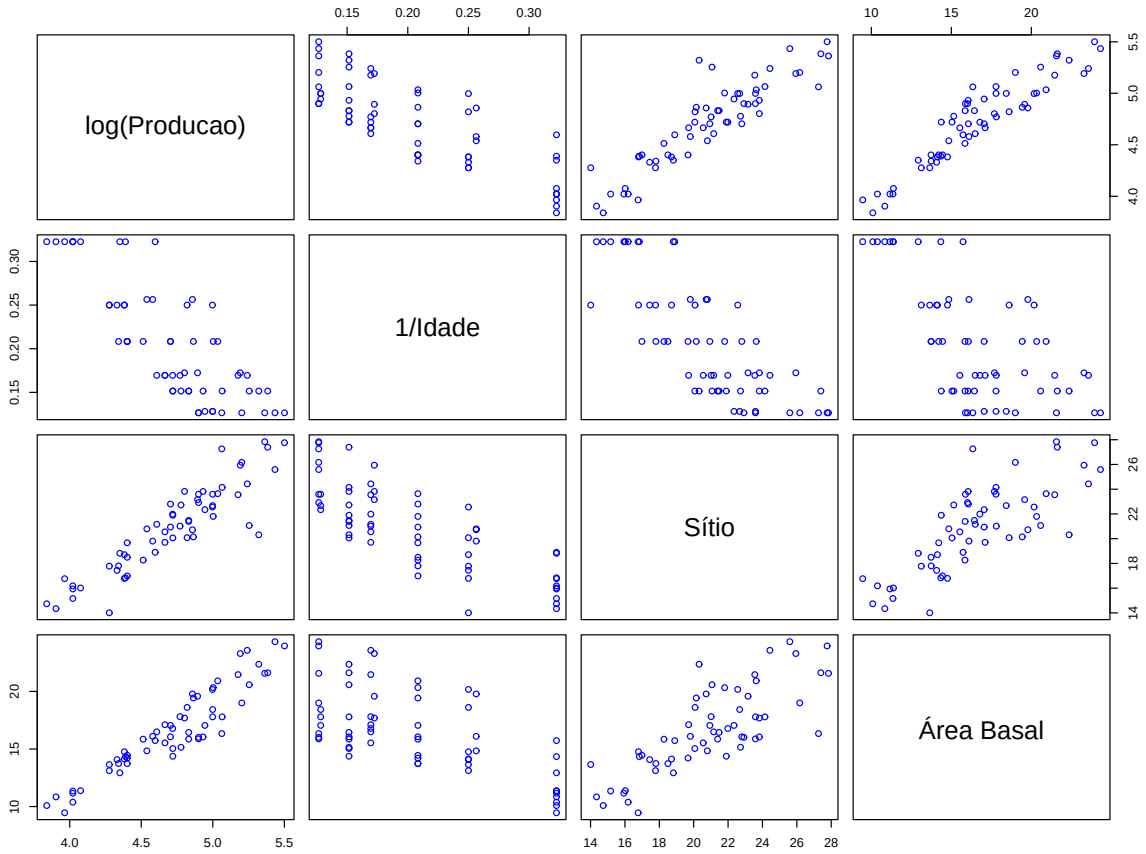


Figura 7.3: Gráficos de dispersão das variáveis utilizadas para a modelagem da produção de floresta de *E. grandis* em primeira rotação na região central do Estado de São Paulo.

  - Por fim, a densidade, medida pela área basal, é uma variável que tem poder explicativo desde que a a idade da floresta e sua capacidade produtiva sejam as mesmas sendo, conseqüentemente, a última variável preditora a ser incluida no modelo.

- Esse fundamento sugere que essa seja uma ordem apropriada para se testar a entrada das variáveis preditoras no modelo de produção.

- O teste F seqüencial desse modelo comprova que as três variáveis preditoras são importantes para o modelo de produção:

Fontes de Variação	**Graus de Liberdade**	**Soma de Quadrados**	**Quadrado Médio**	**Estatística F**	**Valor-p**
Modelo	3	9.6155	3.2052	517.8	$< 2.2 \times 10^{-16}$
1/Idade	1	6.4948	6.4948	1049.18	$< 2.2 \times 10^{-16}$
Sítio	1	1.8681	1.8681	301.78	$< 2.2 \times 10^{-16}$
Área Basal	1	1.2526	1.2526	202.34	$< 2.2 \times 10^{-16}$
Resíduo	56	0.3467	0.0062		
Total	59	9.9621			

- Embora as três variáveis preditoras são importantes, os testes seqënciais não resultam no mesmo valor das estatística F se a ordem das variáveis for mudada no modelo:

Fontes de Variação	**Graus de Liberdade**	**Soma de Quadrados**	**Quadrado Médio**	**Estatística F**	**Valor-p**
Área Basal	1	8.7102	8.7102	1407.060	$< 2.2 \times 10^{-16}$
Sítio	1	0.7466	0.7466	120.611	1.374×10^{-15}
1/Idade	1	0.1586	0.1586	25.628	4.788×10^{-06}
Sítio	1	8.1453	8.1453	1315.807	$< 2.2 \times 10^{-16}$
Área Basal	1	1.3115	1.3115	211.864	$< 2.2 \times 10^{-16}$
1/Idade	1	0.1586	0.1586	25.628	4.788×10^{-06}
Sítio	1	8.1453	8.1453	1315.81	$< 2.2 \times 10^{-16}$
1/Idade	1	0.2176	0.2176	35.15	1.985×10^{-07}
Área Basal	1	1.2526	1.2526	202.34	$< 2.2 \times 10^{-16}$

7.2.3 Outras Formas de Teste F Parcial

- Além do teste F seqüencial, a importância das variáveis preditoras no modelo linear múltiplo pode ser testada de muitas outras formas.
- Uma forma que pode ter grande utilidade é o teste de *grupos de variáveis preditoras* que representam um mesmo *tipo de influência* sobre a variável resposta.
- Suponha um modelo que visa explicaar o crescimento de uma floresta e possui 7 variáveis preditoras.

As três primeiras estão relacionadas com fatores do ambiente físico (efeitos abióticos), enquanto que as quatro últimas representam efeitos de interação entre as árvores na floresta (efeitos bióticos), como o efeito de competição, por exemplo.

- Se o pesquisador desejar testar esses dois tipos de influência, sem se preocupar com a importância de cada variável preditora em particular, a SQM pode ser desdobrada da seguinte maneira:

$$\begin{aligned} SQM(X_1,\ldots,X_7) &= SQM(X_1,X_2,X_3) \\ &\quad + SQM(X_4,X_5,X_6,X_7|X_1,X_2,X_3) \end{aligned}$$

$$\begin{aligned} SQM &= SQM(\text{efeitos abióticos}) \\ &\quad + SQM(\text{efeitos bióticos}|\text{efeitos abióticos}) \end{aligned}$$

- A tabela de análise de variância nesse caso seria

Fonte de Variação	**Graus de Lib.**	**S.Q.**
Modelo	7	SQM
Efeitos Abióticos	3	$SQM(X_1,X_2,X_3)$
Efeitos Bióticos | Efeitos Abióticos	4	$SQM(X_4,X_5,X_6,X_7\|X_1,X_2,X_3)$
Resíduo	$n-8$	SQR
Total	$n-1$	SQT

- As hipóteses nulas testadas nesse caso são:
  - Efeitos Abióticos:

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 = \beta_3 = 0 \\ H_\alpha &: \text{pelo menos um } \beta_j \neq 0, \quad \text{onde } j = 1,2,3. \end{aligned}$$

$$F = \frac{SQM(X_1,X_2,X_3)/3}{SQR/(n-8)}$$

  - Efeitos Bióticos | Efeitos Abióticos:

$$\begin{aligned} H_0 &: \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0 | \, (\beta_1 \neq 0, \beta_2 \neq 0, \beta_3 \neq 0) \\ H_\alpha &: \text{pelo menos um } \beta_j \neq 0, \quad \text{onde } j = 4,5,6,7. \end{aligned}$$

$$\begin{aligned} F &= \frac{SQM(X_4,X_5,X_6,X_7|X_1,X_2,X_3)/4}{SQR/(n-8)} \\ &= \frac{\left[SQR(X_1,X_2,X_3) - SQR(X_1,X_2,X_3,X_4,X_5,X_6,X_7)\right]/4}{SQR(X_1,X_2,X_3,X_4,X_5,X_6,X_7)/(n-8)} \end{aligned}$$

- Note que estatisticamente seria equivalente construir a seguinte tabela de análise de variância:

Fonte de Variação	**Graus de Lib.**	**S.Q.**
Modelo	7	SQM
Efeitos Bióticos	4	$SQM(X_4, X_5, X_6, X_7)$
Efeitos Abióticos | Efeitos Bióticos	3	$SQM(X_1, X_2, X_3 \| X_4, X_5, X_6, X_7)$
Resíduo	$n-8$	SQR
Total	$n-1$	SQT

- Também seria possível realizar todos os testes numa mesma análise

Fonte de Variação	**Graus de Lib.**	**S.Q.**
Modelo	7	SQM
Efeitos Abióticos	3	$SQM(X_1, X_2, X_3)$
Efeitos Bióticos	4	$SQM(X_4, X_5, X_6, X_7)$
Efeitos Bióticos | Efeitos Abióticos	4	$SQM(X_4, X_5, X_6, X_7 \| X_1, X_2, X_3)$
Efeitos Abióticos | Efeitos Bióticos	3	$SQM(X_1, X_2, X_3 \| X_4, X_5, X_6, X_7)$
Resíduo	$n-8$	SQR
Total	$n-1$	SQT

Mas, nesse caso, os testes seriam totalmente redundantes, pois os graus de liberdade dos testes somam mais que os graus de liberdade do modelo.

7.2.4 Coeficiente de Determinação Parcial

- Já foi visto que a *Soma de Quadrados Extra* pode ser vista como uma projeção do resíduo de um modelo ajustado sobre uma *nova* variável preditora.

- Ajustando-se o modelo

$$(y_i - \overline{y}) = \beta_1(x_{1i} - \overline{x}_1) + \varepsilon_i \qquad \Rightarrow \qquad e_{i1} = (y_i - \overline{y}) - b_1(x_{1i} - \overline{x}_1),$$

e em seqüência o modelo

$$e_{i1} = \beta_2(x_{2i} - \overline{x}_2) + \varepsilon_i^*,$$

temos que nesse segundo modelo ajustado o desdobramento da soma de quadrados resulta em

$$\begin{aligned} SQT = &\quad SQM &&+SQR \\ SQR(X_1) = &\quad SQM(X_2|X_1) &&+SQR(X_1, X_2). \end{aligned}$$

- O segundo modelo ajustado terá coeficiente de determinação igual a

$$\begin{aligned} R^2 &= \frac{SQM}{SQT} = \frac{SQM(X_2|X_1)}{SQR(X_1)} \\ &= \frac{SQR(X_1) - SQR(X_1, X_2)}{SQR(X_1)} = 1 - \frac{SQR(X_1, X_2)}{SQR(X_1)}. \end{aligned}$$

- Esse coeficiente é chamado como **Coeficiente de Determinação Parcial** de $Y : X_2$ dado X_1.

- O Coef. de Determinação Parcial pode ser interpretado de duas formas

$R^2_{Y2|1}$: proporção do resíduo do modelo de Y em função de X_1 que pode ser **reduzida** ao se acrescentar a variável X_2 ao modelo.

$R^2_{Y2|1}$: **aumento** da proporção da variabilidade de Y explicada pelo modelo que contém X_1, quando X_2 é acrescentada ao modelo.

- A segunda interpretação fica evidente se olharmos o desdobramento da SQM com as duas variáveis preditoras

$$\begin{aligned} SQM(X_1, X_2) &= SQM(X_1) + SQM(X_2|X_1) \\ \frac{SQM(X_1, X_2)}{SQT} &= \frac{SQM(X_1)}{SQT} + \frac{SQM(X_2|X_1)}{SQT} \\ \frac{SQM(X_1, X_2)}{SQT} &= \frac{SQM(X_1)}{SQT} + \frac{SQM(X_2|X_1)}{SQT}\left[\frac{SQR(X_1)}{SQR(X_1)}\right] \\ \frac{SQM(X_1, X_2)}{SQT} &= \frac{SQM(X_1)}{SQT} + \left(\frac{SQR(X_1)}{SQT}\right)\left(\frac{SQM(X_2|X_1)}{SQR(X_1)}\right) \\ R^2_{Y12} &= R^2_{Y1} + (1 - R^2_{Y1})(R^2_{Y2|1}) \end{aligned}$$

- Como o Coef. de Determinação Parcial tem uma relação direta com a SQ Extra, vale as mesmas generalizações:

$$\begin{aligned} R^2_{Y2|1} \neq R^2_{Y1|2} &= \frac{SQM(X1|X2)}{SQR(X_2)} \\ R^2_{Y3|12} &= \frac{SQM(X_3|X_1, X2)}{SQR(X_1, X_2)} \\ R^2_{Y34|12} &= \frac{SQM(X_3, X_4|X_1, X2)}{SQR(X_1, X_2)} \end{aligned}$$

7.2.5 Teste F Parcial como Perda de Ajuste

- O teste F parcial foi definido em função da SQ Extra, mas também pode ser definido em termos de Coeficiente de Determinação Parcial.
- Voltando ao exemplo de efeitos bióticos e abióticos, a hipótese

$$H_0 : \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0 | \, (\beta_1 \neq 0, \beta_2 \neq 0, \beta_3 \neq 0)$$

pode ser testada pela estatística F definida a partir dos coeficientes de determinação:

$$\begin{aligned} F &= \frac{[SQR(X_1, X_2, X_3) - SQR(X_1, X_2, X_3, X_4, X_5, X_6, X_7)] \, /4}{SQR(X_1, X_2, X_3, X_4, X_5, X_6, X_7)/(n-8)} \left[\frac{SQT}{SQT}\right] \\ &= \frac{\{[SQR(X_1, X_2, X_3)/SQT] - [SQR(X_1, X_2, X_3, X_4, X_5, X_6, X_7)/SQT]\} \, /4}{[SQR(X_1, X_2, X_3, X_4, X_5, X_6, X_7)/SQT]/(n-8)} \\ &= \frac{\{[1 - R^2_{Y123}] - [1 - R^2_{Y1234567}]\} \, /4}{(1 - R^2_{Y1234567})/(n-8)} \\ &= \frac{(R^2_{Y1234567} - R^2_{Y123})/4}{(1 - R^2_{Y1234567})/(n-8)} \end{aligned}$$

- Essa última expressão sugere o que teste F parcial pode ser visto como um teste da importância da **perda de ajuste** ocasionada no modelo com as 7 variáveis preditoras, quando as variáveis X_4, X_5, X_6 e X_7 são excluídas do modelo.
- Portanto, uma forma geral para o teste F parcial é a seguinte:
  - R^2 é o coeficiente de determinação do modelo com p variáveis preditoras.
  - R^2_* é o coeficiente de determinação do modelo que **exclui** um grupo de k variáveis preditoras ($k < p$).
  - A significância da **perda de ajuste** ocasionada pela exclusão pode ser testada pela estatística

$$F = \frac{(R^2 - R^2_*)/k}{(1 - R^2)/(n - p - 1)}$$

7.3 Teste de Hipóteses como Restrições Lineares

7.3.1 Combinação Linear dos Coeficientes de Regressão

- Já foi visto como uma hipótese sobre um coeficiente de regressão no modelo linear múltiplo pode ser testada através do teste t:

$$\begin{aligned} H_0 &: \beta_k = \beta_0 \\ H_\alpha &: \beta_k \neq \beta_0 \end{aligned}$$

$$t = \frac{b_k - \beta_0}{\sqrt{\widehat{\boldsymbol{Var}}\{b_k\}}}$$

  onde β_0 é uma constante hipotetizada e, sob a hipótese nula, a estatística t tem distribuição t de Student com $n - p - 1$ graus de liberdade.

- Suponha agora que, num modelo com duas variáveis preditoras, deseja-se tesar a hipótese nula

$$H_0 : \beta_1 = \beta_2 \qquad \Rightarrow \qquad H_0 : \beta_1 - \beta_2 = 0$$

  contra a hipótese alternativa

$$H_\alpha : \beta_1 \neq \beta_2.$$

- Note que essas são hipóteses sobre uma "*combinação linear* dos coeficientes de regressão do modelo.

- Essa combinação linear pode ser tomada como um *parâmetro*

$$\theta = \beta_1 - \beta_2$$

  e as hipóteses acima tornam-se equivalentes às hipóteses

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_\alpha &: \theta \neq \theta_0 \end{aligned}$$

  onde hipotetizamos que $\theta_0 = 0$.

- Para testarmos essa hipótese via teste t, necessitamos de uma estimativa do parâmetro

$$\widehat{\theta} = b_1 - b_2,$$

  onde b_1 e b_2 são as EQM (ou EMV).

Necessitamos também da variância da estimativa

$$\begin{aligned}\widehat{\boldsymbol{Var}}\{\widehat{\theta}\} &= \widehat{\boldsymbol{Var}}\{b_1\} + \widehat{\boldsymbol{Var}}\{b_2\} + 2\widehat{\boldsymbol{Cov}}\{b_1, b_2\} \\ &= QMR\,[(\mathbf{X}'\mathbf{X})^{-1}]_{22} + QMR\,[(\mathbf{X}'\mathbf{X})^{-1}]_{33} + QMR\,[(\mathbf{X}'\mathbf{X})^{-1}]_{23}\end{aligned}$$

- Sendo $\widehat{\theta}$ uma *combinação linear* das EQM b_1 e b_2, ela também terá distribuição Gaussiana de forma que a estatística

$$t = \frac{\widehat{\theta} - \theta_0}{\sqrt{\widehat{\boldsymbol{Var}}\{\widehat{\theta}\}}}$$

terá, sob H_0, distribuição t de Student com $n - p - 1$ graus de liberade.

7.3.2 Restrições Lineares

- A hipótese $H_0 : \beta_1 = \beta_2$ impõe uma *restrição linear* sobre o modelo original, reduzindo-o a

$$\begin{aligned}y_i &= \beta_0 + \beta_1 x_{i1} + \beta_1 x_{i2} + \varepsilon \\ &= \beta_0 + \beta_1 (x_{i1} + x_{i2}) + \varepsilon\end{aligned}$$

- Essa restrição linear pode ser representada na forma de uma combinação linear de todos os coeficientes de regressão do modelo original

$$H_0 : 0\beta_0 + \beta_1 - \beta_2 = 0,$$

que na forma matricial toma a forma

$$\begin{aligned}H_0 &: \mathbf{r}'\boldsymbol{\beta} = 0 \\ H_0 &: \begin{bmatrix} 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = 0\end{aligned}$$

- Na sua forma mais geral, uma restrição linear é apresentada na forma

$$H_0 : \mathbf{r}'\boldsymbol{\beta} = \theta_0,$$

onde θ_0 é uma constante hipotetizada.

- Alguns exemplos de restrições lineares num modelo linear múltiplo com p variáveis preditoras:

$$H_0 : \beta_k = 0 \quad \Rightarrow \quad \begin{array}{l} \mathbf{r}' = \left[\begin{array}{ccccccccc} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{array} \right] \\ \theta_0 = 0 \end{array}$$

$$H_0 : \beta_j = \beta_k \quad \Rightarrow \quad \begin{array}{l} \mathbf{r}' = \left[\begin{array}{ccccccccccc} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \end{array} \right] \\ \theta_0 = 0 \end{array}$$

$$H_0 : \beta_2 + \beta_3 - 2\beta_4 = 5 \quad \Rightarrow \quad \begin{array}{l} \mathbf{r}' = \left[\begin{array}{ccccccccc} 0 & 0 & 1 & 1 & -2 & 0 & \cdots & 0 \end{array} \right] \\ \theta_0 = 5 \end{array}$$

- Também é possível testar um conjunto de m restrições lineares simultaneamente na forma da hipótese

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \boldsymbol{\theta}_0,$$

onde $\mathbf{R}$ é uma matrix de constantes com m linhas e $p+1$ colunas e $\boldsymbol{\theta}_0$ é um vetor de constantes de comprimento m.

- Por exemplo, num modelo com seis variáveis preditoras, deseja-se testar simultaneamente as restrições

$$H_0 : \beta_2 + \beta_3 = 2;\ \beta_4 + \beta_6 = 0;\ \beta_5 - \beta_6 = 3$$

contra a alternativa

$$H_\alpha : \text{pelo menos uma das restrições acima não ocorre.}$$

A matrix e o vetor de constantes se tornam

$$\mathbf{R} = \left[\begin{array}{cccccc} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{array} \right] \qquad \boldsymbol{\theta}_0 = \left[\begin{array}{c} 2 \\ 0 \\ 3 \end{array} \right].$$

7.3.3 Teste de Wald

- O teste t pode ser generalizado na forma de um *teste de discrepância* para testar qualquer conjunto de restrições lineares.

- A *"discrepância"* para a hipótese

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \boldsymbol{\theta}_0$$

é definida como

$$\mathbf{m} = \mathbf{Rb} - \boldsymbol{\theta}_0,$$

onde $\mathbf{b}$ é o vetor das EQM (ou EMV).

- A variância desse vetor é dada por

$$\boldsymbol{Var}\{\mathbf{m}\} = \boldsymbol{Var}\{\mathbf{Rb} - \boldsymbol{\theta}_0\} = \boldsymbol{Var}\{\mathbf{Rb}\} = \mathbf{R}\boldsymbol{Var}\{\mathbf{b}\}\mathbf{R}' = \sigma^2\ \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'.$$

- Sob H_0, a *estatística de Wald*

$$W = \mathbf{m}'[\boldsymbol{Var}\{\mathbf{m}\}]^{-1}\mathbf{m} = (\mathbf{Rb} - \boldsymbol{\theta}_0)' \left[\sigma^2\ \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right]^{-1} (\mathbf{Rb} - \boldsymbol{\theta}_0)$$

tem distribuição Qui-quadrado com m graus de liberdade.

- A estatística de Wald, entretanto, não é *testável* pois a variância do erro σ^2 é desconhecida.

- Já foi visto uma outra estatística que envolve a variância do erro e tem distribuição Qui-quadrado

$$\frac{(n-p-1)\,QMR}{\sigma^2},$$

que possui $(n-p-1)$ graus de liberdade.

- A razão dessas duas estatísticas Qui-quadrado gera uma estatística que pode ser testada

$$\begin{aligned} F &= \frac{\left[\mathbf{m}'[\boldsymbol{Var}\{\mathbf{m}\}]^{-1}\mathbf{m}\right]\ /\, m}{\{[(n-p-1)\,QMR]\ /\,\sigma^2\}\ /\,(n-p-1)} \\ &= \frac{\left[(\mathbf{Rb} - \boldsymbol{\theta}_0)'[\sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \boldsymbol{\theta}_0)\right]\ /\, m}{\{[(n-p-1)\,QMR]\ /\,\sigma^2\}\ /\,(n-p-1)} \end{aligned}$$

$$F = \frac{(\mathbf{Rb} - \boldsymbol{\theta}_0)'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \boldsymbol{\theta}_0)}{m\,QMR}$$

- Sob H_0, essa estatística assume a forma

$$F = \frac{\chi^2[m]/m}{\chi^2[n-p-1]/(n-p-1)}$$

e, portanto, tem distribuição F com m graus de liberdade no numerador e $(n-p-1)$ graus de liberdade no denominador.

- O teste de Waldo pode ser apresentado sob forma ligeiramente diferente.

Sob H_0, temos que

$$\mathbf{R}\boldsymbol{\beta} = \boldsymbol{\theta}_0 \quad \Rightarrow \quad \mathbf{Rb} - \boldsymbol{\theta}_0 = \mathbf{Rb} - \mathbf{R}\boldsymbol{\beta} = \mathbf{R}(\mathbf{b} - \boldsymbol{\beta}),$$

de modo que estatística fica

$$F = \frac{[\mathbf{R}(\mathbf{b} - \boldsymbol{\beta})]'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}[\mathbf{R}(\mathbf{b} - \boldsymbol{\beta})]}{m\,QMR}$$

7.3.4 Exemplo: Modelagem da Produção em *E. grandis*

- Revisitando o exemplo da modelagem da produção de floresta plantada de *Eucalyptus grandis* consideremos a questão de quantificar a *influência relativa* de cada variável preditora sobre a produção.

Qual variável preditora seria mais influente?

- Sendo a produção ($\log$(produção)) modelada em função da idade ($1/$idade), do sítio e da área basal (g), um pesquisador acredita que
  - idade e sítio têm a mesma influência relativa, pois são variáveis medidas em escalas lineares: ano e metro, respectivamente;
  - área basal tem o dobro de influência relativa de idade e sitio, pois é uma variável medida em escala bidimensional: m^2/ha.

- Para medir a influência relativa não se pode utilizar o modelo linear ajustado diretamente sobre as variáveis, pois os coeficientes de regressão estão nas unidades de medição:

$Y =$ $\log$(produção) em $\log(m^3 ha^{-1}) = 3\log(m) - log(ha)$:

$$\beta_0 : 3\log(m) - log(ha).$$

$X_1 =$ $1/$(idade) em ano^{-1}:

$$\beta_1 : \frac{3\log(m) - log(ha)}{ano^{-1}} = (3\log(m) - log(ha))ano$$

$X_2 =$ sítio em m:

$$\beta_2 : \frac{3\log(m) - log(ha)}{m}$$

$X_3 =$ g em m^2ha^{-1}:

$$\beta_3 : \frac{3\log(m) - log(ha)}{m^2ha^{-1}} = \frac{ha(3\log(m) - log(ha))}{m^2}$$

- Para contornar esse problema é necessário **padronizar** as variáveis e ajustar o modelo sobre as variáveis padronizadas:

$$y_i^* = \frac{y_i - \overline{y}}{\sqrt{\widehat{\boldsymbol{Var}}\{y\}}} \quad x_{i1}^* = \frac{x_{i1} - \overline{x}_1}{\sqrt{\widehat{\boldsymbol{Var}}\{x_1\}}} \quad x_{i2}^* = \frac{x_{i2} - \overline{x}_2}{\sqrt{\widehat{\boldsymbol{Var}}\{x_2\}}} \quad x_{i3}^* = \frac{x_{i3} - \overline{x}_3}{\sqrt{\widehat{\boldsymbol{Var}}\{x_3\}}}$$

- O modelo linear com as variáveis padronizadas fica

$$y_i^* = \beta_1 x_{i1}^* + \beta_2 x_{i2}^* + \beta_3 x_{i3}^* + \varepsilon.$$

  Note que nesse modelo temos:

  – A média de todas as variáveis é nula, logo, o intercepto desaparece, pois a superfície de regressão passa na origem:

$$(\overline{y^*}, \overline{x_1^*}, \overline{x_2^*}, \overline{x_3^*}) = (0, 0, 0, 0).$$

  – A variância de todas as variáveis é unitária (1).

  – Os coeficientes de regressão não têm unidades, pois as variáveis não números "puros".

- O ajuste desse modelo aos dados de *E. grandis* em primeira rotação gera, em essência, os mesmos resultados já vistos, pois a padronização só altera o valor numérico dos coeficientes de regressão:

Fontes de Variação	**Graus de Liberdade**	**Soma de Quadrados**	**Quadrado Médio**	**Estatística F**	**Valor-p**
Modelo	3	56.94693	18.98231	527.29	$< 2.2 \times 10^{-16}$
X_1	1	38.465	38.465	1067.91	$< 2.2 \times 10^{-16}$
X_2	1	11.064	11.064	307.17	$< 2.2 \times 10^{-16}$
X_3	1	7.418	7.418	205.96	$< 2.2 \times 10^{-16}$
Resíduo	57	2.053	0.036		
Total	59	59.000			

COEFICIENTES DE REGRESSÃO

	Estimativa	Erro Padrão
β_1	-0.20773	0.04067
β_2	0.28770	0.05020
β_2	0.57465	0.04004

- As hipóteses levantadas pelo pesquisador implicam nas seguintes restrições lineares sobre o modelo

$$H_0 : \beta_1 + \beta_2 = 0; \qquad 2\beta_1 + \beta_3 = 0; \qquad 2\beta_2 - \beta_3 = 0.$$

que se traduzem nos seguintes matrix e vetor de constantes

$$\mathbf{R} = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \\ 0 & 2 & -1 \end{bmatrix} \qquad \boldsymbol{\theta}_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

(Note o sinal negativa para estimativa de β_1.)

- O teste de Wald resulta na estatística $F = 1.3048$, enquanto que o quantil 95% ($\alpha = 0,05$) da distribuição F com 3 graus de liberdade no numerador e 57 graus de liberdade no denominador é 2.766438.

Conclusão: não se rejeita H_0, i.e., as restrições lineares parecem ser apropriadas.

7.3.5 Revendo o Teste t dos Coeficientes de Regressão

- O teste t dos coeficientes de regressão é um teste padrão utilizado freqüentemente pelos analista e, geralmente, faz parte dos resultados apresentados pela maioria dos softwares estatísticos que realizam análise de regressão.
- Num modelo com p variáveis preditoras teremos p testes t para os coeficientes de regressão, excluindo-se o coeficientes β_0 que geralmente não tem interpretação relevante.
- Uma vez que no modelo linear múltiplo temos diversas variáveis preditoras correlacionadas surgem algumas questões:
  - Como esses testes devem ser interpretados?
  - Eles são redundantes?
  - Qual a relação desses teste com o teste F seqüencial?

- Já foi apresentado que, num modelo linera com p variáveis preditoras, a hipótese sobre um único parâmetro

$$H_0 : \beta_k = 0$$

pode ser realizada na forma de uma restrição linear

$$H_0 : \mathbf{r}'\boldsymbol{\beta} = 0$$

onde os vetores são

$$\mathbf{r}' = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{k-1} \\ \beta_k \\ \beta_{k+1} \\ \vdots \\ \beta_p \end{bmatrix}$$

- Desenvolvendo o teste de Wald para essa restrição linear obtemos

$$\begin{aligned} F &= \frac{(\mathbf{r}'\mathbf{b} - 0)'[\mathbf{r}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{r}']^{-1}(\mathbf{r}'\mathbf{b} - 0)}{(1)\, QMR} \\ &= \frac{b_k \left[[(\mathbf{X}'\mathbf{X})^{-1}]_{kk}\right]^{-1} b_k}{QMR} = \frac{b_k^2}{QMR\left[[(\mathbf{X}'\mathbf{X})^{-1}]_{kk}\right]} \\ F &= \frac{b_k^2}{\widehat{\boldsymbol{Var}}\{b_k\}} \end{aligned}$$

- Sob H_0, essa estatística tem distribuição F com 1 grau de liberdade no numerador e $(n - p - 1)$ graus de liberdade no denominador, sendo idêntica à distribuição t de Student com $(n - p - 1)$ graus de liberade:

$$\sqrt{F} = t = \frac{b_k}{\sqrt{\widehat{\boldsymbol{Var}}\{b_k\}}}$$

- Concluí-se que o teste t de cada coeficiente de regressão é uma restrição linear do modelo, logo os p testes t dos coeficientes de um modelo linear múltiplo representa um conjunto de p restrições lineares.

- Esse conjunto de p restrições é, no entanto, **redundante**.

- A restrição $\beta_k = 0$ reduz o modelo original

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{k-1} x_{i(k-1)} + \beta_k x_{ik} + \beta_{k+1} x_{i(k+1)} + \cdots + \beta_p x_{ip} + \varepsilon$$

para o seguinte modelo

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{k-1} x_{i(k-1)} + \beta_{k+1} x_{i(k+1)} + \cdots + \beta_p x_{ip} + \varepsilon$$

- Dessa forma, a estatística F do teste de Wald pode ser construída de modo equivalente utilizando a soma de quadrados extra

$$F = \frac{SQM(X_k | X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_p)/1}{SQR/(n-p-1)}$$

- Essa forma de apresentação mostra que cada teste t dos coeficientes de regressão testa a contribuição resultante da inclusão da variável preditora correspondente ao coeficiente, **dado que todas as demais variáveis preditoras já estão no modelo**.

- O teste t é, portanto, a forma **mais redundante possível** de testar a importância de cada variável preditora.

- Num ajuste, se **todos** os coeficientes são estatisticamente significativos temos uma forte indicação de que o modelo é apropriado.

  Deve-se procurar sempre um modelo que tenha essa situação.

- Entretanto, freqüentemente num modelo com muitas variáveis preditoras correlacionadas alguns coeficientes não são estatisticamente significativos e, nesse caso, o teste t é uma ferramenta muito limitada para se descobrir quais as variáveis preditoras que devem fazer parte do modelo.

7.3.6 Exemplo: Modelagem da Produção em Floresta de *E. grandis*

- Os testes t dos coeficientes de regressão do modelo de produção:

Variável Preditora	**Estimativa do Coeficiente**	**Erro Padrão da Estimativa**	**Teste** t	**Valor-**p
Intercepto	3.180354	0.146258	21.745	$< 2.0 \times 10^{-6}$
1/Idade	-1.281376	0.253115	-5.062	4.79×10^{-6}
Sítio	0.034047	0.005994	5.680	5.01×10^{-7}
Área Basal	0.065357	0.004595	14.225	$< 2.0 \times 10^{-6}$

- Calculando o teste F parcial para cada variável preditora, assumindo que as demais estão no modelo temos:

Fontes de Variação	**Graus de Liberdade**	**Quadrado Médio**	**Estatística F**	
1/Idade | (Sítio, Área Basal)	1	0.15865	25.628	$= (-5.062)^2$
Sítio | (1/Idade, Área Basal)	1	0.19972	32.263	$= (5.680)^2$
Área Basal | (1/Idade, Sítio)	1	1.25257	202.34	$= (14.225)^2$
Resíduo	56	0.0062		

7.4 Teste de Mudança Estrutural do Modelo

- Uma questão freqüentemente encontrada na aplicação da regressão é se o valor dos coeficientes de um dado modelo são diferentes para dois subconjuntos dos dados.
- No exemplo do modelo de produção de *E. grandis*, o florestal pode estar interessado em saber se esse modelo ajustado para uma floresta em *segunda rotação* teria os mesmos valores para os coeficientes de regressão que o modelo de primeira rotação.
- Outro exemplo comum é saber se um mesmo modelo dendrométrico, como o volume da árvores em função do diâmetro e da altura, pode ser utilizado em diferentes propriedades ou regiões.
- Às vêzes os florestais se referem a esse problema como sendo o "*problema de identidade do modelo*".

7.4.1 Teste por Restrições Lineares

- Consideremos que o conjunto de dados foi dividido em dois subconjuntos, de forma que os dados da variável resposta passam a formar o vetor $\mathbf{y} = [\mathbf{y}_1\, \mathbf{y}_2]$ e a matrix de delineamento passa a ser formada por $\mathbf{X} = [\mathbf{X}_1\, \mathbf{X}_2]$.
- O modelo linear sem restrições para os dois subconjuntos de dados é

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}.$$

- As estimativas de quadrados mínimos para esse modelo são:

$$\begin{aligned}\mathbf{b} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \begin{bmatrix} (\mathbf{X_1'X_1})^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{X_2'X_2})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{X_1'y_1} \\ \mathbf{X_2'y_2} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}\end{aligned}$$

  que são as mesmas estimativas do modelo ajustado aos dois subconjuntos separadamente.

- A SQR do modelo com os dois subconjuntos pode, portanto, ser obtida adicionando-se as SQR dos modelos ajustados separadamente:

$$SQR = \mathbf{e}'\mathbf{e} = \mathbf{e_1'e_1} + \mathbf{e_2'e_2}.$$

- A hipótese nula de ausência de mudança estrutural é representada pela igualdade dos coeficientes de regressão:

$$H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$$

  que pode ser expressa na forma de restrições lineares

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \boldsymbol{\theta},$$

  onde $\mathbf{R} = [\mathbf{I} : -\mathbf{I}]$ e $\boldsymbol{\theta} = \mathbf{0}$.

7.4.2 Exemplo: Modelagem da Produção em Floresta de *E. grandis*

- O objetivo é verificar se há mudança estrutural no modelo entre primeira e segunda rotações.

- Foram observadas $n_1 = 60$ parcelas em primeira rotação e $n_2 = 53$ parcelas em segunda rotação.

- As estimativas dos coeficientes de regressão encontradas foram:

Variável	**Estimativas por Rotação**			
Preditora	**Primeira**		**Segunda**	
Intercepto	3.180354	(0.146258)	4.682942	(0.426937)
1/Idade	-1.281376	(0.253115)	-0.672882	(2.840966)
Sítio	0.034047	(0.005994)	0.009345	(0.011599)
Área Basal	0.065357	(0.004595)	0.037960	(0.005796)

Valores entre parênteses são os erros padrão das estimativas

- As estatística F é calculada por

$$F = \frac{(\mathbf{Rb})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb})}{4QMR} = 27.81945$$

enquanto que o valor tabelado da distribuição F para $\alpha = 0.05$ com

$\Rightarrow$ 4 graus de liberdade no numerador: número de restrições lineares;

$\Rightarrow$ $(60-4)+(53-4) = 105$ graus de liberdade no denominador;

é igual a $F = 2.45821$.

- **Conclusão:** há mudança estrutural no modelo de produção entre a primeira e segunda rotação.

7.4.3 Teste de Wald para Mudança Estrutural

- O teste de mudança estrutural através de restrições lineares tem uma forte pressuposição implícita que freqüentemente não é razoável:

  A variância dos erros é a mesma para os dois subconjuntos.

- Essa pressuposição fica implícita no fato de se ajustar um único modelo e se aplicar as restrições lineares sobre ele.

- No caso de grandes amostras, há uma alternativa que é valida tanto no caso das variâncias serem as mesmas, quanto no caso de serem diferentes.

- Lembremos que as EQM no modelo linear múltiplo de erros esféricos são também estimativas de máxima verossimilhança e, conseqüentemente, têm distribuição assintótica Gaussiana.

- Suponha que $\widehat{\boldsymbol{\theta}}_1$ e $\widehat{\boldsymbol{\theta}}_2$ sejam estimativas com distribuição Gaussiana dos mesmos p parâmetros baseadas em **amostras independentes**, tendo matrizes de covariância $\mathbf{V}_1$ e $\mathbf{V}_2$, respectivamente.

  Sob a hipótese nula

$$H_0 : \boldsymbol{E}\left\{\widehat{\boldsymbol{\theta}}_1\right\} = \boldsymbol{E}\left\{\widehat{\boldsymbol{\theta}}_2\right\} = \boldsymbol{\theta},$$

  temos

$$\boldsymbol{E}\left\{\widehat{\boldsymbol{\theta}}_1 - \widehat{\boldsymbol{\theta}}_2\right\} = \mathbf{0} \qquad \text{e} \qquad \boldsymbol{Var}\{\widehat{\boldsymbol{\theta}}_1 - \widehat{\boldsymbol{\theta}}_2\} = \mathbf{V}_1 + \mathbf{V}_2.$$

- Assim, a **estatística de Wald**

$$W = [\widehat{\boldsymbol{\theta}}_1 - \widehat{\boldsymbol{\theta}}_2]'(\mathbf{V_1} + \mathbf{V_2})^{-1}[\widehat{\boldsymbol{\theta}}_1 - \widehat{\boldsymbol{\theta}}_2]$$

  tem distribuição Qui-quadrado com p graus de liberdade.

- No caso de **grandes amostras independentes** é válido substituir $\mathbf{V}_1$ e $\mathbf{V}_2$ pelas suas respectivas estimativas de modo que a estatística W pode ser efetivamente utilizada para testar a hipótese nula.

7.4.4 Exemplo: Modelagem da Produção em Floresta de *E. grandis*

- No caso do modelo de produção seria razoável assumir a mesma variância do erro para primeira e segunda rotação?

- Os resultados mostram que:

$$\begin{aligned} QMR_1 &= 0.006190343 \\ QMR_2 &= 0.03104989 \end{aligned}$$

  sendo que a estatística de comparação das variâncias resulta em

$$F = \frac{0.03104989}{0.006190343} = 5.01586$$

  sendo que o valor tabela da distribuição F ($\alpha = 0.05$) para 49 por 56 graus de liberdade é 1.576409.

  Portanto, não é razoável assumir que as variâncias dos erros sejam iguais.

- Aplicando a estatística de Wald temos

$$W = [\mathbf{b_1} - \mathbf{b_2}]'(\mathbf{V_1} + \mathbf{V_2})^{-1}[\mathbf{b_1} - \mathbf{b_2}] = 140.3784$$

  enquanto que o quantil 95% ($\alpha = 0.05$) da distribuição Qui-quadrado com 4 graus de liberdade é 9.487729.

- **Conclusão:** mesmo considerando variâncias diferentes há mudança estrutural no modelo de produção entre a primeira e segunda rotação.

7.5 Procedimento Geral para Teste de Hipóteses

- A SQ Extra e as restrições lineares são duas abordagens para testar hipóteses no modelo linear múltiplo que não são equivalentes pois algumas hipóteses só podem ser testadas pela abordagem específica.
- Por exemplo, o teste F seqüencial não pode ser colocado na forma de restrições lineares, e restrições lineares que envolvem contrastes não nulos ($\boldsymbol{\theta} \neq \mathbf{0}$) geralmente não podem ser testadas via SQ Extra.
- Existe, entretanto, um procedimento geral que permite testar qualquer tipo de hipótese.

7.5.1 Modelo Completo versus Modelo Reduzido

- Qualquer teste de hipótese pode ser construido na forma de comparação de dois modelos:

  Modelo Completo: o modelo sem as restrições da hipótese nula sendo testada;

  Modelo Reduzido: o modelo no qual a hipótese nula é assumida como verdadeira.

- Cada um desses modelos terá uma Soma de Quadrados do Resíduo (SQR) que são estimadores da mesma variância σ^2 do erro.
- Assim, se a hipótese nula for verdadeira temos que

$$\begin{aligned} \boldsymbol{E}\{SQR(\text{completo})\} &= \sigma^2 \\ \boldsymbol{E}\{SQR(\text{reduzido})\} &= \sigma^2 \end{aligned}$$

  e, portanto, ambas somas de quadrado terão distribuição Qui-Quadrado com os respectivos graus de liberade

$$\begin{aligned} SQR(\text{completo}) &\sim \chi^2[n-p-1] \\ SQR(\text{reduzido}) &\sim \chi^2[n-p-1+k], \end{aligned}$$

  onde k é

  - o número de restrições lineares impostas ao modelo completo, ou
  - o número de variáveis preditoras excluídas do modelo completo.

- O modelo completo terá, geralmente, uma SQR menor pois envolve um número maior de coeficientes de regressão ou variáveis preditoras.

  Tecnicamente, a SQR do modelo reduzido será *no mínimo* igual a do modelo completo:

$$SQR(\text{completo}) \leq SQR(\text{reduzido}).$$

- Sob a hipótese nula, a diferença entre as somas de quadrado também terá distribuição Qui-Quadrado

$$SQR(\text{reduzido}) - SQR(\text{completo}) \quad \sim \quad \chi^2[k].$$

- A estatística para se testar a hipótese nula, nesse procedimento geral, é

$$F^* \quad = \quad \frac{[SQR(\text{reduzido}) - SQR(\text{completo})] \ / \ k}{SQR(\text{completo}) \ / \ (n-p-1)},$$

  que sob hipótese nula terá distribuição F com k por $(n-p-1)$ graus de liberdade.

7.5.2 Exemplos do Procedimento Geral

Teste F Parcial: para se testar a introdução de um conjunto de duas variáveis preditoras (X_4, X_5) num modelo que já possui três variáveis preditoras, basta comparar os modelos:

Modelo Completo:	$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \varepsilon_i$
Modelo Reduzido:	$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$

Teste F Seqüencial: para testar a indrodução de variáveis preditoras uma-a-uma basta definir uma série de modelos, a partir do modelo mais simples até o modelo mais completo:

Modelo 1:	$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i$
Modelo 2:	$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$
Modelo 3:	$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$
$\ldots$	$\ldots$
Modelo p:	$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_p x_{ip} + \varepsilon_i.$

Os testes são construídos seqüencialmente:

$$\begin{array}{lcll} H_0: & X_2|X_1 & \Rightarrow & \text{Modelo 2} \times \text{Modelo 1} \\ H_0: & X_3|(X_1,X_2) & \Rightarrow & \text{Modelo 3} \times \text{Modelo 2} \\ & \cdots & \cdots & \\ H_0: & X_p|(X_1,X_2,\ldots,X_{p-1}) & \Rightarrow & \text{Modelo p} \times \text{Modelo}(p-1) \end{array}$$

Teste t dos Coeficientes de Regressão: para testar a hipótese $H_0 : \beta_k = 0$, basta comparar os modelos:

- Modelo Completo:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{(k-1)} x_{i(k-1)} + \beta_k x_{ik} + \beta_{(k+1)} x_{i(k+1)} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

- Modelo Reduzido:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{(k-1)} x_{i(k-1)} + \beta_{(k+1)} x_{i(k+1)} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

Teste de Restrições Lineares: qualquer restrição linear pode ser testada definindo os modelos como

Modelo Completo: modelo *sem* as restrições lineares;

Modelo Reduzido: modelo *com* as restrições lineares.

Exemplo: para testar as restrições lineares:

$$H_0 : \beta_2 = -7; \quad \beta_3 = 10$$

temos:

$$\begin{array}{ll} \text{Modelo Completo:} & y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \\ \text{Modelo Reduzido:} & y_i + 7x_{i2} - 10x_{i3} = \beta_0 + \beta_1 x_{i1} + \varepsilon_i \end{array}$$

7.5.3 Aplicação do Critério de Informação de Akaike (AIC)

- Como o procedimento geral define o teste F em termos de modelo completo e reduzido, o AIC pode ser diretamente aplicado nesse procedimento:

$$\begin{aligned} AIC(M_{\text{COMPLETO}}) &= n \ln [SQR(\text{completo})] + 2k_{\text{completo}} \\ AIC(M_{\text{REDUZIDO}}) &= n \ln [SQR(\text{reduzido})] + 2k_{\text{reduzido}} \end{aligned}$$

onde n é o tamanho da amostra e k é o número de parâmetros no modelo.

- A partir desses AICs podemos definir a diferença de AIC:

$$\begin{aligned}\Delta_{\text{AIC}} &= AIC(M_{\text{REDUZIDO}}) - AIC(M_{\text{COMPLETO}}) \\ &= n\left[\ln\left(SQR(\text{reduzido})\right) - \ln\left(SQR(\text{completo})\right)\right] + 2\left[k_{\text{reduzido}} - k_{\text{completo}}\right]\end{aligned}$$

$$\Delta_{\text{AIC}} = n\ln\left[\frac{SQR(\text{reduzido})}{SQR(\text{completo})}\right] - 2\left[k_{\text{completo}} - k_{\text{reduzido}}\right]$$

- A *interpretação direta* dessa diferença geralmente utiliza um nível arbitrário para diferença da log-verossimilhança ($\ln(8)$):
  - Para $\Delta_{\text{AIC}} < 2$ ($\approx ln(8)$) os modelos completo e reduzidos podem ser considerados igualmente plausíveis;
  - Para $\Delta_{\text{AIC}} > 2$ ($\approx \ln(8)$) o modelo completo pode ser considerado superior ao reduzido.

7.6 Exercícios

7.6-1. Um modelo de índice de sítio freqüentemente utilizado por Florestais em florestas plantadas é o seguinte:

$$\log(\texttt{mhdom}) = \beta_0 + \beta_1 1/(\texttt{idade}) + \varepsilon$$

Com base nos resultados apresentados abaixo, realize o teste da mudança estrutural do modelo comparando as duas rotações:

(a) Pelo Teste das Restrições Lineares;

(b) Pelo Teste de Wald;

Dados:
QMR = 0.06227067, QMR_1 = 0.01490523, QMR_2 = 0.05390036;
$n_1 = 342$; $n_2 = 232$;

$$\mathbf{X_1'X_1} = \begin{bmatrix} 342.0000 & 101.52803 \\ 101.5280 & 37.52819 \end{bmatrix} \qquad \mathbf{X_2'X_2} = \begin{bmatrix} 232.0000 & 72.21240 \\ 72.2124 & 24.94308 \end{bmatrix}$$

$$\mathbf{b_1} = \begin{bmatrix} 3.560655 \\ -1.984775 \end{bmatrix} \qquad \mathbf{b_2} = \begin{bmatrix} 3.215512 \\ -2.042366 \end{bmatrix}$$

Discuta de que forma o procedimento geral para teste de hipóteses em modelos lineares poderia ser aplicado a esse problema.

CAPÍTULO 8

REGRESSÃO PADRONIZADA E ANÁLISE DE CAMINHAMENTO

8.1 Limitações da Regressão Linear Múltipla

O método de ajuste tradicional do modelo linear múltiplo apresenta algumas limitações quanto aos resultados obtidos que devem ser considerados:

Erros computacionais :

- O cálculo da inversa de $\mathbf{X}'\mathbf{X}$ é a principal fonte de erro ao se obter as estimativas de quadrados mínimos no modelo linear múltiplo.
- Os prodecimentos de inveresão da matrix pode gerar graves erros de arredondamento principalmente quando:
  - as variáveis preditoras são correlacionadas em um alto grau;
  - quando a **magnitude** da variáveis preditoras é muito diferente de uma para outra.
- Métodos modernos de obter a inversa de $\mathbf{X}'\mathbf{X}$, através de decomposição de Choloski e decomposição por autovalores, reduzem os problemas de cálculo.
- A transformação das variáveis preditoras para uma mesma magnitude também evita erros de arredontamento, pois todas os elementos da matriz $\mathbf{X}'\mathbf{X}$ se reduzem a valores entre -1 e +1.

Coeficientes de Regressão não são comparáveis: Outro problema na RLM tradicional é que a unidade dos coeficientes de regressão estimados dependem das unidades das respectivas variáveis preditoras.

Consequentemente, os coeficientes não são diretamente comparáveis.

No exemplo 7.3.4, na página 7-203, apresentamos uma transformação das variáveis preditoras que elimina as unidades de medidas dos coeficientes.

Essa tranformação é a base da regressão padronizada.

8.2 Regressão Padronizada

8.2.1 Transformação das Variáveis

- A regressão padronizada é uma regressão tradicional realizada com as variáveis tranformadas.
- A transformação é realizada em duas etapas:

  1. Padronização: cada observação de cada variável, incluindo a variável resposta, é subtraída da média amostral e divida pelo desvio padrão amostral:

$$Y: \qquad y_i^* = (y_i - \overline{y})/s_Y, \qquad s_Y = \sqrt{\frac{\sum(y_i - \overline{y})^2}{n-1}}$$

$$X_k: \qquad x_{ik}^* = (x_{ik} - \overline{x}_k)/s_k, \qquad s_k = \sqrt{\frac{\sum(x_{ik} - \overline{x}_k)^2}{n-1}}$$

  onde $k = 1, 2, \ldots, p$.

  2. Controle do Tamanho de Amostra: as variáveis padronizadas são multiplicadas por $1/\sqrt{n-1}$.

$$Z: \qquad z_i = \frac{1}{\sqrt{n-1}}\left(\frac{y_i - \overline{y}}{s_Y}\right) = \frac{y_i - \overline{y}}{\sqrt{\sum(y_i - \overline{y})^2}}$$

$$W_k: \qquad w_{ik} = \frac{1}{\sqrt{n-1}}\left(\frac{x_{ik} - \overline{x}_k}{s_k}\right) = \frac{x_{ik} - \overline{x}_k}{\sqrt{\sum(x_{ik} - \overline{x}_k)^2}}$$

- A transformação, ao subtrair a média e dividir pela raiz quadrada da soma de quadrados, implica, portanto, em tornar todas as variáveis com **média zero** e **variância** igual a $1/(n-1)$.

8.2.2 Os Coeficientes de Regressão

- O modelo linear múltiplo tradicional

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

  é ajustado na forma transformada:

$$z_i = \alpha_1 w_{i1} + \alpha_2 w_{i2} + \cdots + \alpha_p w_{ip} + \eta_i$$

- Como as variáveis transformadas têm média nula (zero) e o modelo de regressão ajustado sempre passa pelo ponto das médias, a regressão padronizada **sempre passa pela origem** e, portanto, não possui intercepto.

- Mas qual a relação entre os coeficientes de regressão na forma tradicional e na regressão padronizada?

$$\begin{aligned}
z_i &= \alpha_1 w_{i1} + \alpha_2 w_{i2} + \cdots + \alpha_p w_{ip} + \eta_i \\
\frac{(y_i - \overline{y})}{s_Y\sqrt{n-1}} &= \alpha_1 \frac{(x_{i1} - \overline{x}_1)}{s_1\sqrt{n-1}} + \alpha_2 \frac{(x_{i2} - \overline{x}_2)}{s_2\sqrt{n-1}} + \cdots + \alpha_p \frac{(x_{ip} - \overline{x}_p)}{s_p\sqrt{n-1}} + \eta_i \\
(y_i - \overline{y}) &= \alpha_1 \frac{s_Y}{s_1}(x_{i1} - \overline{x}_1) + \alpha_2 \frac{s_Y}{s_1}(x_{i2} - \overline{x}_2) + \cdots + \alpha_p \frac{s_Y}{s_p}(x_{ip} - \overline{x}_p) + s_y\sqrt{n-1}\,\eta_i \\
(y_i - \overline{y}) &= \beta_1(x_{i1} - \overline{x}_1) + \beta_2(x_{i2} - \overline{x}_2) + \cdots + \beta_p(x_{ip} - \overline{x}_p) + \varepsilon_i
\end{aligned}$$

- Portanto, a relação entre os coeficientes de regressão é

$$\begin{aligned}
\beta_k &= \left(\frac{s_Y}{s_k}\right)\alpha_k, \quad \text{para } k = 1, 2, \ldots, p; \text{ e} \\
\beta_0 &= \overline{y} - \beta_1\overline{x}_1 - \beta_2\overline{x}_2 - \cdots - \beta_p\overline{x}_p.
\end{aligned}$$

- Os coeficientes da regressão padronizada (α_k) são números puros, i.e., sem unidades de medida.

- Os coeficentes de regressão tradicionais são obtidos multiplicando os coeficientes da regressão padronizada pela razão dos desvios padrão da variável resposta (s_Y) e da respectiva variável preditora (s_k), resultando em coeficientes com a magnitude compatível com a variável preditora e a unidade de medida apropriada.

- Note também que o termo de erro ε_i é transformado:

$$\varepsilon_i = s_Y\sqrt{n-1}\,\eta_i \qquad \Rightarrow \qquad \eta_i = \frac{\varepsilon_i}{s_Y\sqrt{n-1}} = \frac{\varepsilon_i}{\sqrt{\sum(y_i - \overline{y})^2}}$$

$$\textbf{\textit{Var}}\{\varepsilon_i\} = \sigma^2 \qquad \Rightarrow \qquad \textbf{\textit{Var}}\{\eta_i\} = \frac{\sigma^2}{\sum(y_i - \overline{y})^2}$$

8.2.3 Matrizes na Regressão Padronizada

- O sistema de equações lineares da regressão ponderada

$$z_i = \alpha_1 w_{i1} + \alpha_2 w_{i2} + \cdots + \alpha_p w_{ip} + \eta_i$$

em termos matriciais é apresentado por

$$\mathbf{z} = \mathbf{W}\boldsymbol{\alpha} + \boldsymbol{\eta}$$

o qual conduz ao Sistema de Equações Normais na forma

$$\mathbf{W}'\mathbf{W}\boldsymbol{\alpha} = \mathbf{W}'\mathbf{z}.$$

- A **matrix de delineamento** é

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1p} \\ w_{21} & w_{22} & \cdots & w_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{np} \end{bmatrix}$$

e, portanto, a matrix simétrica $\mathbf{W}'\mathbf{W}$ tem o formato

$$\mathbf{W}'\mathbf{W} = \begin{bmatrix} \sum w_{i1}^2 & \sum w_{i1}w_{i2} & \cdots & \sum w_{i1}w_{ip} \\ \sum w_{i2}w_{i1} & \sum w_{i2}^2 & \cdots & \sum w_{i2}w_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ \sum w_{ip}w_{i1} & \sum w_{ip}w_{i2} & \cdots & \sum w_{ip}w_{ip} \end{bmatrix}.$$

- Mas os elementos dessa matrix têm significado especial.

Vejamos o que acontece com uma dada variável preditora k:

$$\sum w_{ik}^2 = \sum \left[\frac{x_{ik} - \overline{w}_k}{s_k\sqrt{n-1}} \right]^2 = \frac{1}{s_k^2} \sum \frac{(x_{ik} - \overline{w}_k)^2}{n-1} = 1$$

Vejamos o que acontece com duas variáveis preditoras k e l:

$$\begin{aligned} \sum w_{ik}w_{il} &= \sum \left(\frac{x_{ik} - \overline{x}_k}{s_k\sqrt{n-1}} \right) \left(\frac{x_{il} - \overline{x}_l}{s_l\sqrt{n-1}} \right) \\ &= \frac{1}{s_k s_l} \sum \frac{(x_{ik} - \overline{x}_k)(x_{il} - \overline{x}_l)}{n-1} \\ &= \frac{\widehat{\boldsymbol{Cov}}\{X_k, X_l\}}{\sqrt{\widehat{\boldsymbol{Var}}\{X_k\}\,\widehat{\boldsymbol{Var}}\{X_l\}}} \\ &= r_{kl} \end{aligned}$$

- A matrix simétrica $\mathbf{W}'\mathbf{W}$ é composta por:
  - elementos unitários (**1**) na diagonal principal, e
  - **coeficientes de correlação** entre as variáveis preditoras nas demais posições:

$$\mathbf{W}'\mathbf{W} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix}.$$

- Já o produto matricial $\mathbf{W}'\mathbf{z}$, possui a forma

$$\mathbf{W}'\mathbf{z} = \begin{bmatrix} \sum w_{i1} z_i \\ \sum w_{i2} z_i \\ \vdots \\ \sum w_{ip} z_i \end{bmatrix} = \begin{bmatrix} r_{y1} \\ r_{y2} \\ \vdots \\ r_{yp} \end{bmatrix}$$

  pois, cada elemento resulta em

$$\begin{aligned} \sum w_{ik} z_i &= \sum \left(\frac{x_{ik} - \overline{x}_k}{s_k \sqrt{n-1}} \right) \left(\frac{y_i - \overline{y}}{s_Y \sqrt{n-1}} \right) \\ &= \frac{1}{s_Y s_k} \sum \frac{(x_{ik} - \overline{x}_k)(y_i - \overline{y})}{n-1} \\ &= \frac{\widehat{\boldsymbol{Cov}}\{Y, X_k\}}{\sqrt{\widehat{\boldsymbol{Var}}\{Y\}\, \widehat{\boldsymbol{Var}}\{X_k\}}} \\ &= r_{yk}. \end{aligned}$$

- O Sistema de Equações Normais, na regressão padronizada, pode portanto ser apresentado na forma de matrix e vetor de correlações:

$$\begin{aligned} \mathbf{W}'\mathbf{W}\boldsymbol{\alpha} &= \mathbf{W}'\mathbf{z} \\ \mathbf{R}_{xx}\boldsymbol{\alpha} &= \mathbf{r}_{yx} \end{aligned}$$

  onde $\mathbf{R}_{xx}$ é a matrix de correlação entre as variáveis preditoras e $\mathbf{r}_{yx}$ é o vetor de correlação entre a variável resposta e as variáveis preditoras.

- As estimativas de quadrados mínimos são obtidas solucionando o sistema de Equações Normais:

$$\widehat{\boldsymbol{\alpha}} = \mathbf{R}_{xx}^{-1}\, \mathbf{r}_{yx}$$

8.2.4 Algumas Propriedades da Regressão Ponderada

Distribuição do Erro: o erro tem distribuição Gassiana:

$$\varepsilon_i \sim N(0, \sigma^2) \qquad \Rightarrow \qquad \eta_i \sim N\left(0, \frac{\sigma^2}{\sum(y_i - \overline{y})^2}\right).$$

Distribuição da Variável Reposta: se a variável resposta original (Y) tem distribuição Gaussiana com

- valor esperado $\boldsymbol{E}\{Y|\mathbf{X} = \mathbf{x}_k\} = \beta_0 + \sum_{j=1}^{p} \beta_j x_{kj}$; e
- variância $\boldsymbol{Var}\{Y|\mathbf{X} = \mathbf{x}_k\} = \sigma^2$,

a nova variável resposta Z tem distribuição Gaussiana com

- valor esperado $\boldsymbol{E}\{Z|\mathbf{W} = \mathbf{w}_k\} = \sum_{j=1}^{p} \alpha_j w_{kj}$; e
- variância $\boldsymbol{Var}\{Z|\mathbf{W} = \mathbf{w}_k\} = \sigma^2/\sqrt{\sum(y_i - \overline{y})^2}$.

Distribuição das Estimativas dos Coeficientes de Regressão: partindo da solução do sistema de Equações Normais:

$$\begin{aligned}
\boldsymbol{Var}\{\widehat{\boldsymbol{\alpha}}\} &= \boldsymbol{Var}\{\mathbf{R}_{xx}^{-1}\mathbf{W}'\mathbf{z}\} \\
&= \mathbf{R}_{xx}^{-1}\mathbf{W}' \left[\boldsymbol{Var}\{\mathbf{z}\}\right] \mathbf{W}\,\mathbf{R}_{xx}^{-1} \\
&= \mathbf{R}_{xx}^{-1}\mathbf{W}' \left[\frac{1}{n-1}\mathbf{I}\right] \mathbf{W}\,\mathbf{R}_{xx}^{-1} \\
&= \frac{1}{n-1}\,\mathbf{R}_{xx}^{-1}\mathbf{W}'\mathbf{W}\,\mathbf{R}_{xx}^{-1} \\
&= \frac{1}{n-1}\,\mathbf{R}_{xx}^{-1}.
\end{aligned}$$

Logo: $\widehat{\boldsymbol{\alpha}} \sim \mathbf{N}\left(\boldsymbol{\alpha},\, [1/(n-1)]\mathbf{R}_{xx}^{-1}\right)$.

Coeficiente de Determinação: como a regressão padronizada não altera a capacidade explicativa do modelo, o valor do coeficiente de determinação deve permanecer inalterado.

Mas, partindo da variância da variável respostas, obtemos:

$$\boldsymbol{Var}\{z_i\} = \frac{1}{n-1} = \frac{SQT_z}{n-1} \quad \Rightarrow \quad SQT_Z = 1,$$

onde SQT_Z indica a soma de quadrados total na regressão padronizada.

Aplicando esse resultado ao cálculo do coeficiente de determinação temos:

$$R^2 = \frac{SQM_Z}{SQT_Z} \quad \Rightarrow \quad SQM_Z = R^2.$$

Lembrando que a SQM é definida matricialmente por $[\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}]$, o coeficiente de determinação pode ser obtido pela expressão

$$\begin{aligned} R^2 &= \mathbf{z}'\mathbf{W}\left[\mathbf{R}_{xx}^{-1}\right]\mathbf{W}'\mathbf{z} \\ &= \mathbf{r}_{yx}^{'}\left[\mathbf{R}_{xx}^{-1}\right]\mathbf{r}_{yx}. \end{aligned}$$

Conclusão: na regressão padronizada a SQM é o próprio coeficiente de determinação e, conseqüentemente,

$$SQR_Z = 1 - R^2.$$

8.3 Análise de Caminhamento

8.3.1 Correlação e Causa: Esquemas Causais

- Existem vários *exemplos clássicos* em estatística que demostram que um alto grau de correlação entre duas variáveis não implica necessariamente na existência de uma relação de *causa-efeito* entre elas.
- Em termos técnicos, as análises estatísticas são neutras no que se refere a qual variável é causa e qual é efeito.
- A aplicação da regressão padronizada a um dado problema pode, no entanto, ser interpetada como *assumindo* um determinado *esquema causal* (esquema de causa-efeito).
- Voltemos ao Sistema de Equações Normais na regressão padronizada:

$$\mathbf{R}_{xx}\boldsymbol{\alpha} = \mathbf{r}_{yx}$$

$$\begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} r_{y1} \\ r_{y2} \\ \vdots \\ r_{yp} \end{bmatrix}.$$

- Abrindo o sistema de equações obtemos

$$\begin{array}{rcccccccc} r_{y1} & = & \alpha_1 & + & \alpha_2 r_{12} & + & \cdots & + & \alpha_p r_{1p} \\ r_{y2} & = & \alpha_1 r_{12} & + & \alpha_2 & + & \cdots & + & \alpha_p r_{2p} \\ \vdots & = & \vdots & + & \vdots & + & \vdots & + & \vdots \\ r_{yp} & = & \alpha_1 r_{1p} & + & \alpha_2 r_{2p} & + & \cdots & + & \alpha_p \end{array}$$

- Esse sistema sugere o seguinte esquema de causa e efeito (figura 8.1):
  - A variável resposta é o efeito.
  - As variáveis preditoras são as causas.
  - A primeira equação do sistema se refere ao efeito total de X_1 sobre Y que é quantificado pelo coeficiente de correlação r_{y1}, o qual é composto por:
    - efeito **direto** de X_1 sobre Y = α_1;
    - efeito **indireto** de X_1 sobre Y via X_2 = $\alpha_2 r_{12}$;
    - efeito **indireto** de X_1 sobre Y via X_3 = $\alpha_2 r_{13}$;
    - e assim sucessivamente até o efeito indireto via X_p.
  - Os efeitos são **aditivos**.
  - Interpreta-se de modo análogo as demais equações.

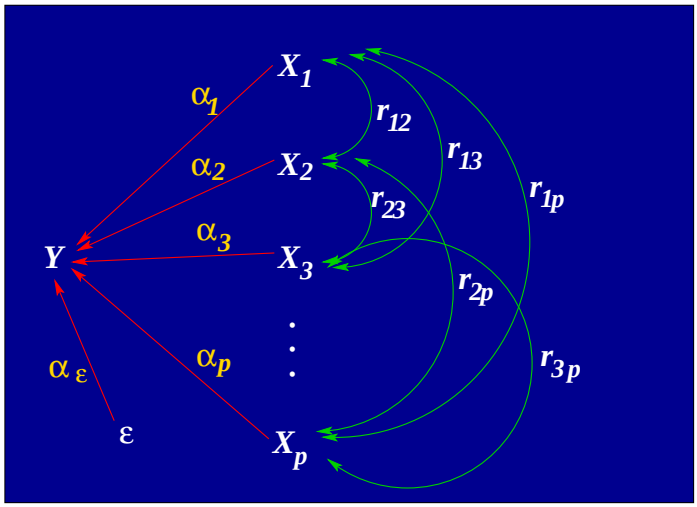


Figura 8.1: Esquema causal para regressão padronizada, assumindo p variáveis preditoras.

- Por fim, é necessário considerar que nem todo comportamento da variável resposta é explicado pelo esquema causal, existindo causas desconhecidos sobre o efeito estudado.

  No esquema da figura 8.1, as causas desconhecidas são designadas pelo termo ε, e a influência dessas causas pelo coeficiente α_ε.

  Como no modelo linear múltiplo, a variação não esplicada pelo é dada pela soma de quadrados do resíduo, podemos estimar esse coeficiente por

$$\widehat{\alpha}_\varepsilon = \sqrt{SQR_Z} = \sqrt{1 - R^2}.$$

8.3.2 Exemplo: Modelagem da Produção de *E. grandis*

- No exemplo da modelagem da produção de *E. grandis*, realizamos uma regressão padronizada para comparar a *magnitude* dos efeitos das variáveis preditoras sobre a produção, i.e., a variável resposta foi o logaritmo da produção.
- Os coeficientes encontrados foram:

VARIÁVEL	COEFICIENTE
1/Idade	-0.20773
Sítio	0.28770
Área Basal	0.57465

- Sendo uma regressão padronizada esses coeficientes representam o efeito direto das variáveis preditoras sobre a variável resposta.
- Para completarmos o equema causal (figura 8.1) necessitamos da matrix de correlações:

	$\log$(Produção)	1/Idade	Sítio	Área Basal
$\log$(Produção)	1.0000	-0.8074	0.9042	0.9351
1/Idade	-0.8074	1.0000	-0.7935	-0.6463
Sítio	0.9042	-0.7935	1.0000	0.7860
Área Basal	0.9351	-0.6463	0.7860	1.0000

- O esquema causal da figura 8.1 pode ser traduzido diretamente num sistema de equações (Equações Normais) utilizando-se a matrix de correlação:

$$\begin{array}{rcrcrcr} -0.8074 & = & \alpha_1 & + & -0.7935\alpha_2 & + & -0.6463\alpha_3 \\ 0.9042 & = & -0.7935\alpha_1 & + & \alpha_2 & + & 0.7860\alpha_3 \\ 0.9351 & = & -0.6463\alpha_1 & + & 0.7860\alpha_2 & + & \alpha_3 \end{array}$$

que em notação matricial se torna:

$$\begin{bmatrix} -0.8074 \\ 0.9042 \\ 0.9351 \end{bmatrix} = \begin{bmatrix} 1.0000 & -0.7935 & -0.6463 \\ -0.7935 & 1.0000 & 0.7860 \\ -0.6463 & 0.7860 & 1.0000 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

com solução

$$\begin{bmatrix} \widehat{\alpha_1} \\ \widehat{\alpha_2} \\ \widehat{\alpha_3} \end{bmatrix} = \begin{bmatrix} -0.20773 \\ 0.28770 \\ 0.57465 \end{bmatrix}.$$

- O coeficiente das causas desconhecidas é estimado a partir do coeficiente de determinação

$$R^2 = \begin{bmatrix} -0.8074 & 0.9042 & 0.9351 \end{bmatrix} \begin{bmatrix} 1.0000 & -0.7935 & -0.6463 \\ -0.7935 & 1.0000 & 0.7860 \\ -0.6463 & 0.7860 & 1.0000 \end{bmatrix}^{-1} \begin{bmatrix} -0.8074 \\ 0.9042 \\ 0.9351 \end{bmatrix}$$

$$R^2 = 0.9652$$

$$\widehat{\alpha}_\varepsilon = \sqrt{1 - R^2} = 0.1865$$

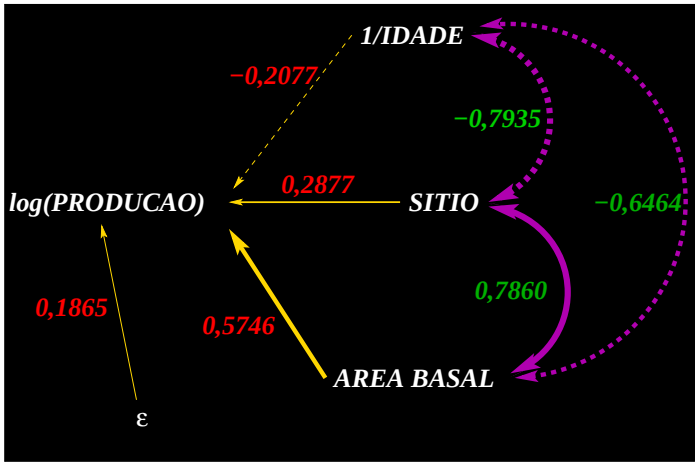


Figura 8.2: Esquema causal modelagem da produção de florestas de *E. grandis* em função da idade, sítio e área basal. Linhas contínuas indicam relação *positiva*, linhas tracejadas indicam relação *negativa*. A espessura da linha é proporcional ao grau do efeito/correlação.

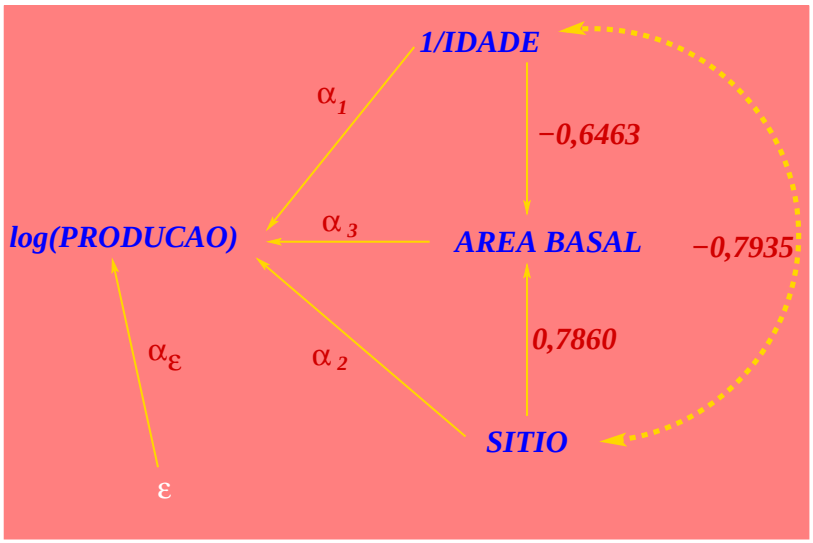


Figura 8.3: Esquema causal alternativop para a modelagem da produção de florestas de *E. grandis* em função da idade, sítio e área basal.

8.3.3 Análise de Caminhamento: Exemplo Aternativo na Modelagem da Produção de *E. grandis*

- Imaginemos um esquema causal alternativo para o problema da modelagem de produção, que segue a figura 8.3.
- Nesse esquema alternativa a área basal é efeito de sítio e idade que são, em última análise, as causas.
- Esse esquema não segue mais a estrutura da regressão padronizada, mas o que é geralmente chamado de *Análise de Caminhamento*.
- O sistema de equações nesse caso fica:

$$\begin{array}{rcccccc} r_{P,1/I} & = & \alpha_1 & + & r_{1/I,S}\alpha_2 & + & (r_{1/I,G} + r_{1/I,S}r_{S,G})\alpha_3 \\ r_{P,S} & = & r_{1/I,S}\alpha_1 & + & \alpha_2 & + & (r_{1/I,G}r_{1/I,G} + r_{S,G})\alpha_3 \\ r_{P,G} & = & 0\,\alpha_1 & + & 0\,\alpha_2 & + & \alpha_3 \end{array}$$

$$\begin{array}{rcccccc} -0.8074 & = & \alpha_1 & + & -0.8074\alpha_2 & + & -0.022609\alpha_3 \\ 0.9042 & = & -0.8074\alpha_1 & + & \alpha_2 & + & 1.298840\alpha_3 \\ 0.9351 & = & 0\,\alpha_1 & + & 0\,\alpha_2 & + & \alpha_3 \end{array}$$

- *Se o sistema for coerente*, a solução matricial será obtida da mesma forma que na regressão padronizada:

$$\begin{bmatrix} -0.8074 \\ 0.9042 \\ 0.9351 \end{bmatrix} = \begin{bmatrix} 1.0000 & -0.8074 & -0.022609 \\ -0.8074 & 1.0000 & 1.298840 \\ 0 & 0 & 1.0000 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

$$\begin{bmatrix} \widehat{\alpha_1} \\ \widehat{\alpha_2} \\ \widehat{\alpha_3} \end{bmatrix} = \begin{bmatrix} -2.978499 \\ -2.715185 \\ 0.935100 \end{bmatrix}$$

- Note que nesse esquema causal a influência direta do sítio (α_2) se torna negativa.
- O poder explicativo do esquema causal também é medido pelo R^2 e coeficiente associado ao termo do erro também pode ser obtido:

$$\begin{aligned} R^2 &= \mathbf{r}_{yx}^{'} \left[\mathbf{R}_{xx}^{-1}\right] \mathbf{r}_{yx}. \\ \widehat{\alpha}_{\varepsilon} &= \sqrt{1 - R^2}. \end{aligned}$$

$$R^2 = \begin{bmatrix} -0.8074 & 0.9042 & 0.9351 \end{bmatrix} \begin{bmatrix} 1.0000 & -0.8074 & -0.022609 \\ -0.8074 & 1.0000 & 1.298840 \\ 0 & 0 & 1.0000 \end{bmatrix}^{-1} \begin{bmatrix} -0.8074 \\ 0.9042 \\ 0.9351 \end{bmatrix}$$

$$R^2 = 0.8241815$$

$$\widehat{\alpha}_{\varepsilon} = \sqrt{1 - 0.8241815} = 0.1758185$$

- Note que o esquema casual alternativo tem um poder explicativo inferior à regressão padronizada.

CAPÍTULO 9
VARIÁVEIS PREDITORAS QUALITATIVAS

- Até agora vimos modelos de regressão onde as variáveis preditoras são todas **quantitativas**
- Entretanto, é possível incorporar variáveis **preditoras qualitativas** através do que matematicamente chamamos de **variáveis indicadoras**

9.1 Uma Variável Preditora Qualitativa

A variável indicadora é utilizada sempre na situação onde:

$$X_1 = I_1 = \begin{cases} 0 \text{ se a condição for A} \\ 1 \text{ se a condição for B} \end{cases}$$

9.1.1 Regra de Criação de Variáveis Indicadoras

Variáveis qualitativas com c classes são sempre representadas por $c - 1$ variáveis indicadoras.

9.1.2 Por que $c - 1$ Variáveis para c Classes?

- Suponha que tenhamos 9 observações, sendo que cada 3 observações foram feitas sob 3 diferentes tratamentos qualitativos.

- Se criarmos uma variável indicadora para cada classe a matrix $\mathbf{x}$ fica:

$$\mathbf{X} = \begin{bmatrix} C_1 & C_2 & C_3 & C_4 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

- A matrix $\mathbf{X}$ pode ser invertida ?

  Não, pois $C_1 = 1 \cdot C_2 + 1 \cdot C_3 + 1 \cdot C4$.

  $\mathbf{X}$ é singular pois suas colunas são linearmente independentes.

- E se utilizarmos apenas 2 variáveis para representar os 3 tratamentos diferentes?

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \Rightarrow y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

  Que deve ser interpretado como:

  β_0 resposta média para o tratamento 1;

  β_1 quantidade que Y é *aumentada* ou *diminuida* da resposta do tratamento 1 como efeito do tratamento 2.

  β_1 quantidade que Y é *aumentada* ou *diminuida* da resposta do tratamento 1 como efeito do tratamento 3.

9.1.3 Situação Hipotética: Efeito de Cipó no Crescimento

A variável resposta crescimento da árvore é modelada em função de uma variável qualitativa que envolve as classes:

- A = árvore sem cipós
- B = árvore com copa coberta por cipós em 50%
- C = árvore com copa completamente coberta de cipós.

O modelo de regressão fica:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

$$x_{i1} = \begin{cases} 1 \text{ se árvore coberta em 50\% por cipós;} \\ 0 \text{ se o contrário.} \end{cases}$$

$$x_{i2} = \begin{cases} 1 \text{ se árvore sem cipós;} \\ 0 \text{ se o contrário.} \end{cases}$$

Assim o modelo para as três cituações fica:

$$\begin{aligned} \text{Árvore coberta em 100\%} &\Rightarrow \boldsymbol{E}\{Y\} = \beta_0 \\ \text{Árvore coberta em 50\%} &\Rightarrow \boldsymbol{E}\{Y\} = \beta_0 + \beta_1 \\ \text{Árvore sem cipós} &\Rightarrow \boldsymbol{E}\{Y\} = \beta_0 + \beta_2 \end{aligned}$$

Portanto

- $\beta_0 =$ crescimento médio das árvores com 100% de cipó
- $\beta_1 =$ aumento (ou diminuição) sobre o crescimento médio das árvores com 100% de cipó referente a redução da cobertura de cipós para 50%.
- $\beta_2 =$ aumento (ou diminuição) sobre o crescimento médio das árvores com 100% de cipó referente a redução da cobertura de cipós para 0%.

Note que com este modelo poderíamos testar hipóteses do tipo:

- A cobertura de cipós reduz o crescimento.

$$H_0 : \beta_1 = \beta_2 = 0 \quad \text{contra} \quad H_\alpha : \beta_1 > 0 \text{ ou } \beta_2 > 0$$

- A redução pela metade da cobertura do cipó resultam num aumento mais que proporcional no crescimento.

$$\begin{aligned} H_0 : \beta_1 = \beta_0 \quad &\text{contra} \quad H_\alpha : \beta_1 > \beta_0 \\ H_0 : \beta_2 = 2\beta_1 \quad &\text{contra} \quad H_\alpha : \beta_2 > 2\beta_1 \end{aligned}$$

9.2 Modelos com Efeitos de Interação

Num modelo que utiliza variáveis quantitativas e qualitativas, é possível ocorrer interação entre elas.

- Suponhamos o seguinte modelo para estimar a produção de florestas de *Eucalyptus*:

$$\log(Y) = \beta_0 + \beta_1 X + \beta_2 I + \varepsilon$$

onde:

Y é a produção da floresta;

X é o inverso da idade da floresta;

e a variável indicadora é $I = \begin{cases} 1 \text{ se } 2^{\text{a}} \text{ rotação} \\ 0 \text{ se } 1^{\text{a}} \text{ rotação} \end{cases}$

- Neste modelo **sem interação** a rotação influencia apenas o intercepto da relação produção - altura das dominantes:

$$\begin{aligned} \text{Floresta em } 1^{\text{a}} \text{ rotação: } \boldsymbol{E}\{Y\} &= \beta_0 + \beta_1 X \\ \text{Floresta em } 2^{\text{a}} \text{ rotação: } \boldsymbol{E}\{Y\} &= (\beta_0 + \beta_2) + \beta_1 X \end{aligned}$$

- O modelo **com interação** seria:

$$\log(Y) = \beta_0 + \beta_1 X + \beta_2 I + \beta_3 (I \cdot X) + \varepsilon$$

Portanto, a rotação interagindo com a altura das dominantes influencia não somente o intercepto mas também a inclinação da curva de regressão:

$$\begin{aligned} 1^{\text{a}} \text{ rotação: } \boldsymbol{E}\{Y\} &= \beta_0 + \beta_1 X \\ 2^{\text{a}} \text{ rotação: } \boldsymbol{E}\{Y\} &= (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X \end{aligned}$$

9.2.1 Exemplo: Produção em Florestas de *E. saligna*

9.3 Modelos mais Complexos

O conceito de variável indicadora pode ser generalizado para situações mais complexas:

1. Variáveis indicadoras com mais de 2 classes.
2. Mais de uma variável qualitativa (var. indicadora).
3. Interações entre variáveis indicadoras.

9.4 Comparação de Duas ou Mais Funções de Regressão

Assumindo o modelo

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 I_i + \beta_3 (I_i \cdot x_i) + \varepsilon_i$$

onde $I = \begin{cases} 1 \text{ p/ situação A} \\ 0 \text{ p/ situação B} \end{cases}$

Podemos testar as hipóteses:

- As linhas de regressão são as mesmas em A e B:

$$F^*_{2,n-4} = \frac{SQM(I, I \cdot X|X)/2}{SQR(X, I, I \cdot X)/n-4}$$

- A inclinação das linhas de regressão são as mesmas em A e B:

$$F^*_{1,n-4} = \frac{SQM(I \cdot X|X, I)/1}{SQR(X, I, I \cdot X)/n-4}$$

9.4.1 Exemplo: Produção em Florestas de *E. saligna*

9.5 Considerações Finais

9.5.1 Por que não utilizar os códigos numéricos geralmente utilizados em variáveis qualitativas?

Suponha que desejemos modelar a diversidade de espécies na regeneração natural de uma floresta nativa após a colheita de madeira.

O impacto da colheita foi classificado em:

Classe	x_1
Baixo impacto	1
Médio impacto	2
Alto impacto	3

O modelo de regressão seria:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

o que leva nas seguintes espectativas quanto ao impacto da colheita

Classe	$\boldsymbol{E}\{Y\}$
Baixo impacto	$\beta_0 + \beta_1(1) = \beta_0 + \beta_1$
Médio impacto	$\beta_0 + \beta_1(2) = \beta_0 + 2\beta_1$
Alto impacto	$\beta_0 + \beta_1(3) = \beta_0 + 3\beta_1$

A implicação muito restritiva é:

$$\begin{aligned} &\boldsymbol{E}\{Y|\text{ médio impacto }\} - \boldsymbol{E}\{Y|\text{ baixo impacto }\} \\ &= \boldsymbol{E}\{Y|\text{ alto impacto }\} - \boldsymbol{E}\{Y|\text{ médio impacto }\} = \beta_1 \end{aligned}$$

Já utilizando as variáveis indicadoras

Classe	x_1	x_2
Baixo impacto	0	0
Médio impacto	1	0
Alto impacto	0	1

temos o modelo

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

onde β_1 mede

$$\boldsymbol{E}\{Y|\text{ médio impacto }\} - \boldsymbol{E}\{Y|\text{ baixo impacto }\}$$

enquanto que β_2 mede

$$\boldsymbol{E}\{Y|\text{ alto impacto }\} - \boldsymbol{E}\{Y|\text{ baixo impacto }\}$$

Sem a necessidade de assumirmos que este dois efeitos são iguais.

9.5.2 Outras Formas de Codificação de Variáveis Indicadoras

A codificação que utiliza os valores 0 e 1, aplicados a $c-1$ variáveis indicadoras representativas de c classes **não é a única** forma de codificação.

Existem outras formas, mas é importante sempre considerar os modelos resultantes para cada classe e as implicações em termos de estimativa dos parâmetros.

Uma forma frequentemente utilizada é:

$$I = \begin{cases} 1 & \text{se condição A} \\ -1 & \text{se condição B} \end{cases}$$

neste caso o modelo

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 I + \varepsilon_i$$

implica nos seguintes "sub-modelos":

$$\begin{aligned} \text{Média entre as situações A e B} &\Rightarrow \boldsymbol{E}\{Y\} = \beta_0 + \beta_1 X \\ \text{Situação A} &\Rightarrow \boldsymbol{E}\{Y\} = (\beta_0 + \beta_2) + \beta_1 X \\ \text{Situação B} &\Rightarrow \boldsymbol{E}\{Y\} = (\beta_0 - \beta_2) + \beta_1 X \end{aligned}$$

CAPÍTULO 10
DIAGNÓSTICO E MULTICOLINEARIDADE

O diagnóstico em regressão linear é a detecção de:

1. adequação das variáveis preditoras;
2. presença de observações discrepantes quanto a
   variávies preditoras (X), e
   variável resposta (Y);
3. presenção de observações influentes;
4. existência de multicolinearidade.

10.1 Adequação de uma Variável Preditora

A adequação de uma nova variável preditora, dado que já existe uma série de variáveis preditoras no modelo, é detectada através de **gráficos de regressão parcial**.

- Ajustar y em função das variáveis preditoras já no modelo e computar os resíduos $e\{y|x_1, \ldots, x_{k-1}\}$.
- Ajustar a nova variável preditora x_k em função das variáveis preditoras já no modelo e computar os resíduos $e\{x_k|x_1, \ldots, x_{k-1}\}$.
- O gráfico dos resíduos de y sobre os resíduos de x_k reflete a importância marginal de x_k.

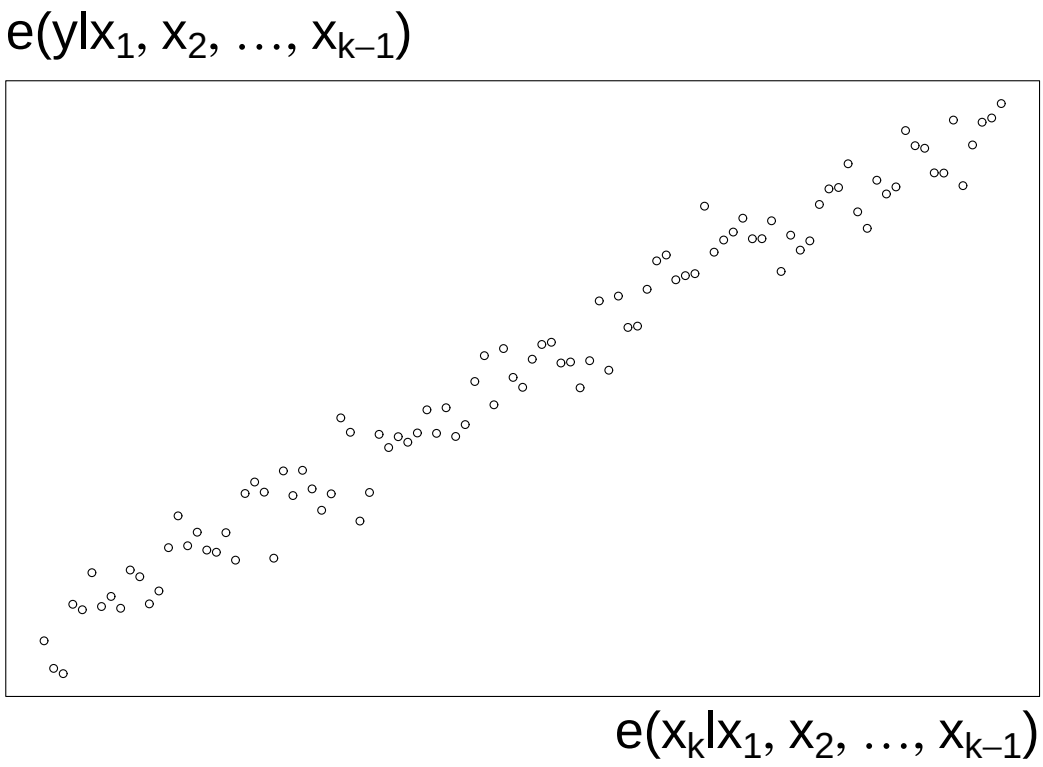


Figura 10.1: Exemplo de gráfico de regressão parcial quando a variável x_k acrescenta poder explicativo ao modelo.

10.2 Observações Discrepantes em Relação às Variáveis Preditoras

Existe diferentes maneiras de uma observação ser discrepante. seguindo a figura 10.2, vejamos:

- Nem todas observações discrepantes têm influência forte sobre a regressão:
  - observações 1 e 2: provavelmente não muito influrentes;
  - observação 3: provavelmente não muito influrente;
  - observações 4 e 5: provavelmente **são muito** influrentes.

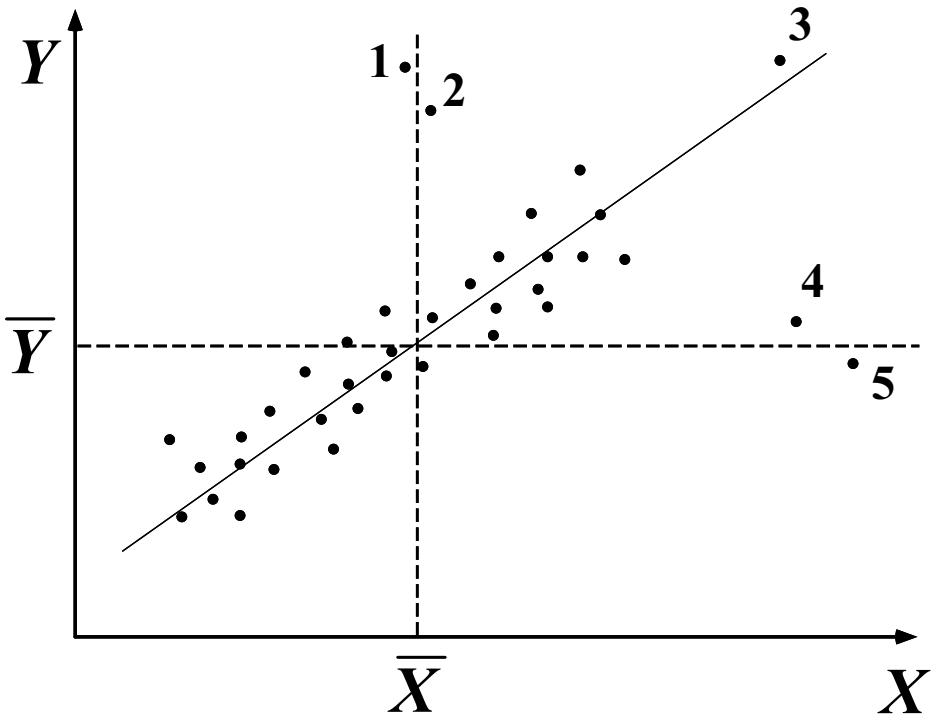


Figura 10.2: exemplo de gráfico de regressão mostrando observações discrepantes (numeradas de 1 a 5) segundo diferentes critérios.

- Quanto temos duas ou mais variáveis preditoras torna-se difícil detectar variáveis discrepantes utilizando apenas gráficos.

10.3 Observações Influentes em Relação às Variáveis Preditoras

10.3.1 Usando a Matrix de Projeção ("*Hat Matrix*")

- Como já foi visto, os resíduos podem ser vistos como uma projeção dos valores observados da variável resposta num hiperplano perpendicular ao hiperplano das variáveis preditoras

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$$

$$\begin{aligned} &= \mathbf{y} - \mathbf{Xb} \\ &= \mathbf{y} - \mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{y} - \mathbf{Py} \\ &= [\mathbf{I} - \mathbf{P}]\mathbf{y} \end{aligned}$$

- Foi mostrado também que

  $$\boldsymbol{Var}\{\mathbf{e}\} = \sigma^2[\mathbf{I} - \mathbf{P}]$$

  assim para um resíduo individual e_i temos

  $$\boldsymbol{Var}\{e_i\} = \sigma^2[1 - P_{ii}]$$

  onde $P_{ii} = \mathbf{x}_i[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{x}_i$ é o i$^{\text{ésimo}}$ elemento da diagonal de $\mathbf{P}$.

- O elemento diagonal P_{ii} é chamado de "alavancamento" ("*leverage*") da i$^{\text{ésima}}$ observação.

- Algumas propriedades o "alavancamento":

  – é um valor entre 0 e 1.
  Como $\mathbf{P}$ é uma matrix idempotente de posto $p + 1$ temos que

  $$\sum P_{ii} = p + 1, \qquad \text{portanto,} \qquad 0 \leq P_{ii} \leq 1.$$

  – Se o valor de P_{ii} é alto, então o valor da $\boldsymbol{Var}\{e_i\}$ é baixo.

- O alavancamento P_{ii} pode ser considerado como uma medida da distância da i$^{\text{ésima}}$ observação do **valor médio** das n observações de X_i.

- A figura 10.3 mostra o "alavancamento" de duas observações no caso de um modelo com duas variáveis preditoras.

- Se P_{ii} é **grande** temos:

  – a i$^{\text{ésima}}$ observação é influente;

  – esta observação exerce considerável influência (alavancamento = leverage) na determinação do valor esperado $\widehat{y}_i$.

- Há duas maneiras para se interpretar P_{ii}:

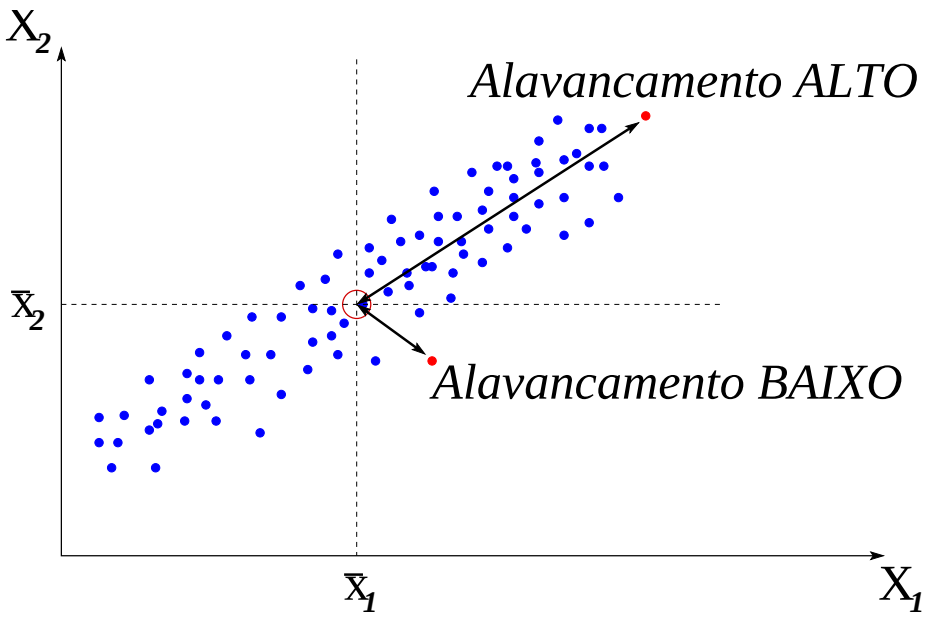


Figura 10.3: Exemplo de "alavancamento" de observações indivdiduais num modelo com duas variáveis preditoras.

1. Os valores esperados são $\widehat{\mathbf{y}} = \mathbf{Py}$, ou seja, cada $\widehat{y}_i$ é uma combinação linear dos valores observados y_i.
   Os P_{ii} são os **pesos** na determinação dos valores esperados.
   Conclusão: quanto maior P_{ii}, mais influente é a observação i.

2. Como ***Var***$\{e_i\} = \sigma^2(1 - P_{ii})$, temos que quanto maior P_{ii}, tanto menor ***Var***$\{e_i\}$ e, conseqüentemente, mais próximo o valor ajustado ($\widehat{y}_i$) estará do valor observado (y_i).

10.3.2 Regras Práticas

Há duas regras práticas para detectar observações influentes:

1. Se $P_{ii} > 2\overline{P}_{ii}$, onde

$$\overline{P}_{ii} = \frac{\sum_{i=1}^{n} P_{ii}}{n} = \frac{p+1}{n},$$

isto é, se

$$P_{ii} \;>\; 2\left(\frac{p+1}{n}\right),$$

então P_{ii} é grande.

2. Existe um grande "salto" entre a magnitude da maioria dos P_{ii} e a magnitude de alguns poucos P_{ii}?

10.4 Observações Discrepantes em Relação a Variável Resposta

Já foi visto que os resíduos podem ser **padronizados** pela transformação:

$$\frac{e_i}{\sqrt{QMR}}$$

os quais podem ser utilizados para estudar a normalidade, pois se

$$\varepsilon_i \sim n(0,\sigma^2), \quad \text{então } \frac{\varepsilon_i}{\sigma} \sim n(0,1)$$

Assim, temos

$$\frac{e_i}{\sqrt{\widehat{\boldsymbol{Var}}\{e_i\}}} \sim t(n-p)$$

e podemos verificar na tabela t de Student se apenas 5% dos resíduos superam em valor absoluto a $t(0.975, n-p)$.

Como a variância do resíduo é

$$\boldsymbol{Var}\{e_i\} = \sigma^2\,(1-P_{ii})$$

utilizaremos como variância estimada

$$\widehat{\boldsymbol{Var}}\{e_i\} = QMR\,(1-P_{ii})$$

para computar os **resíduos Studentizados**:

$$e_i^* \;=\; \frac{e_i}{\sqrt{QMR\,(1-P_{ii})}}$$

utilizando a tabela t de Student podemos detectar observações discrepantes.

10.4.1 Resíduos Studentizados Cancelados

- Uma segunda variante dos resíduos Studentizados é medir o i$^{\text{ésimo}}$ resíduo como

$$d_i \;=\; y_i - \widehat{y}_{(i)}$$

onde $\widehat{y}_{(i)}$ é o valor estimado **excluindo a i$^{\text{ésima}}$ observação** na regressão.

portanto, d_i pode ser chamado do i$^{\text{ésimo}}$ **resíduo cancelado**.

- Note que

$$\widehat{\boldsymbol{Var}}\{d_i\} \;=\; QMR_{(i)}\left(1 + \mathbf{x}_i[\mathbf{X}_{(i)}{}'\mathbf{X}_{(i)}]^{-1}\mathbf{x}_i\right)$$

e o **resíduo Studentizado cancelado** será

$$d_i^* \;=\; \frac{d_i}{\sqrt{\widehat{\boldsymbol{Var}}\{d_i\}}} \sim t(n-p)$$

- **Note que:**
  - se a observação **não** é influente não haverá muita diferença entre e_i e d_i, ou entre e_i e d_i^*.
  - se a observação for muito influente haverá uma grande mudança de e_i para e_i^* ou d_i^*.

- Felizmente, os d_i^* podem ser calculados sem a realização de novas regressões:

$$d_i^* \;=\; e_i\left[\frac{n-p}{SQR(1-P_{ii}) - e_i^2}\right]^{1/2}$$

10.5 Observações Influentes em Relação à Variável Resposta

A partir dos resíduos deletados foram desenvolvidas outras medidas de influência baseadas no cancelamento das observações uma-a-uma.

DFFITS: mede a influência sobre os valores esperados.

$$(\text{DFFITS})_i \;=\; \frac{\widehat{y}_i - \widehat{y}_{i(i)}}{\sqrt{QMR_{(i)}P_{ii}}}$$

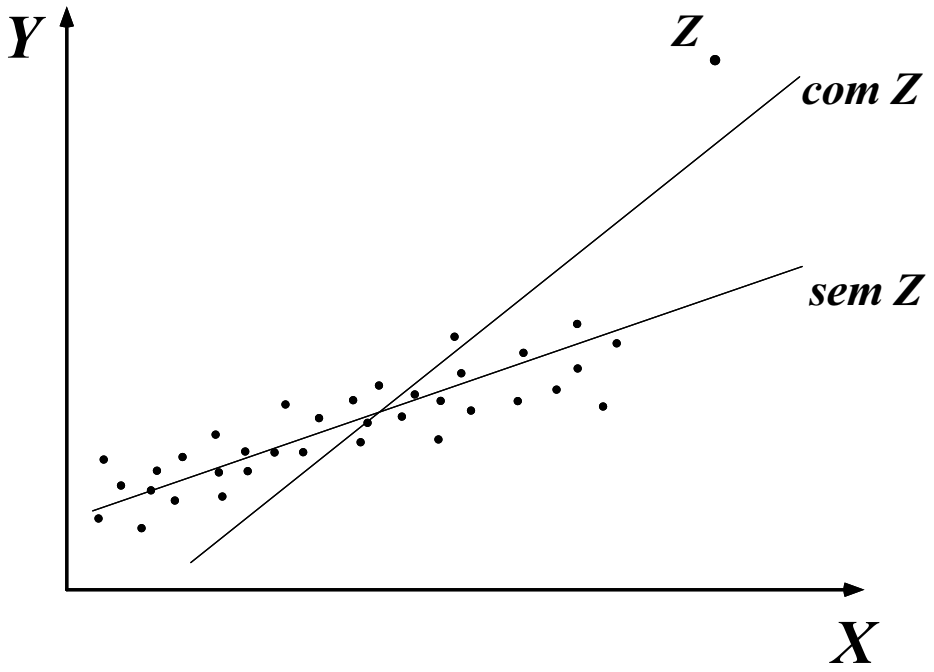


Figura 10.4: Exemplo de gráfico de regressão com observação influente, cuja influência pode ser detectado através do *resíduo deletado*.

ou seja, é a diferença entre o valor estimado com a presentça da i$^{\text{ésima}}$ observação ($\widehat{y}_i$) e o valor estimado quando esta observção é cancelada ($\widehat{y}_{i(i)}$).

Pode ser mostrado que

$$\begin{aligned}(\text{DFFITS})_i &= \left[\frac{n-p}{SQR(1-P_{ii})-e_i^2}\right]^{1/2}\left[\frac{P_{ii}}{1-P_{ii}}\right] \\ &= d_i^*\left[\frac{P_{ii}}{1-P_{ii}}\right]\end{aligned}$$

Regra prática: a observação é influente se

- DFFITS > 1, para conjuntos de dados pequenos ($n < 30$).
- DFFITS $> 2\sqrt{p/n}$, para conjuntos de dados grandes ($n \geq 30$).

DFBETAS: mede a influência sobre as estimativas dos coeficientes de regressão

$$(\text{DFBETAS})_{k(i)} = \frac{b_k - b_{k(i)}}{\sqrt{QMR_{(i)} \cdot C_{kk}}}$$

onde

- b_k estimativa de β_k com base nas n observações;
- $b_{k(i)}$ estimativa de β_k excluindo-se a i$^{\text{ésima}}$ observação ($n-1$ casos);
- C_{kk} é o k$^{\text{ésimo}}$ elemento da diagonal de $[\mathbf{X}'\mathbf{X}]^{-1}$;
- $\sqrt{QMR_{(i)} \cdot C_{kk}}$ é o erro padrão de b_k com base nos $n-1$ casos.

Regra prática: uma observação é influente se

- DFBETAS > 1, para conjuntos de dados pequenos ($n < 30$).
- DFBETAS $> 2\sqrt{n}$, para conjuntos de dados grandes ($n \geq 30$).

Distância de Cook: medida do impacto geral da i$^{\text{ésima}}$ observação sobre todas as estimativas dos coeficientes de regressão.

A região de confiança dos p coeficientes de regressão foi definida como

$$F(p, n-p) = \frac{(\mathbf{b} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\mathbf{b} - \boldsymbol{\beta})}{p \cdot QMR}$$

A distância de Cook tem relação próxima com esta região:

$$D_i = \frac{(\mathbf{b} - \mathbf{b}_{(i)})'\mathbf{X}'\mathbf{X}(\mathbf{b} - \mathbf{b}_{(i)})}{p \cdot QMR}$$

onde $\mathbf{b}_{(i)}$ é o vetor das estimativas dos coeficientes de regressão quando a i$^{\text{ésima}}$ observação é cancelada.

D_i **não** segue a distribuição F, mas

- Se $D_i =$ percentil $< 10 - 20 \Rightarrow p > 0.80 - 0.90 \Rightarrow$ **pequena** influência.
- Se $D_i =$ percentil $> 50 \Rightarrow p < 0.50 \Rightarrow$ **grande** influência.
- Note que para percentis entre 20-50 têm-se uma região "nebulosa".

Na prática, a distância de Cook é calculada por

$$D_i = \frac{e_i^2}{p \cdot QMR}\left[\frac{P_{ii}}{(1-P_{ii})^2}\right]$$

Quanto maior e_i ou P_{ii}, maior é D_i.

10.6 Medidas Remediadoras

- A análise de observações discrepantes e influentes é um componente essencial a uma boa análise de regressão,

  entretanto, ela exige uma boa capacidade de jugamento do analista.

- Para detectar medidas influentes temos:

  matrix $\mathbf{P}$ (valores P_{ii}) que são o "alavancamento";

  resíduos Studentizados cancelados d_i^*;

  distância de Cook D_i.

  Mas o que fazer ao detectarmos uma observação influente nem sempre é muito claro.

- Algumas observações influentes podem ser **observações erradas**,

  Mas uma observação perfeitamente normal também pode ser altamente influente!!

- Neste caso, algumas possiblidades são:

  – A amostragem foi insuficiente na vizinhaça da observação influente,
  $\Rightarrow$ voltar ao campo e coletar mais dados.

  – O modelo é inapropriado.

  – A forma atual das variáveis preditoras não é apropriada ou
  estão faltando variáveis preditoras importantes.

- Frequentemente, a influência de algumas observações poder ser reduzida por

  transformação das variáveis,

  mudança do processo de ajuste do modelo (por exemplo minimização dos desvios absolutos).

10.7 Multicolinearidade

Multicolinearidade = correlação entre as variáveis preditoras.

10.7.1 Efeitos da Multicolinearidade

Efeitos quando as variáveis preditoras não são correlacionadas :

1. As estimativas dos parâmetros para uma dada variável não dependem de quais outras variáveis estão no modelo.
2. A soma de quadrados extra para qualquer variável preditora é igual a SQM para a regressão linear simples com esta variável.

Efeito quando as variáveis preditoras são perfeitamente correlacionadas:

1. Existe um número **infinito** de funções de resposta, todas elas ajustando os dados igualmente bem.
2. Correlação perfeita não impede de se obter um **bom ajuste** aos dados.
3. Como existe infinitas funções de resposta com o mesmo ajuste, não se pode interpretar o conjunto dos coeficientes de regressão como refletindo o efeito das variáveis preditoras.

Efeitos da Multicolinearidade :

1. Nenhum efeito sobre a possibilidade de se obter um bom ajuste.
2. A variância amostral dos coeficientes de regressão é grande:
   — nenhum parâmetro é estatisticamente significativo (não se rejeita $H_0 : \beta_k = 0$);
   — mas o conjunto de variáveis preditoras é estatísticamente significativo (teste F do modelo é significativo).
3. Não é realista avaliar o efeito de uma variável preditora mantendo as demais constantes.
4. A soma de quadrados extra de uma variável preditora depende das demais variáveis presentes no modelo.
5. Testes t simultâneos são problemáticos;
   — eles podem aceitar o conjunto de hipóteses $H_0 : \beta_k = 0$, quando o teste F pela SQ-extra rejeitaria o mesmo conjunto de H_0's.

Lembre-se que o teste t é um teste marginal, análogo ao teste F pela SQ-extra assumindo que todas as demais variáveis preditoras já estão no modelo.

10.7.2 Diagnose Informal da Multicolinearidade

1. Grandes mudanças nas estimativas dos coeficientes quando variáveis são incluídas ou excluidas do modelo.
2. Testes não significativos para variáveis consideradas importantes.
3. Estimativas dos coeficientes de regressão **com sinal errado**.
4. Alta correlação (duas-a-duas) entre as variáveis preditoras.

   Matrix de correlação das variáveis preditoras é útil, mas limitada:

   - uma variável preditora pode ser função de duas ou mais outras variáveis preditoras,
   - correlações triplas, quádruplas, etc.

10.7.3 VIF - Fator de Inflação da Variância

Relembrando a regressão padronizada onde

$$\mathbf{b}' \;=\; [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{R}_{xx}^{-1}\mathbf{r}_{yx}$$

foi mostrado que

$$\boldsymbol{Var}\{\mathbf{b}'\} \;=\; (\sigma')^2\mathbf{R}_{xx}^{-1}$$

onde:

- $(\sigma')^2$ é a variância dos erros na regressão padronizada;
- $\mathbf{R}_{xx}^{-1}$ é a matrix de correlação entre as variáveis preditoras.

Note que:

$$\boldsymbol{Var}\{b'_k\} \;=\; (\sigma')^2\left[\mathbf{R}_{xx}^{-1}\right]_{kk}$$

onde $[\mathbf{R}_{xx}^{-1}]_{kk}$ é o k$^{\text{ésimo}}$ elemento da diagonal principal da matrix $\mathbf{R}_{xx}^{-1}$.

Este elemento é chamado de Fator de Inflação da Variância (VIF - Variance Inflaction Factor):

$$\text{VIF}_k \quad = \quad \left[\mathbf{R}_{xx}^{-1}\right]_{kk} = \frac{1}{1 - R_k^2} \qquad\qquad (k = 1, \ldots, p)$$

onde R_k^2 é o coeficiente de determinação da regressão de X_k sobre todas as outras $p - 1$ variáveis X.

Note que

- $\text{VIF}_k \geq 1$
- À medida que a relação entre X_k e as demais variáveis preditoras X tende à perfeição

  $\Rightarrow R_k^2 \to 1$

  $\Rightarrow \text{VIF}_k \to \infty$

  $\Rightarrow \boldsymbol{Var}\{b_k'\} \to \infty$

Regra prática: Se VIF > 10, então existe clara indicação de que a multicolinearidade está influenciando a precisão das estimativas de quadrados mínimos.

O **valor médio** dos VIF também é importante.

- Pode ser mostrado que

$$\boldsymbol{E}\left\{\sum_{k=1}^{p}(b_k' - \beta_k')^2\right\} \quad = \quad (\sigma')^2 \sum_{k=1}^{p} \text{VIF}_k$$

- Alto valor médio dos VIF

  $\Rightarrow$ grandes diferenças entre valores estimados e verdadeiros dos coeficientes da regressão padronizada

  $\Rightarrow$ **baixa precisão** geral na estimativa dos coeficientes de regressão.

TOLERÂNCIA: muintos softwares estatísticos definem o inverso do VIF como "limite de tolerância" abaixo do qual o modelo de regressão não é ajustado.

$$\text{TOLERANCIA} \quad = \quad \frac{1}{\text{VIF}_k} = 1 - R_k^2$$

MEDIDAS REMEDIADORAS: regressão de "culmieira" (ridge regression).

CAPÍTULO 11

SELEÇÃO DE VARIÁVEIS E MODELOS

11.1 Construindo um Modelo por Regressão

A construação de um modelo por regressão é um **processo interativo** no qual os conceitos e instrumentos vistos nesse curso são integrados

As fases na construção de modelos são:

1. Coleta e preparação dos dados,
2. Redução no número de variáveis preditoras,
3. Seleção do modelo e refinamento,
4. Validação do modelo.

11.1.1 Coleta e Preparação dos Dados

- Reduzir a "longa lista de medidas" de campo para um grupo operacional e eficente (em custo) de medidas.
- Eliminar as variáveis que:
  - não são fundamentais para o problema estudado;
  - são sujeitas a grandes erros de mensuração;
  - duplicam a informação contida em outras variávies.

- **Regra Prática:** 6-10 casos (classes) em cada variável.
- Editar os dados é uma **fase essencial**.

  No caso de "grandes bases de dados", é crítico o uso de um **gerenciador de banco de dados**, preferencialmente um banco de dados **relacional** (linguagem SQL).

11.1.2 Redução do Número de Variáveis Preditoras

- Por "tentatia-e-erro" decidir a forma funcional dos modelos.
- Selecionar um **bom** sub-conjunto de variáveis preditoras (X), incluindo qualquer transformação.

  Com dados **observacionais**, a seleção de um "bom" sub-conjunto e a determinação da forma funcional é geralmente o problema mais difícil na análise de regressão.
- Multicolinearidade pode ser um problema, especialmente quando sugere a exclusão de variáveis preditoras importantes.
- Outro problema típico: pequena amplitude de variação em muitas variáveis preditoras.
- Existe um grande número de abordagens computacionais para construção de modelos de regressão, **mas** o processo de construção de um **modelo útil** exigem um grande dose de "jugamento subjetivo".

11.2 "Todas-Regressões-Possíveis"

Objetivo: identificar um subconjunto pequeno de modelos alternativos (em geral 5 a 10).

Criétrios: vários critérios podem ser utilizados para comparar os modelos e gerar o melhor subconjunto:

- Coeficiente de Determinação (R_p^2),
- Quadrado Médio do Resíduos (QMR_p),
- Cp de Mallow (C_p),
- Soma de Quadrados de Predição ($PRESS_p$).

Notação:

- Número de variáveis preditoras potenciais nos dados = p.

  $\Rightarrow$ modelo com todas variáveis preditoras estima $p+1$ parâmetros.

- Subconjunto das variáveis preditoras = k.

- CRITÉRIO(k) = valor do critério para o modelo com k variáveis preditoras, i.e., estima $k+1$ parâmetros (inclue o intercepto).

- $0 \leq k \leq p$

11.2.1 Critério $R^2(k)$

- Lembrar que:

$$R^2(k) \;=\; \frac{SQM(k)}{SQT} = 1 - \frac{SQR(k)}{SQT}$$

- Note que o denominador é **constante** (SQT) para todas as regressões possíveis e que SQR **nunca pode crescer** quando uma nova variável é incluida no modelo.

- O objetivo é encontrar o ponto onde o acréscimo de mais uma variável não resulta num **aumento substancial** de R^2.

- Para de adicionar variáveis quando o ganho marginal no R^2 é pequeno.

11.2.2 Critério $QMR(k)$ ou R^2_α

- Lembrar que o R^2 ajustado (R^2_α) leva em consideração o número de variáveis preditoras no modelo:

$$R^2_\alpha = 1 - \left(\frac{n-1}{p-1}\right)\frac{SQR}{SQT} = 1 - \frac{QMR}{s^2_y}$$

- Neste critério, observa-se quão rapidamente $QMR(k)$ ou R^2_α são reduzidos com aumento de variáveis preditoras.

- O minímo absoluto do QMR não indica necessariamente o melhor modelo.

  $\Rightarrow$ "excesso de ajuste" (*overfit*) do modelo à amostra.

- Note que o QMR pode **aumentar** à medida que k cresce, caso a redução na soma de quadrados do resíduo não compense a redução nos graus de liberdade:

$$QMR = \frac{SQR}{n-k+1}$$

11.2.3 Critério C_p

- Considere que o com p variáveis preditoras potenciais (modelo completo com $p+1$ parâmetros) foi cuidadosamente escolhido de modo que

$$QMR(X_1, \ldots, X_k)$$

é uma estimativa não viciada de σ^2.

- Então

$$C_p(k) = \frac{SQR(k)}{QMR(X_1, \ldots, X_k)} - (n - 2(k+1))$$

deve ser igual a $k+1$ se o modelo com k variáveis preditoras não possuir nenhum viés.

- Note que se

$$\boldsymbol{E}\{QMR(X_1, \ldots, X_k)\} = \sigma^2 \quad \text{e} \quad \boldsymbol{E}\{SQR(k)\} = (n-k+1)\sigma^2$$

então

$$\begin{aligned} \boldsymbol{E}\{C_p(k)\} &= \frac{SQR(k)}{QMR(X_1, \ldots, X_k)} - (n - 2(k+1)) \\ &= \frac{(n-k-1)\sigma^2}{\sigma^2} - (n - 2(k+1)) \\ &= n - k - 1 - n + 2k + 2 \\ &= k + 1 \quad \text{(número de parâmetros)}. \end{aligned}$$

Portanto, podemos comparar os modelos através de um gráfico de $C_p(k)$ por $k+1$:

- O gráfico permite analisar o *viés* dos modelos:
  - **Viés pequeno** $\Rightarrow$ modelo próximo a linha $C_p(k) = k+1$.
  - **Viés grande** $\Rightarrow$ modelo bem **acima** da linha $C_p(k) = k+1$.

- O C_p contem tanto o *viés* quanto a *variância* e, portanto, está relacionada ao **erro quadrado total médio** ($EQTM$):

$$\begin{aligned} EQTM &= \text{viés}^2 + \text{ variância} \\ EQTM\{\widehat{\theta}\} &= \left(\boldsymbol{E}\left\{\widehat{\theta}\right\} - \theta\right)^2 + \boldsymbol{Var}\{\widehat{\theta}\} \end{aligned}$$

  Formalmente, a estatística $C_p(k)$ estima o critério:

$$\begin{aligned} \Gamma(k) &= \frac{\text{Erro Quadrado Total Médio}}{\text{Variância do Erro Verdadeira } (\sigma^2)} \\ \Gamma(k) &= \frac{\boldsymbol{E}\{SQR(k)\}}{\sigma^2} - (n - 2(k+1)) \end{aligned}$$

  Logo:

$$C_p(k) = \frac{SQR(k)}{QMR(X_1, \ldots, X_k)} - (n - 2(k+1))$$

  e, portanto, o que se deseja do melhor modelo é:

  1. $C_p \approx k + 1$,
  2. C_p pequeno (*Princípio da Parsimônia*).

11.2.4 Algoritmo para Identificar o "Melhor" Subconjunto

- Num problema com p variáveis preditoras potenciais existem 2^{p+1} modelos de regressão possíveis.
- Vários softwares estatísticos apresentam algoritmos para encontrar o "melhor" subconjunto de variáveis preditoras.

  Este algoritmo procurará somente os 5 ou 10 "melhores" subconjuntos utilizando um critérios (em geral um dos critérios acima) **sem ajustar** todos os modelos.
- Uma vez identificados os melhores subconjuntos devemos proceder com
  - análise dos resíduos,
  - análise de observações influentes,
  - os objetivos e conhecimento que o pesquisador tem em mente.

11.3 Regressão Passo-a-Passo (“Stepwise”)

Quando possuimos um grande número de variáveis preditoras, a abrodagem do “melhor subconjunto” pode não ser possível.

A regressão “passo-a-passo” (*stepwise*) é uma abordagem alternativa para se encontrar um subconjunto de variáveis X que seja “razoavelmente bom”.

11.3.1 Algoritmo de Procura

(1) Ajuste o modelo de regressão linear simples para cada uma das $p+1$ variáveis preditoras potenciais.

A variável com o **maior**

$$F_k^* = \frac{QMM(X_k)}{QMR(X_k)}$$

é incluida no modelo primeiro, **se** F_k^* exceder um limite previamente estabelecido: F_{entra}, geralmente o quantil 10% da distribuição F.

Essa variável será designada por X_a.

(2) Ajuste todos os modelos com duas variáveis preditoras compostos pela variável X_a selecionado no passo (1) e mais uma.

Compute:

$$F_k^* = \frac{QMM(X_k|X_a)}{QMR(X_a, X_k)} = \left[\frac{b_k}{s\{b_k\}}\right]^2$$

a variável X_k com o maior F^* é candidata à próxima inclusão.

Se $F_k^* > F_{\text{entra}}$ então X_k entra no modelo (a designaremos por X_b).

(3) Assuindo que X_b entrou no modelo, o modelo é examinado para verificar se alguma variável presente deve ser excluida.

No modelo com duas variáveis se computa

$$F_a^* = \frac{QMM(X_a|X_b)}{QMR(X_a, X_b)}.$$

No modelo com r variáveis, assumindo que X_r foi a última a entrar, se computa

$$F_k^* = \frac{QMM(X_k|X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_r)}{QMR(X_1, \ldots, X_r)} \qquad k = 1, \ldots, r-1.$$

A variável com o **menor** F_k^* é candidata à exclusão.

Se F_k^* for menor que um limite predeterminado, F_{sai}, geralmente o quantil 10% ou 20% da distribuição F, então X_k é excluida.

(4) Se ambas X_a e X_b são mantidas no modelo, repete-se o passo (2) para uma terceira variável.

Sempre seguindo a regra:

- inclui a variável de *fora do modelo* com o maior F_k^* (se maior que o limite F_{entra});
- exclui a variável de *dentro do modelo* com o menor F_k^* (se menor que o limite F_{sai}).

(5) Segue-se os passos (2) a (4) até que nenhuma inclusão/exclusão possa ocorrer.

O modelo resultante é o modelo “passo-a-passo”.

11.3.2 Outros Algoritmos de Regressão “Passo-a-passo”

Passo-a-passo para Frente (“Stepwise Forward”): é o mesmo algoritmo que a regressão passo-a-passo, mas as variáveis **não são excluidas**.

Passo-a-passo para Trás (“Stepwise Backward”): é a eliminação das variáveis a partir do modelo completo.

1. Começar com o modelo completo ($p+1$ variáveis preditoras).
2. Procurar a variável com o **menor** F parcial:

$$F_k^* = \frac{QMM(X_k|X_1,\ldots,X_{k-1},X_{k+1},\ldots,X_p)}{QMR(X_1,\ldots,X_p)}$$

3. Se o menor F_k^* estiver abaixo do limite mínimo, eliminar X_k .
4. Re-ajustar o modelo sem X_k, modelo com $p-1$ variáveis preditoras restantes.
5. Calcular o F parcial e continuar com a exclusão das variáveis até nenhum dos F parciais fique abaixo do limite minímo.

11.3.3 Crítica aos Critérios de Seleção de Variáveis Modelos

- Todos os critérios apresentados anteriormente, assumem a existência de uma **hierarquia** nos modelos, geralmente do modelo mais completo (com todas variáveis preditoras) toma-se uma série de modelos *reduzidos* diminuindo-se o número de variáveis preditoras.
- Para comparar modelos **não hierarquicos** os critérios apresentados são falhos, pois não há como assumir que um dos modelos gera a "*melhor*" estimativa da variância do erro (σ^2).
- O Critério de Akaike (AIC) e suas variações são critérios diferentes desses, pois não assumem a existência de modelos hierárquicos, mas simplesmente de modelos concorrentes.
- AIC é uma estimativa da *Logverossimilhança Negativa dos Modelos* ponderanda para o número de parâmetros estimados, consequentemente, o modelo com menor valor de AIC é o mais apropriado.

11.4 Validação de Modelos

Existem três abordagens básicas para a validação de modelos de regressão:

(1) Coletar dados **novos** e verificar a capacidade preditora do modelo ajustado.

(2) Comparar os resultados do modelo com:

- resultados teóricos esperados,
- resultados empíricos anteriores ou resultados de simulações.

(3) Dados de validação:

- Dividir os dados em **duas amostras** através de seleção aleatória das observações:
  amostra de ajuste,
  amostra de validação.
- Utilizar apenas a amostra de ajuste na construção do modelo.
- Tendo o modelo, verificar a capacidade preditora na amostra de validação.
- **Validação Cruzada:** cruzar os a construção/verificação entre os dois conjuntos de dados.