

Teste da falta de ajustamento

Profª Silvia Nagib Elian
Sala 215 - Bloco A

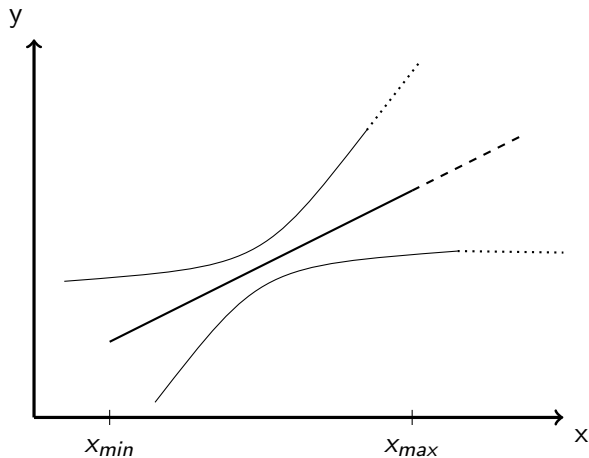
Instituto de Matemática e Estatística
Universidade de São Paulo

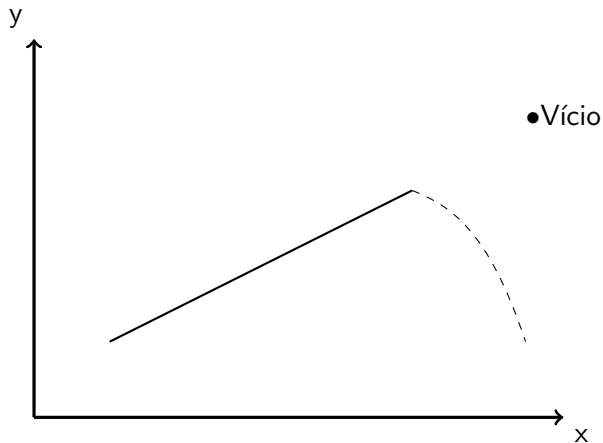
Agenda

1. Exemplo

2. ANOVA para falta de ajustamento

Visualização





Valores

					média	SS	GL
x_1	y_{11}	y_{12}	$\cdots$	y_{1n_1}	$\bar{y}_1$	$\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2$	$n_1 - 1$
x_2	y_{21}	y_{22}	$\cdots$	y_{2n_2}	$\bar{y}_2$	$\sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2$	$n_2 - 1$
$\vdots$		$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_k	y_{k1}	y_{k2}	$\cdots$	y_{kn_k}	$\bar{y}_k$	$\sum_{j=1}^{n_k} (y_{kj} - \bar{y}_k)^2$	$n_k - 1$

onde $\sum_{i=1}^k n_i = n$, $n_i \geq 2$, $i = 1, 2, \dots, k$.

Definições

$$\text{SSE} = \text{soma de quadrados dos resíduos} = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2$$

$$\text{SSEP} = \text{soma de quadrados do erro puro} = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

$$\text{SSFA} = \text{soma de quadrados da falta de ajuste} = \sum_{i=1}^k n_i (\bar{y}_i - \hat{y}_i)^2$$

Resultado

$$\text{SSE} = \text{SSEP} + \text{SSFA}$$

Prova do resultado

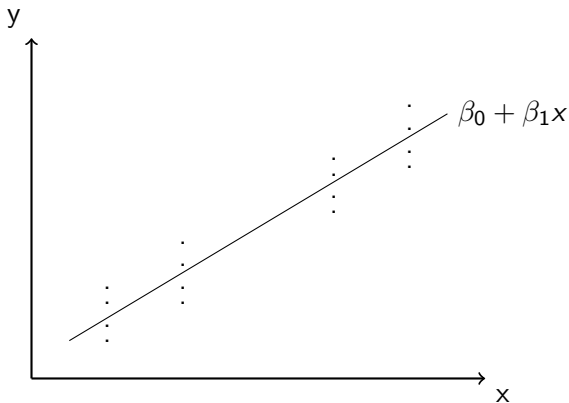
Considere: $e_i = y_{ij} - \hat{y}_i = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \hat{y}_i)$

$$\begin{aligned} \text{SSE} &= \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{y}_i - \hat{y}_i)^2 + 2 \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)(\bar{y}_i - \hat{y}_i) \\ &= \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^k n_i (\bar{y}_i - \hat{y}_i)^2 + 2 \sum_{i=1}^k (\bar{y}_i - \hat{y}_i) \sum_{j=1}^{n_i} \underbrace{(y_{ij} - \bar{y}_i)}_{\rightarrow 0} \\ &= \underbrace{\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}_{\text{SSEP}} + \underbrace{\sum_{i=1}^k n_i (\bar{y}_i - \hat{y}_i)^2}_{\text{SSFA}} \end{aligned}$$

Observação: Se $n_i = 2$,

$$\left(y_1 - \frac{y_1 + y_2}{2} \right)^2 + \left(y_2 - \frac{y_1 + y_2}{2} \right)^2 = \frac{(y_1 - y_2)^2}{2}$$

Quando utilizamos o modelo linear, admitimos:



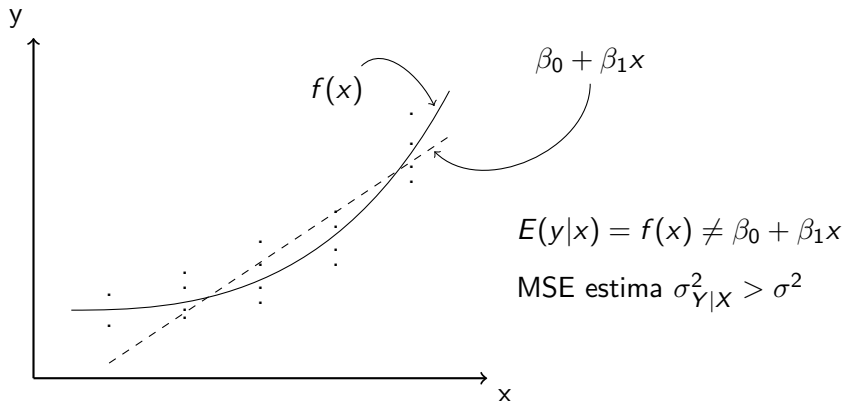
$$E(y|x) = \beta_0 + \beta_1 x$$

$\sigma_{Y|X}^2 \rightarrow$ variância
em torno da reta
 $\beta_0 + \beta_1 x$

MSE estima $\sigma_{Y|X}^2$

Se o modelo é $\beta_0 + \beta_1 x$, $\sigma_{Y|X}^2 = \sigma^2 = \text{Var}(Y)$ e MSE é um estimador não viciado de σ^2 .

Se a situação real for:



Se o modelo não é $\beta_0 + \beta_1 x$, MSE superestima σ^2 .

$$\hat{\sigma}^2 = \frac{\text{SSEP}}{n - k} \quad \text{qualquer que seja o modelo.}$$

ANOVA para falta de ajustamento

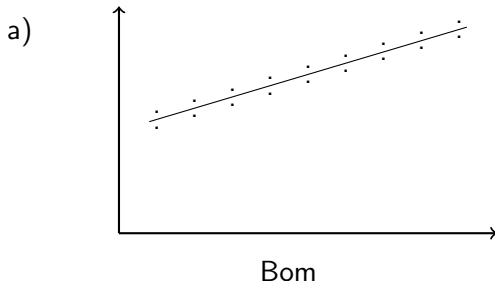
FV	GL	SS	MS	F
falta de ajustamento	k-2	SSFA	MSFA	$F_{FA} = \frac{MSFA}{MSEP}$
erro puro	n-k	SSEP	MSEP	
resíduo	n-2	SSE		

H_0 : o modelo é linear [$E(\epsilon_1) = E(\epsilon_2) = \dots = E(\epsilon_k) = 0$]

Rejeitamos H_0 se $F_{FA} \geq F(k-2, n-k, \alpha)$

ANOVA

FV	GL	SS	MS	F	Hipótese
regressão	1	SSR	MSR	F	$\beta_1 = 0$
resíduo	n-2	SSE			
F. de ajuste	k-2	SSFA	MSFA	$F_{FA} = \frac{MSFA}{MSEP}$	O modelo é linear
E. puro	n-k	SSEP	MSEP		

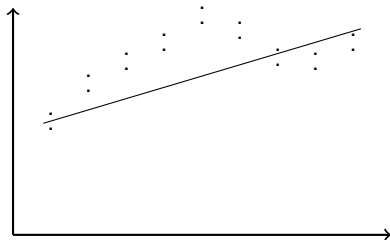


F sig.
 F_{FA} não sig.

ANOVA

FV	GL	SS	MS	F	Hipótese
regressão	1	SSR	MSR	F	$\beta_1 = 0$
resíduo	n-2	SSE			
F. de ajuste	k-2	SSFA	MSFA	$F_{FA} = \frac{MSFA}{MSEP}$	O modelo é linear
E. puro	n-k	SSEP	MSEP		

b)

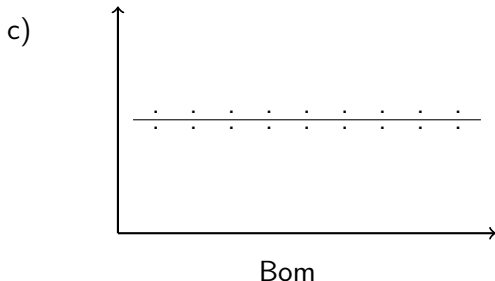


F sig.

F_{FA} sig.

ANOVA

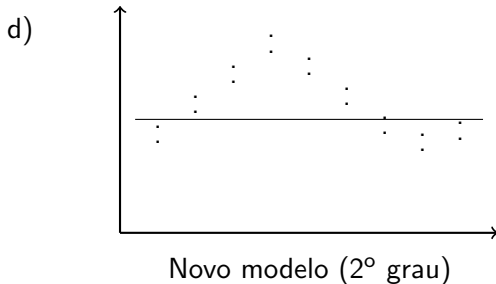
FV	GL	SS	MS	F	Hipótese
regressão	1	SSR	MSR	F	$\beta_1 = 0$
resíduo	n-2	SSE			
F. de ajuste	k-2	SSFA	MSFA	$F_{FA} = \frac{MSFA}{MSEP}$	O modelo é linear
E. puro	n-k	SSEP	MSEP		



F não sig.
 F_{FA} não sig.

ANOVA

FV	GL	SS	MS	F	Hipótese
regressão	1	SSR	MSR	F	$\beta_1 = 0$
resíduo	n-2	SSE			
F. de ajuste	k-2	SSFA	MSFA	$F_{FA} = \frac{MSFA}{MSEP}$	O modelo é linear
E. puro	n-k	SSEP	MSEP		



F não sig.

F_{FA} sig.

- ❶ Obter o modelo. Construir ANOVA. Não testar $H_0 : \beta_1 = 0$ ainda.
- ❷ Construir a ANOVA para a falta de ajustamento. Se F_{FA} sig., ir para 3a. Se F_{FA} não sig., ir para 3b.
- ❸
 - ◇ a) F_{FA} sig: Parar a análise deste modelo e procurar um melhor. Não fazer o teste F e nem construir intervalos de confiança. [As suposições nas quais esses cálculos são baseados não são verdadeiras se houver falta de ajuste].
 - ◇ b) F_{FA} não sig.: $\hat{\sigma}^2 = \text{MSE}$, testar $H_0 : \beta_1 = 0$, construir intervalos de confiança, calcular R^2 , etc. Fazer análise de resíduos.

Montgomery and Peck (pág. 77)

- F_{FA} não significativa e $F(\beta_1 = 0)$ significativa não garantem que o modelo será satisfatório para **previsão**.
- Alguns autores sugerem que para um modelo ser útil para previsão, devemos ter

$$F = \frac{MSR}{MSE} > 4F_c.$$

- Usa-se também a medida:

$$\frac{\hat{y}_{max} - \hat{y}_{min}}{\sqrt{\frac{p\hat{\sigma}^2}{n}}},$$

onde p é o número de parâmetros na equação de regressão (reg. lin. simples $p = 2$), $\hat{\sigma}^2$ é a estimativa de σ^2 independente do modelo. Verifica-se que $\overline{V(\hat{y})} = n^{-1} \sum_{i=1}^n \text{Var}(\hat{y}_i) = \frac{p\sigma^2}{n}$

Exemplo

x_i	y_{ij}					n_i	$\bar{y}_i$	$\sum_j (y_{ij} - \bar{y}_i)^2$	GL
20	96	92	106	100	100	4	98,5	107	3
25	98	104	110	101	101	4	103,25	78,75	3
30	116	106	109	100	100	4	107,75	132,75	3
35	112	105	118	108	108	4	110,75	94,75	3
40	113	112	127	117	117	4	117,25	140,75	3
						20	SSEP = 554		15

$$SSE = 563 \quad SSEP = 554 \quad SSFA = 563 - 554 = 9$$

ANOVA para falta de ajustamento

FV	GL	SS	MS	F
falta de ajustamento	3	9	3	0,0812 (não sig.)
erro puro	15	554	36,93	
resíduo	18	563	563	

H_0 : o modelo é linear.

Não temos evidências contra o modelo linear.