

# Modelos Lineares I

## Regressão Linear Simples (RLS):

### Análise dos Resíduos

(13ª, 14ª e 15ª Aulas)



Professor: Dr. José Rodrigo de Moraes

Universidade Federal Fluminense (UFF)

Departamento de Estatística (GET)

1

### Resíduos do Modelo de Regressão Linear Normal:

- A construção do modelo de RLS está fundamentada em várias hipóteses (ou suposições) que devem ser verificadas:
  - Os erros aleatórios  $\varepsilon_i$ 's são supostamente v.a.'s não correlacionadas com média 0 e variância constante  $\sigma^2$ .
  - Os erros aleatórios  $\varepsilon_i$ 's são normalmente distribuídos.
- A verificação da hipótese distribucional dos erros é de fundamental importância para a realização de inferências sobre os parâmetros  $\beta_0$  e  $\beta_1$  do modelo de RLS. Para verificar tais hipóteses sobre os erros aleatórios  $\varepsilon_i = Y_i - E(Y_i)$ , utiliza-se os resíduos do modelo:  $e_i = Y_i - \hat{Y}_i$ ,  $i=1, 2, \dots, n$ .
- Valores bem ajustados apresentam pequenos resíduos, enquanto que valores mal ajustados apresentam grandes resíduos.

### Propriedades dos resíduos ( $e$ ) do modelo:

- **Média:** A média dos resíduos é igual a zero:

$$\bar{e} = \frac{\sum_{i=1}^n e_i}{n} = \frac{e_1 + e_2 + \dots + e_n}{n} = 0$$

- **Variância:** A variância dos resíduos é dada por:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (e_i - \bar{e})^2}{n-2} = \frac{\sum_{i=1}^n e_i^2}{n-2} = \frac{SQRes}{n-2} = QMRes$$

**OBS:** A variância dos resíduos é denominada "Quadrado Médio dos Resíduos (QMRes)".

3

### Resíduos Padronizados ( $r^p$ ):

- Os resíduos padronizados são úteis para avaliar a adequação do modelo ajustado, bem como para identificar a presença de valores discrepantes ou atípicos (*outliers*). Os resíduos padronizados, denotados por  $r^p$ , são definidos abaixo:

$$r_i^p = \frac{e_i - \bar{e}}{\sqrt{\hat{\sigma}^2}} = \frac{e_i}{\sqrt{QMRes}}$$

**OBS:** Também servem para avaliar a normalidade dos erros e homocedasticidade da variância.

4

### Resíduos estudentizados ( $r^s$ ):

- Pacotes estatísticos calculam ainda os resíduos estudentizados, que são considerados como medidas mais adequadas para o exame de valores discrepantes ou atípicos (*outliers*), pois a variância dos resíduos  $e_i$ 's não é constante. Com efeito,  $e_i$  tem distribuição normal de média 0 e variância  $VAR(e_i) = \sigma^2 (1 - h_{ii})$ . Os resíduos estudentizados, denotados por  $r^s$ , são definidos abaixo:

$$r_i^s = \frac{e_i}{\sqrt{\hat{\sigma}^2 \cdot (1 - h_{ii})}}; \quad \forall i = 1, 2, \dots, n$$

onde:

$h_{ii}$  é o  $i$ -ésimo elemento da diagonal da matriz  $H = X(X'X)^{-1}X'$ , denominada matriz de projeção (ou matriz hat).

### Observações: Resíduos estudentizados ( $r^s$ ):

- Os valores  $h_{ii}$  (*medidas de alavancagem* = "leverage") da matriz  $H$ , tal que  $1/n \leq h_{ii} \leq 1$ , são usados para avaliar se as observações tem potencial para influenciar o ajuste do modelo.
- Em geral:
  - $h_{ii} \approx 1/n \rightarrow$  a obs.  $i$  tem baixo potencial para influenciar o ajuste do modelo.
  - $h_{ii} \approx 1 \rightarrow$  a obs.  $i$  tem alto potencial para influenciar o ajuste do modelo. OBS:  $h_{ii} > 0,5 \rightarrow$  alta alavancagem.
- A vantagem dos resíduos estudentizados é que, se o modelo é satisfatório, eles devem apresentar distribuição aproximadamente normal com média zero e variância constante (e igual a 1).

### Inadequação do modelo: Violações das Hipóteses Básicas

□ A análise dos resíduos (ou dos resíduos padronizados) são de fundamental importância na avaliação de 5 tipos de violações no ajuste de modelo de regressão linear normal:

1. **Não linearidade:** ausência de relação linear entre  $X$  e  $Y$ .
2. **Não normalidade:** os erros aleatórios não são normalmente distribuídos.
3. **Heterocedasticidade:** os erros aleatórios não tem variância constante.
4. **Não independência:** os erros aleatórios não são independentes.
5. **Outliers:** presença de um ou mais valores discrepantes ou atípicos (*outliers*).

7

### Análise Gráfica dos Resíduos do Modelo de RLS:

#### Violações das Hipóteses Básicas

1) **Não linearidade** → ausência de relação linear entre  $X$  e  $Y$

**Gráfico 1.1:** Gráfico de dispersão entre  $e$  vs  $X$ .

**Gráfico 1.2:** Gráfico de dispersão entre  $e$  vs  $\hat{Y}$ .

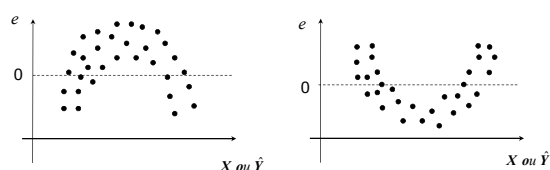
**Gráfico 1.3:** Gráfico de dispersão entre  $X$  e  $Y$  (juntamente com o modelo ajustado).

8

#### 1) Não linearidade

**Gráfico 1.1:** Gráfico de dispersão  $e$  vs  $X$

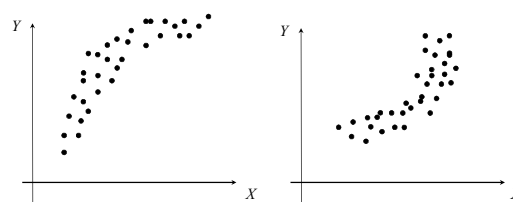
**Gráfico 1.2:** Gráfico de dispersão  $e$  vs  $\hat{Y}$



Situações como essas indicam possíveis inadequações do modelo adotado, e as curvas sugerem que devemos procurar outras funções matemáticas (ex.: modelo quadrático) que melhor expliquem o fenômeno, pois a relação não é linear.<sup>9</sup>

#### 1) Não linearidade

**Gráfico 1.3:** Gráfico de dispersão  $X$  vs  $Y$  (incluindo modelo ajustado):

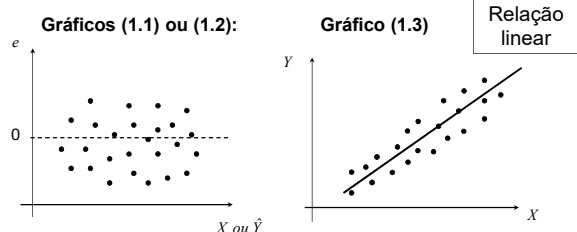


Situações como essas indicam possíveis inadequações do modelo adotado, indicando que a hipótese de linearidade não é satisfeita.

10

### Situações adequadas (*Linearidade*):

Modelo apropriado !!!



Os resíduos estão aleatoriamente distribuídos em torno de zero.

11

### Análise Gráfica dos Resíduos do Modelo de RLS:

#### Violações das Hipóteses Básicas

2) **Não normalidade** → os erros aleatórios não estão normalmente distribuídos.

**Gráfico 2.1:** Histograma dos resíduos.

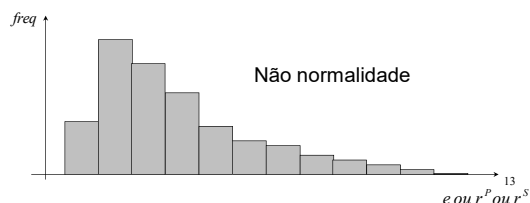
**Gráfico 2.2:** *QQ-Plot* dos resíduos.

12

## 2) Não normalidade dos erros:

### Gráfico 2.1: Histograma dos resíduos:

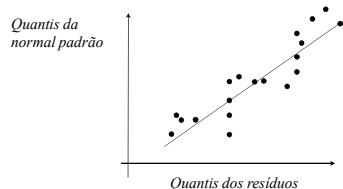
Grandes desvios da normalidade são potencialmente sérios, e portanto medidas precisam ser tomadas para que os resíduos do modelo se comportem pelo menos de maneira aproximadamente normal.



## 2) Não normalidade dos erros:

### Gráfico 2.2: Gráfico dos Quantis (QQ-Plot):

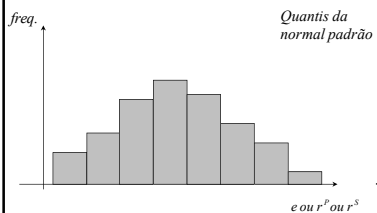
No *QQ-Plot* representa-se em um dos eixos os quantis dos resíduos observados e no outro eixo os quantis teóricos de uma *distribuição normal padrão*.



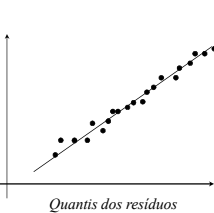
14

## Situações adequadas (Normalidade):

### Gráfico 2.1: Histograma dos resíduos



### Gráfico 2.2: QQ-Plot dos resíduos



15

## Análise Gráfica dos Resíduos do Modelo de RLS: Violações das Hipóteses Básicas

**3) Heterocedasticidade:** os erros aleatórios não tem variância constante.

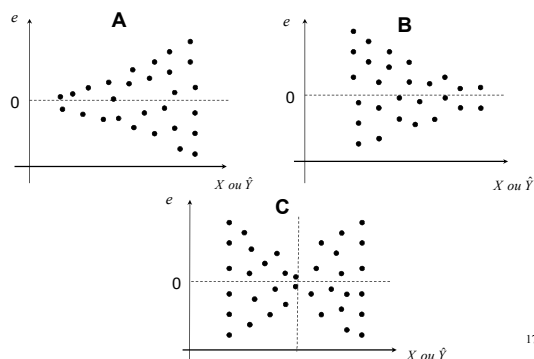
**Gráfico 3.1:** Gráfico de dispersão entre  $e$  vs  $X$ .

**Gráfico 3.2:** Gráfico de dispersão entre  $e$  vs  $\hat{Y}$ .

16

### 3.1) Gráfico de dispersão $e$ vs $X$

### 3.2) Gráfico de dispersão $e$ vs $\hat{Y}$



17

## 3) Heterocedasticidade (continuação):

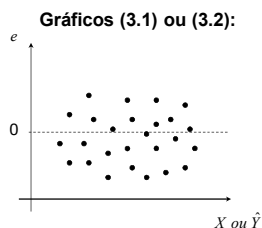
### 3.1) Gráfico de dispersão $e$ vs $X$

### 3.2) Gráfico de dispersão $e$ vs $\hat{Y}$

Situações como essas (A, B ou C) indicam violação da hipótese de homocedasticidade (variância constante), isto é, os erros são heterocedásticos.

18

### Situação adequada (*Homocedasticidade*):



19

### Observações – Violações das hipóteses básicas do modelo:



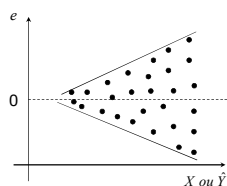
- Feita a análise gráfica dos resíduos e diagnosticada alguma violação das hipóteses básicas do modelo, medidas devem ser tomadas a fim de tornar o modelo mais adequado aos dados observados e as hipóteses feitas.
- Quando a hipótese de homocedasticidade (variância constante) não for satisfeita pode-se aplicar uma transformação nos valores de Y para *estabilizar a variância* e ajusta-se o modelo com base nos dados transformados. Entre as transformações comumente utilizadas para contornar tal violação estão:

20

### Transformações para estabilização da variância:

#### Situação 1:

##### Gráfico dos resíduos

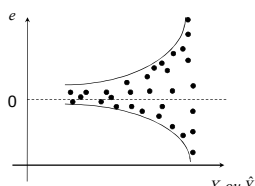


$Y_i \sim \text{Poisson}$

**Transformação:**  $Y_i^* = \sqrt{Y_i}$

#### Situação 2:

##### Gráfico dos resíduos



$Y_i \sim \text{Lognormal}$

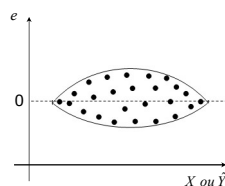
**Transformação:**  $Y_i^* = \log(Y_i)$

21

### Transformações para estabilização da variância:

#### Situação 3:

##### Gráfico dos resíduos



$Y_i \sim \text{binomial (frações - \%)}$ , bastante dispersa.

**Transformação:**  $Y_i^* = \arcsen \sqrt{Y_i}$

22

### Observações importantes – Transformações para estabilização da variância:

- A transformação da variável resposta é um método apropriado para estabilizar a variância dos resíduos.
- Em geral, transformações são usadas para:
  - 1) Estabilizar a variância;
  - 2) Tornar a distribuição aproximadamente normal;
  - 3) Produzir um melhor ajuste do modelo.
- Embora as transformações tenham sido desenvolvidas para *estabilizar a variância*, em muitos casos, também conduzem a *normalização*.



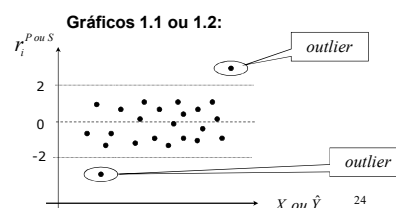
23

### Análise Gráfica dos Resíduos do Modelo de RLS: Violações das Hipóteses Básicas

#### 5) Presença de outliers:

##### Critério para identificação de outliers:

$|r_i^{P \text{ ou } S}| > 2 \rightarrow$  a unidade i é um potencial **outlier**.



24

**Análise Gráfica dos Resíduos do Modelo de RLS:**  
**Violações das Hipóteses Básicas**

**5) Presença de outliers:**

Observações:

- A presença de outlier (valor discrepante ou atípico) pode causar prejuízos para o ajuste de uma reta de regressão (modelo linear), pois a reta pode ser puxada desproporcionalmente para este valor.
- Um outlier pode no entanto conter significativas informações, de forma que a simples exclusão desse ponto poderia causar substancial perda para o ajuste do modelo.

25

**Voltando ao exemplo dos n=30 bovinos:** Dados sobre a concentração da substância  $X$  (mg/L) e ganho de peso  $Y$  (kg):

**Resíduos (brutos) do modelo:**

$$e_i = y_i - \hat{y}_i = y_i - (10,040 + 0,732 x_i); \quad \forall \quad i = 1, 2, \dots, 30$$

**Resíduos padronizados do modelo:**

$$r_i^p = \frac{e_i - \bar{e}}{\sqrt{\hat{\sigma}^2}} = \frac{e_i}{\sqrt{QMRes}} = \frac{e_i}{\sqrt{1,3368}}; \quad \forall \quad i = 1, 2, \dots, 30$$

onde:

$$\hat{\sigma}^2 = QMRes = \frac{\sum_{i=1}^n e_i^2}{n-2} = \frac{(-1,37)^2 + (-1,35)^2 + \dots + (0,87)^2}{30-2} = \frac{37,4294}{28} \approx 1,3368$$

26

**Exemplo:** Dados sobre a concentração da substância  $X$  (mg/L) e ganho de peso  $Y$  (kg) de n=30 bois:

**Cálculo dos resíduos brutos e padronizados**

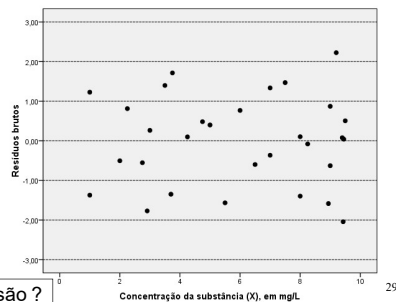
boi i	$X_i$	$Y_i$	$e_i = Y_i - \hat{Y}_i$	$r_i^p = e_i / \sqrt{\hat{\sigma}^2}$
1	1,00	9,40	-1,37	-1,19
2	3,70	11,40	-1,35	-1,17
3	1,00	12,00	1,23	1,06
4	9,00	16,00	-0,63	-0,54
5	2,00	11,00	-0,50	-0,44
6	2,25	12,50	0,81	0,70
7	2,91	10,40	-1,77	-1,53
8	2,75	11,50	-0,55	-0,48
9	3,00	12,50	0,26	0,23
10	3,50	14,00	1,40	1,21
11	3,75	14,50	1,72	1,48
12	9,45	17,00	0,04	0,04
13	4,25	13,25	0,10	0,09
14	7,00	14,80	-0,36	-0,31
15	4,75	14,00	0,48	0,42

**Continuação - Cálculo dos resíduos brutos e padronizados:**

boi i	$x_i$	$y_i$	$e_i = Y_i - \hat{Y}_i$	$r_i^p = e_i / \sqrt{\hat{\sigma}^2}$
16	5,00	14,10	0,40	0,35
17	5,50	12,50	-1,57	-1,35
18	6,00	15,20	0,77	0,66
19	6,50	14,20	-0,60	-0,52
20	7,00	16,50	1,34	1,16
21	7,50	17,00	1,47	1,27
22	8,00	14,50	-1,40	-1,21
23	8,25	16,00	-0,08	-0,07
24	9,40	17,00	0,08	0,07
25	9,43	14,90	-2,04	-1,77
26	8,94	15,00	-1,58	-1,37
27	9,20	19,00	2,23	1,92
28	9,50	17,50	0,51	0,44
29	8,00	16,00	0,10	0,09
30	9,00	17,50	0,87	0,75
Total	177,53	431,15	0,00	0,00

**Exemplo:** Usando os dados dos n=30 bois, faça a análise gráfica dos resíduos (brutos, padronizados e estudentizados):

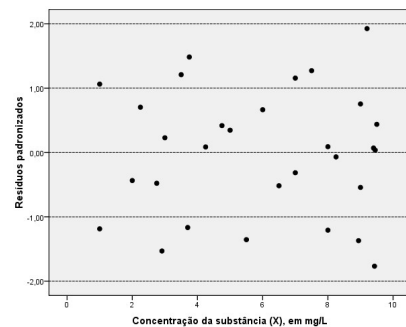
**Gráfico de dispersão entre a concentração da substância (X) e os resíduos brutos (e) do modelo:**



Qual a conclusão ?

29

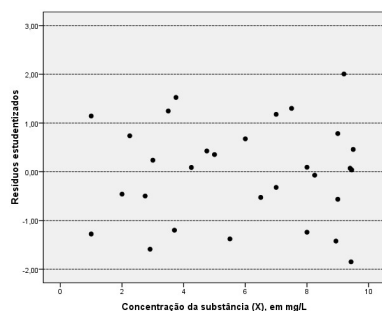
**Gráfico de dispersão entre a concentração da substância (X) e os resíduos padronizados do modelo ( $r^p$ ):**



Qual a conclusão ?

30

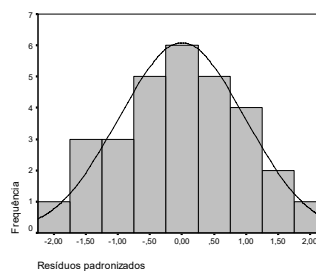
**Gráfico de dispersão entre a concentração da substância (X) e os resíduos estudatizados do modelo ( $r^S$ ):**



Qual a conclusão ?

31

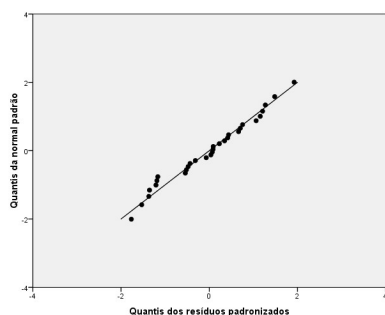
**Histograma dos resíduos padronizados ( $r^P$ ):**



Distribuição aproximadamente normal

32

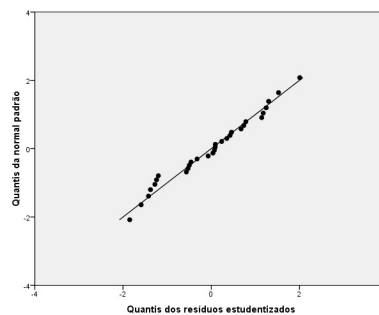
**QQ Plot dos resíduos padronizados ( $r^P$ ):**



Hipótese de normalidade dos erros é satisfeita !!!

33

**QQ Plot dos resíduos estudatizados ( $r^S$ ):**



Hipótese de normalidade dos erros é satisfeita !!!

34

**Gráfico dos resíduos versus variável explicativa (ou valores estimados) - Lembrete:**

- ❑ Se o modelo é adequado e se não existem violações da hipótese de independência ou de variância constante dos erros, os resíduos não devem apresentar nenhum padrão, isto é, eles não devem estar correlacionados com a variável explicativa X, nem com os valores estimados.
- ❑ Construídos os gráficos dos resíduos versus variável explicativa X (ou valores estimados), é preciso analisá-los objetivando identificar a existência ou não de possíveis violações das hipóteses básicas do modelo.

35

**Inadequação do modelo de regressão linear:**

**Violações das hipóteses básicas**

- ❑ Exemplos práticos para ilustrar algumas das violações que podem ocorrer na análise de regressão linear:

**Exemplo 1: Não linearidade da relação entre X e Y**

**Tipos de gráficos usados:**

- Gráfico de dispersão com a reta (modelo) de regressão ajustada(o).
- Gráfico dos resíduos versus variável explicativa X
- Gráfico dos resíduos versus valores ajustados  $\hat{Y}$

36

### Inadequação do modelo: Violações das hipóteses básicas

#### Exemplo 1: Não linearidade da relação entre X e Y

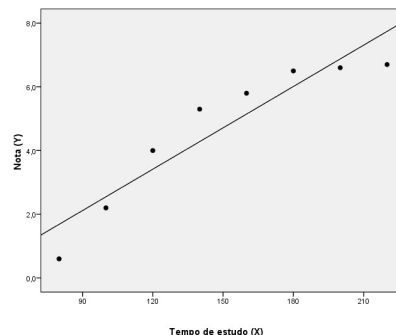
A tabela abaixo fornece informações sobre a *nota dos alunos* e o *tempo de estudo* (em minutos), para  $n=8$  alunos que ficaram em verificação suplementar (VS).

Tabela 1: Dados de 8 alunos

Aluno	Tempo de estudo	Nota
1	80	0,6
2	220	6,7
3	140	5,3
4	120	4,0
5	180	6,5
6	100	2,2
7	200	6,6
8	160	5,8

37

Figura 1: Gráfico de Dispersão entre o tempo de estudo (X) e a nota dos alunos (Y), incluindo a reta de regressão estimada.



Falta de linearidade

38

### Ajuste do modelo

Model Summary<sup>a</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.935 <sup>a</sup>	.875	.854	.8669

a. Predictors: (Constant), Tempo\_estudo  
b. Dependent Variable: Nota\_VS

ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	31,460	1	31,460	41,866	.001 <sup>a</sup>
	Residual	4,509	6	.751		
	Total	35,969	7			

a. Predictors: (Constant), Tempo\_estudo  
b. Dependent Variable: Nota\_VS

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients		t		Sig.		95.0% Confidence Interval for B	
		B	Std. Error	Beta						Lower Bound	Upper Bound
1	(Constant)	-1,779	1,049			-1,696	,141			-4,345	,788
	Tempo_estudo	,043	,007	,935		6,470	,001			,027	,060

a. Dependent Variable: Nota\_VS

39

### Inadequação do modelo: Violações das Hipóteses Básicas

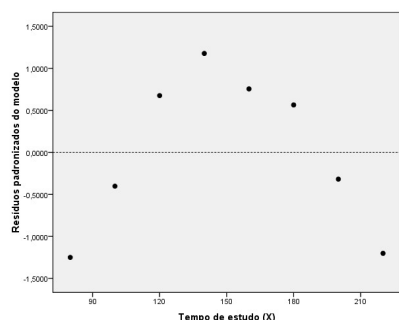
#### Exemplo 1: Não linearidade da relação entre X e Y (continuação)

Tabela 1: Resíduos do modelo

Aluno i	Resíduos ( $e_i$ )	Resíduos padronizados ( $r_i^p$ )
1	-1,0833	-1,2497
2	-1,0417	-1,2017
3	1,0202	1,1769
4	0,5857	0,6757
5	0,4893	0,5644
6	-0,3488	-0,4024
7	-0,2762	-0,3186
8	0,6548	0,7553

40

Figura 2: Gráfico de dispersão entre o tempo de estudo (X) e os resíduos padronizados do modelo ( $r_i^p$ ).



Padrão sistemático dos resíduos

Pergunta: E se utilizássemos os resíduos brutos ou estudatizados ?

### Análise gráfica dos resíduos:

- O gráfico dos "tempos de estudo" *versus* os "resíduos padronizados" sugere a falta de linearidade, pois apresentam um comportamento sistemático, ou seja, os resíduos são negativos para pequenos valores de X, positivos para valores médios de X e também negativos para grandes valores de X.



42

### Inadequação do modelo:

#### Violações das Hipóteses Básicas

##### Exemplo 2: *Outliers e Não linearidade*

Um pesquisador levantou para uma amostra de 30 famílias de um condomínio de classe média, em Niterói, informações sobre a “renda mensal familiar” e a “despesa mensal com alimentação”, ambas as variáveis foram medidas em reais (R\$). O objetivo é relacionar a renda das famílias (X) e a despesa com alimentação (Y).

Os dados são fornecidos na tabela a seguir:

43

### Inadequação do modelo: Violações das Hipóteses Básicas

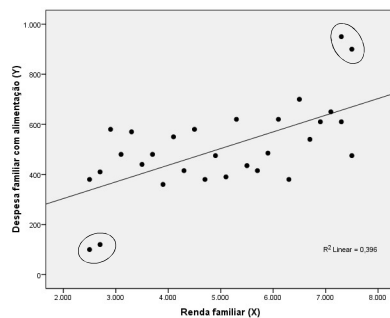
#### Exemplo 2: *Outliers e Não linearidade*

**Tabela 1:** Rendas (X) e despesas com alimentação (Y) de n=30 famílias de classe média de um condomínio, em Niterói.

Família	Renda	Despesa com alimentação	Família	Renda	Despesa com alimentação
1	2.500	380	16	5.500	435
2	2.700	410	17	5.700	415
3	2.900	580	18	5.900	485
4	3.100	480	19	6.100	620
5	3.300	570	20	6.300	380
6	3.500	440	21	6.500	700
7	3.700	480	22	6.700	540
8	3.900	360	23	6.900	610
9	4.100	550	24	7.100	650
10	4.300	415	25	7.300	610
11	4.500	580	26	7.500	475
12	4.700	380	27	2.500	100
13	4.900	475	28	2.700	120
14	5.100	390	29	7.300	950
15	5.300	620	30	7.500	900

44

**Figura 1:** Gráfico de Dispersão entre a renda (X) e a despesa familiar com alimentação (Y).



Existência de 4 valores discrepantes (outliers)

Medida de qualidade do ajuste:  $R^2 = 39,6\%$

45

### Ajuste do modelo (n=30 funcionários):

#### Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,629 <sup>a</sup>	,396	,375	140,186

a. Predictors: (Constant), Renda\_familiar

#### ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	361156,950	1	361156,950	18,377	,000 <sup>a</sup>
	Residual	550259,716	28	19652,133		
	Total	911416,667	29			

a. Predictors: (Constant), Renda\_familiar

b. Dependent Variable: Desp\_alimentação

#### Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients		95.0% Confidence Interval for B		
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	170,654	81,715		2,088	,046	3,267	338,040
	Renda_familiar	,067	,016	,629	4,287	,000	,035	,098

a. Dependent Variable: Desp\_alimentação

46

### Inadequação do modelo: Violações das Hipóteses Básicas

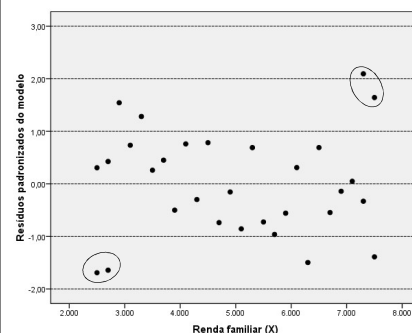
#### Exemplo 2: *Outliers e Não linearidade*

**Tabela 1:** Resíduos padronizados do modelo

Família	Resíduos padronizados	Família	Resíduos padronizados
1	0,3068	16	-0,7248
2	0,4259	17	-0,9624
3	1,5436	18	-0,5579
4	0,7353	19	0,3101
5	1,2824	20	-1,4968
6	0,2602	21	0,6910
7	0,4506	22	-0,5453
8	-0,5004	23	-0,1409
9	0,7601	24	0,0495
10	-0,2979	25	-0,3308
11	0,7842	26	-1,3887
12	-0,7374	27	-1,6906
13	-0,1547	28	-1,6428
14	-0,8559	29	2,0946
15	0,6898	30	1,6430

47

**Figura 2:** Gráfico de dispersão entre as rendas familiares (X) e os resíduos padronizados do modelo ( $r^2$ ).

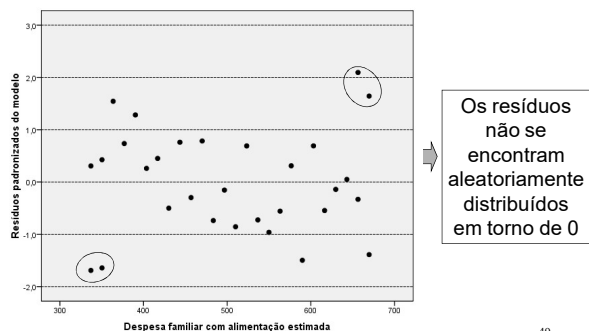


Os resíduos não se encontram aleatoriamente distribuídos em torno de 0

48



**Figura 3: Gráfico de dispersão entre as despesas com alimentação estimadas e os resíduos padronizados do modelo ( $r^2$ ).**



49

#### Análise gráfica dos resíduos:

- Os gráficos sugerem que os 4 pontos indicados nas figuras 2 e 3 são valores discrepantes, e parece comprometer a hipótese de linearidade do modelo.
- Parecem ser pontos que influenciam a inclinação da reta, ou seja, se eliminados podem fazer com que a reta fique paralela ao eixo X ( $\beta_1 \rightarrow 0$ ), levando a não verificação da hipótese de linearidade entre X e Y.
- Checando:** Ajustar o modelo após a exclusão dos referidos 4 valores discrepantes ...

50

#### Inadequação do modelo: Violações das Hipóteses Básicas

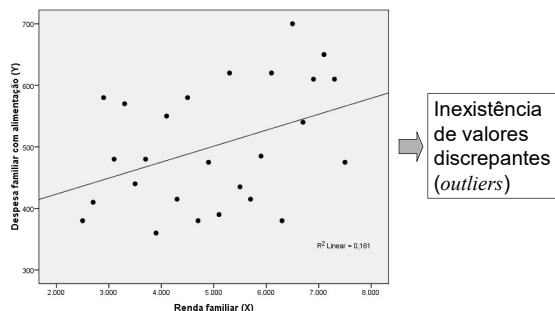
##### Exemplo 2: Outliers e Não linearidade

**Tabela 1: Rendas mensais (X) e despesas com alimentação (Y) de n=26 famílias (sem os outliers).**

Família	Renda mensal	Despesa com alimentação	Família	Salário mensal	Despesa com alimentação
1	2.500	380	16	5.500	435
2	2.700	410	17	5.700	415
3	2.900	580	18	5.900	485
4	3.100	480	19	6.100	620
5	3.300	570	20	6.300	380
6	3.500	440	21	6.500	700
7	3.700	480	22	6.700	540
8	3.900	360	23	6.900	610
9	4.100	550	24	7.100	650
10	4.300	415	25	7.300	610
11	4.500	580	26	7.500	475
12	4.700	380	-	-	-
13	4.900	475	-	-	-
14	5.100	390	-	-	-
15	5.300	620	-	-	-

51

**Figura 4: Gráfico de dispersão entre a renda (X) e a despesa familiar com alimentação (Y).**



Medida de qualidade do ajuste:  $R^2 = 16,1\%$

52

#### Ajuste do modelo (n=26 funcionários):

##### Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,401 <sup>a</sup>	,161	,126	92,505

a. Predictors: (Constant), Renda\_familiar

##### ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	39442,068	1	39442,068	4,609	,042 <sup>a</sup>
	Residual	205373,316	24	8557,222		
	Total	244815,385	25			

a. Predictors: (Constant), Renda\_familiar

b. Dependent Variable: Desp\_alimentação

##### Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	371,325	63,135		5,881	,000	241,020	501,629
	Renda_familiar	,026	,012	,401	2,147	,042	,001	,051

a. Dependent Variable: Desp\_alimentação

53

#### Inadequação do modelo: Violações das Hipóteses Básicas

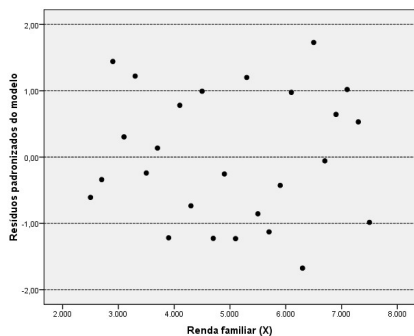
##### Exemplo 2: Outliers e Não linearidade

**Tabela 1: Resíduos do modelo (ajuste sem os outliers)**

Família	Resíduos padronizados	Família	Resíduos padronizados
1	-0,6080	16	-0,8555
2	-0,3398	17	-1,1278
3	1,4418	18	-0,4273
4	0,3046	19	0,9760
5	1,2214	20	-1,6746
6	-0,2400	21	1,7285
7	0,1362	22	-0,0573
8	-1,2171	23	0,6433
9	0,7807	24	1,0196
10	-0,7349	25	0,5311
11	0,9927	26	-0,9845
12	-1,2255	27	-
13	-0,2547	28	-
14	-1,2297	29	-
15	1,2005	30	-

54

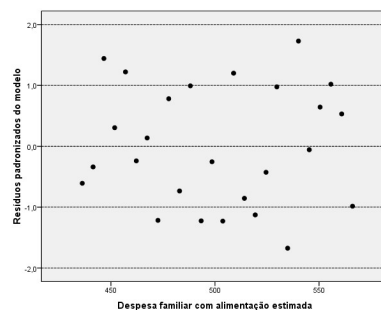
**Figura 4: Gráfico de Dispersão entre as rendas mensais (X) das famílias e os resíduos padronizados do modelo.**



Os resíduos se encontram aleatoriamente distribuídos em torno de zero.

55

**Figura 5: Gráfico de Dispersão entre as despesas estimadas das famílias com alimentação e os resíduos padronizados do modelo.**



Os resíduos se encontram aleatoriamente distribuídos em torno de zero.

56

#### Análise gráfica dos resíduos:

- Caso haja evidências que os quatro valores não pertencem ao conjunto de dados, decorrente de erros de mensuração ou digitação, eles podem ser simplesmente descartados, e a análise baseada no conjunto de dados restantes.
- No caso de dúvidas quanto a este fato, recomenda-se fazer uma análise com e sem os outliers para avaliar a influência desses pontos nos resultados do ajuste do modelo.

57

#### Inadequação do modelo: Violações das Hipóteses Básicas

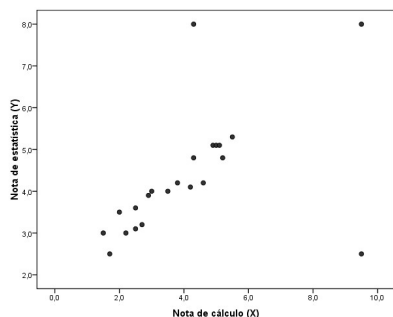
##### Exemplo 3: Outliers e Resíduos não normais

**Tabela 1: Nota de cálculo (X) e nota de estatística (Y) de n=22 alunos de uma mesma turma.**

Aluno	Nota em cálculo	Nota em estatística	Aluno	Nota em cálculo	Nota em estatística
1	1,5	3,0	12	4,2	4,1
2	1,7	2,5	13	4,3	4,8
3	2,0	3,5	14	4,3	8,0
4	2,2	3,0	15	4,6	4,2
5	2,5	3,1	16	4,9	5,1
6	2,5	3,6	17	5,0	5,1
7	2,7	3,2	18	5,1	5,1
8	2,9	3,9	19	5,2	4,8
9	3,0	4,0	20	5,5	5,3
10	3,5	4,0	21	9,5	8,0
11	3,8	4,2	22	9,5	2,5

58

**Figura 1: Gráfico de dispersão entre a nota de cálculo (X) e a nota de estatística (Y).**



Sugere a existência de 3 valores discrepantes (outliers)

59

#### Ajuste do modelo (n=22 alunos):

##### Model Summary<sup>a</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,499 <sup>a</sup>	,249	,212	1,2972

a. Predictors: (Constant), Nota\_calc.X

b. Dependent Variable: Nota\_est.Y

##### ANOVA<sup>a</sup>

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	11,178	1	11,178	6,643	,018 <sup>a</sup>
Residual	33,655	20	1,683		
Total	44,833	21			

a. Predictors: (Constant), Nota\_calc.X

b. Dependent Variable: Nota\_est.Y

##### Coefficients<sup>a</sup>

Model	Unstandardized Coefficients		Standardized Coefficients		t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta				Lower Bound	Upper Bound
1 (Constant)	2,909	,613			4,746	,000	1,630	4,187
Nota_calc.X	,343	,133	,499		2,577	,018	,065	,621

a. Dependent Variable: Nota\_est.Y

60

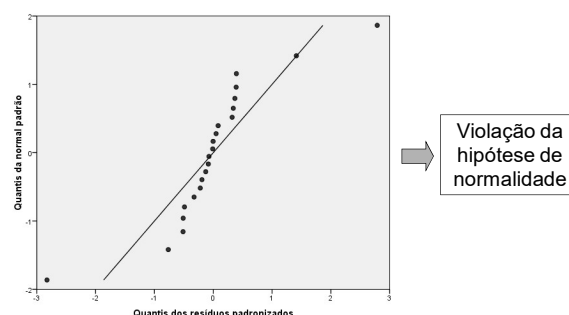
### Inadequação do modelo: Violações das Hipóteses Básicas

#### Exemplo 3: Influência de outliers e resíduos não normais

Tabela 1: Resíduos do modelo – brutos e padronizados

Aluno	Resíduos brutos	Resíduos padronizados	Aluno	Resíduos brutos	Resíduos padronizados
1	-0,4232	-0,3263	12	-0,2494	-0,1922
2	-0,9918	-0,7646	13	0,4163	0,3210
3	-0,0947	-0,0730	14	3,6163	2,7878
4	-0,6633	-0,5114	15	-0,2866	-0,2209
5	-0,6662	-0,5136	16	0,5105	0,3936
6	-0,1662	-0,1281	17	0,4762	0,3671
7	-0,6348	-0,4894	18	0,4419	0,3407
8	-0,0034	-0,0027	19	0,1076	0,0830
9	0,0623	0,0480	20	0,5047	0,3891
10	-0,1093	-0,0842	21	1,8326	1,4128
11	-0,0122	-0,0094	22	-3,6674	-2,8271 <sup>61</sup>

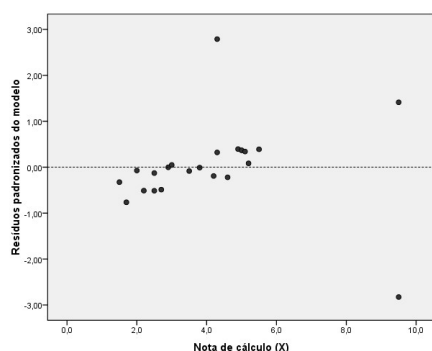
Figura 2: QQ Plot dos resíduos padronizados do modelo.



Violação da hipótese de normalidade

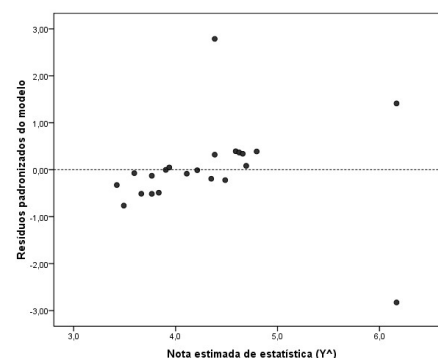
62

Figura 3: Gráfico de dispersão entre as notas de cálculo (X) e os resíduos padronizados do modelo.



63

Figura 4: Gráfico de dispersão entre as notas estimadas de estatística e os resíduos padronizados do modelo.



64

### Análise gráfica dos resíduos:

- Verifica-se violações da hipótese de normalidade dos erros e existência de outliers.
- Próximo passo:** Excluir os 3 valores discrepantes (marcados com o #) e reajustar o modelo.

65

### Inadequação do modelo: Violações das hipóteses básicas

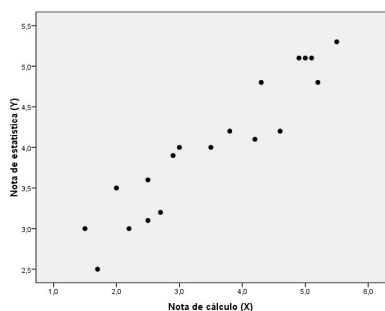
#### Exemplo 3: Influência de Outliers e Resíduos não normais

Tabela 1: Nota de cálculo (X) e nota de estatística (Y) de n=19 alunos de uma mesma turma (sem outliers marcados com #).

Aluno	Nota em cálculo	Nota em estatística	Aluno	Nota em cálculo	Nota em estatística
1	1,5	3,0	12	4,2	4,1
2	1,7	2,5	13	4,3	4,8
3	2,0	3,5	14	4,3	8,0#
4	2,2	3,0	15	4,6	4,2
5	2,5	3,1	16	4,9	5,1
6	2,5	3,6	17	5,0	5,1
7	2,7	3,2	18	5,1	5,1
8	2,9	3,9	19	5,2	4,8
9	3,0	4,0	20	5,5	5,3
10	3,5	4,0	21	9,5	8,0#
11	3,8	4,2	22	9,5	2,5#

66

**Figura 5: Gráfico de Dispersão entre a nota de cálculo (X) e a nota de estatística (Y).**



67

**Ajuste do modelo (n=19 alunos) – sem os outliers:**

Model Summary <sup>a</sup>				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.943 <sup>a</sup>	.890	.883	.2883

a. Predictors: (Constant), Nota\_calc.X

b. Dependent Variable: Nota\_est.Y

ANOVA <sup>b</sup>					
Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	11,384	1	11,384	137,013	,000 <sup>a</sup>
Residual	1,413	17	,083		
Total	12,797	18			

a. Predictors: (Constant), Nota\_calc.X

b. Dependent Variable: Nota\_est.Y

Coefficients <sup>a</sup>							
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
1 (Constant)	1,869	,196		9,543	,000	1,456	2,282
Nota_calc.X	,611	,052	,943	11,705	,000	,501	,721

a. Dependent Variable: Nota\_est.Y

68

### Inadequação do modelo: Violações das hipóteses básicas

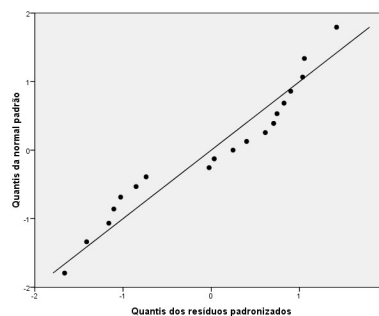
#### Exemplo 2: Influência de outliers e resíduos não normais

**Tabela 1: Resíduos do modelo – brutos e padronizados**

Aluno	Resíduos brutos	Resíduos padronizados	Aluno	Resíduos brutos	Resíduos padronizados
1	0,2149	0,7454	12	-0,3347	-1,1611
2	-0,4073	-1,4131	13	0,3042	1,0554
3	0,4094	1,4202	14	#	#
4	-0,2128	-0,7383	15	-0,4791	-1,6619
5	-0,2961	-1,0272	16	0,2377	0,8245
6	0,2039	0,7074	17	0,1766	0,6126
7	-0,3183	-1,1042	18	0,1155	0,4006
8	0,2595	0,9004	19	-0,2456	-0,8521
9	0,2985	1,0354	20	0,0711	0,2467
10	-0,0070	-0,0244	21	#	#
11	0,0097	0,0336	22	#	#

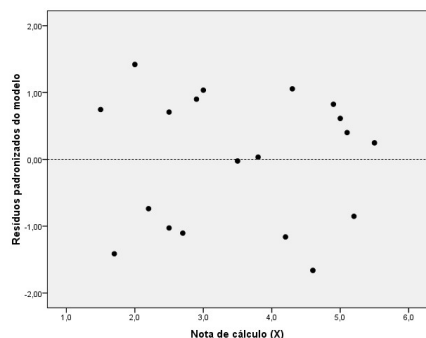
69

**Figura 6: QQ Plot dos resíduos padronizados do modelo.**



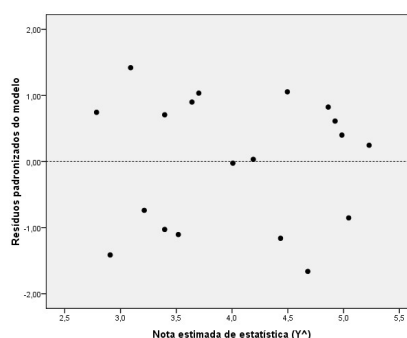
70

**Figura 7: Gráfico de dispersão entre as notas de cálculo (X) e os resíduos padronizados do modelo (r<sup>P</sup>).**



71

**Figura 8: Gráfico de dispersão entre as notas estimadas de estatística e os resíduos padronizados do modelo (r<sup>P</sup>).**



72

**Análise gráfica dos resíduos:**

- ☐ **Pergunta 1:** Quais as conclusões extraídas por você a partir dos gráficos construídos ?



- ☐ **Pergunta 2:** Aplicando alguma transformação na variável resposta Y, a distribuição dos resíduos se aproximaria mais da normal ?

73