



UNIVERSIDADE DE SÃO PAULO

Faculdade de Zootecnia e Engenharia de Alimentos

Mestrado Profissional

Gestão e Inovação na Indústria Animal

**GIA5005 – Estatística Aplicada a Administração**

Prof. César Gonçalves de Lima

[cegdlima@usp.br](mailto:cegdlima@usp.br)

## CORRELAÇÃO E REGRESSÃO LINEAR

### Problemas:

- Podemos calcular o valor de uma grandeza a partir do conhecimento do valor de outras grandezas?
- Como conseguimos quantificar o grau de relacionamento entre duas ou mais variáveis? Elas variam de forma independente ou não?

**Início do estudo:** Construir um **gráfico de dispersão**, plotando os valores das variáveis, duas-a-duas, procurando visualizar alguma relação funcional entre elas.

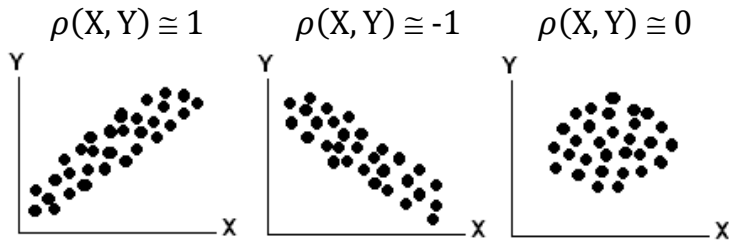
## CORRELAÇÃO LINEAR SIMPLES

O **coeficiente de correlação linear** é uma medida do grau de relacionamento linear entre duas variáveis quantitativas  $X$  e  $Y$  e é definido por:

$$\rho(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X) var(Y)}}, \quad -1 \leq \rho(X, Y) \leq 1$$

- $\rho(X, Y) > 0$  indica que os valores de  $X$  e  $Y$  crescem no mesmo sentido ou que são grandezas diretamente proporcionais.
- $\rho(X, Y) < 0$  indica que os valores  $X$  e  $Y$  crescem em sentidos opostos ou que são grandezas inversamente proporcionais.
- $\rho(X, Y) = 0$  indica que não existe qualquer relação de dependência linear entre estas variáveis.

Situações típicas:



**Figura 11.** Gráficos de dispersão e coeficientes de correlação.

O coeficiente de correlação linear de Pearson é estimado por:

$$r(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad -1 \leq r(X, Y) \leq 1$$

Não se preocupe que as calculadoras científicas e os softwares estatísticos realizam esses cálculos.

---

## Teste de independência das variáveis X e Y

Hipóteses:  $H_0: \rho(X,Y) = 0$  (as variáveis são independentes)

$H_1: \rho(X,Y) \neq 0$  (as variáveis são dependentes)

Estatística:  $T = \frac{r(X,Y)\sqrt{n-2}}{\sqrt{1-r^2(X,Y)}}$  que sob  $H_0$  tem distribuição  $t_{(n-2)}$ .

**Exemplo.** Testar a hipótese de que o peso médio de coelhos ao abate (Y), em quilogramas, é independente do tamanho de ninhada (X), com base nos seguintes dados:

X	4	8	6	1	7	3	7	5
Y	2.125	1.980	2.270	2.300	1.880	2.320	1.860	2.050

Para calcular o coeficiente de correlação fazemos **Stat > Estatísticas Básicas > Correlação**, em **Variáveis** selecionamos as colunas **Ninhada** e **Peso ao Abate**, escolhemos o **Método > Correlação de Pearson** e marcamos **Exibir valor-p**.

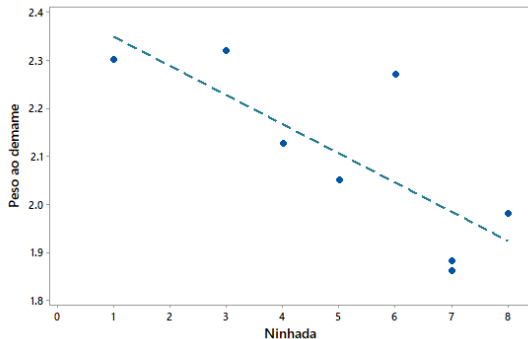


Gráfico de dispersão do peso ao abate (kg) e tamanho de ninhada de coelhos.

Note que os pontos indicam uma relação linear entre as variáveis.

## Correlação: Ninhada; Peso ao abate

Correlações	
Correlação de Pearson	-0.776
Valor-P	0.024

O coeficiente de correlação negativo (-0,776) confirma a existência da dependência linear negativa e relativamente alta entre as variáveis, indi-

cando que elas são inversamente proporcionais: quanto maior o tamanho da ninhada menor deve ser o peso médio ao abate.

Como Valor- $p = 0,024 < 0,05$ , rejeitamos  $H_0: \rho(X, Y) = 0$  e concluímos que existe uma dependência linear negativa e estatisticamente significativa entre o peso médio de coelhos ao abate e o tamanho da ninhada.



## REGRESSÃO LINEAR SIMPLES

**Objetivo:** Procura expressar uma possível relação de dependência entre duas variáveis sob a forma de uma equação matemática.

A equação de regressão pode ser um polinômio (uma reta, parábola ou um polinômio de grau mais elevado), uma função do tipo exponencial (curva logística, von Bertalanfy, de Gompertz etc.) *etc.*

Vamos estudar o **ajuste de uma reta** em problemas envolvendo somente duas variáveis:

$y$ : variável resposta ou variável dependente

$x$ : variável regressora, covariável ou variável independente.

## Modelos de Regressão linear:

**Simples:**  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$$y_i = \beta_0 + \beta_1 \sqrt{x_i} + \varepsilon_i \quad \dots$$

**Múltipla:**  $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i$

$$y_i = \beta_0 + \beta_1 \sqrt{x_{1i}} + \beta_2 \ln(x_{2i}) + \beta_3 \left(\frac{1}{x_{3i}}\right) + \varepsilon_i$$

## Modelos de Regressão não linear:

$$y_i = \beta_0 + e^{\beta_1 x_i} + \varepsilon_i \qquad y_i = \frac{\beta_0}{1 + \beta_1 x_i} + \varepsilon_i$$

$$y_i = \beta_0 (1 - e^{\{-\beta_1 x_i\}}) + \varepsilon_i \quad (\text{Modelo de Mitscherlich}) \quad \dots$$

Para visualizar a relação funcional entre X e Y usamos os **gráficos de dispersão**: a distribuição dos pontos no gráfico pode sugerir qual função explica melhor o comportamento dos dados.

**Exemplo 6.3** Determinar a reta que relaciona a Absorbância (Y) com a concentração de nitrito (X, em mg/100ml) em amostras de mortadela. Os dados experimentais são:

$x_i$	0.5	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0
$y_i$	0.040	0.078	0.145	0.215	0.300	0.340	0.395	0.460	0.560	0.715

Desenhar um gráfico de dispersão para visualizar a (possível) relação funcional entre as duas variáveis.

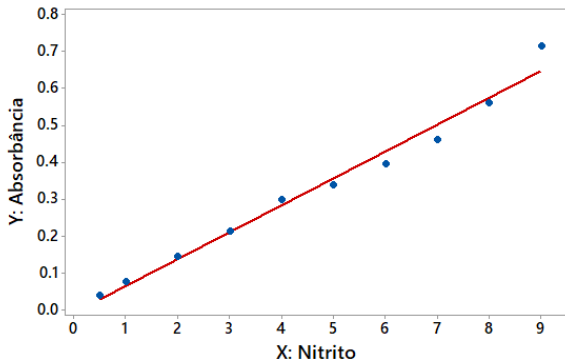


Gráfico de dispersão dos dados de absorbância e quantidade de nitrito

Note que uma reta ( $y = a + bx$ ) parece explicar bem a relação entre as duas variáveis.

## O MODELO DE REGRESSÃO LINEAR SIMPLES

A partir de uma amostra de  $n$  pares de valores  $(x_1, y_1), \dots, (x_n, y_n)$  nós podemos estabelecer o modelo de regressão linear simples:

$$y_i = a + bx_i + \varepsilon_i$$

Em que  $y_i$  é o valor da variável resposta na amostra  $i$ ,  $x_i$  é o valor da variável regressora na amostra  $i$ ,  $a$  e  $b$  são os parâmetros da reta e  $\varepsilon_i$  é um **erro** associado aos valores  $y_i$ .

Os **erros**  $\varepsilon_i$ 's podem resultar de erros de medidas, de digitação, da heterogeneidade das matérias primas utilizadas no experimento, da imprecisão dos equipamentos utilizados nas medições *etc.*

Ao estabelecer o modelo de regressão linear simples nós admitimos que:

- a) A relação entre as variáveis  $X$  e  $Y$  é linear.
- b) Os valores da variável  $X$  não são sujeitos a erros (são valores fixos, escolhidos pelo pesquisador).
- c) A média dos erros é nula, isto é,  $E(\varepsilon_i) = 0$ .
- d)  $var(\varepsilon_i) = \sigma^2$  (constante) e não depende dos valores  $x_i$
- e) A correlação entre os erros de duas observações é nula, isto é,  $corr(\varepsilon_i, \varepsilon_i^*) = 0$ , para  $i \neq i^*$ .
- f) Os erros têm distribuição normal, isto é,  $\varepsilon_i \sim N(0, \sigma^2)$ .

**Estimação dos parâmetros:** é feita utilizando o Método dos Mínimos Quadrados (MMQ), que consiste em obter as estimativas dos parâmetros  $a$  e  $b$  que **minimizam** a soma dos quadrados dos erros:

$$\hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}} \quad \text{e} \quad \hat{a} = \bar{y} - \hat{b}\bar{x}$$

Não se preocupe que calculadoras científicas e softwares estatísticos realizam esses cálculos.

- A **reta ajustada** pode ser escrita como:  $\hat{y}_i = \hat{a} + \hat{b}x_i$
- O **resíduo** da regressão pode ser calculado por:  $\hat{\varepsilon}_i = y_i - \hat{y}_i$

**Modelo bem ajustado  $\Leftrightarrow$  resíduos pequenos ( $\hat{\varepsilon}_i \cong 0$ )**

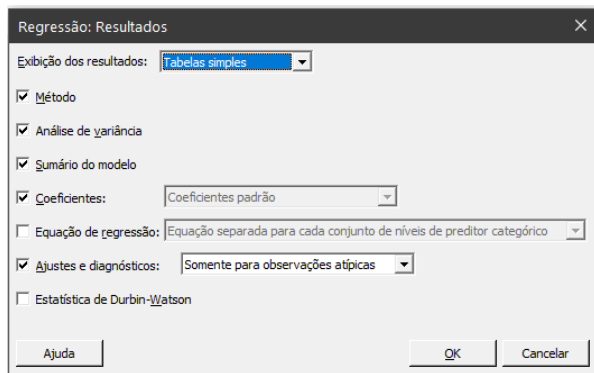
A **qualidade do ajuste** de regressão pode ser avaliada por meio de gráficos de resíduos e do **coeficiente de determinação**:

$$R^2 = (\hat{b})^2 \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = [r(X, Y)]^2$$

- O coeficiente de determinação mede a porcentagem da variação em  $y$ , que é explicada pelo modelo de regressão.
- Como  $0 \leq R^2 \leq 1$ , o ajuste do modelo será tanto melhor quanto mais próximo de 1 estiver o valor de  $R^2$ .



Para ajustar a reta relacionando a Absorbância e a concentração de Nitrito em **Stat > Regressão > Regressão > Ajuste de modelo de regressão**, em **Respostas** selecionar a coluna **Y: Absorbância** e em **Preditores contínuos**, a coluna **X: Nitrito**.



Em **Resultados** podemos escolher quais resultados devem ser exibidos.

Vamos marcar: **Sumário do modelo**, **Coeficientes** e em **Ajustes e diagnósticos** escolher **Somente para observações atípicas**.

#### ↵ Análise de Regressão: Y: Absorbância versus X: Nitrito

##### Sumário do Modelo

S	R2	R2(aj)	R2(pred)
0.0327109	97.95%	97.70%	96.00%

##### Coeficientes

Termo	Coef	EP de Coef	Valor-T	Valor-P	VIF
Constante	-0.0044	0.0198	-0.22	0.830	
X: Nitrito	0.07235	0.00370	19.56	0.000	1.00

##### Ajustados e Diagnósticos para Observações Atípicas

Obs.	Y:		Resíd	Pad	R
	Absorbância	Ajuste			
10	0.7150	0.6468	0.0682	2.59	R

R Resíduo grande

No **Sumário do modelo** o valor  $R^2 \cong 98\%$  indica um ótimo ajuste da reta.

O coeficiente angular positivo da reta (0.07235) indica que as variáveis são diretamente proporcionais e que “ao aumento de 1 *mg* de nitrito na amostra de mortadela corresponde um aumento de 0.07235 unidades na absorbância”.

Ao final, o Minitab indica a observação 10 como uma observação atípica (*outlier*).

## Como avaliara as observações atípicas?

- Quando o ajuste da reta aos dados é muito bom, espera-se que os resíduos do ajuste,  $\hat{\varepsilon}_i = y_i - \hat{y}_i$ , sejam pequenos.
- É difícil decidir se os valores numéricos de  $\hat{\varepsilon}_i$  são pequenos ou grandes!

**Exemplo:** Um resíduo de 2kg num estudo de regressão envolvendo bovinos adultos é pequeno, mas num estudo envolvendo coelhos, é muito grande!

- A solução é trabalhar com **resíduos padronizados** ( $res_{pad}$ ), que teoricamente, têm distribuição normal e a chance de encontrarmos valores superiores a 2 é bem pequena!

- Um tipo de resíduo padronizado consiste em dividir cada um dos resíduos por uma estimativa do seu erro padrão, ou seja,

$$res_{pad(i)} = \frac{\hat{\varepsilon}_i}{ep(\hat{\varepsilon}_i)}$$

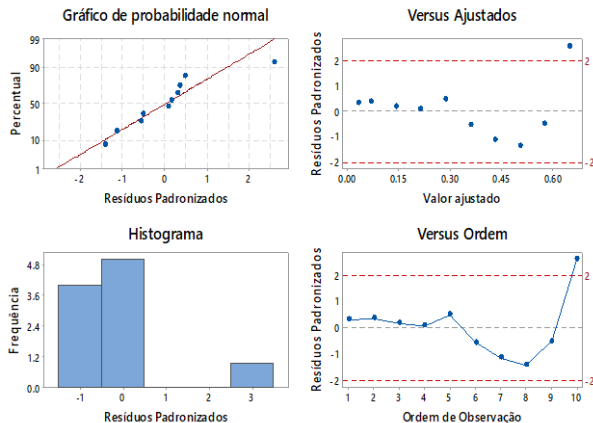
**Regra:** Se  $|res_{pad(i)}| > 2$  indica que o ponto  $(x_i, y_i)$  é um sério candidato a *outlier* (valor atípico, discrepante)

## Dúvidas sobre a observação 10 do exemplo:

- O valor de absorbância deste ponto (0.7150) é plausível?
- Devo descartar este ponto e refazer a análise?

Para visualizar o problema, depois de especificarmos o modelo e antes de rodar a análise, podemos escolher em **Gráficos > Resíduos para gráficos > Padronizado** e em **Gráficos de resíduos** marcar **Quatro em um**.

### Gráficos de Resíduo de Y: Absorbância



**Ideal:** Pontos no **Gráfico de probabilidade normal** estejam próximos da reta vermelha. Pontos distantes da reta são candidatos a *outlier*.

Nos gráficos **Versus Ordem** e **Versus Ajustados** aparece um ponto com resíduo superior a 2 (observação 10)

Como temos poucos pontos, fica mais difícil avaliar a qualidade do histograma.

## DIAGNÓSTICOS EM REGRESSÃO

Na literatura são encontrados três tipos de resíduos: resíduo ordinário, resíduo padronizado e resíduo estudentizado.

### a) Resíduo ordinário:

$$r_i = \hat{\varepsilon}_i = y_i - \hat{y}_i$$

O uso de  $r_i$  em diagnósticos pode ser inadequado devido à sua heterogeneidade de variâncias.

### b) Resíduo padronizado internamente (*Studentized residual*):

$$rsi_i = \hat{\varepsilon}_i / \sqrt{\widehat{var}(\hat{\varepsilon}_i)}$$

Onde  $\widehat{var}(\hat{\varepsilon}_i) = (1 - h_{ii})s_{y/x}^2$ , em que  $h_{ii}$  é o *leverage* da observação  $i$  e  $s_{y/x}^2 = QMResiduo$  é uma estimativa de  $\sigma^2$ .

Estes resíduos ( $rsi_i$ ) são sensíveis à presença de valores atípicos, mas são mais indicados que os resíduos ordinários.

c) **Resíduo estudentizado externamente** (ou *Jackknife residual* ou *RStudent*):

$$rse_i = \hat{\varepsilon}_i / \sqrt{\widehat{var}_{(i)}(\hat{\varepsilon}_i)}$$

em que  $\widehat{var}_{(i)}(\hat{\varepsilon}_i)$  é a estimativa da variância residual calculada sem a observação  $i$ .

Esses resíduos são muito importantes na detecção de pontos discrepantes (*outliers*).



De um modo geral:

- Uma observação é chamada **inconsistente** quando ela destoa da tendência das demais.
- Uma observação é dita **influyente** quando a sua omissão do conjunto de dados resulta em grandes mudanças nas estimativas dos coeficientes de regressão, variância residual, valor- $p$ ,  $R^2$  etc.

Quando a observação está distante das outras ela pode ser considerada influente ou não.

## Estatísticas usadas na identificação de pontos atípicos:

- Medida de *leverage* (alavanca):  $h_{ii}$
- Medida de inconsistência:  $rse_i$
- Medidas de influência geral:  $DFFitS_{(i)}$ ,  $C_{(i)}$  ou  $D_{(i)}$
- Medida de influência sobre os parâmetros:  $DFBetaS_{(i)}$  para  $\beta_j$

Para utilizar as estatísticas no diagnóstico da regressão vamos admitir que:

$p = k + 1$  é o número de parâmetros do modelo

$k$  é o número de variáveis do modelo de regressão

$n$  é o número de indivíduos usados no estudo

**a)  $h_{ii}$  (*leverage*):** avalia a distância de uma observação às demais.

**Regra:** Valores  $h_{ii} \geq 2p/n$  indicam observações que merecem atenção.

**b)  $DFBeta_{(i)}$ :** mede a alteração nas estimativas dos parâmetros ao se retirar a observação  $i$ . É importante quando o coeficiente de regressão tem um significado prático.

**Regra:** Observação com  $DFBeta_{(i)} > 2$  pode ser considerada influente.

**c)  $DFFitS_{(i)}$ :** mede a alteração provocada no valor ajustado pela retirada da observação  $i$ .

**Regra:** Valores  $|DFFitS_{(i)}| > 2\sqrt{p/n}$  podem identificar observações influentes e merecem atenção.

d)  **$D_{(i)}$  (Distância de Cook)**: também mede a alteração nas estimativas dos parâmetros ao se retirar a observação  $i$ . Usa o resíduo estudentizado internamente ( $rsi_i$ ).

**Regra:** Valores extremamente altos de  **$D_{(i)}$**  em relação aos outros podem indicar pontos atípicos em  $x$  ou  $y$ .

**De um modo geral um ponto  $(x_i, y_i)$  é considerado:**

- **Ponto inconsistente:** se apresenta resíduo estudentizado alto, isto é, se  $|rse_i| \geq t_{\{\frac{\gamma}{2n}; n-p-1\}}$ , o percentil da distribuição  $t$  com nível de significância  $100 \gamma\%$ .
- **Ponto de alavanca (*leverage*):** se apresenta  $h_{ii} \geq 2p/n$ . Pode ser classificado como *bom* quando for consistente, ou *ruim*, quando for inconsistente.
- ***Outlier*:** ponto inconsistente, mas com *leverage* baixo, ou seja, ponto com  $rse_i$  alto e  $h_{ii}$  baixo.
- **Ponto influente:** ponto com  $DFFitS_{(i)}$ ,  $DFBetaS_{(i)}$  ou  $D_{(i)}$  alto.

No Exemplo 6.3, que relaciona a absorvância (Y) com a concentração de nitrito (X) em amostras de mortadela, temos:

Obs	Nitrito	Absorb	Ajustado	Res	rsi	rse
1	0.5	0.040	0.03178	0.008219	0.30240	0.28450
2	1.0	0.078	0.06796	0.010043	0.35719	0.33682
3	2.0	0.145	0.14031	0.004693	0.15874	0.14873
4	3.0	0.215	0.21266	0.002343	0.07682	0.07189
5	4.0	0.300	0.28501	0.014993	0.48417	0.45969
6	5.0	0.340	0.35736	-0.017358	-0.56015	-0.53456
7	6.0	0.395	0.42971	-0.034708	-1.13553	-1.15976
8	7.0	0.460	0.50206	-0.042058	-1.41706	-1.53163
9	8.0	0.560	0.57441	-0.014408	-0.50935	-0.48438
10	9.0	0.715	0.64676	0.068241	2.59390	6.08570

**Conclusão:** O ponto (9; 0,715) é considerado **inconsistente** e sério candidato a **outlier**, porque o resíduo padronizado internamente ( $rsi_i$ ) e o padronizado externamente ( $rse_i$ ) são superiores a 2.

**Análise de Variância**

Fonte	GL	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	1	0.247500	0.247500	1273.23	0.000
X: Nitrito	1	0.247500	0.247500	1273.23	0.000
Erro	7	0.001361	0.000194		
Total	8	0.248860			

Podemos excluir a observação 10 do conjunto de dados e refazer a análise, resultando em:

**Sumário do Modelo**

S	R2	R2(aj)	R2(pred)
0.0139423	99.45%	99.38%	99.05%

A qualidade do ajuste da reta é excelente!

**Coefficientes**

Termo	Coef	EP de Coef	Valor-T	Valor-P	VIF
Constante	0.01236	0.00886	1.40	0.205	
X: Nitrito	0.06635	0.00186	35.68	0.000	1.00

O intercepto da reta pode ser considerado nulo e deve ser excluído do modelo.

Para ajustar a reta passando pela origem em **Stat > Regressão > Regressão > Ajuste de modelo de regressão > Modelo** desmarque a opção **Incluir o termo constante no modelo** que aparece no canto inferior esquerdo da janela.

#### Sumário do Modelo

S	R2	R2(aj)	R2(pred)
0.0147452	99.82%	99.80%	99.75%

O  $R^2 = 99,8\%$  indica um ajuste quase perfeito da reta.

#### Coeficientes

Termo	Coef	EP de Coef	Valor-T	Valor-P	VIF
X: Nitrito	0.06856	0.00103	66.45	0.000	1.00

O Valor-p < 0,05 associado ao coeficiente angular da reta indica que ele é significativamente não nulo.



A reta ajustada fica:  $\widehat{Absorbancia} = 0,06856 * Nitrito$



Podemos usar essa equação para **estimar** a quantidade de nitrito em uma nova amostra de mortadela depois de prepará-la em laboratório e medido o seu valor de absorbância no espectrofotômetro,

$$\widehat{Nitrito} = \frac{Absorbancia}{0,06856}$$

**Exemplo:** Necessidade de realizar um diagnóstico detalhado do ajuste de um modelo de regressão linear a um conjunto de dados.

$x_1$	$y_1$	$x_2$	$y_2$	$x_3$	$y_3$	$x_4$	$y_4$
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.74
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.50
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

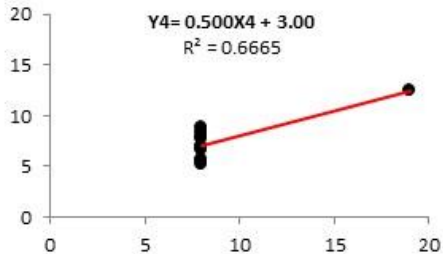
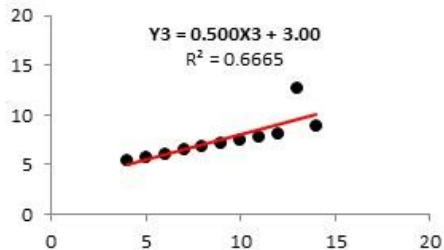
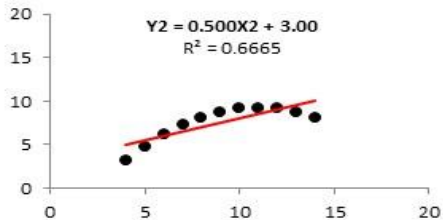
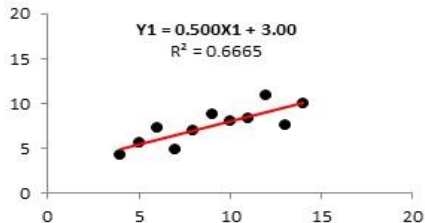
**Fonte:** Chatterjee, S; Price, B. **Regression analysis by Example**. Wiley, 1977, pág. 8.

Conjunto	Reta ajustada	$R^2$
1	$\hat{y}_1 = 3,00 + 0,500x_1$	0,67
2	$\hat{y}_2 = 3,00 + 0,500x_2$	0,67
3	$\hat{y}_3 = 3,00 + 0,500x_3$	0,67
4	$\hat{y}_4 = 3,00 + 0,500x_4$	0,67

### O que podemos comentar sobre os ajustes das retas?

- Pelos valores de  $R^2$  associados a cada ajuste, tem-se a ideia que os ajustes foram bons em todos os conjuntos de dados.

Vamos avaliar (pelo menos!) os gráficos de dispersão dos dados para confirmar (ou não) a “boa” qualidade do ajuste das retas nos quatro conjuntos de dados:



Percebe-se que:

- O ajuste da reta no conjunto 1 (Y1) foi bom e a reta representa bem o comportamento da relação entre as duas variáveis.
- No caso do conjunto 2 (Y2), percebe-se que a relação funcional entre as duas variáveis pode ser mais explicada por uma parábola (polinômio de 2º grau) e não por uma reta.
- O conjunto 3 (Y3) apresenta um ponto atípico (penúltima observação), cuja ordenada destoa das ordenadas dos demais pontos. A presença deste ponto altera (aumenta) a inclinação da reta.
- Não se percebe claramente no conjunto 4 (Y4) qualquer relação funcional entre os seus valores. Só a presença de um ponto bastante atípico (19, 12.50), que destoa muito dos demais pontos, é que possibilitou o ajuste da reta.

**Conclusão:** É muito perigoso achar que o ajuste foi bom baseando-se, exclusivamente no valor do  $R^2$ .

### **Segundo Chatterjee:**

- O método dos mínimos quadrados é bastante robusto.
- Pequenas violações das pressuposições não invalidam as inferências ou as conclusões tiradas da análise, mas violações grosseiras podem distorcer seriamente as conclusões.

**Alerta:** Devemos estar atentos às violações das pressuposições do modelo de regressão, procurando corrigi-las da melhor forma possível.

**IMPORTANTE:** Podemos usar a equação da reta ajustada,  $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$  para fazer **previsões**.

**Estimação pontual:** basta substituir na equação ajustada o valor de  $x_i$  e obter  $\hat{y}_i$ .

**Estimação por intervalo:**

- 1) Intervalo de confiança** para o valor médio  $E(y_i)$
- 2) Intervalo de previsão** para uma nova observação  $y_h$ , relativa a um valor  $x_h$  da variável regressora, que não foi utilizado no estudo, mas está **dentro** dos limites dessa variável.
- 3) Intervalo de predição** para uma nova observação  $y_h$ , relativa a um valor  $x_h$  da variável regressora que não foi utilizado no estudo e está **fora** dos limites dessa variável.

**Exemplo:** No exemplo da mortadela estimar a absorvência de amostras com 5,3 e 8,6 mg/100ml de nitrito, utilizando o modelo final sem intercepto:  $\widehat{Absorvência} = 0,06856 * Nitrito$

Em **Em Stat > Regressão > Regressão > Predizer** escolher **Inserir valores individuais** e nas linhas abaixo de **X: Nitrito** digitar 5.3 e 8.6

Configurações		Predição			
Variável	Configuração	Ajuste	EP do Ajustado	IC de 95%	IP de 95%
X: Nitrito	5.3	0.363	0.0055	(0.351; 0.376)	(0.327; 0.400)

Temos 95% de confiança que a verdadeira Absorvência de amostras com 5,3 mg/100ml de nitrito está entre 0,351 e 0,376.



### Configurações

Variável	Configuração
X: Nitrito	8,6

### Predição

Ajuste	EP do Ajustado	IC de 95%	IP de 95%
0.590	0.0089	(0.569; 0.610)	(0.550; 0.629) X

*X denota um ponto atípico relativo aos níveis dos preditores usados para ajustar o modelo.*

Note que  $x = 8,6$  está fora do intervalo de valores de Nitrito usados no ajuste da reta.

Usando o intervalo de previsão: temos 95% de confiança que a verdadeira absorbância de uma amostra com 8,6 mg/100ml de nitrito está entre 0,550 e 0,629.

Para maiores detalhes sobre **Diagnóstico em Regressão** consultar, por exemplo:

**Apostila “Modelos de Regressão e Covariância”**

Autores: Clarice Garcia Borges Demétrio e Silvio S. Zocchi  
ESALQ/USP

<https://docs.ufpr.br/~taconeli/CE22518/regressao.pdf>