

Regressão

Regressão Linear: Nesta seção o conceito é a descrição da relação entre duas variáveis através do gráfico e da equação da reta que melhor descreve os dados, a Reta de Regressão. A equação é chamada Equação de Regressão.

$$\hat{y} = b_0 + b_1x$$

b_0 e b_1 são estatísticas amostrais para estimarem-se os parâmetros populacionais β_0 e β_1 .

Regressão

	Parâmetro Populacional	Estatística Amostral
Intercepto	β_0	b_0
Inclinação	β_1	b_1
Equação	$y = \beta_0 + \beta_1 x$	$\hat{y} = b_0 + b_1 x$

Determinação das estatísticas

Inclinação $b_1 = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$

Intercepto $b_0 = \bar{y} - b_1 \bar{x}$

Regressão

Exemplo:

Anos	Quantidade (Q) (1.000 un.)	Preço (P) (\$ 1.000)	x^2	xy
1	2	4	16	8
2	1	6	36	6
3	3	3	9	9
4	1	5	25	5
5	4	1	1	4
6	3	2	4	6

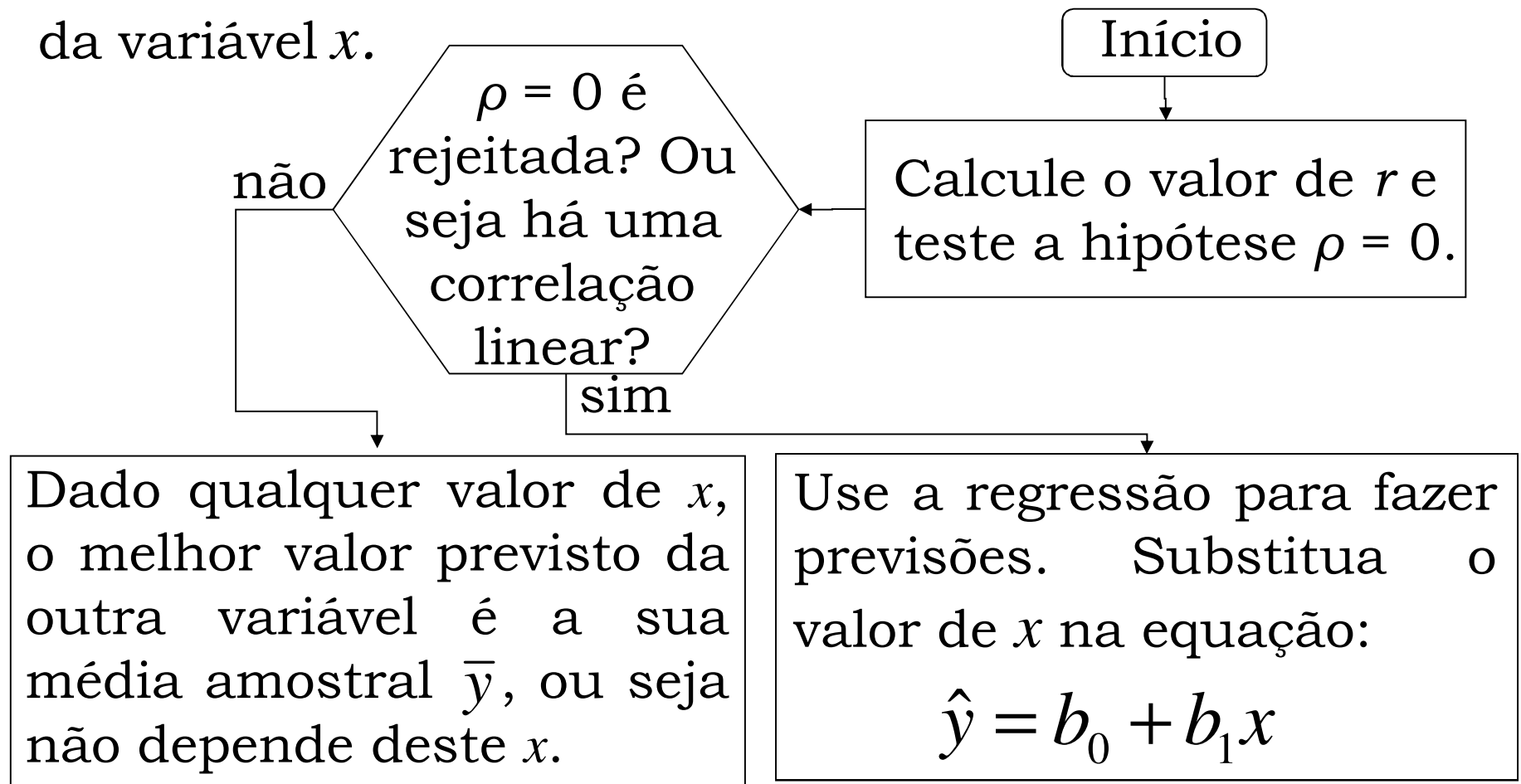
14 21 91 38

$$b_1 = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} = \frac{-66}{105} = \boxed{-0,6286}$$

$$b_0 = \bar{y} - b_1\bar{x} = \boxed{4,5333}$$

Uso da regressão para previsão

A equação de regressão pode ser útil para prever o valor de uma variável, dado algum valor particular da outra variável. Importante: não ultrapasse os limites da variável x .



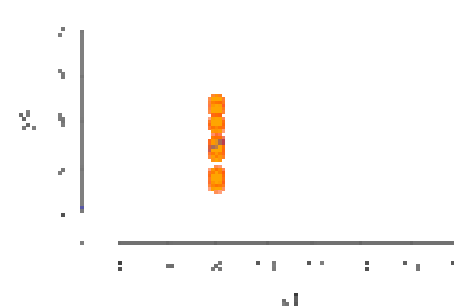
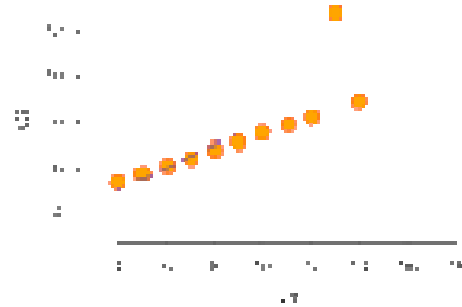
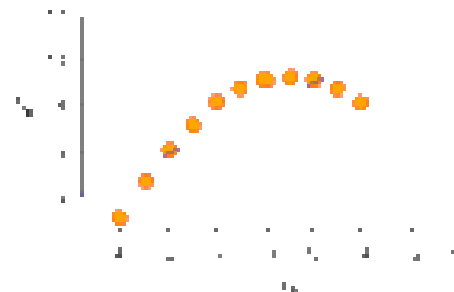
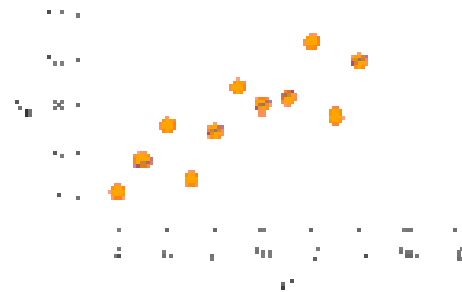
Dado qualquer valor de x , o melhor valor previsto da outra variável é a sua média amostral \bar{y} , ou seja não depende deste x .

Use a regressão para fazer previsões. Substitua o valor de x na equação:

$$\hat{y} = b_0 + b_1x$$

Cuidados: Observe o Quarteto de Anscomb

Anscombe's Quartet							
I		II		III		IV	
x	y	x	y	x	y	x	y
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	7,71
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



Estatística com HP12C

Para dados **univariados**, para cada observação:

Digite: x , $\Sigma+$

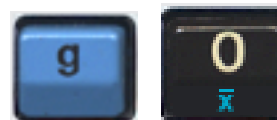
Para dados **bivariados**, para cada par

Digite: y , [enter], x , $\Sigma+$

Anos	Quantidade (Q) (1.000 un.)	Preço (P) (\$ 1.000)
1	2	4
2	1	6
3	3	3
4	1	5
5	4	1
6	3	2

Registros	Valor
R_1	n
R_2	Σx
R_3	Σx^2
R_4	Σy
R_5	Σy^2
R_6	Σxy

Média de x :



Média de y :



Desvio padrão x :



Desvio padrão y :



Correlação Linear:

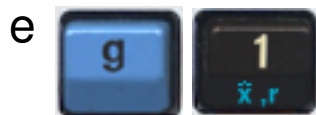


Estimativa linear:

\hat{y} : **digite um valor de x**



\hat{x} : **digite um valor de y**



Anos	Quantidade (Q) (1.000 un.)	Preço (P) (\$ 1.000)	x^2	xy
1	2	4	16	8
2	1	6	36	6
3	3	3	9	9
4	1	5	25	5
5	4	1	1	4
6	3	2	4	6
	14	21	91	38

$$b_1 = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} = \frac{-66}{105} = \boxed{-0,6286}$$

$$b_0 = \bar{y} - b_1\bar{x} = \boxed{4,5333}$$

$$\hat{y} = b_0 + b_1x$$

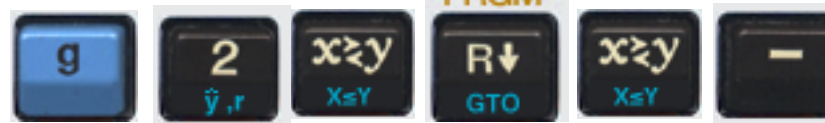
Determine b_0

Estimativa de y para x = 0: **Digite 0** e



Na seqüência, determine b_1

Digite 1 em seguida:



Regressão Múltipla (Seção 10.5)

A tabela abaixo apresenta as séries históricas (15 meses) de valores de Custos Indiretos de Fabricação (CIF), Horas de Mão-de-Obra Direta (HMOD) e Horas-Máquina (HM). Fonte: FIPECAFI – Contabilometria, Atlas, 2004.

Período	CIF	HMOD	HM
1	350	4	10
2	400	8	14
3	470	12	16
4	550	10	26
5	620	15	31
6	380	7	12
7	290	6	13
8	490	10	21
9	580	11	26
10	610	13	24
11	560	12	23
12	420	8	12
13	450	11	19
14	510	12	19
15	380	5	11

O modelo de regressão Múltipla é uma extensão lógica do Modelo de regressão Simples.

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

Em geral:

No caso do exemplo da tabela ao lado:

$$\hat{CIF} = b_0 + b_1 HMOD + b_2 HM$$

Resolução pelo Excel:

Ferramentas – Análise de Dados – **Correlação**

Ferramentas – Análise de Dados – **Regressão**

Regressão Múltipla (Seção 10.5)

RESUMO DOS RESULTADOS

$$\hat{C}IF = b_0 + b_1 HMOD + b_2 HM$$

<i>Estatística de regressão</i>	
R múltiplo	0,941
R-Quadrado	0,885
R-quadrado ajustado	0,866
Erro padrão	36,827
Observações	15

$$\hat{C}IF = 184,884 + 11,746 HMOD + 9,369 HM$$

ANOVA

	<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>F de sig.</i>
Regressão	2	125218,94	62609,47	46,165	0,000
Resíduo	12	16274,39	1356,20		
Total	14	141493,33			

	<i>Coefficientes</i>	<i>Erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>	<i>95% inf.</i>	<i>95% sup.</i>
Interseção	184,884	31,762	5,821	0,000	115,68	254,09
HMOD	11,746	5,835	2,013	0,067	-0,97	24,46
HM	9,369	2,825	3,317	0,006	3,21	15,52

Regressão Múltipla (Seção 10.5)

Exemplo do livro FIPECAFI, 2007. Utilização do Software SPSS

Base de dados: Exame – 500 Melhores & Maiores, ano 2001

Variáveis - Indicadores Financeiros:

RENTAT:	Rentabilidade do Ativo
RENTPL:	Rentabilidade do Patrimônio Líquido
ALOPER:	Alavancagem Operacional
MARVEN:	Margem de Vendas
ALFIN:	Alavancagem Financeira
LUPRE:	Lucro ou Prejuízo

Variável dependente RENTPL

Objetivo: Descobrir as relações mais relevantes através de correlações e regressões.

No SPSS: **Correlação:** Correlate – Bivariate – Pearson

Regressão: Regression – Linear (Method: stepwise)