UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
CADERNOS DE MATEMÁTICA E ESTATÍSTICA
SÉRIE B: TRABALHO DE APOIO DIDÁTICO

ANÁLISE DE REGRESSÃO E CORRELAÇÃO

JOÃO RIBOLDI
DINARA WESTPHALEN XAVIER FERNANDEZ

SÉRIE B, Nº 30
PORTO ALEGRE, JUNHO 1995

P R E F Á C I O

As presentes notas destinam-se ao apoio didático da disciplina AGRP01 - Análise Estatística dos Cursos de Pós-Graduação em Agronomia. Surgiram da experiência acumulada ao longo dos anos e tem por objetivo servir como um guia aos conteúdos abordados e não como um limitante dos assuntos, não prescindindo, evidentemente, da consulta de bibliografia especializada para complementacão.

Apesar de serem de objetivo especifico, podem também servir como texto de apoio didático a outras disciplinas a nível de graduação e pós-graduação.

Agradecemos a todos que colaboraram na organização destas notas e em especial aos bolsistas Stela, Flávio e Anna Christina pelo trabalho de digitação.

Porto Alegre, 17 de Junho de 1995.

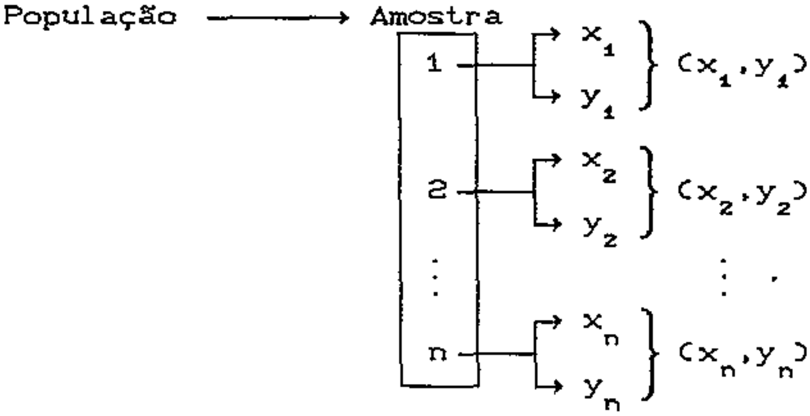
Prof. João Riboldi

Í N D I C E

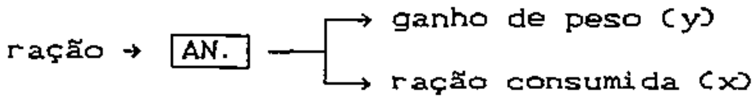
ANÁLISE DE REGRESSÃO E CORRELAÇÃO

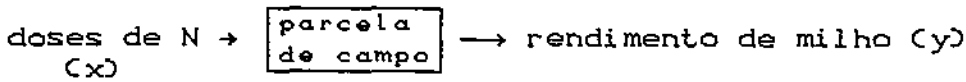
1 - Introdução

Em muitas ocasiões se obtém, para cada elemento amostral, duas medidas, correspondentes às variáveis x e y.



Análise de Regressão





O objetivo da obtenção desses pares de valores é o estudo do relacionamento entre as variáveis x e y que pode ser feito através de uma **análise de regressão** ou de uma **análise de correlação**.

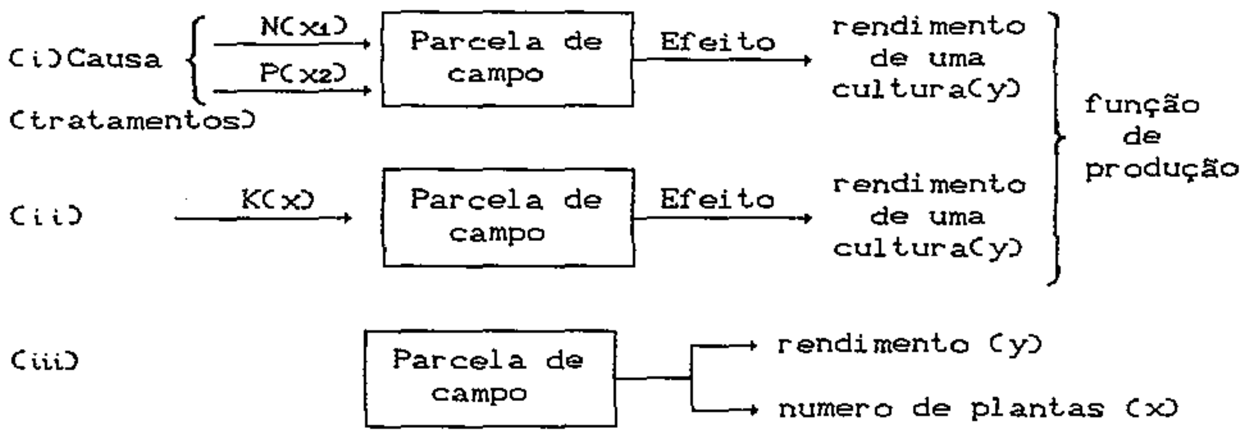
Na **análise de regressão** a variável y é considerada dependente e a variável x independente e se estabelece um relacionamento funcional de y para x , ou seja **y=f(x)**. A função matemática y=f(x) é expressa através de uma equação chamada **equação de regressão**.

Em estudos de correlação, verifica-se o grau de associação entre duas variáveis x e y (relacionamento linear), sem que haja necessariamente relação de dependência entre elas.

<u>Exs:</u> (i) Altura de irmão (x) — Altura de irmã (y)

(ii) Teor de fósforo no solo
- → pelo método A (x)
- → pelo método B (y)

2 - ANALISE DE REGRESSÃO



(i) $\underbrace{y}_{\text{variável dependente}} = f\underbrace{(x_1, x_2)}_{\text{variáveis independentes}}$

(ii) e (iii) $y = f(x)$

· Interessa estabelecer o relacionamento funcional entre os dois tipos de variáveis .

REGRESSÃO LINEAR SIMPLES

Relação entre y e x é dada por uma reta

$$\hat{y} = \underbrace{a}_{\text{intercepto da reta}} + \underbrace{b}_{\substack{\text{coeficiente de regressão linear} \\ \text{(inclinação da reta)}}}x$$

$$b = \frac{\text{covariância } (x, y)}{\text{variância } (x)} = \frac{SP_{xy}}{SQ_x}$$

SP_{xy} = soma dos produtos dos desvios de x e y

$$SP_{xy} = \Sigma(x-\bar{x})(y-\bar{y}) = \Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}$$

$n \longrightarrow$ numero de pares de valores

SQ_x = soma dos quadrados dos desvios de x

$$SQ_x = \Sigma(x-\bar{x})^2 = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$$

$$\text{Covariância } (x,y) = \frac{SP_{xy}}{\underbrace{n-1}_{GL}}$$

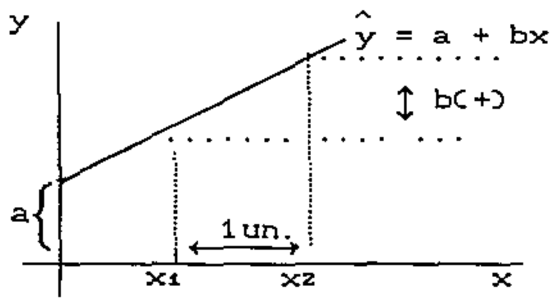
$\downarrow$

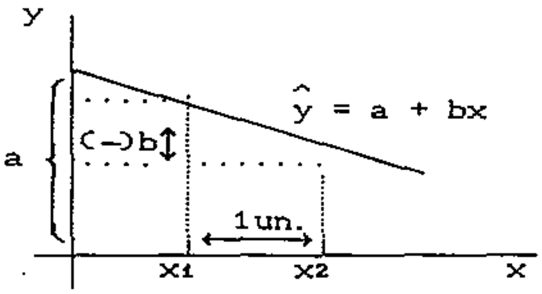
Medida de variação simultânea de x e y

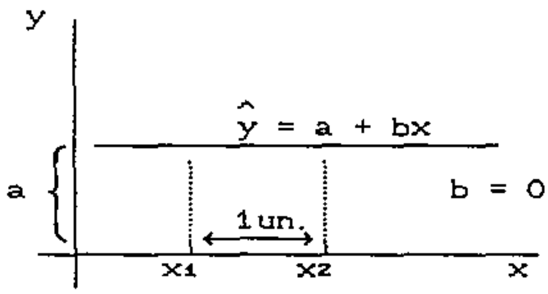
$$a = \bar{y} - b\bar{x}$$

a : Intercepto da reta. Ponto onde a reta corta o eixo dos y. É o valor estimado para y quando x = 0 .

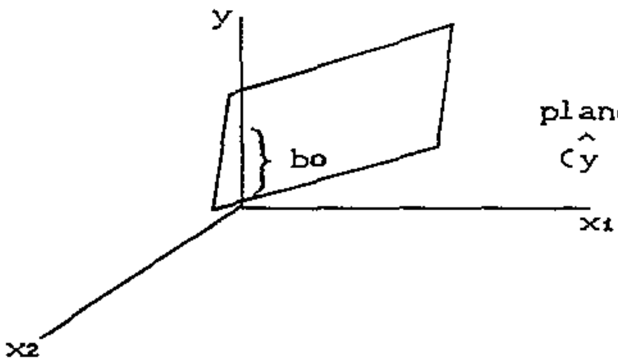
b : Inclinação da reta. Coeficiente de regressão linear . Representa quanto aumenta ou diminui y quando x cresce de uma unidade .







<u>REGRESSÃO MÚLTIPLA</u>



plano de regressão
$(\hat{y} = b_0 + b_1x_1 + b_2x_2)$
$\rightarrow$ Intercepto
b_1 e b_2 coeficientes
de regressão parciais

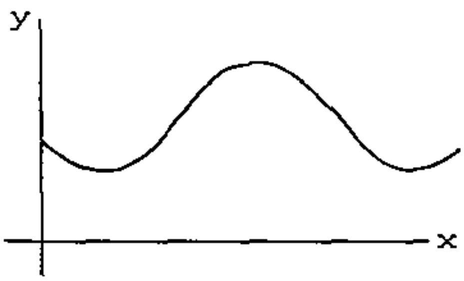
b_1 : Quantidade que afeta y quando x_1 varia de uma unidade permanecendo constante x_2 .

REGRESSÃO QUADRATICA

$$\hat{y} = a + bx + cx^2$$

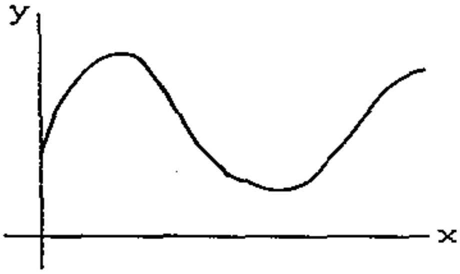
(curva de resposta)

REGRESSÃO CÚBICA OU DE 3º GRAU.



$$\hat{y} = a + bx + cx^2 + dx^3$$

REGRESSÃO QUARTICA OU DE 4º GRAU



$$\hat{y} = a + bx + cx^2 + dx^3 + ex^4$$

SUPERFÍCIE DE RESPOSTA

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_{11}x_1^2 + b_{22}x_2^2 + b_{12}x_1x_2$$

3 - REGRESSÃO LINEAR SIMPLES

É o caso mais simples de regressão, em que se estabelece o relacionamento entre duas variáveis X e Y através de uma reta, onde variável X (variável fixa): X_1 X_2 ... X_n

variável Y (variável aleatória): Y_1 Y_2 ... Y_n

Procura-se uma equação de regressão (uma reta) que melhor estime o relacionamento entre X e Y.

3.1. Ajuste de uma Regressão Linear Simples

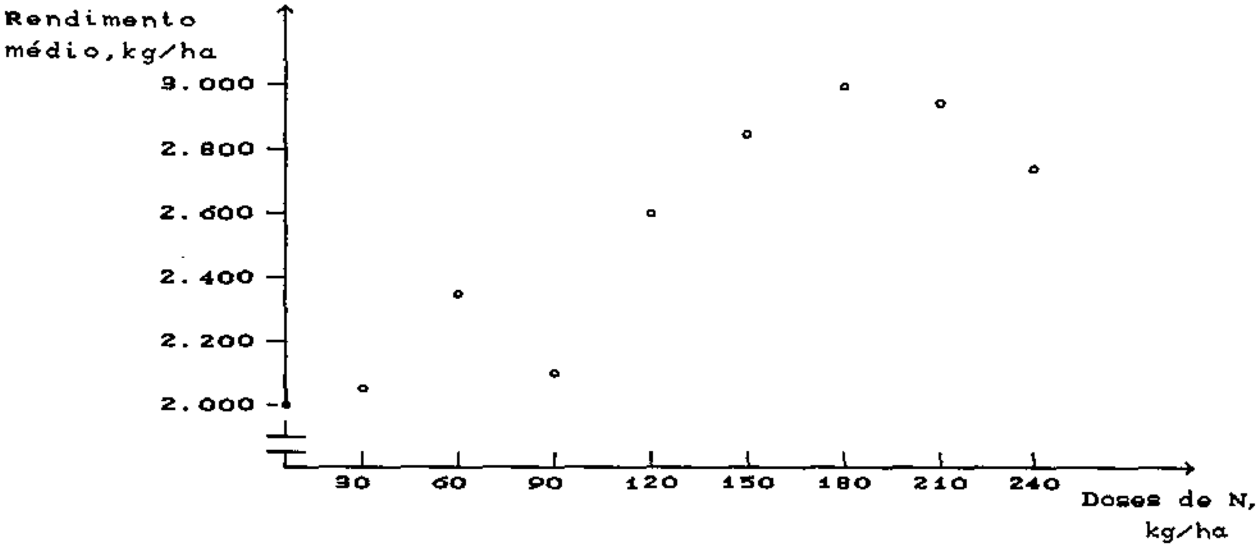
EXEMPLO 1:

Em um experimento realizado com o objetivo de verificar o efeito de adubação nitrogenada sobre o rendimento do milho, obteve-se estes resultados:

Doses de N, kg/ha (X)	Rendimento médio kg/ha (Y)	$\hat{Y}$	$Y - \hat{Y}$	$(Y - \hat{Y})^2$
0	2.000	1.994,2	5,8	33,64
30	2.080	2.126,2	-46,2	2.134,44
60	2.300	2.258,2	41,8	1.747,24
90	2.150	2.390,2	-240,2	57.696,04
120	2.600	2.522,2	77,8	6.052,84
150	2.820	2.654,2	165,8	27.489,64
180	3.000	2.786,2	213,8	45.710,44
210	2.950	2.918,2	31,8	1.011,24
240	2.800	3.050,2	-250,8	62.600,04
Total 1080	22.700	-	0,2	204.475,56

Etapas para verificar o relacionamento, através de uma análise de regressão, entre X e Y.

(a) Diagrama de dispersão ou diagrama de pontos.



(b) Estabelecimento da Equação de Regressão

$$\hat{Y} = a + bX \qquad b = \frac{SPxy}{SQx} \qquad a = \bar{Y} - b\bar{X}$$

$$SPxy = \sum xy - \left(\sum x\right)\left(\sum y\right) \Big/ n \;=\; 0\times2.000 + 30\times2.080 + \ldots +$$

$$240\times2.800 - (1.080)(22.700) \;/\; 9 \;=\; 2.960,40 - 2.724.000 = 236.400$$

$$SQx = \sum x^2 - \left(\sum x\right)^2 \Big/ n = 0^2 + 30^2 + \ldots + 240^2 - (1.080)^2 / \; 9$$
$$= 54.000$$

$$SQy = \sum y^2 - \left(\sum y\right)^2 \Big/ n = 2.000^2 + 2.080^2 + \ldots + 2.800^2 -$$
$$(22.700)^2 \;/\; 9 = 1.239.355,6$$

$$\bar{X} = \frac{\sum x}{n} = \frac{1.080}{9} = 120 \qquad\qquad \bar{Y} = \frac{\sum y}{n} = \frac{22.700}{9} = 2.522,2$$

$$b = \frac{236.400}{54.000} = 4,4 \text{ kg/ha} \;;$$

Estima-se em 4,4 kg/ha o acréscimo no rendimento de milho a cada kg/ha de N aplicado ao solo, ou seja, 440 kg/ha a cada 100 kg/ha de N.

$$a = 2.522,2 - 4,4\,(120) = 2.522,2 - 528 = 1.994,2$$

$$a = 1.994,2 \text{ kg/ha}$$

Para x = 0 , ou seja, sem aplicar N ao solo, estima-se um rendimento de 1.994,2 kg/ha de milho. Portanto, a equação de regressão é dada por:

$$\hat{Y} = 1.994,2 + 4,4X$$

onde $\hat{Y}$ = rendimento estimado de milho, em kg/ha e

X = doses de N, em kg/ha.

(c) Construção da reta sobre o diagrama de pontos, usando-se a equação de regressão: $\hat{Y} = 1.994,2 + 4,4X$. Para contruirmos a reta sobre o diagrama de pontos, necessitamos localizar 2 pontos, ou seja, faz-se:

$X = 0$; $\hat{Y} = 1.994,2 + 4,4 \times 0 = 1.994,2$

$x = 240$; $\hat{Y} = 1.994,2 + 4,4 \times 240 = 1.994,2 + 1.056 = 3.050,2$

Verificação do cálculo da equação e traçado da reta: a reta deverá pasar pelo ponto $(\bar{X},\bar{Y})$.

(d) Variância dos erros de Estimativa

$$s^2_{y/x} = \frac{\Sigma\,(y-\hat{y})^2}{n-2} = \frac{204.475,56}{9-2} = 29.210,79 \qquad s^2_{y/x} = 29.210,79$$

(e) Teste de significância do coeficiente de regressão (β = coeficiente de regressão linear verdadeiro ou populacional)

Ho: $\beta = 0$
Ha: $\beta \neq 0$

$$t = \frac{b}{s_b} = \frac{4,4}{0,73} = 6,027$$

$$s_b = \sqrt{s^2_b} = \sqrt{0,54} = 0,73$$

$$s^2_b = \frac{s^2_{y/x}}{SQx} = \frac{29.210,79}{54.000} = 0,54$$

$$|t| = 6,027 \quad > \quad t_{.05(7)} = 2,365$$

O coeficiente de regressão linear é significativo.

Existe uma tendência linear de acréscimo no rendimento de milho com a aplicação de N ao solo.

(f) Intervalo de confiança para o coeficiente de regressão linear verdadeiro.

$$\text{IC } 95\% \text{ para } \beta = b \pm t_{.05(7)}\, s_b$$

$$= 4{,}4 \pm 2{,}365 \cdot 0{,}73$$

$$= 4{,}4 \pm 1{,}7 \quad \begin{cases} \to 6{,}1 \\ \to 2{,}7 \end{cases}$$

IC 95% para β é [2,7 ; 6,1 kg/ha]

Temos uma confiança de 95% de que o acréscimo verdadeiro no rendimento de milho, para cada kg/ha de N aplicado ao solo, esteja entre 2,7 e 6,1 kg/ha.

3.2. MÉTODO DOS QUADRADOS MÍNIMOS

$y \to$ valor observado

$\hat{y} \to$ valor estimado pela equação da reta dada por $\hat{y} = a + bx$. Deseja-se que os desvios ou erros de estimativa $y-\hat{y}$ sejam um mínimo, ou que a soma dos quadrados dos desvios $\sum(y-\hat{y})^2$ seja um mínimo, ou seja que $z=\sum(y-\hat{y})^2= \sum(y-a-bx)^2$ seja mínimo.

Ter-se-á um mínimo para z quando $\frac{\partial z}{\partial a} = 0$ e $\frac{\partial z}{\partial b} = 0$

$$\left.\begin{aligned} \frac{\partial z}{\partial a} &= 2\sum (y-a-bx)(-1) = 0 \\ \frac{\partial z}{\partial b} &= 2\sum (y-a-bx)(-x) = 0 \end{aligned}\right\} \to \begin{aligned} &\sum y - na - b\sum x = 0 \\ &\sum xy - a\sum x - b\sum x^2 = 0 \end{aligned}$$

$$\therefore \left.\begin{aligned} na + b\sum x &= \sum y \\ a\sum x + b\sum x^2 &= \sum xy \end{aligned}\right\} \text{ Sistema de Equações Normais}$$

Tem-se um sistema com 2 equações e 2 incógnitas e utilizando-se qualquer método algébrico de resolução obtém-se:

$$b = \frac{\sum xy - (\sum x)(\sum y)/n}{\sum x^2 - (\sum x)^2/n} = \frac{SPxy}{SQx}$$

$$a = \bar{y} - b\bar{x}$$

Graus de Liberdade para $s^2_{y/x}$

GL = n-2 pois existem 2 retrições

(i) Da 1ª equação $\sum(y-\hat{y}) = 0$, pois

$$na+b\sum x = \sum y \Leftrightarrow \sum(a+bx) = \sum y \Leftrightarrow \sum[y-(a+bx)] = 0 \Leftrightarrow \sum(y-\hat{y}) = 0$$

(ii) Da 2ª equação $\sum x(y-\hat{y}) = 0$, pois

$$a\sum x + b\sum x^2 = \sum xy \Leftrightarrow \sum(ax+bx^2) = \sum xy \Leftrightarrow \sum x(a+bx) = \sum xy$$

$$\Leftrightarrow \sum x[y-(a+bx)] = 0 \Leftrightarrow \sum x(y-\hat{y}) = 0$$

3.3. ESTIMATIVA DE y POR INTERVALO

(a) Intervalo de confiança para um valor Y (faixa de confiança para valores futuros)

$$IC\ 100(1-\alpha)\%\ p/\ Y = \hat{y} \pm t_{\alpha(n-2)} \sqrt{\underbrace{\left[1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{SQx}\right] s^2_{y/x}}_{\Delta}}$$

Para o exemplo 1 seja $x = 150 = x_o$

$$\hat{y} = a+bx = 1994,2 + 4,4x = 1994,2 + 4,4\ (150) = 2654,2$$

$$IC\ 95\%\ p/\ y = \hat{y} \pm t_{.05(7)} \sqrt{\Delta}$$

$$= 2654,2 \pm 2,365 \sqrt{\left[1 + \frac{1}{9} + \frac{(150-120)^2}{54000}\right] 29210,79}$$

$$= 2654,2 \pm 429,3 \rightarrow \begin{cases} 3083,5 \\ 2224,9 \end{cases}$$

IC 95% p/ Y = [2224,9 kg/ha ; 3083,5 kg/ha]

Estima-se que o rendimento verdadeiro de milho de uma parcela com aplicação de 150 kg/ha de N esteja entre 2224,9 e 3083,5 kg/ha , com uma confiança de 95% .

(b) Intervalo de confiança para a média de todos os possíveis valores de y para um dado x (faixa de confiança para a reta de regressão).

$$IC\ 100(1-\alpha)\% = \hat{y} \pm t_{\alpha(n-2)} \sqrt{\left[\frac{1}{n} + \frac{(x_o - \bar{x})^2}{SQx}\right] s^2_{y/x}}$$

Para o exemplo 1, seja $x = x_o = 150 \Rightarrow \hat{y} = 2654,2$

$$IC\ 95\%\ p/\ E(y) = 2654,2 \pm 2,365 \sqrt{\left[\frac{1}{9} + \frac{(150-120)^2}{54000}\right] 29210,79}$$

$$= 2654,2 \pm 144,5 \rightarrow \begin{cases} 2798,7 \\ 2509,7 \end{cases}$$

IC 95% p/ E(y) = [2509,7 kg/ha ; 2798,7 kg/ha]

Estima-se que o rendimento médio verdadeiro de milho para uma aplicação de 150 kg/ha de N está entre 2509,7 e 2798,7 kg/ha com uma confiança de 95% .

3.4. APLICAÇÕES DA EQUAÇÃO DE REGRESSÃO

(i) Conhecer se y depende de x

b significativo ⇒ y depende de x

b não significativo ⇒ y não depende de x

(ii) Conhecer se existe relação de causa e efeito entre x e y

Uso da análise de regressão para comparar médias de tratamentos.

(iii) Determinar a forma da curva de regressão.

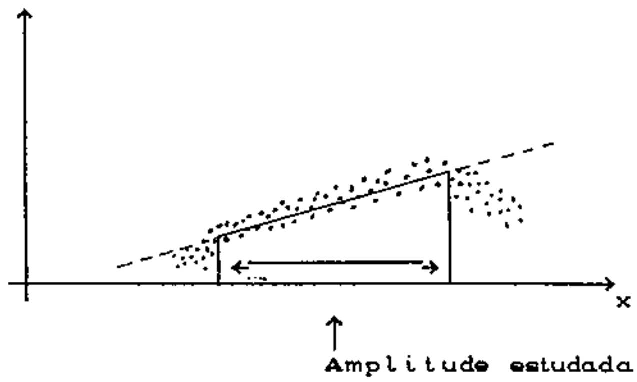
(iv) Fazer previsões de y partindo de x

Interpolação: prever y para valores x dentro da amplitude estudada.

Extrapolação: prever y para valores x fora da amplitude estudada.

CUIDADO! Deveria sempre ser evitada.

Risco da Extrapolação



(v) Ajustar valores y em função de uma variável relacionada x (análise de covariância)

animal → y = ganho de peso
animal → x = peso inicial

animal → y = ganho de peso
animal → x = ração consumida

parcela de campo → y = rendimento
parcela de campo → nº de plantas na parcela

$y_{iaj} = y_i - b(x_i - \bar{x})$

observado

4 - MODELO MATEMATICO PARA A REGRESSÃO LINEAR SIMPLES

$y_i = \alpha + \beta x_i + \varepsilon_i$

βx_i ⟶ Valor da variável independente x

ε_i ⟶ Erro ⟶ Erro de mensuração de y
⟶ Contribuição de outras variáveis não incluídas no modelo.
⟶ Relação pode não ser estritamente linear.

α, β ⟶ Parâmetros

y_i ⟶ valor observado de y

$\alpha \longrightarrow$ Intercepto

$\beta \longrightarrow$ Coeficiente de Regressão Linear verdadeiro

$i = 1, 2, \ldots, n$

PRESSUPOSIÇÕES

(i) A relação entre x e y é linear

(ii) Os valores x são fixos, isto é, x não é variável aleatória

(iii) $\varepsilon_i \cap N(0, \sigma^2)$ [Normalidade e Homocedasticidade]

(iv) O número de observações y é maior do que o número de parâmetros do modelo

As expressões de a e b poderiam ser deduzidas a partir do modelo e considerando que $\hat{y} = a + bx$ e tendo-se e_i (estimativa de ε_i), onde

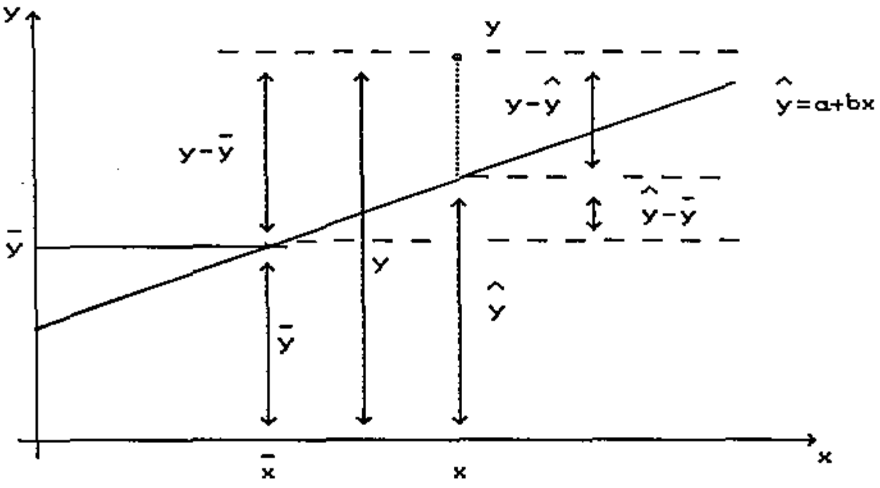
$$e_i = y - \hat{y} = (y - a + bx)$$

$$Z = \sum e_i^2 = \sum (y - \hat{y})^2 = \sum (y - a + bx)^2$$

5 - ANÁLISE DE VARIÂNCIA PARA A REGRESSÃO

(TESTE F PARA A REGRESSÃO)

(DECOMPOSIÇÃO DA SOMA DE QUADRADOS DE Y)



$$Y_i = \hat{y}_i + (y_i - \hat{y}_i)$$

$$y_i - \bar{y} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$$

$$\sum (y_i - \bar{y})^2 = \sum [(\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)]^2$$

$$\underbrace{\sum (y_i - \bar{y})^2}_{\text{SQTotal } (SQ_y)} = \underbrace{\sum (\hat{y}_i - \bar{y})^2}_{\text{SQRegressão}} + \underbrace{\sum (y_i - \hat{y}_i)^2}_{\text{SQDesvios da Regressão}} + \underbrace{2\sum (\hat{y}_i - \bar{y})(y_i - \hat{y}_i)}_{0}$$

SQTotal (SQ_y) $\updownarrow$ $\sum y_i^2 - (\sum y_i)^2/n$

SQRegressão: Redução da Soma de Quadrados de y devida a regressão de y para x

SQDesvios da Regressão: ↑ Forma direta de cálculo

Forma usual:
SQDesvios da Regressão =SQTotal - SQRegressão

$$\text{SQRegressão} = \sum (\hat{y} - \bar{y})^2 = \sum (a + bx - \bar{y})^2 = \sum (\bar{y} - b\bar{x} + bx - \bar{y})^2$$

$$= \sum [b(x - \bar{x})]^2 = b^2 \sum (x - \bar{x})^2$$

$$= b^2\ SQ_x = \left[\frac{SP_{xy}}{SQ_x}\right]^2 SQ_x = \frac{(SP_{xy})^2}{(SQ_x)^2}\ SQ_x$$

$$= \frac{(SP_{xy})^2}{SQ_x}$$

GL

GL Total = n - 1

GL Regressão = p - 1 = k

GL Desvios da regressão = (n - 1) - (p - 1)

= GL Total - GL Regressão

n = numero de pares de valores .

p = numero de parâmetros do modelo

k = numero de variáveis independentes

No caso da Regressão Linear p = 2 e k = 1 e então

GL Regressão = 1

GL Desvios da Regressão = n-1-1 = n-2

TABELA DE ANALISE DE VARIANCIA

Causas da Variação	GL	SQ	QM	F
Regressão	1	SQRegressão	QMRegressão	$\frac{QMReg}{QMDReg}$
Desvios da Regressão	n - 2	SQDReg	QMDReg	
Total	n - 1	SQTotal		

Para o exemplo 1

Causas da Variação	GL	SQ	QM	F
Regressão Linear	1	1034906,67	1034906,67	35,43**
Desvios da Regressão	7	204448,93	29206,99 $\sigma^2_{y/x}$	
Total	8	1239355,60		

$F_{.01(1,7)} = 12,25$

$Ho: \beta = 0$ $\quad$ $Ha: \beta \neq 0$

$F = 35,43 > F_{.01(1,7)} = 12,25$ O coeficiente de Regressão Linear é significativo.

A aplicação de doses crescentes de N ao solo incrementam linearmente o rendimento de milho.

1 GL $\Rightarrow$ $F = 35,43$; $t = 6,027$; $F = t^2$

Esse procedimento é válido para dados pareados, isto é, nos casos onde para cada valor de x tem-se um valor de y.

6 - REGRESSÃO LINEAR NA ANÁLISE DE VARIANCIA DO EFEITO DE TRATAMENTOS

Sempre que os tratamentos representam níveis de um ou mais fatores quantitativos a técnica recomendada a ser utilizada na decomposição da SQ de tratamentos é a Análise de Regressão. O esquema básico de análise é o do esquema experimental somente com a subdivisão da SQT, onde SQT → SQ Regressão
→ SQ Desvios da Regressão

DCC

t tratamento, r repetições

C.Variação	GL
Tratamentos	t - 1
Regressão	1
Desvio da Regressão	t - 2
Erro Exp.	t(r - 1)
Total	rt - 1

DBC

t tratamentos, r blocos

C. Variação	GL
Blocos	r - 1
Tratamentos	t - 1
Regressão	1
Desvios da Regressão	t - 2
Erro Exp.	(t - 1)(r - 1)
Total	rt - 1

SQ Tratamentos, SQ Total, SQ Erro Exp. e SQ Blocos da maneira usual

$$\text{SQ Regressão} = \frac{(SP_{xy})^2}{r\ SQ_x}$$

SQ Desvios da Regressão = SQT - SQRegressão

$$\text{Coeficiente de Regressão Linear : } b = \frac{SP_{xy}}{rSQ_x}$$

$$a = \bar{y} - b\bar{x}$$

O teste F para Desvios da Regressão nesse caso é o teste do ajuste da Regressão Linear para os dados.

EXEMPLO 2:

Em um experimento de adubação fosfatada em milho, instalado em blocos casualizados com 4 repetições, em que foram usadas as doses 0 , 25 , 50 e 75 kg/ha de P_2O_5, obteve-se estes resultados, em kg/parcela:

Blocos	P_2O_5 (x) 0	25	50	75	Totais
1	3,30	10,20	12,00	19,28	44,68
2	4,80	9,75	13,32	16,13	44,00
3	5,44	7,18	15,00	18,22	45,84
4	4,18	9,23	17,25	20,75	51,41
Totais de Tratamentos	17,72	36,36	57,57	74,28	185,93
Médias de Tratamentos	4,43	9,09	14,39	18,57	

Tabela de Análise de Variância

Causas de Variação	GL	SQ	QM	F
Blocos	3	8,53		
Tratamentos	3	456,34		
Regressão Linear	1	455,49	455,49	158,21 **
Desvios da Regressão	2	0,85	0,425	0,25
Erro Experimental	9	25,92	2,88	
Total	15	490,79		

A Regressão Linear do rendimento de milho para doses de P_2O_5 é significativa.

A aplicação de doses crescentes de P_2O_5 incrementa linearmente o rendimento de milho.

A não significância dos desvios regressão indica que a regressão linear se ajusta aos dados.

EQUAÇÃO DE REGRESSÃO

$b = (SP_{xy})/rSQ_x = 2386{,}125/4(3175) = 0{,}19$

$a = \bar{y} - b\bar{x} = 11{,}62 - (0{,}19)(37{,}5) = 4{,}50$

$\hat{y} = 4{,}50 + 0{,}19x$

b = 0,19 → Estima-se em 0,19 kg/parcela o acréscimo no rendimento de milho para cada kg/ha a mais de P_2O_5 aplicado ao solo.

a = 4,50 ⇒ Estima-se em 4,50 kg/parcela o rendimento de milho sem aplicação de P_2O_5 ao solo.

7 - COEFICIENTE DE DETERMINAÇÃO

(1) Para dados pareados

[Para cada valor de x um valor de y]

$$r^2 = \frac{\text{SQRegressão}}{\underbrace{\text{SQTotal}}_{SQ_y}}$$

Representa a proporção da variação total (SQ_y) explicada pela regressão de y para x

r^2 indica a utilidade de usar a regressão para se fazer ajustamentos ou previsões.

SQRegressão = $r^2 SQ_y$

SQDesvios da regressão = $(1 - r^2)SQ_y$

EXEMPLO 1

Adubação Nitrogenada em Milho

$$r^2 = \frac{\text{SQRegressão}}{\text{SQTotal}} = \frac{1034906,67}{1239355,60} = 0,84$$

84% da variação no rendimento de milho é explicada pela variação do rendimento para doses crescentes de N aplicadas ao solo.

(2) Para dados com repetição

[Para cada valor de x vários valores de y]

$$r^2 = \frac{\text{SQRegressão}}{\underbrace{\text{SQTratamentos}}_{SQT}}$$

Representa a proporção da variação devida a tratamentos explicada pela regressão de y para x

SQRegressão = r^2SQT

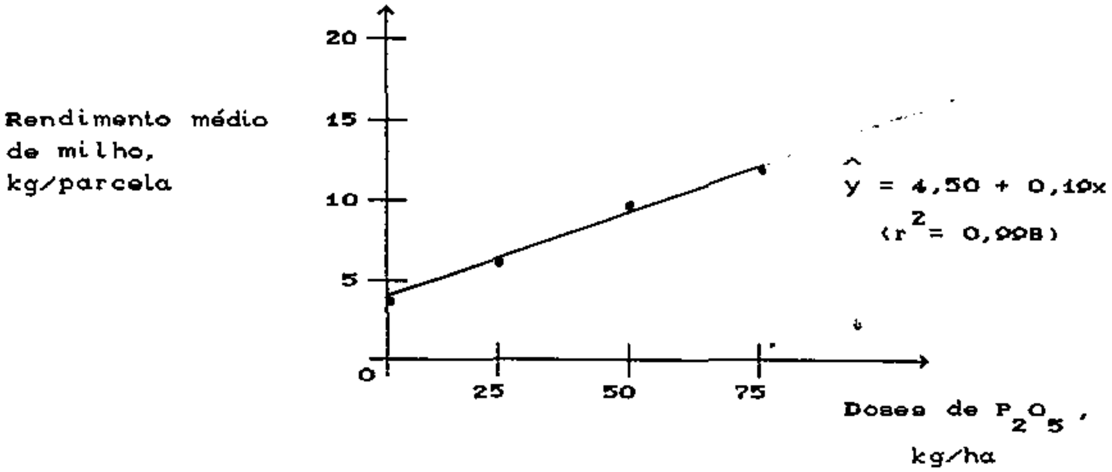
SQDesvios da Regressão = $(1 - r^2)$SQT

EXEMPLO 2: Adubação fosfatada em milho

$$r^2 = \frac{\text{SQRegressão}}{\text{SQT}} = \frac{455,49}{456,34} = 0,998$$

99,8% da variação no rendimento do milho devida a tratamentos é explicada pela influência linear das doses de P_2O_5 aplicadas ao solo.

x	$\bar{y}$	$\hat{y}$	$\bar{y} - \hat{y}$
0	4,43	4,50	-0,07
25	9,09	9,25	-0,26
50	14,39	14,00	0,39
75	18,57	18,75	-0,18
			-0,02



8- VERIFICAÇÃO DA ADEQUABILIDADE DO MODELO DE REGRESSÃO LINEAR SIMPLES

1. Análise de Resíduos:

Exemplo 3: x = taxa de mistura
y = % impureza de um produto produzido num processo químico.

x	20	22	24	26	28	30	32	34	36	38	40	42
y	8,4	9,5	11,8	10,4	13,3	14,8	13,2	14,7	16,4	16,5	18,9	18,

$\bar{y} = 13,87 \qquad \bar{x} = 31 \qquad n = 12$

$$b = \frac{SPxy}{SQx} = \frac{261,20}{572} = 0,4566$$

$$a = \bar{y} - b\bar{x} = 13,87 - 0,4566(31) = -0,2879$$

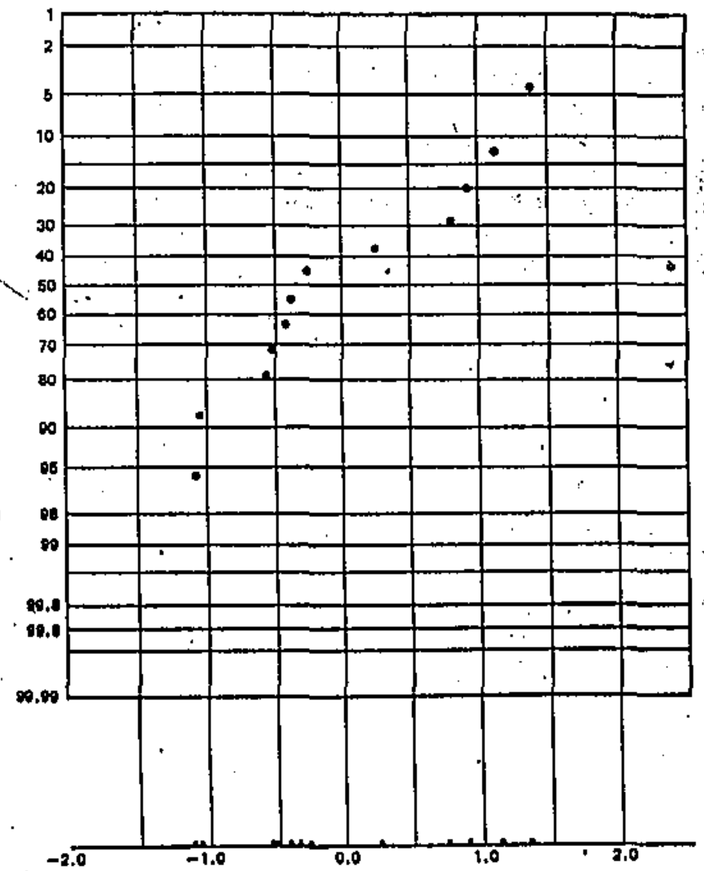
$$\therefore \hat{y} = -0,2879 + 0,4566x$$

Resíduos: $e_j = y_j - \hat{y}_j$

x	y_j	$\hat{y}_j$	$e_j = \hat{y}_j - y_j$
20	8,4	8,8441	-0,4441
22	9,5	9,7573	-0,2573
24	11,8	10,6705	1,1295
26	10,4	11,5837	-1,1295
28	13,3	12,4969	0,8031
30	14,8	13,4101	1,3899
32	13,2	14,3233	-1,1233
34	14,7	15,2365	-0,5365
36	16,4	16,1497	0,2503
38	16,5	17,0629	-0,5629
40	18,9	17,9761	0,9239
42	18,5	18,8893	-0,3893

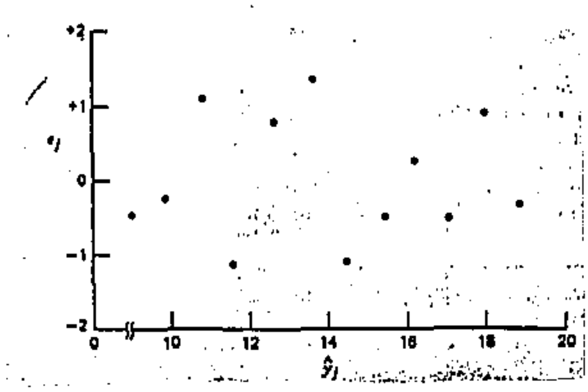
Análise de Resíduos:

- Gráfico de probabilidade normal.
- e_j vs $\hat{y}_j$
- e_j vs variáveis independentes (x)
- e_j vs variáveis não incluídas no modelo.



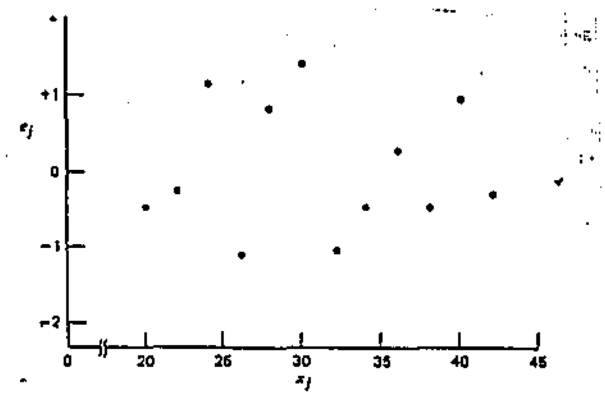
Muito próximo de uma reta.

Gráfico de Probabilidade Normal dos Resíduos



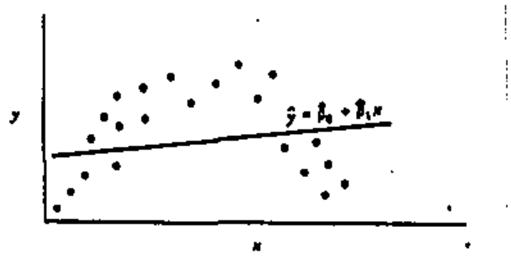
Sem estrutura de funil.

Resíduos e_j versus valores ajustados $\hat{y}_j$



Nenhuma relação evidente.

Resíduos e_j versus x_j



Modelo Quadrático ajustar-se-ia bem melhor.

Modelo de Regressão Linear com falta de ajuste

2. <u>Teste de Falta de Ajuste</u> (Lack of Fit Teste)

Necessita-se para cada x vários valores y, ou seja repetições.

$$\begin{array}{ll} x_1 & y_{11}, y_{12}, \ldots, y_{1r} \\ x_2 & y_{21}, y_{22}, \ldots, y_{2r} \\ \vdots & \\ x_t & y_{t1}, y_{t2}, \ldots, y_{tr} \end{array}$$

Ho: O modelo se ajusta adequadamente aos dados.

Ha: O modelo não se ajusta adequadamente aos dados.

SQ Resíduo = SQ Erro Puro + SQ Falta de Ajuste

↑ SQ Erro Experimental

SQ↑Desvios da Regressão

DCC

C. Variação	GL
Regressão	1
Resíduo	rt-2
Erro Exper.	t(r-1)=rt-t
Desvios da Reg.	t-2
Total	rt-1

Teste de Ho

$$F = \frac{\text{QM Desvios de Reg}}{\text{QM Erro Exp.}}$$

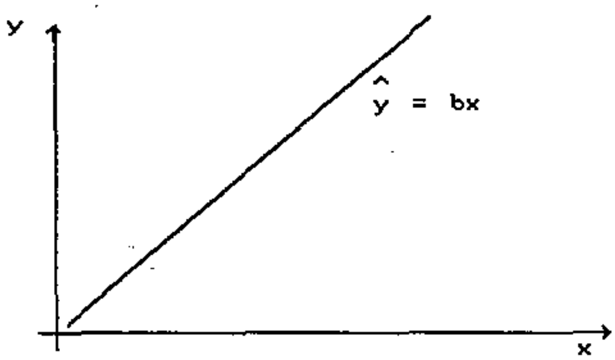
3. <u>Coeficiente de Determinação:</u> r^2 ou R^2

- Usar com cautela, especialmente em regressão polinomial (por 3 pontos ajusta-se perfeitamente uma curva de 2º grau).
- R^2 cresce adicionando variáveis ao modelo.
- O crescimento no R^2 não deve ser compensado pela perda de GL para o resíduo provocando QM do resíduo maior.

9 - CASOS ESPECIAIS NA ANÁLISE DE REGRESSÃO

1. Regressão através da origem

Em muitas situações dada a natureza dos fatores nos interessa obter uma reta que passa pela origem, como por exemplo na calibração de instrumentos.



onde $b = \dfrac{\sum xy}{\sum x^2}$

Para se testar a hipótese Ho: $\beta=0$ vs Ha: $\beta\neq 0$ calcula-se

$$t = \frac{b}{s_b} \qquad \text{onde} \quad s_b = \sqrt{\frac{s^2_{y/x}}{\sum x^2}}$$

← variância dos erros de estimativa

↖ erro padrão do coeficiente de regressão

$$\text{onde} \quad s^2_{y/x} = \frac{\sum (y-\hat{y})^2}{n-1}$$

Mesmo que a reta deva passar pela origem, pode-se calcular $a=\bar{y}-b\bar{x}$ que representa o "erro zero" do instrumento em calibração e para o qual deverá ser feito um ajustamento. O valor de $\underline{a}$ pode ser testado, ou seja, a hipótese Ho: $\alpha=0$ é testada calculando-se

$$t = \frac{a}{s_a}$$ → erro padrão de $\underline{a}$

$$\text{onde} \quad s_a = \sqrt{\frac{s^2_{y/x}\,(\sum x^2)}{n\ SQx}}$$

Quando t NS → aceita-se a pressuposição de que a reta passa pela origem.

2. Regressão quando x e y são aleatórios

Exemplos:

(i) Peso (x) e altura (y) de indivíduos tomados ao acaso.

(ii) Produção de uma cultura (y) e infestação de nematóides (x).

Nestes casos tanto x como y podem ser variável dependente ou independente.

$$\begin{cases} y \rightarrow \text{dependente} \\ x \rightarrow \text{independente} \end{cases} \Rightarrow \quad b_{y.x} = \frac{SPxy}{SQx}$$

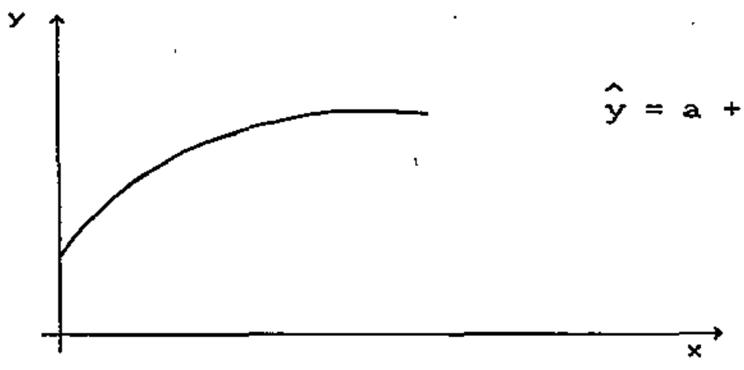
$$\begin{cases} x \rightarrow \text{dependente} \\ y \rightarrow \text{independente} \end{cases} \Rightarrow \quad b_{x.y} = \frac{SPxy}{SQy}$$

Duas retas de regressão distintas podem ser construídas, sendo que geralmente uma só tem interesse prático.

3. Regressão entre x e y não-lineares

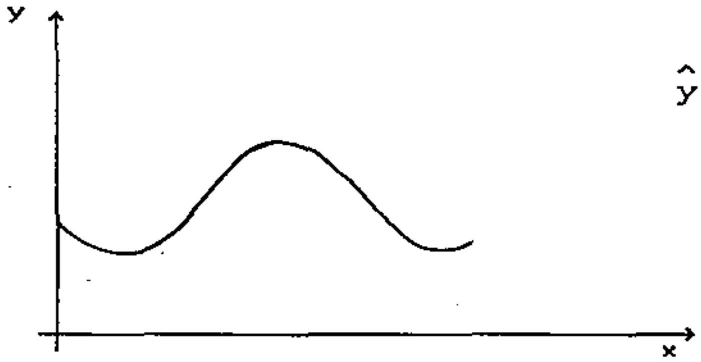
(a) Regressões Polinomiais.

(i) Regressão Quadrática



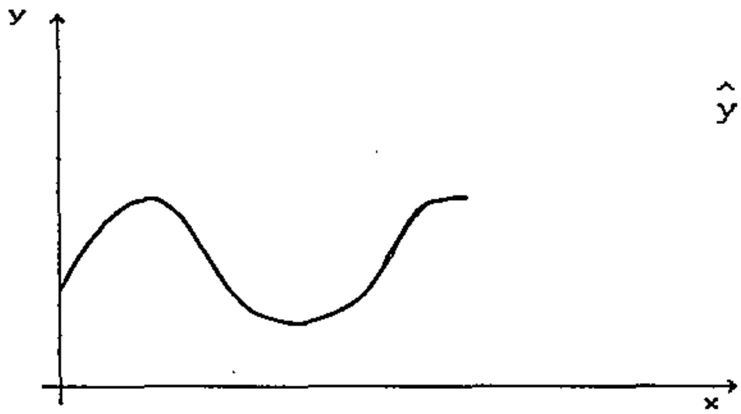
$$\hat{y} = a + bx + cx^2$$

(ii) Regressão Cúbica



$$\hat{y} = a + bx + cx^2 + dx^3$$

(iii) Regressão Quártica

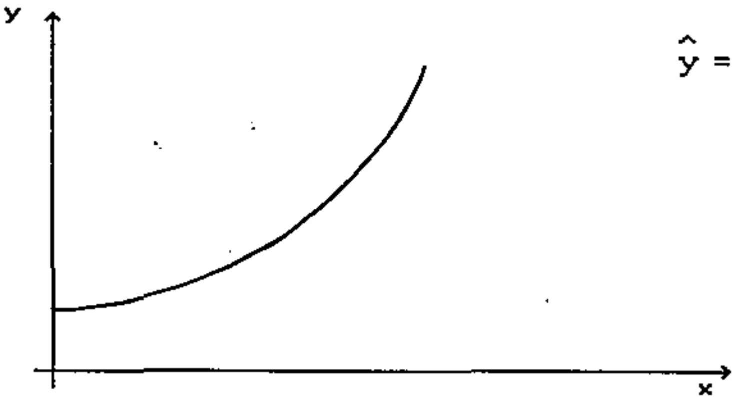


$$\hat{y} = a + bx + cx^2 + dx^3 + ex^4$$

Ajuste das equações, procedimento similar ao utilizado na regressão linear simples.

(b) Regressões Exponenciais

(i) Curva de crescimento exponencial



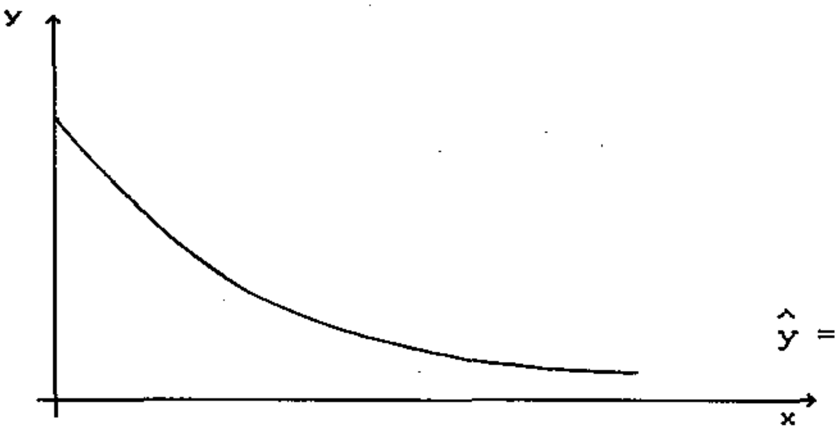
$$\hat{y} = ab^{x}$$

Obtenção da equação:

$$\underbrace{\log \hat{y}}_{\hat{y}} = \underbrace{\log a}_{a} + \underbrace{\log bx}_{b}$$

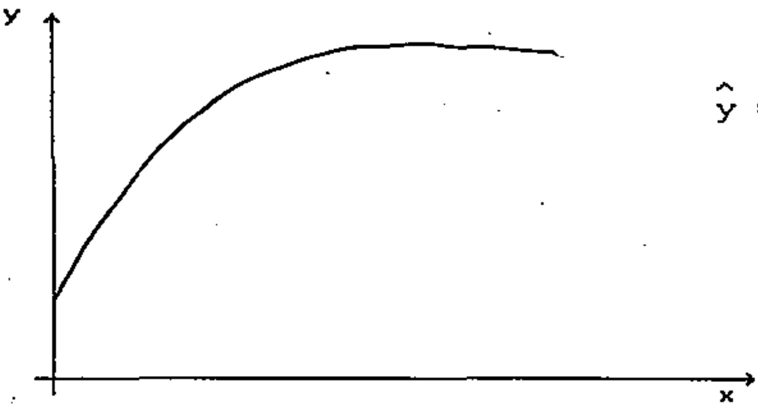
$$\hat{y} = a + bx$$

(ii) Curva de decadência exponencial



$$\hat{y} = ab^{-x}$$

(iii) Regressão Assintótica



$$\hat{y} = a - b(e^{-cx})$$

Equação de Mitscherlich

Utilizada no estudo de fenômenos de adubação.

10 - ANALISE DE REGRESSÃO APLICADA À TRANSFORMAÇÃO DE DADOS

10.1. BUSCA DA TRANSFORMAÇÃO

Seja $E_{(y)} = \mu$ a média de y e supõe-se que o desvio padrão de y é proporcional a uma potência da média de y tal que

$$\boxed{\sigma_y \quad \propto \quad \mu^{\alpha}}$$

Desejamos encontrar uma transformação de y que produz uma variância constante. Supõe-se que a transformação é uma potência dos dados originais, por exemplo $y^{*} = y^{\lambda}$.

Então pode-se mostrar que

$$\boxed{\sigma^{*}_y \quad \propto \quad \mu^{\lambda+\alpha-1}}$$

Logo se tomarmos $\lambda=1-\alpha$, então a variância dos dados transformados y^{*} é constante.

As transformações mais comuns são sintetisadas a seguir:

Relação entre σ_y e μ	α	$\lambda=1-\alpha$	Transformações
$\sigma_y \propto$ constante	0	1	Nenhuma
$\sigma_y \propto \mu^{1/2}$	1/2	1/2	Raíz Quadrada ← dados ∩ Poisson
$\sigma_y \propto \mu$	1	0	Logarítmo
$\sigma_y \propto \mu^{3/2}$	3/2	-1/2	Raíz quadrada recíproca
$\sigma_y \propto \mu^{2}$	2	-1	Recíproca

10.2. SELEÇÃO EMPÍRICA DE α

Em muitas situações experimentais, onde existem repetições, podemos empiricamente estimar α dos dados. Desde que no i-ésimo tratamento

$$\boxed{\sigma_{yi} \quad \propto \quad \mu_i^{\alpha} = \theta\ \mu_i^{\alpha}}$$

onde θ é uma constante de proporcionalidade ($\sigma_{yi} = \theta\ \mu_i^{\alpha}$).

Usando logaritmo obtemos

$$\boxed{\log \sigma_{yi} = \log \theta + \alpha \log \mu_i}$$

Assim um gráfico de $\log \sigma_{yi}$ e $\log \mu_i$, produziria uma reta com inclinação α.

Desde que não conhecemos σ_{yi} e $\log \mu_i$, utiliza-se a média $\bar{y}_{i.}$ e o desvio padrão s_i do tratamento i como estimadores de μ_i e σ_{yi} e a inclinação da linha reta ajustada como estimativa de α, ou seja

$$\boxed{\underbrace{\log s_i}_{y} = \log \theta + \hat{\alpha} \underbrace{\log \bar{y}_{i.}}_{x}}$$

Exemplo 4:

Um engenheiro está interessado em determinar se quatro diferentes métodos de estimação da freqüência de ocorrência de enchentes produzem equivalentes estimativas do pico de vazão quando aplicados à mesma bacia de um rio. Cada procedimento é usado 6 vezes na bacia do rio e os resultados de vazão foram os seguintes:

Método de estimação	Observações						$\bar{y}_{i.}$	s_i
1	0,34	0,12	1,23	0,70	1,75	0,12	0,71	0,66
2	0,91	2,94	2,14	2,36	2,86	4,55	2,63	1,09
3	6,31	8,37	9,75	6,09	9,82	7,24	7,93	1,66
4	17,15	11,52	10,95	17,20	14,35	16,82	14,72	2,77

ANOVA para dados originais

C. Variação	GL	SQ	QM	F
Métodos	3	708,3471	236,1157	76,07 * *
Erro	20	62,0811	3,1041	
Total	23	770,4282		

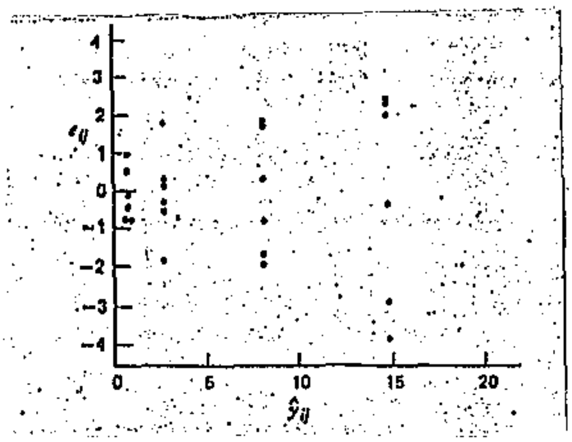
ANOVA para dados transformados $y_{ij}^{*} = \sqrt{y_{ij}}$

C. Variação	GL	SQ	QM	F
Métodos	3	32,6842	10,8947	81,05 **
Erro	20	2,6884	0,1344	
Total	23	35,3726		

. Resíduos $e_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_{i.}$

Método	Observações						$\hat{y}_{ij} = \bar{y}_{i.}$
1	-0,37	-0,59	0,52	-0,01	1,04	-0,59	0,71
2	-1,72	0,31	-0,49	-0,27	0,23	1,92	2,63
3	-1,62	0,44	1,82	-1,84	1,89	-0,69	7,93
4	2,43	-2,90	-3,77	2,48	-0,37	2,10	14,72

. Gráfico resíduos e_{ij} versus valores ajustados $\hat{y}_{ij}$



Estrutura de funil, os resíduos crescem com a medida que crescem $\hat{y}_{ij}$.

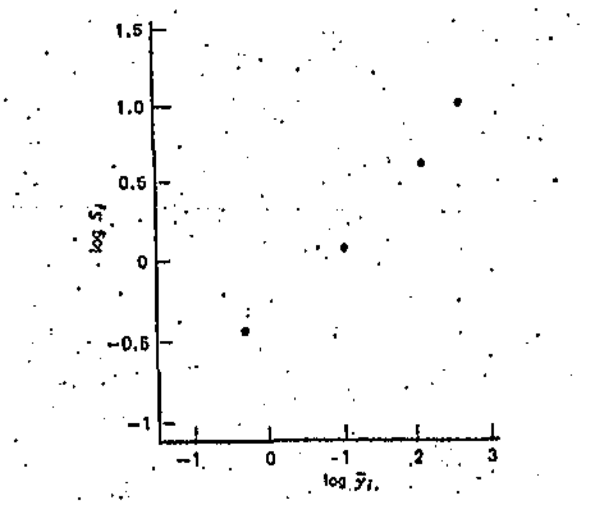
↓

Indicador de Variância não constante, ou seja, de Heterocedasticidade

. $\log \bar{y}_{i.}$ e $\log s_i$

$\bar{y}_{i.}$	s_i	$\overbrace{\log \bar{y}_{i.}}^{x}$	$\overbrace{\log s_i}^{y}$
0,71	0,66	-0,1487	-0,1805
2,63	1,09	0,4200	0,0374
7,93	1,66	0,8993	0,2201
14,72	2,77	1,1679	0,4425

Gráfico $\log s_i$ versus $\log \bar{y}_{i.}$



. Equação de Regressão: $\hat{y} = -0{,}1342 + 0{,}4517x$

$\hat{y} = \log \theta + \hat{\underline{\alpha}}\, x$ $\qquad \hat{\alpha} = 0{,}4517 \cong 1/2$

$\uparrow$ $\log s_i$ $\qquad \uparrow$ $\log \bar{y}_{i.}$ $\qquad \downarrow$ Transformação raíz quadrada

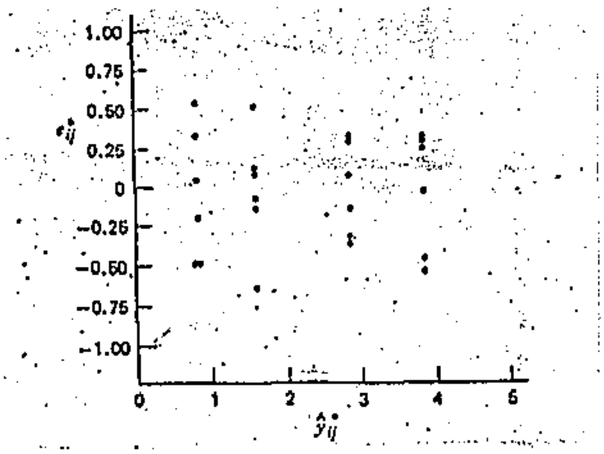
. Transforma-se os dados para $y_{ij}* = \sqrt{y_{ij}}$

Método	Observações						$\bar{y}_{i.}*$
1	0,5831	0,3464	1,1091	0,8367	1,3229	0,3464	0,7574
2	0,9539	1,7146	1,4629	1,5362	1,6912	2,1331	1,5820
3	2,5120	2,8931	3,1225	2,4678	3,1337	2,6907	2,8033
4	4,1413	3,4380	3,3091	4,1473	3,7881	4,1012	3,8208

. Resíduos $e_{ij}* = y_{ij}* - \hat{y}_{ij}* = y_{ij}* - \bar{y}_{i.}*$

Método	Observações						$\hat{y}_{ij}* = y_{i.}*$
1	-0,1743	-0,4110	0,3517	0,0793	0,5655	-0,4110	0,7574
2	-0,6281	0,1326	-0,1191	-0,0458	0,1092	0,5511	1,5820
3	-0,2913	0,0898	0,3192	-0,3355	0,3304	-0,1126	2,8033
4	0,3205	-0,3828	-0,5117	0,3265	-0,0327	0,2804	3,8208

. Gráfico resíduos $e_{ij}*$ versus valores ajustados $\hat{y}_{ij}*$



Não se verifica estrutura de funil, não se evidenciando relação entre $e_{ij}*$ e $\hat{y}_{ij}*$.

. Análise de variância com os valores transformados $y_{ij}^* = \sqrt{y_{ij}}$

. ANOVA, técnicas de complementação com dados transformados. Apresentação de resultados com dados originais.

A transformação utilizada simplesmente para atender as suposições do modelo.

11- ANÁLISE DE REGRESSÃO PARA COMPARAÇÃO DE MÉDIAS DE TRATAMENTOS
(Ajuste de regressões curvilineares pela técnica de polinômios ortogonais)

Sempre que os tratamentos representam níveis de um ou mais fatores quantitativos, a técnica mais recomendável (por ser mais informativa e mais eficiente para identificar diferenças entre médias de tratamentos) na decomposição da SQtratamentos é a análise de regressão.

O esquema básico de análise de variância é o do delineamento experimental adotado, somente com subdivisão da SQtratamentos, em regressão linear, regressão quadrática; isto é em tantos componentes quantos forem os GL de tratamentos.

A equação de regressão a ser ajustada incluirá todos os componentes até o de mais alto grau significativo, mesmo que no intervalo exista algum componente não significativo.

O procedimento se simplifica quando os níveis de x forem equidistantes, através da utilização de polinômios ortogonais, cujos coeficientes (C_{ji}) são tabelados, por exemplo em Pimentel Gomes, 1985.

<u>Exemplo 5:</u>

Num experimento de adubação potássica de cana-de-açucar utilizou-se 4 níveis de K_2O (0 ; 60 ; 120 e 180 Kg/ha) em DCC com 8 repetições por tratamento.

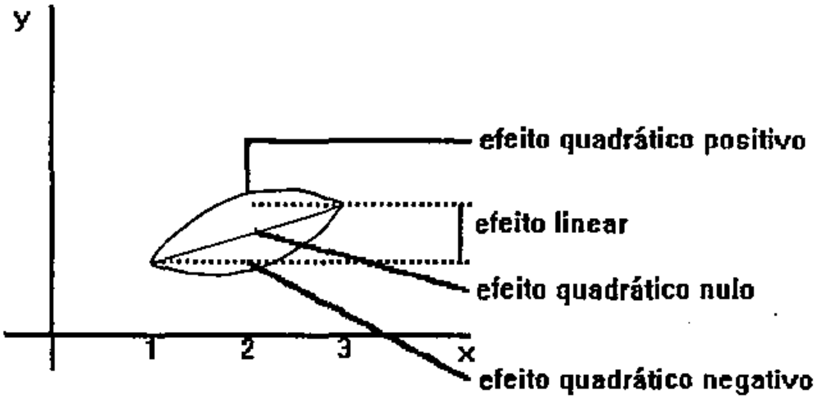
A análise de variância é a seguinte:

Causas de Variação	GL	SQ	QM	F
Tratamentos	3	638,50	212,83	3,20*
Resíduo	28	1860,60	66,45	

CV = 21,2%
F.05(3,28) = 2,96

$$\text{GLtratamentos} = 3 \begin{cases} \text{componente linear (K') ou reg. linear} \rightarrow 1 \text{ GL} \\ \text{componente quadrático (K'') ou reg. quadrática} \rightarrow 1 \text{ GL} \\ \text{componente cúbico (K''') ou reg. cúbica} \rightarrow 1 \text{ GL} \end{cases}$$

Para três níveis a situação seria:



$$2\ \text{GL} \begin{cases} \text{efeito linear : linearidade} \rightarrow 1\ \text{GL} \\ \text{efeito quadrático : curvatura} \rightarrow 1\ \text{GL} \end{cases}$$

níveis de x	coeficientes dos componentes linear	quadrático
1	-1	1
2	0	-2
3	1	1

Para o exemplo tem-se:

(x) trat.	totais($y_{i.}$) de tratamentos	coeficientes (c_{ji}) 1°grau	2°grau	3°grau	$c_{ji}y_{i.}$ 1°grau	2°grau	3°grau
0	302	-3	1	-1	-906	302	-302
60	370	-1	-1	+3	-370	-370	1110
120	400	1	-1	-1	400	-400	-1200
180	348	3	1	+1	1044	348	348
total	1420	0	0	0	168	-120	-44
	K	20	4	20			
	M	2	1	10/3			

K: soma dos quadrados dos coeficientes
M: constante para tornar os coeficientes inteiros

$$SQK' = SQ\,Regressão\,1^\circ grau = \frac{(168)^2}{8x20} = 176,40$$

$$SQK'' = SQ\,Regressão\,2^\circ grau = \frac{(-120)^2}{8x4} = 450,00$$

$$SQK''' = SQ\,Regressão\,3^\circ grau = \frac{(-44)^2}{8x20} = 12,10$$

$$SQC_j = \frac{(\sum c_{ji} y_{i.})^2}{\underset{\displaystyle \sum_j C_{ji}^2}{\underset{\downarrow}{rK}}}$$

C.VARIAÇÃO	GL	SQ	QM	F
K' (reg.linear)	1	176,40	176,40	2,65
K'' (reg.quadrática)	1	450,00	450,00	6,77*
K''' (reg.cúbica)	1	12,10	12,10	0,18
Tratamentos	(3)	(638,50)		
Resíduo (erro)	28	1860,60	66,45	

$F_{.05(1,28)} = 4,20$

A resposta da cana-de-açúcar à adubação potássica é quadrática, indicando que para doses baixas de K_2O aplicadas ao solo, o rendimento da cana apresenta um acréscimo linear; à medida que doses de K_2O crescem, o acréscimo vai se tornando menor tendendo a estabilizar nas doses mais altas.

COEFICIENTE DE DETERMINAÇÃO

$$R^2 = \frac{SQRlinear + SQRquadrática}{SQtratamentos} = \frac{SQK' + SQK''}{SQK} = \frac{176,40 + 450}{638,50} = 0,98$$

98% da variação no rendimento de cana-de-açúcar devido a aplicação de K_2O é explicada pela reg. quadrática do rendimento de cana para doses de K_2O aplicadas ao solo.

AJUSTE DA EQUAÇÃO DE REGRESSÃO

$\hat{y} = \bar{y} + B_1 M_1 P_1 + B_2 M_2 P_2 + \ldots$

| média geral

$$B_j = \frac{\sum_i c_{ji} y_{i.}}{rK}$$

B_j coeficiente de regressão de grau j

P_j polinômio ortogonal de grau j

No exemplo

$$\hat{y} = \bar{y} + B_1 M_1 P_1 + B_2 M_2 P_2 + \ldots$$

$\bar{y} = \dfrac{1420}{32} = 44,375$ $\qquad M_1 = 2$

$M_2 = 1$

$B_1 = \dfrac{168}{8(20)} = 1,05$

$B_2 = \dfrac{-120}{8(4)} = -3,75$

$P_1 = x = \dfrac{x - \bar{x}}{q}$ ← médias dos níveis

← diferença entre os níveis

$P_2 = x^2 - \dfrac{n^2 - 1}{12} = x^2 - \frac{5}{4}$; n = n° de níveis, no exemplo n=4

$\hat{Y} = 44,375 + 1,05(2)x - 3,75(1)(x^2 - \frac{5}{4})$

substituindo $\mathrm{X} = \dfrac{\mathrm{X} - 90}{60} = \dfrac{\mathrm{X}}{60} - \dfrac{3}{2}$ obteremos

$\hat{Y} = 37,475 + 0,2225X - 0,001042X^2$

X	$\bar{Y}$	$\hat{Y}$	$\bar{Y} - \hat{Y}$
0	37,75	37,4750	0,2750
60	46,25	47,0738	-0,8238
120	50,00	49,1702	0,8298
180	43,50	43,7642	-0,2642
			0,0168

DOSE ÓTIMA

$$\hat{y} = a + bX + cX^2$$

Ponto de máximo : (i) $\dfrac{d\hat{y}}{dX} = 0 \quad \Rightarrow X^*$ (ponto crítico)

(ii) $\dfrac{d\hat{y}^2}{dX^2} \; < 0 \;\; \Rightarrow$ ponto de máximo

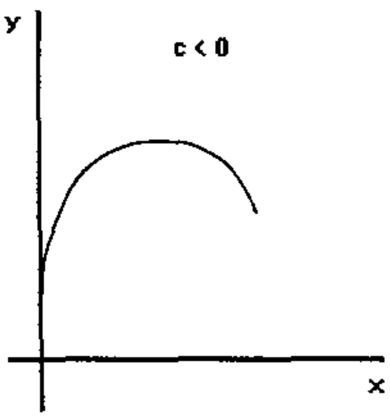
$> 0 \;\; \Rightarrow$ ponto de mínimo

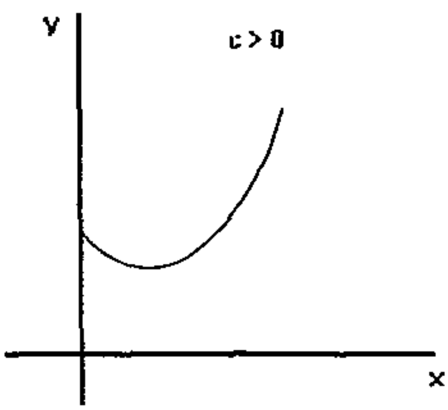
$$\frac{d\hat{y}}{dX} = b + 2cX = 0 \quad \Rightarrow X^* = \frac{-b}{2c}$$

$$\frac{d\hat{y}^2}{dX^2} = 2c$$

$\therefore$ c < 0 $\quad \Rightarrow$ ponto de máximo

c > 0 $\quad \Rightarrow$ ponto de mínimo





No exemplo

c = -0,001042
b = 0,2225

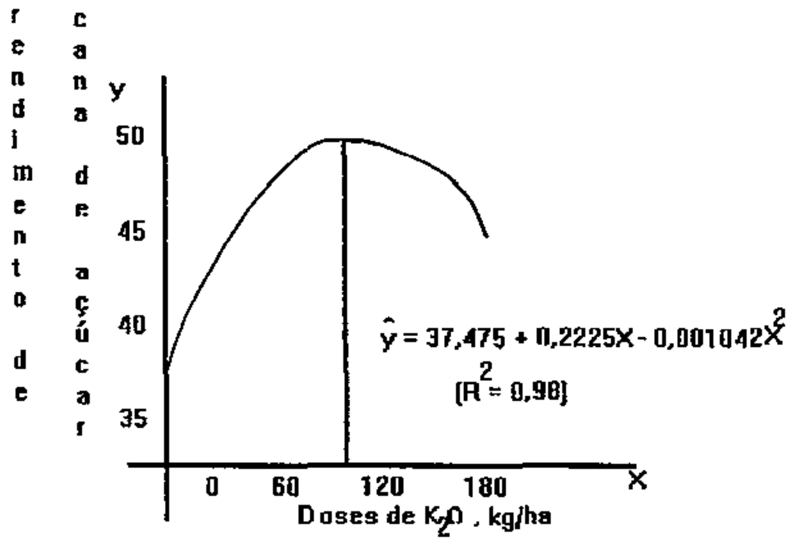
$$X^{*} = \frac{-0,2225}{2(-0,001042)} = 106,76 \cong 107$$

É ponto de máximo pois c < 0

$\therefore$ dose ótima 106,76(107) kg/ha de K_2O que nos leva a um rendimento máximo. (melhor tratamento)

- O ponto determinado é o ponto de máxima eficiência técnica

- Poder-se-ia determinar o ponto de máxima eficiência econômica.



para X = 107 $\hat{y}$ = 49,3526

PONTO DE MÁXIMA EFICIÊNCIA ECONÔMICA

Deseja-se tornar máxima a receita líquida

$$L = w\hat{y} - uX - m$$

custos fixos

dose do nutriente

preço unitário do nutriente

estimativa de produção dada pela curva de resposta

preço do produto

Para determinação de máximo, deriva-se e iguala-se a zero:

$$\frac{dL}{dX} = w\hat{y}' - u = 0 \ ; \quad \text{como} \quad \hat{y} = a + bX + cX^2$$

$$\hat{y}' = b + 2cX \ ,$$

então temos

$$w(b + 2cX) - u = o \Rightarrow b + 2cX - \frac{u}{w} = 0 \Rightarrow 2cX = \frac{u}{w} - b$$

Logo $X^* - \dfrac{\frac{u}{w} - b}{2c}$, X^* depende da relação de preços

dose de máxima eficiência econômica

EXEMPLO 6:

Um experimento usando 3 variedades de cana-de-açúcar (v_1, v_2, v_3) e 3 níveis de nitrogênio (n_1, n_2, n_3), foi conduzido no Delineamento Blocos Casualizados com 4 repetições. Os níveis de nitrogênio foram, respectivamente, 170, 240 e 310 kg/ha . Os rendimentos de cana foram:

Blocos	Tratamentos								
	v_1n_1	v_1n_2	v_1n_3	v_2n_1	v_2n_2	v_2n_3	v_3n_1	v_3n_2	v_3n_3
1	70,5	67,3	79,9	58,6	64,3	64,4	65,8	64,1	56,3
2	67,5	75,9	72,8	65,2	48,3	67,3	68,3	64,8	54,7
3	63,9	72,2	64,8	70,2	74,0	78,0	72,7	70,9	66,2
4	64,2	60,5	86,3	51,8	63,6	72,0	67,6	58,3	54,4

Utilizando o pacote estatístico SANEST (Sistema de Análise Estatística) obteve-se os seguintes resultados:

```
**************************************************************
*        SANEST - SISTEMA DE ANALISE ESTATISTICA             *
*  Autores: Elio Paulo Zonta  -  Amauri Almeida Machado      *
*            Departamento de Estatistica - UFRGS             *
*    ANALISE DA VARIAVEL RENDIMEN - ARQUIVO: FATORIAL        *
**************************************************************
```

CODIGO DO PROJETO: FATORIAL

RESPONSAVEL: RIBOLDI

DELINEAMENTO EXPERIMENTAL: DBC

OBSERVACOES NAO TRANSFORMADAS

NOME DOS FATORES

FATOR	NOME
A	VARIEDAD
B	NITROGEN

QUADRO DA ANALISE DE VARIANCIA

CAUSAS DA VARIACAO	G.L.	S.Q.	Q.M.	VALOR F	PROB.>F
BLOCOS	3	200.6898129	66.8966043	1.5236	0.23306
VARIEDAD	2	319.3756054	159.6878027	3.6370	0.04067
NITROGEN	2	56.5422232	28.2711116	0.6439	0.53846
VAR*NIT	4	559.7863193	139.9465798	3.1873	0.03067
RESIDUO	24	1053.7684641	43.9070193		
TOTAL	35	2190.1624249			

MEDIA GERAL = 66.322220

COEFICIENTE DE VARIACAO = 9.991 %

TESTE DE TUKEY PARA MEDIAS DE VARIEDAD
DENTRO DE N1 DO FATOR NITROGEN

NUM.ORDEM	NUM.TRAT.	NOME	NUM.REPET.	MEDIAS	MEDIAS ORIGINAIS	5%	1%
1	3	V3	4	68.599998	68.599998	a	A
2	1	V1	4	66.524994	66.524994	a	A
3	2	V2	4	61.450001	61.450001	a	A

TESTE DE TUKEY PARA MEDIAS DE VARIEDAD
DENTRO DE N2 DO FATOR NITROGEN

NUM.ORDEM	NUM.TRAT.	NOME	NUM.REPET.	MEDIAS	MEDIAS ORIGINAIS	5%	1%
1	1	V1	4	68.975006	68.975006	a	A
2	3	V3	4	64.524994	64.524994	a	A
3	2	V2	4	62.550003	62.550003	a	A

TESTE DE TUKEY PARA MEDIAS DE VARIEDAD
DENTRO DE N3 DO FATOR NITROGEN

NUM.ORDEM	NUM.TRAT.	NOME	NUM.REPET.	MEDIAS	MEDIAS ORIGINAIS	5%	1%
1	1	V1	4	75.950005	75.950005	a	A
2	2	V2	4	70.425003	70.425003	a	AB
3	3	V3	4	57.900002	57.900002	b	B

MEDIAS SEGUIDAS POR LETRAS DISTINTAS DIFEREM ENTRE SI AO NIVEL DE SIGNIFICANCIA INDICADO
D.M.S. 5% = 11.69531 - D.M.S. 1% = 15.07469

REGRESSAO POLINOMIAL PARA OS NIVEIS DE NITROGEN
DENTRO DE V1 DO FATOR VARIEDAD

QUADRO DA ANALISE DE VARIANCIA

CAUSAS DA VARIACAO	G.L.	S.Q.	Q.M.	VALOR F	PROB.>F
REGRESSAO LINEAR	1	177.6616527	177.6616527	4.04632	0.05290
REGRESSAO QUADR.	1	13.6503338	13.6503338	0.31089	0.58856
RESIDUO	24	1053.7684641	43.9070193		

E Q U A C O E S   P O L I N O M I A I S

	X	X^2
Y = 54.326174	0.0673215	
Y = 79.413683	-0.1543105	0.00046173

MEDIAS AJUSTADAS PELAS EQUACOES DE REGRESSAO

NIVEIS	MEDIAS OBS.	MEDIAS ORIG.	LINEAR	QUADR.
170.000	66.5250	66.5250	65.7708	66.5250
240.000	68.9750	68.9750	70.4833	68.9750
310.000	75.9500	75.9500	75.1958	75.9500
COEF. DETERMINACAO			0.9286	1.0000

REGRESSAO POLINOMIAL PARA OS NIVEIS DE NITROGEN
DENTRO DE V2 DO FATOR VARIEDAD

QUADRO DA ANALISE DE VARIANCIA

CAUSAS DA VARIACAO	G.L.	S.Q.	Q.M.	VALOR F	PROB.>F
REGRESSAO LINEAR	1	161.1013322	161.1013322	3.66915	0.06439
REGRESSAO QUADR.	1	30.6003960	30.6003960	0.69694	0.58308
RESIDUO	24	1053.7684641	43.9070193		

E Q U A C O E S   P O L I N O M I A I S

	X	X^2
Y = 49.422617	0.0641072	
Y = 86.984680	-0.2677295	0.00069133

MEDIAS AJUSTADAS PELAS EQUACOES DE REGRESSAO

NIVEIS	MEDIAS OBS.	MEDIAS ORIG.	LINEAR	QUADR.
170.000	61.4500	61.4500	60.3208	61.4500
240.000	62.5500	62.5500	64.8083	62.5500
310.000	70.4250	70.4250	69.2958	70.4250
COEF. DETERMINACAO			0.8404	1.0000

REGRESSAO POLINOMIAL PARA OS NIVEIS DE NITROGEN
DENTRO DE V3 DO FATOR VARIEDAD

QUADRO DA ANALISE DE VARIANCIA

CAUSAS DA VARIACAO	G.L.	S.Q.	Q.M.	VALOR F	PROB.>F
REGRESSAO LINEAR	1	228.9798694	228.9798694	5.21511	0.02980
REGRESSAO QUADR.	1	4.3349585	4.3349585	0.09873	0.75398
RESIDUO	24	1053.7684641	43.9070193		

E Q U A C O E S P O L I N O M I A I S

		X	X^2
Y =	82.017850	-0.0764285	
Y =	67.880162	0.0484688	-0.00026020

MEDIAS AJUSTADAS PELAS EQUACOES DE REGRESSAO

NIVEIS	MEDIAS OBS.	MEDIAS ORIG.	LINEAR	QUADR.
170.000	68.6000	68.6000	69.0250	68.6000
240.000	64.5250	64.5250	63.6750	64.5250
310.000	57.9000	57.9000	58.3250	57.9000
COEF. DETERMINACAO			0.9814	1.0000

12 - ANÁLISE DE CORRELAÇÃO

Correlação é a medida do grau em que duas variáveis variam juntas (= intensidade de associação).

Na correlação :

Relação de dependência $\nLeftrightarrow$ relação de causa e efeito

$$\underbrace{y}_{\text{variável dependente}} = f(\underbrace{x}_{\text{variável independente}})$$

Na correlação :

Não existe relação de dependência . Busca-se verificar somente o grau de associação entre as variáveis . Por exemplo: altura de irmãs e irmãos . As variáveis não apresentam relação de dependência mas certamente estão associadas pelo mecanismo da hereditariedade .

Coeficiente de Correlação

A medida de associação entre as variáveis x e y é dada pelo coeficiente de correlação r onde :

$$r = \frac{\text{covariância }(x,y)}{\sqrt{\text{variância}(x)\cdot\text{variância}(y)}} = \frac{SPxy\ /\ n-1}{\sqrt{\dfrac{SQx}{n-1}\cdot\dfrac{SQy}{n-1}}} =$$

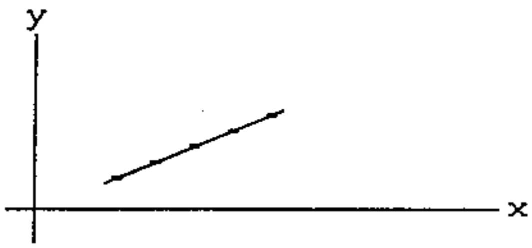
$$= \frac{SPxy}{\sqrt{SQx\cdot SQy}}$$

$$SPxy = \Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n} \quad ; \quad SQx = \Sigma x^2 - \frac{(\Sigma x)^2}{n} \qquad SQy = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$$

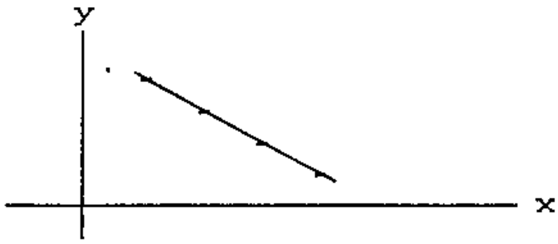
r é um valor adimencional pois independe da unidade de medida das variáveis .

$-1 \leq r \leq 1$

Para verificar a associação entre as variáveis x e y , o diagrama de dispersão dá uma idéia ; permite observar o grau e o tipo de correlação existente .

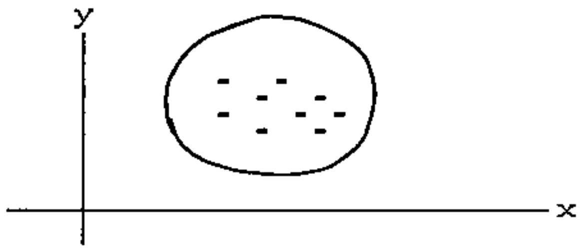


r = 1

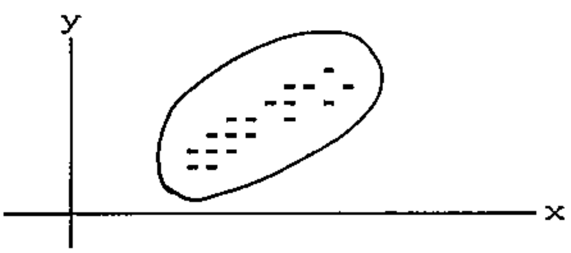


r = -1

r = 0



Em geral , $-1 < r < 1$ onde os pontos no diagrama de dispersão estão dentro da elipse .



A correlação entre x e y será maior quanto menor o eixo menor da elipse .

Teste de significância do coeficiente da correlação

Hipóteses : $H_o : \rho = 0$ (não existe correlação ou a correlação entre x e y é nula)

$H_a : \rho \neq 0$

Se rejeitarmos $H_o \rightarrow \rho \neq 0 \rightarrow$ as variáveis estão associadas e r é uma estimativa do grau de associação entre elas.

Para se testar Ho pode-se utilizar dois procedimentos :

(i) Teste t : $t = \dfrac{r}{s_r}$ → erro padrão do coeficiente de correlação .

$$s_r = \sqrt{(1-r)^2/n-2}$$

n = numero de pares de valores

g.l = n-2

Rejeita-se Ho quando $|t_{calculado}| > t_{\alpha(n-2)}$

(ii) Pela tabela de r :

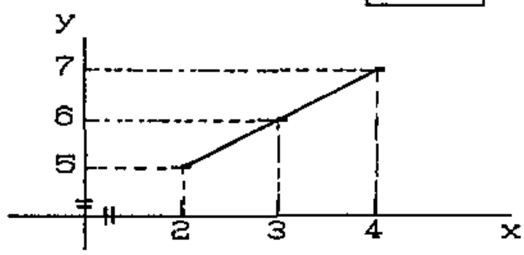
Rejeita-se Ho se $|r_{calculado}| > r_{\alpha(n-2)}$

Exemplos

(i)

x	y
2	5
3	6
4	7

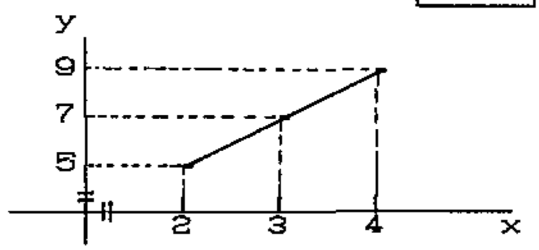
$\boxed{r = 1}$



(ii)

x	y
2	5
3	7
4	9

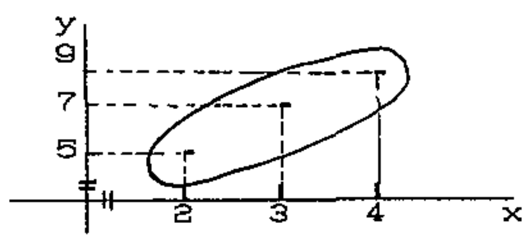
$\boxed{r = 1}$



(iii)

x	y
2	5
3	7
4	8

$0 < r < 1$

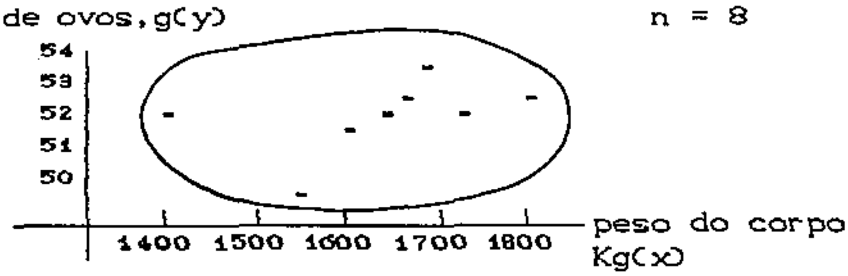


Exemplo 7:

Num estudo sobre seleção de matéria prima destinada a formação de populações básicas para o melhoramento de galinhas, obteve-se os seguintes resultados :

Plantel	peso do corpo, Kg(x)	peso do ovo, g(y)
1	1,615	52,13
2	1,809	58,90
3	1,677	52,85
4	1,684	53,58
5	1,651	52,40
6	1,784	52,46
7	1,569	49,60
8	1,457	52,27
total	13,246	419,19

(a) Diagrama de dispersão



(b) Cálculo do coeficiente de correlação r

$SQ_x = 1,615^2 + 1,809^2 + \ldots + 1,457^2 - (13,246)^2/8 = 0,0898935$

$SQ_y = 52,13^2 + 52,90^2 + \ldots + 52,27^2 - (419,19)^2/8 = 10,2382875$

$SP_{xy} =$

$$(1,615)(52,13)+(1,809)(52,90)+\ldots+(1,457)(52,27)-\frac{(13,246)(419,19)}{8}$$

$$= 0,3662075$$

$$r = \frac{SP_{xy}}{\sqrt{SQ_x \cdot SQ_y}} = \frac{0,3662075}{\sqrt{(0,0898935 \times 10,2382875)}} = 0,382$$

(c) $H_o : \rho = 0$

$H_a : \rho \neq 0$ $\quad r_{0,05(6)} = 0,707 \quad$ GL = n-2 = 6

$|r| = 0,382 < r_{tabelado} = 0,707 \Rightarrow$ não se rejeita $H_o \Rightarrow$ r é não significativo

As evidências amostrais não são suficientes para comprovar que exista associação entre o peso do corpo e o peso do ovo em galinhas .

Desta forma não podemos selecionar para peso do corpo quando o objetivo é o peso do ovo.

<u>Considerações sobre a interpretação do coeficiente de correlação r</u>

- causa e efeito

$$r = 0,90 \begin{cases} x = \text{salário de prof. universitários} \\ y = \text{consumo de bebidas alcóolicas} \\ \qquad \text{ou} \\ \quad \text{consumo de automóveis} \end{cases}$$

x_1 = aumento geral do padrão de vida no país

$$r = 0,80 \begin{cases} x = \text{altura de alunos} \\ y = \text{notas de provas de estatística} \end{cases}$$

Não há causa e efeito e não há terceira variável

<u>Utilidade prática de correlação e regressão</u>

<u>r</u> estima a intensidade da relação linear entre y e x , sendo que para muitas situações para uma etapa inicial de investigação pode satisfazer .

Mas quem responde perguntas do tipo :

- Quanto varia y devido a uma modificação em x ?
- Qual a função (curva) que relaciona y com x ?
- Com que precisão pode-se prever y partindo de x ?

É a <u>análise de regressão</u>

<u>Em experimentos</u> :

- Explicação de efeito de fatores : análise de regressão
- Relação entre variáveis dependentes : análise de correlação

<u>Relação entre os coeficientes de correlação e de regressão</u>

r : coeficiente de correlação

r^2: coeficiente de determinação ; $\hat{y} = a + bx$

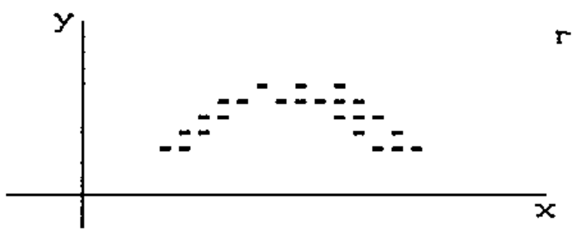
$r = \sqrt{r^2}$ ou $r^2 = (r)^2$

$$r = \frac{SPxy}{\sqrt{(SQx)(SQy)}} = \frac{SPxy}{\sqrt{SQx}\,\sqrt{SQy}} \cdot \frac{\sqrt{SQx}}{\sqrt{SQy}}$$

$$= \underbrace{\frac{SPxy}{SQx}}_{b} \cdot \frac{\sqrt{SQx}}{\sqrt{SQy}} = b\sqrt{\frac{SQx}{SQy}}$$

$$\therefore r = b\sqrt{\frac{SQx}{SQy}} \quad \text{ou} \quad b = r\sqrt{\frac{SQy}{SQx}}$$

No caso



r = 0

Relação quádrica entre x e y

$\hat{y} = a + bx + cx^2$

$R^2 = 0,80$

13 - REGRESSÃO LINEAR MÚLTIPLA

13.1 - Introdução:

A regressão múltipla se aplica ao caso em que se considera mais do que uma variável independente na estimativa da variável dependente Y, ou seja:

$$\underset{\text{variável dependente}}{Y} = f(\underbrace{X_1 , X_2 , \ldots, X_k}_{\text{variáveis independentes}})$$

Exemplos:

(i) Produção agrícola

$$Y = f(X_1 , X_2 , X_3) \quad \text{Função de Produção}$$

X_1 ↑ N, X_2 ↑ P, X_3 ↑ K

Y ↑ Rendimento de trigo

(ii) Nutrição Animal

$$Y = f(X_1 , X_2)$$

X_2 ↑ Ração consumida

X_1 ↑ Peso inicial

Y ↑ Ganho de peso

(iii) Social e Econômico

$$Y = f(X_1 , X_2 , X_3) \quad \text{Função de Demanda}$$

↑ Rênda relativa

↑ Nível educacional

↑ Urbanização relativa

Y ↑ Utilização de um produto em determinada área geográfica

13.2 - Modelo Matemático e Pressuposições:

O modelo matemático da regressão linear múltipla é dado por $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i$ onde $i = 1, 2, \dots, n$; $X_1, X_2, \dots, X_k$ são variáveis independentes; Y = variável dependente; $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ são parâmetros do modelo.

Pressuposições:

(a) A variável dependente Y é função linear das variáveis independentes $X_1, X_2, \dots, X_k$.

(b) Os valores das variáveis independentes são fixos (se não forem não há maiores problemas).

(c) $\varepsilon_i \cap N(0, \sigma^2)$

(d) ε_i são independentes.

13.3 - Relação linear entre 3 variáveis:

13.3.1. Equação de Regressão

Considerando as variáveis Y, X_1 e X_2 sob a hipótese de relacionamento linear entre as 3 variáveis pode-se escrever

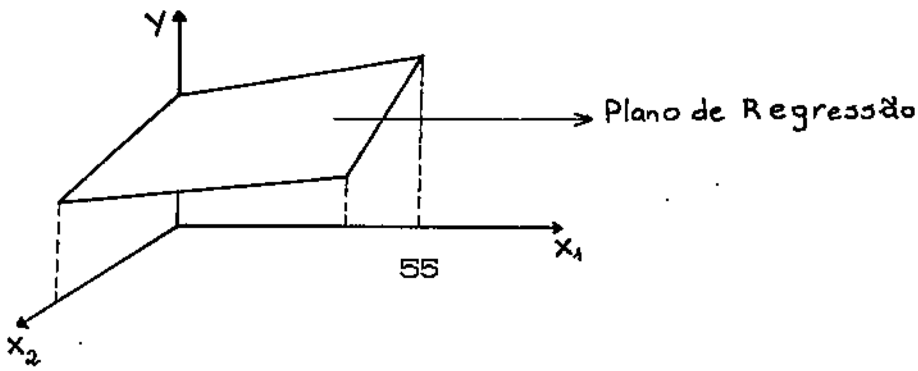
$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

onde $i = 1, 2, \dots, n$ e Y é função linear dos X.

Baseados nesse modelo linear estima-se a relação denominada de regressão de Y sobre X_1 e X_2 , dada por

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_{2i} X_{2i}$$

Nesse caso a equação de regressão representa um plano no espaço tridimensional, isto é



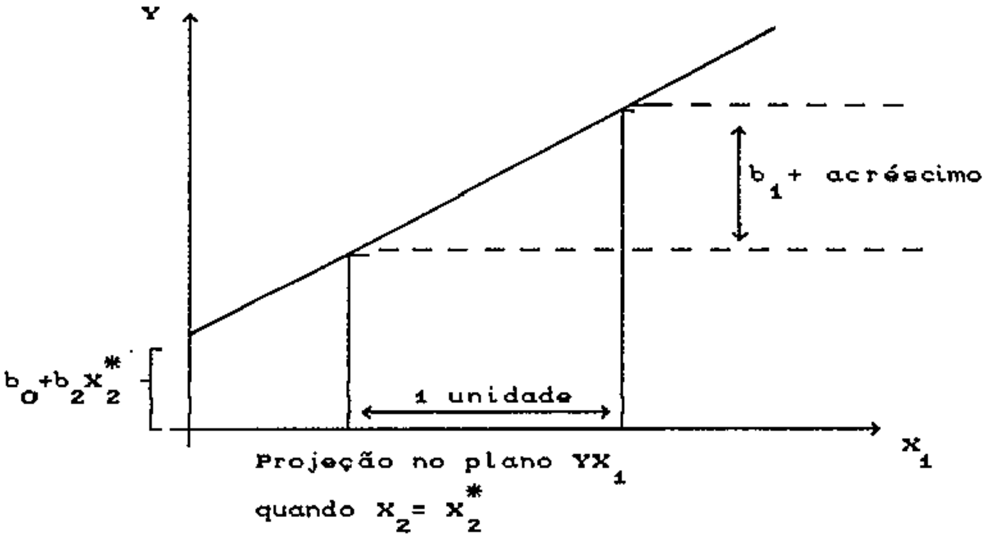
Na Equação de Regressão

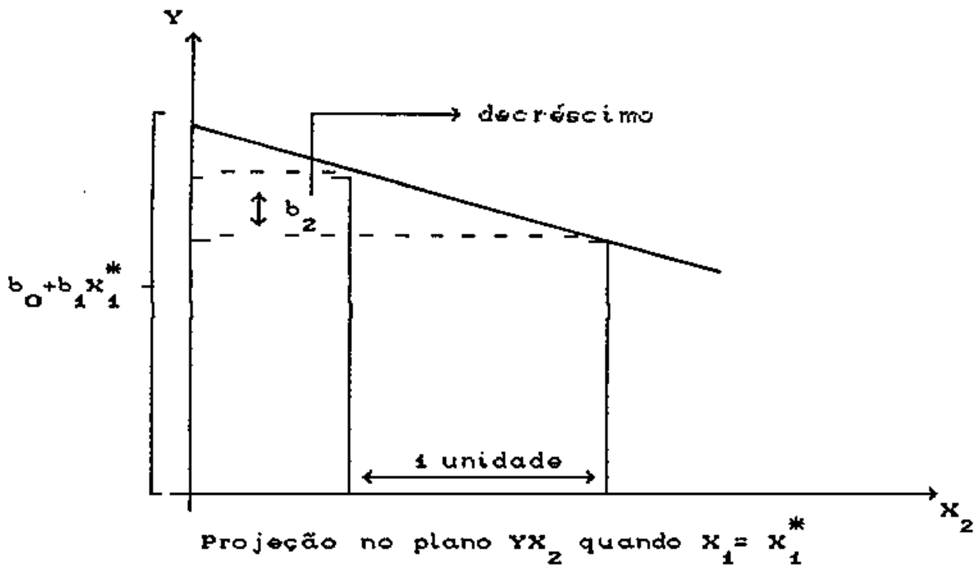
b_0= intercepto do plano de regressão com eixo dos Y, e é uma estimativa do parâmetro β_0; ou seja, estimativa para Y quando X_1 e X_2 são iguais a zero (X_1= 0 , X_2= 0) b_1 e b_2 são chamados coeficientes de regressão parciais, onde

b_1= quantidade que afeta Y dada uma variação de uma unidade em X_1 , permanecendo constante X_2 .

b_2= quantidade que afeta Y dada uma variação unitária em X_2 , permanecendo constante X_1 .

b_1 e b_2 consistem na declividade das retas produzidas pela projeção do plano de regressão sobre os planos YX_1 e YX_2 respectivamente, ou seja





13.3.2. <u>Ajustamento do plano de regressão:</u>

(= Obtenção dos coeficientes da equação de regressão)

A estimativa da relação entre as variáveis Y, X_1 e X_2 é dada por:

$$\hat{Y} = b_0 + b_1X_{1i} + b_2X_{2i}$$

Os desvios ou erros de estimativa são expressos por $e_i = Y_i - \hat{Y}_i$ e tem-se que a soma dos quadrados dos desvios é:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - b_0 - b_1X_{1i} - b_2X_{2i})^2$$

Deseja-se obter o melhor plano de regressão, ou seja, o plano de regressão que melhor represente o relacionamento entre as variáveis Y, X_1 e X_2 , isto é, o plano de regressão que minimize a soma de quadrados dos desvios ou erros de estimativa.

A equação de regressão é dada em função dos coeficientes b_0, b_1 e b_2 ; portanto deseja-se estabelecer as expressões de cálculo desses coeficientes de forma que os desvios sejam mínimos.

O cálculo de b_o , b_1 e b_2 requer um conjunto de três equações normais, obtidas pelo método dos quadrados mínimos (de forma análoga a regressão linear simples) e pela resolução, das quais obtém-se as seguintes expressões de cálculo dos coeficientes:

$$b_o = \overline{Y} - b_1\overline{X}_1 - b_2\overline{X}_2 \qquad b_2 = \frac{(SP_{Y2})(SQ_1)-(SP_{Y1})(SP_{12})}{(SQ_1)(SQ_2)-(SP_{12})^2}$$

$$b_1 = \frac{(SP_{Y1})(SQ_2)-(SP_{Y2})(SP_{12})}{(SQ_1)(SQ_2)-(SP_{12})^2}$$

1 = Variável X_1

2 = Variável X_2

Obtenção e Solução das equações normais:

$$\hat{y} = b_o + b_1x_1 + b_2x_2$$

$$y-\hat{y} = y - b_o - b_1x_1 - b_2x_2$$

$$z = \sum (y-\hat{y})^2 = \sum (y-b_o-b_1x_1-b_2x_2)^2$$

$$\left.\begin{aligned} \frac{\partial z}{\partial b_o} &= 2\sum (y-b_o-b_1x_1-b_2x_2)(-1) = 0 \\ \frac{\partial z}{\partial b_1} &= 2\sum (y-b_o-b_1x_1-b_2x_2)(-x_1) = 0 \\ \frac{\partial z}{\partial b_2} &= 2\sum (y-b_o-b_1x_1-b_2x_2)(-x_2) = 0 \end{aligned}\right\} \div (-2)$$

(1) $\sum y - nb_o - b_1 \sum x_1 - b_2 \sum x_2 = 0 \rightarrow nb_o + b_1 \sum x_1 + b_2 \sum x_2 = \sum y$

$\div n \qquad b_o + b_1 \dfrac{\sum x_1}{n} + b_2 \dfrac{\sum x_2}{n} = \dfrac{\sum y}{n} \rightarrow b_o + b_1\bar{x}_1 + b_2\bar{x}_2 = \bar{y}$

$\therefore \; b_o = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2$

$$(2)\ \Sigma x_1y - b_0\Sigma x_1 - b_1\Sigma x_1^2 - b_2\Sigma x_1x_2 = 0$$

$$\rightarrow \quad b_0\Sigma x_1 + b_1\Sigma x_1^2 + b_2\Sigma x_1x_2 = \Sigma x_1y$$

$$(\bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2)\Sigma x_1 + b_1\Sigma x_1^2 + b_2\Sigma x_1x_2 = \Sigma x_1y$$

$$\overset{\uparrow}{\frac{\Sigma y}{n}} \quad \overset{\uparrow}{\frac{\Sigma x_1}{n}} \quad \overset{\uparrow}{\frac{\Sigma x_2}{n}}$$

$$\frac{\Sigma x_1\Sigma y}{n} - b_1\frac{(\Sigma x_1)^2}{n} - \frac{b_2(\Sigma x_1)(\Sigma x_2)}{n} + b_1\Sigma x_1^2 + b_2\Sigma x_1x_2 = \Sigma x_1y$$

$$b_1\underbrace{\left[\Sigma x_1^2 - \frac{(\Sigma x_1)^2}{n}\right]}_{SQx_1} + b_2\underbrace{\left[\Sigma x_1x_2 - \frac{(\Sigma x_1)(\Sigma x_2)}{n}\right]}_{SPx_1x_2} = \underbrace{\Sigma x_1y - \frac{(\Sigma x_1)(\Sigma y)}{n}}_{SPyx_1}$$

$$SPx_1x_2 = SP_{12} \qquad SPx_1 = SPy_1$$

$$b_1\ SQx_1 + b_2\ SPx_1x_2 = SPyx_1 \qquad SPyx_2 = SPy_2$$

$$b_1\ SQ_1 + b_2SP_{12} = SPy_1 \qquad SQx_2 = SQ_2$$

$$(3)\ \Sigma x_2y - b_0\ \Sigma x_2 - b_1\ \Sigma x_1x_2 - b_2\ \Sigma x_1^2 = 0$$

$$\rightarrow \quad b_0\ \Sigma x_2 + b_1\ \Sigma x_1x_2 + b_2\ \Sigma x_2^2 = \Sigma x_2y$$

Por analogia

$$b_1\ SP_{12} + b_2\ SQ_2 = SPy_2$$

Portanto tem-se o sistema de 2 equações e duas incógnitas

$$b_1\ SQ_1 + b_2\ SP_{12} = SPy_1$$

$$b_1\ SP_{12} + b_2\ SQ_2 = SPy_2$$

Utilizando-se qualquer método para a solução obtém-se:

$$b_1 = \frac{(SPy_1)(SQ_2)-(SP_{12})(SPy_2)}{(SQ_1)(SQ_2)-(SP_{12})^2} \qquad b_2 = \frac{(SQ_1)(SPy_2)-(SPy_1)(SP_{12})}{(SQ_1)(SQ_2)-(SP_{12})^2}$$

13.3.3. Contribuição Relativa das Variáveis Indpendentes

Como geralmente se trabalha com valores de magnitude e de unidades de medida distinta para se estabelecer a contribuição de cada uma das variáveis não se pode usar diretamente os coeficientes de regressão parciais, e por isso lança-se mão dos coeficientes de regressão padronizados (b_i') que são coeficientes de regressão independentes da escala de medida utilizada para as variáveis independentes.

No caso de 2 variáveis independentes os coeficientes de regressão padronizados são dados por:

$$b_1' = b_1 \frac{s_1}{s_Y} \quad \text{ou} \quad b_1' = b_1 \sqrt{\frac{SQ_Y}{SQ_1}} \Rightarrow b_1 = b_1' \sqrt{\frac{SQ_Y}{SQ_1}}$$

$$b_2' = b_2 \frac{s_2}{s_Y} \quad \text{ou} \quad b_2' = b_2 \sqrt{\frac{SQ_2}{SQ_Y}} \Rightarrow b_2 = b_2' \sqrt{\frac{SQ_Y}{SQ_2}}$$

$$\text{e } b_1' = \frac{r_{Y1} - r_{Y2} r_{12}}{1 - r_{12}^2} \quad ; \quad b_2' = \frac{r_{Y2} - r_{Y1} r_{12}}{1 - r_{12}^2}$$

onde: b_1, b_2 são os coeficientes de regressão parciais

s_Y = desvio padrão da variável dependente Y

s_1 = desvio padrão da variável independente X_1

s_2 = desvio padrão da variável independente X_2

r_{Y1} = coeficiente de correlação simples entre Y e X_1

r_{Y2} = coeficiente de correlação simples entre Y e X_2

r_{12} = coeficiente de correlação simples entre X_1 e X_2

13.3.4. <u>Análise de Variância para a Regressão Múltipla</u>

Variação Total (SQTotal ou SQ_Y)	Variação de Y explicada pela influência linear de X_1 e X_2 (SQRegressão)	Variação de Y não explicada pela influência linear de X_1 e X_2 (Variação Residual) (SQDesvios da Regressão)

$$SQ\ Total = SQ_Y = \Sigma Y^2 - \left(\Sigma Y\right)^2 / n$$

$$SQRegressão = b_1 SP_{Y1} + b_2 SP_{Y2}$$

$$SQDesvios\ da\ Regressão = SQ\ Total - SQRegressão$$

Tabela de Análise de Variância

Causas de Variação	GL	SQ	QM	F
Regressão	k ou p-1 (2)	SQRegressão	QMRegressão	$\frac{QMRegr.}{QMDes.Reg.}$
Desvios da Regressão	n-k-1 ou n-p (n-3)	SQ Des.Reg.	QM Des.Reg.	
Total	n-1	SQ Total		

onde k = número de variáveis independentes
p é o número de coeficientes da equação; entre parênteses os GL para o caso de duas variáveis independentes (três coeficientes na equação)

Coeficiente de Determinação Múltiplo (R^2 ou $R^2_{Y.12}$)

$$R^2 = \frac{SQRegressão}{SQ\ Total} \quad ; \quad R^2 = r_{Y1}b'_1 + r_{Y2}b'_2$$

R^2 representa a proporção da variação de Y explicada pela influência linear das variáveis independentes X_1 e X_2.

As hipóteses que se formulam no teste F da regressão múltipla são:

Ho: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$ (Ho: $\beta_1 = \beta_2 = 0$ no caso 3 variáveis)

Ha: $\beta_i \neq 0$ para pelo menos um i.

Rejeita-se Ho quando Fcalculado > F_{α}(GL Regressão, GL Desvios da Regressão) e nesse caso a regressão linear sobre X_1 e X_2 é significativa.

Para se testar hipóteses do tipo $\beta_1 = 0$; usa-se o desdobramento dos GL da regressão na análise de variância utilizando o teste F e para o caso de 2 variáveis independentes ter-se-ia:

Causas de Variação	GL
Regressão	2
Influência de X_1	1
Influência de X_2	1
Desvios da Regressão	n-3
Total	n-1

O teste F para a influência de X_1, consiste no teste de Ho: $\beta_1 = 0$; e o teste F para influência de X_2 , consiste no teste de Ho: $\beta_2 = 0$.

Na tabela de análise de variância, se X_1 e X_2 forem independentes entre si (não correlacionados ou coeficiente de correlação nulo entre elas), então: SQInfluência linear de X_1+ SQInfluência linear de X_2= SQRegressão; caso contrário essa igualdade não se verifica.

Se X_1 e X_2 forem independentes entre si:

$$\text{SQInfluência linear de } X_1 = \frac{(SP_{Y1})^2}{SQ_1}$$

$$\text{SQInfluência linear de } X_2 = \frac{(SP_{Y2})^2}{SQ_2}$$

Se X_1 e X_2 não forem independentes entre si (são correlacionados), calcula-se a SQInfluência linear de X_1 e a

SQInfluência linear de X_2 através do cálculo das SQ devida a adição das variáveis, ou seja, SQ de X_1 consiste na adição da soma de quadrados devida a inclusão de X_1 depois de X_2 já estar inclusa na equação; em relação a SQ de X_2 o procedimento é inverso.

Pode-se utilizar para testar hipóteses do tipo Ho: $\beta_i = 0$ ao invés do teste F o teste t onde

$$t = \frac{b_i}{\underbrace{s_{bi}}_{\text{erro padrão do coeficiente de regressão parcial}}}$$

Para o caso de 2 variáveis independentes tem-se:

$$t = \frac{b_1}{s_{b1}} \quad ; \quad s_{b1} = \sqrt{\underbrace{s^2 Y/12}_{\text{variância dos erros de estimativa}} \left[\frac{SQ_1}{SQ_1\ SQ_2 - (SP_{12})^2} \right]}$$

$$s^2 Y/12 = \frac{(1-R^2)\ SQ\ Total}{n-3} = \frac{SQ\ Desvios\ da\ Regressão}{n-3}$$

$$= QM\ Desvios\ da\ Regressão$$

$$e\ t = \frac{b_2}{s_{b2}} \quad ; \quad s_{b2} = \sqrt{s^2 Y/12 \left[\frac{SQ_1}{SQ_1\ SQ_2 - (SP_{12})^2} \right]}$$

Rejeita-se Ho quando $|t\ calculado| > t_{\alpha(n-3)}$

Nesse caso $F = t^2$; pois se tem 1 GL no numerador.

Na análise de variância utilizando a regressão linear múltipla quando existem valores repetidos de Y para a combinação de valores X_1 e X_2 , como é comum em caso de experimentos com fatores quantitativos, pode-se manter (o que é comumente mantido) o esquema de análise do delineamento empregado e desdobra-se a

SQTratamentos em SQRegressão e SQDesvios da Regressão testando com QME, possibilitando com isso um teste da falta de ajuste da regressão, através do teste de Desvios da Regressão.

13.4. EXEMPLO

Exemplo 8: Y: Rendimentos médios de soja, q/ha

X_1: Doses de calcário (C), t/ha

X_2: Doses de P_2O_5 (P), kg/ha

Calcário (C) t/ha X_1	P_2O_5 (P) kg/ha X_2	Rendimento médio de soja, q/ha Y
0	0	12
0	100	14
0	200	18
0	300	20
0	400	18
4	0	17
4	100	17
4	200	22
4	300	17
4	400	20
8	0	20
8	100	22
8	200	24
8	300	26
8	400	30

$\Sigma y = 297$

$\bar{y} = 19{,}8$

$\bar{x}_1 = 4$

$\bar{x}_2 = 200$

$SQy = 294{,}4$

$SQx_1 = SQ_1 = 160$

$SQx_2 = SQ_2 = 300000$

$SPx_1x_2 = SP_{12} = 0$

$SPyx_1 = SP_{y1} = 160$

$SPyx_2 = SP_{y2} = 4800$

Equação de Regressão:

$$b_1 = \frac{(SP_{y1})(SQ_2) - (SP_{y2})(SP_{12})}{(SQ_1)(SQ_2) - (SP_{12})^2} = \frac{(160)(300000) - (4800)(0)}{(160)(300000) - (0)^2} = 1 \text{ q/ha}$$

$$b_2 = \frac{(SP_{y2})(SQ_1) - (SP_{y1})(SP_{12})}{(SQ_1)(SQ_2) - (SP_{12})^2} = \frac{(4800)(160) - (160)(0)}{(160)(300000) - (0)^2} = 0{,}016 \text{ q/ha}$$

ou também:

$$r_{y1} = \frac{SP_{y1}}{\sqrt{(SQ_y)(SQ_1)}} = \frac{160}{\sqrt{(294{,}4)(160)}} = 0{,}737210$$

$$r_{y2} = \frac{SP_{y2}}{\sqrt{(SQ_y)(SQ_2)}} = \frac{160}{\sqrt{(294{,}4)(300000)}} = 0{,}520754$$

$$b'_1 = \frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2} = \frac{0{,}737210 - (0{,}510754)(0)}{1 - 0} = 0{,}737210$$

$$b'_2 = \frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2} = 0{,}520754$$

Como $r_{12} = 0 \Rightarrow b'_1 = r_{y1}$ e $b'_2 = r_{y2}$

$$b_1 = b'_1 \sqrt{\frac{SQ_y}{SQ_1}} = 0{,}737210 \sqrt{\frac{294{,}4}{160}} = 1 \text{ q/ha}$$

$$b_2 = b'_2 \sqrt{\frac{SQ_y}{SQ_2}} = (0{,}510754) \sqrt{\frac{294{,}4}{300000}} = 0{,}016 \text{ q/ha}$$

$$b_0 = \bar{y} - b_1\bar{x}_1 - b_1\bar{x}_2 = 19{,}8 - (1)(4) - (0{,}016)(200) = 12{,}6 \text{ q/ha}$$

$\therefore \bar{y} = 12{,}6 + x_1 + 0{,}016x_2 \rightarrow$ Equação de Regressão Múltipla

<u>Interpretação dos Coeficientes da Equação de Regressão</u>

$b_0 = 12{,}6$ q/ha $\rightarrow$ Em condições de Fertilidade Natural, isto é, sem aplicação de calcário e fósforo, estima-se em 12,6 q/ha o rendimento de soja.

$b_1 = 1$ q/ha $\rightarrow$ Estima-se em 1 q/ha o acréscimo no rendimento de soja a cada tonelada por ha de calcário aplicada ao solo, mantendo-se constante o fósforo.

$b_2 = 0{,}016$ q/ha $\rightarrow$ Estima-se em 0,016 q/ha (ou 1,6 kg/ha) o acréscimo no Rendimento de soja a cada kg/ha de fósforo aplicado ao solo, mantendo-se constante o calcário.

<u>Análise de Variância</u>

<u>SQ</u> SQ Total $= SQ_y = 294{,}4$

SQ Regressão $= b_1 SP_{y1} + b_2 SP_{y2} = (1)(160) + (0{,}016)(4800)$

$= 160 + 76{,}8 = 236{,}8$

SQ Desvios da Regressão = SQ Total − SQ Regressão

$= 294{,}4 - 236{,}8 = 57{,}6$

<u>GL</u> Total $= n - 1 = 15 - 1 = 14$

Regressão $= k = 2$ ou Regressão $= p - 1 = 3 - 1 = 2$

Desvios da Regressão = 14 - 2 = 12

Causas de Variação	GL	SQ	QM	F
Regressão	2	236,8	118,4	26,67**
X_1	1	160,0	160,0	33,33**
X_2	1	76,8	76,8	16,00
Desvios de Regressão	12	57,6	4,8	
Total	14	294,4		

Teste F para Regressão

Ho: $\beta_1 = \beta_2 = 0$

Ha: $\beta_i \neq 0$ para pelo menos um i

$F = 24,67 > F_{.01}(2,12) = 6,93$ → Rejeita-se Ho

A Regressão Linear Múltipla do rendimento de soja para doses de calcário e fósforo aplicados ao solo, explica fração significativa da variação do rendimento de soja.

Coeficiente de Determinação

$$R^2 = \frac{SQRegressão}{SQTotal} = \frac{236,8}{294,4} = 0,80$$

80% da variação no rendimento de soja é explicada pela Regressão Linear Múltipla do Rendimento para doses de calcário e fósforo aplicados ao solo.

20% ⟶ Outros fatores, dentre os quais interação de calcário e fósforo, termos quadráticos,...

Influência de cada variável independente

$$\text{SQ Influência de } x_1 = \frac{(SP_{y1})^2}{SQ_1} = \frac{(160)^2}{160} = 160$$

$$\text{SQ Influência de } x_2 = \frac{(SP_{y2})^2}{SQ_2} = \frac{(4800)^2}{300000} = 76,8$$

$Ho^{(i)}$: $\beta_1 = 0$ vs Ha: $\beta_1 \neq 0$

$F = 33,33 > F_{.01}(1,12) = 9,33 \Rightarrow$ Rejeita-se $Ho^{(i)} \Rightarrow \beta_1 \neq 0$

x_1 importante na estimação de y

↑ calcário ↑ Rendimento de soja

$Ho^{(ii)}$: $\beta_2 = 0$ vs Ha: $\beta_2 \neq 0$

$F = 16,00 > F_{.01}(1,12) = 9,33 \Rightarrow$ Rejeita-se $Ho^{(ii)} \Rightarrow \beta_2 \neq 0$

x_2 importante na estimação de y

↑ Fósforo ↑ Rendimento de soja

Influência de cada variável independente pelo teste t

$Ho^{(i)}$: $\beta_1 = 0$ vs Ha: $\beta_1 \neq 0$

$$t = \frac{b_1}{s_{b1}} = \frac{1}{0,1733} = 5,7735 ** $$

$$s_{b1} = \sqrt{\underbrace{s^2y/12}_{\text{QMDR}}\left[\frac{SQ_2}{(SQ_1)(SQ_2)-(SP_{12})^2}\right]} = \sqrt{4,8\left[\frac{300000}{(160)(300000)-(0)^2}\right]}$$

$$= 0,1732$$

$|t| = 5,7735 > t_{.01(12)} = 3.055 \rightarrow$ Rejeita-se $Ho^{(i)} \rightarrow \beta_1 \neq 0$

$t^2 = 33,33 = F$

$Ho^{(ii)}$: $\beta_2 = 0$ vs Ha: $\beta_2 \neq 0$

$$t = \frac{b_2}{s_{b2}} = \frac{0,016}{0,004} = 4 ** $$

$$s_{b2} = \sqrt{s^2 y/12 \left[\frac{SQ_2}{(SQ_1)(SQ_2)-(SP_{12})^2}\right]} = \sqrt{4,8\left[\frac{160}{(160)(300000)-(0)^2}\right]}$$

$$= 0,004$$

$|t| = 4 > t_{.01(12)} = 3.055 \Rightarrow$ Rejeita-se $Ho^{(ii)} \Rightarrow \beta_2 \neq 0$

$t^2 = 16 = F$

<u>Contribuição Relativa de x_1 e x_2</u>

Coeficientes de Regressão Parciais Padronizadas

$b'_1 = 0,73721$ ← Calcário mais importante para fins de previsão de rendimento de soja

$b'_2 = 0,51075$ Calcário 44% mais eficiente que o fósforo.

13.5. <u>REGRESSÃO MÚLTIPLA</u>

Qualquer caso de Regressão Múltipla, como qualquer tipo de regressão linear nos parâmetros, pode ser considerado como um caso particular da regressão linear múltipla, no que se refere a estimação dos parâmetros do modelo.

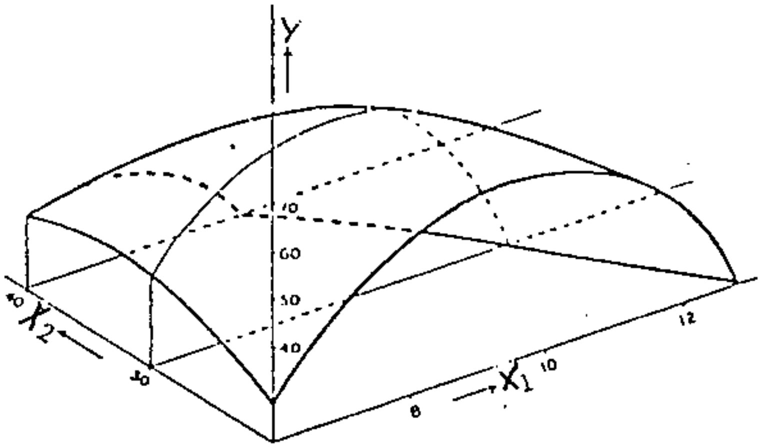
Comuns são modelos não lineares no estudo do relacionamento de variáveise um desses modelos é o modelo quadrático muito utilizado para funções de produção.

Esses modelos quadráticos são também chamados de modelos de superfície de resposta, onde os mesmos são expressos da seguinte maneira, em relação a equação estimada

(i) Para 2 variáveis

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_{11}X_1^2 + b_{22}X_2^2 + b_{12}X_1X_2$$

Como tem-se componentes quadráticos a equação de regressão fornece não mais um plano e sim uma superfície de resposta, que é representada a seguir.



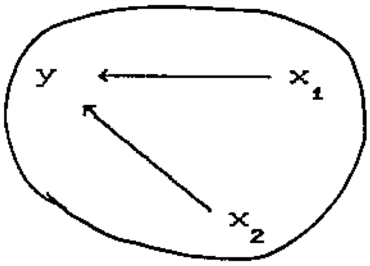
(ii) Para 3 variáveis

$$\hat{Y} = b_0 + b_1X_1 + b_1X_2 + b_3X_3 + b_{11}X_1^2 + b_{22}X_2^2 + b_{33}X_3^2 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3$$

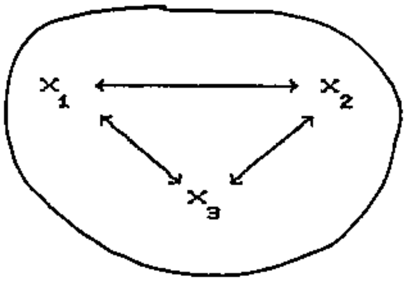
Não só a regressão múltipla constitui um caso particular da regressão linear múltipla como também a regressão linear simples ou as regressões curvilineares poilinomiais, bem como qualquer modelo que possa ser transformado em linear, constituem casos particulares de regressão linear múltipla.

14 - CORRELAÇÃO PARCIAL E MÚLTIPLA

14.1. Correlação Parcial: Se quisermos explicar a variação de Y através da influência relativa de X_1 e X_2, um estudo de Regressão é que se adequa, através da relação de dependência $Y = f(X_1, X_2)$.



São basicamente os casos onde Y é variável aleatória, isto é, sujeita a erro, e X_1 e X_2 são variáveis fixas.



Se quisermos verificar simplesmente a associação entre as variáveis, onde não existe uma relação funcional de dependência entre as variáveis, um estudo de correlação é apropriado e nesse caso a correlação é parcial.

São os casos em que X_1, X_2 e X_3 são variáveis aleatórias, isto é, sujeitas a erro de determinação.

Para uma distribuição a três variáveis o estudo de associação entre elas, deve ser feito através da correlação parcial, uma vez que procura-se uma associação líquida entre as variáveis, eliminando aquilo que se entende por correlação espúrea (associação devida a 3ª variável).

Assim sendo pode-se indagar se uma correlação entre X_1 e X_2 é devida meramente a uma associação destas variáveis com X_3, ou se existe na realidade uma associação líquida entre X_1 e X_2 que é medida através do coeficiente de correlação parcial.

Ao se calcular o coeficiente de correlação parcial entre X_1 e X_2 removemos a influência linear de X_3 de cada uma das variáveis e verifica-se que correlação existe entre os desvios não explicados que restam.

Os coeficientes de correlação parcial são expressos da seguinte maneira onde : 1 = variável X_1 ; 2 = variável X_2 e 3 = variável X_3 .

r12.3 = Correlação parcial entre X_1 e X_2 quando X_3 é constante

r13.2 = Correlação parcial entre X_1 e X_3 quando X_2 é constante

r23.1 = Correlação parcial entre X_2 e X_3 quando X_1 é constante

As expressões de cálculo dos coeficientes de correlação parciais são dadas por:

$$r12.3 = \frac{r12 - r13\ r23}{\sqrt{(1-r13^2)(1-r23^2)}}$$

$$r13.2 = \frac{r13 - r12\ r23}{\sqrt{(1-r12^2)(1-r23^2)}}$$

$$r23.1 = \frac{r23 - r12\ r13}{\sqrt{(1-r12^2)(1-r13^2)}}$$

Para se verificar a significância dos coeficientes de correlação parciais usa-se os valores tabelados de r ao nível α de significância com (n-3) GL.

Exemplo 9:

X_1= ocorrência de insetos em armadilhas luminosas

X_2= temperatura média

X_3= precipitação

1 = X_1 r12 = 0,41

2 = X_2 r13 = 0,42 n = 120

3 = X_3 r23 = 0,48

$$r12.3 = \frac{0{,}41 - (0{,}42)(0{,}48)}{\sqrt{(1-0{,}42^2)(1-0{,}48^2)}} = 0{,}25^*$$

$r._{05(117)} = 0{,}195$

Existe uma associação positiva entre a ocorrência de insetos e a temperatura média independentemente da precipitação.

$$r13.2 = \frac{0{,}42 - (0{,}41)(0{,}48)}{\sqrt{(1-0{,}41^2)(1-0{,}48^2)}} = 0{,}28^*$$

Existe uma associação positiva entre a ocorrência de insetos e a precipitação independentemente da temperatura média.

$r12.3 = 0{,}25 < r12 = 0{,}41$ A associação entre a ocorrência de insetos e a temperatura média é em parte devida a ação da precipitação.

$r13.2 = 0{,}28 < r13 = 0{,}42$ A associação entre a ocorrência de insetos e a precipitação é em parte devida a influência da temperatura.

14.2. Coeficiente de Correlação Múltipla:

Considerando as variáveis X_1, X_2 e X_3, o coeficiente de correlação múltipla (R ou $R_{1.23}$) é dado por .

$$R = \sqrt{R^2} \quad \text{ou} \quad R = \sqrt{1 - (1-r12^2)(1-r13^2)}$$

onde r12 = coeficiente de correlação simples entre X_1 e X_2, r13.2 = coeficiente de correlação parcial entre X_1 e X_3.

$0 \leq R \leq 1$; enquanto que coeficientes de correlação simples ou parciais variam de −1 a +1.

R ⟶ fornece informação sobre a influência das variáveis sobre alguma de interesse primeiro. Para se verificar a significância do coeficiente de correlação múltipla utiliza-se tabela específica, similar a utilização para o teste do coeficiente de correlação simples (r), considerando o maior nº. de variáveis envolvidas.

15 - B I B L I O G R A F I A

CAMPOS, H., 1984. **Estatística Aplicada à Experimentação com Cana-de-açúcar.** FEALQ.

COCHRAN, W.G. & G.M. COX, 1957. **Experimental Designs.** John Wiley.

GOMEZ, K.A. & A.A.GOMEZ, 1984. **Statistical Procedures for Agricultural Research.** John Wiley.

MARKUS, R., 1977. **Elementos de Estatística Aplicada, Partes I e II.** Faculdade de Agronomia/UFRGS.

MONTGOMERY, D.C., 1991. **Design and Analysis of Experiments.** John Wiley.

OSTLE, B. & R.W. MENSING, 1975. **Statistics in Research.** Iowa State University.

PIMENTEL GOMES, F., 1985. **Curso de Estatística Experimental.** Liv. Nobel.

SNEDECOR, G.M. & W.G. COCHRAN, 1980. **Statistical Methods.** Iowa State University.

STEEL, R.G. & J.H.TORRIE, 1980. **Principles and Procedures of Statistics.** McGraw-Hill.

Publicações do Instituto de Matemática da UFRGS

Cadernos de Matemática e Estatística

<u>Série B: Trabalho de Apoio Didático</u>

1. Elsa Mundstock - Curso Básico Sobre Wordstar 3.45 - MAR/89

2. Jaime B. Ripoll - Introdução ao Cálculo Diferencial Via Funções de Uma Variável Real - OUT/89

3. Edmund R. Puczylowski - Dimension of Modular Lattices - JUN/90

4. Marcos Sebastiani - Geometrias Não Euclidianas - JUL/90

5. Sandra R. C. Pizzatto - Cálculo Numérico - AGO/91

6. Vera Clotilde G. Carneiro - Elementos de Cálculo para Biologia - AGO/91

7. Elsa Mundstock - Iniciação ao SPSS/PC - SET/91

8. Elisa Hagg, Loiva C. de Zeni, Maria Alice Gravina e Vera Carneiro - Notas da 1ª Oficina de Matemática da UFRGS - JAN/92

9. Paulo Werlang de Oliveira, Elisabete Rambo, Suzana Lima dos Santos, Coordenação: Profª Maria Alice Gravina - A Tartaruga no Espaço Tridimensional - FEV/92

10. Silvio Possoli - Análise Multivariada - JUL/92

11. Dinara Westphalen Fernandez - Números Índices - OUT/92

12. Maria Teresinha Albanese - Coeficiente de Fidedignidade de um Instrumento de Medida - OUT/92

13. Vera Clotilde Carneiro e Sérgio Cláudio Ramos - Gráficos na Escola - DEZ/92

14. João Riboldi - Elementos Básicos de Estatística - JAN/93

15. Paulo W. de Oliveira e M. Alice Gravina - Logo: Manual do Usuário - MAR/93

16. Ruben Markus, Elsa C. de Mundstock, Dinara W. X. Fernandez e João Riboldi - Exercícios de Métodos Estatísticos - AGO/93

17. Loiva C. de Zeni e M. Alice Gravina - Sugestões de Atividades no Ambiente Logo para a Exploração de Conteúdos Matemáticos dos Currículos Escolares de 1º e 2º Grau - SET/93

18. João Riboldi - Delineamentos Experimentais de Campo, Parte 1 - SET/93

19. Marlusa Benedetti, Patrícia P. Gil, Shirley I. Techera, Angela Andreotti, Milene Milan, Marlise Moraes, Luciana Santos, Augustinho Zimmermann, Coordenação: Profª Maria Alice Gravina - Atividades em Geometria Usando Recortes - OUT/93

20. João Riboldi - Delineamentos Experimentais de Campo, Parte 2 - OUT/93

21. Anne C. Rutsatz, Edina R. de C. Alexandre, Gorete Losada, M. Alice Gravina, Rosamary P. Disconzi, Shirley Techera e Vera C. G. Carneiro - O Pensamento e a Linguagem da Algebra: Tabelas, Gráficos e Equações - DEZ/93

22. Dinara W. X. Fernandez - Estatística Descritica I - JAN/94

23. João Riboldi - Planejamento e Análise de Experimentos, Parte 1 - FEV/94

24. Dinara W. X. Fernandez - Estatística Descritiva II - AGO/94

25. João Riboldi e Lidia do C. S. C. do Nascimento - Metodologia de Superfície de Resposta: Uma Abordagem Introdutória - NOV/94

26. Júlio Cesar Ruiz Claeyssen - Métodos Operacionais e Computacionais em Algebra Matricial - MAR/95

27. João Riboldi - Análise de Variância - MAR/95

28. João Riboldi - Experimentos Fatoriais - MAI/95

29. João Riboldi - Planejamento e Análise de Experimentos, Parte 2 - JUN/95

30. João Riboldi e Dinara W. X. Fernandez - Análise de Regressão e Correlação - JUN/95

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
NÚCLEO DE ATIVIDADES EXTRACURRICULARES

Os Cadernos de Matemática e Estatística publicam as seguintes séries:

Série A: Trabalho de Pesquisa

Série B: Trabalho de Apoio Didático

Série C: Colóquio de Matemática SBM/UFRGS

Série D: Trabalho de Graduação

Série F: Trabalho de Divulgação

Série G: Textos para Discussão

Toda correspondência com solicitação de números publicados e demais informações deverá ser enviada para:

NAEC - NÚCLEO DE ATIVIDADES EXTRACURRICULARES
INSTITUTO DE MATEMÁTICA - UFRGS
AV. BENTO GONÇALVES, 9500 - PRÉDIO 43111
CEP 91509 - 900 AGRONOMIA - POA/RS
FONE: 336 92 22 OU 339 13 55 OU 228 16 33
RAMAL 6197
FAX: 336 15 12