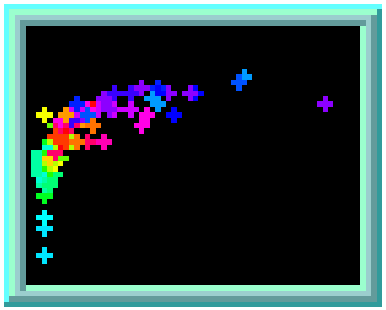


Regressão

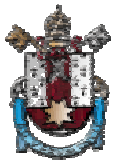
Prof. Lorí Viali, Dr.

viali@pucrs.br

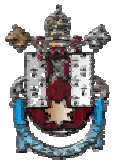
http://www.pucrs.br/~viali/

2

Em muitas situações duas ou mais variáveis estão relacionadas e surge então a necessidade de determinar a natureza deste relacionamento.



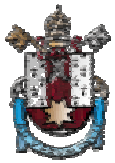
A análise de regressão é uma técnica estatística para modelar e investigar o relacionamento entre duas ou mais variáveis.



De fato a regressão pode ser dividida em dois problemas:

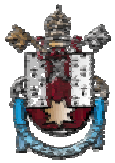
(i) o da especificação e

(ii) o da determinação.



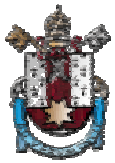
A especificação

O problema da especificação é descobrir dentre os possíveis modelos (linear, quadrático, exponencial, etc.) qual o mais adequado.



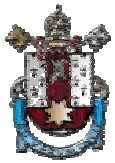
A determinação

O problema da determinação é uma vez definido o modelo (linear, quadrático, exponencial, etc.) estimar os parâmetros da equação.

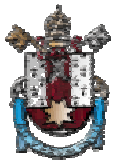


O modelo

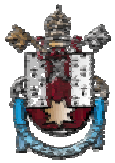
Normalmente é suposto que exista uma variável Y (dependente ou resposta), que está relacionada a "k" variáveis (independentes ou regressoras) X_i $(i = 1, 2, ..., k)$.



A variável resposta Y é aleatória, enquanto que as variáveis regressoras X_i são normalmente controladas. O relacionamento entre elas é caracterizado por uma equação denominada de "equação de regressão"



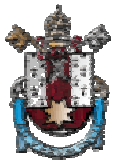
Quando existir apenas uma variável regressora (X) tem-se a regressão simples, se Y depender de duas ou mais variáveis regressoras, então tem-se a "regressão múltipla".



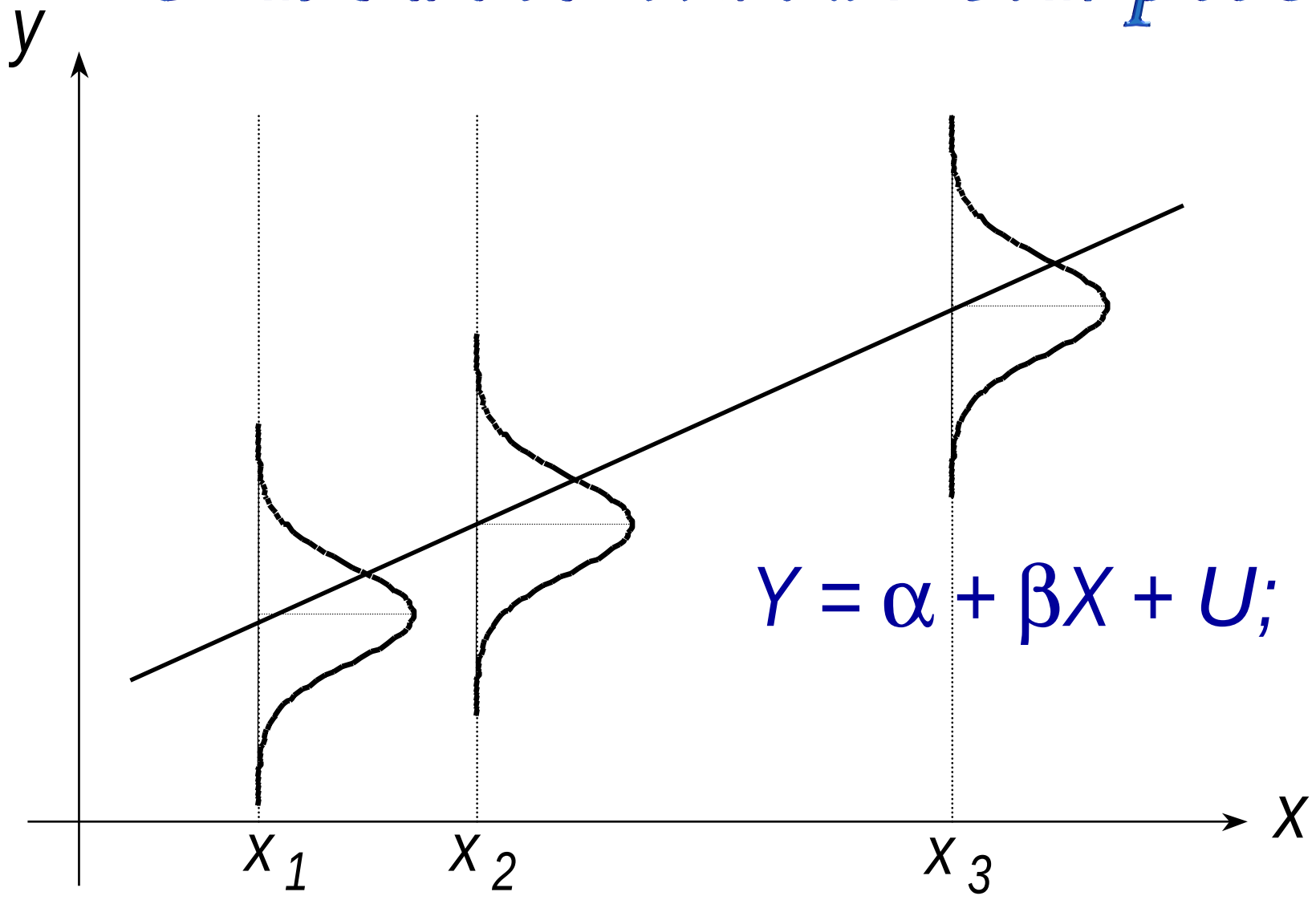
O modelo considerado

Vamos supor que a regressão é do tipo simples e que o o modelo seja linear, isto é, vamos supor que a equação de regressão seja do tipo:

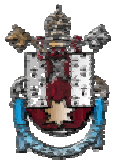
$$Y = \alpha + \beta X + U$$



O modelo linear simples

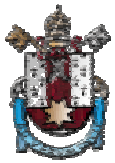


O termo "U" é o termo erro, isto é, "U" representa outras influências sobre a variável Y, além da exercida pela variável "X". A variação residual (termo U) é suposto de média zero e desvio constante e igual a σ.



Ou ainda pode-se admitir que o modelo fornece o valor médio de Y, para um dado "x", isto é,

$$E(Y/x) = \alpha + \beta X$$



Em resumo, as hipóteses são:

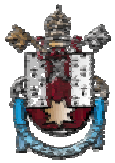
$Y = \alpha + \beta X + U;$

$E(Y/x) = \alpha + \beta X$, *isto é*, $E(U) = 0$

$V(Y/x) = \sigma^2;$

$Cov(Ui, Uj) = 0$, *para* $i \neq j;$

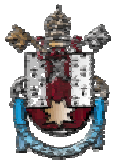
A variável X permanece fixa em observações sucessivas e os erros U são normalmente distribuídos.



A equação de regressão

O modelo suposto $E(Y/x) = \alpha + \beta X$ é populacional.

Vamos supor que se tenha n pares de observações, digamos: (x_1, y_1), (x_2, y_2), ..., (x_2, y_2) e que através deles queremos estimar o modelo acima.

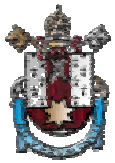


A equação de regressão

A reta estimada será representada por:

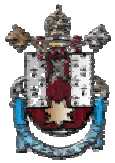
$$\hat{Y} = a + bX \quad ou \quad Y = a + bX + E$$

Onde "a" é um estimador de α e "b" é um estimador de β, sendo $\hat{Y}$ um estimador de $E(Y/x)$.



O método utilizado

Existem diversos métodos para a determinação da reta desejada. Um deles, denominado de ***MMQ*** *(****M****étodos dos* ***M****ínimos Quadrados), consiste em minimizar a "soma dos quadrados das distâncias da reta aos pontos".*

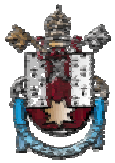


Tem-se:

$$Y_i = a + bx_i + E_i,$$

Então:

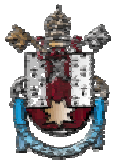
$$E_i = Y_i - (a + bx_i)$$



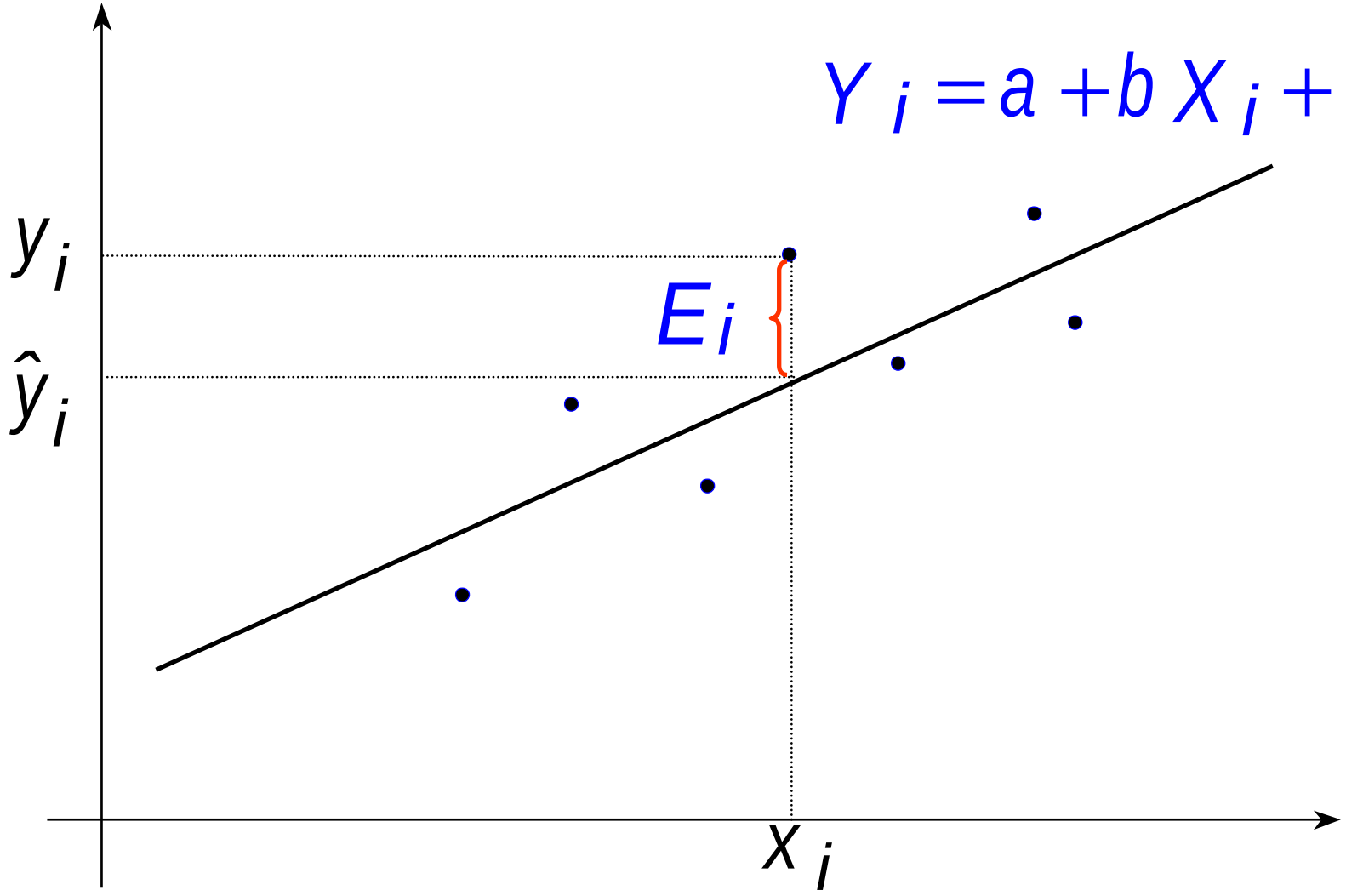
Deve-se minimizar:

$$\phi = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 =$$

$$= \sum_{i=1}^{n} (Y_i - a - b X_i)^2$$



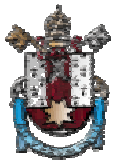
O método dos mínimos quadrados



Derivando parcialmente tem-se:

$$\frac{\partial \phi}{\partial a} = -2 \sum_{i=1}^{n} (Y_i - a - b X_i)$$

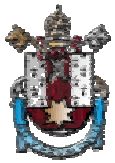
$$\frac{\partial \phi}{\partial b} = -2 \sum_{i=1}^{n} x_i (Y_i - a - b X_i)$$



Igualando as derivadas parciais a zero vem:

$$\sum_{i=1}^{n}(Y_i - a - b X_i) = 0$$

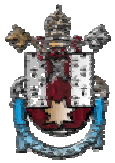
$$\sum_{i=1}^{n} x_i(Y_i - a - b X_i) = 0$$



Isolando as incógnitas, tem-se:

$$\sum Y_i = na + b\sum X_i$$

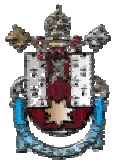
$$\sum X_i Y_i = n\sum X_i + b\sum X_i^2$$



Resolvendo para “a” e “b”, segue:

$$b = \frac{\sum X_i y_i - n\overline{X}\,\overline{Y}}{\sum X_i^2 - n\,\overline{X}^2} = \frac{S_{XY}}{S_{XX}}$$

$$a = \overline{Y} - b\overline{X}$$

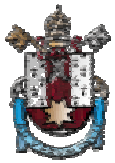


Lembrando que:

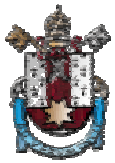
$$S_{XY} = \sum X_i Y_i - n\bar{X}\bar{Y}$$

$$S_{XX} = \sum X_i^2 - n\bar{X}^2$$

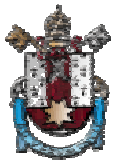
$$S_{YY} = \sum Y_i^2 - n\bar{Y}^2$$



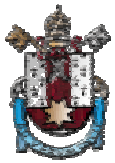
Exemplo



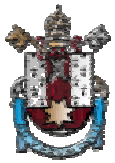
Um engenheiro químico está investigando o efeito da temperatura de operação do processo no rendimento do produto. O estudo resultou nos dados da tabela, ao lado. Determinar a linha de regressão.



Temperatura, C^0 (X)	Rendimento (Y)
100	45
110	51
120	54
130	61
140	66
150	70
160	74
170	78
180	85
190	89



Da mesma forma que para calcular o coeficiente de correlação é necessário a construção de três novas colunas. Uma para X^2, uma para Y^2 e outra para XY.



X	Y	XY	X	Y
100	45	4500	10000	2025
110	51	5610	12100	2601
120	54	6480	14400	2916
130	61	7930	16900	3721
140	66	9240	19600	4356
150	70	10500	22500	4900
160	74	11840	25600	5476
170	78	13260	28900	6084
180	85	15300	32400	7225
190	89	16910	36100	7921
1450	673	101570	218500	47225

Tem-se:

$$n = 10 \quad \sum X = 1450 \quad \sum Y = 673$$

$$\bar{X} = 145 \quad \bar{Y} = 67,3 \quad \sum XY = 101570$$

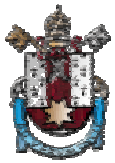
$$\sum X^2 = 218500 \qquad \sum Y^2 = 47225$$

Então:

$$S_{XY} = \sum X_i Y_i - n\bar{X}\bar{Y} =$$

$$= 101570 - 10.145.67,3 =$$

$$= 3985$$



$$S_{XX} = \sum X_i^2 - n\bar{X}^2 =$$

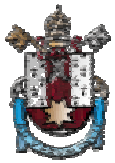
$$= 218500 - 10.145^2 =$$

$$= 8250$$

$$S_{YY} = \sum Y_i^2 - n\bar{Y}^2 =$$

$$= 47225 - 10.67,3^2 =$$

$$= 1932,10$$



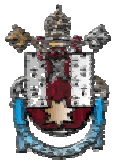
A equação de regressão, será, então:

$$b = \frac{S_{XY}}{S_{XX}} = \frac{3985}{8250} = 0,4830 \cong 0,48$$

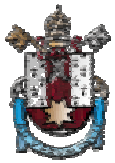
$$a = \bar{Y} - b\bar{X} = 67,30 - 0,4830 .145 =$$

$$= -2,7394 \cong -2,74$$

$$\hat{Y} = -2,74 + 0,48\,x$$



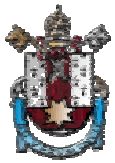
A pergunta que cabe agora é: este modelo representa bem os pontos dados? A resposta é dada através do erro padrão da regressão.



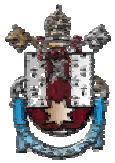
Variância Residual
e
Erro Padrão da Regressão

O objetivo do MMQ é minimizar a variação residual em torno da reta de regressão. Uma avaliação desta variação é dada por:

$$S^2 = \frac{\sum E^2}{n-2} = \frac{\sum (Y - a - bX)^2}{n-2}$$



O cálculo da variância residual, por esta expressão, é muito trabalhoso, pois é necessário primeiro determinar os valores previstos. Entretanto é possível obter uma expressão que não requeira o cálculo dos valores previstos, isto é, de $\hat{Y} = a + bX$



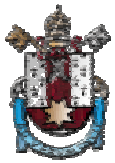
Desenvolvendo o numerador da expressão, vem:

$$\sum(Y-a-bX)^2=\sum[Y-(\bar{Y}-b\bar{X})-bX]^2=$$

$$=\sum[Y-\bar{Y}+b\bar{X}-bX]^2=\sum[Y-\bar{Y}-b(X-\bar{X})]^2=$$

$$=\sum(Y-\bar{Y})^2-2b\sum(X-\bar{X})(Y-\bar{Y})+b^2\sum(X-\bar{X})^2=$$

$$=S_{YY}-2bS_{XY}+b^2S_{XX}$$



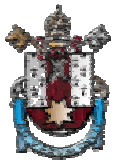
Uma vez que:

$$\sum (X - \bar{X})(Y - \bar{Y}) =$$

$$= \sum X_i Y_i - n\bar{X}\bar{Y} = S_{XY}$$

$$\sum (X - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2 = S_{XX}$$

$$\sum (Y - \bar{Y})^2 = \sum Y_i^2 - n\bar{Y}^2 = S_{YY}$$



Deste modo, tem-se:

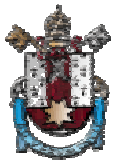
$$\sum(Y - a - bX)^2 = S_{YY} - 2b\,S_{XY} + b^2 S_{XX}$$

Mas: $$b = \frac{S_{XY}}{S_{XX}} \Rightarrow S_{XY} = b\,S_{XX}$$

Então:

$$\sum(Y - a - bX)^2 = S_{YY} - 2b\,S_{XY} + b^2 S_{XX} =$$

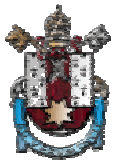
$$= S_{YY} - 2b^2 S_{XX} + b^2 S_{XX} = S_{YY} - b^2 S_{XX}$$



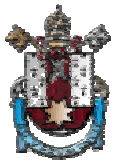
Finalmente:

$$s = \sqrt{\frac{\sum E^2}{n-2}} = \sqrt{\frac{\sum (Y - a - bX)^2}{n-2}} =$$

$$= \sqrt{\frac{S_{YY} - b^2 S_{XX}}{n-2}} = \sqrt{\frac{S_{YY} - b\, S_{XY}}{n-2}}$$



Exemplo



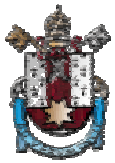
Considerando os valores do exemplo anterior, determinar o erro padrão da regressão.

Tem-se:

$$S_{YY} = 1932,10$$

$$S_{XX} = 8250$$

$$b = \frac{S_{XY}}{S_{XX}} = \frac{3985}{8250} = 0,4830$$

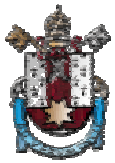


Então:

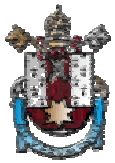
$$s = \sqrt{\frac{S_{YY} - b\,S_{XY}}{n-2}} =$$

$$= \sqrt{\frac{1932{,}10 - \frac{3985}{8250}.3985}{10-2}} =$$

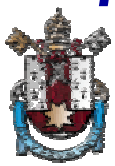
$$= 0{,}9503 \cong 0{,}95$$



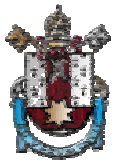
Distribuições das Estimativas



Os valores de "a" e "b" são estimadores de "α" e "β". As propriedades estatísticas destes estimadores são úteis para testar a adequação do modelo. Eles são variáveis aleatórias uma vez que são combinações lineares dos Y_i que são, por sua vez, variáveis aleatórias.



As principais propriedades de interesse são a média (expectância), a variabilidade (erro padrão) e a distribuição de probabilidade de cada um dos estimadores.



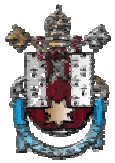
Comportamento de "a"

(i) Expectância

$$E(a) = E\left(\bar{Y} - b\bar{X}\right) = \ldots = \alpha$$

(ii) Variância

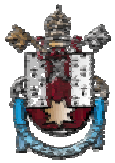
$$V(a) = V\left(\bar{Y} - b\bar{X}\right) = \ldots = \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}}\right)$$



Portanto a distribuição da estatística "a", será:

$$a \sim N\left(\alpha,\ \sigma\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}}\right)$$

Como o valor "σ" não é conhecido e precisa ser estimado por "s", então, de fato, utiliza-se a distribuição t_{n-2}.



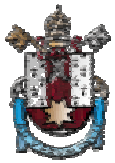
Comportamento de "b"

(i) Expectância

$$E(b) = E\left(\frac{S_{XY}}{S_{XX}}\right) = \ldots = \beta$$

(ii) Variância

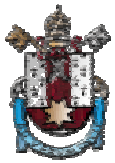
$$V(b) = V\left(\frac{S_{XY}}{S_{XX}}\right) = \ldots = \frac{\sigma^2}{S_{XX}}$$



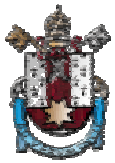
Portanto a distribuição da estatística "b", será:

$$b \sim N\left(\beta, \frac{\sigma}{\sqrt{S_{xx}}}\right)$$

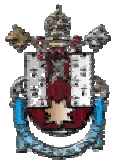
Como o valor "σ" não é conhecido e precisa ser estimado por "s", então, de fato, utiliza-se a distribuição t_{n-2}.



Intervalos de Confiança para os parâmetros da regressão



Da mesma forma que foram obtidos IC para a média, a proporção e a variância de uma população, pode-se determinar intervalos para os parâmetros "α" e "β" da regressão.

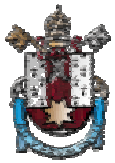


IC para o coeficiente linear " α"

O IC de "$1-\alpha$" de confiança para o coeficiente linear "α" é dado por:

$$P\left(a - t_{n-2} S \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}}} \leq \alpha \leq a + t_{n-2} S \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}}}\right) =$$

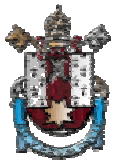
$$= 1 - \alpha$$



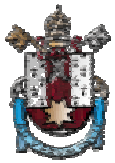
IC para o coeficiente angular " β"

O IC de "1 – α" de confiança para o coeficiente da regressão "β" é dado por:

$$P\left(b - t_{n-2}\frac{S}{\sqrt{S_{XX}}} \leq \beta \leq b + t_{n-2}\frac{S}{\sqrt{S_{XX}}}\right) = 1 - \alpha$$

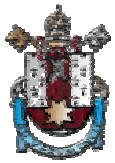


Exemplo



Determinar intervalos de confiança de 95% para os parâmetros da equação de regressão, utilizando os dados do exercício anterior.

$$\hat{Y} = -2,74 + 0,48\,x$$



Dados

$S_{YY} = 1932,10$

$S_{XX} = 8250$

$S_{XY} = 3985$

$\overline{X} = 145$

$\overline{Y} = 67,30$

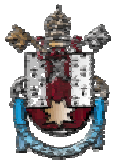
$a = -2,7394$

$b = 0,4830$

$s = 0,9503$

$n = 10$

$1 - \alpha = 95\%$



O IC de "1- α" para o Coef. Linear

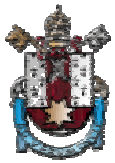
"α" é dado por: $a \pm t_{n-2} S \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}$

Então:

$$-2{,}7394 \pm 2{,}306.0{,}9503 \sqrt{\frac{1}{10} + \frac{145^2}{8250}}$$

$$-2{,}7394 \pm 3{,}5663$$

$$[-6{,}31;\ 0{,}83]$$



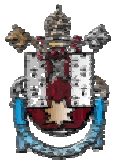
O IC de "1- α" para o Coef. Angular "β" é dado por: $b \pm t_{n-2} \frac{S}{\sqrt{S_{xx}}}$

Então: $0,4830 \pm 2,306 . \frac{0,9503}{\sqrt{8250}}$

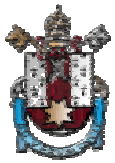
$0,4830 \pm 2,306$

$[0,4589;\ 0,5071]$

$[0,46;\ 0,51]$



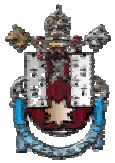
Intervalos de Confiança para o valor médio e para um valor individual de Y



Da mesma forma que foram obtidos IC para os parâmetros da regressão, pode-se obter IC para os valores estimados de Y para um dado x. Vamos considerar dois casos:

(a) Considerando somente a incerteza da linha de regressão;

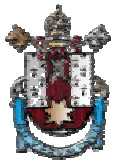
(b) Considerando a incerteza da linha mais a variação da variável Y.



IC para um valor médio de Y, dado x

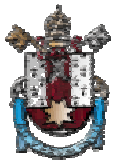
Para construir o IC de "1 – α" para o valor médio de Y, dado x, é necessário conhecer sua distribuição. Tem-se:

$$\hat{Y} \sim N\left(\alpha+\beta x;\ \sigma\sqrt{\frac{1}{n}+\frac{(X-\bar{X})^2}{S_{XX}}}\right)$$



Então IC de "1 – α" de confiança para o um valor médio de Y, dado x ,é:

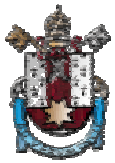
$$\hat{Y} \pm t_{n-2} S \sqrt{\frac{1}{n} + \frac{(X - \bar{X})^2}{S_{XX}}}$$



IC para um valor médio individual de Y, dado x

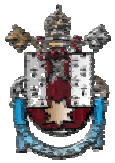
Uma estimativa do valor individual de Y é dado por "a + bx" e a distribuição desta estimativa será dada por:

$$\hat{Y} \sim N(0; \sigma\sqrt{1+\frac{1}{n}+\frac{(X-\bar{X})^2}{S_{XX}}})$$

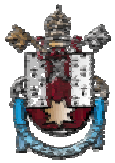


Então IC de "1 – α" de confiança para o um valor individual de Y, dado x , será:

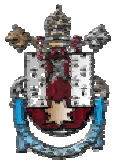
$$\hat{Y} \pm t_{n-2} S \sqrt{1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{S_{XX}}}$$



Exemplo



Determinar intervalos de confiança de 95% para os valores médio e individual de Y, na hipótese de x = 200.



Dados

$S_{YY} = 1932,10$

$S_{XX} = 8250$

$S_{XY} = 3985$

$\overline{X} = 145$

$\overline{Y} = 67,30$

$x = 200$

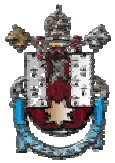
$a = -2,7394$

$b = 0,4830$

$s = 0,9503$

$n = 10$

$1 - \alpha = 95\%$



O IC de "1- α" para o valor médio de Y, dado "x" é:

$$\hat{Y} \pm t_{n-2} S \sqrt{\frac{1}{n} + \frac{(X - \bar{X})^2}{S_{XX}}}$$

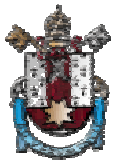
Então:

$$\hat{y} = -2,7394 + 0,4830 .200 = 93,8606$$

$$93,8606 \pm 2,306.0,9503 \sqrt{\frac{1}{10} + \frac{(200 - 145)^2}{8250}}$$

$$93,8606 \pm 1,4970$$

[92,36; 95,36]



O IC de "1- α" para o valor individual de Y , dado "x" é:

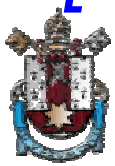
Então:

$$\hat{Y} \pm t_{n-2} S \sqrt{1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{S_{XX}}}$$

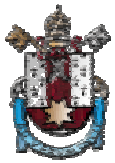
$$93{,}8606 \pm 2{,}306 . 0{,}9503 \sqrt{1 + \frac{1}{10} + \frac{(200 - 145)^2}{8250}}$$

$$93{,}8606 \pm 2{,}6539$$

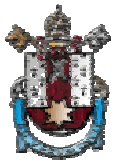
$$[91{,}21;\ 96{,}51]$$



Testes de Hipóteses para os parâmetros da regressão



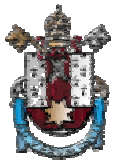
Da mesma forma que foram testados todos os parâmetros até então pode-se testar os parâmetros "α" e "β" da regressão.



Teste para o coeficiente linear " α"

A variável teste para testar o coeficiente linear é dado por:

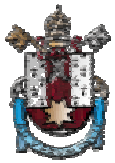
$$t_{n-2} = \frac{a - \alpha}{S\sqrt{\frac{1}{n} + \frac{\overline{X}^2}{S_{XX}}}}$$



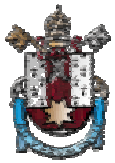
Teste para o coeficiente angular " β"

A variável teste para testar o coeficiente da regressão "β" é dada por:

$$t_{n-2} = \frac{b - \beta}{\frac{S}{\sqrt{S_{xx}}}}$$

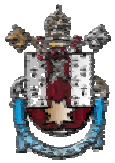


Exemplo



(a) Testar, a 1% de significância, se é possível afirmar que a linha de regressão, do exemplo dado, não passa pela origem.

(b) Testar se é possível, a 1% de significância, afirmar que existe regressão positiva entre as duas variáveis.



Dados

$$a = -2,7394$$

$$b = 0,4830$$

$$s = 0,9503$$

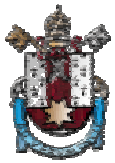
$$n = 10$$

$$1 - \alpha = 1\%$$

$$S_{YY} = 1932,10$$

$$S_{XX} = 8250$$

$$S_{XY} = 3985$$



Solução:

Hipóteses:

H_0: $\alpha = 0$

H_1: $\alpha \neq 0$

(A)

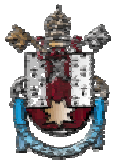
Dados:

$n = 10$

$a = -2,739$

$\alpha = 1\%$

Trata-se de um teste bilateral para o coeficiente linear da regressão.

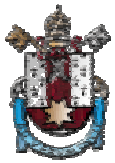


A variável teste é:

$$t_{n-2} = \frac{a - \alpha}{s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}}}$$

Então:

$$t_8 = \frac{-2{,}739 - 0}{0{,}9503\sqrt{\frac{1}{10} + \frac{145^2}{8250}}} = -1{,}771$$

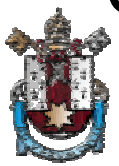


O valor crítico t_c *é tal que:* $P(|T| > t_c) = \alpha$

Então $t_c = -3{,}355$. *Assim RC =* $[-3{,}355; \infty)$

DECISÃO e CONCLUSÃO:

Como $t_8 = -1{,}771 \in RC$ *ou* $-1{,}771 > -3{,}355$. *Aceito* H_0, *isto é, a 1% de significância, não se pode afirmar que a linha de regressão não passe pela origem.*



Solução:

Hipóteses:

$H_0: \beta = 0$

$H_1: \beta > 0$

(B)

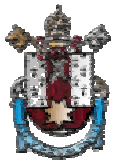
Dados:

$n = 10$

$b = 0,4830$

$\alpha = 1\%$

Trata-se de um teste unilateral para o coeficiente angular da regressão.

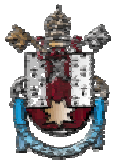


A variável teste é:

$$t_{n-2} = \frac{b - \beta}{\frac{S}{\sqrt{S_{xx}}}}$$

Então:

$$t_8 = \frac{0,4830 - 0}{0,9503 / \sqrt{8250}} = 46,165$$

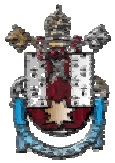


O valor crítico t_c *é tal que:* $P(T > t_c) = \alpha$

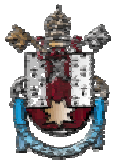
Então $t_c = 2{,}896$. *Assim RC =* $[2{,}896; \infty)$

DECISÃO e CONCLUSÃO:

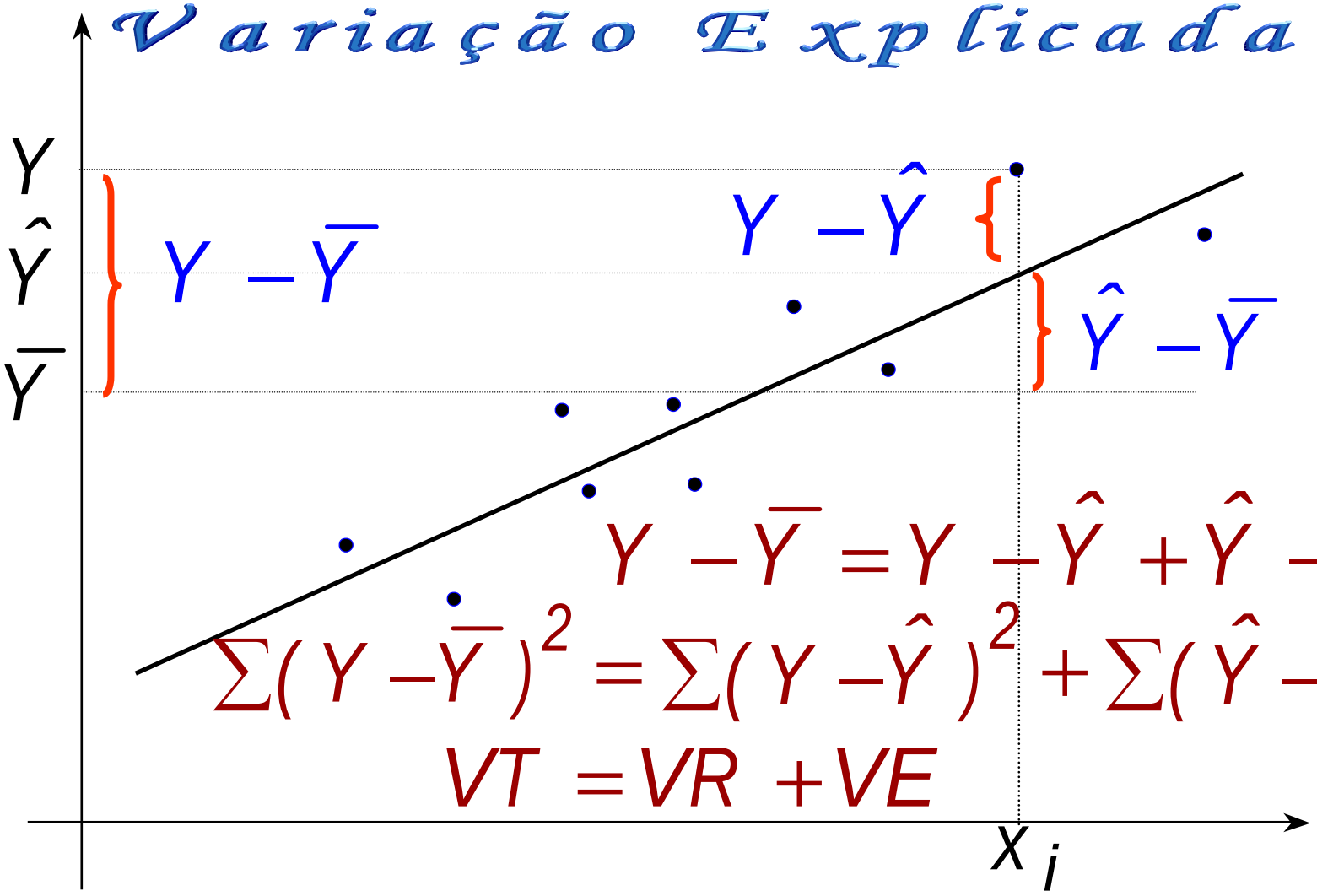
Como $t_8 = 46{,}165 \in RC$ *ou* $46{,}165 > 2{,}896$. *Rejeito* H_0, *isto é, a 1% de significância, pode-se afirmar que existe regressão entre as duas variáveis.*



Decomposição da Variação



Variação Total = Variação Não-Explicada + Variação Explicada



$$Y - \bar{Y} = Y - \hat{Y} + \hat{Y} - \bar{Y}$$

$$\Sigma(Y - \bar{Y})^2 = \Sigma(Y - \hat{Y})^2 + \Sigma(\hat{Y} - \bar{Y})^2$$

$$VT = VR + VE$$

(a) Variação Total: VT

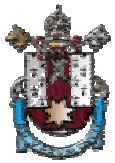
$$VT = \sum \left(Y - \bar{Y}\right)^2 = S_{YY}$$

(b) Variação Residual: VR

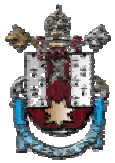
$$VR = \sum \left(Y - \hat{Y}\right)^2 = S_{YY} - b^2 S_{XX} = VT - VE$$

(c) Variação Explicada: VE

$$VE = \sum \left(\hat{Y} - \bar{Y}\right)^2 = b^2 S_{XX}$$



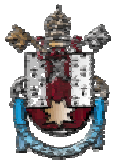
Uma maneira de medir o grau de aderência (adequação) de um modelo é verificar o quanto da variação total de Y é explicada pela reta de regressão.



Para isto, toma-se o quociente entre a variação explicada, VE, pela variação total ,VT:

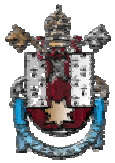
$$R^2 = VE / VT$$

Este resultado é denominado de "Coeficiente de Determinação".



$$R^2 = \frac{VE}{VT} = \frac{b^2 S_{XX}}{S_{YY}} = \frac{b S_{XY}}{S_{YY}} = \frac{S_{XY}^2}{S_{XX} S_{YY}}$$

Este resultado mede o quanto as variações de uma das variáveis são explicadas pelas variações da outra variável.



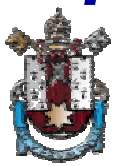
Ou ainda, ele mede a parcela da variação total que é explicada pela reta de regressão, isto é:

$$VE = b^2 S_{XX} = R^2 S_{YY}$$

A variação residual corresponde a:

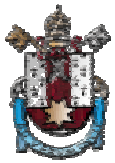
$$VR = (1 - R^2) S_{YY}$$

Assim $1 - R^2$ é o Coeficiente de Indeterminação.

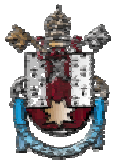


Exercício

O % de impurezas no gás oxigênio produzido por um processo de destilação supõem-se que esteja relacionado com o % de hidrocarbono no condensador principal do processador. Os dados de um mês de operação produziram a seguinte tabela



X	Y	X	Y
1,02	86,91	1,46	96,73
1,11	89,85	1,55	99,42
1,43	90,28	1,55	98,66
1,11	86,34	1,55	96,07
1,01	92,58	1,40	93,65
0,95	87,33	1,15	87,31
1,11	86,29	1,01	95,00
0,87	91,86	0,99	96,85
1,43	95,61	0,95	85,20
1,02	89,86	0,98	90,56

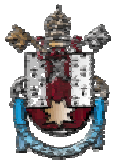


(a) Ajuste um modelo linear aos dados;

(b) Teste a existência da regressão;

(c) Determine o valor de R^2 para este modelo;

(d) Determine um IC, de 95%, para o valor da pureza, na hipótese do % de hidrocarbono ser 1,20% .



Até a próxima ...