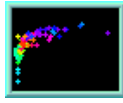


Regressão



Prof. Lorí Viali, Dr.

viali@mat.ufrgs.br

<http://www.mat.ufrgs.br/~viali/>

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



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



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.*



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



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.



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



O modelo

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



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



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**"



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



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**".



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



O modelo considerado

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

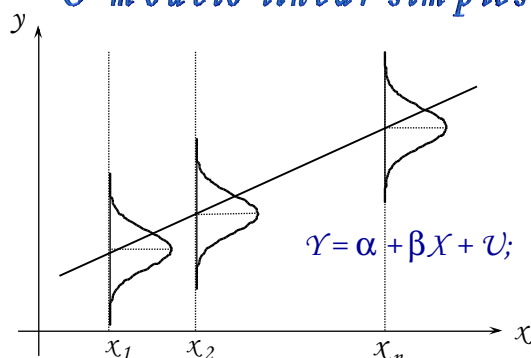
$$Y = \alpha + \beta X + U$$



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



O modelo linear simples



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



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 .



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



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:

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

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

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

$$\text{Cov}(U_i, U_j) = 0, \text{ para } i \neq j;$$

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: $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ e que através deles queremos estimar o modelo acima.

A equação de regressão

A reta estimada será representada por:

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

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** (Métodos dos Mínimos Quadrados), consiste em minimizar a "soma dos quadrados das distâncias da reta aos pontos".*

Tem-se:

$$Y_i = a + b x_i + E_i$$

Então:

$$E_i = Y_i - (a + b x_i)$$

Deve-se minimizar:

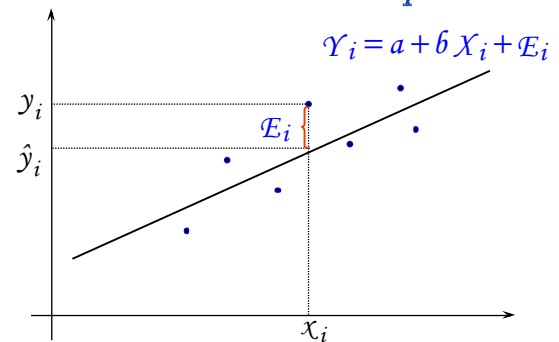
$$\begin{aligned}\phi &= \sum_{i=1}^n \mathcal{E}_i^2 = \sum_{i=1}^n (\mathcal{Y}_i - \hat{\mathcal{Y}}_i)^2 = \\ &= \sum_{i=1}^n (\mathcal{Y}_i - a - bX_i)^2\end{aligned}$$



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



O método dos mínimos quadrados



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Derivando parcialmente tem-se:

$$\begin{aligned}\frac{\partial \phi}{\partial a} &= -2 \sum_{i=1}^n (\mathcal{Y}_i - a - bX_i) \\ \frac{\partial \phi}{\partial b} &= -2 \sum_{i=1}^n x_i (\mathcal{Y}_i - a - bX_i)\end{aligned}$$



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Igualando as derivadas parciais a zero vem:

$$\begin{aligned}\sum_{i=1}^n (\mathcal{Y}_i - a - bX_i) &= 0 \\ \sum_{i=1}^n x_i (\mathcal{Y}_i - a - bX_i) &= 0\end{aligned}$$



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Isolando as incógnitas, tem-se:

$$\begin{aligned}\sum \mathcal{Y}_i &= na + b \sum X_i \\ \sum X_i \mathcal{Y}_i &= n \sum X_i + b \sum X_i^2\end{aligned}$$



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Resolvendo para “a” e “b”, segue:

$$\begin{aligned}b &= \frac{\sum X_i \mathcal{Y}_i - n \bar{X} \bar{\mathcal{Y}}}{\sum X_i^2 - n \bar{X}^2} = \frac{S_{X\mathcal{Y}}}{S_{XX}} \\ a &= \bar{\mathcal{Y}} - b \bar{X}\end{aligned}$$



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



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° (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 ²	Y ²
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:

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

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

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

Então:

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

$$= 101570 - 10 \cdot 145 \cdot 67,3 =$$

$$= 3985$$

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

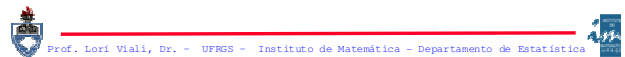
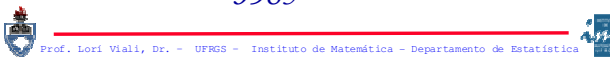
$$= 218500 - 10 \cdot 145^2 =$$

$$= 8250$$

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

$$= 47225 - 10 \cdot 67,3^2 =$$

$$= 1932,10$$



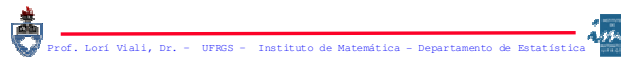
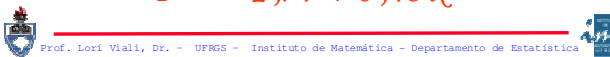
A equação de regressão, será, então:

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

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

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

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

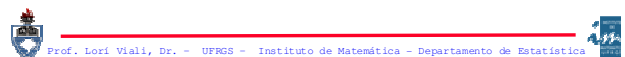
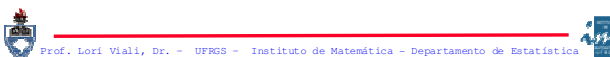


A pergunta que cabe agora é: este modelo representa bem os pontos dados? A resposta é dada através do erro padrão da regressão.

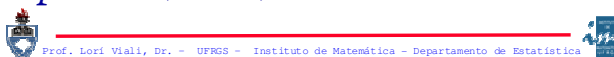
Variância Residual e Erro Padrão da Regressão

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

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



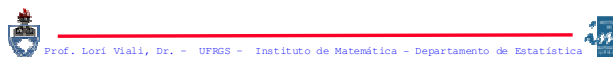
O cálculo da variância residual, por esta expressão, é muito trabalhoso, pois é necessário primeiro determinar os valores previstos. Entretanto é possível obter uma expressão que não requeira o cálculo dos valores previstos, isto é, de $\hat{Y} = a + bX$



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

Desenvolvendo o numerador da expressão, vem:

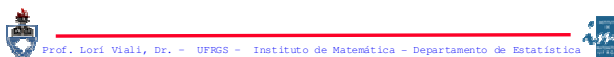
$$\begin{aligned}\sum (Y - a - bX)^2 &= \sum [Y - (\bar{Y} - b\bar{X}) - b(X - \bar{X})]^2 = \\ &= \sum [Y - \bar{Y} + b\bar{X} - b(X - \bar{X})]^2 = \sum [Y - \bar{Y} - b(X - \bar{X})]^2 = \\ &= \sum (Y - \bar{Y})^2 - 2b \sum (X - \bar{X})(Y - \bar{Y}) + b^2 \sum (X - \bar{X})^2 = \\ &= S_{YY} - 2b S_{XY} + b^2 S_{XX}\end{aligned}$$



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

Uma vez que:

$$\begin{aligned}\sum (X - \bar{X})(Y - \bar{Y}) &= \\ &= \sum X_i Y_i - n \bar{X} \bar{Y} = S_{XY} \\ \sum (X - \bar{X})^2 &= \sum X_i^2 - n \bar{X}^2 = S_{XX} \\ \sum (Y - \bar{Y})^2 &= \sum Y_i^2 - n \bar{Y}^2 = S_{YY}\end{aligned}$$



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

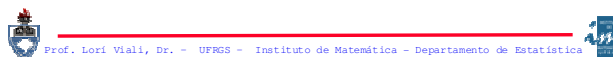
Deste modo, tem-se:

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

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

Então:

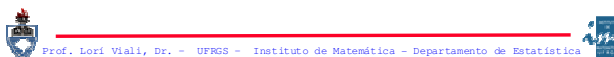
$$\begin{aligned}\sum (Y - a - bX)^2 &= S_{YY} - 2b S_{XY} + b^2 S_{XX} = \\ &= S_{YY} - 2b^2 S_{XX} + b^2 S_{XX} = S_{YY} - b^2 S_{XX}\end{aligned}$$



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

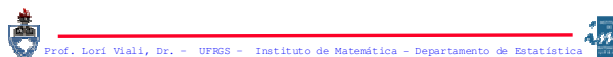
Finalmente:

$$\begin{aligned}s &= \sqrt{\frac{\sum E^2}{n-2}} = \sqrt{\frac{\sum (Y - a - bX)^2}{n-2}} = \\ &= \sqrt{\frac{S_{YY} - b^2 S_{XX}}{n-2}} = \sqrt{\frac{S_{YY} - b S_{XY}}{n-2}}\end{aligned}$$



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

E x e m p l o



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

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

Tem-se:

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

$$S_{XX} = 8250$$

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



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Então:

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

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

$$= 0,9503 \approx 0,95$$



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Distribuições das Estimativas



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



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.



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



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



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Comportamento de “a”

(i) Expectância

$$E(a) = E(\bar{Y} - b\bar{X}) = \dots = \alpha$$

(ii) Variância

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



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Portanto a distribuição da estatística "a", será:

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

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



Prof. Lori Viali, Dr. - UFERSA - Instituto de Matemática - Departamento de Estatística



Comportamento de "b"

(i) Expectância

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

(ii) Variância

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



Prof. Lori Viali, Dr. - UFERSA - Instituto de Matemática - Departamento de Estatística



Portanto a distribuição da estatística "b", será:

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

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



Prof. Lori Viali, Dr. - UFERSA - Instituto de Matemática - Departamento de Estatística



Intervalos de Confiança para os parâmetros da regressão



Prof. Lori Viali, Dr. - UFERSA - Instituto de Matemática - Departamento de Estatística



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.



Prof. Lori Viali, Dr. - UFERSA - Instituto de Matemática - Departamento de Estatística



IC para o coeficiente linear " α "

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

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



Prof. Lori Viali, Dr. - UFERSA - Instituto de Matemática - Departamento de Estatística



IC para o coeficiente angular " β "

O IC de " $1 - \alpha$ " de confiança para o coeficiente da regressão " β " é dado por:

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



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Exemplo



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



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

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



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Dados

$$S_{YY} = 1932,10 \quad a = -2,7394$$

$$S_{XX} = 8250 \quad b = 0,4830$$

$$S_{XY} = 3985 \quad s = 0,9503$$

$$\bar{X} = 145 \quad n = 10$$

$$\bar{Y} = 67,30 \quad 1 - \alpha = 95\%$$



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



O IC de " $1 - \alpha$ " 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:

$$-2,7394 \pm 2,306 \cdot 0,9503 \sqrt{\frac{1}{10} + \frac{145^2}{8250}}$$

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

$$[-6,31; 0,83]$$



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



O IC de " $1 - \alpha$ " para o Coef. Angular " β " é dado por:

$$b \pm t_{n-2} \frac{S}{\sqrt{S_{XX}}}$$

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

$$0,4830 \pm 2,306$$

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

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



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



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



Prof. Lori Viali, Dr. - UFGRS - Instituto de Matemática - Departamento de Estatística



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 .



Prof. Lori Viali, Dr. - UFGRS - Instituto de Matemática - Departamento de Estatística



IC para um valor médio de Y , dado x

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

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



Prof. Lori Viali, Dr. - UFGRS - Instituto de Matemática - Departamento de Estatística



Então IC de " $1 - \alpha$ " de confiança para o um valor médio de Y , dado x , é:

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



Prof. Lori Viali, Dr. - UFGRS - Instituto de Matemática - Departamento de Estatística



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:

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



Prof. Lori Viali, Dr. - UFGRS - Instituto de Matemática - Departamento de Estatística



Então IC de " $1 - \alpha$ " de confiança para o um valor individual de Y , dado x , será:

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



Prof. Lori Viali, Dr. - UFGRS - Instituto de Matemática - Departamento de Estatística



Exemplo

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

Dados

$$\begin{aligned} S_{YY} &= 1932,10 & a &= -2,7394 \\ S_{XX} &= 8250 & b &= 0,4830 \\ S_{XY} &= 3985 & s &= 0,9503 \\ \bar{X} &= 145 & n &= 10 \\ \bar{Y} &= 67,30 & 1-\alpha &= 95\% \\ x &= 200 \end{aligned}$$

O IC de " $1-\alpha$ " 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:

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

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

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

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

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

Então:

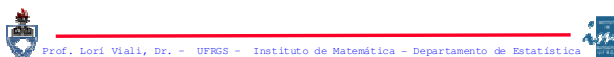
$$93,8606 \pm 2,306 \cdot 0,9503 \sqrt{1 + \frac{1}{10} + \frac{(200-145)^2}{8250}}$$

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

$$[91,21; 96,51]$$

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:

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



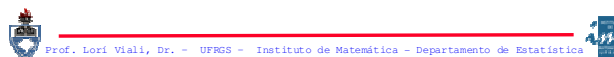
Teste para o coeficiente angular " β "

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

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

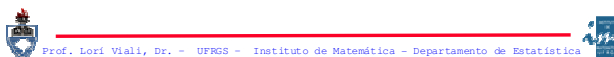


Exemplo



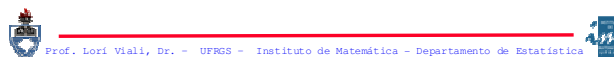
(a) Testar, a 1% de significância, se é possível afirmar que a linha de regressão, do exemplo dado, não passa pela origem.

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



Dados

$$\begin{aligned} a &= -2,7394 & S_{YY} &= 1932,10 \\ b &= 0,4830 & S_{XX} &= 8250 \\ s &= 0,9503 & S_{XY} &= 3985 \\ n &= 10 \\ 1 - \alpha &= 1\% \end{aligned}$$



Solução:

Hipóteses:

$$H_0: \alpha = 0 \quad (\mathcal{A})$$

$$H_1: \alpha \neq 0$$

Dados:

$$n = 10$$

$$a = -2,739$$

$$\alpha = 1\%$$

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



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



A variável teste é:

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

Então:

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



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



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

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

DECISÃO e CONCLUSÃO:

Como $t_8 = -1,771 \in \mathcal{RC}$ 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.



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Solução:

Hipóteses:

$$H_0: \beta = 0 \quad (\mathcal{B})$$

$$H_1: \beta > 0$$

Dados:

$$n = 10$$

$$b = 0,4830$$

$$\alpha = 1\%$$

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



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



A variável teste é:

$$t_{n-2} = \frac{b - \beta}{S \sqrt{\frac{1}{S_{XX}}}}$$

Então:

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



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



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

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

DECISÃO e CONCLUSÃO:

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



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



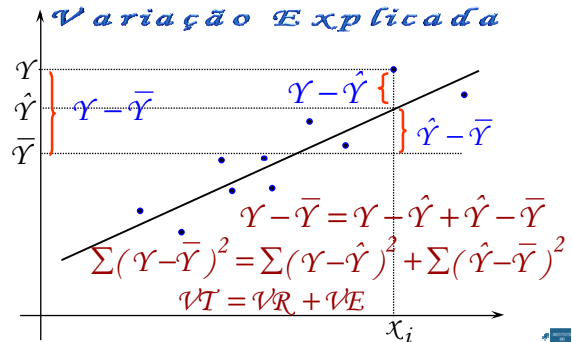
Decomposição da Variação



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



$$\text{Variação Total} = \text{Variação Não-Explicada} + \text{Variação Explicada}$$



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



(a) Variação Total: $\mathcal{VT}$

$$\mathcal{VT} = \sum (\gamma - \bar{\gamma})^2 = S_{\gamma\gamma}$$

(b) Variação Residual: $\mathcal{VR}$

$$\mathcal{VR} = \sum (\gamma - \hat{\gamma})^2 = S_{\gamma\gamma} - b^2 S_{XX} = \mathcal{VT} - \mathcal{VE}$$

(c) Variação Explicada: $\mathcal{VE}$

$$\mathcal{VE} = \sum (\hat{\gamma} - \bar{\gamma})^2 = b^2 S_{XX}$$



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



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



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Para isto, toma-se o quociente entre a variação explicada, $\mathcal{VE}$, pela variação total, $\mathcal{VT}$:

$$\mathcal{R}^2 = \mathcal{VE} / \mathcal{VT}$$

Este resultado é denominado de “Coeficiente de Determinação”.



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



$$\mathcal{R}^2 = \frac{\mathcal{VE}}{\mathcal{VT}} = \frac{b^2 S_{XX}}{S_{\gamma\gamma}} = \frac{b S_{X\gamma}}{S_{\gamma\gamma}} = \frac{S_{X\gamma}^2}{S_{XX} S_{\gamma\gamma}}$$

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



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



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.



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



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



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



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



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



- Ajuste um modelo linear aos dados;
- Teste a existência da regressão;
- Determine o valor de R^2 para este modelo;
- Determine um IC, de 95%, para o valor da pureza, na hipótese do % de hidrocarbono ser 1,20% .



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

