

Capítulo 15- Análise de regressão linear

15.1 Introdução

A análise de regressão linear é feita pelo método dos mínimos quadrados

Dado uma série de n números, podemos estabelecer uma correlação entre eles através de uma equação de uma reta conforme Figura (15.1).

$$Y = aX + b + \varepsilon$$

Sendo:

Y= variável dependente

X= variável independente

b= intercepto da linha, isto é, o ponto na qual a linha reta corta o eixo y

a= declividade da linha reta

ε = erro randômico

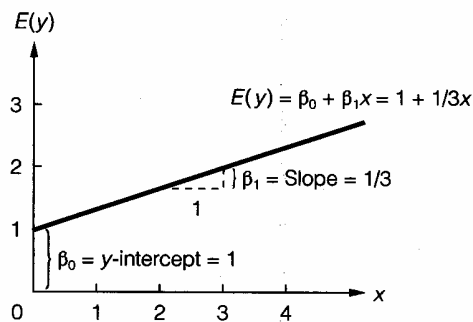


Figura 15.1- Equação de uma linha reta
Fonte: Sincich, 1993

Que é obtida da seguinte maneira:

$$a = (n \cdot \sum XY - \sum X \cdot \sum Y) / (n \sum X^2 - \sum X \cdot \sum X)$$

$$b = (\sum Y - a \sum X) / n$$

n= número de dados

Exemplo 15.1

Vamos mostrar um problema de Sincich, 1993. Trata-se de crianças de várias idades onde foi feito eletroencefalograma EEG em crianças normais conforme Tabela (15.1). Verificar se há correlação do crescimento do pico do EEG com a idade.

Queremos achar uma equação de uma reta da forma $Y = aX + b$

Tabela 15.1-Dados

X Idade	Y Pico de frequência	X ²	X . Y
2	2	4	4
3	5	9	15
4	7	16	28
5	10	25	50
6	11	36	66
ΣX=20	ΣY=35	ΣX²=90	ΣXY=163

$$a = (n \cdot \Sigma XY - \Sigma X \cdot \Sigma Y) / (n \Sigma X^2 - \Sigma X \cdot \Sigma X)$$

$$n=5$$

$$a = (5 \cdot 163 - 20 \cdot 35) / (5 \cdot 90 - 20 \cdot 20) = 2,3$$

$$b = (\Sigma Y - a \Sigma X) / n$$

$$b = (35 - 2,3 \cdot 20) / 5 = -2,2$$

A equação da reta obtida é: $Y = 2,3 X - 2,2$

15.2 Estimativa do desvio padrão s

Vamos ver como se faz a estimativa do desvio padrão s.

Tabela 15.2- Cálculo de s

Y real	y calculado	(Y-y) ²
20	2,4	0,16
15	4,7	0,09
25	7	7,89E-31
22	9,3	0,49
30	11,6	0,36
		Σ 1,10
SSE=Σ(Y-y)² =1,10		

$$SSE = 1,10$$

Isto significa que qualquer outra reta escolhida o valor dos mínimos quadrados será maior que 1,10

$$s^2 = \text{SSE} / (n-2)$$

$$s^2 = 1,10 / (5-2) = 1,10/3 = 0,3666$$

$$s = 0,6055$$

O valor de s é o desvio padrão estimado

O valor $2s = 2 \times 0,6055 = 1,211$ mostra a faixa de variação dos valores, isto é, que os valores estão normalmente dentro desta faixa.

15.3 Coeficiente de determinação R^2

O coeficiente de determinação $0 \leq R^2 \leq 1$, sendo que o ideal é quando $R^2 = 1$, isto para uma curva perfeita, mas como estamos usando dados reais isto nunca vai ser alcançado, pois sempre haverá variáveis que não foram consideradas e trariam um acréscimo no consumo. Portanto, levando-se em conta as variáveis consideradas, haverá sempre um resíduo.

Tabela 15.3-Dados para o coeficiente de determinação do Exemplo (15.1)

y_{calc}	(Y-y)²	y²
2,4	0,16	4
4,7	0,09	25
7	7,89E-31	49
9,3	0,49	100
11,6	0,36	121
	Σ 1,10	Σ 299

$$R^2 = 1 - \text{SSE} / \text{SS}_{yy}$$

Sendo:

$R^2 =$ coeficiente de terminação

$$\text{SSE} = \sum (Y - y_c)^2 = 1,10$$

$$\text{SS}_{yy} = \sum y^2 - (\sum y)^2 / n = 299 - 35^2 / 5 = 54$$

$$R^2 = 1 - \text{SSE} / \text{SS}_{yy}$$

$$R^2 = 1 - 1,10 / 54 = 1 - 0,02037 = 0,9796$$

Foi obtido um coeficiente de determinação R^2 igual a **0,9796**, significando que 97,96% da soma total dos quadrados dos desvios das 5 amostras podem ser explicados usando-se o modelo adotado.

Exemplo 15.2

Queremos uma equação do custo por metro de rede em reais C na forma: $C = a D^b$, sendo D= diâmetro do tubo em mm. Os dados estão na Tabela (15.4).

Os valores a e b deverão ser obtidos por análise de regressão linear.

Tabela 15.4- Custo por metro de material e mão de obra conforme o diâmetro do tubo de ferro fundido base dezembro de 2007

X	Y
Diâmetro (mm)	Custo R\$/metro
80	114,00
100	124,57
150	148,19
200	180,77
250	250,80
300	290,12
400	351,69
500	471,81
600	638,43
700	764,42

Como $C = a D^b$ não é um equação linear podemos artificialmente linearizá-la aplicando logaritmo para os dois lados da equação:

$$C = a D^b$$

$$\text{Log } C = \text{Log } (a D^b) = \text{log } (a) + b \times \text{log } (D)$$

$$\text{Log } C = b \times \text{log } (D) + \text{log } (a)$$

Que é uma equação linear da forma:

$$Y = \text{log } C$$

$$X = \text{log } (D)$$

$$b = \text{log } (a)$$

Tabela 15.5- Cálculos

X	Y	X	Y		
diâmetro	Custo R\$	log (D)	Log(custo)	X ²	X . Y
80	114,00	1,90	2,06	3,62	3,91
100	124,57	2,00	2,10	4,00	4,19
150	148,19	2,18	2,17	4,74	4,72
200	180,77	2,30	2,26	5,29	5,19
250	250,80	2,40	2,40	5,75	5,75
300	290,12	2,48	2,46	6,14	6,10
400	351,69	2,60	2,55	6,77	6,63
500	471,81	2,70	2,67	7,28	7,22
600	638,43	2,78	2,81	7,72	7,79
700	764,42	2,85	2,88	8,09	8,20
		24,18	24,35	59,41	59,71
		ΣX=24,18	ΣY=24,35	ΣX ² =59,41	ΣXY=59,71

$$a = (n \cdot \Sigma XY - \Sigma X \cdot \Sigma Y) / (n \Sigma X^2 - \Sigma X \cdot \Sigma X)$$

$$n=10$$

$$a = (10 \cdot 59,71 - 24,18 \cdot 24,35) / (10 \cdot 59,41 - 24,18 \cdot 24,18) = 0,89$$

$$b = (\Sigma Y - a \Sigma X) / n$$

$$b = (24,35 - 0,89 \cdot 24,18) / 10 = 0,29$$

$$Y = 0,89 X + 0,29$$

$$\text{Mas Log } C = b \times \log(D) + \log(a)$$

$$\text{Portanto } \log(a) = 0,29 \quad \text{e} \quad b = 0,89$$

O valor de a é obtido com o conceito de logaritmo:

$$10^{0,29} = 1,93 = a$$

$$C = a D^b$$

$$C = a D^{0,89}$$

15.4 Análise de Regressão Múltipla

Temos uma regressão linear múltipla quando admitimos que o valor da variável dependente é função linear de duas ou mais variáveis independentes. O modelo estatístico de uma regressão linear múltipla é:

$$Y_{\text{calc}} = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + \beta_k \cdot X_k + \varepsilon$$

Onde Y é a variável dependente e X_1, X_2, \dots, X_k são as variáveis independentes, ε é a componente incontroleável e randômica do erro do modelo adotado e $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ são os coeficientes.

Deve-se lembrar que os símbolos X_1, X_2, \dots, X_k podem representar termos de alta ordem. Por exemplo, X_2 pode representar X_1^2 , X_3 pode representar $X_1 \cdot X_2$ e assim por diante. Por anamorfose os vários modelos estatísticos podem se transformar em lineares, inclusive com o uso de logaritmos.

Segundo Sincich, 1993, a Análise de Regressão Múltipla deve seguir as seguintes *etapas*:

- 1) *Hipótese de que o modelo é linear;*
- 2) *Assumir que os erros são randômicos;*
- 3) *Estimar os coeficientes, $\beta_0, \beta_1, \beta_2, \dots, \beta_k$;*
- 4) *Verificar se o modelo é útil em prever Y;*
- 5) *Verificar se a etapa 2 está satisfeita;*
- 6) *Se decidimos que o modelo é útil e que segue as hipóteses admitidas use o modelo para estimar o valor Y.*

15.5 Softwares de estatística para cálculo da análise de regressão linear múltipla

Não serão mostradas todas as fórmulas usadas na Análise de Regressão Múltipla, pois para isso existem numerosos softwares e livros de estatística.

Um deles é Statistical Package for Business, Economics, and the Social Sciences. É o *ASP* fornecido pela firma americana DMC Software, Inc. Um outro é o *Minitab* versão 6.1 feito por David D. Krueger e Ruth K. Meyer.

Outro software é o SPSS versão 8 para estudantes para no máximo 50 variáveis e 1500 linhas de dados.

Usamos a *Planilha Excel da Microsoft* com o livro do Lapponi, 1995, bem como cálculo de matrizes no Excel.

Usando o Excel da Microsoft foi usada função estatística **PROJ.LIN**.

15.6 Custos dos tubos

Na otimização de Lagrange feita por Thomas M. Waslki é necessário possuímos o custo do material em função do diâmetro em milímetros e portanto vamos apresentar preços atuais de dezembro de 2007 para tubos de PVC e ferro fundido.

Os tubos de ferro fundido variam de 80mm a 700mm que são usuais em aplicações de redes de distribuição conforme Tabela (15.6).

Tabela 15.6- Preço médio por metro linear de tubo de ferro incluso um registro e uma curva em cada 1000 metros de rede e mais mão de obra de 40% sobre o custo dos materiais para dezembro de 2007

Diâmetro (mm)	Custo para 1000m + 1registro e +1curva	ferro fundido	Mão obra 40%	R\$
80mm	R\$ 81.427,30	R\$ 81,43	R\$ 32,57	114,00
100mm	R\$ 88.976,20	R\$ 88,98	R\$ 35,59	124,57
150mm	R\$ 105.849,00	R\$ 105,85	R\$ 42,34	148,19
200mm	R\$ 129.118,00	R\$ 129,12	R\$ 51,65	180,77
250mm	R\$ 179.140,00	R\$ 179,14	R\$ 71,66	250,80
300mm	R\$ 207.227,00	R\$ 207,23	R\$ 82,89	290,12
400mm	R\$ 251.205,00	R\$ 251,21	R\$ 100,48	351,69
500mm	R\$ 337.006,00	R\$ 337,01	R\$ 134,80	471,81
600mm	R\$ 456.021,00	R\$ 456,02	R\$ 182,41	638,43
700mm	R\$ 546.016,00	R\$ 546,02	R\$ 218,41	764,42

Os tubos de plásticos variam de 50mm a 400mm conforme Tabela (15.7)

Tabela 15.7- Preço médio por metro linear de tubo de PVC um registro e uma curva em cada 1000 metros de rede e mais mão de obra de 40% sobre o custo dos materiais para dezembro de 2007.

Diâmetros	Custo	Custo/metro		
		PVC 6,3	MO (40%)	Mat+MO
50mm	R\$ 4.758,43	R\$ 4,76	R\$ 1,90	R\$ 6,66
75mm	R\$ 10.059,02	R\$ 10,06	R\$ 4,02	R\$ 14,08
100mm	R\$ 16.736,12	R\$ 16,74	R\$ 6,69	R\$ 23,43
			R\$ -	R\$ -
		PVC 12		
150mm	R\$ 24.749,00	R\$ 24,75	R\$ 9,90	R\$ 34,65
200mm	R\$ 42.218,00	R\$ 42,22	R\$ 16,89	R\$ 59,11
250mm	R\$ 62.440,00	R\$ 62,44	R\$ 24,98	R\$ 87,42
300mm	R\$ 91.827,00	R\$ 91,83	R\$ 36,73	R\$ 128,56
400mm	R\$ 197.205,00	R\$ 197,21	R\$ 78,88	R\$ 276,09

Na Tabela (15.8) estão as especificações dos materiais e o custo das peças.

Tabela 15.8- Especificação e custos dos materiais dezembro 2007

Válvula de gaveta , série métrica oval, corpo, tampa e cunha em ferro fundido dúctil, NBR 6916 classe 42012, anéis da cunha e corpo em bronze fundido ASTM B62, haste não ascendente com rosca trapezoidal em aço inox ASTM A 276 Gr.410, porca da haste em latão fundido, junta copro/tampa em borracha natural ASTM D 2000, gaxeta em amianto grafitado, extremidades com bolsas para junta elástica JE fornecido com anéis de borracha,, pressão de trabalho de 1,6 MPa, acionamento com cabeçote. Padrão construtivo NBR 12430.			
	DN-400	R\$ 6.637,00	peça
	DN-500	R\$ 9.770,00	peça
	DN-600	R\$ 20.360,00	peça
	DN-700	R\$ 29.370,00	peça
Tubo em Ferro Fundido Dúctil, centrifugado, Série K-9 , conforme NBR 7675 revisada em 2005, com ponta e bolsa de junta elástica do tipo JE2GS de acordo com a NBR 13747 com respectivo anel de borracha conforme NBR 7676.			
	DN-080 x 6,0 m	R\$ 81,00	metro
	DN-100 x 6,0 m	R\$ 88,50	metro
Tubo em Ferro Fundido Dúctil, centrifugado, Série K-7 , conforme NBR 7675 revisada em 2005, com ponta e bolsa de junta elástica do tipo JE2GS de acordo com a NBR 13747 com respectivo anel de borracha conforme NBR 7676.			
	DN-150 x 6,0 m	R\$ 105,00	metro
	DN-200 x 6,0 m	R\$ 128,00	metro
	DN-250 x 6,0 m	R\$ 177,00	metro
	DN-300 x 6,0 m	R\$ 204,00	metro
	DN-400 x 6,0 m	R\$ 243,00	metro
	DN-500 x 6,0 m	R\$ 324,00	metro
	DN-600 x 6,0 m	R\$ 431,00	metro
	DN-700 x 7,0 m	R\$ 510,00	metro
Tubo em PVC 6,3 extrudado, com ponta e bolsa de Junta Elástica Integrada, com anel de borracha incorporado, em barras de 6,0m de comprimento, para aplicações sob pressão de serviço de 1,0 MPa, classe 20, fabricado de acordo com a ABNT NBR 5647-1 e NBR 5647-2, para redes de distribuição de água potável. O tubo deve apresentar duas faixas indelévels em sua ponta para a marcação dos posicionamentos de montagem, sendo uma na posição de acoplamento máximo e outra na posição final da junta elástica. O tubo deve apresentar em toda a sua extensão, de forma indelével, a identificação do fabricante,			
	DN-050	R\$ 4,58	metro
	DN-075	R\$ 9,84	metro

	DN-100	R\$	16,40	metro
Tubo em PVC 12 com diâmetro externo equivalente ao de tubos de ferro fundido (DEFoFo), PN=1,0 MPa, com Junta Elástica Integrada, fabricado de acordo com a NBR 7665/2007. Os tubos devem ser fornecidos com duas marcas indeléveis em sua ponta para posicionamento de montagem, sendo uma no seu acoplamento máximo e outra na posição final da junta elástica.				
	DN-150	R\$	23,90	metro
	DN-200	R\$	41,10	metro
	DN-250	R\$	60,30	metro
	DN-300	R\$	88,60	metro
	DN-400	R\$	189,00	metro

15.7 Bibliografia

- SHIMIZU, TAMIO. *Pesquisa Operacional em engenharia, economia e administração*. Editora Guanabara dois, 1984, 360 páginas.
- SINCICH, TERRY. *Statistics by example*. 5a ed. New York, 1993, 1006 páginas;
- TOMAZ, PLINIO. *Previsão de consumo de água*; São Paulo, 2000, Navegar Editora, 250 páginas
- WALSKI, M. THOMAS. *Analysis of water distribution systems*. Florida, 1992, 275páginas. ISBN 0-89464-624-9