Login (/user/login) |

2.2 ESTIMAÇÃO DOS PARÂMETROS DO MODELO

(/)Início (/) / Análise de Regressão (/analise-de-regressao) / Regressão Linear Múltipla (/analise-de-regressao/regressao-linear-multipla) / 2.2 Estimação dos Parâmetros do Modelo

Suponha que temos n observações $(n>p)$ da variável resposta e das p variáveis explicativas. Assim, y_i é o valor da variável resposta na i-ésima observação enquanto que x_{ij} é o valor da variável x_j na i-ésima observação, $j=1,\dots,p$. Os dados de um MRLM podem ser representados da seguinte forma:

y	x_1	x_2	$\dots$	x_p
y_1	x_{11}	x_{12}	$\dots$	x_{1p}
y_2	x_{21}	x_{22}	$\dots$	x_{2p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_n	x_{n1}	x_{n2}	$\dots$	x_{np}

Tabela 2.2.1: Representação dos dados.

em que cada observação satisfaz

$$Y_i=\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\ldots+\beta_p x_{ip}+\varepsilon_i, \quad i=1,\ldots,n.$$

2.2.1 Método dos Mínimos Quadrados

ANÁLISE DE REGRESSÃO (/ANALISE-DE-REGRESSAO)

- 1. Regressão Linear Simples (/analise-de-regressao/regressao-linear-simples)
- 2. Regressão Linear Múltipla (/analise-de-regressao/regressao-linear-multipla)
  - 2. Regressão Linear Múltipla (/analise-de-regressao/regressao-linear-multipla)
  - 2.1 Modelo Estatístico (/analise-de-regressao/21-

O objetivo é minimizar a função

$$L=\sum_{i=1}^{n}\varepsilon_i^2=\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1x_{i1}-\beta_2x_{i2}-\ldots-\beta_px_{ip})^2.$$

Derivando L em função dos β's obtemos

(/)

$$\frac{\partial L}{\partial\beta_0}=-2\sum_{i=1}^{n}[Y_i-\beta_0-\beta_1x_{i1}-\beta_2x_{i2}-\cdots-\beta_px_{ip}],$$

$$\frac{\partial L}{\partial\beta_j}=-2\sum_{i=1}^{n}[Y_i-\beta_0-\beta_1x_{i1}-\beta_2x_{i2}-\cdots-\beta_px_{ip}]x_{ji},\quad j=1,2,\ldots,p.$$

Igualando as derivadas parciais a zero e substituindo $\beta_0,\beta_1,\ldots,\beta_p$ por $\widehat{\beta}_0,\widehat{\beta}_1,\ldots,\widehat{\beta}_p$, temos o sistema de equações

$$\begin{cases} n\widehat{\beta}_0+\widehat{\beta}_1\sum_{i=1}^{n}x_{i1}+\widehat{\beta}_2\sum_{i=1}^{n}x_{i2}+\ldots+\widehat{\beta}_p\sum_{i=1}^{n}x_{ip}=\sum_{i=1}^{n}Y_i\\ \widehat{\beta}_0\sum_{i=1}^{n}x_{i1}+\widehat{\beta}_1\sum_{i=1}^{n}x_{i1}^2+\widehat{\beta}_2\sum_{i=1}^{n}x_{i1}x_{i2}+\ldots+\widehat{\beta}_p\sum_{i=1}^{n}x_{i1}x_{ip}=\sum_{i=1}^{n}x_{i1}Y_i\\ \vdots\\ \widehat{\beta}_0\sum_{i=1}^{n}x_{ip}+\widehat{\beta}_1\sum_{i=1}^{n}x_{ip}x_{i1}+\widehat{\beta}_2\sum_{i=1}^{n}x_{ip}x_{i2}+\ldots+\widehat{\beta}_p\sum_{i=1}^{n}{}^2_{ip}=\sum_{i=1}^{n}x_{ip}Y_i.\end{cases}$$

Resolvendo este sistema, obtemos os estimadores de mínimos quadrados $\widehat{\beta}_0,\ldots,\widehat{\beta}_p$ dos parâmetros do modelo em questão.

2.2.2 Representação matricial do MRLM

Notemos que os estimadores de mínimos quadrados dos parâmetros do "Modelo 2.2" (/content/21-modelo-estatístico#MRLM) podem ser facilmente encontrados considerando a notação matricial dos dados, que é de fácil manipulação. Desta forma, considerando a entrada de dados apresentada na Tabela 2.2.1, o modelo de Regressão Linear Múltipla pode ser escrito como

$$Y = X\beta + \varepsilon,$$

com

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \text{ e } \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

em que

- Y é um vetor $n \times 1$ cujos componentes corresponde às n respostas;
- X é uma matriz de dimensão $n \times (p+1)$ denominada matriz do modelo;
- ε é um vetor de dimensão $n \times 1$ cujos componentes são os erros e
- β é um vetor $(p+1) \times 1$ cujos elementos são os coeficientes de regressão.

O método de mínimos quadrados tem como objetivo encontrar o vetor $\widehat{\beta}$ que minimiza

$$L = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta) =$$

$$= Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta = Y'Y - 2\beta'X'Y + \beta'X'X\beta,$$

sendo que $Y'X\beta = \beta'X'Y$ pois o produto resulta em um escalar. A notação X' representa o transposto da matriz X enquanto que Y' e β' representam os transpostos dos vetores Y e β, respectivamente. Usando a técnica de derivação (em termos matriciais) obtemos

$$\frac{\partial L}{\partial \beta} = -2X'Y + 2X'X\beta.$$

Igualando a zero e substituindo o vetor β pelo vetor $\widehat{\beta}$, temos

$$(X'X)\widehat{\beta} = X'Y.$$

Em geral, a matriz $(X'X)$ é não singular, ou seja, tem determinante diferente de zero, e portanto é invertível. Desta forma, conclui-se que os estimadores para os parâmetros $\beta_j, \quad j=0,\dots,p$ são dados pelo vetor

$$\widehat{\beta}=(X'X)^{-1}X'Y.$$

Portanto, o modelo de regressão linear ajustado e o vetor de resíduos são respectivamente

$$\widehat{Y}=X\widehat{\beta} \quad \text{e} \quad e=Y-\widehat{Y}=Y-\widehat{Y}.$$

Ao substituirmos os estimadores de mínimos quadrados, obtemos que $\widehat{Y}=HY$ no qual $H=X(X'X)^{-1}X'$ é a matriz chapéu, ou matriz de projeção do vetor de respostas Y no vetor de respostas ajustadas $\widehat{Y}$.

Exemplo 2.2.1

Com os dados do exemplo na "Motivação 2" (/content/2-regressão-linear-múltipla#motivacao2), obter as estimativas dos parâmetros do "Modelo 2.2" (/content/21-modelo-estatístico#MRLM).

clique aqui para efetuar o download dos dados utilizados nesse exemplo (/sites/default/files/analise_regressao/planilhas/RegMult1.xls)

Solução:

Sejam a variável resposta Ganho (Y) e as variáveis explicativas Tempo (x_1) e Dose (x_2).

Temos que $n=14, \sum\limits_{i=1}^{14} Y_i=17.495;$ $\sum\limits_{i=1}^{14} x_{i1}=3.155$ e $\sum\limits_{i=1}^{14} x_{i2}=60,72.$ Além disso, $\sum\limits_{i=1}^{14} x_{i1}^2=716.425;$ $\sum\limits_{i=1}^{14} x_{i2}^2=264,2584;$ $\sum\limits_{i=1}^{14} x_{i1}x_{i2}=13.683,50;$ $\sum\limits_{i=1}^{14} x_{i1}Y_i=4.001.120$ e $\sum\limits_{i=1}^{14} x_{i2}Y_i=75.738,30.$

As equações normais serão

$$\left\{\begin{array}{c} n\widehat{\beta}_0+\widehat{\beta}_1\sum\limits_{i=1}^{n}x_{i1}+\widehat{\beta}_2\sum\limits_{i=1}^{n}x_{i2}=\sum\limits_{i=1}^{n}Y_i \\ \widehat{\beta}_0\sum\limits_{i=1}^{n}x_{i1}+\widehat{\beta}_1\sum\limits_{i=1}^{n}x_{i1}^2+\widehat{\beta}_2\sum\limits_{i=1}^{n}x_{i1}x_{i2}=\sum\limits_{i=1}^{n}x_{i1}Y_i \\ \widehat{\beta}_0\sum\limits_{i=1}^{n}x_{i2}+\widehat{\beta}_1\sum\limits_{i=1}^{n}x_{i1}x_{i2}+\widehat{\beta}_2\sum\limits_{i=1}^{n}x_{i2}^2=\sum\limits_{i=1}^{n}x_{i2}Y_i \end{array}\right.$$

Substituindo os valores para esse exemplo, temos

$$\left\{\begin{array}{c} 14\,\widehat{\beta}_0\;+\;3.155\,\widehat{\beta}_1\;+\;60,72\,\widehat{\beta}_2\;=\;17.495 \\ 3.155\,\widehat{\beta}_0\;+\;716.425\,\widehat{\beta}_1\;+\;13.683,50\,\widehat{\beta}_2\;=\;4.001.120 \\ 60,72\,\widehat{\beta}_0\;+\;13.683,50\,\widehat{\beta}_1\;+\;264,2584\,\widehat{\beta}_2\;=\;75.738,30 \end{array}\right.$$

Resolvendo o sistema, obtemos

$\widehat{\beta}_0=-520,08,\widehat{\beta}_1=10,78$ e $\widehat{\beta}_2=-152,15.$

Na representação matricial temos

$$(X'X)=\begin{bmatrix} 14,00 & 3.155,00 & 60,72 \\ 3.155,00 & 716.425,00 & 13.683,50 \\ 60,72 & 13.683,50 & 264,26 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 30,24760 & -0,04172 & -4,78994 \\ -0,04172 & 0,00018 & 0,00004 \\ -4,78994 & 0,00004 & 1,10244 \end{bmatrix} \quad \text{e}$$

$$X'y=\begin{bmatrix} 17.495,0 \\ 4.001.120,0 \\ 75.738,3 \end{bmatrix}.$$

Logo, as estimativas $\widehat{\beta}$ são dadas por

$$\widehat{\beta}=(X'X)^{-1}X'y=\begin{bmatrix} -520,08 \\ 10,78 \\ -152,15 \end{bmatrix}.$$

e portanto, $\widehat{y}=-520,08+10,78x_1-152,15x_2$.

Não podemos afirmar que a variável x_1 aumenta o ganho, pois as variáveis x_1 e x_2 estão em unidades diferentes. No exemplo 2.2.3 abaixo, as variáveis foram transformadas para que fiquem na mesma unidade.

Exemplo 2.2.2

Interpretar os resultados do "Exemplo 2.2.1" (/content/23-estimação-dos-parâmetros-do-modelo-0#exemplo241).

(/)

Solução:

Com os dados do exemplo na "Motivação 2" (/content/2-regressão-linear-múltipla#motivacao2), obtemos o modelo

$$E(Y \mid x) = -520,08 - 152,15\, Dose + 10,78\, Tempo.$$

Se o $Tempo = 225$ (constante), então

$$E(Y \mid x) = 1.905,42 - 152,15\, Dose.$$

Assim,

$\beta_1 = -152,15$ indica que a cada acréscimo de uma unidade na $Dose$ a resposta média decrescerá $152,15$ unidades;

$\beta_2 = 10,78$ indica um acréscimo na resposta média de $10,78$ unidades para cada acréscimo de uma unidade na variável $Dose$.

2.2.3 Transformação de dados

Em determinadas situações é usual transformarmos as variáveis explicativas dos dados originais para que fiquem na mesma unidade (escala), facilitando a interpretação dos resultados. Suponha que $\xi_1, \xi_2, \ldots, \xi_p$ são as variáveis explicativas originais do modelo. A expressão para transformar as variáveis explicativas ξ é dada por

$$x_{ij} = \frac{\xi_{ij} - \frac{[max(\xi_{ij}) + min(\xi_{ij})]}{2}}{\frac{[max(\xi_{ij}) - min(\xi_{ij})]}{2}},$$

em que j indica a variável que está sendo transformada, $j = 1, \dots, p$. Desta forma, temos que os valores das variáveis explicativas transformadas estão todos entre -1 e 1.

Notemos que os valores das estimativas dos parâmetros do modelo de regressão linear múltipla não são os mesmos considerando as variáveis originais e as variáveis transformadas.

Exemplo 2.2.3

Considerando os dados do Exemplo 2.2.1, transformar as variáveis explicativas para que fiquem na mesma unidade (escala). Então, ajustar o modelo de regressão linear múltipla aos dados transformados.

clique aqui para efetuar o download dos dados utilizados nesse exemplo (/sites/default/files/analise_regressao/planilhas/RegMult1.xls)

Considerando que ξ_1 e ξ_2 são as variáveis explicativas no conjunto de dados original, temos que as variáveis explicativas transformadas, denotadas por x_1 e x_2 são dadas por

- Para x_{i1}:

$$x_{i1} = \frac{\xi_{i1} - 225}{30}.$$

- Para x_{i2}:

$$x_{i2} = \frac{\xi_{i2} - 4,36}{0,36}.$$

Observação	Tempo (min) (ξ_1)	Dose de íons 10^{14} (ξ_2)	x_1	x_2	Ganho (y)
1	195	4	-1	-1	(/)1.004
2	255	4	1	-1	1.636
3	195	4,6	-1	0,6667	852
4	255	4,6	1	0,6667	1.506
5	225	4,2	0	-0,4444	1.272
6	225	4,1	0	-0,7222	1.270
7	225	4,6	0	0,6667	1.269
8	195	4,3	-1	-0,1667	903
9	255	4,3	1	-0,1667	1.555
10	225	4	0	-1	1.260
11	225	4,7	0	0,9444	1.146
12	225	4,3	0	-0,1667	1.276
13	225	4,72	0	1	1.225
14	230	4,3	0,1667	-0,1667	1.321

Tabela 2.2.2: Dados Transformados.

Considerando os dados transformados, temos que a matriz X e o vetor y são dados respectivamente por

$$X = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 0,6667 \\ 1 & 1 & 0,6667 \\ 1 & 0 & -0,4444 \\ 1 & 0 & -0,7222 \\ 1 & 0 & 0,6667 \\ 1 & -1 & -0,1667 \\ 1 & 1 & -0,1667 \\ 1 & 0 & -1 \\ 1 & 0 & 0,9444 \\ 1 & 0 & -0,1667 \\ 1 & 0 & 1 \\ 1 & 0,1667 & -0,1667 \end{bmatrix}, \qquad y = \begin{bmatrix} 1004 \\ 1636 \\ 852 \\ 1506 \\ 1272 \\ 1270 \\ 1269 \\ 903 \\ 1555 \\ 1260 \\ 1146 \\ 1276 \\ 1225 \\ 1321 \end{bmatrix}.$$

Assim, a matriz $X'X$ é

$$(X'X) = \begin{bmatrix} 14 & 0,1667 & -0,8889 \\ 0,1667 & 6,0278 & -0,0278 \\ -0,8889 & -0,0278 & 7,0556 \end{bmatrix}$$

(/)

$$(X'X)^{-1} = \begin{bmatrix} 0,0720 & -0,0019 & 0,0091 \\ -0,0019 & 0,1660 & 0,0004 \\ 0,0091 & 0,0004 & 0,1429 \end{bmatrix} \quad \text{e}$$

$$X'y = \begin{bmatrix} 17495 \\ 2158,167 \\ -1499,722 \end{bmatrix}.$$

Portanto, as estimativas $\widehat{\beta}$ são

$$\widehat{\beta} = (X'X)^{-1}X'y = \begin{bmatrix} 1242,31 \\ 323,43 \\ -54,77 \end{bmatrix}.$$

Assim, a equação da regressão é dada por

$$\widehat{y} = 1.242,31 + 323,43\ x_1 - 54,77\ x_2.$$

Usando o **software Action** temos os seguintes resultados:

- Considerando a escala original dos dados

	Coeficientes			
Preditor	*Estimativa*	*Desvio Padrão*	*Estat. T*	*P-Valor*
Intercepto	-520,0766769	192,1070916	-2,707222688	0,020391846
Tempo	10,78115789	0,474319633	22,72973146	1,34912E-10
Dose_de_íons	-152,1488747	36,67543895	-4,148522255	0,001620499

- Considerando a transformação

Preditor	Coeficientes Estimativa	Desvio Padrão	Estat. T	P-valor
Intercepto	1242,314755	9,374475111	132,520993[illegible]	[illegible]
x1	323,4347368	14,229589	22,72973146	[illegible]
x2	-54,77359489	13,20315802	-4,14852[illegible]	[illegible]

(/manual-modelos/modelo-de-regressao-linear-simples-sobre-tratamento-termico-0)

(/)

Para entender como executar essa função do ***Software Action***, você pode consultar o **manual do usuário.** (/manual-modelos/modelo-de-regressao-linear-simples-sobre-tratamento-termico-0)

‹ 2.1 Modelo Estatístico (/analise-de-regressao/21-modelo-estatistico)

acima (/analise-de-regressao/regressao-linear-multipla)

2.3 Propriedades dos Estimadores › (/analise-de-regressao/23-propriedades-dos-estimadores)

Dúvidas sobre esse conteúdo? Comente:

0 Comentários Portal Action Entrar

Recomendar 2 Compartilhar Ordenar por Mais votados

Iniciar a discussão...

(/)

FAZER LOGIN COM

OU REGISTRE-SE NO DISQUS ?

Nome

Seja o primeiro a comentar.

Inscreva-se Adicione o Disqus no seu siteAdicionar DisqusAdicionar

SOBRE O PORTAL ACTION

O Portal Action é mantido por *Estatcamp - Consultoria Estatística e Qualidade* e por *DIGUP - Desenvolvimento de Sistemas e Consultoria Estatística*, com o objetivo de disponibilizar uma ferramenta estatística em conjunto com uma fonte de informação útil aos profissionais interessados.

ESTATCAMP (http://www.estatcamp.com.br)

LINKS IMPORTANTES

- Home (http://www.portalaction.com.br/)
- Action Stat (/sobre-o-action)
  - Sobre o Action Stat (/sobre-o-action)
  - Action Stat Pro (/action-stat-pro)
  - Action Stat Quality (/action-stat-quality)
  - Action Stat Pharma (/action-stat-pharma)

- Download (/content/download-action)
- Manual do Usuário (/manual-action-[illegible])
- FAQ (/faq-page)
- Ambiente de Aprendizado (/ambiente-virtual-de-aprendizado)
- Serviços (/servicos)
- Loja (http://loja.portalaction.com.br)
- Contato (/contato)

FACEBOOK

CONTATO

Maestro Joao Seppe, 900, São Carlos - SP | CEP 13561-180

Telefone: (16) 3376-2047

E-Mail: estatistica@estatcamp.com.br