

Exercício 5

A Tabela 1 apresenta a concentração de NbOCl3 em um reator de fluxo em pistão como função de diversas variáveis controláveis. Os dados podem ser obtidos em ftp://ftp.wiley.com/public/sci_tech_med/introduction_linear_regression, em um arquivo de extensão .zip, sob o nome data-table-B6.XLS.

Tabela 1: Concentração de NbOCl3 (g-mol/L) (y) em um reator de fluxo em pistão e variáveis controláveis: concentração de COCl2 (x_1 , em g-mol/L); tempo (x_2 , em segundos); densidade molar (x_3 , em g-mol/L) e fração molar de CO2 (x_4 , em g-mol/L).

y	x1	x2	x3	x4
0,00045	0,0105	90,9	0,0164	0,0177
0,00045	0,011	84,6	0,0165	0,0172
0,000473	0,0106	88,9	0,0164	0,0157
0,000507	0,0116	488,7	0,0187	0,0082
0,000457	0,0121	454,4	0,0187	0,007
0,000452	0,0123	439,2	0,0187	0,0065
0,000453	0,0122	447,1	0,0186	0,0071
0,000426	0,0122	451,6	0,0187	0,0062
0,001215	0,0123	487,8	0,0192	0,0153
0,001256	0,0122	467,6	0,0192	0,0129
0,001145	0,0094	95,4	0,0163	0,0354
0,001085	0,01	87,1	0,0162	0,0342
0,001066	0,0101	82,7	0,0162	0,0323
0,001111	0,0099	87	0,0163	0,0337
0,001364	0,011	516,4	0,019	0,0161
0,001254	0,0117	488,4	0,0189	0,0149
0,001396	0,011	534,5	0,0189	0,0163
0,001575	0,0104	542,3	0,0189	0,0164
0,001615	0,0067	98,8	0,0163	0,0379
0,001733	0,0066	84,8	0,0162	0,036
0,002753	0,0044	69,6	0,0163	0,0327
0,003186	0,0073	436,9	0,0189	0,0263
0,003227	0,0078	406,3	0,0192	0,02
0,003469	0,0067	447,9	0,0192	0,0197
0,001911	0,0091	58,5	0,0164	0,0331
0,002588	0,0079	394,3	0,0177	0,0674
0,002635	0,0068	461	0,0174	0,077
0,002725	0,0065	469,2	0,0173	0,078

- (a) Formalize e estime um modelo de regressão linear múltiplo com intercepto relacionando a concentração de NbOCl3 à concentração de COCl2 (x_1) e fração molar (x_4).

Resolução

O modelo de regressão linear múltiplo a ser utilizado é da forma $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_4 x_{4i} + e_i$ supondo $\mathbb{E}(e_i) = 0$ e $Var(e_i) = \sigma^2$ e e_1, \dots, e_n independentes. Relacionando a concentração de NbOCl3 à concentração de COCl2 (x_1) e fração molar (x_4).

Logo temos nossa reta estimada: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_4 x_{4i} = 0.004833 - 0.34498x_{1i} - 0.00014x_{4i}$

- (b) Para quais valores de x_1 e x_4 , dentro da região experimental, a resposta estimada \hat{Y} é máxima? Qual é esse máximo?

Resolução

Os valores de x_1 e x_4 , dentro da região experimental são: $x_1 = 0.0067$ e $x_4 = 0.0197$

O valor máximo de \hat{Y} é:

```
max(model$fitted.values)
```

```
## [1] 0.003310685
```

- (c) Estude as covariáveis e argumente a favor ou contra existência de multicolinearidade no modelo ajustado.

Resolução

Podemos notar a existência de multicolinearidade no modelo ajustado $\hat{Y} = 0.004833 - 0.34498x_{1i} - 0.00014x_{4i}$, pois analisando a matriz

$$(X^T X) = \begin{pmatrix} 28.0000 & 0.27030000 & 0.74120000 \\ 0.2703 & 0.00274965 & 0.00632464 \\ 0.7412 & 0.00632464 & 0.03005396 \end{pmatrix}$$

Podemos notar elementos muito próximos de zero (uma possível singularidade), além de que o $\det(X^T X) = 2.166 * 10^{-5}$ também é muito próximo de zero, dando a entender que existe alguma dependência linear entre as linhas da matriz $(X^T X)$, e também temos que a correlação entre as covariáveis é $\text{cor}(x_1, x_4) = -0.686$ sendo mais um fator a favor da multicolinearidade, por ter uma relação forte negativa entre as covariáveis, assim trazendo grandes consequências no cálculo da matriz $(X^T X)^{-1}$, pois seus elementos ficariam muito grandes, consequentemente levando a um modelo pouco confiável e possíveis análises errôneas.