

Exemplo sobre uma reação química: análise de correlação canônica

Prof. Caio Azevedo

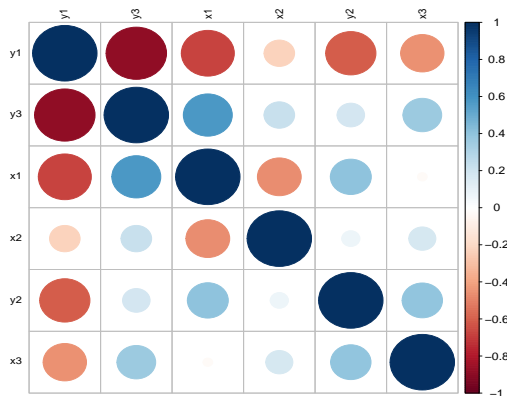
- Os dados são fornecidos em Rencher (Methods of multivariate analysis, 2002, p. 365) e correspondem a um experimento envolvendo uma reação química.
- As variáveis observadas são classificadas como dados de entrada $\mathbf{X}^{(1)} = (x_1, x_2, x_3)'$ e dados de resultado do experimento $\mathbf{X}^{(2)} = (y_1, y_2, y_3)'$.

- As variáveis observadas foram as seguintes:
 - x_1 : temperatura
 - x_2 : concentração
 - x_3 : tempo
 - y_1 : porcentagem de material inicial inalterado
 - y_2 : porcentagem convertida para o produto desejado
 - y_3 : porcentagem de um derivado não desejado

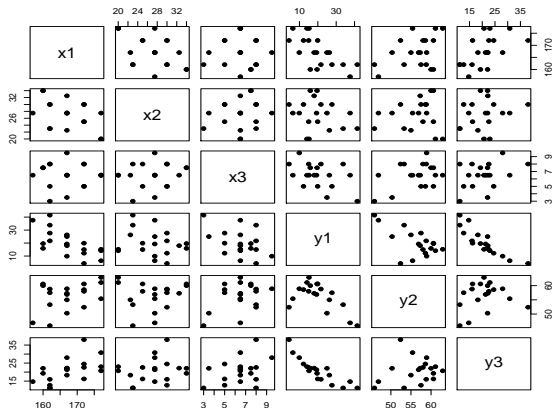
Medidas descritivas

	y_1	y_2	y_3	x_1	x_2	x_3
Média	20.18	56.34	20.78	167.32	27.18	6.50
Var.	99.30	22.25	45.27	38.67	17.95	2.67
DP	9.96	4.72	6.73	6.22	4.24	1.63
Mínimo	4.30	45.90	11.20	157.00	20.00	3.00
Mediana	19.30	57.50	21.30	167.00	27.50	6.50
Máximo	41.50	62.90	38.00	177.00	34.00	9.50

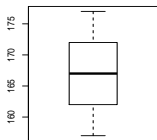
Gráfico de correlações



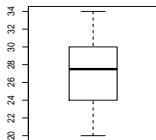
Gráficos de dispersão



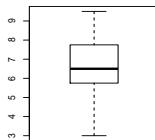
Boxplots



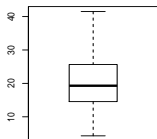
x1



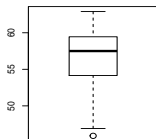
x2



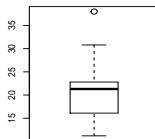
x3



y1

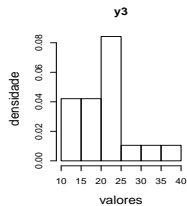
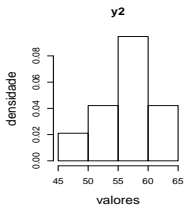
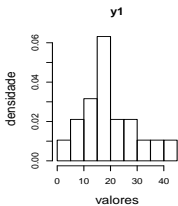
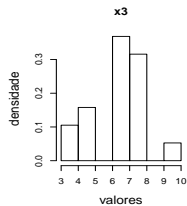
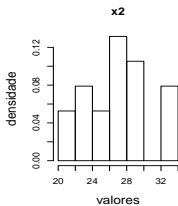
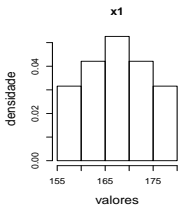


y2

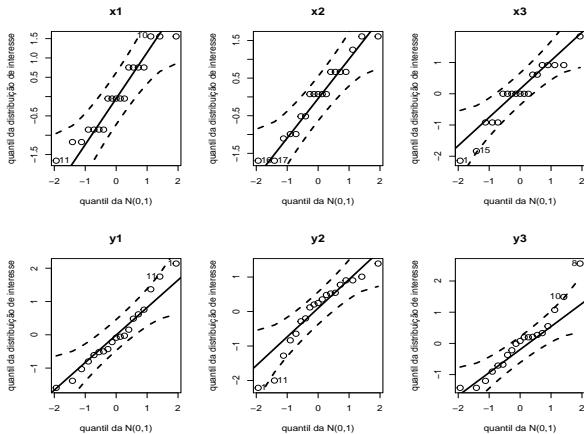


y3

Histogramas



Gráficos de quantis-quantis



Análise de correlações canônicas

- Utilizou-se a função “cc” do pacote “CCA”.
- O objetivo é encontrar combinações lineares de $\mathbf{X}^{(1)} = (x_1, x_2, x_3)'$ e $\mathbf{X}^{(2)} = (y_1, y_2, y_3)'$ com correlação máxima.
- As três maiores correlações canônicas foram 0.98, 0.30 e 0.06.

- Os valores dos coeficientes das combinações lineares foram

$$\mathbf{A} = \begin{bmatrix} -1.00 & -0.63 & -0.35 \\ -0.43 & -0.52 & 0.95 \\ -0.31 & 0.81 & 0.06 \end{bmatrix} \quad \text{e} \quad \mathbf{B} = \begin{bmatrix} 1.70 & 0.33 & 0.58 \\ 6.24 & 3.44 & 4.80 \\ 3.76 & 1.02 & 3.63 \end{bmatrix}$$

- Por exemplo, o primeiro par de variáveis canônicas é dado por:

$$\hat{U}_1 = -x_1 - 0.63x_2 - 0.35x_3,$$

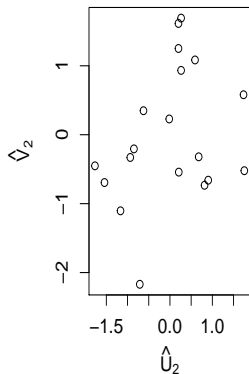
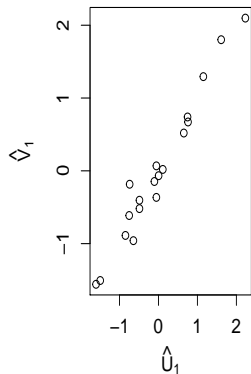
$$\hat{V}_1 = 1.70y_1 + 0.33y_2 + 0.58y_3.$$

Correlações entre variáveis originais e variáveis canônicas

variável	\hat{U}_1	\hat{V}_1	\hat{U}_2	\hat{V}_2
x_1	-0.7	-0.69	-0.21	-0.06
x_2	-0.23	-0.23	-0.15	-0.04
x_3	-0.44	-0.43	0.87	0.26
y_1	0.98	0.99	-0.02	-0.08
y_2	-0.59	-0.6	0.16	0.53
y_3	-0.85	-0.86	-0.02	-0.07

- Proporção da variância total das variáveis padronizadas explicada pelo primeiro par de variáveis canônicas:
 - Para $X^{(1)}$: 24.61%
 - Para $X^{(2)}$: 69.90%
- Proporção da variância total das variáveis padronizadas explicada pelos dois primeiros pares de variáveis canônicas:
 - Para $X^{(1)}$: 52.21%
 - Para $X^{(2)}$: 79.47%

Gráficos de dispersão das variáveis canônicas



Observações

- Pela proporção da variação total explicada, notamos que parece ser mais adequado utilizar os dois primeiros pares de variáveis canônicas, ao invés do primeiro par somente.
- \hat{U}_1 e \hat{V}_1 parecem se tratar de índices relacionados à quantidade de material que não é convertida no resultado desejado (permanece inalterado ou gera outro produto).
- \hat{U}_2 fornece um índice relacionado ao tempo da reação.
- \hat{V}_2 fornece um índice relacionado à proporção de material que resultou no produto desejado.