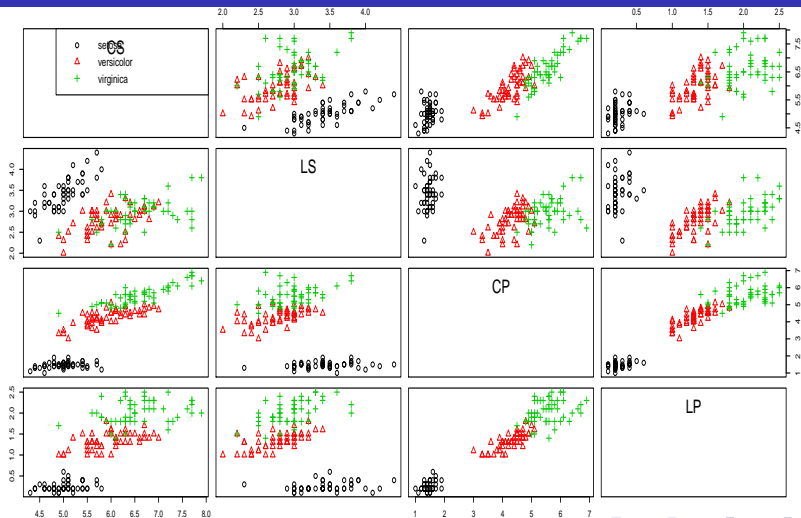


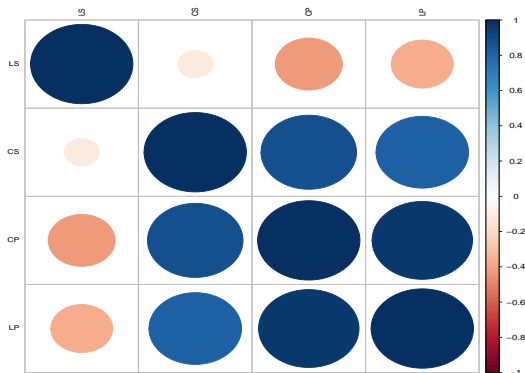
Análise de Correlação Canônica: parte 2

Prof. Caio Azevedo

Voltando ao Exemplo 1: iris de Fisher



Correlações



Voltando ao Exemplo 1: iris de Fisher

- Como as variáveis relativas à sépala se relacionam com as aquelas relativas à pétala?
- Como as espécies se comportam em relação à essas variáveis?
- Neste caso podemos obter dois pares (no máximo) de variáveis canônicas.
- Correlações canônicas: $\tilde{\rho}_1^* = 0,940$, $\tilde{\rho}_2^* = 0,124$.
- Utilizou-se a função “cc” do pacote “CCA”.

Variável	$U_1^{(Z)}$	$\text{Corr}(U_1^{(Z)}, \mathbf{X}^{(1)})$	$U_2^{(Z)}$	$\text{Corr}(U_2^{(Z)}, \mathbf{X}^{(1)})$
CS	-0,89	-0,93	0,48	0,37
LS	0,37	0,48	0,94	0,88

$$\hat{U}_1^{(Z)} = -0,89CS + 0,37LS$$

$$\hat{U}_2^{(Z)} = 0,48CS + 0,94LS$$

Variável	$V_1^{(Z)}$	$Corr(V_1^{(Z)}, \mathbf{X}^{(2)})$	$V_2^{(Z)}$	$Corr(V_2^{(Z)}, \mathbf{X}^{(2)})$
CP	-1,50	-0,99	-3,39	0,14
LP	0,53	-0,91	3,67	0,40

$$\widehat{V}_1^{(Z)} = -1,50CP + 0,53LP$$

$$\widehat{V}_2^{(Z)} = -3,39CP + 3,67LP$$

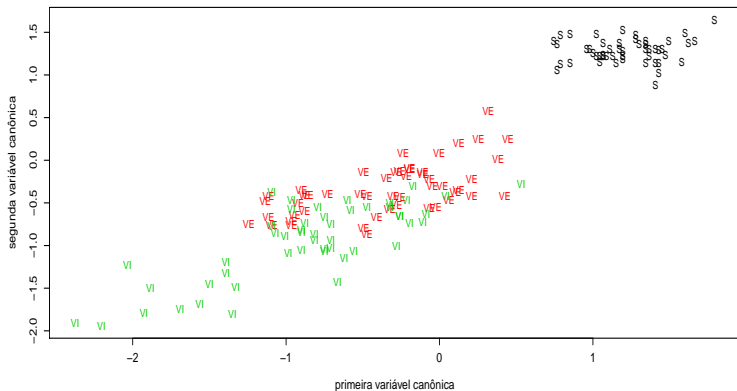
Aproximação das matrizes de correlações

	CS	LS
CS	0,14	0,33
LS	0,33	0,77

	CP	LP
CP	0,02	0,06
LP	0,06	0,16

	CP	LP
CS	0,01	0,02
LS	0,02	0,04

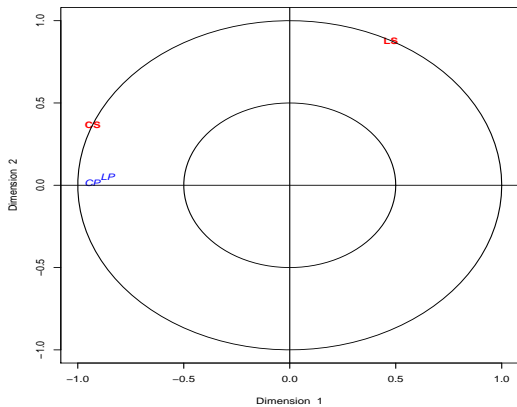
Dispersão entre as variáveis canônicas



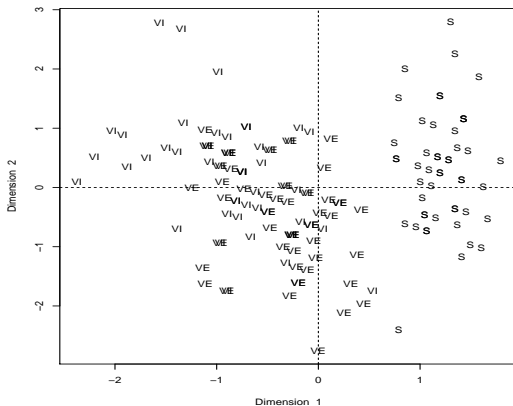
Gráficos das variáveis e das observações

- Figura 1 (gráfico das variáveis) : dispersão entre $Corre(U_1, \mathbf{X}_1^{(1)}) \times Corre(U_2, \mathbf{X}_1^{(1)})$ (vermelho) e $Corre(V_1, \mathbf{X}^{(2)}) \times Corre(V_2, \mathbf{X}^{(2)})$ (azul).
- Figura 2 (gráfico das observações) : dispersão entre U_1 e U_2 .

Gráficos das variáveis



Gráficos das observações



Exemplo 8: Cereais

- Quantidade de nutrientes, vitaminas e outros compostos alimentares em porções (100 gramas?).
- Várias marcas de cereal.
- Variáveis escolhidas: calorias, proteína, gordura, fibra, sódio, carboidrato, açúcar, potássio.
- Como as variáveis “nutricionais” (calorias, proteína, gordura, fibra, carboidrato e açúcar) se relacionam com as variáveis “elementais” (sódio e potássio)?

Exemplo 8: Cereais

- Como as marcas se comportam em relação à essas variáveis?
- Marcas: G = General Mills, K = Kellogs, N = Nabisco , P = post, Q = Quaker Oats, R = Ralston Purine.
- Disponível no pacote MASS do “R” sob o nome de “UScereal” .

Medidas descritivas: variáveis nutricionais

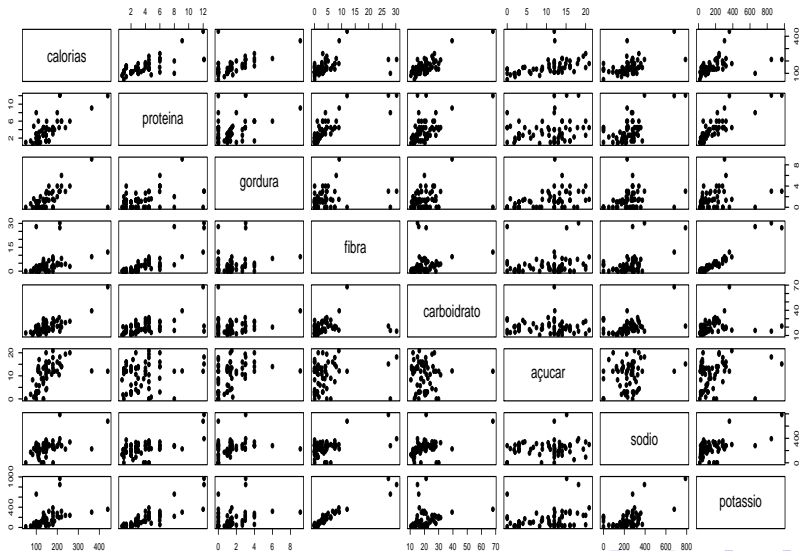
M.R.	Variável					
	calorias	proteina	gordura	fibra	carboidrato	açucar
Média	149,41	3,68	1,42	3,87	19,97	10,05
Var.	3895,24	6,98	2,71	37,62	71,71	34,05
DP	62,41	2,64	1,65	6,13	8,47	5,84
CV(%)	41,77	71,74	115,80	158,45	42,41	58,06
Mínimo	50,00	0,75	0,00	0,00	10,53	0,00
Mediana	134,33	3,00	1,00	2,00	18,67	12,00
Máximo	440,00	12,12	9,09	30,30	68,00	20,90

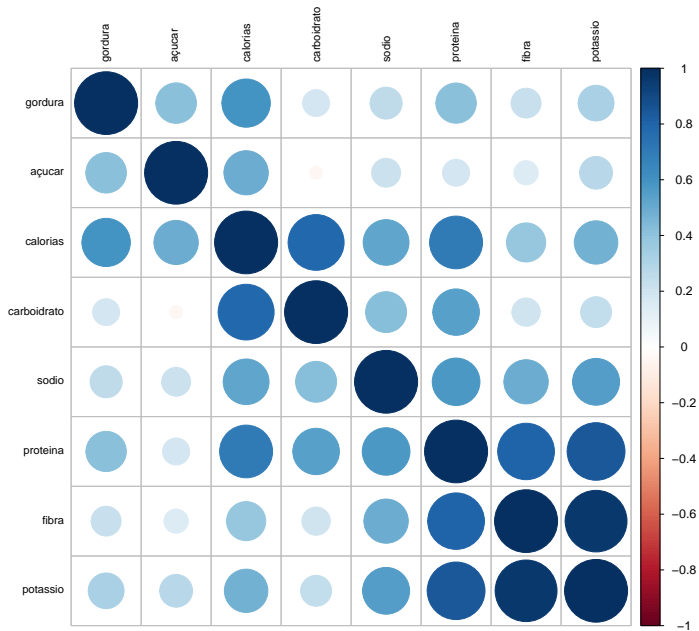
M.R.: medida - resumo



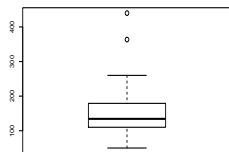
Medidas descritivas: variáveis elementais

M.R.	Variável	
	sodio	potassio
Média	237,84	159,12
Var.	17064,10	32503,97
DP	130,63	180,29
CV(%)	54,92	113,30
Mínimo	0,00	15,00
Mediana	232,00	96,59
Máximo	787,88	969,70

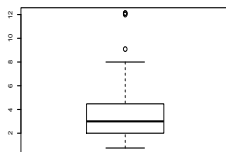




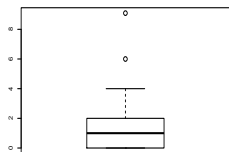
Box plot: variáveis nutricionais



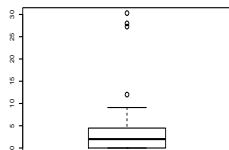
calorias



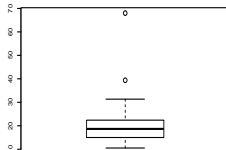
proteina



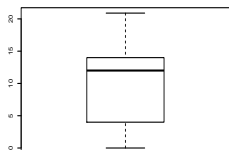
gordura



fibra

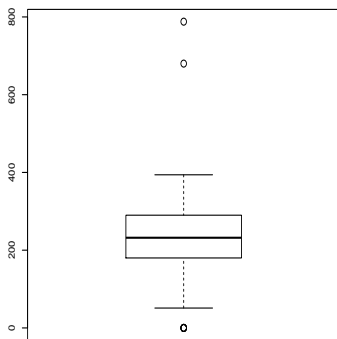


carboidrato

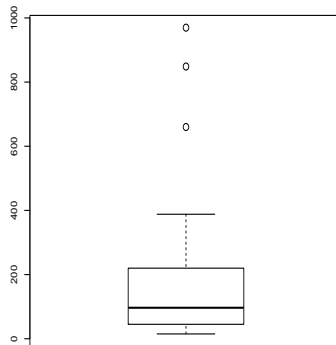


açucar

Box plot: variáveis elementais

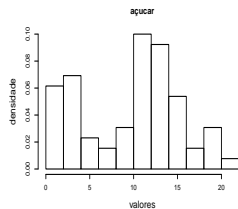
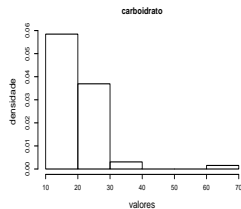
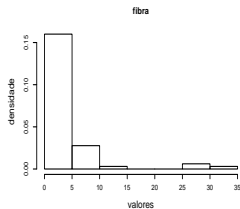
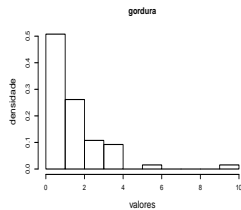
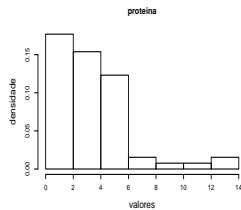
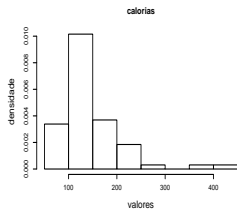


sodio

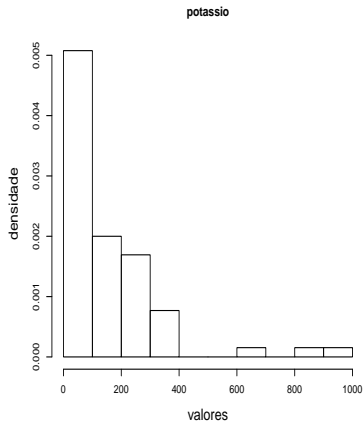
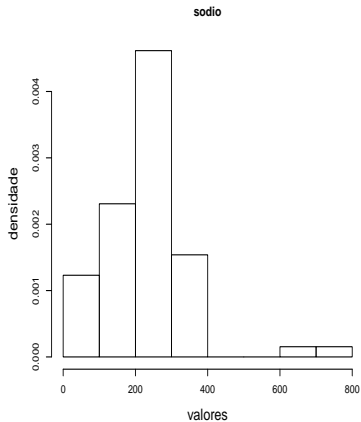


potassio

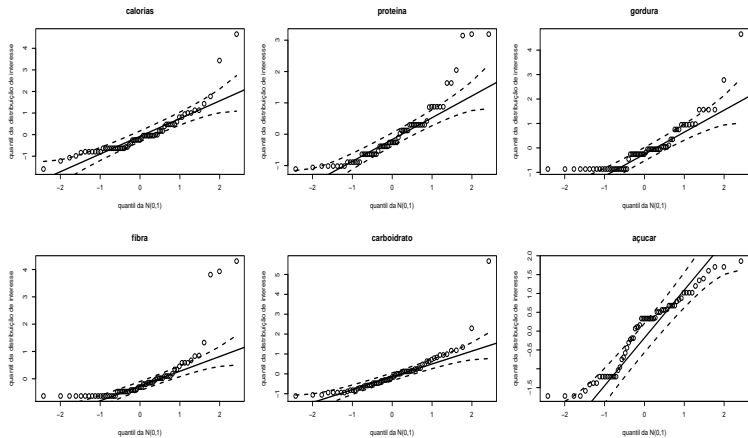
Histograma: variáveis nutricionais



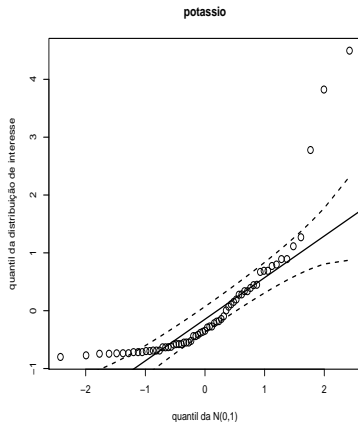
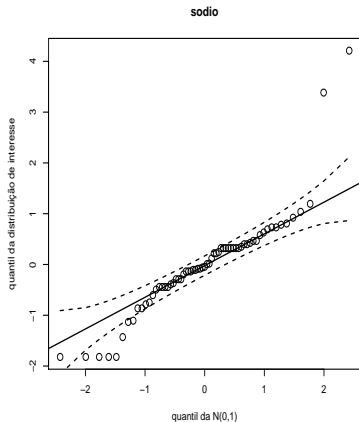
Histograma: variáveis elementais



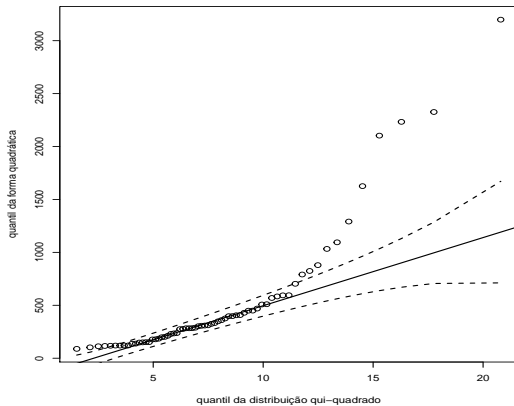
QQ-plot: variáveis nutricionais



QQ-plot: variáveis elementais



QQ-plot: todas as variáveis



Medidas descritivas por fabricante: calorias

Marca	Medida resumo							
	média	dp	var.	cv	min.	med.	max.	n
G	137,79	45,00	2024,69	32,66	73,33	136,67	260,00	22
K	149,67	45,77	2095,24	30,58	100,00	146,67	238,81	21
N	160,26	44,91	2017,24	28,03	134,33	134,33	212,12	3
P	194,76	122,90	15104,38	63,10	82,71	146,67	440,00	9
Q	135,85	55,90	3124,89	41,15	50,00	149,25	200,00	5
R	124,85	20,82	433,60	16,68	97,35	133,33	149,25	5

Medidas descritivas por fabricante: proteína

Marca	Medida resumo							
	média	dp	var.	cv	min.	med.	max.	n
G	2,88	1,63	2,65	56,47	1,00	2,67	6,00	22
K	3,92	2,55	6,52	65,18	1,00	3,75	12,12	21
N	7,03	4,41	19,47	62,81	4,48	4,48	12,12	3
P	4,70	3,66	13,40	77,90	0,75	4,48	12,00	9
Q	3,46	3,30	10,88	95,33	1,00	1,33	8,00	5
R	2,60	1,32	1,74	50,75	0,88	2,67	4,48	5

Medidas descritivas por fabricante: gordura

Marca	Medida resumo							
	média	dp	var.	cv	min.	med.	max.	n
G	1,78	1,18	1,38	65,87	0,67	1,33	4,00	22
K	1,02	1,57	2,48	153,82	0,00	0,00	6,00	21
N	1,01	1,75	3,06	173,21	0,00	0,00	3,03	3
P	1,78	2,92	8,53	163,93	0,00	1,14	9,09	9
Q	1,93	1,16	1,35	60,14	0,00	2,00	2,99	5
R	0,60	0,82	0,67	136,93	0,00	0,00	1,49	5

Medidas descritivas por fabricante: fibra

Marca	Medida resumo							
	média	dp	var.	cv	min.	med.	max.	n
G	1,65	1,62	2,61	97,97	0,00	1,80	5,00	22
K	5,07	7,87	62,01	155,36	0,00	1,33	28,00	21
N	13,58	14,50	210,21	106,74	4,48	5,97	30,30	3
P	5,38	4,61	21,22	85,70	0,00	7,46	12,00	9
Q	1,60	1,81	3,29	113,57	0,00	1,00	4,00	5
R	2,36	2,73	7,43	115,67	0,00	1,33	5,97	5

Medidas descritivas por fabricante: carboidrato

Marca	Medida resumo							
	média	dp	var.	cv	min.	med.	max.	n
G	17,69	5,06	25,62	28,61	12,00	15,67	28,00	22
K	19,80	5,62	31,59	28,39	11,00	20,00	31,34	21
N	24,45	8,09	65,45	33,08	15,15	28,36	29,85	3
P	24,28	18,35	336,82	75,58	10,53	17,33	68,00	9
Q	17,38	6,39	40,77	36,74	12,00	16,00	28,00	5
R	22,82	1,93	3,71	8,44	20,35	22,39	25,37	5

Medidas descritivas por fabricante: açúcar

Marca	Medida resumo							
	média	dp	var.	cv	min.	med.	max.	n
G	10,05	5,27	27,80	52,46	0,80	12,00	20,00	22
K	10,80	6,12	37,42	56,65	0,00	12,00	20,00	21
N	6,06	10,50	110,19	173,21	0,00	0,00	18,18	3
P	12,71	4,99	24,88	39,24	5,68	12,12	20,90	9
Q	9,59	5,94	35,31	61,96	0,00	11,00	16,00	5
R	4,97	2,88	8,28	57,86	1,77	4,48	8,96	5

Medidas descritivas por fabricante: sódio

Marca	Medida resumo							
	média	dp	var.	cv	min.	med.	max.	n
G	240,39	64,79	4198,12	26,95	140,00	226,00	373,33	22
K	242,79	162,36	26359,88	66,87	0,00	266,67	787,88	21
N	131,31	227,44	51729,41	173,21	0,00	0,00	393,94	3
P	253,73	179,46	32206,46	70,73	51,14	227,27	680,00	9
Q	201,44	116,80	13642,75	57,98	0,00	223,88	293,33	5
R	277,50	48,99	2399,62	17,65	212,39	280,00	343,28	5

Medidas descritivas por fabricante: potássio

Marca	Medida resumo							
	média	dp	vari	cv	min.	med.	max.	n
G	109,20	72,16	5207,31	66,08	25,00	87,00	280,00	22
K	184,96	238,28	56779,29	128,83	20,00	59,70	969,70	21
N	412,18	378,14	142993,22	91,74	179,10	208,96	848,48	3
P	203,87	150,27	22582,25	73,71	26,32	283,58	388,06	9
Q	93,69	85,21	7260,49	90,95	15,00	46,67	220,00	5
R	103,28	76,87	5909,25	74,43	25,00	106,67	186,57	5

- Neste caso podemos obter dois pares (no máximo) de variáveis canônicas.
- Correlações canônicas: $\tilde{\rho}_1^* = 0,983$, $\tilde{\rho}_2^* = 0,395$.
- A função “cc” não permite obter todas as variáveis canônicas.

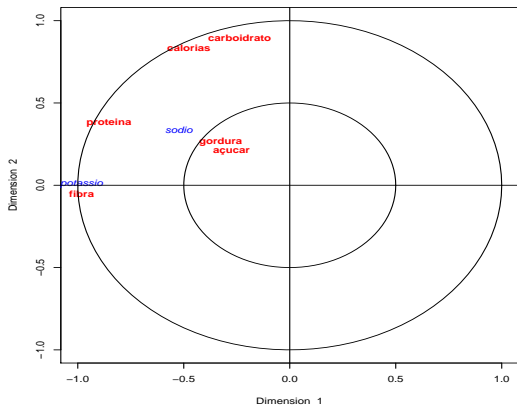
Variável	$U_1^{(Z)}$	$Corr(U_1^{(Z)}, \mathbf{X}^{(1)})$
calorias	0,77	-0,48
proteina	-0,27	-0,85
gordura	-0,22	-0,33
fibra	-0,85	-0,98
carboidrato	-0,50	-0,24
açucar	-0,41	-0,27

Variável	$V_1^{(Z)}$	$Corr(V_1^{(Z)}, \mathbf{X}^{(2)})$
sodio	1,20	-0,53
potassio	-0,64	<-0,99

Gráficos das variáveis e das observações

- Figura 1 (gráfico das variáveis) : dispersão entre $Corre(U_1, \mathbf{X}_1^{(1)}) \times Corre(U_2, \mathbf{X}_1^{(1)})$ (vermelho) e $Corre(V_1, \mathbf{X}^{(2)}) \times Corre(V_2, \mathbf{X}^{(2)})$ (azul).
- Figura 2 (gráfico das observações) : dispersão entre U_1 e U_2 .

Gráficos das variáveis



Gráficos das variáveis e das observações

