Análise de correlações canônicas)

Análise de correlações canônicas)

Número de variáveis canônicas)

Dados)

Análise de correlações canônicas

Prof. Caio Azevedo

24 de novembro de 2009

- Construir variáveis que preservem uma parte da variabilidade e que concentrem, em poucas delas, a estrutura de correlação das variáveis originais.
- Variáveis canônicas:
  - $\mathbf{U}_{(p \times p)} = \mathbf{A}_{(p \times p)} \mathbf{X}^{(1)}_{(p \times 1)}$ e $\mathbf{U}_{(q \times q)} = \mathbf{B}_{(q \times q)} \mathbf{X}^{(2)}_{(q \times 1)}$.
  - $\mathbf{U}^{(\mathbf{Z})}_{(p \times p)} = \mathbf{A}^{(\mathbf{Z})}_{(p \times p)} \mathbf{Z}^{(1)}_{(p \times 1)}$ e $\mathbf{U}^{(\mathbf{Z})}_{(q \times q)} = \mathbf{B}^{(\mathbf{Z})}_{(q \times q)} \mathbf{Z}^{(2)}_{(q \times 1)}$.
- Em geral, trabalha-se com as variáveis canônicas obtidas a partir das variáveis padronizadas: interpretação, menor influência das médias, me nor influência das variâncias.

- Equações:

$$
\begin{aligned}
U_1^{(Z)} &= a_{11}^{(Z)} Z_1^{(1)} + a_{12}^{(Z)} Z_2^{(1)} + \ldots + a_{1p}^{(Z)} Z_p^{(1)} \\
U_2^{(Z)} &= a_{21}^{(Z)} Z_1^{(1)} + a_{22}^{(Z)} Z_2^{(1)} + \ldots + a_{2p}^{(Z)} Z_p^{(1)} \\
&\vdots \\
U_p^{(Z)} &= a_{p1}^{(Z)} Z_1^{(1)} + a_{p2}^{(Z)} Z_2^{(1)} + \ldots + a_{pp}^{(Z)} Z_p^{(1)}
\end{aligned}
$$

$$
\begin{aligned}
V_1^{(Z)} &= b_{11}^{(Z)} Z_1^{(2)} + b_{12}^{(Z)} Z_2^{(2)} + \ldots + b_{1q}^{(Z)} Z_q^{(2)} \\
V_2^{(Z)} &= b_{21}^{(Z)} Z_1^{(2)} + b_{22}^{(Z)} Z_2^{(2)} + \ldots + b_{2q}^{(Z)} Z_q^{(2)} \\
&\vdots \\
V_q^{(Z)} &= b_{q1}^{(Z)} Z_1^{(2)} + b_{q2}^{(Z)} Z_2^{(2)} + \ldots + b_{qq}^{(Z)} Z_q^{(2)}
\end{aligned}
$$

- Magnitude dos "coeficientes canônicos".
- Correlação entre as variáveis canônicas e as variáveis originais: $Cor(\mathbf{Z}^{(1)}, \mathbf{U}^{(\mathbf{Z})}) = \mathbf{A}^{(\mathbf{Z})}\boldsymbol{\rho}_{11}$, $Cor(\mathbf{Z}^{(2)}, \mathbf{V}^{(\mathbf{Z})}) = \mathbf{B}^{(\mathbf{Z})}\boldsymbol{\rho}_{22}$.
- Proporção da variabilidade de cada v.c. explicada pelas variáveis originais $(\lambda_1^2, \lambda_2^2, ...\lambda_k^2)'$.
- Proporção da soma das variâncias das variáveis originais explicada pelas v.c.'s $Cov(\mathbf{Z}^{(1)}) = \mathbf{A}^{(\mathbf{Z})-1}\mathbf{A}'^{(\mathbf{Z})-1}$, $Cov(\mathbf{Z}^{(2)}) = \mathbf{B}^{(\mathbf{Z})-1}\mathbf{B}'^{(\mathbf{Z})-1}$. Com k variáveis canônicas, $\sum_{i=1}^{p} Var(Z_i^{(1)}) = p \cong tr(\mathbf{A}_k^{'(\mathbf{Z})-1}\mathbf{A}_k^{(\mathbf{Z})-1})$ e $\sum_{i=1}^{p} Var(Z_i^{(2)}) = p \cong tr(\mathbf{B}_k^{'(\mathbf{Z})-1}\mathbf{B}_k^{(\mathbf{Z})-1})$.

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

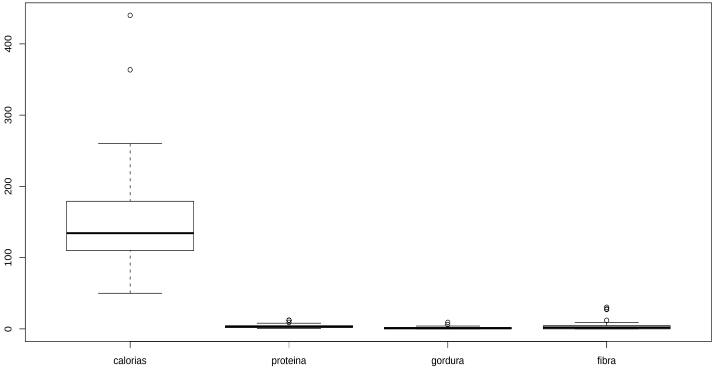
- Medidas descritivas: variáveis nutricionais

Estatística	Cal	Proteina	Gordura	Fibra
Média	149.4	3.69	1.42	3.87
D.P.	62.41	2.64	1.65	6.13
C.V (%)	41.72	71.74	115.80	6.13
Mínimo	50.00	0.76	0.00	0.00
Máximo	440.00	12.12	9.09	30.03

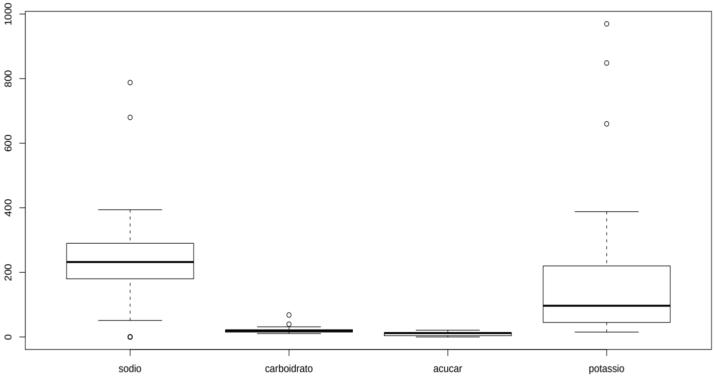
- Medidas descritivas: variáveis elementais

Estatística	**Sódio**	**Carb**	**Açucar**	**Potássio**
Média	237.80	19.97	10.05	159.10
D.P.	130.63	8.46	5.83	180.29
C.V (%)	54.92	42.41	58.05	113.30
Mínimo	0.00	10.53	0.00	15.00
Máximo	787.90	68.00	20.90	969.70

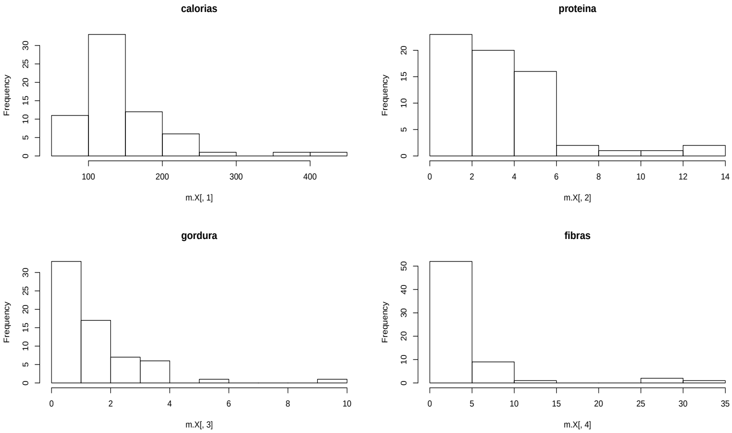
Box plot: variáveis nutricionais



Box plot: variáveis elementais



Histograma: variáveis elementais



Histograma: variáveis elementais

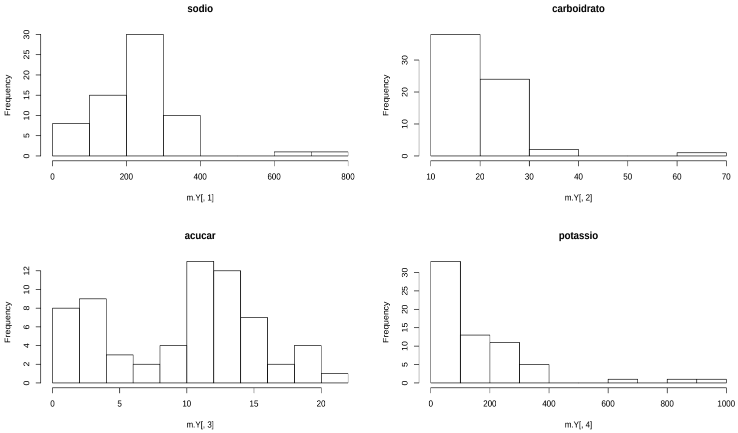
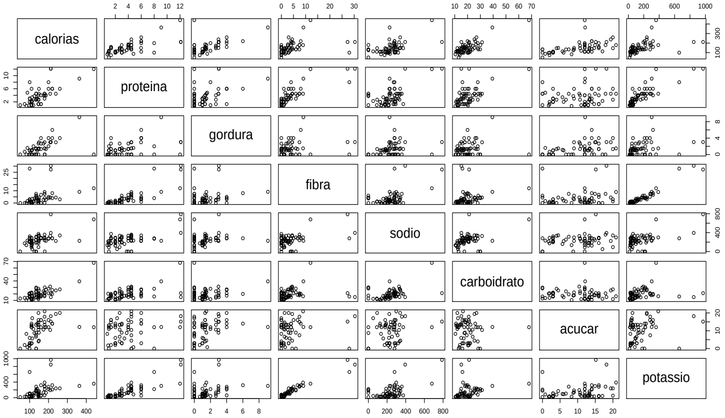


Gráfico de dispersão



- Solução considerando dois pares:
- Correlação canônicas: $\widehat{\lambda}_1 = 0.99$, $\widehat{\lambda}_2 = 0.97$
- Percentual da variabilidade explicada de cada variável canônica $(U_1^{(Z)}, V_1^{(Z)}) = \widehat{\lambda}_1^2 = 0.97$ e $(U_2^{(Z)}, V_2^{(Z)}) = \widehat{\lambda}_2^2 = 0.95$.
- Proporção das somas das variâncias explicadas: $\mathbf{Z}^{(1)} = 0.75$ e $\mathbf{Z}^{(2)} = 0.59$.

- Coeficientes Canônicos:

$$
\begin{aligned}
U_1^{(Z)} &= 1.31(0.77)\text{cal} - 0.23(0.17)\text{prot} - 0.32(0.25)\text{gordura} - 0.47(0.22)\text{fibra} \\
V_1^{(Z)} &= 0.03(0.24)\text{sodio} + 0.89(0.76)\text{carb} + 0.57(0.41)\text{acucar} - 0.52(-0.13)\text{potassio} \\
U_2^{(Z)} &= 0.19(0.59)\text{cal} + 0.15(0.91)\text{prot} - 0.01(0.34)\text{gordura} + 0.77(0.96)\text{fibra} \\
V_2^{(Z)} &= -0.02(0.58)\text{sodio} + 0.17(0.39)\text{carb} - 0.01(0.24)\text{acucar} + 0.96(0.98)\text{potassio}
\end{aligned}
$$

- A variável canônica $U_1^{(Z)}$ é um contraste entre caloria e proteína - gordura - fibra.
- A variável canônica $V_1^{(Z)}$ é um contraste entre açucar e carboidrato - potássio.
- A variável canônica $U_2^{(Z)}$ é uma média ponderada entre caloria-proteína-fibra.
- A variável canônica $V_2^{(Z)}$ é uma média ponderada entre carboidrato e potássio.

- Carboidrato, açucar e potássio influenciam (conjuntamente), no comportamento de todas variáveis nutricionais.
- Carboidrato e potássio influenciam (conjuntamente), no comportamento de quantidade de caloria, proteína e fibra.
- Gordura e açucar parecem ter algum tipo de influência mútua.
- Um modelo de regressão normal linear simples, entre $U_1^{(Z)}$ e $V_1^{(Z)}$ com interceprto nulo, revela que o aumento de uma unidade da variável $V_1^{(Z)}$ imlplica, em média, no aumento de uma unidade na variável $U_1^{(Z)}$. O modelo se ajustou adequadamente aos dados (estrutura linear e suposição de normalidade).

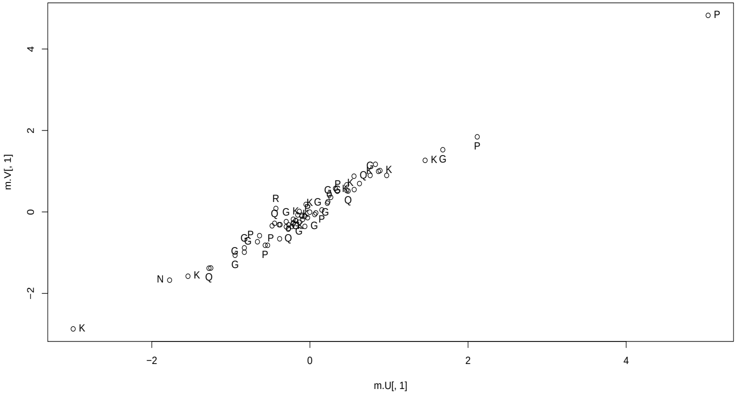
Análise de correlações canônicas)

Análise de correlações canônicas)

Número de variáveis canônicas)

Dados)

- Dipersão entre $U_1^{(Z)}$ e $V_1^{(Z)}$



Análise de correlações canônicas)

Análise de correlações canônicas)

Número de variáveis canônicas)

Dados)

- Dipersão entre $U_1^{(Z)}$ e $V_2^{(Z)}$