Análise de componentes principais: parte 1

Prof. Caio Azevedo

Introdução

- Alguns dos objetivos mais comuns na resoluçao de problemas (multivariados/multidimensionais) são:
  - Reduzir a dimensionalidade (número de variáveis) da matriz de dados.
  - Investigar estruturas multivariadas: dependência/correlação.
  - Identificar padrões de comportamento e interpretações relativas aos problemas levantados.
  - Identificar relações de causalidade e/ou associação.
- A análise de componentes principais (ACP), desenvolvida por Pearson (1901) e, posteriormente, revisitada (e modificada) por Hotteling (1933), pode ser útil em relação aos objetivos mencionados anteriormente. Veja também (artigo).

Introdução

- Objetivo da ACP (veja também aqui): construir variáveis não correlacionadas que retenham a maior parte da estrutura de variabilidade e correlação, a partir da variáveis originais, através de transformações lineares.
- Benefícios:
  - Pode-se usar metodologias de análise univariada.
  - Pode-se trabalhar com um menor número de variáveis.
  - Pode-se obter detalhes do comportamento dos dados os quais são difíceis de serem deduzidos a partir das variáveis originais.

Introdução

- Transformações lineares:
    - São as mais simples de serem construídas.
    - Permitem interpretações, em geral, mais fáceis (modelos não lienares).
    - É possível estabelecer conexões e relacões com os modelos de regressão lineares (MRNLH, modelos mistos, modelos hierárquicos).

Introdução

- Aplicações:
  - Psicometria.
  - Biometria.
  - Econometria.
  - Problemas com poucas observações e muitas variáveis ("large p and small n") (artigo).
  - Seleção de (cov)variáveis (modelos de regressão) (artigo).

Estrutura estatística (conceituação teórica)

- Comecemos com um único vetor aleatório (sem a matriz de dados).
- $\boldsymbol{X} = (X_1, ..., X_p)' \sim D_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, que $D_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ representa uma distribuição p-variada (em princípio qualquer) com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$.
- A matriz de correlações será denotada, como antes, por $\boldsymbol{\rho}$.
- Pode-se considerar, em princípio, qualquer tipo de correlação embora, em princípio, utilizaremos a correlação de Pearson.

Cont.

- Objetivo, encontrar $\boldsymbol{A}_{(p \times p)}$ (não estocástica), a fim de obter $\boldsymbol{Y} = \boldsymbol{AX} = (Y_1, ..., Y_p)'$, de sorte que $\mathcal{V}(Y_1) \geq \mathcal{V}(Y_2) \geq ... \geq \mathcal{V}(Y_p)$ e que $Corre(Y_i, Y_j) = 0 \ \forall i \neq j, i, j = 1, ..., p$, e que $\mathcal{V}(Y_i)$ seja máximo $i = 1, 2, .., p$.
- Seja $\boldsymbol{a}_i'$ a i-ésima linha da matriz $\boldsymbol{A}$. Temos que

$$
\begin{aligned}
Y_1 &= \boldsymbol{a}_1'\boldsymbol{X} = a_{11}X_1 + a_{12}X_2 + ... + a_{1p}X_p \\
Y_2 &= \boldsymbol{a}_2'\boldsymbol{X} = a_{21}X_1 + a_{22}X_2 + ... + a_{2p}X_p \\
\vdots &= \vdots \qquad \vdots \\
Y_p &= \boldsymbol{a}_p'\boldsymbol{X} = a_{p1}X_1 + a_{p2}X_2 + ... + a_{pp}X_p
\end{aligned}
$$

Cont.

Assim

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} = \begin{bmatrix} \boldsymbol{a}_1' \\ \boldsymbol{a}_2' \\ \vdots \\ \boldsymbol{a}_p' \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pp} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

- Portanto $\mathcal{V}(Y_i) = \boldsymbol{a}_i'\boldsymbol{\Sigma}\boldsymbol{a}_i$ e $Cov(Y_i, Y_j) = \boldsymbol{a}_i'\boldsymbol{\Sigma}\boldsymbol{a}_j$, $\forall i \neq j$.
- Claramente $\mathcal{V}(Y_i)$ pode ser aumentada de modo ilimitado multiplicando-se cada vetor $\boldsymbol{a}_i$ por uma constante positiva (o que não afetaria o fato de que $\boldsymbol{a}_i'\boldsymbol{\Sigma}\boldsymbol{a}_j = 0$). Assim, iremos restringir nossa atenção a vetores com norma unitária ($\boldsymbol{a}_i'\boldsymbol{a}_i = 1$).

Procedimento

- Primeira componente principal: combinação linear $\boldsymbol{a}_1'\boldsymbol{X}$ que maximiza $\mathcal{V}(\boldsymbol{a}_1'\boldsymbol{X})$ sujeito à $\boldsymbol{a}_1'\boldsymbol{a}_1 = 1$.
- Segunda componente principal: combinação linear $\boldsymbol{a}_2'\boldsymbol{X}$ que maximiza $\mathcal{V}(\boldsymbol{a}_2'\boldsymbol{X})$ sujeito à $\boldsymbol{a}_2'\boldsymbol{a}_2 = 1$ e $Cov(\boldsymbol{a}_1'\boldsymbol{X}, \boldsymbol{a}_2'\boldsymbol{X}) = 0$.

$$\vdots$$

- i-ésima componente principal: combinação linear $\boldsymbol{a}_i'\boldsymbol{X}$ que maximiza $\mathcal{V}(\boldsymbol{a}_i'\boldsymbol{X})$ sujeito à $\boldsymbol{a}_i'\boldsymbol{a}_i = 1$ e $Cov(\boldsymbol{a}_i'\boldsymbol{X}, \boldsymbol{a}_k'\boldsymbol{X}) = 0$, para $k < i$.

Resultado (ver Johnson & Wichern)

- Sejam $(\lambda_1, \boldsymbol{e}_1)$, $(\lambda_2, \boldsymbol{e}_2)$,...,$(\lambda_p, \boldsymbol{e}_p)$, em que $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$ são os autovalores associados à $\boldsymbol{\Sigma}$, e $\boldsymbol{e}_i$ os respectivos autovetores ortonormalizados, ou seja $\boldsymbol{e}_i'\boldsymbol{e}_i = 1$ e $\boldsymbol{e}_i'\boldsymbol{e}_j = 0$, $\forall i \neq j$, $i, j = 1, 2, ..., p$.
- A i-ésima componente principal é dada por

$$\boldsymbol{Y}_i = \boldsymbol{e}_i'\boldsymbol{X}, i = 1, 2, ..., p. \quad (1)$$

- Assim, temos que

$$\mathcal{V}(Y_i) = \boldsymbol{e}_i'\boldsymbol{\Sigma}\boldsymbol{e}_i = \lambda_i, i = 1, 2, ..., p \quad (2)$$

$$Cov(Y_i, Y_j) = \boldsymbol{e}_i'\boldsymbol{\Sigma}\boldsymbol{e}_j = 0, \forall i \neq j, i, j = 1, 2, ..., p \quad (3)$$

Prova

- Com relação à (2) e (3), seja $\boldsymbol{E}$ uma matriz em que as linhas são os autovetores $\boldsymbol{e}_1, \boldsymbol{e}_2, .., \boldsymbol{e}_p$ (definidos anteriormente), ou seja (ortonormalizados)

$$\boldsymbol{E} = \begin{bmatrix} e_{11} & e_{12} & \dots & e_{1p} \\ e_{21} & e_{22} & \dots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{p1} & e_{p2} & \dots & e_{pp} \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}_1' \\ \boldsymbol{e}_2' \\ \vdots \\ \boldsymbol{e}_p' \end{bmatrix}$$

- Vimos, anteriormente, que $\boldsymbol{\Sigma}$ for uma matriz positiva definida então $\boldsymbol{\Sigma} = \boldsymbol{E}'\boldsymbol{\Lambda}\boldsymbol{E}$, em que $\boldsymbol{\Lambda} = diag(\lambda_1, ...\lambda_p)$ (autovalores), e as colunas da matriz $\boldsymbol{E}'$ são formadas pelos respectivos autovetores orto-normalizados.

- Assim, temos que

$$Cov(\boldsymbol{Y}) = Cov(\boldsymbol{EX}) = \boldsymbol{E}\,Cov(\boldsymbol{X})\boldsymbol{E}' = \boldsymbol{E}\boldsymbol{\Sigma}\boldsymbol{E}' = \boldsymbol{EE}'\boldsymbol{\Lambda}\boldsymbol{EE}' = \boldsymbol{\Lambda}$$

pois,

$$\begin{aligned}
\boldsymbol{EE}' &= \begin{bmatrix} \boldsymbol{e}_1' \\ \boldsymbol{e}_2' \\ \vdots \\ \boldsymbol{e}_p' \end{bmatrix} \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \ldots & \boldsymbol{e}_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}_1'\boldsymbol{e}_1 & \boldsymbol{e}_1'\boldsymbol{e}_2 & \ldots & \boldsymbol{e}_1'\boldsymbol{e}_p \\ \boldsymbol{e}_2'\boldsymbol{e}_1 & \boldsymbol{e}_2'\boldsymbol{e}_2 & \ldots & \boldsymbol{e}_2'\boldsymbol{e}_p \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{e}_p'\boldsymbol{e}_1 & \boldsymbol{e}_p'\boldsymbol{e}_2 & \ldots & \boldsymbol{e}_p'\boldsymbol{e}_p \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix} = \boldsymbol{I}_p
\end{aligned}$$

Prova

- A prova de (1), vem do que fato de que $max_{\boldsymbol{a}\neq\boldsymbol{0}} \frac{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{a}}$ é obtido quando $\boldsymbol{a} = \boldsymbol{e}_1$ e o $min_{\boldsymbol{a}\neq\boldsymbol{0}} \frac{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{a}}$ é obtido quando $\boldsymbol{a} = \boldsymbol{e}_p$ (veja página 80, do livro do Johnson and Wichern (2007), 7^a edição). Como $\boldsymbol{e}_1'\boldsymbol{e}_1 = 1$, temos que $\max_{\boldsymbol{a}\neq\boldsymbol{0}} \frac{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{a}} = \max_{\boldsymbol{a}\neq\boldsymbol{0}} \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}$ e o mesmo vale para o mínimo.
- Similarmente, $\max_{\boldsymbol{a}\neq\boldsymbol{0}\perp\boldsymbol{e}_1,\boldsymbol{e}_2,..,\boldsymbol{e}_k} \frac{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{a}}$ é obtido quando $\boldsymbol{a} = \boldsymbol{e}_{k+1}$.

Componentes principais e variáveis originais

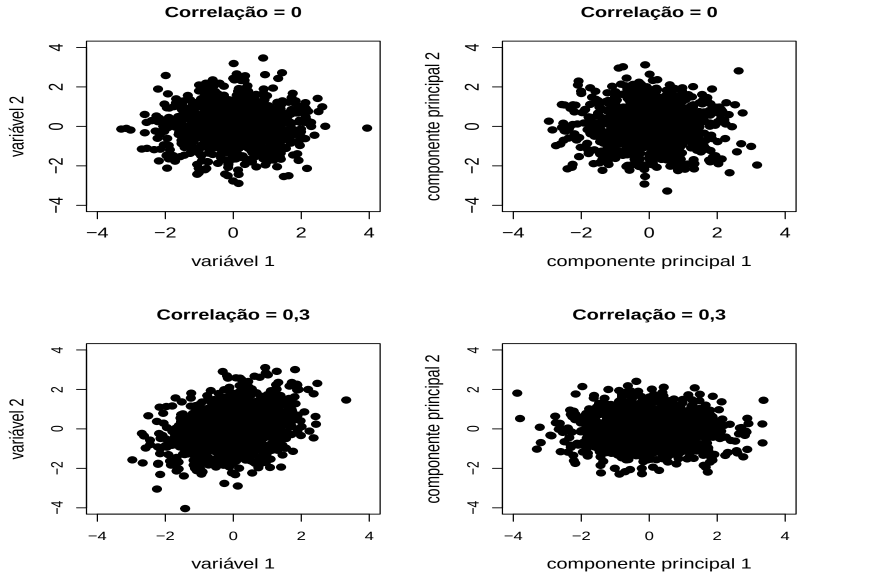
- Defina $\boldsymbol{a}_j = [0\ 0 \ldots \underbrace{1}_{\text{posição } j} \ldots 0\ 0]$. Veja também aqui e aqui.
- Temos que $Cov(Y_i, X_j) = Cov(\boldsymbol{e}_i'\boldsymbol{X}, \boldsymbol{a}_j'\boldsymbol{X}) = \boldsymbol{e}_i' Cov(\boldsymbol{X}, \boldsymbol{X})\boldsymbol{a}_j = \boldsymbol{e}_i'\boldsymbol{\Sigma}\boldsymbol{a}_j = \lambda_i \boldsymbol{e}_i'\boldsymbol{a}_j = \lambda_i e_{ij}$. (Exercício: provar que $\boldsymbol{e}_i'\boldsymbol{\Sigma} = \lambda_i \boldsymbol{e}_i'$).
- Assim $Corre(Y_i, X_j) = \frac{Cov(Y_i, X_j)}{\sqrt{\mathcal{V}(Y_i)}\sqrt{\mathcal{V}(X_j)}} = \frac{\lambda_i e_{ij}}{\sqrt{\lambda_i}\sigma_j} = \frac{\sqrt{\lambda_i} e_{ij}}{\sigma_j}$, em que $\sigma_j = DP(X_j)$.
- Pode-se mostrar que $tr(\boldsymbol{\Sigma}) = \sum_{i=1}^{p} \mathcal{V}(X_i) = \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} \mathcal{V}(Y_i)$

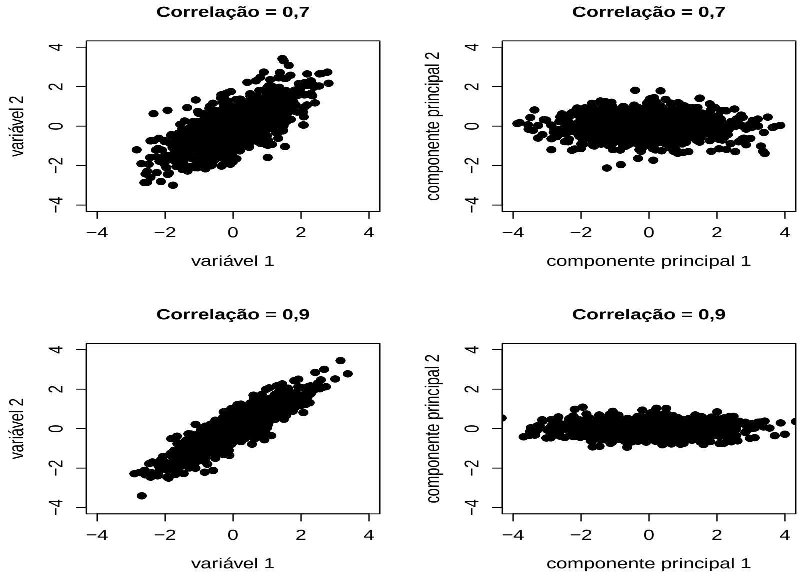
Comentários

- Há problemas em se trabalhar com a matriz de covariâncias ($\boldsymbol{\Sigma}$). As componentes principais tendem a ser (muito) influenciadas pela variabilidade (escala) das variáveis.
- Quanto maior a variabilidade de uma determinada variável, mais ele influenciará as componentes.
- Alternativa: trabalhar com a matriz de correlações ($\boldsymbol{\rho}$) (o que é equivalente a trabalhar com variáveis com variância unitária).
- Seja $\boldsymbol{Z} = (Z_1, ..., Z_p)'$, $Z_i = \frac{X_i - \mu_i}{\sigma_i}$. Então $Cov(\boldsymbol{Z}) = Corre(\boldsymbol{Z}) = Corre(\boldsymbol{X}) = \boldsymbol{\rho}$.
- Nesse caso as variáveis, portanto as componentes principais, serão adimensionais.

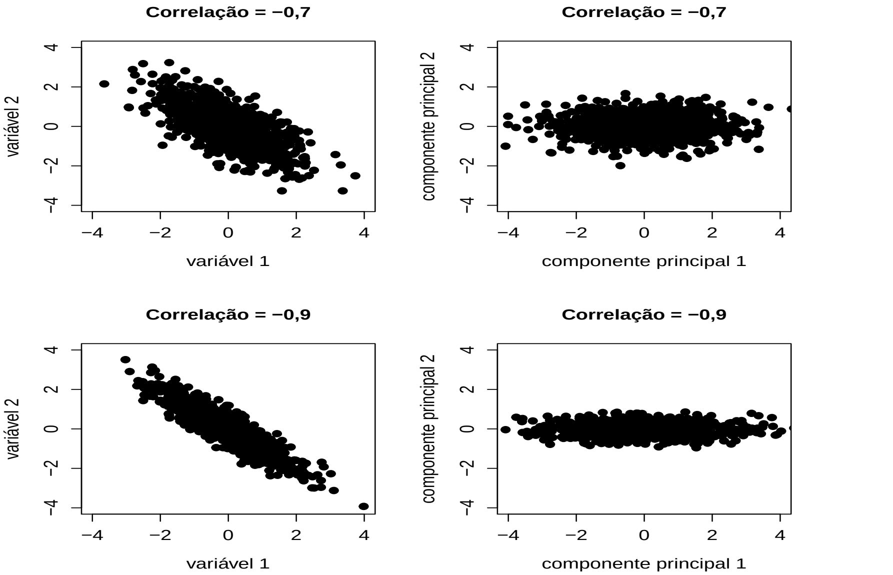
Comentários

- O procedimento de obtenção das componentes principais continua o mesmo mas, nesse caso, trabalha-se com a matriz de correlações.
- Nesse caso, temos que $Cov(Y_i, Z_j) = Cov(\boldsymbol{e}_i'\boldsymbol{Z}, \boldsymbol{a}_j'\boldsymbol{Z}) = \boldsymbol{e}_i' Cov(\boldsymbol{Z}, \boldsymbol{Z})\boldsymbol{a}_j = \boldsymbol{e}_i'\boldsymbol{\rho}\boldsymbol{a}_j = \lambda_i \boldsymbol{e}_i'\boldsymbol{a}_j = \lambda_i e_{ij}$. (Exercício: provar que $\boldsymbol{e}_i'\boldsymbol{\rho} = \lambda_i \boldsymbol{e}_i'$).
- Além disso, $Corre(Y_i, Z_j) = \frac{Cov(Y_i, Z_j)}{\sqrt{\mathcal{V}(Y_i)}\sqrt{\mathcal{V}(Z_j)}} = \frac{\lambda_i e_{ij}}{\sqrt{\lambda_i}1} = \sqrt{\lambda_i} e_{ij}$.
- Como determinar o número de componentes principais e como utilizá-las?
- Nos próximos três slides, discutiremos, através de dados simulados, alguns aspectos relevantes.


Correlação = 0
variável 2
variável 1
Correlação = 0
componente principal 2
componente principal 1
Correlação = 0,3
variável 2
variável 1
Correlação = 0,3
componente principal 2
componente principal 1
−4
−2
0
2
4



Correlação = 0,7
variável 2
variável 1
Correlação = 0,7
componente principal 2
componente principal 1
Correlação = 0,9
variável 2
variável 1
Correlação = 0,9
componente principal 2
componente principal 1



Correlação = −0,7
variável 1
variável 2
Correlação = −0,7
componente principal 1
componente principal 2
Correlação = −0,9
variável 1
variável 2
Correlação = −0,9
componente principal 1
componente principal 2
−4
−2
0
2
4

Estimação das componentes principais

- Dada uma matriz de dados $\boldsymbol{X}_{(n \times p)}$

Indivíduo	Variável 1	Variável 2	...	Variável p
1	X_{11}	X_{12}	...	X_{1p}
2	X_{21}	X_{22}	...	X_{2p}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	X_{n1}	X_{n2}	...	X_{np}

estima-se a matriz de covariâncias ($\widetilde{\boldsymbol{\Sigma}}$) ou a matriz de correlações amostrais ($\widetilde{\boldsymbol{\rho}}$), conforme visto anteriormente. Ou seja $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{S}^2$ (aqui, slide 10), e $\widehat{\boldsymbol{\rho}} = \widehat{\boldsymbol{D}}^{-1}\boldsymbol{S}\widehat{\boldsymbol{D}}^{-1}$ (aqui slide 11).

Estimação das componentes principais

- Obtem-se os auto-valores e auto-vetores ortonormalizados, digamos $(\widetilde{\lambda}_1, \widetilde{\boldsymbol{e}}_1)$, $(\widetilde{\lambda}_2, \widetilde{\boldsymbol{e}}_2)$,...,$(\widetilde{\lambda}_p, \widetilde{\boldsymbol{e}}_p)$ a partir de $\widetilde{\boldsymbol{\Sigma}}$ ou $\widetilde{\boldsymbol{\rho}}$ (por exemplo, usando a função "eigen" do R).
- Assim $\widetilde{\mathcal{V}}(Y_{ij}) = \widetilde{\lambda}_i$.
- Cada componente principal é calculada como $y_{ij} = \widetilde{\boldsymbol{e}}_j \boldsymbol{x}_i$, em que $\boldsymbol{x}_i, i = 1, .., n, j = 1, ..., p$ é a i-ésima linha da matriz de dados observada.
- Alternativamente, podemos estimar as componentes através de $y_{ij} = \widetilde{\boldsymbol{e}}_j \boldsymbol{z}_i$, em que $\boldsymbol{z}_i$ corresponde às variáveis padronizadas (default da função "princomp" do R).

Exemplo 4: índices de criminalidade de estados americanos

- Número de prisões efetuadas (por 100.000 habitantes) em 50 estados americanos em 1973.
- Variáveis: assalto, violência sexual (VS), assassinato e porcentagem de moradores na área urbana (PMAU).
- Digitar *USArrests* no programa *R*.

Exemplo 4: banco (matriz) de dados

Estado	Assassinato	Assalto	PMAU	VS
Alabama	13,2	236,0	58	21,2
Alaska	10,0	263,0	48	44,5
Arizona	8,1	294,0	80	31,0
⋮	⋮	⋮	⋮	⋮
Wyoming	6,8	161	60	15,6

Medidas resumo

Estatística	**Assassinato**	**Assalto**	**PMAU**	**VS**
Média	7,79	170,76	65,54	21,23
Variância	18,97	**6.945,17**	209,52	87,73
DP	4,36	83,34	14,47	9,37
CV(%)	55,93	48,80	22,09	44,11
Mínimo	0,80	45,00	32,00	7,30
Mediana	7,25	159,00	66,00	20,10
Máximo	17,40	337,00	91,00	46,00
Assimetria	0,37	0,22	-0,21	0,75
Curtose	-0,95	-1,15	-0,87	0,08

Diagrama de dispersão múltiplo

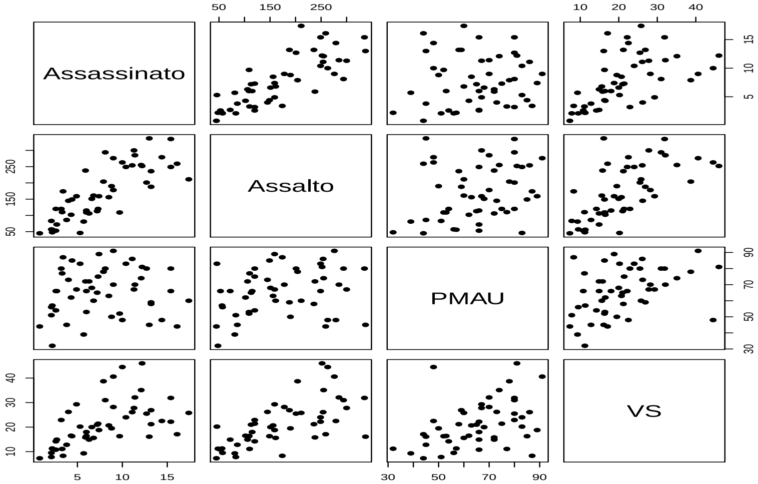
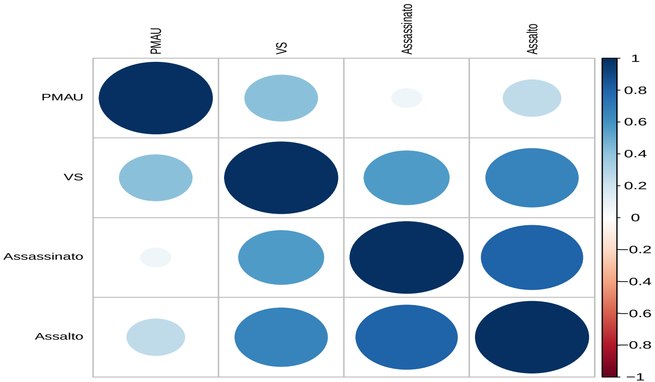
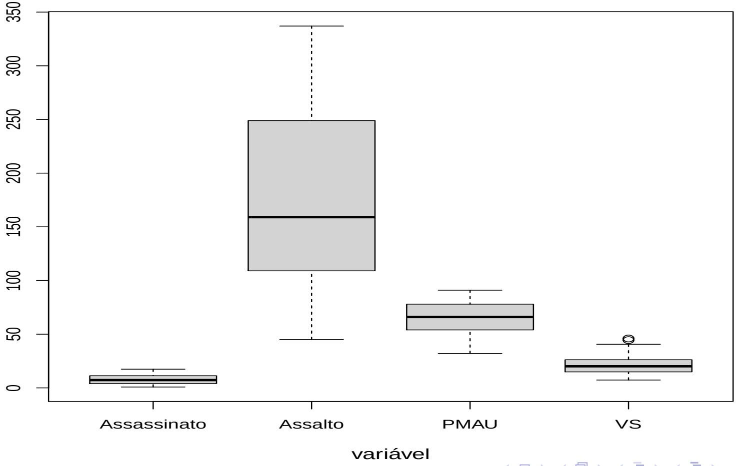


Gráfico de correlações

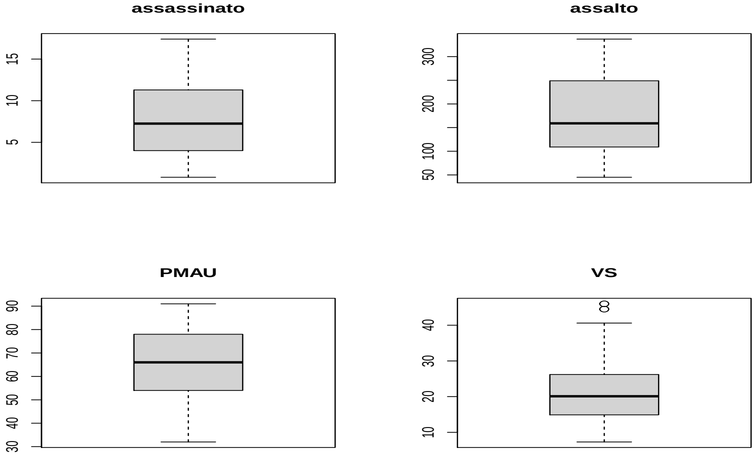


Box-plot

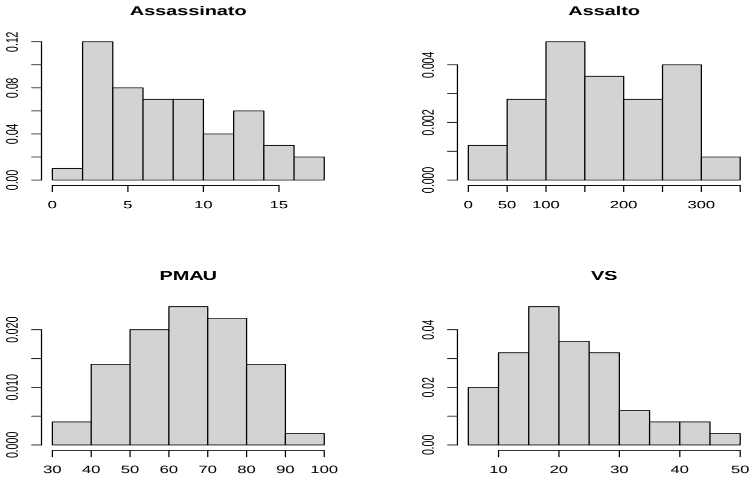


0
50
100
150
200
250
300
350
Assassinato
Assalto
PMAU
VS
variável

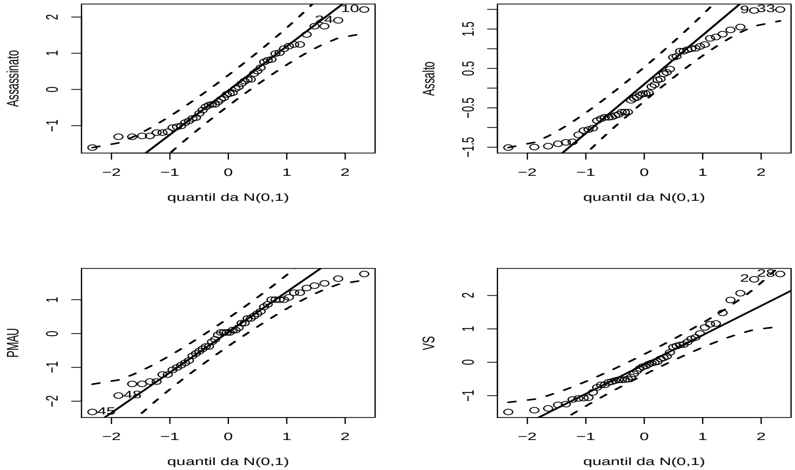
Box-plot: por variável



Histogramas



QQplots


Assassinato
quantil da N(0,1)
Assalto
quantil da N(0,1)
PMAU
quantil da N(0,1)
VS
quantil da N(0,1)

Matriz de covariâncias amostrais

- Matriz de covariâncias:

	Assassinato	**Assalto**	**PMAU**	**VS**
Assassinato	18,97	291,06	4,38	22,99
Assalto	291,06	**6945,16**	312,27	519,26
PMAU	4,38	312,27	209,51	55,76
VS	22,99	519,26	55,76	87,72

Matriz de covariâncias amostrais

- Autovalores

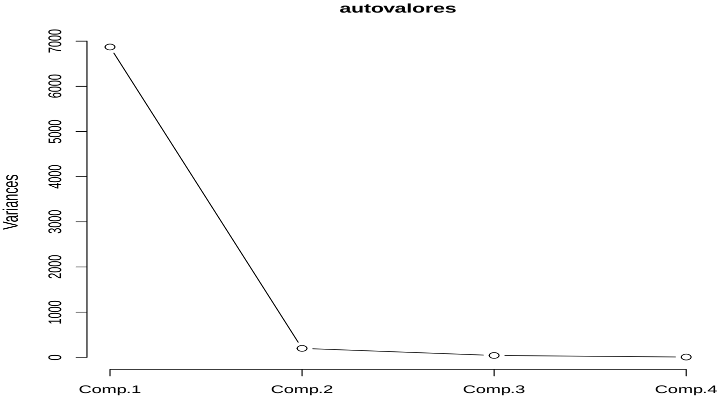
  7011,11 201,99 42,11 6,16

- Autovetores

-0,04170432	0,04482166	0,07989066	0,99492173
-0,99522128	0,05876003	-0,06756974	-0,03893830
-0,04633575	-0,97685748	-0,20054629	0,05816914
-0,07515550	-0,20071807	0,97408059	-0,07232502

0,04170432	-0,04482166	-0,07989066	-0,99492173
0,99522128	-0,05876003	0,06756974	0,03893830
0,04633575	0,97685748	0,20054629	-0,05816914
0,07515550	0,20071807	-0,97408059	0,07232502

Scree-plot: Matriz de covariâncias amostrais



- Variância explicada

	Comp. 1	**Comp. 2**	**Comp. 3**	**Comp. 4**
PVE (%)	96,55	2,78	< 0.01	< 0.01
PVEA (%)	96,55	99,33	99,91	100,00

- Componentes principais

	Comp. 1	**Comp. 2**
Assassinato	0,04 (0,80)	-0,04 (-0,14)
Assalto	0,99 (0,99)	-0,06 ($<$-0,01)
PMAU	0,04 (0,26)	0,98 (0,96)
VS	0,07 (0,67)	0,20 (0,30)

- Apenas duas variáveis são retidas
- Solução: Utilizar a matriz de correlações.

Matriz de correlações

- Matriz de correlações:

	Assassinato	**Assalto**	**PMAU**	**VS**
Assassinato	1,00	0,80	0,07	0,56
Assalto	0,80	1,00	0,25	0,66
PMAU	0,07	0,25	1,00	0,41
VS	0,56	0,66	0,41	1,00

Matriz de correlações

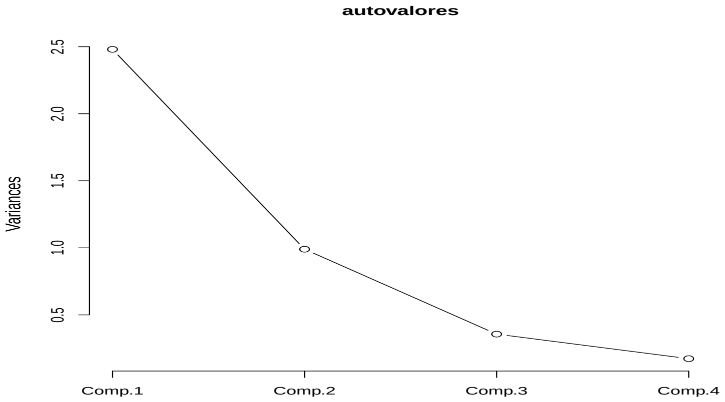
- Autovalores

  2,48 0,98 0,35 0,17

- Autovetores

0,5358995	-0,4181809	0,3412327	-0,64922780
0,5831836	-0,1879856	0,2681484	0,74340748
0,2781909	0,8728062	0,3780158	-0,13387773
0,5434321	0,1673186	-0,8177779	-0,08902432

Matriz de correlações



Matriz de correlações

- Variância explicada

	Comp. 1	Comp. 2	Comp. 3	Comp. 4
PVE (%)	62,00	24,74	8,91	4,32
PVEA (%)	62,00	86,75	95,67	100,00

- Componentes principais

	Comp. 1 (Y_1)	Comp. 2 (Y_2)
Assassinato (Z_1)	0,53 (0,84)	-0,41 (-0,41)
Assalto (Z_2)	0,58 (0,91)	-0,18 (-0,19)
PMAU (Z_3)	0,27 (0,43)	0,86 (0,87)
VS(Z_4)	0,54 (0,85)	0,16 (0,17)

Equações das componentes principais

- Temos que ($j = 1, 2, ..., 50,$ estados)

$$
\begin{aligned}
Y_{j1} &= 0,53Z_1 + 0,58Z_2 + 0,27Z_3 + 0,54Z_4 \\
Y_{j2} &= -0,41Z_1 - 0,18Z_2 + 0,86Z_3 + 0,16Z_4
\end{aligned}
$$

Matriz de correlações/cmonandos no R

- Primeira componente: "escore ponderado" entre as variáveis, com maiores pesos para as variáveis criminais.
- Segunda componente: é um contraste (diferença) entre PMAU e VS e as outras variáveis criminais, com maior peso para a variável PMAU.
- Função "princomp". Resultado salvo num objeto chamado result.

Comando	**Significado**
result$loadings	coeficientes que geram as CP's (geralmente os autovetores negativos)
result$scores	estimativa das componentes principais
screeplot(result)	screeplot

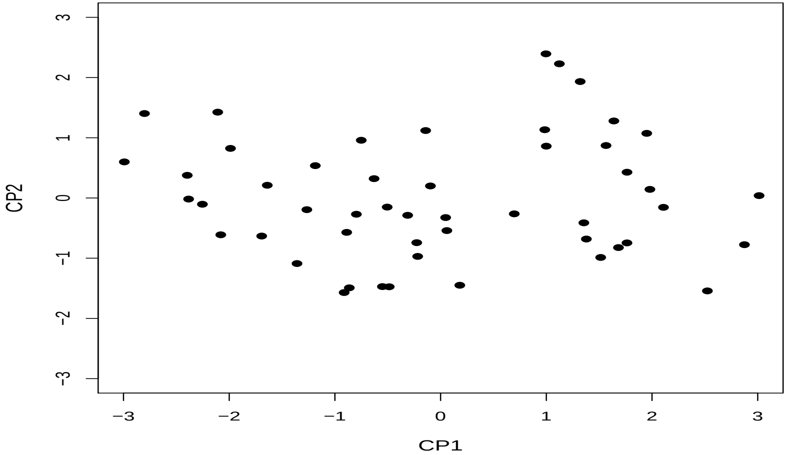
Interpretações

- Valores positivos para Y_1 e Y_2:
  - Valores acima da média para (todas) as variáveis (com maior magnitude para PMAU e VS).
  - Valores acima da média para PMAU e VS e valores abaixo da média para Assassinato e Assalto (com maior magnitude para PMAU e VS).
- Valores negativos para Y_1 e Y_2:
  - Valores abaixo da média para (todas) as variáveis (com menor magnitude para PMAU e VS).
  - Valores abaixo da média para PMAU e VS e valores acima da média para Assassinato e Assalto (com menor magnitude para PMAU e VS).

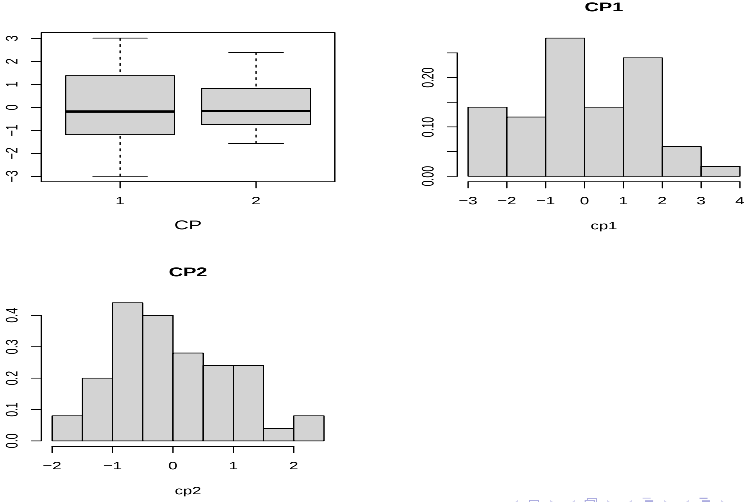
Interpretações

- Valores positivos para Y_1 e negativos Y_2:
  - Valores acima da média para (todas) as variáveis (com menor magnitude para PMAU e VS).
  - Valores abaixo da média para PMAU e VS e valores acima da média para Assassinato e Assalto (com maior magnitude para Assassinato e Assalto).
- Valores negativos para Y_1 e positivos Y_2:
  - Valores abaixo da média para (todas) as variáveis (com menor magnitude para PMAU e VS).
  - Valores acima da média para PMAU e VS e valores abaixo da média para Assassinato e Assalto (com menor magnitude para Assassinato e Assalto).

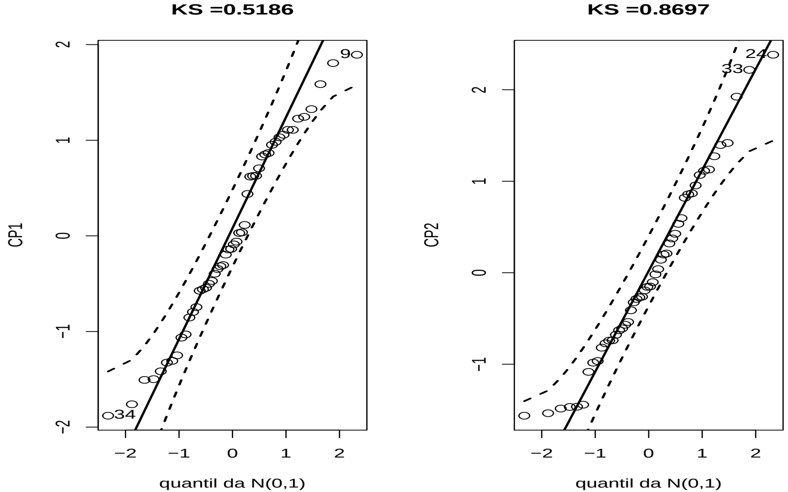
Dispersão entre a primeira e a segunda componentes



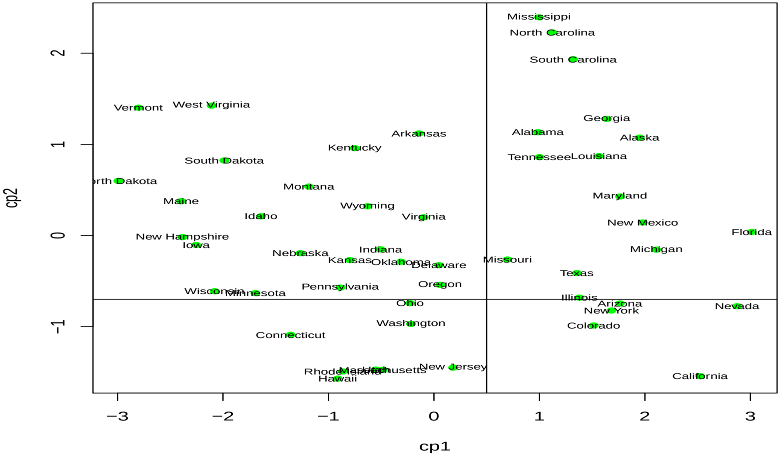
Box-plot e histograma das duas primeiras componentes



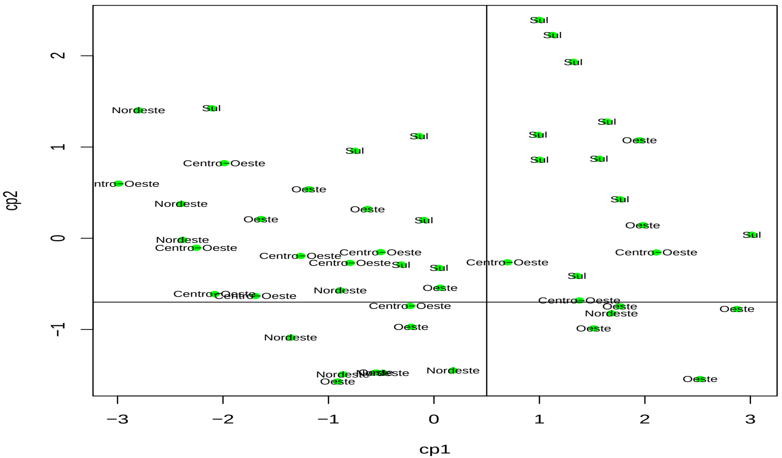
QQ plots com testes de KS



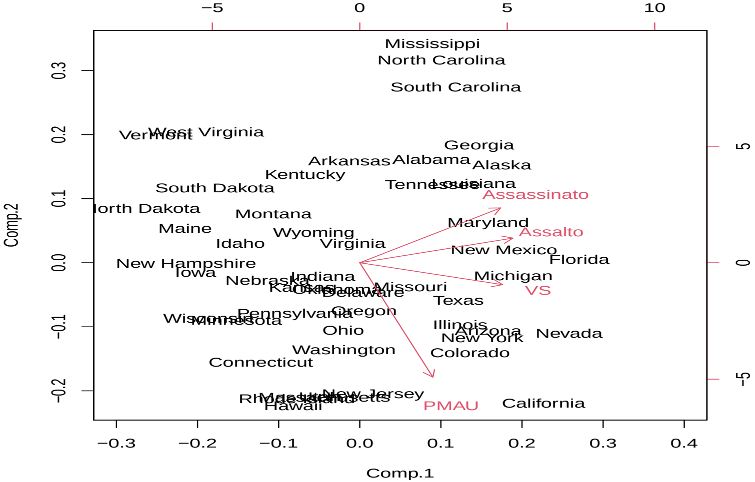
Dispersão entre as componentes por estado



Dispersão entre as componentes por região



Biplot (duas componentes)



Biplot

- Um gráfico com quatro eixos.
- Os pontos representam as observações enquanto que os vetores (setas) representam as variáveis.
- Os eixos da parte de baixo e da esquerda se referem aos valores das duas componentes principais divididos pelos respectivos desvios-padrão (das componentes) vezes a raiz quadrada do tamanho da amostra ($y^*_{ij} = \frac{y_{ij}}{\sqrt{\lambda_i n}}$).
- Os eixos das partes de cima e da direita representam os valores dos coeficientes das componentes multiplicados pelos respectivos desvios-padrão (das componentes) vezes a raiz quadrada do tamanho da amostra ($e^*_{ij} = e_{ij}\sqrt{\lambda_i n}$).

Biplot: algumas interpretações

- As variáveis Assassinato, Assalto e VS apresentam maior correlação entre si do que com a variável PMAU.
- Estados como Montana, Virgina, Wyoming etc possuem menor PMAU. Enquanto que California, Colorado, Illinois etc apresentam maior.
- Estados que apresentam menores índices (em geral) de criminalidade: Idaho, Indiana, Nesbraska, New Hampshire etc.
- Estados que apresentam menores índices de criminalidade (para pelo menos um tipo de violência): New Mexico, Michigan, Florida, Maryland, Texas etc.