Universidade de Brasília

DEPARTAMENTO DE ESTATÍSTICA

27 junho 2023

Prova 2

Prof. Dr. George von Borries

Análise Multivariada 1

Aluno: Bruno Gondim Toledo | Matrícula: 15/0167636

Página 01/11

Lista 6 - Análise fatorial exploratória

Questão 44

Ex. 9.2 | Johnson & Wichern

Use the information in Exercise 9.1.

(a) Calculate communalities $h_i^2, i = 1, 2, 3$ and interpret these quantities.

(b) Calculate $Corr(Z_i, F_1)$ for $i = 1, 2, 3$. Which variable might carry the greatest weight in "naming" the common factor? Why?

Soluções:

a) As comunalidades são dadas por: $\sum_{j=1}^{n} l_{ij}^2 = h_i^2$. Para a matriz $\mathbf{L} = [.9 \quad .7 \quad .5]$, temos que as comunalidades são:

$h_1^2 = .81$
$h_2^2 = .49$
$h_3^2 = .25.$

Como as comunalidades são quantidades de variâncias de cada variável explicada pelos fatores, quanto maior for a comunalidade, maior será o poder de explicação daquela variável pelo fator. A comunalidade h_i^2 assume valores no intervalo [0,1]. Desejamos, em geral, valores acima de 0.5. Neste caso, temos que $h_1^2 > 0.5$, enquanto $h_2^2, h_3^2 < 0.5$. Entretanto $h_2^2 \approx 0.5$, temos que h_2^2 também pode ser utilizada.

b) Como $Cov(\mathbf{X}, \mathbf{F} = \mathbf{L})$, e $Cov(\mathbf{X_i}, \mathbf{F_j}) = \ell ij$ (Resultado 2., pag. 484 J&W) [1], e $Cov(\mathbf{X_1}, \mathbf{F_1}) = Corr(\mathbf{X_1}, \mathbf{F_1})$ (pag. 486 J&W) [1] sabemos que $Cor(\mathbf{Z_i}, \mathbf{F_1})$, para $i = 1, 2, 3$ será $\ell_{i1} = [.9 \quad .7 \quad .5] = \mathbf{L}$. Isso indica que a variável Z_1 carrega a maior carga fatorial, dado seu maior valor absoluto (comparando também com o último resultado encontrado sobre comunalidade).

Questão 53

Ex. 9.22 | Johnson & Wichern

a) Escore dos fatores para m = 2 para:

Método regressivo:

```
##          Factor1     Factor2
##  [1,] -0.6874886 -0.67456264
##  [2,] -1.2235869 -0.82550084
##  [3,] -0.8225668  0.03634364
##  [4,]  0.4166648 -0.27745404
##  [5,] -0.2239543  0.49607112
##  [6,] -0.4199121 -0.63308417
##  [7,] -0.6094400 -0.21108559
##  [8,]  1.9036734  0.96272772
##  [9,] -0.2158057  0.50416779
## [10,]  0.8004769  0.55881769
```

Método mínimos quadrados ponderados:

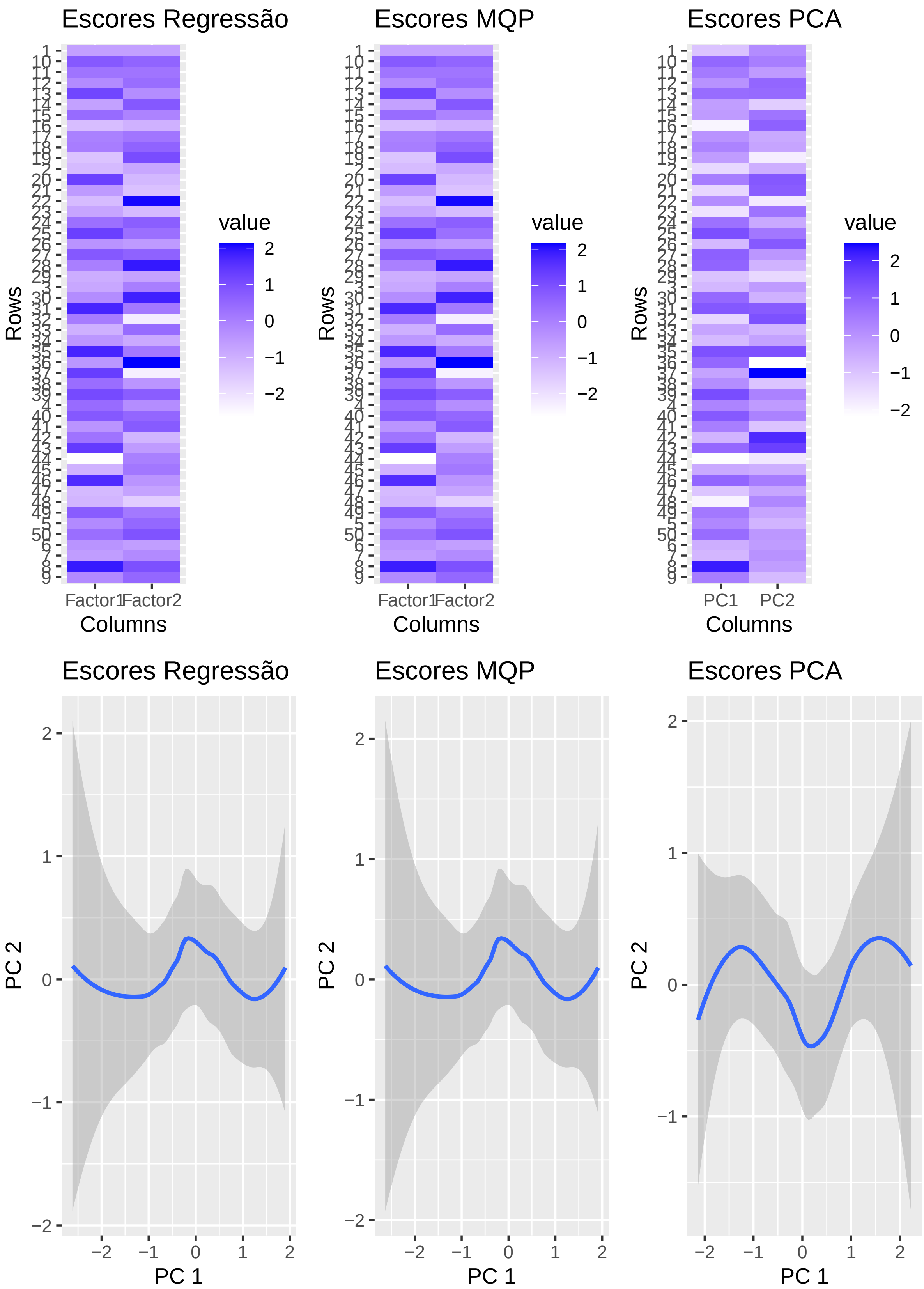
```
##          Factor1     Factor2
##  [1,] -0.6905620 -0.68971146
##  [2,] -1.2290570 -0.84403931
##  [3,] -0.8262441  0.03715982
##  [4,]  0.4185275 -0.28368489
##  [5,] -0.2249555  0.50721151
##  [6,] -0.4217893 -0.64730149
##  [7,] -0.6121645 -0.21582599
##  [8,]  1.9121838  0.98434793
##  [9,] -0.2167704  0.51549001
## [10,]  0.8040555  0.57136719
```

b) Encontrar os escores dos fatores pelo método de análise de componente principal:

```
##              PC1         PC2
##  [1,] -0.9857690  0.14148631
##  [2,] -1.3981720 -0.54988531
##  [3,] -0.7355783 -0.16795544
##  [4,]  0.2960623 -0.17622658
##  [5,]  0.2352160 -0.67722122
##  [6,] -0.5617045 -0.18077413
##  [7,] -0.7212336  0.01658193
##  [8,]  2.2230288 -0.20267945
##  [9,]  0.4323279 -0.79740021
## [10,]  0.8878426  0.42719694
```

c) Comparar os três métodos:

Visualizando graficamente os escores:



Pela visualização dos gráficos, podemos perceber que os escores do método regressivo e de mínimos quadrados ponderados se assemelham bastante, enquanto os escores da PCA diferem um pouco desses últimos dois, mas ainda se mantendo numa escala parecida.

Lista 7 - Normal Multivariada

Questão 55

```
##            [,1]       [,2]
##  [1,] 2.357074  2.0946752
##  [2,] 2.337701  1.4864535
##  [3,] 3.104502  3.5643835
##  [4,] 2.089002  4.1984057
##  [5,] 3.231104  1.7970795
##  [6,] 3.612742  0.6433665
##  [7,] 3.767353 -2.0199445
##  [8,] 3.405614  1.2193625
##  [9,] 2.856354  3.1555661
## [10,] 4.784329 -0.3236111
```

Outros algorítmos...

```
##            [,1]      [,2]
##  [1,] 2.800896 3.3898442
##  [2,] 2.678677 2.7801475
##  [3,] 1.624445 3.3665972
##  [4,] 4.083305 1.8620278
##  [5,] 3.430604 1.8308119
##  [6,] 4.172041 0.4064622
##  [7,] 4.806869 1.5691417
##  [8,] 4.144648 0.5365565
##  [9,] 3.509041 0.7210338
## [10,] 4.151782 1.7644548
```

```
##             [,1]       [,2]
##  [1,] 2.7320424  2.9048127
##  [2,] 4.7926536 -0.1361930
##  [3,] 0.8974694  6.9849363
##  [4,] 2.5421927  3.3045247
##  [5,] 3.5942901  0.2395318
##  [6,] 1.0528061  5.7303391
##  [7,] 2.2290414  2.1305702
##  [8,] 1.7660094  5.8904050
##  [9,] 1.9641529  5.5431037
## [10,] 3.6063348 -0.5278945
```

```
##            [,1]      [,2]
##  [1,] 3.498224 1.2931100
##  [2,] 4.435514 0.4272262
##  [3,] 4.548098 0.8916664
##  [4,] 2.073462 3.6488088
##  [5,] 1.915099 3.4380123
##  [6,] 1.439808 4.6273818
##  [7,] 2.808718 1.9400238
##  [8,] 3.496199 0.7007631
##  [9,] 3.109454 2.3072431
## [10,] 3.552331 0.6253776
```

Questão 58

$\mathbf{X} \sim \mathbf{N}(\mu, \mathbf{\Sigma})$ tal que

$$\mu = \begin{bmatrix} 2 \\ 2 \end{bmatrix}; \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

E $\mathbf{A} = [1 \quad 1]; \mathbf{B} = [1 \quad -1]$. Mostre que $\mathbf{AX}$ é independente de $\mathbf{BX}$
Solução:
Como $\mathbf{\Sigma_{12}} = 0$, temos que os vetores $\mathbf{Y_1}$ e $\mathbf{Y_2}$ amostras de $\mathbf{X}$ sao independentes.

Seja

$$\mathbf{Y}_1 = \mathbf{AX} = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = X \ ;$$

e

$$\mathbf{Y}_2 = \mathbf{BX} = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ -X_2 \end{bmatrix} \neq X \ .$$

Que são independentes por $\Sigma_{12} = 0$. Logo, $Cov(\mathbf{Y_1}, \mathbf{Y_2}) = 0$
$\square$

Lista 8 - Correlação canônica

Questão 72

Ex. 10.2 | Johnson & Wichern

Os vetores aleatórios $\mathbf{X^{(1)}}, \mathbf{X^{(2)}}$ $(\mathbf{2} \times \mathbf{1})$ têm vetor de médias e variâncias conjuntas

$$\mu = \begin{bmatrix} \mu^{(1)} \\ -- \\ \mu^{(2)} \end{bmatrix} = \begin{bmatrix} -3 \\ 2 \\ -- \\ 0 \\ 1 \end{bmatrix}; \mathbf{\Sigma} = \begin{bmatrix} \Sigma_{11} & | & \Sigma_{12} \\ -- & -|- & -- \\ \Sigma_{21} & | & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} 8 & 2 & | & 3 & 1 \\ 2 & 5 & | & -1 & 3 \\ -- & -- & -|- & -- & -- \\ 3 & -1 & | & 6 & -2 \\ 1 & 3 & | & -2 & 7 \end{bmatrix}$$

a)

Calcular as correlações canônicas ρ_1^*, ρ_2^*.

```
## [1] 0.3046268 0.2399638
```

Logo, as correlações canônicas serão $\rho_1 = \sqrt{0,3046268} \approx 0.552$ e $\rho_2 = \sqrt{0,2399638} \approx 0.489$

b)

Determinar os pares de variáveis canônicas (U_1, V_1) e (U_2, V_2).

```
##            [,1]      [,2]
## [1,] -0.3168206 0.3622269
```

```
##            [,1]       [,2]
## [1,] -0.3647058 0.09506271
```

Portanto, $U_1 = -0,316X_1^{(1)} + 0,362X_2^{(1)}$,$V_1 = -0,364X_1^{(2)} + 0,095X_2^{(2)} = (U_1, V_1)$.

```
##           [,1]      [,2]
## [1,] 0.1962489 0.3016851
```

```
##           [,1]     [,2]
## [1,] 0.2262746 0.385821
```

Enquanto que (U_2, V_2) será: $U_2 = 0,196x_1^{(1)} + 0,301X_2^{(2)}$,$V_2 = 0,226X_1^{(2)} + 0,385X_2^{(2)}$.
□

c)

Seja $\mathbf{U} = [U_1, U_2]'$ e $\mathbf{V} = [V_1, V_2]'$. Avalie:

$$\mathbf{E}\begin{bmatrix} U \\ -- \\ V \end{bmatrix} \text{ e } \mathbf{Cov}\begin{bmatrix} U \\ -- \\ V \end{bmatrix} = \begin{bmatrix} \Sigma_{UU} & | & \Sigma_{UV} \\ -- & -|- & -- \\ \Sigma_{VU} & | & \Sigma_{VV} \end{bmatrix}$$

.

```
##           [,1]
## [1,] 1.674915
```

```
##            [,1]
## [1,] 0.01462369
```

```
##            [,1]
## [1,] 0.09506271
```

```
##           [,1]
## [1,] 0.385821
```

Portanto,

$$\mathbf{E}\begin{bmatrix} U \\ -- \\ V \end{bmatrix} = \begin{bmatrix} 1,674 \\ 0,014 \\ -- \\ 0,095 \\ 0,385 \end{bmatrix}$$

Enquanto que

$$\mathbf{Cov}\begin{bmatrix} U \\ -- \\ V \end{bmatrix} = \begin{bmatrix} \Sigma_{UU} & | & \Sigma_{UV} \\ -- & -|- & -- \\ \Sigma_{VU} & | & \Sigma_{VV} \end{bmatrix} = \begin{bmatrix} 1 & 0 & | & \rho_1^* & 0 \\ 0 & 1 & | & 0 & \rho_2^* \\ -- & -- & -|- & -- & -- \\ \rho_1^* & 0 & | & 1 & 0 \\ 0 & \rho_2^* & | & 0 & 1 \end{bmatrix}$$

será um resultado trivial.

Comparar os resultados com as propriedades do resultado 10.1.

Neste caso, da definição 10.1 do livro [1], e definição 10.5; $E(X^{(1)}) = \mu^{(1)}$, temos que $E(U) = E(a'X)$. Pela lineariedade da esperança, temos que $a'E(X) = E(U) = a'E(\mu)$, que foi justamente o resultado utilizado para as contas. Já para a covariância, existe a definição em 10.1, porém o resultado é trivial visto que Σ_{ij} irá retornar sempre uma matriz diagonal; se $i = j$, o valor da diagonal será 1 (a variável por ela mesma); e para $i \neq j$, o resultado será as correlações canônicas calculadas para i e j.
□

Questão 73

Ex. 10.9 | Johnson & Wichern (itens (a) e (c))

Foram aplicados para $n = 140$ alunos da sétima série quatro testes, tais que $\mathbf{X_1^{(1)}}$ = velocidade de leitura; $\mathbf{X_2^{(1)}}$ = habilidade de leitura; $\mathbf{X_1^{(2)}}$ = velocidade em aritmética; $\mathbf{X_2^{(2)}}$ = habilidade em aritmética. A correlação da performance medida foi:

$$\mathbf{R} = \begin{bmatrix} R_{11} & | & R12 \\ -- & -|- & -- \\ R21 & | & R22 \end{bmatrix} = \begin{bmatrix} 1.0 & 0,6328 & | & 0,2412 & 0,0586 \\ 0,6328 & 1 & | & -0,0553 & 0,0655 \\ --- & --- & -|- & --- & --- \\ 0,2412 & -0,0553 & | & 1 & 0,4248 \\ 0,0586 & 0,0655 & | & 0,4248 & 1 \end{bmatrix}$$

.

a)

Encontrar todas as correlações e variáveis canônicas amostrais

```
## [1] 0.155634923 0.004740029
```

Logo, as correlações canônicas amostrais serão $\hat{\rho}_1 = \sqrt{0,155634923} \approx 0.394$ e $\hat{\rho}_2 = \sqrt{0,004740029} \approx 0.068$.

```
##             [,1]     [,2]
## [1,] -1.256845 1.025317
```

```
##             [,1]      [,2]
## [1,] 0.2970177 0.7852413
```

```
##             [,1]      [,2]
## [1,] -1.104472 0.4527216
```

```
##               [,1]     [,2]
## [1,] -0.01818009 1.007587
```

Enquanto que os pares canônicos amostrais serão: $\hat{U}_1 = -1,256X_1^{(1)} + 1,025X_2^{(1)}; \hat{V}_1 = -1,104X_1^{(2)} + 0,452X_2^{(2)}$ e $\hat{U}_2 = 0,297X_1^{(1)} + 0,785X_2^{(1)}; \hat{V}_2 = -0,018X_1^{(2)} + 1,007X_2^{(2)}$

c)

Avaliar as matrizes de erros aproximados para $\mathbf{R_{11}}, \mathbf{R_{22}}$ e $\mathbf{R_{12}}$ determinadas pelo primeiro par de variáveis canônicas $\hat{U}_1, \hat{V}_1$.

O cálculo das matrizes resultou em:

$\mathbf{R_{11}} =$

```
##              [,1]      [,2]
## [1,] 0.05289325 0.2238204
## [2,] 0.22382037 0.9471067
```

$\mathbf{R_{22}} =$

```
##                 [,1]        [,2]
## [1,]  0.0002708725 -0.01645598
## [2,] -0.0164559754  0.99972913
```

$\mathbf{R_{12}} =$

```
##                 [,1]       [,2]
## [1,] -0.0002605992 0.01583185
## [2,] -0.0011027382 0.06699326
```

Referências:

[1] JOHNSON, Richard A; WICHERN, Dean W. APPLIED MULTIVARIATE STATISTICAL ANALYSIS. 6ª Edição. Pearson, 2007.

[2] von Borries, George. Material de aula disponível no Aprender3; Notas de aula e códigos. Análise Multivariada 1. Universidade de Brasília, 2023.

[3] RENCHER, Alvin C; CHRISTENSEN, William F. METHODS OF MULTIVARIATE ANALYSIS. 3ª Edição. WILEY, 2012.