Lista Entrega 5

Davi Wentrick Feijó - 200016806

2023-06-05

Questao 1 (9.1)

A questao nos dá a matriz de covariancia ρ e a matriz de erros Ψ

A matriz ρ

1.00	0.63	0.45
0.63	1.00	0.35
0.45	0.35	1.00

A matriz Ψ

0.19	0.00	0.00
0.00	0.51	0.00
0.00	0.00	0.75

Sabemos que na analise fatorial temos a seguinte relacao:

$$\Sigma = LL^T + \Psi$$
$$LL^T = \Sigma - \Psi$$

Calculando LL^T

0.81	0.63	0.45
0.63	0.49	0.35
0.45	0.35	0.25

Podemos encontrar a comunalidade na diagonal da matriz LL^T já que subtraimos o Ψ

x
0.81
0.49
0.25

Com essas informacoes podemos escrever nossa matriz Σ como:

$$\Sigma = LL^T + \Psi$$

```r
p_construido = LLT + psi
```

1.00	0.63	0.45
0.63	1.00	0.35
0.45	0.35	1.00

Questao 2 (9.2)

A) As comunalidades sao:

```r
comu
```

```
## [1] 0.81 0.49 0.25
```

Podemos perceber que F1 detem a maior comunalidade logo é o fator que mais explica a variancia dos dados

B) Sabemos que:

$$Cor(X, Y) = \frac{Cov(X, Y)}{S_x S_y}$$

$$Cov(X, F) = L$$

Logo

$$Cor(X_i, F_i) = \frac{Cov(X_i, F_i)}{S_i S_f} = \frac{L_i}{S_x S_f}$$

```r
cor_xf = Lestimado[1]/(1*comu[1])
```

```
## [1] -1.141896
```

Questao 3 (9.3)

A) Para realizar por meio de componentes principais primeiro precisamos encontrar os autovalores e autovetores da matriz de correlacao aplicando a decompisicao espectral em ρ dada na questao 9.1

$$\rho = CDC^T$$

```
eigen_p = eigen(p)

autoval <- eigen_p$values

autovet <- eigen_p$vectors

D <- matrix(0, nrow = 3, ncol = 3)
diag(D) <- sqrt(autoval)
```

```
autoval
```

```
## [1] 1.9632830 0.6794930 0.3572239
```

```
autovet
```

```
##            [,1]       [,2]       [,3]
## [1,] -0.6250027  0.2186276  0.7493822
## [2,] -0.5931510  0.4910833 -0.6379726
## [3,] -0.5074875 -0.8432314 -0.1772492
```

Em seguida podemos encontrar nossa matriz L

$$L = CD^{1/2}$$

Aqui temos nossa matriz dos loadings

```
Lestimado
```

```
## [1] -0.8757363 -0.8311066 -0.7110772
```

Para calcular a matriz Ψ temos que seguir a equacao:

$$\Psi = \Sigma - LL^T$$

Na diagonal obteremos nosso Ψ

```
psiestimado <- diag(p-LLT)
```

```
psiestimado
```

```
## [1] 0.2330860 0.3092618 0.4943692
```

Para comparar com os resultados anteriores podemos aproximar a matrix Σ de correlacoes por meio da formula:

$$\Sigma = LL^T + \Psi$$

```
##           [,1]      [,2]      [,3]
## [1,] 0.9569140 0.7278302 0.6227161
## [2,] 0.7278302 1.2007382 0.5909810
## [3,] 0.6227161 0.5909810 1.2556308
```

B) A variancia explicada é:

```
## [1] 0.6544277 0.2264977 0.1190746
```

Podemos notar que a primeira componente exxplica 65% da variancia dos dados

Questao 4 (9.19)

	x1	x2	x3	x4	x5	x6	x7
x1	1.0000000	0.9260758	0.8840023	0.5720363	0.7080738	0.6744073	0.9273116
x2	0.9260758	1.0000000	0.8425232	0.5415080	0.7459097	0.4653880	0.9442960
x3	0.8840023	0.8425232	1.0000000	0.7003630	0.6374712	0.6410886	0.8525682
x4	0.5720363	0.5415080	0.7003630	1.0000000	0.5907360	0.1469074	0.4126395
x5	0.7080738	0.7459097	0.6374712	0.5907360	1.0000000	0.3859502	0.5745533
x6	0.6744073	0.4653880	0.6410886	0.1469074	0.3859502	1.0000000	0.5663721
x7	0.9273116	0.9442960	0.8525682	0.4126395	0.5745533	0.5663721	1.0000000

A) Vamos usar a função principa() para obter a analise fatorial com m=2 e m=3

m=2

```
AF2
```

```
## Principal Components Analysis
## Call: principal(r = cor_data, nfactors = 2, rotate = "none", n.obs = 50,
##     covar = F)
## Standardized loadings (pattern matrix) based upon correlation matrix
##     PC1   PC2   h2    u2 com
## x1 0.97 -0.11 0.96 0.041 1.0
## x2 0.94  0.03 0.89 0.110 1.0
## x3 0.94  0.01 0.89 0.107 1.0
## x4 0.66  0.65 0.85 0.147 2.0
## x5 0.78  0.28 0.69 0.305 1.3
## x6 0.65 -0.62 0.81 0.194 2.0
## x7 0.91 -0.19 0.87 0.127 1.1
##
##                        PC1  PC2
## SS loadings           5.03 0.93
## Proportion Var        0.72 0.13
## Cumulative Var        0.72 0.85
## Proportion Explained  0.84 0.16
## Cumulative Proportion 0.84 1.00
##
## Mean item complexity =  1.3
## Test of the hypothesis that 2 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.08
##  with the empirical chi square  11.93  with prob <  0.15
##
## Fit based upon off diagonal values = 0.99
```

Vale notar que a primeira componente explica 72% enquanto que a segunda 13%. Isso indica que adicionar uma 3 variavel nao deve adicionar muita informação m=3

```
AF3
```

```
## Principal Components Analysis
## Call: principal(r = cor_data, nfactors = 3, rotate = "none", n.obs = 50,
##     covar = F)
## Standardized loadings (pattern matrix) based upon correlation matrix
##     PC1   PC2   PC3   h2    u2 com
## x1 0.97 -0.11 -0.05 0.96 0.039 1.0
## x2 0.94  0.03 -0.31 0.99 0.013 1.2
## x3 0.94  0.01  0.14 0.91 0.087 1.0
## x4 0.66  0.65  0.32 0.95 0.045 2.4
## x5 0.78  0.28  0.00 0.69 0.305 1.3
## x6 0.65 -0.62  0.43 0.99 0.012 2.7
## x7 0.91 -0.19 -0.31 0.97 0.033 1.3
##
##                        PC1  PC2  PC3
## SS loadings           5.03 0.93 0.50
## Proportion Var        0.72 0.13 0.07
## Cumulative Var        0.72 0.85 0.92
## Proportion Explained  0.78 0.14 0.08
## Cumulative Proportion 0.78 0.92 1.00
##
## Mean item complexity =  1.6
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.04
##  with the empirical chi square  3.95  with prob <  0.27
##
## Fit based upon off diagonal values = 1
```

Como haviamos discutido antes, acabou que a 3 componente adicionou 7%

B) Aqui temos as mesmas analises porem rotacionadas com o metodos “varimax”

```
AF2_rotated
```

```
## Principal Components Analysis
## Call: principal(r = cor_data, nfactors = 2, rotate = "varimax", n.obs = 50,
##     covar = F)
## Standardized loadings (pattern matrix) based upon correlation matrix
##     RC1   RC2   h2    u2 com
## x1 0.79  0.58 0.96 0.041 1.8
## x2 0.67  0.66 0.89 0.110 2.0
## x3 0.68  0.65 0.89 0.107 2.0
## x4 0.04  0.92 0.85 0.147 1.0
## x5 0.38  0.74 0.69 0.305 1.5
## x6 0.90 -0.01 0.81 0.194 1.0
## x7 0.80  0.48 0.87 0.127 1.6
##
##                        RC1  RC2
## SS loadings           3.13 2.84
## Proportion Var        0.45 0.41
## Cumulative Var        0.45 0.85
## Proportion Explained  0.52 0.48
## Cumulative Proportion 0.52 1.00
##
## Mean item complexity =  1.6
## Test of the hypothesis that 2 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.08
##  with the empirical chi square  11.93  with prob <  0.15
##
## Fit based upon off diagonal values = 0.99
```

```r
AF3_rotated
```

```
## Principal Components Analysis
## Call: principal(r = cor_data, nfactors = 3, rotate = "varimax", n.obs = 50,
##     covar = F)
## Standardized loadings (pattern matrix) based upon correlation matrix
##     RC1  RC2  RC3   h2    u2 com
## x1 0.78 0.39 0.45 0.96 0.039 2.1
## x2 0.91 0.36 0.19 0.99 0.013 1.4
## x3 0.62 0.55 0.48 0.91 0.087 2.9
## x4 0.21 0.95 0.05 0.95 0.045 1.1
## x5 0.55 0.61 0.15 0.69 0.305 2.1
## x6 0.29 0.06 0.95 0.99 0.012 1.2
## x7 0.91 0.18 0.33 0.97 0.033 1.3
##
##                        RC1  RC2  RC3
## SS loadings           3.07 1.89 1.51
## Proportion Var        0.44 0.27 0.22
## Cumulative Var        0.44 0.71 0.92
## Proportion Explained  0.48 0.29 0.23
## Cumulative Proportion 0.48 0.77 1.00
##
## Mean item complexity =  1.7
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.04
##  with the empirical chi square  3.95  with prob <  0.27
##
## Fit based upon off diagonal values = 1
```

Apos aplicar a rotação podemos perceber que a informação fica mais bem distribuida entre as componentes. O objetivo é simplificar a analise.

C) Vamos obter as comunalidades, a variancia especifica e a matriz L sem a rotação varimax

```
## Resultados para m=2

## A comunalidade é:
##  0.95 0.89 0.89 0.44 0.61 0.42 0.84 0.01 0 0 0.42 0.08 0.39 0.04

## A diagonal da matriz psi é:
##  0.04 0.11 0.11 0.15 0.31 0.19 0.13

## A matriz LLT:

##       x1    x2    x3    x4    x5    x6    x7
## x1 0.959 0.914 0.918 0.573 0.731 0.698 0.910
## x2 0.914 0.890 0.891 0.641 0.747 0.594 0.856
## x3 0.918 0.891 0.893 0.630 0.743 0.607 0.862
## x4 0.573 0.641 0.630 0.853 0.701 0.028 0.479
## x5 0.731 0.747 0.743 0.701 0.695 0.331 0.661
## x6 0.698 0.594 0.607 0.028 0.331 0.806 0.713
## x7 0.910 0.856 0.862 0.479 0.661 0.713 0.873

## Matriz de correlação aproximada:

##        x1     x2     x3     x4     x5     x6     x7
## x1  0.000  0.012 -0.034 -0.001 -0.023 -0.024  0.017
## x2  0.012  0.000 -0.049 -0.099 -0.001 -0.129  0.088
## x3 -0.034 -0.049  0.000  0.071 -0.105  0.034 -0.009
## x4 -0.001 -0.099  0.071  0.000 -0.111  0.119 -0.066
## x5 -0.023 -0.001 -0.105 -0.111  0.000  0.055 -0.086
## x6 -0.024 -0.129  0.034  0.119  0.055  0.000 -0.147
## x7  0.017  0.088 -0.009 -0.066 -0.086 -0.147  0.000

## Resultados para m=3

## A comunalidade é:
##  0.92 0.84 0.84 0.29 0.48 0.27 0.76 0 0 0 0.27 0.02 -0.24 -0.01 0 -0.03 0 0.03 0 0.08 -0.03

## A diagonal da matriz psi é:
##  0.04 0.01 0.09 0.05 0.31 0.01 0.03

## A matriz LLT:

##       x1    x2    x3    x4    x5    x6    x7
## x1 0.961 0.931 0.911 0.556 0.731 0.676 0.927
## x2 0.931 0.987 0.846 0.541 0.745 0.461 0.952
## x3 0.911 0.846 0.913 0.675 0.743 0.669 0.818
## x4 0.556 0.541 0.675 0.955 0.703 0.163 0.381
## x5 0.731 0.745 0.743 0.703 0.695 0.333 0.660
## x6 0.676 0.461 0.669 0.163 0.333 0.988 0.583
## x7 0.927 0.952 0.818 0.381 0.660 0.583 0.967
```

```
## Matriz de correlação aproximada:
```

```
##        x1     x2     x3     x4     x5     x6     x7
## x1  0.000  0.012 -0.034 -0.001 -0.023 -0.024  0.017
## x2  0.012  0.000 -0.049 -0.099 -0.001 -0.129  0.088
## x3 -0.034 -0.049  0.000  0.071 -0.105  0.034 -0.009
## x4 -0.001 -0.099  0.071  0.000 -0.111  0.119 -0.066
## x5 -0.023 -0.001 -0.105 -0.111  0.000  0.055 -0.086
## x6 -0.024 -0.129  0.034  0.119  0.055  0.000 -0.147
## x7  0.017  0.088 -0.009 -0.066 -0.086 -0.147  0.000
```

D) Vamos testar se o numero de fatores é suficiente para representar o banco

- Hipótese nula (H_0): $\Sigma = LL^T + \Psi$
- Hipótese alternativa (H_1): $\Sigma \neq LL^T + \Psi$

Podemos usar a propria função principal para fazer o calculo do teste, mas para isso temos que adicionar o numero de observaçoes.

```
## Teste para m=2
```

```
## A estatica qui-quadrado obtida foi: 11.9336
```

```
## O p-valor obtido foi: 9.682959e-32
```

```
## Teste para m=3
```

```
## A estatica qui-quadrado obtida foi: 3.946427
```

```
## O p-valor obtido foi: 3.703147e-17
```

A partir do p-valor temos evidencias para rejeitar H0 indicando que 2 e 3 fatores nao sao suficiente para estimar a matriz Σ logo teriamos que usar mais fatores para representar melhor.