Entrega 1 Multivariada George

Davi Wentrick Feijó - 200016806

2023-04-20

```r
pacman::p_load(tidyverse,knitr,ggcorrplot,Matrix,psych)


matrix_corr <- function(r,p) {
  matriz  = matrix(1,r,r)
  matriz[lower.tri(matriz)] = p
  matriz[upper.tri(matriz)] = p

  return(matriz)
}
p = c(0.1, 0.3, 0.5, 0.7, 0.9)
```

Exercício 13 da Lista 2: Considere uma matriz de correlação (r × r) com a mesma correlação (ρ) em todas as células fora da diagonal. Encontre os autovalores e autovetores desta matriz quando r = 2, 3, 4. Generalize seus resultados para qualquer número r de variáveis. Como exemplo, faça ρ = 0.1, 0.3, 0.5, 0.7, 0.9.

A seguir os resultados das matrizes $X_{2\times 2}$ com $x_{ii} = 1$ e $x_{ij} = \rho$:

```
## Autovalores e Autovetores de uma matriz 2 x 2 com p = 0.1

## eigen() decomposition
## $values
## [1] 1.1 0.9
##
## $vectors
##           [,1]       [,2]
## [1,] 0.7071068 -0.7071068
## [2,] 0.7071068  0.7071068

## Autovalores e Autovetores de uma matriz 2 x 2 com p = 0.3

## eigen() decomposition
## $values
## [1] 1.3 0.7
##
## $vectors
##           [,1]       [,2]
## [1,] 0.7071068 -0.7071068
## [2,] 0.7071068  0.7071068
```

```
## Autovalores e Autovetores de uma matriz 2 x 2 com p = 0.5

## eigen() decomposition
## $values
## [1] 1.5 0.5
##
## $vectors
##           [,1]       [,2]
## [1,] 0.7071068 -0.7071068
## [2,] 0.7071068  0.7071068

## Autovalores e Autovetores de uma matriz 2 x 2 com p = 0.7

## eigen() decomposition
## $values
## [1] 1.7 0.3
##
## $vectors
##           [,1]       [,2]
## [1,] 0.7071068 -0.7071068
## [2,] 0.7071068  0.7071068

## Autovalores e Autovetores de uma matriz 2 x 2 com p = 0.9

## eigen() decomposition
## $values
## [1] 1.9 0.1
##
## $vectors
##           [,1]       [,2]
## [1,] 0.7071068 -0.7071068
## [2,] 0.7071068  0.7071068
```

A seguir os resultados das matrizes $X_{3\times 3}$ com $x_{ii} = 1$ e $x_{ij} = \rho$:

```
## Autovalores e Autovetores de uma matriz 3 x 3 com p = 0.1

## eigen() decomposition
## $values
## [1] 1.2 0.9 0.9
##
## $vectors
##           [,1]       [,2]       [,3]
## [1,] 0.5773503  0.0000000  0.8164966
## [2,] 0.5773503 -0.7071068 -0.4082483
## [3,] 0.5773503  0.7071068 -0.4082483

## Autovalores e Autovetores de uma matriz 3 x 3 com p = 0.3

## eigen() decomposition
## $values
## [1] 1.6 0.7 0.7
##
## $vectors
##            [,1]       [,2]       [,3]
## [1,] -0.5773503  0.8164966  0.0000000
## [2,] -0.5773503 -0.4082483 -0.7071068
## [3,] -0.5773503 -0.4082483  0.7071068

## Autovalores e Autovetores de uma matriz 3 x 3 com p = 0.5

## eigen() decomposition
## $values
## [1] 2.0 0.5 0.5
##
## $vectors
##            [,1]       [,2]       [,3]
## [1,] -0.5773503  0.0000000  0.8164966
## [2,] -0.5773503 -0.7071068 -0.4082483
## [3,] -0.5773503  0.7071068 -0.4082483

## Autovalores e Autovetores de uma matriz 3 x 3 com p = 0.7

## eigen() decomposition
## $values
## [1] 2.4 0.3 0.3
##
## $vectors
##           [,1]       [,2]        [,3]
## [1,] 0.5773503  0.3555207  0.73503175
## [2,] 0.5773503 -0.8143165 -0.05962589
## [3,] 0.5773503  0.4587958 -0.67540586

## Autovalores e Autovetores de uma matriz 3 x 3 com p = 0.9
```

```
## eigen() decomposition
## $values
## [1] 2.8 0.1 0.1
##
## $vectors
##           [,1]       [,2]        [,3]
## [1,] 0.5773503  0.4586617  0.67549693
## [2,] 0.5773503 -0.8143284  0.05946423
## [3,] 0.5773503  0.3556666 -0.73496116
```

A seguir os resultados das matrizes $X_{4\times 4}$ com $x_{ii} = 1$ e $x_{ij} = \rho$:

```
## Autovalores e Autovetores de uma matriz 4 x 4 com p = 0.1

## eigen() decomposition
## $values
## [1] 1.3 0.9 0.9 0.9
##
## $vectors
##      [,1]       [,2]          [,3]         [,4]
## [1,] -0.5  0.5099852  0.000000e+00  0.699939347
## [2,] -0.5  0.4899141  8.756053e-17 -0.714131779
## [3,] -0.5 -0.4999496 -7.071068e-01  0.007096216
## [4,] -0.5 -0.4999496  7.071068e-01  0.007096216

## Autovalores e Autovetores de uma matriz 4 x 4 com p = 0.3

## eigen() decomposition
## $values
## [1] 1.9 0.7 0.7 0.7
##
## $vectors
##      [,1]       [,2]       [,3]       [,4]
## [1,] -0.5  0.8660254  0.0000000  0.0000000
## [2,] -0.5 -0.2886751  0.0000000  0.8164966
## [3,] -0.5 -0.2886751 -0.7071068 -0.4082483
## [4,] -0.5 -0.2886751  0.7071068 -0.4082483

## Autovalores e Autovetores de uma matriz 4 x 4 com p = 0.5

## eigen() decomposition
## $values
## [1] 2.5 0.5 0.5 0.5
##
## $vectors
##      [,1]       [,2]       [,3]       [,4]
## [1,] -0.5  0.8660254  0.0000000  0.0000000
## [2,] -0.5 -0.2886751 -0.5773503 -0.5773503
## [3,] -0.5 -0.2886751 -0.2113249  0.7886751
## [4,] -0.5 -0.2886751  0.7886751 -0.2113249

## Autovalores e Autovetores de uma matriz 4 x 4 com p = 0.7

## eigen() decomposition
## $values
## [1] 3.1 0.3 0.3 0.3
##
## $vectors
##      [,1]       [,2]       [,3]       [,4]
## [1,] -0.5  0.0000000  0.0000000  0.8660254
## [2,] -0.5 -0.5773503 -0.5773503 -0.2886751
## [3,] -0.5 -0.2113249  0.7886751 -0.2886751
## [4,] -0.5  0.7886751 -0.2113249 -0.2886751
```

```
## Autovalores e Autovetores de uma matriz 4 x 4 com p = 0.9

## eigen() decomposition
## $values
## [1] 3.7 0.1 0.1 0.1
##
## $vectors
##      [,1]       [,2]       [,3]          [,4]
## [1,] -0.5  0.8660254  0.0000000  0.000000e+00
## [2,] -0.5 -0.2886751 -0.8164966 -1.751211e-16
## [3,] -0.5 -0.2886751  0.4082483  7.071068e-01
## [4,] -0.5 -0.2886751  0.4082483 -7.071068e-01
```

Dado o vetor aleatorio $X^\top = [X_1, X_2, X_3, X_4]$ e o vetor de médias $\mu_X^\top = [3, 2, -2, 0]$ e a seguinte matriz de variancia-covariancia:

```r
#montando as matrizes no R
A = c(1,-1,0,0,1,1,-2,0,1,1,1,-3)
medias = c(3,2,-2,0)
matriz_A = matrix(A,ncol = 4,byrow = TRUE)
sigma_x = diag(3,nrow = 4)
matriz_medias = matrix(medias,ncol = 1)
```

$$\Sigma_X = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 1 & -2 & 0 \\ 1 & 1 & 1 & -3 \end{bmatrix}$$

Encontre a $E(AX)$ Pela linearidade da esperanca temos:

$$E(AX) = AE(X) = A\mu_X \quad E(X) = \mu_X = (\mu_X^\top)^\top$$

Logo teremos a seguinte equacao:

$$E(AX) = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 1 & -2 & 0 \\ 1 & 1 & 1 & -3 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 2 \\ -2 \\ 0 \end{bmatrix}$$

```r
matriz_A %*% matriz_medias
```

```
##      [,1]
## [1,]    1
## [2,]    9
## [3,]    3
```

$$E(AX) = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 1 & -2 & 0 \\ 1 & 1 & 1 & -3 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 2 \\ -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 9 \\ -3 \end{bmatrix}$$

Encontre a $Cov(AX)$ Sabemos que:

$$Cov(AX) = A \cdot \Sigma_X \cdot A^\top$$

$$Cov(AX) = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 1 & -2 & 0 \\ 1 & 1 & 1 & -3 \end{bmatrix} \cdot \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -2 & 0 \\ -3 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 18 & 0 \\ 0 & 0 & 36 \end{bmatrix}$$

```r
matriz_A %*% sigma_x %*% t(matriz_A)
```

```
##      [,1] [,2] [,3]
## [1,]    6    0    0
## [2,]    0   18    0
## [3,]    0    0   36
```

Quais pares de combinacao linear tem covariancia igual a zero? Os pares (1,2);(1,3);(2,1);(2,3);(3,1);(3,2), ou seja todos os pares com $i \neq j$ apresentaram $\sigma = 0$

Exercício 28 da Lista 2: Considere o seguinte conjunto de dados de Pacientes em Tratamento de Hemodiálise.

Represente graficamente e através de medidas descritivas. Matriz de covariancia

	idade	proteina	energia	albumina	imc
idade	174.9393939	-2.8203693	-3736.92562	0.5121212	19.3448864
proteina	-2.8203693	0.2079267	152.26715	-0.0321449	-1.4953153
energia	-3736.9256250	152.2671531	249960.54006	-11.0733438	-573.0939687
albumina	0.5121212	-0.0321449	-11.07334	0.0613258	0.4797727
imc	19.3448864	-1.4953153	-573.09397	0.4797727	22.0800568

Matriz de variancias

	idade	proteina	energia	albumina	imc
idade	174.9393939	-2.8203693	-3736.92562	0.5121212	19.3448864
proteina	-2.8203693	0.2079267	152.26715	-0.0321449	-1.4953153
energia	-3736.9256250	152.2671531	249960.54006	-11.0733438	-573.0939687
albumina	0.5121212	-0.0321449	-11.07334	0.0613258	0.4797727
imc	19.3448864	-1.4953153	-573.09397	0.4797727	22.0800568

Matriz de correlacao

	idade	proteina	energia	albumina	imc
idade	1.0000000	-0.4676350	-0.5651125	0.1563535	0.3112593
proteina	-0.4676350	1.0000000	0.6679060	-0.2846658	-0.6978749
energia	-0.5651125	0.6679060	1.0000000	-0.0894379	-0.2439439
albumina	0.1563535	-0.2846658	-0.0894379	1.0000000	0.4123006
imc	0.3112593	-0.6978749	-0.2439439	0.4123006	1.0000000

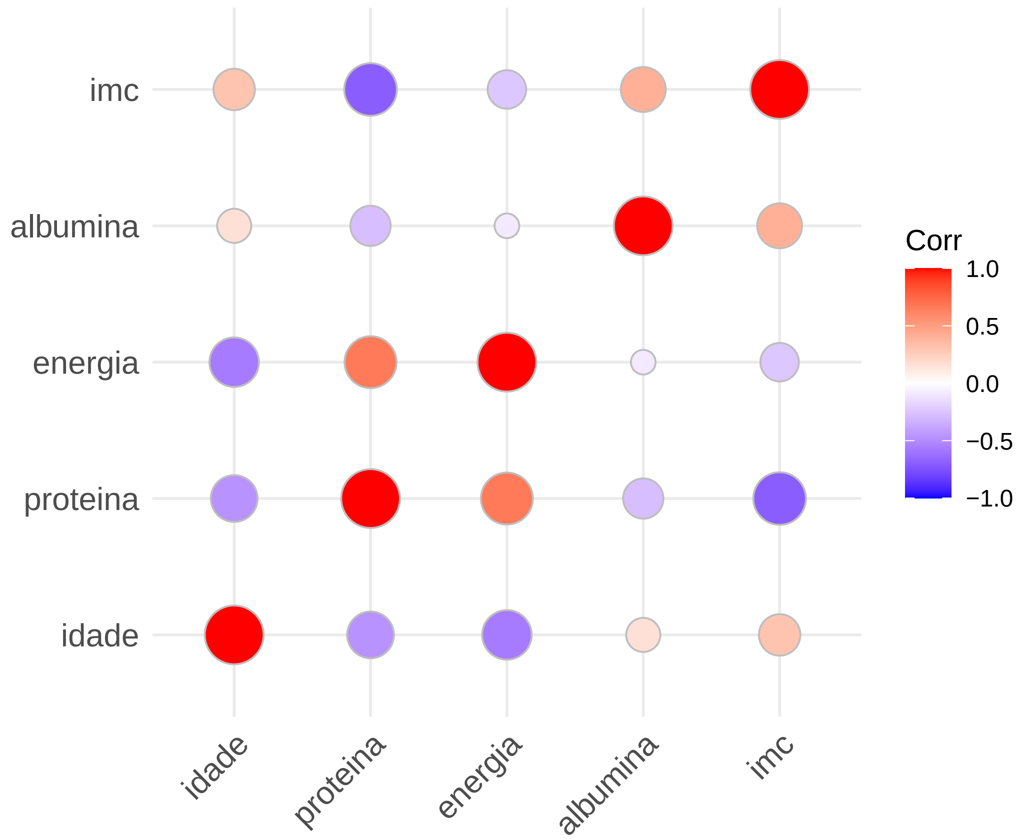

Figure 1: Gráfico de correlacao entre as variaveis

```r
p <- as.matrix(dados)
p_svd = svd(p)

u = p_svd$u
v = p_svd$v
d = diag(p_svd$d)
```

```r
#2 valores singulares
(p_svd_2 <- svd(p,nu=2,nv=2))
```

Obtenha a decomposição espectral e verifique se existe indicação de uma possível redução da dimensão do estudo em questão. Justifique.

```
## $d
## [1] 8952.322440  152.222259   28.608980    2.767222    1.287161
##
## $u
##              [,1]         [,2]
##  [1,] -0.30592194 -0.324267973
##  [2,] -0.09226951  0.270009424
##  [3,] -0.14612013  0.085552992
##  [4,] -0.10353565  0.188927788
##  [5,] -0.31131788 -0.389764025
##  [6,] -0.13673702  0.200854095
##  [7,] -0.22770684 -0.170074647
##  [8,] -0.11869546  0.104305485
##  [9,] -0.18531142  0.073147111
## [10,] -0.18466507 -0.095650959
## [11,] -0.18293641  0.125870242
## [12,] -0.20611534 -0.172401493
## [13,] -0.17241102  0.073998455
## [14,] -0.13580762  0.132118095
## [15,] -0.16222091  0.064305096
## [16,] -0.19969861 -0.023795158
## [17,] -0.22064851 -0.200170061
## [18,] -0.13964746  0.188375045
## [19,] -0.11054700  0.261901520
## [20,] -0.11933414  0.131218754
## [21,] -0.10839331  0.246625504
## [22,] -0.09367140  0.245270087
## [23,] -0.13394239  0.101143632
## [24,] -0.20313418 -0.154326317
## [25,] -0.13848258  0.035620244
## [26,] -0.24060312 -0.123122442
## [27,] -0.19231188  0.025547528
## [28,] -0.21305452  0.003909098
## [29,] -0.10983205  0.179571733
## [30,] -0.11418641  0.145078879
## [31,] -0.11494705  0.084176526
## [32,] -0.16394515  0.221944065
```

```
## [33,] -0.16280525  0.126666699
##
## $v
##                [,1]          [,2]
## [1,] -0.0286968729  0.9432313346
## [2,] -0.0007260725  0.0008550346
## [3,] -0.9994775101 -0.0319941722
## [4,] -0.0024219141  0.0448994730
## [5,] -0.0146561766  0.3275275997
```

```
(u2 <- round(p_svd_2$u,1))
```

```
##       [,1] [,2]
##  [1,] -0.3 -0.3
##  [2,] -0.1  0.3
##  [3,] -0.1  0.1
##  [4,] -0.1  0.2
##  [5,] -0.3 -0.4
##  [6,] -0.1  0.2
##  [7,] -0.2 -0.2
##  [8,] -0.1  0.1
##  [9,] -0.2  0.1
## [10,] -0.2 -0.1
## [11,] -0.2  0.1
## [12,] -0.2 -0.2
## [13,] -0.2  0.1
## [14,] -0.1  0.1
## [15,] -0.2  0.1
## [16,] -0.2  0.0
## [17,] -0.2 -0.2
## [18,] -0.1  0.2
## [19,] -0.1  0.3
## [20,] -0.1  0.1
## [21,] -0.1  0.2
## [22,] -0.1  0.2
## [23,] -0.1  0.1
## [24,] -0.2 -0.2
## [25,] -0.1  0.0
## [26,] -0.2 -0.1
## [27,] -0.2  0.0
## [28,] -0.2  0.0
## [29,] -0.1  0.2
## [30,] -0.1  0.1
## [31,] -0.1  0.1
## [32,] -0.2  0.2
## [33,] -0.2  0.1
```

```
(v2 <- round(p_svd_2$v,1))
```

```
##      [,1] [,2]
## [1,]    0  0.9
## [2,]    0  0.0
## [3,]   -1  0.0
```

```
## [4,]    0  0.0
## [5,]    0  0.3
```

```r
(d2 <- round(diag(p_svd_2$d[1:2]),1))
```

```
##        [,1]  [,2]
## [1,] 8952.3   0.0
## [2,]    0.0 152.2
```

```r
#3 valores singulares
(p_svd_3 <- svd(p,nu=3,nv=3))
```

```
## $d
## [1] 8952.322440  152.222259   28.608980    2.767222    1.287161
##
## $u
##              [,1]         [,2]         [,3]
##  [1,] -0.30592194 -0.324267973 -0.004576321
##  [2,] -0.09226951  0.270009424 -0.156544762
##  [3,] -0.14612013  0.085552992  0.126220554
##  [4,] -0.10353565  0.188927788 -0.075358535
##  [5,] -0.31131788 -0.389764025 -0.002396906
##  [6,] -0.13673702  0.200854095  0.109545685
##  [7,] -0.22770684 -0.170074647  0.099737899
##  [8,] -0.11869546  0.104305485 -0.047453598
##  [9,] -0.18531142  0.073147111 -0.119230332
## [10,] -0.18466507 -0.095650959 -0.180048752
## [11,] -0.18293641  0.125870242  0.096807969
## [12,] -0.20611534 -0.172401493 -0.125480326
## [13,] -0.17241102  0.073998455 -0.098456162
## [14,] -0.13580762  0.132118095  0.122030593
## [15,] -0.16222091  0.064305096 -0.097020989
## [16,] -0.19969861 -0.023795158  0.011285469
## [17,] -0.22064851 -0.200170061  0.006249183
## [18,] -0.13964746  0.188375045  0.326442996
## [19,] -0.11054700  0.261901520 -0.001275511
## [20,] -0.11933414  0.131218754 -0.157428931
## [21,] -0.10839331  0.246625504 -0.346651365
## [22,] -0.09367140  0.245270087 -0.029318877
## [23,] -0.13394239  0.101143632  0.115356715
## [24,] -0.20313418 -0.154326317 -0.107212768
## [25,] -0.13848258  0.035620244 -0.072719555
## [26,] -0.24060312 -0.123122442  0.302352684
## [27,] -0.19231188  0.025547528 -0.116306345
## [28,] -0.21305452  0.003909098 -0.138528845
## [29,] -0.10983205  0.179571733 -0.105376925
## [30,] -0.11418641  0.145078879 -0.334474526
## [31,] -0.11494705  0.084176526  0.042811863
## [32,] -0.16394515  0.221944065  0.530549812
## [33,] -0.16280525  0.126666699  0.104146462
##
## $v
##               [,1]          [,2]         [,3]
```

```
## [1,] -0.0286968729  0.9432313346  0.329861784
## [2,] -0.0007260725  0.0008550346  0.033223677
## [3,] -0.9994775101 -0.0319941722  0.004453749
## [4,] -0.0024219141  0.0448994730 -0.058062868
## [5,] -0.0146561766  0.3275275997 -0.941645506
```

```r
(u3 <- round(p_svd_3$u,1))
```

```
##       [,1] [,2] [,3]
##  [1,] -0.3 -0.3  0.0
##  [2,] -0.1  0.3 -0.2
##  [3,] -0.1  0.1  0.1
##  [4,] -0.1  0.2 -0.1
##  [5,] -0.3 -0.4  0.0
##  [6,] -0.1  0.2  0.1
##  [7,] -0.2 -0.2  0.1
##  [8,] -0.1  0.1  0.0
##  [9,] -0.2  0.1 -0.1
## [10,] -0.2 -0.1 -0.2
## [11,] -0.2  0.1  0.1
## [12,] -0.2 -0.2 -0.1
## [13,] -0.2  0.1 -0.1
## [14,] -0.1  0.1  0.1
## [15,] -0.2  0.1 -0.1
## [16,] -0.2  0.0  0.0
## [17,] -0.2 -0.2  0.0
## [18,] -0.1  0.2  0.3
## [19,] -0.1  0.3  0.0
## [20,] -0.1  0.1 -0.2
## [21,] -0.1  0.2 -0.3
## [22,] -0.1  0.2  0.0
## [23,] -0.1  0.1  0.1
## [24,] -0.2 -0.2 -0.1
## [25,] -0.1  0.0 -0.1
## [26,] -0.2 -0.1  0.3
## [27,] -0.2  0.0 -0.1
## [28,] -0.2  0.0 -0.1
## [29,] -0.1  0.2 -0.1
## [30,] -0.1  0.1 -0.3
## [31,] -0.1  0.1  0.0
## [32,] -0.2  0.2  0.5
## [33,] -0.2  0.1  0.1
```

```r
(v3 <- round(p_svd_3$v,1))
```

```
##      [,1] [,2] [,3]
## [1,]    0  0.9  0.3
## [2,]    0  0.0  0.0
## [3,]   -1  0.0  0.0
## [4,]    0  0.0 -0.1
## [5,]    0  0.3 -0.9
```

```r
(d3 <- round(diag(p_svd_3$d[1:3]),1))
```

```
##        [,1]  [,2] [,3]
## [1,] 8952.3   0.0  0.0
## [2,]    0.0 152.2  0.0
## [3,]    0.0   0.0 28.6
```

```r
var_explicada_2 = tr(d2)/tr(d) * 100
var_explicada_3 = tr(d3)/tr(d) * 100
```

```
## escolhendo 2 valores singulares a variancia explicada é: 99.64203
```

```
## escolhendo 3 valores singulares a variancia explicada é: 99.95504
```

```
## Uma diferenca de: 0.3130059
```

Logo podemos perceber que uma reducao é possivel, e muito eficaz ja que retem mais de 99% da variancia total do banco de dados!