

Análise de Componentes Principais (PCA) e Análise Fatorial

Alison Cordeiro Sousa

Contexto

Este projeto realizou uma Análise de Componentes Principais (PCA) e Análise Fatorial em dados econômicos contendo 50 observações e 13 variáveis numéricas relacionadas a setores da economia, como Agricultura, Mineração, Construção, Manufatura, entre outros. O objetivo principal foi reduzir a dimensionalidade do conjunto de dados, identificar padrões subjacentes e facilitar a interpretação dos fatores que explicam a maior parte da variabilidade nos dados originais.

Utilizamos métodos clássicos de PCA com cálculo da variância explicada e cargas fatoriais, além de aplicar rotações ortogonais (Varimax) e oblíquas (Promax) para melhorar a interpretabilidade dos componentes extraídos. A adequação dos dados para a análise fatorial foi avaliada por meio dos testes de Kaiser-Meyer-Olkin (KMO) e Bartlett, que indicam se a estrutura dos dados é adequada para essas técnicas.

Metodologia

O procedimento adotado consistiu nas seguintes etapas:

1. **Pré-processamento dos dados:** Seleção apenas das variáveis numéricas e análise estatística descritiva inicial para compreensão das distribuições e estrutura dos dados.
2. **Análise de Componentes Principais (PCA):** Extração dos componentes principais utilizando correlação entre variáveis e avaliação da variância acumulada para definir o número ideal de componentes.
3. **Rotações dos Componentes:** Aplicação das rotações Varimax (ortogonal) e Promax (obliqua) para facilitar a interpretação das cargas fatoriais.
4. **Análise Fatorial:** Estimação de um modelo com 3 fatores, utilizando rotação Varimax e cálculo dos scores fatoriais.
5. **Testes de Adequação:** Aplicação do teste KMO para medir a adequação amostral e do teste de esfericidade de Bartlett para verificar a existência de correlações significativas entre as variáveis.

Código em R e Análise

```
# --- Limpeza do ambiente
rm(list = ls())

# --- Definir diretório de trabalho
setwd("C:/Users/PC GAMER/Downloads/data")

# --- Carregar bibliotecas necessárias
library(REdaS)      # Para KMO e Bartlett
library(psych)      # Análise fatorial, rotação
library(ggplot2)    # Gráficos

# --- Carregar dados
dados <- read.csv("pca.csv")

# --- Selecionar apenas colunas numéricas
dados_num <- dados[sapply(dados, is.numeric)]

# --- Estrutura e resumo dos dados numéricos
str(dados_num)
```

```
## 'data.frame':    50 obs. of  13 variables:
## $ Ag      : num  2 1.5 1.7 5.1 2.1 ...
## $ Mining  : num  1.5 22.4 1.3 1 0.6 ...
## $ Constr  : num  4.2 4.1 5.8 4 3.3 ...
## $ Manuf   : num  10.5 1.1 11.5 12.8 9 ...
## $ Manuf_nd: num  11.8 3.7 3 11.8 5 ...
## $ Transp  : num  2.9 12.1 2.8 4.4 2.6 ...
## $ Comm    : num  2.9 2 2.2 2.4 2.5 ...
## $ Energy  : num  3.6 1.5 2.7 4.2 1.8 ...
## $ TradeW  : num  6.3 2.9 6.3 6.1 6.8 ...
## $ TradeR  : num  9.9 6.5 10.5 10.2 8.9 ...
## $ RE      : num  12.8 10.7 18.9 11.4 22.7 ...
## $ Services: num  16.1 11.9 20.2 14.8 23.1 ...
## $ Govt    : num  15.5 19.6 13 11.8 11.5 ...
```

```
summary(dados_num)
```

##	Ag	Mining	Constr	Manuf
## Min.	: 0.500	Min. : 0.000	Min. :2.900	Min. : 0.800
## 1st Qu.:	1.025	1st Qu.: 0.200	1st Qu.:3.825	1st Qu.: 6.250
## Median :	1.800	Median : 0.450	Median :4.200	Median :10.400
## Mean :	2.480	Mean : 2.624	Mean :4.338	Mean : 9.784
## 3rd Qu.:	2.525	3rd Qu.: 1.650	3rd Qu.:4.675	3rd Qu.:12.375

```
## Max. :10.600 Max. :31.600 Max. :8.400 Max. :21.400
## Manuf_nd Transp Comm Energy
## Min. : 1.700 Min. : 1.500 Min. :1.300 Min. :1.000
## 1st Qu.: 4.500 1st Qu.: 2.650 1st Qu.:1.900 1st Qu.:2.500
## Median : 7.150 Median : 3.200 Median :2.100 Median :2.950
## Mean : 7.696 Mean : 3.476 Mean :2.398 Mean :3.112
## 3rd Qu.:10.500 3rd Qu.: 3.875 3rd Qu.:2.875 3rd Qu.:3.600
## Max. :16.700 Max. :12.100 Max. :5.700 Max. :7.500
## TradeW TradeR RE Services
## Min. :2.900 Min. : 6.000 Min. :10.40 Min. : 9.60
## 1st Qu.:5.825 1st Qu.: 8.600 1st Qu.:13.15 1st Qu.:16.15
## Median :6.300 Median : 8.900 Median :16.20 Median :18.40
## Mean :6.348 Mean : 9.002 Mean :17.09 Mean :18.71
## 3rd Qu.:7.275 3rd Qu.: 9.850 3rd Qu.:19.15 3rd Qu.:20.77
## Max. :9.100 Max. :11.500 Max. :35.40 Max. :32.30
## Govt
## Min. : 9.00
## 1st Qu.:10.90
## Median :12.25
## Mean :12.93
## 3rd Qu.:14.55
## Max. :21.30
```

```
# --- Matriz de correlação (com observações completas)
cor_mat <- cor(dados_num, use = "complete.obs")
print(round(cor_mat, 2))
```

```
## Ag Mining Constr Manuf Manuf_nd Transp Comm Energy TradeW TradeR
## Ag 1.00 -0.06 0.08 0.03 -0.15 0.28 -0.18 0.04 0.25 0.09
## Mining -0.06 1.00 -0.02 -0.42 -0.14 0.61 -0.19 0.39 -0.55 -0.40
## Constr 0.08 -0.02 1.00 -0.13 -0.32 0.08 -0.02 0.01 -0.09 0.40
## Manuf 0.03 -0.42 -0.13 1.00 0.20 -0.36 -0.32 -0.05 0.27 0.19
## Manuf_nd -0.15 -0.14 -0.32 0.20 1.00 -0.18 -0.10 0.07 0.04 -0.12
## Transp 0.28 0.61 0.08 -0.36 -0.18 1.00 -0.05 -0.06 -0.21 -0.15
## Comm -0.18 -0.19 -0.02 -0.32 -0.10 -0.05 1.00 -0.17 0.33 0.12
## Energy 0.04 0.39 0.01 -0.05 0.07 -0.06 -0.17 1.00 -0.27 0.03
## TradeW 0.25 -0.55 -0.09 0.27 0.04 -0.21 0.33 -0.27 1.00 0.17
## TradeR 0.09 -0.40 0.40 0.19 -0.12 -0.15 0.12 0.03 0.17 1.00
## RE -0.30 -0.41 -0.25 -0.18 -0.13 -0.50 0.12 -0.38 0.04 -0.31
## Services -0.32 -0.46 0.32 -0.16 -0.46 -0.42 0.31 -0.31 0.24 0.20
## Govt 0.11 0.23 0.18 -0.41 -0.24 0.43 0.19 0.05 -0.34 0.29
## RE Services Govt
## Ag -0.30 -0.32 0.11
## Mining -0.41 -0.46 0.23
## Constr -0.25 0.32 0.18
## Manuf -0.18 -0.16 -0.41
## Manuf_nd -0.13 -0.46 -0.24
```

```
## Transp    -0.50    -0.42    0.43
## Comm      0.12     0.31    0.19
## Energy    -0.38    -0.31    0.05
## TradeW    0.04     0.24   -0.34
## TradeR    -0.31     0.20    0.29
## RE        1.00     0.52   -0.35
## Services  0.52     1.00   -0.18
## Govt      -0.35    -0.18    1.00
```

```
# --- Análise de Componentes Principais (PCA)
pca_modelo <- princomp(dados_num, scores = TRUE, cor = TRUE)

# --- Variância explicada por componente
summary(pca_modelo)
```

```
## Importance of components:
##
##              Comp.1    Comp.2    Comp.3    Comp.4    Comp.5
## Standard deviation  1.7987525  1.4954801  1.3999420  1.1663403  1.07583525
## Proportion of Variance 0.2488854  0.1720354  0.1507567  0.1046423  0.08903242
## Cumulative Proportion 0.2488854  0.4209209  0.5716776  0.6763199  0.76535232
##
##              Comp.6    Comp.7    Comp.8    Comp.9    Comp.10
## Standard deviation  0.93184458  0.85116719  0.78471605  0.5641253  0.4851322
## Proportion of Variance 0.06679495  0.05572966  0.04736764  0.0244798  0.0181041
## Cumulative Proportion 0.83214726  0.88787692  0.93524456  0.9597244  0.9778285
##
##              Comp.11    Comp.12    Comp.13
## Standard deviation  0.38943836  0.36945813  8.279806e-03
## Proportion of Variance 0.01166633  0.01049995  5.273476e-06
## Cumulative Proportion 0.98949478  0.99999473  1.000000e+00
```

```
# --- Cargas fatoriais dos componentes
loadings(pca_modelo)
```

```
##
## Loadings:
##      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10
## Ag      0.135      0.385  0.373  0.411  0.245  0.433  0.277  0.152  0.217
## Mining  0.470      -0.260      0.164 -0.276 -0.148      0.116
## Constr      -0.393  0.257 -0.350  0.196      -0.370  0.499  0.371
## Manuf    -0.183  0.376  0.375 -0.147  0.111 -0.198 -0.151 -0.500  0.387
## Manuf_nd      0.459      -0.465 -0.217 -0.102  0.592 -0.102
## Transp   0.418 -0.147      0.365  0.143 -0.169 -0.302      -0.264  0.108
## Comm    -0.152 -0.316      0.343 -0.550  0.267 -0.145 -0.101  0.436  0.383
## Energy   0.247  0.138      -0.416 -0.202  0.689  0.199      -0.116
## TradeW   -0.315      0.290  0.442      0.353 -0.254      -0.253 -0.455
## TradeR      -0.261  0.507 -0.227 -0.252 -0.143  0.146 -0.100 -0.515  0.432
## RE       -0.363      -0.447      0.173      0.359  0.106      0.169
```

```

## Services -0.380 -0.384 -0.127 -0.183  0.125  0.101 -0.133          -0.222 -0.204
## Govt      0.289 -0.369          -0.295 -0.306  0.428 -0.121  0.171 -0.548
##          Comp.11 Comp.12 Comp.13
## Ag              0.286  0.206
## Mining -0.475    0.282  0.500
## Constr -0.141 -0.256
## Manuf    0.138          0.406
## Manuf_nd 0.142  0.122  0.338
## Transp   0.507 -0.407  0.144
## Comm
## Energy   0.282 -0.286
## TradeW   -0.322 -0.210  0.112
## TradeR   -0.181          0.106
## RE        -0.127 -0.496  0.451
## Services  0.458  0.460  0.320
## Govt              0.238
##
##          Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
## SS loadings    1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
## Proportion Var 0.077  0.077  0.077  0.077  0.077  0.077  0.077  0.077  0.077
## Cumulative Var 0.077  0.154  0.231  0.308  0.385  0.462  0.538  0.615  0.692
##          Comp.10 Comp.11 Comp.12 Comp.13
## SS loadings    1.000  1.000  1.000  1.000
## Proportion Var 0.077  0.077  0.077  0.077
## Cumulative Var 0.769  0.846  0.923  1.000

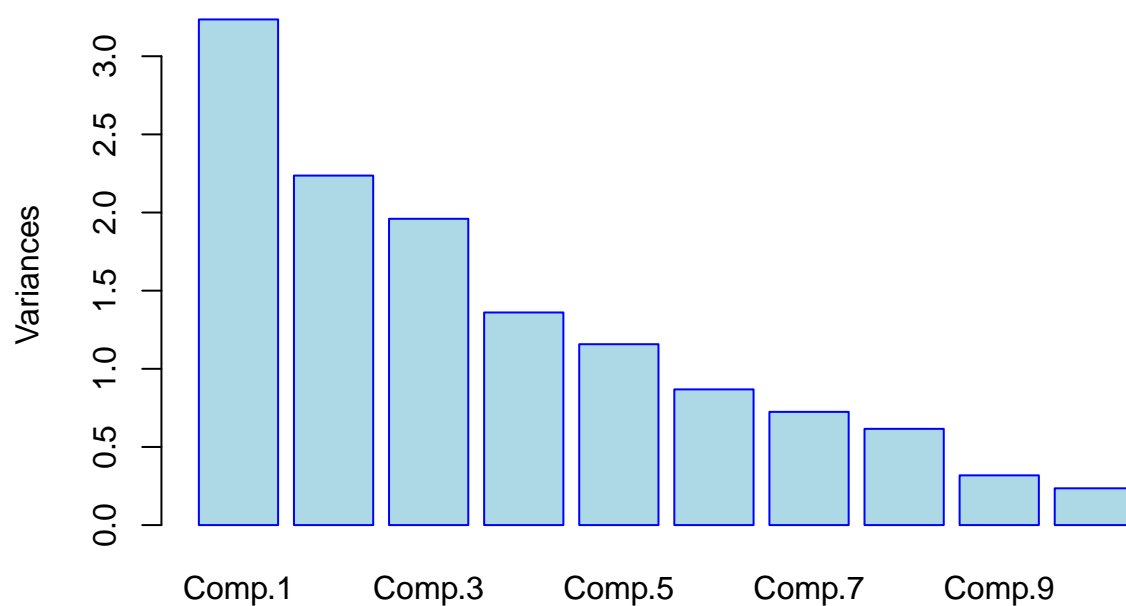
```

```

# --- Scree plot para visualização dos autovalores
screeplot(pca_modelo, type = "barplot", main = "Gráfico Scree - Autovalores",
          col = "lightblue", border = "blue")

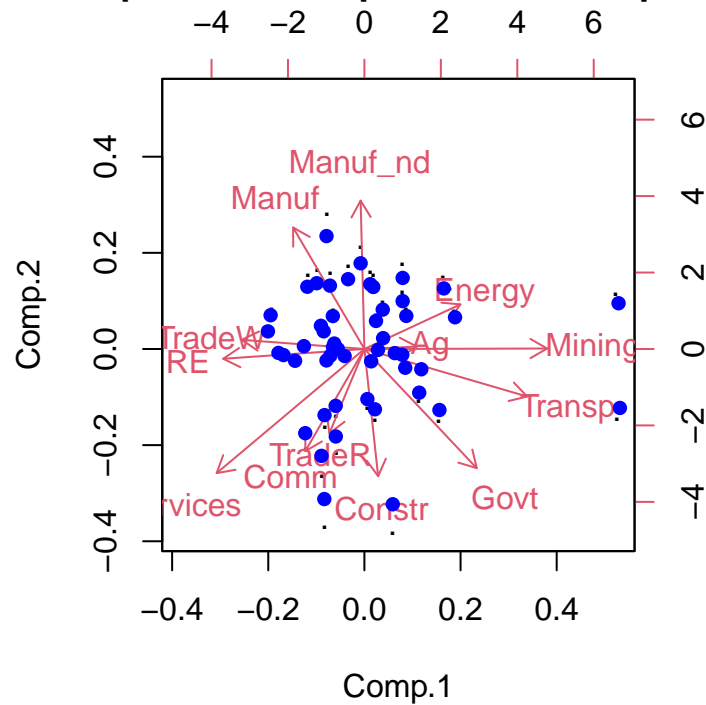
```

Gráfico Scree – Autovalores



```
# --- Biplot dos dois primeiros componentes principais
biplot(pca_modelo, choices = 1:2, main = "Biplot dos Componentes Principais",
       xlab = rep(".", nrow(dados_num)))
points(pca_modelo$scores[1:50, 1], pca_modelo$scores[1:50, 2], pch = 16, col = "blue")
```

Biplot dos Componentes Principais



```
# --- Scores das primeiras 10 observações
head(round(pca_modelo$scores, 3), 10)
```

```
##      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10
## [1,]  0.490  0.284  0.921 -0.088 -1.754 -0.257  0.312  0.133  0.167 -0.128
## [2,]  6.685 -1.542 -2.725  2.094  0.962 -2.426 -1.242 -0.835 -0.026  0.033
## [3,] -0.752 -1.488  0.864 -1.239  0.701 -0.402 -0.257 -0.173  0.132  0.332
## [4,]  1.000  1.253  1.797  0.160 -0.629  0.311  0.280  0.252 -0.270  0.914
## [5,] -1.814 -0.308 -1.073  0.621  0.587 -0.193  0.377 -0.685 -0.602  0.150
## [6,] -1.123 -2.797 -0.127  0.834 -1.592  0.581 -1.043 -0.162  1.599  1.525
## [7,] -2.454  0.884 -2.261  0.177  1.015  0.097  0.216 -0.535  0.067  0.096
## [8,] -0.992  2.954 -4.026 -0.220  0.478 -1.240  1.563  2.347  0.387  0.469
## [9,] -1.548 -2.203  0.143 -0.171 -0.147  0.468  0.188 -0.052 -1.270  0.645
## [10,] -0.992 -0.300  0.055  2.181 -1.876  0.877 -0.943  0.163  0.116  0.085
##      Comp.11 Comp.12 Comp.13
## [1,]  -0.141   0.118  -0.002
## [2,]   0.606  -0.425   0.004
## [3,]  -0.676  -0.535  -0.006
## [4,]   0.628  -0.031   0.003
## [5,]  -0.050   0.578  -0.007
## [6,]  -0.021   0.203  -0.003
## [7,]  -0.096  -0.110   0.000
```

```
## [8,] -0.248 -0.842  0.000
## [9,] -0.421 -0.296 -0.008
## [10,] -0.143 -0.318 -0.002
```

```
# --- Rotações para melhor interpretação dos componentes
```

```
cat("Rotação Varimax:\n")
```

```
## Rotação Varimax:
```

```
varimax_rot <- varimax(pca_modelo$loadings[, 1:3])
print(varimax_rot)
```

```
## $loadings
##
## Loadings:
##          Comp.1 Comp.2 Comp.3
## Ag          0.179  0.361
## Mining    0.511  0.144
## Constr          -0.240  0.404
## Manuf    -0.439  0.338
## Manuf_nd -0.164  0.391 -0.179
## Transp    0.392          0.195
## Comm          -0.359
## Energy    0.133  0.250
## TradeW   -0.405          0.139
## TradeR   -0.227 -0.115  0.518
## RE          -0.321 -0.470
## Services -0.142 -0.534
## Govt      0.315 -0.151  0.323
##
##          Comp.1 Comp.2 Comp.3
## SS loadings    1.000  1.000  1.000
## Proportion Var  0.077  0.077  0.077
## Cumulative Var  0.077  0.154  0.231
##
## $rotmat
##          [,1]      [,2]      [,3]
## [1,]  0.8361463  0.4624026  0.2950307
## [2,] -0.3050432  0.8390411 -0.4505093
## [3,] -0.4558596  0.2866946  0.8426139
```

```
cat("Rotação Promax:\n")
```

```
## Rotação Promax:
```



```
promax_rot <- promax(pca_modelo$loadings[, 1:3])
print(promax_rot)
```

```
## $loadings
##
## Loadings:
##      Comp.1 Comp.2 Comp.3
## Ag          0.199  0.378
## Mining      0.522  0.164 -0.131
## Constr      -0.213  0.401
## Manuf       -0.458  0.323  0.148
## Manuf_nd    -0.158  0.372 -0.156
## Transp      0.380  0.104  0.160
## Comm        -0.357
## Energy      0.125  0.261
## TradeW     -0.421          0.181
## TradeR     -0.271          0.547
## RE          -0.356 -0.476
## Services   -0.132 -0.545
## Govt        0.295 -0.115  0.292
##
##      Comp.1 Comp.2 Comp.3
## SS loadings  1.036  1.009  1.042
## Proportion Var 0.080  0.078  0.080
## Cumulative Var 0.080  0.157  0.237
##
## $rotmat
##      [,1]      [,2]      [,3]
## [1,] 0.8165019 0.5226146 0.2251165
## [2,] -0.2859022 0.7963844 -0.4050635
## [3,] -0.5361016 0.3184902 0.9096131
```

```
# --- Análise Fatorial com 3 fatores e rotação Varimax
fa_modelo <- factanal(dados_num, factors = 3, rotation = "varimax", scores = "regression")
print(fa_modelo)
```

```
##
## Call:
## factanal(x = dados_num, factors = 3, scores = "regression", rotation = "varimax")
##
## Uniquenesses:
##      Ag      Mining      Constr      Manuf Manuf_nd      Transp      Comm      Energy
##      0.863      0.005      0.550      0.518      0.533      0.536      0.874      0.782
##      TradeW      TradeR      RE Services      Govt
##      0.653      0.420      0.005      0.284      0.755
##
```

```
## Loadings:
##          Factor1 Factor2 Factor3
## Ag          0.154   0.335
## Mining       0.825  -0.551  -0.108
## Constr       0.116           0.656
## Manuf       -0.117   0.641  -0.238
## Manuf_nd           0.401  -0.552
## Transp       0.659  -0.144
## Comm        -0.198           0.294
## Energy       0.461
## TradeW      -0.344   0.471
## TradeR           0.527   0.547
## RE          -0.847  -0.498  -0.171
## Services    -0.613  -0.191   0.551
## Govt         0.337           0.363
##
##          Factor1 Factor2 Factor3
## SS loadings    2.747   1.799   1.674
## Proportion Var  0.211   0.138   0.129
## Cumulative Var  0.211   0.350   0.479
##
## Test of the hypothesis that 3 factors are sufficient.
## The chi square statistic is 393.94 on 42 degrees of freedom.
## The p-value is 1.01e-58
```

```
# --- Scores fatoriais para as 10 primeiras observações
head(fa_modelo$scores, 10)
```

```
##          Factor1   Factor2   Factor3
## [1,]  0.4087868  0.9692599 -0.06812808
## [2,]  2.7145736 -2.0939938 -0.18119779
## [3,] -0.3857527 -0.3275125  0.91811013
## [4,]  0.5451977  1.3786039 -0.35350998
## [5,] -0.8789713 -0.6820351  0.05649415
## [6,] -0.1433143 -0.1753478  1.30233201
## [7,] -1.5477407 -1.3044088 -0.92444286
## [8,] -2.2911314 -2.0343099 -3.02983078
## [9,] -0.8551863 -0.7367485  1.18818263
## [10,] -0.1326606  0.5084247 -0.07918391
```

```
# --- Testes de adequação
# KMO
kmo_result <- KMOS(dados_num)
print(kmo_result)
```

```
##
```

```
## Kaiser-Meyer-Olkin Statistics
##
## Call: KMOS(x = dados_num)
##
## Measures of Sampling Adequacy (MSA):
##      Ag      Mining      Constr      Manuf      Manuf_nd      Transp      Comm      Energy
## 0.0344615 0.1225392 0.0412464 0.0636196 0.0425694 0.0996285 0.0400061 0.0415722
##      TradeW      TradeR      RE      Services      Govt
## 0.0696772 0.0535185 0.0966035 0.1073965 0.0705293
##
## KMO-Criterion: 0.0689331
```

```
# Bartlett
bartlett_result <- bart_spher(dados_num)
print(bartlett_result)
```

```
## Bartlett's Test of Sphericity
##
## Call: bart_spher(x = dados_num)
##
##      X2 = 609.261
##      df = 78
## p-value < 2.22e-16
```