UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ADMINISTRAÇÃO

# ANÁLISE DE COMPONENTES PRINCIPAIS, ANÁLISE FATORIAL: Exemplos em STATA.

**Prof. Dr. Evandro Marcos Saidel Ribeiro**

**RESUMO**

**A1 – Análise de Componentes Principais**

Atualmente, pesquisadores defrontam-se com dezenas ou centenas de variáveis diferentes em suas análises. É frequente haver redundância entre diversas dimensões, levando a problemas de multicolinearidade.

A análise de componentes principais é um método que explora a interdependência em dados multivariados. Se houver redundância substancial no conjunto de dados, pode ser possível explicar a maior parte das informações num conjunto menor de dimensões.

**A1.1 Covariância e Correlação**

**Matriz de Covariância**

Organize os dados na forma de matriz, por exemplo, na coluna 1 liste as 5 observações da variável $x_1$ e na coluna 2 liste as 5 observações da variável $x_2$:

$$\mathbf{X} = \begin{bmatrix} 2 & 1 \\ 9 & 8 \\ 7 & 5 \\ 3 & 4 \\ 4 & 2 \end{bmatrix}. \tag{1}$$

Calcule as médias das variáveis:

$$\bar{x}_1 = 5 \qquad \bar{x}_2 = 4\,. \tag{2}$$

Utilize a matriz obtida em (1) e, para cada coluna subtraia a média, obtendo a matriz de desvios $\mathbf{X}_{\text{desv}}$:

$$\mathbf{X}_{\text{desv}} = \begin{bmatrix} 2 & 1 \\ 9 & 8 \\ 7 & 5 \\ 3 & 4 \\ 4 & 2 \end{bmatrix} - \begin{bmatrix} 5 & 4 \\ 5 & 4 \\ 5 & 4 \\ 5 & 4 \\ 5 & 4 \end{bmatrix} = \begin{bmatrix} -3 & -3 \\ 4 & 4 \\ 2 & 1 \\ -2 & 0 \\ -1 & -2 \end{bmatrix}. \tag{3}$$

A matriz de covariância, $\mathbf{S}$, é obtida pela expressão:

$$\mathbf{S} = \frac{1}{n-1}\mathbf{X}'_{\text{desv}}\mathbf{X}_{\text{desv}} \tag{4}$$

- $n$ é o número de observações, que no nosso exemplo é igual a 5;
- $\mathbf{X}'_{\text{desv}}$ é a matriz transposta da matriz obtida em (3). Então, para o exemplo dado a matriz $\mathbf{S}$ é:

$$\mathbf{S} = \frac{1}{4}\begin{bmatrix} -3 & 4 & 2 & -2 & -1 \\ -3 & 4 & 1 & 0 & -2 \end{bmatrix}\begin{bmatrix} -3 & -3 \\ 4 & 4 \\ 2 & 1 \\ -2 & 0 \\ -1 & -2 \end{bmatrix},$$

que leva a: $\mathbf{S} = \frac{1}{4}\begin{bmatrix} 34 & 29 \\ 29 & 30 \end{bmatrix}$, ou seja

$$\mathbf{S} = \begin{bmatrix} 8{,}50 & 7{,}25 \\ 7{,}25 & 7{,}50 \end{bmatrix}. \tag{5}$$

Note que para duas variáveis a matriz de covariância será 2x2:

$$\mathbf{S} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}. \tag{6}$$

**Matriz de correlação**

A matriz de correlação é escrita na forma:

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1k} \\ r_{21} & 1 & \cdots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \cdots & 1 \end{bmatrix}. \tag{7}$$

O cálculo de cada elemento desta matriz pode ser obtido da matriz $\mathbf{S}$ por:

$$r_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}}\sqrt{S_{jj}}}\,. \tag{8}$$

A matriz R também é simétrica, ou seja, $r_{ij} = r_{ji}$.

Para matrizes 2x2:

$$\mathbf{R}=\begin{bmatrix} 1 & \frac{S_{12}}{\sqrt{S_{11}}\sqrt{S_{22}}} \\ \frac{S_{12}}{\sqrt{S_{11}}\sqrt{S_{22}}} & 1 \end{bmatrix}. \quad (9)$$

No exemplo, considerando (1) e (5), a matriz de correlação será:

$$\mathbf{R}=\begin{bmatrix} 1 & 0{,}908025 \\ 0{,}908025 & 1 \end{bmatrix}. \quad (10)$$

**A1.2 Autovalores e Autovetores**

**1º passo**: Determinar autovalores. Para a matriz de covariância $\mathbf{S}$ os autovalores são:

$$\lambda=\frac{(s_{11}+s_{22})}{2}\pm\frac{\sqrt{(s_{11}+s_{22})^2-4(s_{11}s_{22}-s_{12}^2)}}{2} \quad (11)$$

No caso específico da matriz $\mathbf{S}$ obtida em (5), os autovalores são:

$$\lambda_1=15{,}267221 \qquad \lambda_2=0{,}732779\ . \quad (12)$$

Os autovalores da matriz de correlação $\mathbf{R}$ são:

$$\lambda=1\pm r_{12}\,. \quad (13)$$

No caso específico da matriz $\mathbf{R}$ obtida em (10), os autovalores são:

$$\lambda_1=1{,}908025 \qquad \lambda_2=0{,}091975\ . \quad (14)$$

**2º Passo**: Determinar os autovetores da matriz de covariância.

No caso de uma matriz 2x2 deve-se resolver o sistema:

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}\begin{bmatrix} v_{1i} \\ v_{2i} \end{bmatrix}=\lambda_i\begin{bmatrix} v_{1i} \\ v_{2i} \end{bmatrix} \quad (15)$$

para cada autovalor. Para resolver o sistema podemos considerar $v_{ii}=1$.

- Considerando o primeiro autovalor $\lambda_1$:

$$v_{11}=1, \quad (16\text{-a})$$

então $v_{21}$ é obtida pela expressão:

$$v_{21}=\frac{s_{11}+s_{21}-\lambda_1}{\lambda_1-s_{22}-s_{12}}. \quad (16\text{-b})$$

o resultado é:

$$\begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix}=\begin{bmatrix} 1 \\ 0{,}93341 \end{bmatrix} \quad (16\text{-c})$$

- Considerando o segundo autovalor $\lambda_2$:

$$v_{22}=1, \quad (17\text{-a})$$

então $v_{12}$ é obtida pela expressão:

$$v_{12}=\frac{\lambda_2-s_{22}-s_{12}}{s_{11}+s_{21}-\lambda_2}. \quad (17\text{-b})$$

o resultado é:

$$\begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix}=\begin{bmatrix} -0{,}93341 \\ 1 \end{bmatrix} \quad (17\text{-c})$$

**3º Passo**: Normalizar os autovetores $\mathbf{e}$: Divida cada componente dos autovetores obtidos pelo seu módulo. No caso, o módulo dos dois vetores obtidos é iguala a 1,367938. O resultado obtido é:

$$\begin{bmatrix} e_{11} \\ e_{21} \end{bmatrix}=\begin{bmatrix} 0{,}731027 \\ 0{,}682348 \end{bmatrix} \quad (18\text{-a})$$

$$\begin{bmatrix} e_{12} \\ e_{22} \end{bmatrix}=\begin{bmatrix} -0{,}682348 \\ 0{,}731027 \end{bmatrix} \quad (18\text{-b})$$

Os vetores normalizados apresentados nas expressões (18-a) e (18-b) acima são autovetores da matriz de covariância.

**A1.3 Componentes Principais**

Componentes Principais são escritos em termos das variáveis originais através de ponderações, dadas pelas componentes dos autovetores $\mathbf{e}$. Se tivermos duas variáveis originais teremos dois componentes principais: $\mathbf{CP}=\mathbf{e}'\,\mathbf{X}$

Sendo, $\mathbf{e}=\begin{bmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{bmatrix}$ ou seja $\mathbf{e}'=\begin{bmatrix} e_{11} & e_{21} \\ e_{12} & e_{22} \end{bmatrix}$.

Assim, $\begin{bmatrix} CP_1 \\ CP_2 \end{bmatrix}=\begin{bmatrix} e_{11} & e_{21} \\ e_{12} & e_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ e então:

$$\begin{aligned} CP_1 &= e_{11}x_1+e_{21}x_2 \\ CP_2 &= e_{12}x_1+e_{22}x_2 \end{aligned}\ . \quad (19)$$

No exemplo específico, utilizando os resultados obtidos para os autovetores, temos:

$$\begin{aligned} CP_1 &= 0{,}731027\ x_1+0{,}682348\ x_2 \\ CP_2 &= -0{,}682348\ x_1+0{,}731027\ x_2 \end{aligned} \quad (20)$$

**A1.4 Variância explicada**

O Percentual da variação explicada é dado por

$$CP_i=\frac{\lambda_i}{\sum\lambda}\times 100\,. \quad (21)$$

No exemplo em questão, considerando os autovalores (12), obtemos:

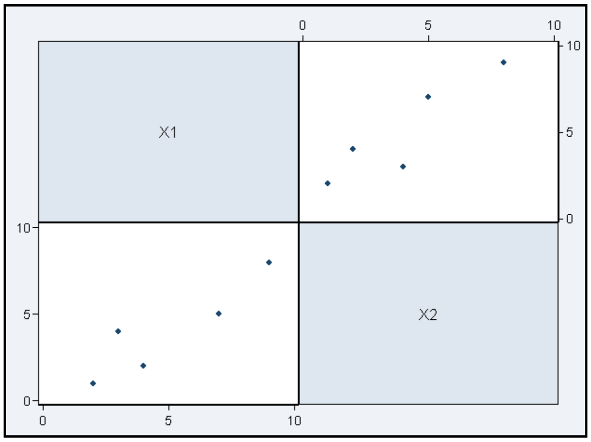
Variação explicada por $CP_1$ = 95,42 %, (22-a)

Variação explicada por $CP_2$ = 4,58 %. (22-b)

### A1.5 PCA: Exemplo no STATA

Após inserir os dados do exemplo (Equação 1) no Stata, para obter gráficos de dispersão, considere o link: Graphics ▶ Scatterplot Matrix. O resultado é presentado na Figura 1.

**Figura 1**. Scatterplot Matrix para variáveis X1 e X2, obtido no STATA.




Neste exemplo observa-se uma correlação linear positiva entre as variáveis X1 e X2.

**Covariância e Correlação – Comando no STATA:**

Statistics ▶ Summaries, tables and tests
▶ Summary and descriptive statistics
▶ Correlation and Covariances

Os resultados de covariância e de correlação são apresentados na Figura 2 a seguir.

**Figura 2**. Covariância e Correlação calculados no STATA.

. **correlate, covariance**
(obs=5)

| | X1 | X2 |
|---|---|---|
| X1 | **8.5** | |
| X2 | **7.25** | **7.5** |

. **correlate**
(obs=5)

| | X1 | X2 |
|---|---|---|
| X1 | **1.0000** | |
| X2 | **0.9080** | **1.0000** |

Observe que os valores apresentados para a matriz de covariância (correlate, covariance na Figura 2) e os valores apresentados para a matriz de correlação (correlate, na Figura 2) são os mesmos apresentados nas expressões (5) e (10), respectivamente.

**Componentes Principais – Comando no STATA**

Statistics ▶ Multivariate analysis
▶ Factor and principal componentes analysis
▶ Principal componentes analysis (PCA)

Nas Figuras 3 e 4 a seguir são apresentados os resultados da PCA considerando a matriz de correlação (Figura 3) e a matriz de covariância (Figura 4).

**Figura 3**. Resultado PCA, matriz de correlação.

| Component | Eigenvalue | Differe |
|---|---|---|
| Comp1 | **1.90803** | **1.** |
| Comp2 | **.0919748** | |

Principal components (eigenvectors)

| Variable | Comp1 | Comp2 |
|---|---|---|
| X1 | **0.7071** | **0.7071** |
| X2 | **0.7071** | **-0.7071** |

**Figura 4**. Resultado PCA, matriz de covariância.

| Component | Eigenvalue | Differe |
|---|---|---|
| Comp1 | **15.2672** | **14.5** |
| Comp2 | **.732779** | |

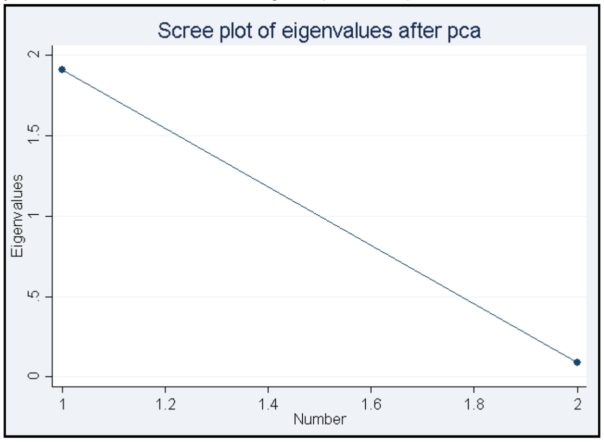
Principal components (eigenvectors)

| Variable | Comp1 | Comp2 | U |
|---|---|---|---|
| X1 | **0.7310** | **-0.6823** | |
| X2 | **0.6823** | **0.7310** | |

Observe que o resultado PCA calculado para matriz de covariância (Figura 4) é equivalente ao resultado obtido em (20).

**Gráfico Scree plot**:
Statistics ▶ Multivariate analysis
▶ Factor and principal componentes analysis
▶ Postestimation

**Figura 5**. Gráfico Scree plot para PCA calculada pela matriz de correlação (STATA)




**A1.E – Exercícios de Componentes Principais**

**A1.E1 Produto Estadual Bruto, nos estados americanos. (Gross State Product – GSP).**
Dados sobre produto estadual bruto estão disponíveis em dois arquivos: gsp_raw (expresso em milhões de dólares) e gsp_share (expresso como uma proporção do total do GSP para cada estado). Cada arquivo consiste em 50 linhas (uma para cada estado) e 14 colunas (a primeira é a identidade do estado e as 13 restantes correspondendo às diferentes áreas de atividade econômica).
A1.E1.a Usando os dados do gsp_share faça uma análise de componentes principais.

A1.E1.b O que acontecerá se você remover as observações correspondentes ao Alasca e Wyoming? Repita a análise e esquematize os scores para os primeiros dois componentes principais. Como você interpreta os resultados?

**Respostas, A1.E1:**
Com a base gsp.raw:
**. pca col1 col2 col3 col4 col5 col6 col7 col8 col9 col10 col11 col12 col13**

**Tabela 1.** Autovalores da matriz de correlação para os dados gsp_raw

| Component | Eigenvalue | Difference |
|---|---|---|
| Comp1 | 10.9443 | 9.96488 |
| Comp2 | .979435 | .579379 |
| Comp3 | .400057 | .0573802 |
| Comp4 | .342677 | .203464 |
| Comp5 | .139213 | .0698184 |
| Comp6 | .0693941 | .0292465 |
| Comp7 | .0401476 | .00680468 |
| Comp8 | .0333429 | .00825054 |
| Comp9 | .0250924 | .0145441 |
| Comp10 | .0105483 | .00305999 |
| Comp11 | .00748827 | .00143102 |
| Comp12 | .00605725 | .00382676 |
| Comp13 | .00223049 | . |

**Tabela 2.** Autovalores da matriz de correlação para os dados gsp.share.

| Component | Eigenvalue | Difference |
|---|---|---|
| Comp1 | 3.23551 | .99905 |
| Comp2 | 2.23646 | .276623 |
| Comp3 | 1.95984 | .599488 |
| Comp4 | 1.36035 | .202928 |
| Comp5 | 1.15742 | .289087 |
| Comp6 | .868334 | .143849 |
| Comp7 | .724486 | .108706 |
| Comp8 | .615779 | .297542 |
| Comp9 | .318237 | .0828841 |
| Comp10 | .235353 | .083691 |
| Comp11 | .151662 | .0151629 |
| Comp12 | .136499 | .136431 |
| Comp13 | .0000685552 | . |

Nas Tabelas 1 e 2 são pode-se ver a diferença entre o cálculo com os dados do arquivo gsp_raw (formado com valores absolutos do produto estadual bruto) e o cálculo com os dados do arquivo gsp_share (formado com os dados proporcionais para cada estado. As Figuras 6 e 7 apresentam os gráficos Scree plot para os cálculos comas bases gsp_raw e gsp_share. Após o cálculo, digite **.screeplot** para obter o gráfico ScreePlot.

**Figura 6**. Gráfico Scree plot para PCA calculada pela matriz de correlação gsp_raw.

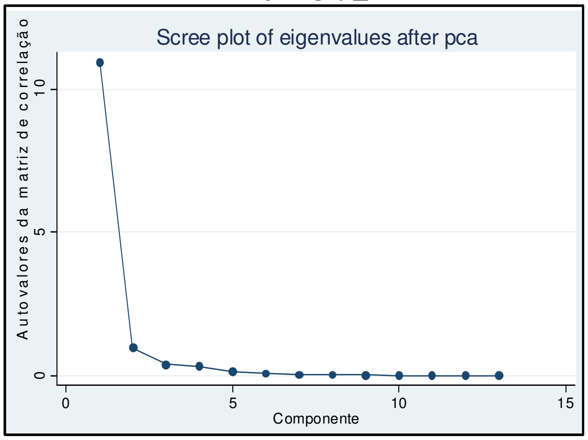
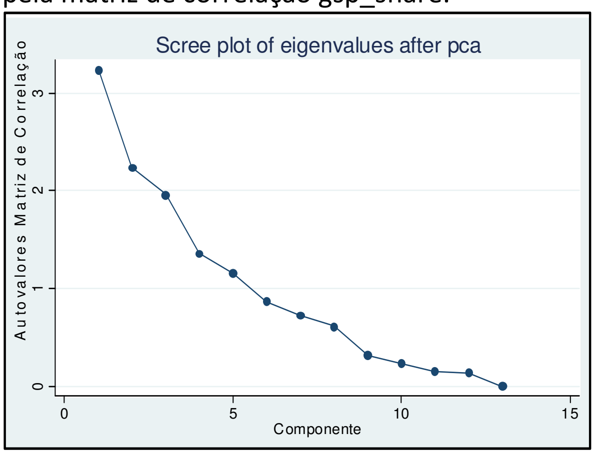



**Figura 7**. Gráfico Scree plot para PCA calculada pela matriz de correlação gsp_share.




Para obter os scores como novas variáveis devemos utilizar o comando **predict** e depois o nome das componentes. Por exemplo para criar as primeiras duas componentes e gravá-las como novas variáveis podemos digitar:
**predict pc1 pc2, score**

**A1.E2 Opinião Política.**
Um pesquisador coletou informações sobre 100 respondentes acerca de seis variáveis de opinião política apresentadas a seguir:

x1 O governo deveria investir mais dinheiro em escolas
x2 O governo deveria investir mais dinheiro para reduzir o desemprego
x3 O governo deveria controlar os grandes negócios
x4 O governo deveria acelerar o fim da discriminação racial através de transporte escolar
x5 O governo deveria zelar para que as minorias obtenham suas respectivas quotas de emprego
x6 O governo deveria expandir o programa Head Start (www.acf.hhs.gov/programs/ohs)

A matriz de correlação para esses dados é apresentada na Tabela 1

**Tabela 1**. Matriz de correlação para os dados de opinião política.

| | x1 | x2 | x3 | x4 | x5 | x6 |
|---|---|---|---|---|---|---|
| x1 | 1.00 | | | | | |
| x2 | .601 | 1.00 | | | | |
| x3 | .498 | .475 | 1.00 | | | |
| x4 | .192 | .220 | .208 | 1.00 | | |
| x5 | .196 | .191 | .201 | .433 | 1.00 | |
| x6 | .347 | .298 | .245 | .320 | .421 | 1.00 |

A1.E2.a Analise esses dados usando a análise de componentes principais. Quanta variação é explicada pelos dois primeiros componentes principais? Como você interpretaria os dois componentes?

**Respostas, A1.E2:**

Transformando os dados em matriz no STATA:
mkmat x1 x2 x3 x4 x5 x6, matrix(MatrizOp)

Obtendo a matriz completa:
matrix Opinion = MatrizOp + MatrizOp' - I(6)

Nomes das linhas e colunas da matriz:
matrix rownames MatrizOp = x1 x2 x3 x4 x5 x6
matrix colnames MatrizOp = x1 x2 x3 x4 x5 x6

Observando os valores da matriz:
matrix list MatrizOp

Componentes Principais:
pcamat Opinion, n(6)
Resultados:

**Figura 8**. Resultado da pca a partir da matriz de correlação pelo comando **pcamat**.

```
Principal components/correlation

    Rotation: (unrotated = principal)
```

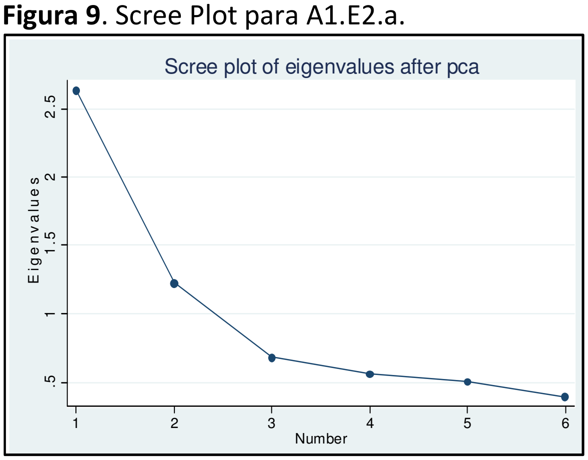
| Component | Eigenvalue | Difference |
|---|---|---|
| Comp1 | 2.63335 | 1.4067 |
| Comp2 | 1.22665 | .543098 |
| Comp3 | .683555 | .123086 |
| Comp4 | .560469 | .0555001 |
| Comp5 | .504969 | .113962 |
| Comp6 | .391007 | . |

```
Principal components (eigenvectors)
```

| Variable | Comp1 | Comp2 | Comp3 |
|---|---|---|---|
| x1 | 0.4629 | -0.3701 | -0.1216 |
| x2 | 0.4532 | -0.3647 | 0.0356 |
| x3 | 0.4202 | -0.3334 | 0.3075 |
| x4 | 0.3415 | 0.4846 | 0.6146 |
| x5 | 0.3544 | 0.5463 | -0.0333 |
| x6 | 0.4019 | 0.2925 | -0.7145 |

O Scree Plot obtido é apresentado na Figura 9 a seguir.

**Figura 9**. Scree Plot para A1.E2.a.




Pelo resultado observa-se que a variância explicada pelos dois primeiros componentes é 64,33%, ou seja:

$$(2.63335 + 1.22665)/6 = 0.643333$$

**A2 Análise Fatorial**

**A2.1 Características gerais**

O objetivo é descrever um **conjunto de *k* variáveis originais** ($x_1$, $x_2$, ... $x_k$) através da criação de um número menor de variáveis (**fatores**).

Cada variável é descrita em termos de fatores:

$$x_i = a_{i1}F_1 + a_{i2}F_2 + ... + a_{in}F_n \qquad (23)$$

$a_{ij}$ são cargas fatoriais.

A soma das cargas ao quadrado, para uma determinada variável é igual a comunalidade da variável:

$$h_i^2 = \sum_{j=1}^{n} a_{ij}^2 \qquad (24)$$

Uma regra útil para considerar uma variável na análise é que a comunalidade deve ser maior do que 0,7.

A medida de unicidade (uniqueness) é dada por um menos a comunalidade, ou seja:

$$u_i = 1 - h_i^2 \qquad (25)$$

Se a soma das cargas fatoriais (23) ao quadrado, for feita para um índice relacionado a um fator, então obtemos o autovalor correspondente ao fator em questão, ou seja

$$\lambda_j = \sum_{i=1}^{k} a_{ij}^2 \qquad (26)$$

Como resultado, cada Fator pode ser escrito em termos das variáveis sendo possível assim obter uma função para o Score Fatorial *j*:

$$F_j = b_{j1}z_1 + b_{j2}z_2 + ... + b_{jk}z_k \qquad (27)$$

Para obter o cálculo correto do Score Fatorial as variáveis devem ser padronizadas.

Assim, em (27) $z_i = (x_i - \bar{x}_i)/s_i$ é o Score-z.

**A2.2 Exemplo Completo no STATA**

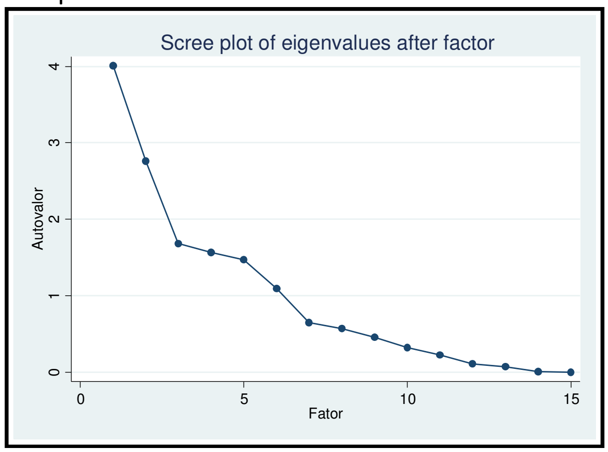
Como exemplo é considerado o "Exemplo Completo" apresentado no livro "Análise Multivariada para cursos de Administração, Ciências Contábeis e Economia" (Corrar, Paulo e Dias Filho).

A base de dados está disponível no arquivo **Exemplo_CAP02_CORRAR.dta**

Os passos para a Análise Fatorial são os seguintes:

- **factor ICOM-ILGE, pcf**. Este comando executa a análise fatorial a partir de componentes principais. Utiliza todas as variáveis, desde ICOM até ILGE.
- **screeplot, ytitle(Autovalor) xtitle(Fator)**. Com este comando obtemos o screeplot.

**Figura 10**. Scree plot obtido no exemplo completo com 15 variáveis.




- **estat kmo**. Um comando *postestimation* importante é o "estat kmo". Com este comando podemos analisar a Medida de Adequação da Amostra (MSA – Measure of Sampling Adequacy) de Kaiser-Meyer-Olkin. Os valores variam de zero a um. Valores baixos podem indicar que a variável em questão não contribui para a análise e deve ser removida do modelo. Em geral, valores abaixo de 0,5 são considerados inaceitáveis.

**Tabela 2**. Medidas KMO para o Exemplo Completo com 15 variáveis

| Variable | kmo |
|---|---|
| ICOM | 0.5314 |
| ICOA | 0.2277 |
| ICAP | 0.7240 |
| IEND | 0.8243 |
| IRPG | 0.5147 |
| IIMR | 0.4153 |

continua...

... continuação da Tabela 2:

| | |
|---|---|
| ISIN | 0.3286 |
| ICOL | 0.7098 |
| IDAD | 0.6237 |
| ILPG | 0.5138 |
| IRPL | 0.4852 |
| PRPL | 0.7094 |
| IALI | 0.6393 |
| ILCO | 0.5713 |
| ILGE | 0.6020 |
| Overall | 0.5689 |

- **rotate**. Este comando proporciona a rotação (tendo como padrão o método varimax). A rotação permite observar um padrão mais claro das cargas fatoriais.

**Tabela 3**. Cargas fatorias rotacionadas para o resultado da análise fatorial do exemplo completo, com 9 variáveis.

| Variable | Factor1 | Factor2 | Factor3 |
|---|---|---|---|
| ICOM | 0.9701 | -0.0728 | 0.1635 |
| ICAP | 0.3034 | -0.8375 | -0.1733 |
| IEND | -0.0047 | 0.7487 | -0.3004 |
| IDAD | 0.9403 | -0.1181 | 0.2198 |
| ILPG | 0.9777 | -0.0768 | 0.1607 |
| PRPL | -0.0224 | 0.8246 | -0.2391 |
| IALI | -0.0687 | 0.9352 | 0.1387 |
| ILCO | 0.2150 | 0.0028 | 0.9684 |
| ILGE | 0.2307 | -0.0068 | 0.9616 |

Os valores de comunalidade devem ser analisados a partir da medida de unicidade (25), que fica na coluna à direita das cargas fatoriais

**Tabela 4**. Unicidade

| Variable | Factor1 | Factor3 | Uniqueness |
|---|---|---|---|
| ICOM | 0.9701 | 0.1635 | 0.0268 |
| ICAP | 0.3034 | -0.1733 | 0.1765 |
| IEND | -0.0047 | -0.3004 | 0.3492 |
| IDAD | 0.9403 | 0.2198 | 0.0536 |
| ILPG | 0.9777 | 0.1607 | 0.0123 |
| PRPL | -0.0224 | -0.2391 | 0.2624 |
| IALI | -0.0687 | 0.1387 | 0.1015 |
| ILCO | 0.2150 | 0.9684 | 0.0160 |
| ILGE | 0.2307 | 0.9616 | 0.0220 |

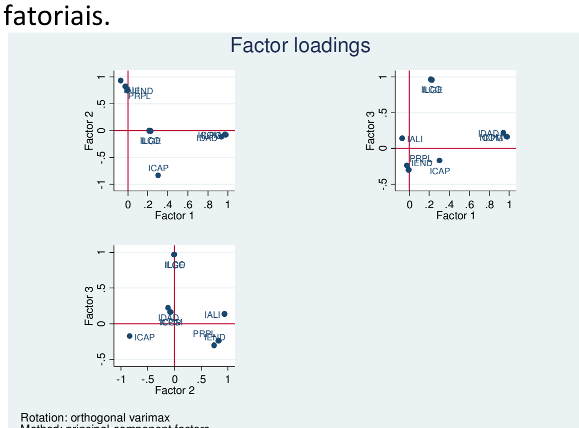
- **predict f1 f2 f3**. Com este comando são obtidos os seguintes resultados:
  a) Os coeficientes dos scores fatoriais. Ou seja, os valores de $b_{ji}$ da expressão (27);
  b) Os scores fatoriais, f1, f2 e f3, calculados para cada caso. Os scores resultantes são disponibilizados como novas variáveis na base de dados

**Tabela 5**. Coeficientes dos Scores Fatoriais

| Variable | Factor1 | Factor2 | Factor3 |
|---|---|---|---|
| ICOM | 0.35505 | 0.04432 | -0.06765 |
| ICAP | 0.08751 | -0.28645 | -0.14326 |
| IEND | 0.10009 | 0.27355 | -0.15513 |
| IDAD | 0.33093 | 0.02572 | -0.03327 |
| ILPG | 0.35822 | 0.04338 | -0.07033 |
| PRPL | 0.08884 | 0.30040 | -0.11960 |
| IALI | 0.01852 | 0.33923 | 0.08812 |
| ILCO | -0.06779 | 0.02062 | 0.47840 |
| ILGE | -0.06118 | 0.01817 | 0.47230 |

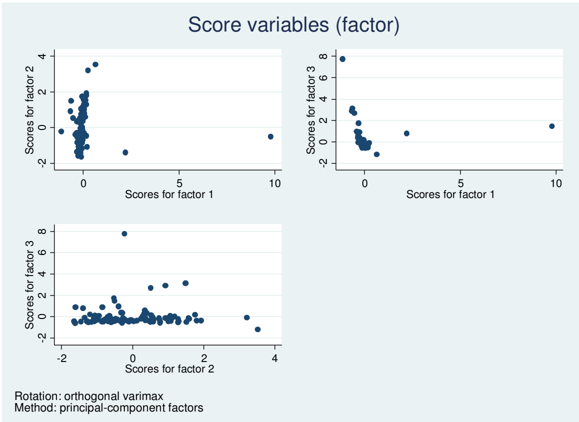
- **loadingplot, factors(3) combined xline(0) yline(0) aspect(1)**. Este comando apresenta gráficos das cargas fatoriais.

**Figura 11**. Loadingplot. Gráfico com cargas fatoriais.




- **scoreplot, factors(3) combined**. Este comando apresenta gráficos dos cores fatoriais f1, f2 e f3 obtidos pelo comando predict.

**Figura 12**. Scoreplot. Gráfico com scores fatoriais




É testada então esta noção prévia para determinar se ela é consistente com o padrão de correlação nos dados.

**A2.3 Análise Fatorial Exploratória vs Análise Fatorial Confirmatória**

Na forma como realizada nas seções anteriores a Análise Fatorial é considerada uma "Análise Fatorial Exploratória".
Nesta seção são feitas considerações sobre os dois tipos de análises.

**A2.3.1 Análise Fatorial Exploratória**
Na Análise Fatorial Exploratória deixamos que cada variável tenha uma determinada carga em cada fator e utilizamos a rotação para identificar estruturas. Ou seja, obtemos uma solução interpretável a partir de rotações. A meta é inferir a estrutura fatorial a partir do padrão de correlação nos dados.

Este tipo de análise foi descrito na seção anterior quando, a partir de 15 indicadores investigamos o padrão de correlações e levantamos a estrutura de fatores. Ou seja, quais indicadores deveriam fazer parte da modelagem, quais deveriam ser retirados do modelo.

**A2.3.2 Análise Fatorial Confirmatória**
Neste caso começamos a análise com uma forte noção prévia da estrutura do modelo fatorial.
Neste caso já partimos do pressuposto que existe uma solução única para o modelo, ou seja, não há necessidade de solução rotacionada.

\- - - - - - - - - - - - -

**Evandro Marcos Saidel Ribeiro**
**E-mail:** esaidel@usp.br

\- - - - - - - - - - - - -

**Bibliografia:**
• J. Lattin, J.D. Carroll, P.E. Green. Análise de dados Multivariados, Cengage Learning, 2011.
• L.P. Fávero, P. Belfiore, F.L. da Silva, B.L. Chan, Análise de Dados: Modelagem Multivariada para tomada de decisões, Elsevier, 2009.
• L.J. Corrar, E. Paulo, J.M. Dias Filho, Análise Multivariada para cursos de Administração, Ciências Contábeis e Economia, Atlas, 2007.
• R.A. **Johnson**, Applied Multivariate Statistical Analysis, Prentice Hall, 1992