Análise Multivariada
Análise Fatorial Exploratória

Prof. George von Borries

Departamento de Estatística
Universidade de Brasília

2023

George von B

Introdução

Karl Pearson

Desenvolvido no Século XX com estudos de Karl Pearson e Charles Spearman para medição de inteligência.

- Psicometria: avaliação de construtos (variáveis latentes).
  - Construto é o conjunto de habilidades e/ou conhecimentos que podem ser plausivelmente argumentados e/ou teoricamente justificados como esperados. Ex. Inteligência, habilidades específicas.
- **Objetivo:** descrever as relações de covariância de muitas variáveis aleatórias em termos de poucas quantidades aleatórias não observáveis, chamadas de fatores.
- **Importante:** se queremos reduzir a dimensão dos dados para análises adicionais (regressão, por exemplo) devemos utilizar componentes principais.

George von Borries

Modelo de Fatores Ortogonais

$$\mathrm{X}_1 = \quad \mu_1 + \ell_{11}\mathbf{F_1} + \ell_{12}\mathrm{F}_2 + \ldots + \ell_{1m-1}\mathrm{F}_{m-1} + \ell_{1m}\mathrm{F}_m + \epsilon_1$$

$$\mathrm{X}_2 = \quad \mu_2 + \ell_{21}\mathrm{F}_1 + \ell_{22}\mathrm{F}_2 + \ldots + \ell_{2m-1}\mathbf{F_{m-1}} + \ell_{2m}\mathrm{F}_m + \epsilon_2$$

$$\vdots \qquad \vdots$$

$$\mathrm{X}_p = \quad \mu_p + \ell_{p1}\mathbf{F_1} + \ell_{p2}\mathrm{F}_2 + \ldots + \ell_{p\ m-1}\mathbf{F_{m-1}} + \ell_{pm}\mathrm{F}_m + \epsilon_p$$

$\mathrm{F_i}$ (azul e vermelho) $\equiv$ importância de $\mathrm{F}_i, i = 1, \ldots, m$ na explicação de $\mathrm{X}_k, k = 1, \ldots, p$.

X_1 e X_p pertencem a um construto[1].
X_2 e X_p pertencem a outro construto.

[1] (Michaelis) Conceito ou construção teórica, puramente mental, elaborada ou sintetizada com base em dados simples, a partir de fenômenos observáveis, que auxilia os pesquisadores a analisar e entender algum aspecto de um estudo ou ciência. Exemplo: Habilidade mental, inteligência, empatia.

George von Borries

- Em **AF** queremos verificar se os dados são consistentes com a estrutura imaginada de construtos (fatores comuns), ou seja:
    - Em **AF** as variáveis originais são expressas como combinação linear de fatores (variáveis latentes).
      Em **ACP** as CPs (variáveis latentes) são expressas como combinação linear das variáveis originais.
    - Em **AF** nos investigamos a estrutura de covariâncias ou correlações entre as variáveis.
      Em **ACP** procuramos explicar uma grande parte da variância total das variáveis.
    - Em **AF** a variância de cada variável é decomposta em **variância comum** (comunalidade ou *communality*) e **variância única** (*uniqueness*).
- Na **Análise Fatorial Exploratória** procuramos estruturas latentes (não observadas) nos dados.

George von Borries

Modelo Fatorial Simples

Este modelo contém um Fator Comum.

$$
\begin{aligned}
X_1 - \mu_1 &= \ell_{11} F_1 + \epsilon_1 \\
X_2 - \mu_2 &= \ell_{21} F_1 + \epsilon_2 \\
&\vdots \\
X_p - \mu_p &= \ell_{p1} F_1 + \epsilon_p
\end{aligned}
$$

Note que F_1 não é observado. Por isso chamamos de **variável latente**.

(✓) μ_i: média de X_i.
(✓) ℓ_{i1}: *loadings*.
(✓) F_1: fator comum (latente) não observado com média 0 e variância 1.
(✓) ϵ_i: erros independentes, com média 0 e variância ψ_i.
(✓) F_1 independente dos erros.

Suponha $p = 4$. Então,

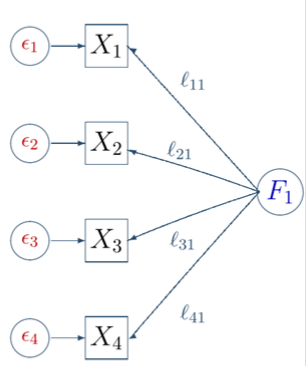


Figura de C.J. Anderson, 2013.

George von Borries

Na teoria clássica de testes (avaliações),

escore verdadeiro (t_i) = escore observado (X_i) - erro puro de medida (e_i)

- Se o modelo fatorial é adequado para os escores observados, o modelo deve ser adequado também para os verdadeiros escores.
- A singularidade do modelo fatorial para escores verdadeiros contém erros específicos devido a variáveis particulares (itens) selecionados, isto é,

$$\mathrm{t}_i = \ell_{i1}\mathrm{F}_1 + \mathrm{s}_i$$

- E resolvendo para X_i,

$$\mathrm{t}_i = \ell_{i1}\mathrm{F}_1 + s_i = X_i - e_i \rightarrow X_i = \ell_{i1}\mathrm{F}_1 + \overbrace{\mathrm{s}_i + e_i}^{\epsilon_i}$$

Fonte: C.J. Anderson, 2013.

George von Borries

- A variável única ϵ_i contém
  - Erros de medição "puros" (e_i)
  - Erros específicos (s_i)

- As variáveis observadas serão correlacionadas porque dependem de F_1.

- O modelo de fator comum tem implicações na matriz de covariâncias (ou correlações). Desta forma, os dados necessários são (novamente) a matriz de covariâncias (ou correlações) dos dados.

George von Borries

Reescrevendo o Modelo Fatorial Simples

Considere,

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} \quad \mathbf{L} = \begin{pmatrix} \ell_{11} \\ \ell_{21} \\ \vdots \\ \ell_{p1} \end{pmatrix} \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_p \end{pmatrix}$$

$$\mathrm{E}(\mathrm{F}_1) = 0 \quad e \quad \mathrm{Var}(\mathrm{F}_1) = \mathbf{\Phi} = \phi_{11} = 1$$

E assim, $\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\mathrm{F}_1 + \boldsymbol{\epsilon}$.

Considere ainda, $\mathbf{\Sigma}$ (matriz de variância-covariância) das variáveis observadas e $\mathbf{\Psi}$ (matriz diagonal de variâncias dos erros)

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \dots & \sigma_{pp} \end{pmatrix} \quad \mathbf{\Psi} = \begin{pmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{pmatrix}$$

George von Borries

Resultados para o Modelo Fatorial Simples

Seja $\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\mathrm{F}_1 + \boldsymbol{\epsilon}$:

- $\mathrm{E}(\mathbf{X} - \boldsymbol{\mu}) = \mathbf{L}\mathrm{E}(\mathrm{F}_1) + \mathrm{E}(\boldsymbol{\epsilon}) = \mathbf{0}$
- Variância-covariância:

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \mathrm{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^{\mathsf{T}}] \\
&= \mathrm{E}[(\mathbf{L}\mathrm{F}_1 + \boldsymbol{\epsilon})(\mathbf{L}\mathrm{F}_1 + \boldsymbol{\epsilon})^{\mathsf{T}}] \\
&= \mathrm{E}[(\mathbf{L}\mathrm{F}_1 + \boldsymbol{\epsilon})(\mathrm{F}_1\mathbf{L}^{\mathsf{T}} + \boldsymbol{\epsilon}^{\mathsf{T}})] \\
&= \mathbf{L}\,\overbrace{\mathrm{E}[\mathrm{F}_1^2]}^{\phi_{11}=1}\,\mathbf{L}^{\mathsf{T}} + \mathbf{L}\,\overbrace{\mathrm{E}[\mathrm{F}_1\boldsymbol{\epsilon}^{\mathsf{T}}]}^{\mathbf{0}} + \overbrace{\mathrm{E}[\boldsymbol{\epsilon}\mathrm{F}_1]}^{\mathbf{0}}\,\mathbf{L}^{\mathsf{T}} + \overbrace{\mathrm{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathsf{T}}]}^{\boldsymbol{\Psi}} \\
&= \mathbf{L}\mathbf{L}^{\mathsf{T}} + \boldsymbol{\Psi}
\end{aligned}
$$

George von Borries

Logo,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \dots & \sigma_{pp} \end{pmatrix} = \begin{pmatrix} \ell_{11}^2 + \psi_1 & \ell_{11}\ell_{21} & \dots & \ell_{11}\ell_{p1} \\ \ell_{11}\ell_{21} & \ell_{21}^2 + \psi_2 & \dots & \ell_{21}\ell_{p1} \\ \vdots & \vdots & \ddots & \vdots \\ \ell_{11}\ell_{p1} & \ell_{21}\ell_{p1} & \dots & \ell_{p1}^2 + \psi_p \end{pmatrix}$$

sendo $\ell_{i1}^2 = h_i^2$ a **comunalidade** da variável i.

Independente do modelo de estimação utilizado,

$$\boldsymbol{\Sigma} - \boldsymbol{\Psi} = \mathbf{L}\mathbf{L}^{\mathsf{T}} = (\boldsymbol{e}_1\sqrt{\lambda_1})(\sqrt{\lambda_1}\boldsymbol{e}_1^{\mathsf{T}})$$

ou

$$\boldsymbol{\Sigma} = \mathbf{L}^*\mathbf{L}^{*\mathsf{T}} = (\boldsymbol{e}_1^*\sqrt{\lambda_1^*})(\sqrt{\lambda_1^*}\boldsymbol{e}_1^{*\mathsf{T}}) \quad \text{e} \quad \tilde{\boldsymbol{\Psi}} = \mathrm{diag}(\boldsymbol{\Sigma} - \mathbf{L}^*\mathbf{L}^{*\mathsf{T}})$$

George von Borries

Exemplo 1: Análise Fatorial de Stock-Price - taxas de retorno semanais de 5 ações. Exemplo 9.4 de Johnson e Wichern.

Programa `AFexemplos.R` - Exemplo 1

```
> library(psych)
> cor(stock)

             JPMorgan  Citibank WellsFargo     Shell     Exxon
JPMorgan    1.0000000 0.6322878  0.5104973 0.1146019 0.1544628
Citibank    0.6322878 1.0000000  0.5741424 0.3222921 0.2126747
WellsFargo  0.5104973 0.5741424  1.0000000 0.1824992 0.1462067
Shell       0.1146019 0.3222921  0.1824992 1.0000000 0.6833777
Exxon       0.1544628 0.2126747  0.1462067 0.6833777 1.0000000

> af1stock <- principal(stock, nfactors = 1, rotate = 'none',
                        covar = FALSE)
> af1stock

Principal Components Analysis
Call: principal(r = stock, nfactors = 1, rotate = "none", covar = FALSE)
Standardized loadings (pattern matrix) based upon correlation matrix

            PC1   h2   u2 com
JPMorgan   0.73 0.54 0.46   1
Citibank   0.83 0.69 0.31   1
WellsFargo 0.73 0.53 0.47   1
Shell      0.60 0.37 0.63   1
Exxon      0.56 0.32 0.68   1

                PC1
SS loadings    2.44
Proportion Var 0.49
```

Exemplo 1: Análise Fatorial de Stock-Price - taxas de retorno semanais de 5 ações. Exemplo 9.4 de Johnson e Wichern.

Programa `AFexemplos.R` - Exemplo 1

```
> resaf1stock <- (round(cor(stock)
                  - af1stock$loadings %*% t(af1stock$loadings)
                  - diag(af1stock$uniquenesses),3))

> resaf1stock
           JPMorgan Citibank WellsFargo  Shell  Exxon
JPMorgan      0.000    0.024     -0.021 -0.328 -0.258
Citibank      0.024    0.000     -0.029 -0.180 -0.255
WellsFargo   -0.021   -0.029      0.000 -0.257 -0.263
Shell        -0.328   -0.180     -0.257  0.000  0.343
Exxon        -0.258   -0.255     -0.263  0.343  0.000
```

- O método de estimação para o Modelo Fatorial foi Componentes Principais[2].
- `PC1` indica as estimativas $\tilde{\mathbf{L}}$ dos loadings de F_1.
- `h2` indica as estimativas de comunalidade, i.e, $\tilde{h}_i^2 = \tilde{\ell}_{i1}^2$.
- `u2` indica as estimativas variâncias específicas $\tilde{\psi}_i$.
- Note que o a solução com um fator aproxima bem a matriz de correlações (e também a de covariâncias).
- Ainda, os maiores valores de F_1 estão relacionados a ações de bancos. Talvez a inclusão de mais um fator melhore esta visualização. Neste caso precisamos de um modelo mais amplo, i.e., com mais fatores.

[2] Veremos mais sobre notação e métodos de estimação em FA nas próximas aulas.

George von Borries

Modelo de Fatores Ortogonais

O modelo de Fatores Ortogonais (geral) permite o acréscimo de mais fatores ao modelo. Considere, por exemplo, $p = 4$ variáveis e $m = 2$ fatores.

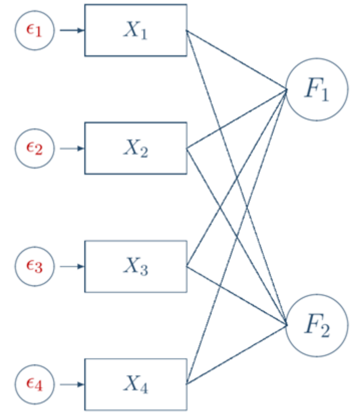


Figura de C.J. Anderson, 2013.

George von Borries

Modelo Geral:

$$
\begin{aligned}
\mathrm{X}_1 &= \mu_1 + \ell_{11}\mathrm{F}_1 + \ell_{12}\mathrm{F}_2 + \ldots + \ell_{1m}\mathrm{F}_m + \epsilon_1 \\
\mathrm{X}_2 &= \mu_2 + \ell_{21}\mathrm{F}_1 + \ell_{22}\mathrm{F}_2 + \ldots + \ell_{2m}\mathrm{F}_m + \epsilon_2 \\
\vdots\ & \quad \vdots \\
\mathrm{X}_p &= \mu_p + \ell_{p1}\mathrm{F}_1 + \ell_{p2}\mathrm{F}_2 + \ldots + \ell_{pm}\mathrm{F}_m + \epsilon_p
\end{aligned}
$$

ou,

$$(\mathbf{X} - \boldsymbol{\mu})_{p\times 1} = \mathbf{L}_{p\times m}\mathbf{F}_{m\times 1} + \boldsymbol{\epsilon}_{p\times 1}$$

sendo

- ℓ_{ij} são os *loadings* e representam o grau de relacionamento linear entre X_i e F_j, $i = 1, \ldots, p$ e $j = 1, \ldots, m$.
- F_j são os fatores, ou variáveis latentes (não observáveis).
- ϵ_i são os erros únicos (não observáveis).

George von Borries

Modelo Geral:

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{LF} + \boldsymbol{\epsilon}$$

com as seguintes suposições:

- $\mathrm{E}(\mathbf{F}) = \mathbf{0}$;
- $\mathrm{Cov}(\mathbf{F}) = \boldsymbol{\Phi} = \mathbf{I}$;
- $\mathrm{E}(\boldsymbol{\epsilon}) = \mathbf{0}$;
- $\mathrm{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Psi} = \mathrm{diag}(\psi_i)$;
- $\mathrm{Cov}(\boldsymbol{\epsilon}, \mathbf{F}) = \mathrm{E}(\boldsymbol{\epsilon}\mathbf{F}') = \mathbf{0}$.

Chamamos

⇒ **L** de matriz de *loadings* dos Fatores.

⇒ $\sum_{j=1}^{m} \ell_{ij}^2 = h_i^2$ de Comunalidade do item i.

⇒ ψ_i de variância específica do item i.

Modelo + Suposições = Modelo de Fatores Ortogonais

George von Borries

Resultado 1:

$$
\begin{aligned}
(\mathbf{X}-\mu)(\mathbf{X}-\mu)^{\mathsf{T}} &= (\mathbf{LF}+\epsilon)(\mathbf{LF}+\epsilon)^{\mathsf{T}} \\
&= (\mathbf{LF}+\epsilon)((\mathbf{LF})^{\mathsf{T}}+\epsilon^{\mathsf{T}}) \\
&= \mathbf{LF}(\mathbf{LF})^{\mathsf{T}}+\epsilon(\mathbf{LF})^{\mathsf{T}}+\mathbf{LF}\epsilon^{\mathsf{T}}+\epsilon\epsilon^{\mathsf{T}}
\end{aligned}
$$

E assim,

$$
\begin{aligned}
\underset{\text{(dados)}}{\mathbf{\Sigma}} = \mathrm{Cov}(\mathbf{X}) &= \mathrm{E}[(\mathbf{X}-\mu)(\mathbf{X}-\mu)^{\mathsf{T}}] \\
&= \mathrm{E}[\mathbf{LF}(\mathbf{LF})^{\mathsf{T}}+\epsilon(\mathbf{LF})^{\mathsf{T}}+\mathbf{LF}\epsilon^{\mathsf{T}}+\epsilon\epsilon^{\mathsf{T}}] \\
&= \mathbf{L}\,\overbrace{\mathrm{E}(\mathbf{FF}^{\mathsf{T}})}^{\mathbf{I}}\,\mathbf{L}^{\mathsf{T}}+\overbrace{\mathrm{E}(\epsilon\mathbf{F}^{\mathsf{T}})}^{\mathbf{0}}\,\mathbf{L}^{\mathsf{T}}+\mathbf{L}\,\overbrace{\mathrm{E}(\mathbf{F}\epsilon^{\mathsf{T}})}^{\mathbf{0}}+\mathrm{E}(\epsilon\epsilon^{\mathsf{T}}) \\
&= \underbrace{\mathbf{LL}^{\mathsf{T}}+\mathbf{\Psi}}_{\text{estrutura não observável}}
\end{aligned}
$$

George von Borries

Resultado 2:

$$
\begin{aligned}
\mathrm{Cov}(\mathbf{X}, \mathbf{F}) &= \mathrm{E}(\mathbf{X}\mathbf{F}^\top) - \mathrm{E}(\mathbf{X})\mathrm{E}(\mathbf{F}) \\
&= \mathrm{E}(\mathbf{X}\mathbf{F}^\top) = \mathrm{E}[(\mathbf{X} - \mu)\mathbf{F}^\top] \\
&= \mathrm{E}(\mathbf{L}\mathbf{F} + \epsilon)\mathbf{F}^\top] = \mathrm{E}[\mathbf{L}\mathbf{F}\mathbf{F}^\top + \epsilon\mathbf{F}^\top] \\
&= \mathbf{L}\mathrm{E}(\mathbf{F}\mathbf{F}^\top) + \mathrm{E}[\epsilon\mathbf{F}^\top] = \mathbf{L}
\end{aligned}
$$

- A covariância entre os dados originais e os fatores comuns é dada pela matriz de *loadings* **L**.
- Resultados (1) e (2) formam a estrutura de covariâncias para o modelo de Fatores Ortogonais.

George von Borries

Resultado 3:

Como $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^\top + \boldsymbol{\Psi}$,

$$\mathrm{Var}(\mathrm{X}_i) = \sigma_{ii} = \ell_{i1}^2 + \ell_{i2}^2 + \ldots + \ell_{im}^2 + \psi_i = h_i^2 + \psi_i, \quad i = 1, \ldots, p.$$

- h_i^2: variância de termos comuns - comunalidade - associada aos F_i que aparecem em todas as equações do modelo.
- ψ_i: variância específica - associada a ϵ_i que é específico da i-ésima equação do modelo.

George von Borries

Resultado 4:

A matriz $\mathbf{L}$ no modelo Fatorial não é única.
Seja $\mathbf{H}$ uma matriz ortogonal, i.e., $\mathbf{H}\mathbf{H}^\top = \mathbf{I}$ e considere o modelo fatorial,

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\mathbf{F} + \boldsymbol{\epsilon} = \mathbf{L}\mathbf{H}\mathbf{H}^\top\mathbf{F} + \boldsymbol{\epsilon} = \mathbf{L}^*\mathbf{F}^* + \boldsymbol{\epsilon}$$

Então,

$$\begin{aligned} \boldsymbol{\Sigma} &= \mathbf{L}\mathbf{L}^\top + \boldsymbol{\Psi} = \mathbf{L}\mathbf{H}\mathbf{H}^\top\mathbf{L}^\top + \boldsymbol{\Psi} \\ &= \mathbf{L}\mathbf{H}(\mathbf{L}\mathbf{H})^\top + \boldsymbol{\Psi} = \mathbf{L}^*\mathbf{L}^{*\top} + \boldsymbol{\Psi} \end{aligned}$$

- Isto implica que diferentes *loadings* podem reproduzir a matriz de covariância de $\mathbf{X}$.
- Note ainda que $h_i^{*2} = h_i^2$: a comunalidade não é afetada pela transformação.

Conclusão: podemos fazer a rotação dos *loadings* sem afetar as suposições ou propriedades do modelo.

George von Borries

Métodos de Estimação

- Seja $\mathbf{S}$ um estimador da matriz de covariância populacional $\boldsymbol{\Sigma}$.
- Se os elementos fora da diagonal de $\mathbf{S}$ são pequenos ou os da matriz $\mathbf{R}$ são próximos de zero, as variáveis não são relacionadas e a análise fatorial não será útil.

  Isto porque os fatores específicos são dominantes e não poderemos determinar um número pequeno de fatores communs importantes.

1. **Método de Componentes Principais:**

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} = \lambda_1 \boldsymbol{e}_1 \boldsymbol{e}_1^{\mathsf{T}} + \lambda_2 \boldsymbol{e}_2 \boldsymbol{e}_2^{\mathsf{T}} + \ldots + \lambda_p \boldsymbol{e}_p \boldsymbol{e}_p^{\mathsf{T}}$$

$$= \left[\sqrt{\lambda_1}\boldsymbol{e}_1 \ldots \sqrt{\lambda_p}\boldsymbol{e}_p\right]\left[\sqrt{\lambda_1}\boldsymbol{e}_1 \ldots \sqrt{\lambda_p}\boldsymbol{e}_p\right]^{\mathsf{T}} = \underset{(p\times p)}{\mathbf{L}}\underset{(p\times p)}{\mathbf{L}^{\mathsf{T}}}$$

com $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$.

George von Borries

1. **Método de Componentes Principais:**

Queremos estimar $\boldsymbol{\Sigma}$ com apenas m dos fatores, i.e.,

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} \approx \lambda_1 \boldsymbol{e}_1 \boldsymbol{e}_1^{\mathsf{T}} + \lambda_2 \boldsymbol{e}_2 \boldsymbol{e}_2^{\mathsf{T}} + \ldots + \lambda_m \boldsymbol{e}_m \boldsymbol{e}_m^{\mathsf{T}}$$

$$= \left[\sqrt{\lambda_1}\boldsymbol{e}_1 \ldots \sqrt{\lambda_m}\boldsymbol{e}_m\right]\left[\sqrt{\lambda_1}\boldsymbol{e}_1 \ldots \sqrt{\lambda_m}\boldsymbol{e}_m\right]^{\mathsf{T}} = \underset{(p\times m)}{\mathbf{L}}\underset{(m\times p)}{\mathbf{L}^{\mathsf{T}}}$$

- $(\boldsymbol{\Sigma})\ \mathbf{S} \approx \hat{\mathbf{L}}\hat{\mathbf{L}}^{\mathsf{T}} + \hat{\boldsymbol{\Psi}}$.
- Utilizamos a decomposição espectral ou $\mathbf{S} = \mathbf{C}\mathbf{D}\mathbf{C}^{\mathsf{T}}$ e escrevemos

$$\mathbf{S} = \mathbf{C}\mathbf{D}\mathbf{C}^{\mathsf{T}} = \mathbf{C}\mathbf{D}^{1/2}\mathbf{D}^{1/2}\mathbf{C}^{\mathsf{T}} = (\mathbf{C}\mathbf{D}^{1/2})(\mathbf{C}\mathbf{D}^{1/2})^{\mathsf{T}} = \hat{\mathbf{L}}\hat{\mathbf{L}}^{\mathsf{T}}.$$

- Note que $\mathbf{C}\mathbf{D}^{1/2}$ é $(p \times p)$ e $\hat{\mathbf{L}}$ é $(p \times m)$, com $m < p$, selecionando os m primeiros autovetores de $\mathbf{C}$ (denominado $\mathbf{C}_1$) correspondentes aos m maiores autovalores correspondentes de $\mathbf{D}$ (denominado $\mathbf{D}_1$).

George von Borries

- Então, $\hat{\mathbf{L}} = \mathbf{C}_1\mathbf{D}_1^{1/2} = (\sqrt{\lambda_1}\mathbf{c}_1, \ldots, \sqrt{\lambda_m}\mathbf{c}_m)$, com $\lambda_1 > \ldots > \lambda_m$.
- Note que o termo **Componentes Principais** se deve ao fato de que os *loadings* do Fator j são proporcionais aos coeficientes na CP_j.

  Não calculamos qualquer CP no método!
- Definimos $\hat{\psi}_i = s_{ii} - \sum_{j=1}^m \hat{\ell}_{ij}^2$ e assim $\mathbf{S} \approx \hat{\mathbf{L}}\hat{\mathbf{L}}^\top + \hat{\mathbf{\Psi}}$.
- $\hat{h}_i^2 = \sum_{j=1}^m \hat{\ell}_{ij}^2$, sendo $\hat{\ell}_{ij}$ pertencente a linha i e coluna j de $\hat{\mathbf{L}}$.
- Note que,

$$\sum_{j=1}^m \hat{\ell}_{ij}^2 = \sum_{i=1}^m (\sqrt{\lambda_j}\mathbf{c}_{ij})^2 = \lambda_j \sum_{i=1}^m c_{ij}^2 = \lambda_j$$

  e proporção da variância total devido ao fator j será

$$\frac{\sum_{j=1}^m \hat{\ell}_{ij}^2}{\mathrm{tr}(\mathbf{S})} = \frac{\lambda_j}{\mathrm{tr}(\mathbf{S})}.$$

- A qualidade do ajuste será obtida por $\mathbf{S} - (\hat{\mathbf{L}}\hat{\mathbf{L}}^\top + \hat{\mathbf{\Psi}})$.

George von Borries

2. Método do Fator Principal:

- Em vez de desconsiderar $\boldsymbol{\Psi}$, utilizamos uma estimativa inicial $\hat{\boldsymbol{\Psi}}$ e fazemos

$$\mathbf{S} - \hat{\boldsymbol{\Psi}} \approx \hat{\mathbf{L}}\hat{\mathbf{L}}^{\mathrm{T}} \quad \text{ou} \quad \mathbf{R} - \hat{\boldsymbol{\Psi}} \approx \hat{\mathbf{L}}\hat{\mathbf{L}}^{\mathrm{T}}$$

  sendo $\hat{\mathbf{L}}_{(p \times m)}$ calculado como no método anterior.

- Note que

$$\mathbf{S} - \hat{\boldsymbol{\Psi}} = \begin{pmatrix} \hat{h}_1^2 & s_{12} & \dots & s_{1p} \\ s_{21} & \hat{h}_2^2 & \dots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \dots & \hat{h}_p^2 \end{pmatrix} \quad e \quad \mathbf{R} - \hat{\boldsymbol{\Psi}} = \begin{pmatrix} \hat{h}_1^2 & r_{12} & \dots & r_{1p} \\ r_{21} & \hat{h}_2^2 & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & \hat{h}_p^2 \end{pmatrix}$$

- Estimativas iniciais são dadas por $\hat{h}_i^2 = \mathrm{R}_i^2 = 1 - \frac{1}{r^{ii}}$ ou $\hat{h}_i^2 = s_{ii} - \frac{1}{s^{ii}}$ com R_i^2 o quadrado do coeficiente de correlação múltipla entre X_i e as outras $p-1$ variáveis.

George von Borries

- Estimativas iniciais são dadas por $\hat{h}_i^2 = \mathrm{R}_i^2 = 1 - \frac{1}{r^{ii}}$ ou $\hat{h}_i^2 = s_{ii} - \frac{1}{s^{ii}} = s_{ii}\mathrm{R}_i^2$, sendo

  - R_i^2 o quadrado do coeficiente de correlação múltipla entre X_i e as outras $p-1$ variáveis;
  - r^{ii} é o elemento i da diagonal de $\mathbf{R}^{-1}$,
  - s_{ii} é o elemento i da diagonal de $\mathbf{S}$ e
  - s^{ii} o elemento i da diagonal de $\mathbf{S}^{-1}$.

- Tanto **R** como **S** devem ser não singulares.
- Quando **R** é singular, podemos fazer $\hat{h}_i^2$ igual a maior correlação na linha i de **R** (estimativa inicial).
- Após obter $\hat{h}_i^2$, calculamos $\hat{\mathbf{L}}$ pelos autovalores e autovetores de $\mathbf{S} - \hat{\mathbf{\Psi}}$ e $\mathbf{R} - \hat{\mathbf{\Psi}}$.
- Em seguida recalculamos $\hat{h}_i^2 = \hat{\ell}_{i1}^2 + \ldots + \hat{\ell}_{im}^2$.

George von Borries

- Proporção da variância explicada pelo fator j,

$$\frac{\lambda_j}{\mathrm{tr}(\mathbf{S}-\hat{\mathbf{\Psi}})} \text{ ou } \frac{\lambda_j}{\mathrm{tr}(\mathbf{R}-\hat{\mathbf{\Psi}})} \equiv \frac{\lambda_j}{\sum_{i=1}^{p}\lambda_i}$$

- Quando $\mathbf{S}-\hat{\mathbf{\Psi}}$ ou $\mathbf{R}-\hat{\mathbf{\Psi}}$ não são positivas definidas alguns autovalores serão negativos podendo resultar em proporção explicada superior a 1.
- Loadings não podem ser obtidos por $\lambda_i \boldsymbol{e}_i$ quando $\lambda_i < 0$ $(\sqrt{\lambda_i} = ?)$. Uma alternativa é fazer $\lambda_i = 0$.
- Um método iterativo pode ser realizado atualizando $\hat{h}_i^2 = \sum_{j=1}^{m}\hat{\lambda}_{ij}$ substituindo na diagonal de $\mathbf{S}-\hat{\mathbf{\Psi}}$ ou $\mathbf{R}-\hat{\mathbf{\Psi}}$ e obtendo novas estimativas para $\mathbf{L}$. O processo continua até as estimativas de comunalidade $(\hat{h}_i^2)$ convergirem.
- Os métodos da Componente Principal e Fator Principal irão apresentar valores semelhantes quando:

  As correlações são elevadas, resultando em m pequeno ou número de variáveis, p, é elevado.

George von Borries

3 Método da Máxima Verossimilhança:

- Suposição: $\mathbf{\Sigma}$ ou $\mathbf{R}$ não singular.
- Método é invariante em escala, i.e., $\mathbf{\Sigma}$ ou $\mathbf{R}$ podem ser utilizados.
- Quando $\mathbf{F}_j$ e $\boldsymbol{\epsilon}_j$ são conjuntamente normais, as observações $\mathbf{X}_j - \boldsymbol{\mu} = \mathbf{L}\mathbf{F}_j + \boldsymbol{\epsilon}_j$ são normais e

$$
\begin{aligned}
\mathrm{L}(\mathbf{X}_1, \ldots, \mathbf{X}_n; \boldsymbol{\mu}, \mathbf{\Sigma}) &= \prod_{j=1}^{n} \left\{ \frac{1}{(2\pi)^{p/2} |\mathbf{\Sigma}|^{1/2}} \exp\left(\frac{-(\mathbf{x}_j - \boldsymbol{\mu})^\mathsf{T} \mathbf{\Sigma}^{-1} (\mathbf{x}_j - \boldsymbol{\mu})}{2} \right) \right\} \\
&= \frac{1}{(2\pi)^{np/2} |\mathbf{\Sigma}|^{n/2}} \exp\left(-\sum_{j=1}^{n} \frac{(\mathbf{x}_j - \boldsymbol{\mu})^\mathsf{T} \mathbf{\Sigma}^{-1} (\mathbf{x}_j - \boldsymbol{\mu})}{2} \right)
\end{aligned}
$$

- Ver notas de aula para desenvolvimento.
- $\ln \mathrm{L}(\mathbf{X}_1, \ldots, \mathbf{X}_n; \boldsymbol{\mu}, \mathbf{\Sigma})$ é resolvido por algoritmos numéricos iterativos[3].

[3] Ver Johnson e Wichern (2007) Suplemento 9A ou Mardia et al. (1979) Capítulo 9.

George von Borries

O métodos de estimação iterativos do Fator Principal e Máxima Verossimilhança podem resultar no **Caso Heywood**.

Caso Heywood: quando comunalidades (h_i^2) são superiores a 1.

$\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^\top + \mathbf{\Psi}$ pode ser p.d. mesmo com um ou mais elementos de $\mathbf{\Psi}$ nulos ou negativos.

- Solução **Heywood:** h_i^2 é igualado a 1 e assim $\Psi_i = 0$.
- Solução **Ultraheywood:** mantem $h_i^2 > 1$ até convergir (se ocorrer). Neste caso $\Psi_i = 1 - h_i^2 < 0$ (variância específica negativa!).
- Razões para ocorrência do Caso Heywood: n pequeno, p pequeno, muitos/poucos fatores, escolha ruim de prioris para h_i^2, modelo inadequado, "outliers" nos dados.

Outros Métodos de Estimação:

- Método da Análise de Imagem (Guttman, 1953).
- Mínimos Quadrados não ponderados.
- Análise do Fator Alfa.

George von Borries

Quantos Fatores utilizar?

- **Scree Plot:** gráfico de $\lambda_i \times i$ (de $\boldsymbol{S} ou \boldsymbol{R}$) com os autovalores em ordem decrescente.

- Reter fatores responsáveis por um certo % da variância total.

- Reter fatores com $\lambda_i > \sum_{i=1}^{p} \lambda_i / p$.

- **Testar a hipótese de número de Fatores (m) adequado:** este teste assume normalidade multivariada dos dados (amostrais).

  $\mathrm{H}_0 : \mathbf{\Sigma} = \mathbf{L}\mathbf{L}^{\mathsf{T}} + \mathbf{\Psi} \quad \text{vs}$

  $\mathrm{H}_1 : \mathbf{\Sigma} \neq \mathbf{L}\mathbf{L}^{\mathsf{T}} + \mathbf{\Psi}$

  sendo $\mathbf{L}$ é $(p \times m)$.

  A estatística do teste (razão de verossimilhança) é

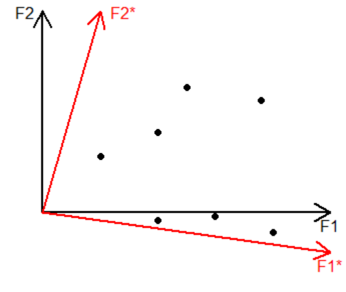
$$\left(n - \frac{2p + 4m + 11}{6}\right) \ln\left(\frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}^{\mathsf{T}} + \hat{\mathbf{\Psi}}|}{|\mathbf{S}|}\right),$$

  que tem distr. aproximada χ^2_ℓ, $\ell = [(p-m)^2 - p - m]/2$ e $\hat{\mathbf{L}}$, $\hat{\mathbf{\Psi}}$ são as estimativas de máxima verossimilhança.

George von Borries

Rotação de Fatores

- É o processo de transformação dos *loadings* através da multiplicação por uma matriz ortogonal (Resultado 4 no Slide 19).
- A rotação tem o único objetivo de encontrar uma estrutura mais simples e fornecer uma interpretação não observada com os *loadings* originais.



- Esta rotação não altera as propriedades do modelo.

George Borries

Rotação Ortogonal[4]

- **Quartimax**: maximiza a variância da soma dos quadrados dos *loadings* dos fatores transformados (Neuhaus e Wrigley, 1954).
- **Varimax**: maximiza a variância dos quadrados dos *loadings* em cada coluna (Kaizer, 1958).
  Seja $\tilde{\ell}^*_{ij} = \hat{\ell}^*_{ij}/hath_i$. Varimax seleciona a transformação ortogonal **T** que maximiza

$$\mathrm{V} = \frac{1}{p}\sum_{j=1}^{m}\left[\sum_{i=1}^{p}\tilde{\ell}^{*4}_{ij} - \frac{\left(\sum_{i=1}^{p}\tilde{\ell}^{*2}_{ij}\right)^2}{p}\right]$$

  O efeito é espalhar o quadrado dos *loadings* em cada fator.

Rotação Não Ortogonal (Oblíqua)

- É utilizada uma matriz de transformação **Q** não singular tal que

$$\mathbf{F}^* = \mathbf{Q}^\mathsf{T}\mathbf{F} \;\text{ e }\; Cov(\mathbf{F}^*) = \mathbf{Q}^\mathsf{T}\mathbf{I}\mathbf{Q} = \mathbf{Q}^\mathsf{T}\mathbf{Q} \neq \mathbf{I}.$$

- Fatores são correlacionados, distâncias e ângulos não são preservados $\Rightarrow$ comunalidades são alteradas.
- Rotações: HK (Harris e Kaiser, 1964), PROMAX (Hendrickson e White, 1964).

[4] Ler Seção 13.5 de Rencher.

George von Borries

Estimação de Fatores Comuns

Ideia:

Obter escores dos fatores $\hat{\mathbf{f}}_i = \{\hat{\mathrm{f}}_{i1}, \ldots, \hat{\mathrm{f}}_{im}\}$, $i = 1, \ldots, n$ que são estimativas dos fatores inerentes a cada observação.

Usos:

- Identificar outiliers.
- Passo intermediário para outras análises: MANOVA, redução de dimensão em séries temporais (Diffusion Index Models[5]).

Procedimentos:

- Estimar $\mathbf{F}'s$ como funções dos valores observados x.
- Obter primeiro $\hat{\ell}_{ij}$ e $\hat{\psi}_i$ e tratar estas estimativas como valores reais.
- Métodos: Mínimos Quadrados Ponderados, Componentes Principais, Regressão.

[5] Ver Tsay, R.S. (2014) Multivariate Time Series Analysis (with R and Financial Applications), Wiley.

George von Borries

Mínimos Quadrados Ponderados:

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\epsilon} \text{ com } \mathrm{Var}(\epsilon_i) = \psi_i \quad i = 1, \ldots, p.$$

Queremos obter $\hat{\mathbf{f}}$ que minimiza,

$$\sum_{i=1}^{p} \frac{\epsilon_i^2}{\psi_i} = \boldsymbol{\epsilon}^\top \boldsymbol{\Psi}^{-1} \boldsymbol{\epsilon} = (\mathbf{X} - \boldsymbol{\mu} - \mathbf{Lf})^\top \boldsymbol{\Psi}^{-1} (\mathbf{X} - \boldsymbol{\mu} - \mathbf{Lf})$$

que tem solução,

$$\hat{\mathbf{f}} = (\mathbf{L}^\top \boldsymbol{\Psi}^{-1} \mathbf{L})^{-1} \mathbf{L}^\top \boldsymbol{\Psi}^{-1} (\mathbf{X} - \boldsymbol{\mu})$$

$$\equiv (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{Y}.$$

Então,

$$\hat{\mathbf{f}}_j = (\hat{\mathbf{L}}^\top \hat{\boldsymbol{\Psi}}^{-1} \hat{\mathbf{L}})^{-1} \hat{\mathbf{L}}^\top \hat{\boldsymbol{\Psi}}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}})$$

George von Borries

Componentes Principais:

Neste caso assumimos Ψ_i iguais.

$$\hat{\mathbf{f}}_j = (\hat{\mathbf{L}}^\mathsf{T}\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}^\mathsf{T}(\mathbf{x}_j - \bar{\mathbf{x}})$$

ou

$$\hat{\mathbf{f}}_j = (\hat{\mathbf{L}}_z^\mathsf{T}\hat{\mathbf{L}}_z)^{-1}\hat{\mathbf{L}}_z^\mathsf{T}\mathbf{z}_j$$

para dados padronizados (matriz de correlações). Mas,

$$\hat{\mathbf{L}} = [\sqrt{\hat{\lambda}_1}\hat{\mathbf{e}}_1 \ldots \sqrt{\hat{\lambda}_m}\hat{\mathbf{e}}_m]$$

e

$$\hat{\mathbf{f}}_j = \begin{pmatrix} \frac{1}{\sqrt{\hat{\lambda}_1}}\hat{\mathbf{e}}_1^\mathsf{T}(\mathbf{x}_j - \bar{\mathbf{x}}) \\ \frac{1}{\sqrt{\hat{\lambda}_2}}\hat{\mathbf{e}}_2^\mathsf{T}(\mathbf{x}_j - \bar{\mathbf{x}}) \\ \vdots \\ \frac{1}{\sqrt{\hat{\lambda}_m}}\hat{\mathbf{e}}_m^\mathsf{T}(\mathbf{x}_j - \bar{\mathbf{x}}) \end{pmatrix}$$

Os fatores tem média **0** e variância-covariância **I**.

George von Borries

Procedimentos para Análise Fatorial no R

- Pacote **stats**: função **factanal**.
- Pacote **psych**: função **fa**.
- Pacote **nFactors**.
- Pacote **FactoMiner**: diversas funções.
- Ver `AFexemplos.R`

Observação: a `PROC FACTOR` do `SAS` tem um grande número de métodos de estimação, rotação e obtenção de fatores comuns.

George von Borries

Resumo de Características da AFE e ACP

Análise Fatorial Exploratória	**Componentes Principais**
1. Técnica de redução que pode ser testada	1. Técnica descritiva para redução de dados
2. Verifica padrões de correlação	2. Verifica variância dos dados
3. Livre de escala	3. Depende de escala (**R**, **S**)
4. Fatores correlacionados ou não	4. Componentes não correlacionadas
5. Fatores são combinações lineares de partes comuns de variáveis não observáveis	5. Combinação linear das variáveis observadas
6. Escores nos fatores podem ser estimados	6. Escores exatos podem ser calculados

George von Borries