# Análise Fatorial: parte 2

Prof. Caio Azevedo

# Predição dos fatores

- Uma vez estimadas as cargas fatoriais ($\boldsymbol{L}$) e a matriz de variâncias-covariâncias dos fatores específicos ($\Psi$) tem-se o interesse em predizer os fatores.
- Os valores preditos dos escores, em geral, são chamados de escores fatoriais.
- Veremos dois métodos: mínimos quadrados ponderados e método de regressão.
- Tanto as estimativas de $\boldsymbol{L}$ e $\Psi$, obtidas pelo método das componentes principais quanto por máxima verossimilhança, podem ser usadas.

# Cont.

- Lembremos que, em geral, se utiliza a matriz de correlações.
- Em geral, também, trabalhamos com as variáveis originais padronizadas (média 0 e variância 1) para gerar os escores.
- Naturalmente, para um mesmo método de predição de escores, os resultados podem ser diferentes consoante o método de estimação usado para as matrizes $\boldsymbol{L}$ e $\Psi$.
- Os escores fatoriais são usados, em geral, como variáveis resposta (ou explicativas) em análises subsequentes bem como para caracterizar (entender) melhor os dados.

# Mínimos quadrados ponderados

- Lembrando que o modelo original é dado por

  $\boldsymbol{X}_i - \boldsymbol{\mu} = \boldsymbol{L}\boldsymbol{F}_i + \boldsymbol{\xi}_i, i = 1, 2, ..., n.$

- O objetivo é minimizar a soma de quadrados dos erros ponderada (link), ou seja, minimizar

$$Q = \sum_{j=1}^{p} \frac{\xi_{ij}^2}{\psi_j} = \boldsymbol{\xi}_i' \Psi^{-1} \boldsymbol{\xi}_i$$

  (para cada indivíduo).

- Temos ainda que $Q = (\boldsymbol{x}_i - \boldsymbol{\mu} - \boldsymbol{L}\boldsymbol{f}_i)' \, \Psi^{-1} \, (\boldsymbol{x}_i - \boldsymbol{\mu} - \boldsymbol{L}\boldsymbol{f}_i)$.

## Cont.

- Derivando-se Q com relação à $\boldsymbol{f}_i$ igualando-se à zero e resolvendo o sistema, obtem-se

$$\widehat{\boldsymbol{F}}_i^* = \left(\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\right)^{-1}\boldsymbol{L}'\Psi^{-1}\left(\boldsymbol{X}_i - \boldsymbol{\mu}\right).$$

- Susbtituindo-se $\Psi$, $\boldsymbol{L}$ e $\boldsymbol{\mu}$ pelos respectivos estimadores, vem que

$$\widehat{\boldsymbol{F}}_i = \left(\widehat{\boldsymbol{L}}'\widehat{\Psi}^{-1}\widehat{\boldsymbol{L}}\right)^{-1}\widehat{\boldsymbol{L}}'\widehat{\Psi}^{-1}\left(\boldsymbol{X}_i - \overline{\boldsymbol{X}}\right).$$

- Se $\Psi$ e $\boldsymbol{L}$ tiverem sido estimados pelo método de máxima verossimilhança, então $\boldsymbol{L}'\Psi^{-1}\boldsymbol{L} = \Delta$, (se $\widehat{\Delta} = \widehat{\boldsymbol{L}}'\widehat{\Psi}^{-1}\widehat{\boldsymbol{L}}$), logo

$$\widehat{\boldsymbol{F}}_i = \widehat{\Delta}^{-1}\widehat{\boldsymbol{L}}'\widehat{\Psi}^{-1}\left(\boldsymbol{X}_i - \overline{\boldsymbol{x}}\right).$$

# Cont.

- Note ainda que $\mathcal{E}(\widehat{\boldsymbol{F}}_i^*) = \left(\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\right)^{-1}\boldsymbol{L}'\Psi^{-1}(\mathcal{E}(\boldsymbol{X}_i) - \boldsymbol{\mu}) = 0$.
- Além disso,

$$
\begin{aligned}
Cov(\widehat{\boldsymbol{F}}_i^*) &= \left(\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\right)^{-1}\boldsymbol{L}'\Psi^{-1}Cov(\boldsymbol{X}_i - \boldsymbol{\mu})\Psi^{-1}\boldsymbol{L}\left(\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\right)^{-1} \\
&= \left(\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\right)^{-1}\boldsymbol{L}'\Psi^{-1}\left(\boldsymbol{L}\boldsymbol{L}' + \Psi\right)\Psi^{-1}\boldsymbol{L}\left(\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\right)^{-1} \\
&= \left(\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\right)^{-1}\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\left(\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\right)^{-1} \\
&+ \left(\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\right)^{-1}\boldsymbol{L}'\Psi^{-1}\Psi\Psi^{-1}\boldsymbol{L}\left(\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\right)^{-1} \\
&= \boldsymbol{I} + \left(\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\right)^{-1}
\end{aligned}
$$

- Se $\Psi$ e $\boldsymbol{L}$ tiverem sido estimados pelo método de máxima verossimilhança, então $Cov(\boldsymbol{F}_i^*) = \boldsymbol{I} + \Delta^{-1}$. Logo, $\mathcal{E}(\widehat{\boldsymbol{F}}_i) \approx 0$ e $Cov(\widehat{\boldsymbol{F}}_i) \approx \boldsymbol{I} + \widehat{\Delta}^{-1}$ ou $\left(\boldsymbol{I} + \left(\widehat{\boldsymbol{L}}'\widehat{\Psi}^{-1}\widehat{\boldsymbol{L}}\right)^{-1}\right)$.

# Método da regressão

- Supõe-se que, i=1,2,...,n

$$\left[\begin{array}{c} \boldsymbol{X}_i - \boldsymbol{\mu} \\ \boldsymbol{F}_i \end{array}\right] \sim N_{(p+m)}\left[\left(\begin{array}{c} 0_{(p\times 1)} \\ 0_{(m\times 1)} \end{array}\right), \left(\begin{array}{cc} \Sigma = \boldsymbol{L}\boldsymbol{L}' + \Psi & \boldsymbol{L} \\ \boldsymbol{L}' & \boldsymbol{I} \end{array}\right)\right]$$

que é equivalente a assumir que $\boldsymbol{F} \sim N_m(0, \boldsymbol{I})$ e $\boldsymbol{\xi} \sim N_p(0, \Psi)$.

- Usando o resultado (visto anteriormente) sobre distribuições condicionais de normais multivariadas, temos que $\boldsymbol{F}|\boldsymbol{x} \sim N(\boldsymbol{\mu}_{\boldsymbol{F}}, \Psi_{\boldsymbol{F}})$, em que

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{F}} &= \boldsymbol{L}'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) = \boldsymbol{L}'\left(\boldsymbol{L}\boldsymbol{L}' + \Psi\right)^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \\ \Psi_{\boldsymbol{F}} &= \boldsymbol{I} - \boldsymbol{L}'\Sigma^{-1}\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{L}'\left(\boldsymbol{L}\boldsymbol{L}' + \Psi\right)^{-1}\boldsymbol{L} \end{aligned}$$

# Cont.

- Então, pode-se usar como preditor para $\boldsymbol{F}_i$:

$$\widehat{\boldsymbol{F}}_i^* = \boldsymbol{L}' \left(\boldsymbol{L}\boldsymbol{L}' + \Psi\right)^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu})$$

- Substituindo-se $\Psi$, $\boldsymbol{L}$ e $\boldsymbol{\mu}$ pelos respectivos estimadores, vem que:

$$\widehat{\boldsymbol{F}}_i = \widehat{\boldsymbol{L}}' \left(\widehat{\boldsymbol{L}}\widehat{\boldsymbol{L}}' + \widehat{\Psi}\right)^{-1} (\boldsymbol{X}_i - \overline{\boldsymbol{X}})$$

- Note ainda que $\mathcal{E}(\widehat{\boldsymbol{F}}_i^*) = \boldsymbol{L}' \left(\boldsymbol{L}\boldsymbol{L}' + \Psi\right)^{-1} (\mathcal{E}(\boldsymbol{X}_i) - \boldsymbol{\mu}) = 0$.

# Cont.

- Além disso,

$$
\begin{aligned}
Cov(\widehat{\boldsymbol{F}}_i^*) &= \boldsymbol{L}'\left(\boldsymbol{L}\boldsymbol{L}'+\Psi\right)^{-1} Cov\left(\boldsymbol{X}_i-\boldsymbol{\mu}\right)\left(\boldsymbol{L}\boldsymbol{L}'+\Psi\right)^{-1}\boldsymbol{L} \\
&= \boldsymbol{L}'\left(\boldsymbol{L}\boldsymbol{L}'+\Psi\right)^{-1}\left(\boldsymbol{L}\boldsymbol{L}'+\Psi\right)\left(\boldsymbol{L}\boldsymbol{L}'+\Psi\right)^{-1}\boldsymbol{L} \\
&= \boldsymbol{L}'\left(\boldsymbol{L}\boldsymbol{L}'+\Psi\right)^{-1}\boldsymbol{L} \\
&= \boldsymbol{L}'\left(\Psi^{-1}-\Psi^{-1}\boldsymbol{L}\left(\boldsymbol{I}+\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\right)^{-1}\boldsymbol{L}'\Psi^{-1}\right)\boldsymbol{L} \\
&= \boldsymbol{L}'\Psi^{-1}\boldsymbol{L}-\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\left(\boldsymbol{I}+\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\right)^{-1}\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}
\end{aligned}
$$

- Se $\boldsymbol{L}$ e $\Psi$ tiverem sido estimados por mv, então $\boldsymbol{L}'\Psi^{-1}\boldsymbol{L} = \Delta$. Logo:

$$
Cov(\widehat{\boldsymbol{F}}_i^*) = \Delta - \Delta\left(\boldsymbol{I}+\Delta\right)^{-1}\Delta
$$

# Cont.

- Portanto, $\mathcal{E}(\widehat{\boldsymbol{F}}_i) \approx 0$ e $Cov(\widehat{\boldsymbol{F}}_i) \approx \widehat{\boldsymbol{L}}' \left(\widehat{\boldsymbol{L}}\widehat{\boldsymbol{L}}' + \widehat{\Psi}\right) \widehat{\boldsymbol{L}}$.
- No caso de $\boldsymbol{L}$ e $\Psi$ terem sido estimados por mv, então $Cov(\widehat{\boldsymbol{F}}_i) \approx \widehat{\Delta} - \widehat{\Delta}\left(\boldsymbol{I} + \widehat{\Delta}\right)^{-1} \widehat{\Delta}$.
- Para ambos os métodos, quando se se considera as variáveis padronizadas, além de se substituir $\Sigma$ por $\boldsymbol{\rho}$, utiliza-se $\boldsymbol{z}_i = \boldsymbol{D}^{-1/2}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})$ em que $\boldsymbol{D}$ é uma matriz diagonal com as variâncias amostrais de $\boldsymbol{X}_i$ em sua diagonal.
- As estimativas (valores preditos) são obtivos substituindo-se os estimadores pelas respectivas estimativas.

# MQP $\times$ MR

- Utilizando-se a relação $\boldsymbol{L}'\left(\boldsymbol{L}\boldsymbol{L}' + \Psi\right)^{-1} = (\boldsymbol{I} + \boldsymbol{L}'\Psi^{-1}\boldsymbol{L})^{-1}\boldsymbol{L}'\Psi^{-1}$, pode-se provar que

$$\widehat{\boldsymbol{F}}_i^{*MQP} = \left(\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\right)^{-1}\left(\boldsymbol{I} + \boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\right)\widehat{\boldsymbol{F}}_i^{*R} = \left(\boldsymbol{I} + \left(\boldsymbol{L}'\Psi^{-1}\boldsymbol{L}\right)^{-1}\right)\widehat{\boldsymbol{F}}_i^{*R}$$

- Sob as estimativas de mv de $\boldsymbol{L}$ e $\Psi$, temos que

$$\widehat{\boldsymbol{F}}_i^{*MQP} = \left(\boldsymbol{I} + \Delta^{-1}\right)\widehat{\boldsymbol{F}}_i^{*R}$$

## Resumo: estimação de $\boldsymbol{L}$ e $\Psi$

- A partir de uma matriz de dados $\boldsymbol{X}_{(n \times p)}$, calcula-se a matriz $\boldsymbol{S}^2$ (covariâncias) ou a matriz $\boldsymbol{R}$ (correlações).
- Componentes principais: calcula-se $(\widetilde{\lambda}_1, \widetilde{\boldsymbol{e}}_1')', ..., (\widetilde{\lambda}_m, \widetilde{\boldsymbol{e}}_m')'$ através de $\boldsymbol{S}^2$ ou $\boldsymbol{R}$. Preferencialmente $\boldsymbol{R}$.
- Teremos $\widetilde{\boldsymbol{L}}$ e $\widetilde{\Psi}$ a partir de $\boldsymbol{S}^2$ e $\widetilde{\boldsymbol{L}}_Z$ e $\widetilde{\Psi}_Z$ a partir de $\boldsymbol{R}$. Em que $\begin{bmatrix} \widetilde{\psi}_1 & 0 & \ldots & 0 \\ 0 & \widetilde{\psi}_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \widetilde{\psi}_n \end{bmatrix} = \widetilde{\Psi} = \text{diag}(\boldsymbol{S}^2 - \widetilde{\boldsymbol{L}}\widetilde{\boldsymbol{L}}')$, ou $\widetilde{\Psi}_Z = \text{diag}(\boldsymbol{R} - \widetilde{\boldsymbol{L}}_Z\widetilde{\boldsymbol{L}}_Z')$

# Resumo

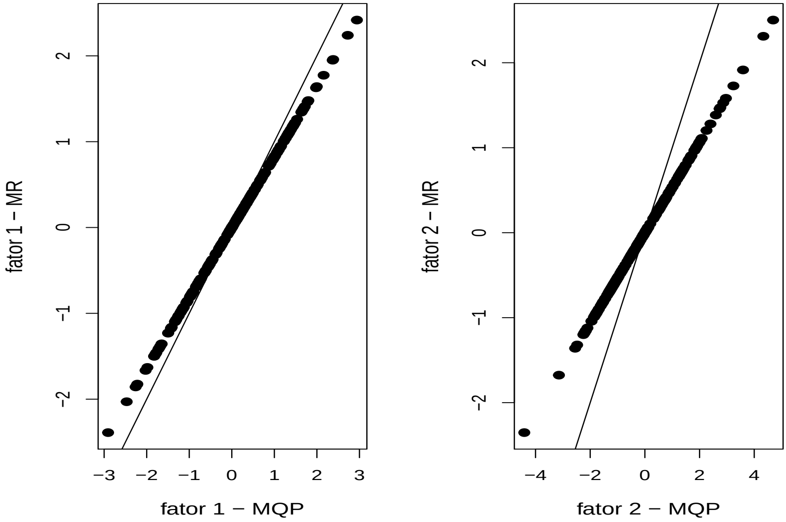
- A partir de uma matriz de dados $\boldsymbol{X}_{(n \times p)}$, maximiza-se à verossimilhança em função de $\boldsymbol{\mu}$, $\boldsymbol{L}$, $\Psi$, com as devidas restrições de identificabilidade. Obtem-se assim, $\widetilde{\boldsymbol{\mu}} = \overline{\boldsymbol{x}}$, $\widetilde{\boldsymbol{L}}$, $\widetilde{\Psi}$, $\widetilde{\Sigma} = \widetilde{\boldsymbol{L}}\widetilde{\boldsymbol{L}}' + \widetilde{\Psi}$ e $\widetilde{\boldsymbol{D}}$.
- Para a matriz de correlações trabalhamos com $\boldsymbol{R}$, ou seja, utilizamos $\boldsymbol{Z}$, obtendo-se: $\widetilde{\boldsymbol{L}}_Z$, $\widetilde{\Psi}_Z$ e $\widetilde{\boldsymbol{R}} = \widetilde{\boldsymbol{L}}_Z\widetilde{\boldsymbol{L}}_Z' + \widetilde{\Psi}_Z$.
- $\boldsymbol{Z} = (\boldsymbol{X} - \overline{\boldsymbol{x}})\boldsymbol{D}^{-1/2}$.

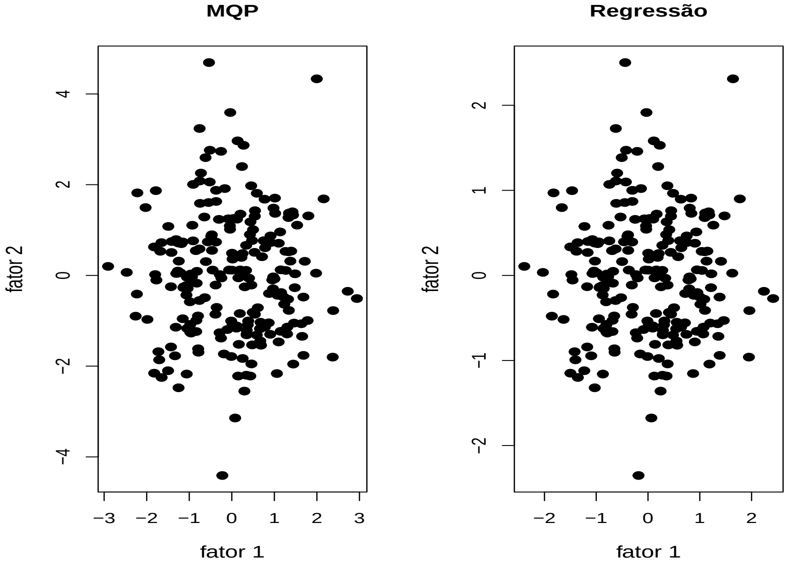
# Resumo

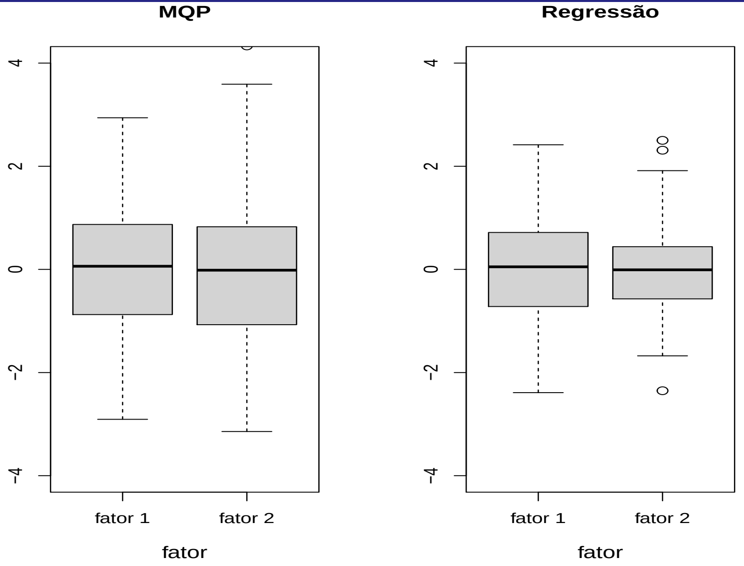
- Considera-se $\widetilde{\boldsymbol{L}}$, $\widetilde{\Psi}$ ou $\widetilde{\boldsymbol{L}}_Z$, $\widetilde{\Psi}_Z$, como valores conhecidos (componentes principais ou máxima verossimilhança).
- MQP:
  - $\widetilde{\boldsymbol{F}}_i = \left(\widetilde{\boldsymbol{L}}'\widetilde{\Psi}^{-1}\widetilde{\boldsymbol{L}}\right)^{-1}\widetilde{\boldsymbol{L}}'\widetilde{\Psi}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})$
  - $\widetilde{\boldsymbol{F}}_i = \left(\widetilde{\boldsymbol{L}}_Z'\widetilde{\Psi}_Z^{-1}\widetilde{\boldsymbol{L}}_Z\right)^{-1}\widetilde{\boldsymbol{L}}_Z'\widetilde{\Psi}_Z^{-1}\boldsymbol{z}_i$
- Regressão:
  - $\widetilde{\boldsymbol{F}}_i = \widetilde{\boldsymbol{L}}'\left(\widetilde{\boldsymbol{L}}\widetilde{\boldsymbol{L}}' + \widetilde{\Psi}\right)^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})$.
  - $\widetilde{\boldsymbol{F}}_i = \widetilde{\boldsymbol{L}}_Z'\left(\widetilde{\boldsymbol{L}}_Z\widetilde{\boldsymbol{L}}_Z' + \widetilde{\Psi}_Z\right)^{-1}\boldsymbol{z}_i$.

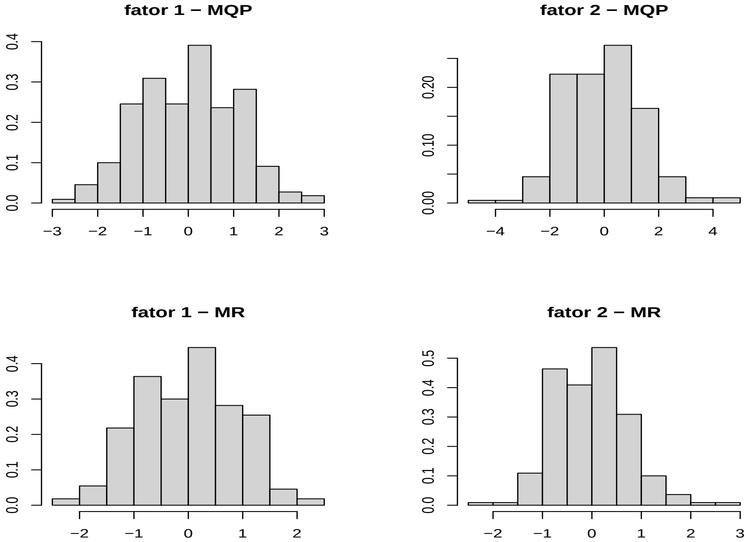
# Voltando ao Exemplo 5

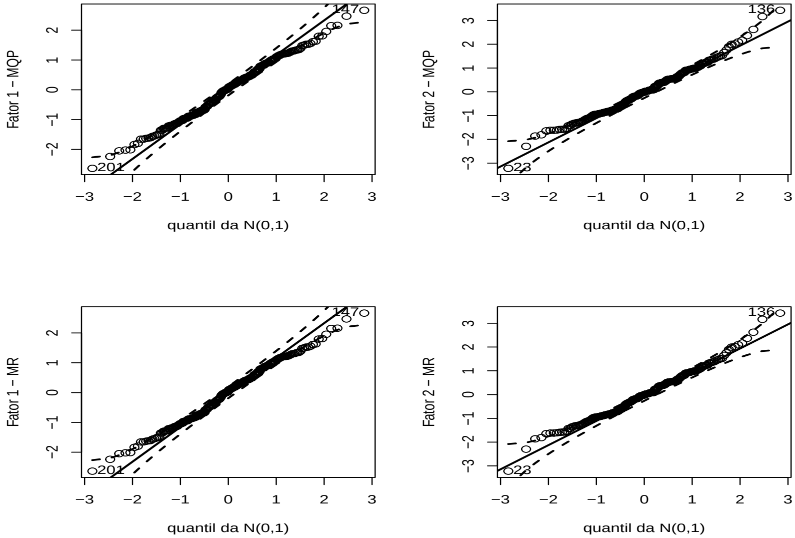
| Estatística | MQP | | Regressão | |
|---|---|---|---|---|
| | Fator 1 | Fator 2 | Fator 1 | Fator 2 |
| Média | 0,00 | 0,00 | 0,00 | 0,00 |
| Var. | 1,18 | 1,64 | 0,85 | 0,61 |
| DP | 1,09 | 1,28 | 0,92 | 0,78 |
| Mínimo | -3,56 | -3,94 | -3,01 | -2,41 |
| Mediana | 0,03 | 0,03 | 0,02 | 0,02 |
| Máximo | 2,79 | 3,02 | 2,37 | 1,84 |

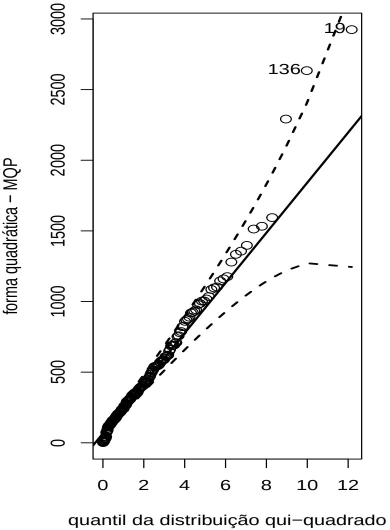

fator 1 – MR
fator 1 − MQP
−3 −2 −1 0 1 2 3
−2 −1 0 1 2
fator 2 – MR
fator 2 − MQP
−4 −2 0 2 4
−2 −1 0 1 2


MQP
Regressão
fator 1
fator 2
−3 −2 −1 0 1 2 3
−4 −2 0 2 4
−2 −1 0 1 2
−2 −1 0 1 2


MQP
Regressão
4
2
0
-2
-4
fator 1
fator 2
fator


fator 1 − MQP
fator 2 − MQP
fator 1 − MR
fator 2 − MR


Fator 1 – MQP
147
201
quantil da N(0,1)
Fator 2 – MQP
136
23
quantil da N(0,1)
Fator 1 – MR
147
201
quantil da N(0,1)
Fator 2 – MR
136
23
quantil da N(0,1)



forma quadrática – MQP
0
500
1000
1500
2000
2500
3000
0 2 4 6 8 10 12
quantil da distribuição qui−quadrado
19
136
forma quadrática – Regressão
0
500
1000
1500
2000
2500
3000
0 2 4 6 8 10 12
quantil da distribuição qui−quadrado
19
136

# Rotação dos fatores (cargas fatoriais)

- Vimos anteriormente que se $T_{(m \times m)}$ é uma matriz ortogonal ($\boldsymbol{T}^{-1} = \boldsymbol{T}'$) e definindo $\boldsymbol{L}^* = \boldsymbol{L}\boldsymbol{T}$, então, tem-se que:

$$\boldsymbol{L}^*(\boldsymbol{L}^*)' + \Psi = (\boldsymbol{L}\boldsymbol{T})(\boldsymbol{L}\boldsymbol{T})' + \Psi = \boldsymbol{L}\boldsymbol{T}\boldsymbol{T}'\boldsymbol{L}' + \Psi = \boldsymbol{L}\boldsymbol{L}' + \Psi$$

- Ou seja, para uma dada estimativa das cargas fatoriais, uma outra pode ser obtida de modo a presevar a representação originalmente obtida para a matriz de covariâncias (correlações).
- Natualmente, se usarmos $\boldsymbol{L}^*$ ao invés de $\boldsymbol{L}$, outros valores para os escores fatoriais serão obtidos.

## Cont.

- Há também rotações não ortogonais, por exemplo as oblíquas, as quais, quase nunca, fornecem a mesma representação da matriz de covariâncias (correlações).
- De álgebra matricial, sabemos que uma transformação (ortogonal ou não) corresponde a uma rotação (ou reflexão) dos eixos coordenados.
- Por esta razão, uma transformação (ortogonal ou não) das cargas dos fatores, e a transformação subsequente dos fatores, é chamada de rotação dos fatores.

# Cont.

- Como as cargas originais dos fatores podem não ser (facilmente) interpretáveis, uma prática comum é rotacioná-las até que uma estrutura de cargas "mais simples" seja atingida.
- Uma situação ideal é buscar por um padrão de cargas tal que cada variável tenha cargas altas em um único fator e cargas pequenas ou moderadas nos demais fatores.
- Porém nem sempre é possível chegar à essa estrutura simples.

## Cont.

- Quando m=2 ou os fatores comuns são considerados dois de cada vez, a transformação para uma estrutura mais simples pode frequentemente ser determinada graficamente.
- Os fatores comuns (não correlacionados) são olhados como vetores unitários no plano Cartesiano cujos eixos coordenados correspondem a cada um dos dois fatores.
- Um gráfico dos pares das cargas dos dois fatores $(\widetilde{\boldsymbol{l}}_{i1}, \widetilde{\boldsymbol{l}}_{i2}), i = 1, 2, ..., p$ produz $p$ pontos, cada um correspondendo à uma variável.
- Os eixos coordenados podem, então, ser "visualmente rotacionados" através um ângulo, digamos $\phi$.

# Exemplo com dois fatores: rotação ortogonal

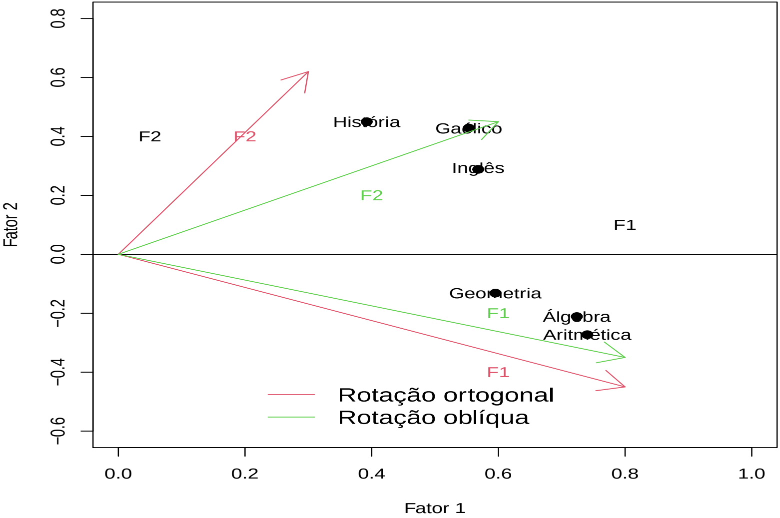
- Suponha que, para um dado exemplo, decidimos considerar dois fatores. Podemos definir a seguinte transformação ortogonal

$$\widehat{\boldsymbol{L}}^*_{(p\times 2)} = \widehat{\boldsymbol{L}}_{(p\times 2)} \boldsymbol{T}_{(2\times 2)}$$

$$\text{em que} \begin{cases} \boldsymbol{T} = \begin{bmatrix} cos\phi & sen\phi \\ -sen\phi & cos\phi \end{bmatrix}, \text{rotação no sentido horário} \\ \boldsymbol{T} = \begin{bmatrix} cos\phi & -sen\phi \\ sen\phi & cos\phi \end{bmatrix}, \text{rotação no sentido anti-horário} \end{cases}$$

- Estimativas das cargas fatoriais (Exemplo 5) : máxima verossimilhança

| **Var.** | **Fator 1** | **Fator 2** | **Comun.** $(\sum_{j=1}^{m} l_{ij}^2)$ | **Var. Específ.** $(\psi_i)$ |
|---|---|---|---|---|
| Gaélico | 0,553 | 0,429 | 0,490 | 0,510 |
| Inglês | 0,568 | 0,288 | 0,406 | 0,594 |
| História | 0,392 | 0,450 | 0,356 | 0,644 |
| Aritmética | 0,740 | -0,273 | 0,623 | 0,377 |
| Álgebra | 0,724 | -0,211 | 0,569 | 0,431 |
| Geometria | 0,595 | -0,132 | 0,372 | 0,628 |


0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
Fator 2
F2
F2
História
Gaélico
Inglês
F2
F1
Geometria
F1
Álgebra
Aritmética
F1
Rotação ortogonal
Rotação oblíqua
0.0
0.2
0.4
0.6
0.8
1.0
Fator 1

- Estimativas das cargas fatoriais, rotação ortogonal (Exemplo 5) : máxima verossimilhança

| **Var.** | **Fator 1** | **Fator 2** | **Comun.** ($\sum_{j=1}^{m} l_{ij}^2$) | **Var. Específ.** ($\psi_i$) |
|---|---|---|---|---|
| Gaélico | 0,366 | 0,701 | 0,625 | 0,375 |
| Inglês | 0,568 | 0,316 | 0,422 | 0,578 |
| História | 0,165 | 0,581 | 0,365 | 0,635 |
| Aritmética | 0,815 | -0,020 | 0,665 | 0,335 |
| Álgebra | 0,772 | 0,049 | 0,599 | 0,401 |
| Geometria | 0,617 | 0,072 | 0,386 | 0,614 |

- Fator 1: fator relativo à Matemática. Fator 2: fator relativo às ciências humanas. Rotação considerando $\phi = 20^o$ na matriz $\boldsymbol{T}$ anterior (rotação no sentido horário).

# Exemplo com dois fatores

- Nesse caso (m=2), aglomerações de variáveis são freqüentemente aparentes graficamente, e essas aglomerações permitem-nos identificar os fatores comuns, sem ter que inspecionar as magnitudes das cargas.
- Por outro lado, quando $m \geq 2$, padrões não são facilmente visualizados e as magnitudes das cargas rotacionadas devem ser verificadas para se encontrar uma interpretação útil dos dados originais.
- A escolha de uma matriz ortogonal que satisfaz uma medida analítica de estrutura simples será considerada.

# Rotação dos fatores (cargas fatoriais)

- Rotação ortogonal: Seja $T_{(m\times m)}$ uma matriz ortogonal ($\boldsymbol{T}^{-1} = \boldsymbol{T}'$) e defina $\boldsymbol{L}^* = \boldsymbol{L}\boldsymbol{T}$, assim:

$$\boldsymbol{L}^*(\boldsymbol{L}^*)' + \Psi = (\boldsymbol{L}\boldsymbol{T})(\boldsymbol{L}\boldsymbol{T})' + \Psi = \boldsymbol{L}\boldsymbol{T}\boldsymbol{T}'\boldsymbol{L}' + \Psi = \boldsymbol{L}\boldsymbol{L}' + \Psi$$

- Rotação oblíqua: Seja $T_{(m\times m)}$ uma matriz não-ortogonal e defina $\boldsymbol{L}^* = \boldsymbol{L}\boldsymbol{T}$, assim:

$$\boldsymbol{L}^*(\boldsymbol{L}^*)' + \Psi = (\boldsymbol{L}\boldsymbol{T})(\boldsymbol{L}\boldsymbol{T})' + \Psi = \boldsymbol{L}\boldsymbol{T}\boldsymbol{T}'\boldsymbol{L}' + \Psi \neq \boldsymbol{L}\boldsymbol{L}' + \Psi$$

# Rotações ortogonais e oblíquas

- Rotação ortogonal:
  - Corresponde à uma rotação rígida.
  - Preserva a representação da matriz de covariâncias obtida pela solução original.
  - Preserva a ortogonalidade dos fatores.
- Rotações oblíquas:
  - Corresponde à uma rotação não-rígida.
  - Não preserva a representação da matriz de covariâncias obtida pela solução original.
  - Não preserva a ortogonalidade dos fatores.

# Rotação ortogonal varimax

- O método de rotação varimax usa como critério uma medida analítica (simples) denominada de critério varimax ou critério normal varimax.
- Sejam $\widetilde{l}_{ij}^{*} = \widetilde{l}_{ij}/\widetilde{h}_i, i = 1, ..., p; j = 1, ..., m$ as cargas fatoriais a serem rotacionadas escalonadas pela comunalidade. O método varimax seleciona a matriz $\boldsymbol{T}$ de modo a maximizar

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

- Essencialmente $V \propto$ $\sum_{j=1}^{m}$(variância do quadrado das cargas escalonadas do j-ésimo fator).

# Cont.

- Escalonar os coeficientes rotacionados $\widetilde{l}_{ij}^{*}$ tem o efeito de atribuir a variáveis com comunalidades pequenas mais peso na determinação de uma estrutura simples.
- Após a transformação $\boldsymbol{T}$ ter sido determinada, as cargas $\widetilde{l}_{ij}^{*}$ são multiplicadas por $\widetilde{h}_i$ para que as comunalidades originais sejam preservadas.
- Portanto, espera-se encontrar tanto grupos de coeficientes de elevada magnitude como grupos de coeficientes de magnitude desprezível em cada coluna da matriz de cargas rotacionadas.

- Estimativas das cargas fatoriais, rotação varimax (Exemplo 5) : máxima verossimilhança

| **Var.** | **Fator 1** | **Fator 2** | **Comun.** $(\sum_{j=1}^{m} l_{ij}^2)$ | **Var. Específ.** $(\psi_i)$ |
|---|---|---|---|---|
| Gaélico | 0,250 | 0,750 | 0,625 | 0,375 |
| Inglês | 0,511 | 0,402 | 0,422 | 0,578 |
| História | 0,071 | 0,600 | 0,365 | 0,635 |
| Aritmética | 0,808 | 0,110 | 0,665 | 0,335 |
| Álgebra | 0,755 | 0,170 | 0,599 | 0,401 |
| Geometria | 0,598 | 0,169 | 0,386 | 0,614 |

- Fator 1: fator relativo à Matemática. Fator 2: fator relativo às ciências humanas.

# Rotações oblíquas

- Rotações ortogonais são apropriadas para um modelo fatorial no qual os fatores são supostos não correlacionados.
- Em algumas áreas de pesquisa é comum (de interesse) considerar rotações oblíquas (não ortogonais), além das não ortogonais.
- As primeiras (oblíquas) são sugeridas após olhar as cargas estimadas e não a partir dos postulados básicos do modelo. Apesar disso, uma rotação oblíqua pode ser útil.

# Rotações oblíquas

- Se olhamos os $m$ fatores comuns como eixos coordenados, o ponto com as coordenadas das cargas da j-ésima variável sobre os $m$ fatores comuns representa a posição dessa variável no espaço fatorial (como no gráfico anterior relativo ao exemplo de rotação).
- Supondo que as variáveis são agrupadas em conglomerados não sobrepostos, uma rotação ortogonal para uma estrutura mais simples corresponde a uma rotação rígida dos eixos coordenados tal que os eixos, após a rotação, passam tão próximos dos conglomerados quanto possível.

# Rotações oblíquas

- Uma rotação oblíqua, para simplificar a estrutura das cargas, corresponde à uma rotação não rígida do sistema de coordenadas tal que os eixos rotacionados (não mais ortogonais dois a dois) passam (próximos) dos conglomerados.
- Uma rotação oblíqua busca expressar cada variável em termos de um número mínimo de fatores, preferivelmente um único fator.

# Rotações oblíqua: promax

- Para rotações oblíquas, o método mais popular é o método promax que tem a vantagem de ser rápido e conceitualmente simples.
- O método busca ajustar uma matriz de cargas fatoriais ($\boldsymbol{L}$) (alvo) que tenha uma estrutura simples.
- A matriz alvo pode ser baseada, por exemplo, na matriz gerada pelo método de varimax.

- Estimativas das cargas fatoriais, rotação promax (Exemplo 5) : máxima verossimilhança

| Var. | Fator 1 | Fator 2 | Comun. ($\sum_{j=1}^{m} l_{ij}^2$) | Var. Específ. ($\psi_i$) |
|---|---|---|---|---|
| Gaélico | -0,004 | 0,793 | 0,629 | 0,371 |
| Inglês | 0,417 | 0,324 | 0,279 | 0,721 |
| História | -0,147 | 0,669 | 0,469 | 0,531 |
| Aritmética | 0,859 | -0,090 | 0,746 | 0,254 |
| Álgebra | 0,777 | -0,007 | 0,604 | 0,396 |
| Geometria | 0,603 | 0,034 | 0,364 | 0,636 |

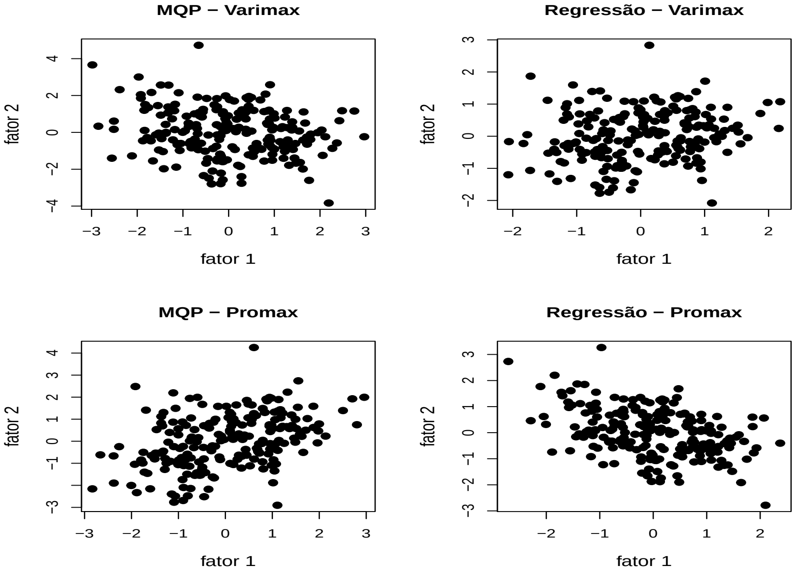
- Fator 1: fator relativo às Ciências Matemáticas. Fator 2: fator relativo à Ciências Humanas.

- Comparação entre as estimativas (originais e rotacionadas) (Exemplo 5) : máxima verossimilhança

| | Original | | Varimax | | Promax | |
|---|---|---|---|---|---|---|
| **Var.** | **Fator 1** | **Fator 2** | **Fator 1** | **Fator 2** | **Fator 1** | **Fator 2** |
| Gaélico | 0,583 | 0,533 | 0,250 | 0,750 | -0,004 | 0,793 |
| Inglês | 0,641 | 0,103 | 0,511 | 0,402 | 0,417 | 0,324 |
| História | 0,354 | 0,490 | 0,071 | 0,600 | -0,147 | 0,669 |
| Aritmética | 0,760 | -0,297 | 0,808 | 0,110 | 0,859 | -0,090 |
| Álgebra | 0,742 | -0,218 | 0,755 | 0,170 | 0,777 | -0,007 |
| Geometria | 0,604 | -0,143 | 0,598 | 0,169 | 0,603 | 0,034 |

# Predição dos escores fatoriais

- De posse de alguma solução rotacionada, a predição dos fatores é feita de igual modo ao que foi anteriormente apresentado.
- Os fatores preditos, obtidos usando-se uma rotação ortogonal das cargas fatoriais, continuam sendo ortogonais.
- Os fatores preditos, obtidos usando-se uma rotação oblíqua das cargas fatoriais, já não mais são ortogonais.
- No gráfico a seguir temos os escores preditos a partir das cargas rotacioandas via varimax e promax.


MQP − Varimax
Regressão − Varimax
MQP − Promax
Regressão − Promax
fator 1
fator 2