Multivariate Statistical Analysis

rafael.poerschke@gmail.com

26 de maio de 2023

Resumo

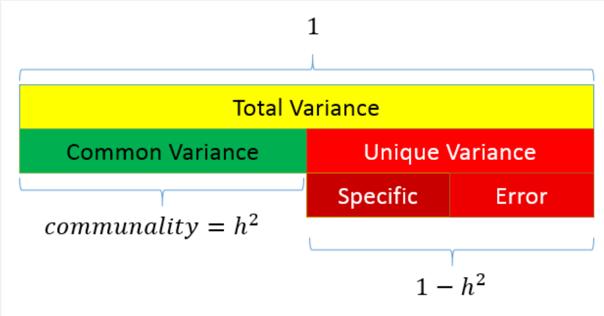
Rudimentos de álgebra para Análise Multivariada, Análise dos Componentes Principais e Análise Fatorial Exploratória.

1 Introduction

1.1 Análise Fatorial Exploratória (AFE)

Na Análise Fatorial, a ideia central é de que existem variáveis de um conjunto que podem ser agrupadas/condensadas em um grupo segundo suas correlações. As variáveis de maior relação deverão ficar em um único grupo, e as demais, com baixa correlação com as primeiras, ficariam em outro grupo. Embora seja possível, em ambos os casos, reduzir a dimensão dos dados a fim de torná-los mais administráveis, a AF é uma extensão da ACP mais elaborada (JOHNSON and WICHERN, 1992).

Se o objetivo da análise fatorial exploratória é modelar a relação entre variáveis, então ela irá se debruçar mais sobre as variâncias e covariâncias do que nas médias. Por essa razão, a análise fatorial assume que a variância original do conjunto pode ser particionada em dois tipos, uma parcela comum e outra parte denominada de variância única. A variância comum é caracterizada pela parcela da variância entre as variáveis de um fator. A comunalidade h^2 é então uma definição matemática da variância comum, que varia entre 0 e 1, e, quanto mais próxima de 1, maior a parcela da variância original explicada pelos fatores estimados. Classificamos como variância única a parcela não explicada, e, portanto, não compartilhada. Essa, por sua vez, pode ser dividida entre variância específica, que corresponde a parcela de variância ligada com uma variável em particular, ao passo que o restante denominamos de erro - o que o modelo não foi capaz de explicar -. A figura que segue auxilia a visualizar esses conceitos.



Fonte: Bruin (2011)

Figura 1: Ilustração do Modelo

Na Análise de Componentes Principais, os dados serão as "p" direções ortogonais da variância original, além de assumir que não existe variância única, isto é, a variância total seria igual a variância comum - comunalidade igual a 1. Esse conjunto pode ser encontrado tomando-se os "m " principais autovetores extraídos da matriz de variância-covariância $\boldsymbol{\Sigma}$ e média μ. Como definimos anteriormente, cada componente principal é um vetor com dimensão " p", existindo então "p " autovetores. Assume-se, assim, que cada vetor de p-variáveis terá uma estrutura linear

$$\mathbf{X}_{p\times 1} = \mu_{p\times 1} + \mathbf{L}_{p\times m}\mathbf{F}_{m\times 1} + \epsilon_{p\times 1} \tag{2.1}$$

O modelo (2.1) postula que a variação das variáveis observadas são dependentes de algumas variáveis aleatórias não observáveis e estimáveis $F_1, F_2, \ldots, F_q$, chamadas de *fatores comuns*, com um adicional de "p" fontes de variações $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$, chamados de *fatores específicos* ou *únicos* – essa denominação faz referência a sua associação única com cada variável.

Note que a expressão (1.6) e (2.1) são semelhantes, pois, para atingir a segunda, partimos da primeira. Os valores das variâncias e covariâncias observados em $\mathbf{X}$ estão agora no lado esquerdo da expressão, e não mais aparecem em termos das observações. Agora, escrevemos que cada variável tem uma carga associada aos pesos da matriz quadrada $\mathbf{W}_{(p\times p)}$, que no caso da análise fatorial pode ter " m " fatores, com "$m \leq p$ " e doravante denotada por $\mathbf{L}_{(p\times m)}$. Algebricamente, o modelo de AFE é definido pelo conjunto de equações

$$
\begin{aligned}
X_1 - \mu_1 &= \ell_{11}F_1 + \ell_{12}F_2 + \cdots + \ell_{1m}F_m + \epsilon_1 \\
X_2 - \mu_2 &= \ell_{21}F_1 + \ell_{22}F_2 + \cdots + \ell_{2m}F_m + \epsilon_2 \\
&\vdots \\
X_p - \mu_p &= \ell_{p1}F_p + \ell_{p2}F_2 + \cdots + \ell_{pm}F_m + \epsilon_p
\end{aligned} \tag{2.2}
$$

Os coeficientes ℓ_{ij} são chamados de *cargas fatoriais* da i-ésima variável com o j-ésimo fator, e, esse coeficiente representa os efeitos diretos. Sendo assim, ℓ_{12} é a carga da variável X_1 oriunda do fator F_2, assim, $\mathbf{L}$ é a matriz $p \times m$ das cargas fatoriais. Vale ressaltar que o i-ésimo fator específico ϵ_i é associado apenas com a i-ésima resposta de X_i. Os "p" desvios $X_1 - \mu_1, X_2 - \mu_2, \cdots, X_p - \mu_p$ são expressos em termos de $p + m$ variáveis aleatórias $F_1, F_2, \cdots, F_q, \epsilon_1, \epsilon_2, \cdots, \epsilon_p$ que são não observáveis. O sistema pode ser escrito como

$$
\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_j \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_j \\ \vdots \\ \mu_p \end{bmatrix} + \begin{bmatrix} \ell_{11} & \ell_{12} & \cdots & \ell_j & \cdots & \ell_{1m} \\ \ell_{21} & \ell_{22} & \cdots & \ell_j & \cdots & \ell_{2m} \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ \ell_{j1} & \ell_{j2} & \cdots & \ell_{ji} & \cdots & \ell_{pm} \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ \ell_{p1} & \ell_{k2} & \cdots & \ell_j & \cdots & \ell_{pm} \end{bmatrix} . \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_j \\ \vdots \\ F_m \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_j \\ \vdots \\ \epsilon_p \end{bmatrix} . \tag{2.3}
$$

Vale ressaltar que estamos considerando que $E[\mathbf{X}] = \mu$, $Cov(\mathbf{X}) = \mathbf{\Sigma}$, mas se adotarmos o modelo (2.1), $E[\mathbf{X}] = 0$, e, por hipótese em modelos estatísticos, implica também em $E[\epsilon] = 0$. Segundo Jolliffe (2003), por conveniência tomamos $E[\mathbf{F}] = 0$. Já os fatores comuns são considerados oblíquos, tal que $Cov(\mathbf{F}) = E[\mathbf{F}\mathbf{F}^T]$ irá resultar na matriz identidade $\mathbf{I}$, onde podemos deduzir que $Cov(\epsilon) = E[\mathbf{F}\epsilon^T]$, isto é, os fatores comuns não mantêm relação com os fatores específicos. Ainda, $Cov(\epsilon)$ será uma matriz diagonal $\mathbf{\Psi}$ de dimensão $p \times p$, pois assumimos que os termos de erros não são correlacionados, tal que

$$
\mathbf{\Psi} = \begin{bmatrix} \psi_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \psi_{pp} \end{bmatrix} . \tag{2.4}
$$

Note ainda que com base em (2.1) e nas médias podemos reescrever $\mathbf{\Sigma}$ em termos de $\mathbf{L}$ e $\boldsymbol{\Psi}$, pois

$$\begin{aligned}
Cov(\mathbf{X}) = \boldsymbol{\Sigma}_{p\times p} &= E[(\mathbf{X}-\mu)(\mathbf{X}-\mu)^T] \\
&= E[(\mathbf{LF}+\epsilon)(\mathbf{LF}+\epsilon)^T] \\
&= E[\mathbf{LFL}^T\mathbf{F}^T + \mathbf{LF}\epsilon^T + \epsilon\mathbf{L}^T\mathbf{F}^T + \epsilon\epsilon^T] \\
&= \mathbf{L}E[\mathbf{FF}^T]\mathbf{L}^T + \mathbf{L}E[\mathbf{F}\epsilon^T] + \mathbf{L}^T E[\epsilon\mathbf{F}^T] + E[\epsilon\epsilon^T] \\
&= \mathbf{LIL}^T + \mathbf{L}0 + 0 + \boldsymbol{\Psi} \\
&\Longrightarrow \mathbf{S} = \boldsymbol{\Sigma}_{p\times p} \quad \Longleftrightarrow \quad \mathbf{S} = \boxed{\mathbf{LL}^T + \boldsymbol{\Psi}}
\end{aligned} \tag{2.5}$$

Adicionar nota e adicionar label na última implicação:

$Cov(X_i, X_k) = \ell_{i1}\ell_{k1} + \cdots + \ell_{im}\ell_{km}\,, \ \forall \ i \neq k$

$\Longrightarrow Cov(\mathbf{X}, \mathbf{F}) = E\left[(\mathbf{X}-\mu)\,(\mathbf{F}-E\,(\mathbf{F}))^T\right]$

$= E\left[(\mathbf{X}-\mu)\,\mathbf{F}^T\right]$

considerando que $(\mathbf{X}-\mu) = \mathbf{LF} + \boldsymbol{\Psi}$, podemos escrever então

$(\mathbf{X}-\mu)\,\mathbf{F}^T = (\mathbf{LF} + \boldsymbol{\Psi})\,\mathbf{F}^T$, tal que tomando o valor esperado, teremos

$E\left[(\mathbf{X}-\mu)\,\mathbf{F}^T\right] = E\left[(\mathbf{LF} + \boldsymbol{\Psi})\,\mathbf{F}^T\right]$

$= E\left(\mathbf{LFF}^T\right) + E\left(\boldsymbol{\Psi}\mathbf{F}^T\right)$

$= \mathbf{L}E\left(\mathbf{FF}^T\right) + 0$

$\Longrightarrow Cov(\mathbf{X}, \mathbf{F}) = \mathbf{L} \quad \text{ou} \quad Cov(X_i, F_j) = \ell_{ij}$

Teorema 1 : Suposições para o Modelo com Variáveis Padronizadas

$$\vec{Z}_{p\times 1} = \mathbf{L}_{p\times q}\vec{F}_{q\times 1} + \vec{\epsilon}_{p\times 1},$$

supondo que

$$E(\vec{F}) = 0; \qquad \text{Var}(\vec{F}) = \mathbf{I}$$

$$E(\vec{\epsilon}) = \vec{0}; \qquad \text{Var}(\vec{\epsilon}) = \Psi = \text{diag}(\psi_1, \cdots, \boldsymbol{\psi}_m)$$

é o modelo fatorial ortogonal, já que os m fatores são ortogonais entre si. Além disso, o modelo supõe que $\vec{F}$ e $\vec{\epsilon}$ são não correlacionados.

Sendo assim, a variância original pode ser quebrada em duas partes. A proporção da variância original explicada da i-ésima variável pelo “ m ” fator comum é chamada de

i-ésima *comunalidade* (devido aos fatores), denotada por h_i^2, então

$$\underbrace{\sigma_{ii}^2}_{Var(X_i)} = \underbrace{\ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{ij}^2}_{\text{comunalidade}} + \underbrace{\psi_i}_{\text{variância específica}}, \qquad \forall\ j = 1, \cdots, m; \tag{2.6}$$

ou ainda,

$$h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{ij}^2 \quad \Longleftrightarrow \quad h_i^2 = \sum_{j=1}^{m} \ell_{ij}^2 \tag{2.7}$$

e

$$\sigma_{ii}^2 = h_i^2 + \psi_i, \qquad \forall\ i = 1, 2, \cdots, p. \tag{2.8}$$

Logo, a i-ésima comunalidade representa a variância da i-ésima variável que é compartilhada com as demais variáveis do fator. Em particular, h^2 representa o grau de dependência dessa variável com o fator. A expressão (2.7) é então a soma do quadrado das cargas fatoriais da i-ésima variável em função do "m" fator comum. A segunda parte de (2.6) é a chamada variância específica ψ_i, e explica a variabilidade de X_i não compartilhada com as outras variáveis.

Nesse ponto, fica mais intuitivo o método, uma vez que agora é possível estimar esses fatores por meio do método de Componentes Principais e/ou Máxima Verossimilhança.

2.1.1 Estimação: O método dos Componentes Principais (e Fatores Principais)

Principal factor analysis is an eigenvalue and eigenvector technique similar in many respects to principal components analysis (see Chapter 3) but operating not directly on S (or R) but on what is known as the reduced covariance matrix , S, defined as Uma vez que fatoramos a matriz de variâncias-covariâncias, será possível a estimação dos pares de autovalores-autovetores $(\lambda_i, \vec{e}_i)$. Para prosseguir iremos partir da equação (2.5),

$$\mathbf{\Sigma}_{p\times p} = \mathbf{L}\mathbf{L}^T + \mathbf{\Psi}$$

com a matriz $\mathbf{\Sigma}$ e $\mathbf{L}$ conhecidas, é possível comparar com a variância original para se ter em conta quanto da parcela da variância cumulativa pode ser reunida nos "q" fatores escolhidos. A estratégia adotada para a escolha do número de fatores é similar a ACP e, assim, iremos também observar os autovalores maiores que a unidade e o *scree plot* dos mesmos. A ideia é reter a maior parte da variância dos dados originais, assim, uma variância acumulada que explique 90% ou mais do total seria ideal. Com isso, por definição

$$\widehat{\mathbf{\Sigma}}_{p\times p} = \mathbf{S}_{p\times p} \tag{2.9}$$

no qual

$$\widehat{\mathbf{\Sigma}}_{p\times p} = \widehat{\mathbf{L}}\widehat{\mathbf{L}}^T + \widehat{\mathbf{\Psi}} \tag{2.10}$$

Logo, podemos escrever

$$(\mathbf{S} - \mathbf{\Sigma})^2 \Longrightarrow (\mathbf{S} - \mathbf{\Sigma})^T(\mathbf{S} - \mathbf{\Sigma}). \tag{2.11}$$

Exemplo 1: Verificando a relação $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T + \mathbf{\Psi}$ para dois fatores

1. (Exemplo 9.1 - (JOHNSON and WICHERN, 1992)) Considere a matriz de covariâncias $\mathbf{\Sigma}$

$$\mathbf{\Sigma} = \begin{bmatrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{bmatrix}$$

A igualdade
$$\begin{bmatrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} 4 & 7 & -1 & 1 \\ 1 & 2 & 6 & 8 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$
ou

$$\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T + \mathbf{\Psi}$$

deve ser verificada pela álgebra matricial. Ainda, $\mathbf{\Sigma}$ tem uma estrutura de um modelo de $m = 2$ fatores ortogonais. Assim

$$\mathbf{L} = \begin{bmatrix} \ell_{11} & \ell_{12} \\ \ell_{21} & \ell_{22} \\ \ell_{31} & \ell_{32} \\ \ell_{41} & \ell_{42} \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{bmatrix},$$

$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_1 & 0 & 0 & 0 \\ 0 & \psi_2 & 0 & 0 \\ 0 & 0 & \psi_3 & 0 \\ 0 & 0 & 0 & \psi_4 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

a comunalidade de X_1 é

$$h_1^2 = \ell_{11}^2 + \ell_{12}^2 = 4^2 + 1^2 = 17$$

a variância de X_1 pode ser decomposta em

$$\sigma_{11} = \left(\ell_{11}^2 + \ell_{12}^2\right) + \psi_1 = h_1^2 + \psi_1$$

ou

$$\underbrace{19}_{\text{variância}} = \underbrace{4^2 + 1^2}_{\text{comunalidade}} + \underbrace{2}_{\substack{\text{variância}\\\text{específica}}} = 17 + 2$$

E o mesmo pode ser feito para a segunda variável. ■

Sendo objetivo minimizar a expressão (2.11), que pode ser tida como a matriz residual $\mathbf{G}$, tal que procuramos a solução para

$$\begin{aligned} \mathbf{G} &= \mathrm{Tr}\left([(\mathbf{S} - \boldsymbol{\Sigma})^T(\mathbf{S} - \boldsymbol{\Sigma})]\right) \\ &= ||(\mathbf{S} - \boldsymbol{\Sigma})^2|| \\ &= ||\left(\mathbf{S} - \left(\mathbf{L}\mathbf{L}^T + \boldsymbol{\Psi}\right)\right)|| \end{aligned} \tag{2.12}$$

Lembrando que $\boldsymbol{\Sigma} = \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T + \cdots + \lambda_p e_p e_p^T$, e utilizando (2.10), tem-se

$$\boldsymbol{\Sigma} = \underbrace{\left[\sqrt{\lambda_1}\, e_1; \sqrt{\lambda_2}\, e_2; \cdots; \sqrt{\lambda_p}\, e_p\right]}_{\mathbf{L}} \cdot \underbrace{\begin{bmatrix} \sqrt{\lambda_1}\, e_1^T; \\ \sqrt{\lambda_2}\, e_2^T; \\ \vdots \\ \sqrt{\lambda_p}\, e_p^T \end{bmatrix}}_{\mathbf{L}^T}. \tag{2.13}$$

Contudo, como o objetivo é diminuir o número de variáveis, condensadas em " m "

fatores (2.13) passará a ser representada por

$$\boldsymbol{\Sigma} = \underbrace{\left[\sqrt{\lambda_1}\, e_1; \sqrt{\lambda_2}\, e_2; \cdots; \sqrt{\lambda_m}\, e_m\right]}_{\mathbf{L}} \cdot \underbrace{\begin{bmatrix} \sqrt{\lambda_1}\, e_1^T; \\ \sqrt{\lambda_2}\, e_2^T; \\ \vdots \\ \sqrt{\lambda_m}\, e_m^T \end{bmatrix}}_{\mathbf{L}^T} = \underset{(p\times m)}{\mathbf{L}} \quad \underset{(m\times p)}{\mathbf{L}^T}. \tag{2.14}$$

Considerando "m" fatores, é possível escrever o vetor de fatores $F_1 = \sqrt{\hat{\lambda}_1}\, \vec{e}_1$, sendo que $\vec{e}_1 = \begin{bmatrix} e_{11} \\ e_{21} \\ \vdots \\ e_{p1} \end{bmatrix}$ até $F_m = \sqrt{\hat{\lambda}_m}\, e_m$, tal que $e_1 = \begin{bmatrix} e_{1p} \\ e_{2p} \\ \vdots \\ e_{mp} \end{bmatrix}$. Como salientado anteriormente, o pacote PSYCH não apresenta os autovetores, e sim o vetor multiplicado pela raiz quadrada de sua raiz característica. Isso pode ser verificado se o leitor tomar o componente PC1 do quadro (8) e multiplicá-lo pela raiz quadrada do seu respectivo autovalor. Esse resultado aparece no quadro (10).

Após a obtenção de $\mathbf{L}$, restará apenas obter $\boldsymbol{\Psi}$ através de (2.5), tal que

$$\widehat{\boldsymbol{\Psi}} = \operatorname{diag}(\mathbf{S} - \widehat{\mathbf{L}}\widehat{\mathbf{L}}^T) \tag{2.15}$$

Sendo assim, restará encontrar a solução para (2.12). A estimação dos fatores foi descrita pelo método dos componentes principais ou também chamado de fatores principais. O método de estimação dos fatores pode diferir entre os pacotes, alguns seguem o método dos componentes principais, enquanto outros estimam os vetores por máxima verossimilhança. O quadro que segue resume o método até aqui descrito.

RESUMINDO

Solução para o Modelo Fatorial por Componentes Principais

A análise dos componentes principais de uma matriz amostral $\mathbf{S}$ é especificada em termos dos pares ordenados de autovalores-autovetores $\left(\hat{\lambda}_1, \hat{\vec{e}}_1\right), \left(\hat{\lambda}_2, \hat{\vec{e}}_2\right), \cdots, \left(\hat{\lambda}_p, \hat{\vec{e}}_p\right)$, em que $\hat{\lambda}_1 \geq \hat{\lambda}_2 \cdots \geq \hat{\lambda}_p$. Como temos $m < p$ o número de fatores será inferior ao número de autovetores. Assim a matriz que contém as cargas fatoriais das variáveis com o fatores $\left(\tilde{\ell}_{ij}\right)$ é dada por

$$\tilde{\mathbf{L}} = \left[\sqrt{\hat{\lambda_1}}\,\hat{\vec{e_1}}; \quad \sqrt{\hat{\lambda_2}}\,\hat{\vec{e_2}}; \quad \cdots; \quad \sqrt{\hat{\lambda_m}}\,\hat{\vec{e_m}}\right] \tag{2.16}$$

As variâncias específicas são retiradas da diagonal principal da matriz $\mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}^T$, logo

$$\tilde{\mathbf{\Psi}} = \begin{bmatrix} \tilde{\psi}_1 & 0 & \cdots & 0 \\ 0 & \tilde{\psi}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{\psi}_p \end{bmatrix} \quad \text{com} \quad \tilde{\psi}_i = s_{ii} - \sum_{j=1}^{m} \tilde{\ell}_{ij}^2 \tag{2.17}$$

As comunalidades são estimadas por

$$\tilde{h}_i^2 = \tilde{\ell}_{i1}^2 + \tilde{\ell}_{i2}^2 + \cdots + \tilde{\ell}_{im}^2 \tag{2.18}$$

A análise dos componentes principais para análise fatorial da matriz amostral de correlações é obtida com a estimação sobre **R** ao invés de **S**.

Ainda existem outras maneiras para se estimar o modelo de fatores, dependendo do pacote, isso pode variar. Outrossim, não é sempre que conseguimos essa estimação, uma vez que pode não haver convergência para um vetor solução. Antes de vermos um exemplo no R, seguimos com um úlitmo exercício retirado do livro texto de JOHNSON and WICHERN (1992).

Exemplo 2: Análise Fatorial com dados das preferências do consumidor

1. Em um estudo sobre as preferências do consumidor, baseado em uma amostra aleatória de consumidores, algumas perguntas sobre os atributos de um produto foram efetuadas. As respostas foram dividas em uma escala semântica de cinco itens, que foram tabulados para construção de uma matriz de correlações. A matriz derivada do estudo é

$$
\mathbf{R} = \begin{array}{c}
\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{bmatrix}
1,00 & ,02 & ,96 & ,42 & ,01 \\
,02 & 1,00 & ,13 & ,71 & ,85 \\
,96 & ,13 & 1,00 & ,50 & ,11 \\
,42 & ,71 & ,50 & 1,00 & ,79 \\
,01 & ,85 & ,11 & ,79 & 1,00
\end{bmatrix}
\end{array}
\begin{array}{l}
\textbf{Atributo (Variável)} \\
1\ \text{Gosto} \\
2\ \text{Custo-benefício} \\
3\ \text{Sabores} \\
4\ \text{Utilidade como lanche} \\
5\ \text{Ótima fonte de energia}
\end{array}
$$

É possível notar que as variáveis 1 e 3 e as variáveis 2 e 5 formam um grupo de alta afinidade. Ainda, a variável 4 parece mais próxima do grupo (2,5) do que de (1,3). Esse resultado preliminar indica que poderemos ter de três a dois grupos com relações lineares significativas.

Os dois primeiros autovalores de **R**, $\hat{\lambda}_1 = 2,85$ e $\hat{\lambda}_2 = 1,81$, são os únicos maiores que a unidade. Ainda, se tomarmos $m = 2$ como fatores, teremos acumulado uma proporção de quase 95% da variância do conjunto normalizado

$$
\frac{\hat{\lambda}_1 + \hat{\lambda}_2}{p} = \frac{2,85 + 1,81}{5} = ,93
$$

Então, a carga fatorial estimada, as comunalidades, e a variância específica são

$$
\tilde{\mathbf{L}}\tilde{\mathbf{L}}^T + \tilde{\boldsymbol{\Psi}} = \begin{bmatrix} ,56 & ,82 \\ ,78 & -,53 \\ ,65 & ,75 \\ ,94 & -,20 \\ ,80 & -,54 \end{bmatrix} \begin{bmatrix} ,56 & ,78 & ,65 & ,94 & ,80 \\ ,82 & -,53 & ,75 & -,10 & -,54 \end{bmatrix} +
\begin{bmatrix} ,02 & 0 & 0 & 0 & 0 \\ 0 & ,12 & 0 & 0 & 0 \\ 0 & 0 & ,02 & 0 & 0 \\ 0 & 0 & 0 & ,11 & 0 \\ 0 & 0 & 0 & 0 & ,07 \end{bmatrix}
$$

$$
= \begin{bmatrix} 1,00 & ,01 & ,97 & ,44 & ,00 \\ & 1,00 & ,11 & ,79 & ,91 \\ & & 1,00 & ,53 & ,11 \\ & & & 1,00 & ,81 \\ & & & & 1,00 \end{bmatrix}
$$

Como o leitor poderá verificar, o modelo quase reproduziu na íntegra a matriz original **R**. Com isso, podemos dizer que o modelo fatorial, com 2 fatores, foi eficiente uma vez que se aproximou de reproduzir a variabilidade dos dados originais. As comunalidades indicam ainda que os dois fatores são responsáveis por uma parcela significativa da variância de cada variável.

Tabela 1 - Resumo das estatísticas

	Carga Fatorial Estimada $\tilde{\ell}_{i\times j} = \sqrt{\hat{\lambda}_i}\hat{e}_{i\times j}$		**Comunalidades**	**Variância Única**
Variável	F_1	F_2	$\tilde{h}_i^2 = \tilde{l}_{i1}^2 + \cdots + \tilde{l}_{im}^2$	$\tilde{\psi}_i = 1 - \tilde{h}_i^2$
1. Gosto	0,56	0,82	0,98	0,02
2. Custo-benefício	0,78	-0,53	0,88	0,12
3. Sabores	0,65	0,75	0,98	0,02
4. Utilidade como lanche	0,94	-0,11	0,89	0,11
5. Ótima fonte de energia	0,80	-0,54	0,93	0,07
Autovalores	2,85	1,81		
Proporção acumulada da Variância total	0,571	0,932		

A ideia aqui ainda não é interpretar o resultado dos fatores e suas cargas, contudo, esse exemplo será retomado quando o assunto for rotação dos fatores (exemplo 6). Mesmo que seja possível uma interpretação dos resultados, uma rotação nos fatores irá ajudar a ressaltar melhor as relações entre as variáveis e os fatores estimados, portanto esperemos até o próximo exemplo.

2.1.2 Rotação dos Fatores

Até aqui, mostramos a transformação ortogonal, que em determinados casos foi ortonormal, através dos autovalores-autovetores da matriz de variâncias-covariâncias (ou correlações). Segundo JOHNSON and WICHERN (1992), esse é um tipo de rotação rígida, pois engendra apenas uma projeção dos eixos ordenados. Nesse sentido, os autores chamam essa transformação de Rotação Fatorial. Outras transformações podem ser aplicadas para ajustar ainda mais o foco sobre os dados. Antes de tratar de tipos específicos de rotação, vamos entender o processo por trás de uma transformação linear.

Se $\widehat{\mathbf{L}}$ é uma matriz $p \times m$ de cargas fatoriais obtidas pelo método escolhido, então $\widehat{\mathbf{L}}^*$

$$\widehat{\mathbf{L}}^* = \widehat{\mathbf{L}}\mathbf{T}, \quad \text{em que} \quad \mathbf{T}\mathbf{T}^T = \mathbf{T}^T\mathbf{T} = \mathbf{T}\mathbf{T}^{-1} = \mathbf{I} \tag{2.19}$$

é uma matriz $p \times m$ de cargas rotacionadas. Ainda, observe que, nesse caso, a matriz de variâncias-covariâncias (ou correlações) se mantém sem alterações

$$\widehat{\mathbf{L}}\widehat{\mathbf{L}}^T + \widehat{\mathbf{\Psi}} = \widehat{\mathbf{L}}\mathbf{T}\mathbf{T}^T\widehat{\mathbf{L}} + \widehat{\mathbf{\Psi}} = \widehat{\mathbf{L}}^*\widehat{\mathbf{L}}^{*T} + \widehat{\mathbf{\Psi}} \tag{2.20}$$

e, conforme a equação (2.20) nos mostra, a matriz residual, $\mathbf{S}_n - \widehat{\mathbf{L}}\widehat{\mathbf{L}}^T - \widehat{\mathbf{\Psi}} = \mathbf{S}_n - \widehat{\mathbf{L}}^*\widehat{\mathbf{L}}^{*T} - \widehat{\mathbf{\Psi}}$, permanece igualmente sem mudanças. Ainda, a matriz das variâncias específicas $\widehat{\mathbf{\Psi}}$ e comunalidades seguem o mesmo comportamento.

Essa transformação em duas dimensões, quando $m = 2$, é facilmente demonstrada gráfica e matricialmente. Suponha a matriz **B0** do exemplo que segue, formada por dois vetores ortogonais $(\widehat{\ell}_{i1}\widehat{\ell}_{i2})$, isto é, perpendiculares entre si. As novas coordenadas podem ser obtidas rotacionando os eixos - baseado em um ângulo θ - tal que os novos vetores $\widehat{\ell}^*_{ij}$ são determinados pela relação disposta em (2.21). Em duas dimensões, uma rotação segue

$$
\mathbf{T}_\theta \mathbf{T}_\theta^{-1} = \underbrace{\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}}_{\substack{\text{sentido} \\ \text{anti-horário}}} \underbrace{\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}}_{\substack{\text{sentido} \\ \text{horário}}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I} \tag{2.21}
$$

A matriz de rotação anti-horaria pode ser utilizada para ilustrar o processo de transformação de uma matriz de cargas fatoriais, com $m = 2$, que será rotacionada em θ graus. Assim, suponha que queiramos rotacionar os eixos dessa matriz de dimensão 2, tal que cada coluna da mesma represente os fatores extraídos de um conjunto qualquer de dados. Partindo de (2.19), temos

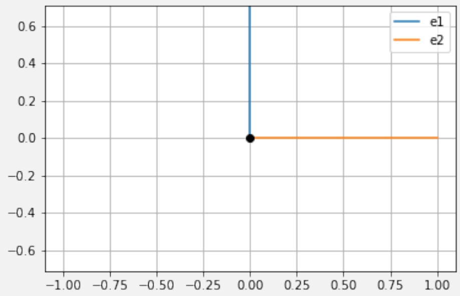
$$
\widehat{\mathbf{L}}^{*T} = \mathbf{T}\,\widehat{\mathbf{L}}^{\mathbf{T}} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \widehat{\ell}_{11} & \widehat{\ell}_{21} \\ \widehat{\ell}_{12} & \widehat{\ell}_{22} \end{bmatrix}.
$$

Esse processo irá realocar os fatores (colunas que foram transpostas) em novos eixos, rotacionando os $\widehat{\ell}_{ij}$ que representam as cargas fatoriais da i-ésima variável do j-ésimo fator, e os fatores após a rotação passam a ser escritos como

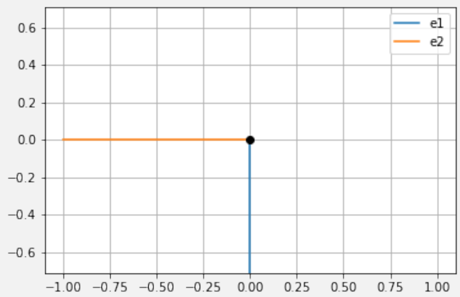
$$
\widehat{\ell}_{i1}^{*T} = \widehat{\ell}_{i1}\cos\theta - \widehat{\ell}_{i2}\sin\theta, \qquad \widehat{\ell}_{i2}^{*T} = \widehat{\ell}_{i1}\sin\theta + \widehat{\ell}_{i2}\cos\theta
$$

O leitor atento irá perceber que os fatores rotacionados têm elementos cruzados em função da multiplicação, isto é, as cargas fatoriais do Fator 1 irão carregar em sua nova orientação uma parcela do Fator 2, e o mesmo vale para os elementos do Fator 2. É exatamente esse fato que vai nos permitir escolher um tipo de rotação que seja apropriado ao nosso objetivo, isto é, aplicar uma rotação que seja capaz de redistribuir a variância da variável entre os fatores da forma que ajude na interpretação dos resultados. Essa relação logo será retomada quando estudarmos os tipos de rotação fatorial mais utilizados.

Para reforçar, podemos fazer uso de exemplo gráfico, o qual é de grande valia por seu apelo visual. Para tal, optamos utilizar o exemplo por meio de Python, que graficamente apresenta uma estrutura melhor que a linguagem R. O exemplo pode ser vislumbrado nas figuras que seguem, à esquerda, temos a matriz identidade **B0** original, e na figura 4b, temos a mesma matriz rotacionada 180 graus no sentido anti-horário.



(a) Matriz BO antes da Rotação



(b) Matriz rotacionada 180 anti-horário

Figura 2: Ilustração dos fatores da matriz B0

1: Código no Python - Exemplo com Matrizes

```
#Rotation Matrix
# importando a biblioteca numpy do python
import numpy
# importando a biblioteca matplotlib do python

theta = np.radians(180)     #transforma graus em radianos
c, s = np.cos(theta), np.sin(theta)
R = np.array(((c, -s), (s, c)))
print(R)

B0 = np.array([
    [0, 1],  #always in line
    [1, 0]
    ], dtype=np.float64)

print(B0)
```

```python

# Rotacionando B0
new = B0 @ R
print(new)
M = new
```

Contudo, a relação em (2.21) é raramente implementada uma vez que em análise multivariada as relações são superiores a duas dimensões e graficamente seria impossível demonstrar tal transformação para $m > 3$.

Hair et al. (2006) dividem as técnicas de rotação em duas famílias. O primeiro caso, que inclui as rotações ortogonais, já foi discutido parcialmente. Esse caso caracteriza-se por manter a independência entre os fatores, gerando novos vetores perpendiculares entre si. Ainda, existe o caso das rotações do tipo oblíqua (directoblimin e PROMAX, entre outras), que são semelhantes às ortogonais, porém permitem fatores correlacionados ao invés de manterem a independência entre eles após a rotação.

Das rotações ortogonais, podemos dizer que seu objetivo é reduzir o número de linhas, ou colunas, a fim de manter uma carga alta de uma variável em um fator, tornando-a menor possível nos demais fatores. No caso do tipo ortogonal QUARTIMAX, como o objetivo é simplificar as linhas da matriz, muitas variáveis podem ter uma carga alta no mesmo fator. Ao se concentrar na simplificação das linhas, a QUARTIMAX acaba concentrando a maior parte da explicação da variância no primeiro fator (DARTON, 1980). Já o método VARIMAX, que se concentra na simplificação das colunas da matriz fatorial, tende a gerar em um fator cargas altas e algumas cargas próximas de 0 - exatamente como gostaríamos. Finalmente, o método EQUIMAX está entre ambos ao se concentrar na simplificação de parte das linhas ou das colunas e, segundo JOHNSON and WICHERN (1992), tem sido pouco utilizado quando comparado aos demais.

Kaiser (1974) chama o método VARIMAX de Normal Varimax, e define $\tilde{\ell}^*_{ij} = \frac{\hat{\ell}^*_{ij}}{\hat{h}_i}$ como o coeficiente rotacionado pela raiz quadrada da comunalidade. Então a tranformação do método normal VARIMAX para o fator j, denotado por $\tilde{\ell}^*_j$, é definida pela função ϕ, que tem por critério maximizar a soma das variâncias do quadrado das cargas de cada coluna da matriz de cargas fatoriais, sendo que cada linha foi normalizada por sua comunalidade; isto é

$$\phi = \sum_{j=1}^{m}\sum_{i=1}^{p}\left(\tilde{\ell}_{ij}^{*2} - \overline{\tilde{\ell}}_{j}^{*}\right)^2 = \frac{1}{p}\sum_{j=1}^{m}\left[\sum_{i=1}^{p}\tilde{\ell}_{ij}^{*4} - \left(\sum_{i=1}^{p}\tilde{\ell}_{ij}^{*2}\right)^2\frac{1}{p}\right], \quad \forall\ j = 1, 2, \cdots, m. \tag{2.22}$$

quanto maior possível, em que

$$h_i^2 = \sum_{j=1}^{m}\tilde{\ell}_{ij}^{*2}, \quad i = 1, \cdots, p, \tag{2.23}$$

é a comunalidade da variável X_i.

O aumento da carga nos coeficientes $\tilde{\ell}_{ij}^{*}$ irá afetar as variáveis com baixa comunalidade, dando a elas um maior peso relativo, e, após a transformação, as cargas $\tilde{\ell}_{ij}^{*}$ devem ser multiplicadas por $\widehat{h}_i$ para que a comunalidade original permaneça. A expressão (2.22) nos mostra que ϕ corresponde a maximizar o quadrado das cargas fatoriais em cada fator tanto quanto possível. Com isso, segundo JOHNSON and WICHERN (1992), espera-se encontrar um grupo (um vetor das cargas fatoriais) com coeficientes grandes e outros negligenciáveis em cada coluna da matriz rotacionada $\widehat{L}^{*}$. Diversos pacotes computacionais realizam essa operação, o que facilita, e muito, a vida do pesquisador. Cabe ressaltar, que essa nova matriz rotacionada irá diferir conforme o método de estimação dos fatores variar (componentes principais, máxima verossimilhança).

Exemplo 3: Análise Fatorial com dados das preferências do consumidor e Rotação VARIMAX

1. Seguimos com os dados e resultados do exemplo (5), sobre o estudo das preferências do consumidor. Após a rotação VARIMAX, temos uma nova coluna na tabela, agora com os fatores rotacionados.

   Tabela 2 - Resumo das estatísticas com fatores rotacionados (VARIMAX)

Variável	Carga Fatorial Estimada F_1	F_2	Carga Fatorial Rotacionada F_1^*	F_2^*	Comunalidades	Variância Específica
1. Gosto	0,56	0,82	0,02	0,99	0,98	0,02
2. Custo-benefício	0,78	-0,53	0,94	-0,01	0,88	0,12
3. Sabores	0,65	0,75	0,13	0,98	0,98	0,02
4. Utilidade como lanche	0,94	-0,104	0,84	0,43	0,89	0,11
5. Ótima fonte de energia	0,80	-0,54	0,97	-0,02	0,93	0,07
Proporção acumulada da Variância total	0,571	0,932	0,507	0,932		

Se observarmos com atenção o fator 1, antes e depois da rotação, as variáveis (2) Custo-Benefício e (5) Fonte de Energia ganharam mais peso no fator, ao passo que (3) Sabor e (1) Gosto perderam importância. Esses mesmos (1) Gosto e (3) Sabor, formaram um grupo mais coeso dentro do Fator 2. Com isso, poderíamos dizer que o Fator 1 representa a percepção *nutricional*, enquanto o Fator 2 seria o fator relacionado à percepção do *sabor* pelo consumidor.

2.1.3 Então e Agora: os Escores Fatoriais

A etapa seguinte à rotação dos fatores escolhidos se baseia em calcular os *escores fatoriais* utilizando as cargas fatoriais e suas correlações com os fatores. Os escores fatoriais servem como *proxies* das variáveis latentes e posição da observação nesse fator. Conceitualmente, um escore fatorial é o valor que devemo observar a relação dos indivíduos com os fatores. Esses escores devem ser estimados com muito cuidado. Suponha, agora, que estamos satisfeitos com os quatro m-fatores que estimamos e ainda queremos preparar os dados para uma análise subsequente. Sendo os fatores não-observáveis, precisamos gerar ou prever os escores fatoriais. Para tal fim, existem três métodos bem difundidos: regressão, Bartlett e Anderson-Rubin.

Quando comparados, a regressão irá maximizar a correlação entre os escores fatoriais e os fatores. Contudo, mesmo havendo uma solução ortogonal, ainda podemos encontrar escores fatoriais correlacionados. Uma opção para contornar essa característica seria estimar por Bartlett, tornando os escores fatoriais correlacionados com seu fator e não com os demais, sendo estimativas não-viesadas dos fatores originais. Ausência de viés deve ser entendida como a situação que, em uma amostragem repetida dos escores fatoriais, as médias dos escores serão iguais a média do fator a qual ele se relaciona. Já na opção por Anderson-Rubin (A-R), por sua vez, produz escores que não mantêm relação com os fatores e nem com os demais escores produzidos. O método de Anderson-Rubin pressupõe inexistência de relação entre os escores fatoriais, o que não seria adequado para rotações do tipo oblíquas. Assim, o escores gerados em A-R são viesados.

Sendo assim, se a rotação dos fatores escolhida for ortogonal, qualquer dos três métodos pode ser uma opção. Para escores não-viesados, seria ideal uma rotação ortogonal e estimação por Bartlett. Contudo, se o objetivo for maximizar a relação dos escores com os fatores, regressão seria melhor opção, podendo ser utilizado para rotações oblíquas. Caso a ideia seja produzir escores se relação entre eles, A-R seria apropriado.

2.1.4 Exemplo aplicado no R: dados do Censo Agropecuário 1994/95

O pacote básico do R apresenta opções de análise fatorial exploratória, bem como existem outras opções disponíveis. Podemos ainda destacar os pacotes PSYCH e FactorMiner, ou, se o leitor preferir, poderá estimar sua própria regressão a fim de extrair os fatores desejados.

O pacote *default* do R já nos fornece uma opção com o comando "factanal", o qual irá estimar o número de fatores desejado pelo método da Máxima-Verossimilhança (Maximum Likelihood Estimation - MLE). É recomendável que o leitor consulte a documentação do R para aprender mais sobre a sintaxe da função "factanal".

Seguindo no nosso exemplo, já havíamos definido que quatro ou cinco fatores iriam reter 90% ou mais da variância do conjunto. Aqui, vamos tentar encontrar uma solução para o número de quatro (4) fatores sem rotação, para, posteriormente, rotacionarmos esses resultados. O quadro que segue apresenta um breve comentário da função e traz seus resultados a seguir.

1: Código no R - Exemplo com os COREDES agropecuários do RS (1994/95)

```
############################################################################################
# Estimando pelo pacote default do R - Maximum Likelihood Factor Analysis
############################################################################################
# Exemplo com entrada de 4 fatores e sem rotação
# Algumas vezes o default da parte inferior do intervalo de unicidade durante a
↪  otimização, por exemplo, aumentando do default 0.005 para 0.01 funcionou neste
↪  caso.

fit <- factanal(data, 4, rotation = "none", lower = 0.01)

############################################################################################
# Chamando os fatores - quatro colunas de loads
############################################################################################
loads <- fit$loadings[,1:4]
loads
```

```
> fit <- factanal(data, 4, rotation = "none", lower = 0.01)

> loads <- fit$loadings[,1:4]
> loads
          Factor1      Factor2     Factor3      Factor4
gado    0.8946692 -0.421474484  0.11585265  0.008002959
pop     0.5464798  0.509748267  0.45826994 -0.218868509
pea     0.6579030  0.380546877  0.62197217 -0.136455667
asspec  0.4674445  0.387899563  0.28635054  0.189351545
assagr  0.3463977  0.583517580  0.19361976  0.106221894
trator  0.8852697  0.343539422 -0.04846889  0.287781668
adubos  0.4098123  0.601038554  0.64528468  0.010346896
irriga  0.7623914 -0.002016925  0.04557246  0.074868406
recveg  0.9269895  0.169403531 -0.29531673 -0.127688314
recani  0.9063167 -0.237590724  0.16055804  0.031242781
financ  0.8991506  0.054224875 -0.23933102 -0.073359465
maqcol  0.8191416  0.425015773 -0.17707828  0.267813620
valveg  0.9367113  0.208340051 -0.22975649 -0.131491040
valani  0.9441096 -0.268039396  0.13674094  0.041253423
arexpl  0.9277722 -0.353555487  0.06963031  0.023453907
```

```
distrg -0.2866997  0.080780082 -0.03839641 -0.390396357
renda   0.2598517  0.184572751 -0.25813253  0.467011991
```

Aqui, já podemos perceber que os fatores são diferentes dos componentes principais gerados anteriormente. Mas esses resultados não parecem animadores, uma vez que apenas três dos fatores trouxeram variáveis com relações consideráveis. O Fator 4, por exemplo, não apresentou nenhuma variável que tenha correlação forte com ele. A variável renda e distância do porto de Rio Grande (distrg) não mostraram relações significativas com os fatores estimados. Para tentar melhorar o foco, vamos utilizar a função - rotation - com a opção do método VARIMAX. Os resultados estão dispostos no quadro que segue.

2: Código no R - Exemplo com os COREDES agropecuários do RS (1994/95)

```
##########################################################################################
# Estimando pelo pacote default do R com rotação VARIMAX
##########################################################################################
fit <- factanal(data, 4, rotation = "varimax", lower = 0.01)

##########################################################################################
# Chamando os fatores - quatro colunas de loads
##########################################################################################
loads <- fit$loadings[,1:4]
loads
```

```
> fit <- factanal(data, 4, rotation = "varimax", lower = 0.01)

> loads <- fit$loadings[,1:4]
> loads
           Factor1     Factor2    Factor3     Factor4
gado    0.96318854  0.09166948 0.21205619  0.10225584
pop     0.21452366  0.83487171 0.24978001 -0.10477117
pea     0.41396950  0.89138788 0.11691837 -0.05923549
asspec  0.18334274  0.59651884 0.13954937  0.27910177
assagr -0.04750461  0.62845100 0.24622238  0.22676652
trator  0.45055145  0.47792985 0.53155880  0.52241837
adubos  0.09461166  0.96604833 0.01209050  0.05784383
irriga  0.59104797  0.27810471 0.33716066  0.22038509
```

```
recveg  0.53735315  0.22769370 0.79001545  0.16372560
recani  0.87039357  0.24700954 0.25374491  0.14706850
financ  0.59805727  0.17683199 0.67324269  0.17853664
maqcol  0.32168908  0.42545798 0.62393797  0.53001437
valveg  0.53615789  0.30073018 0.76750308  0.15474729
valani  0.91206131  0.22292222 0.27384200  0.16446383
arexpl  0.93642189  0.11647042 0.28255094  0.14448128
distrg -0.27063409 -0.05087770 0.04130924 -0.40628553
renda   0.02045104  0.01819023 0.22652244  0.57815314
```

A rotação parece ter melhorado os resultados para análise, muito embora não tenha resolvido por completo os problemas ressaltados anteriormente. Mas as variáveis gado (rebanho bovino), pop (população) e pea (população economicamente ativa) ganharam novos pesos e uma distribuição melhor entre os Fatores 1 e 2. O Fator 1 parece agora reunir variáveis ligadas à pecuária, pois conta com variáveis como o efetivo bovino (gado), receita das atividades com pecuária de corte (recani), valor da produção oriunda da pecuária (valani) e área explorada (arexpl). Como podemos ver, pea e pop ficaram mais ligadas ao Fator 2 após a rotação, ao passo que antes não era possível encontrar uma correlação alta dessas com os fatores. É interessante notar que o Fator 3, antes sem possuir variáveis com expressão, ganhou a presença das variáveis maqcol (número de máquinas colheitadeiras), recveg (receita vegetal), valveg (valor vegetal) e financ (valores financiados). Talvez o maior problema seja a falta de relação da distância do porto de Rio Grande e da cidade continua, uma vez que ele não tem uma relação significativa com nenhum fator. O Fator 4, em específico, não apresentando variáveis com elevada correlação. Por esse ponto de vista, ele seria dispensável, muito embora os autovalores mostrem ser suficiente a presença de quatro fatores para descrever a variância dos dados.

Após essa breve análise, vamos passar aos resultados apresentados pelo pacote PSYCH. A estimação aqui foi feita pelo *default* da função "fa", isto é, a estimativa dos fatores segue o método dos mínimos quadrados ordinários. Se pensarmos na regressão linear como tal, ou mesmo no método anterior - MLE[1] - já podemos esperar que os resultados não sejam os mais apropriados, pois, pelas características dessas variáveis, os erros da regressão deverão

[1]Maximum Likelihood assume uma distribuição multivariada normal do conjunto, quando essa hipótese é violada, uma estimativa apropriada seria por Roboust Maximum Likelihood (MLR). Ambos são mais apropriados para dados do tipo

violar algum dos pressupostos previstos pelo teorema de Gauss-Markov. Contudo, vamos analisar os resultados que o pacote produz sobre a matriz de correlações - observe que a função "facanal"utilizava a matriz de dados originais.

3: Código no R - Exemplo com os COREDES agropecuários do RS (1994/95)

```
##########################################################################################################
# Estimando pelo pacote PSYCH por Mínimos Quadrados
##########################################################################################################
#Five alternatives are provided in psych, all of them are included in the #fa function and are called by specifying the factor method (e.g., #fm=“minres”, fm=“pa”, fm=“”wls”, fm=”gls” and fm=”ml”).
#Factoring method fm="minres" will do a minimum residual as will #fm="uls". Both of these use a first derivative. fm="ols" differs very #slightly from "minres" in that it minimizes the entire residual matrix #using an OLS procedure but uses the empirical first derivative. This #will be slower. fm="wls" will do a weighted least squares (WLS) #solution, fm="gls" does a generalized weighted least squares (GLS), #fm="pa" will do the principal factor solution, fm="ml" will do a maximum #likelihood factor analysis. fm="minchi" will minimize the sample size #weighted chi square when treating pairwise correlations with different #number of subjects per pair. fm ="minrank" will do a minimum rank factor #analysis. "old.min" will do minimal residual the way it was done prior #to April, 2017 (see discussion below). fm="alpha" will do alpha factor #analysis as described in Kaiser and Coffey (1965)

fit <- fa(r = R, nfactors = 4)
##########################################################################################################
# Chamando os fatores
##########################################################################################################
loads <- fit$loadings[,1:4]
loads
```

```
> fit <- fa(r = R, nfactors = 4)
Warning message:
In fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs,  :
  The estimated weights for the factor scores are probably incorrect.  Try a different factor score estimation method.
> loads <- fit$loadings[,1:4]
> loads
```

```
                MR3         MR1          MR2         MR4
gado   -0.01898016  1.01808453 -0.031479183 -0.04759454
pop     0.25379943 -0.10160962  0.860100913 -0.22012260
pea    -0.02607958  0.25722241  0.897362226 -0.03908360
asspec  0.07357120  0.18535202  0.380919057  0.56702392
assagr  0.28476445 -0.18439023  0.486616738  0.33368393
trator  0.75828128  0.13078972  0.138290151  0.21427919
adubos -0.06148208 -0.01469871  0.936649486  0.21475386
irriga  0.38382131  0.38891032  0.148108436 -0.13945079
recveg  0.91722725  0.10396251 -0.009075043 -0.12986347
recani  0.03761497  0.88672805  0.095459165  0.10059903
financ  0.75946495  0.25533476 -0.040369993 -0.11383119
maqcol  0.91383599 -0.06576339  0.051793062  0.23412186
valveg  0.88622236  0.10734666  0.066782995 -0.12408519
valani  0.07923622  0.92113868  0.052374666  0.04538699
arexpl  0.09868687  0.93876533 -0.043690165 -0.02131907
distrg -0.08065286 -0.26605704  0.049662267 -0.16641260
renda   0.45572875 -0.04409380 -0.286257339  0.42385544
```

O interessante no pacote PSYCH é que ele fornece alguns comentários junto dos resultados, sugerindo se a estimação dos fatores foi ou não eficiente pelo método escolhido. Contudo, ele não sugere um método, logo, precisamos achar o mais apropriado para tal. Como salientamos, é possível que nossa estimativa estivesse violando alguma hipótese do teorema de Gauss-Markov, e isso foi confirmado pelo comentário do software. O leitor irá notar que ele sugere alguma falha na estimação, mas ainda é muito vago em apontar a razão. Nesse sentido, o pacote oferece a opção de estimarmos os fatores por outros métodos descritos na documentação e presentes no quadro acima. Se o leitor refletir um pouco, irá perceber que uma opção factível seria uma estimativa por Mínimos Quadrados Ponderados. Vamos testar acrescentando à linha de comando a informação do método fatorial desejado, isso é feito pelo acréscimo da parcela fm = "wls"para Mínimos Quadrados Ponderados. Os resultados podem ser encontrados no quadro que segue.

4: Código no R - Exemplo com os COREDES agropecuários do RS (1994/95)

```
##########################################################################################
# Estimando pelo pacote PSYCH por Mínimos Quadrados Ponderados
```

```r
########################################################################################################
fit <- fa(r = R, nfactors = 4, fm = "wls", rotate = "none" )

loads <- fit$loadings[,1:4]
loads
```

```
> fit <- fa(r = R, nfactors = 4, fm = "wls", rotate = "none")

> loads <- fit$loadings[,1:4]
> loads
             WLS1        WLS2        WLS3         WLS4
gado    0.7995574 -0.39707853 -0.18327513 -0.034758772
pop     0.6344374  0.47083162 -0.27269991  0.031933493
pea     0.7288960  0.43503825 -0.29189333  0.014054098
asspec  0.5792522  0.48765015  0.17709380  0.027080188
assagr  0.4844357  0.61783838  0.14420390 -0.139551796
trator  0.9111796  0.07021452  0.19429519  0.004903423
adubos  0.5422019  0.66809712 -0.15151051 -0.010541531
irriga  0.7613059 -0.16926648 -0.10428125 -0.184272365
recveg  0.8724153 -0.16428792  0.09183902  0.148190382
recani  0.8570880 -0.22551270 -0.13963438  0.021354590
financ  0.8614709 -0.23967426  0.06016279  0.116645459
maqcol  0.8336906  0.08570175  0.28326153  0.077490767
valveg  0.8956266 -0.11589507  0.07169035  0.136100250
valani  0.8766755 -0.28401904 -0.13352832 -0.020752682
arexpl  0.8371549 -0.37724442 -0.13826630 -0.017513092
distrg -0.3269868  0.09489752 -0.16752924  0.730492841
renda   0.2936523 -0.02107428  0.71435396  0.094961191
```

O erro que apareceu em Mínimos Quadrados Ordinários não está presente na opção pela estimativa por Mínimos Quadrados Ponderados. Contudo, antes de prosseguirmos na interpretação, vamos rotacionar os fatores pelo método VARIMAX a fim de melhorar os resultados para uma análise mais objetiva. Para tal, vamos apenas acrescentar à linha de comando a informação, por meio do comando "rotate", da opção desejada. O quadro que segue traz apenas os resultados, uma vez que o resultado da linha de comando digitada aparece e o leitor já está mais familiarizado com o R e sua sintaxe.

5: Código no R - Exemplo com os COREDES agropecuários do RS (1994/95)

```
##############################################################################################
# Estimando pelo pacote PSYCH por Mínimos Quadrados Ponderados com Rotação VARIMAX
##############################################################################################
> fit <- fa(r = R, nfactors = 4, fm = "wls", rotate = "Varimax")

> loads <- fit$loadings[,1:4]
> loads
```

```
            WLS1        WLS2         WLS3         WLS4
gado    0.90033972  0.07310739 -0.050433944  0.115158058
pop     0.34800061  0.75608066 -0.077376577 -0.028787732
pea     0.44706768  0.77411057 -0.082624400 -0.001866452
asspec  0.22143890  0.66178636  0.340757598  0.048158603
assagr  0.06494693  0.73813523  0.269305311  0.187308867
trator  0.70582578  0.45791911  0.383996636  0.132739239
adubos  0.14814382  0.86049043  0.028887050  0.013354627
irriga  0.72700757  0.23941504  0.009041239  0.259089573
recveg  0.82041381  0.25366948  0.284156026 -0.017945906
recani  0.86202047  0.23885077  0.029954621  0.066206303
financ  0.85067448  0.19009632  0.238584888  0.009755392
maqcol  0.62564580  0.41655328  0.468064742  0.067770601
valveg  0.81788180  0.31027774  0.271937011 -0.009427367
valani  0.90268491  0.19759275  0.026352100  0.113621160
arexpl  0.91685250  0.09942358  0.005557549  0.109688467
distrg -0.22552807 -0.05815233 -0.078525040 -0.785621378
renda   0.15565093 -0.01163896  0.759514514  0.069065452
```

Fazendo a rotação dos fatores, a redistribuição das cargas pelo método VARIMAX ressaltou a relação de algumas variáveis com os fatores e alterou alguns sinais. Especialmente o Fator 4, mostrou-se que a distância do porto de Rio Grande agora tem forte relação. Esse fator pode ser agora visto como o fator *espacial* da análise. E o sinal negativo vai ao encontro da teoria, sugerindo que quanto maior a distância do município maior os custos com transporte. O Fator 3, quando comparado ao método da MLE, parece ter mudado em parte suas características. Ele passou a ter apenas a variável renda (parcela Renda do IDESE calculado pela FEE), sugerindo ser uma fator que explica o peso desses municípios em relação a capacidade de geração de renda. Contudo, o Fator 1 ainda contém uma parcela grande de variáveis correlacionadas, que misturam a produção pecuária com agrícola. Nesse sentido, talvez a opção por cinco fatores talvez possa tornar as relações entre as variáveis com os fatores e entre si mais clara. Vejamos os resultados dos fatores rotacionados no quadro a seguir.

6: Código no R - Exemplo com os COREDES agropecuários do RS (1994/95)

```
##########################################################################################
# Estimando pelo pacote PSYCH por Mínimos Quadrados Ponderados e 5 fatores
↪  rotacionados (VARIMAX)
##########################################################################################
fit <- fa(r = R, nfactors = 5, fm = "wls", rotate = "Varimax" )

loads <- fit$loadings[,1:5]
loads
```

```
            WLS1        WLS2        WLS5        WLS3         WLS4
gado    0.93106618  0.08116621  0.24567853 -0.01079748  0.098973477
pop     0.20475843  0.80038892  0.31252374 -0.19743247  0.002724358
pea     0.39899155  0.84333386  0.19110680 -0.05393408 -0.009938117
asspec  0.22802149  0.63432092  0.12744484  0.58245966  0.009383182
assagr -0.06744467  0.69582240  0.27870987  0.26933601  0.251924766
trator  0.45641711  0.39939528  0.69445517  0.24539440  0.208901672
adubos  0.09703351  0.91324068  0.11419449  0.10593967  0.023279020
irriga  0.56659369  0.23216092  0.46185234 -0.19291385  0.379509101
recveg  0.56074842  0.19060749  0.75266447  0.05907173 -0.007416980
recani  0.88923457  0.24816746  0.25428903  0.13337700  0.023879094
financ  0.60969100  0.13036171  0.72128622  0.01309998  0.026778911
magcol  0.34425053  0.33794678  0.73938122  0.30381185  0.117864414
valveg  0.56002417  0.25460635  0.74580021  0.05166813  0.003767913
valani  0.90722091  0.20002125  0.30631288  0.07174342  0.100007805
arexpl  0.92039048  0.09569265  0.31091804  0.02706025  0.095257244
distrg -0.21518760 -0.04036079 -0.07469911 -0.07604765 -0.789273089
renda   0.03738737 -0.09426502  0.37861735  0.65175450  0.094212796
```

Se a intenção de adicionar um quinto fator era dividir as variáveis que estavam agrupadas no antigo Fator 1, o resultado parece promissor. O Fator 1 manteve as variáveis gado, irriga, recani, valani e arexpl, enquanto o Fator 5 passou a abrigar as variáveis trator, recveg, financ, magcol e valveg. Isso nos leva a deduzir que o Fator 5 é aquele que guarda as variáveis com características relacionadas à agricultura, ao passo que o Fator 1 manteve as variáveis relacionadas à pecuária, mais a variável irrigação[2]. O Fator 2, ainda tem as variáveis pop e pea, bem como asspec, assagr e adubos. Nos Fatores 3 e 4 não houve mudança, permanecendo renda e distrg alocadas nesses fatores, respectivamente. Mesmo que tenhamos adicionado um quinto fator, os resultados parecem mais nítidos e corroboram com as teorias de Graziando da Silva[3].

Finalmente, o leitor deve estar se perguntando, o quão próximo da matriz de covariância-variâncias essa estimativa está. O conceito de matriz residual advém da ideia de variância específica (*uniqueness*), tal que esse último que varia entre 0 e 1, e é referido como o erro correspondente a parte da variância que não pode ser explicada pela combinação linear dos fatores. Lembre que a matriz de variância estimada é dada pela expressão (NUMERO $\hat{\mathbf{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}^T + \hat{\mathbf{\Psi}}$, e de posse da matriz de escores fatoriais e de variância específica, podemos comparar o resultado do modelo $(\hat{\mathbf{\Sigma}})$ com $(\mathbf{\Sigma})$. Tal que isso pode ser sumarizado no código que segue.

7: Código no R - Exemplo com os COREDES agropecuários do RS (1994/95)

```
Lambda <- food.fa$loadings.        #matriz de escores fatoriais
Psi <- diag(food.fa$uniquenesses) #matriz de var. específica
S <- food.fa$correlation          #matriz de correlações original
Sigma <- Lambda %*% t(Lambda) + Psi

Sigma
```

2.2 Métodos de Estimação dos Parâmetros

A proporção da variância de cada variável a ser explicada pelos fatores nada mais é que o quadrado da carga fatorial da variável estimada em cada fator. Essa carga é conhecida por *comunalidade* e é definida pela notação h^2 como vimos. Essa será a proporção

[2]O leitor pode estar se perguntando a razão dessa variável estar no Fator 1 não no Fator 5, mas a resposta é simples quando observamos os escores fatoriais de algumas cidades que se relacionam positivamente com esses fatores. Nesse sentido, o uso de irrigação à época era comum em áreas de produção de arroz, as quais estão localizadas nos município que reúnem os maiores rebanhos bovinos, bem como dimensões territoriais.

[3]A agropecuária precisou se voltar para um crescimento da produção que fosse oriundo de rendimentos físicos, calcados na produtividade e racionalização do uso do solo (GRAZIANO DA SILVA, 1996)

da variância de uma variável explicada por todos os fatores. Por sua vez, um autovalor é uma solução algébrica para proporção da variância explicada dada por cada um dos fatores. O autovalor é a soma do quadrado das cargas fatoriais do respectivo fator. Portanto, ele é a medida da importância relativa de cada fator. Sendo assim, já vimos que, matematicamente, a variância total de um conjunto de variáveis é igual ao número total de variáveis.

Matematicamente, os métodos para estimação dos fatores diferem na maneira de como alocar os fatores que melhor reproduzem a matriz de correlações original. Na literatura, existe uma clara dicotomia entre métodos, variando entre Máxima verossimilhança (ou *Maximum Likelihood* - ML) e mínimos quadrados. Os estimadores por ML buscam reproduzir a matriz populacional que mais se aproxima da matriz amostral, enquanto em mínimos quadrados o método procura reproduzir a matriz de correlações amostral.

O foco em reproduzir a matriz populacional via ML levanta duas hipóteses para sua operacionalização: (a) os dados são uma amostra aleatória da população; (b) as variáveis mensuradas seguem uma distribuição normal multivariada. De outro lado, em mínimos quadrados não se assume uma distribuição desejável para o conjunto (FABRIGAR and WEGNER, 2009).

Em geral, mínimos quadrados superam em performance o ML quando os fatores são relativamente fracos, quando o modelo definido esta equivocado (muitos fatores especificados), ou amostra é pequena (n = 100). Assim, Briggs and MacCallum (2003) recomendam o uso de OLS para análise fatorial para aumentar a verossimilhança nos fatores comuns quando recuperados. Então, o ML seria apropriado para grandes amostras com dados de distribuição normal e fatores fortes, enquanto OLS seria preferível para amostras pequenas com distribuição não normal e fatores mais fracos (de Winter et al., 2009; Watson, 2017)

Alguns pesquisadores preferem a estimação dos fatores por ML pois esse método procura por uma generalização da população além de permitir que os parâmetros tenham um maior rigor nos testes estatísticos (RMSEA[4] - *based maximum-likelihood method.* Uma discussão mais apurada sobre comparação de modelos pode ser vista em Matsunaga (2010), Park et al. (2002), Zwick and Velicer (1986) e Turner (1998). De outro lado, outros pesquisadores preferem o método dos mínimos quadrados uma vez que não presume hipóteses da distribuição e são sensíveis a fatores mais fracos. Osborne and Banjanovic (2016, p. 26)

[4]Método utilizado em grandes amostras, o RMSEA e amplamente utilizado em Modelos de Equações Estruturais a fim de fornecer o grau de ajuste da amostra mediante a utilização de estatísticas chi-quadrado Steiger (1980)

concluíram que existe um consenso na literatura que o ML é preferido para amostras que exibem uma distribuição multivariada normal enquanto o mínimos quadrados é útil na violação dessa hipótese. Outros autores vão ao encontro dessa ideia, e mais detalhes podem ser consultados em Schmitt (2011), Bandalos and Gerstner (2016) e Sakaluk and Short (2017). Conforme lembra Tabachnick et al. (2007), os diferentes métodos de estimativa produzem resultados muito semelhantes na maior parte dos casos, ficando a cargo do pesquisador e do objeto de sua pesquisa a decisão de qual será o modelo mais apropriado para representar a realidade de sua amostra.

2.2.1 Métodos de Mínimos Quadrados Ponderados

Sobre a mudança de método, a explicação pode estar nos erros, isto é, estamos dizendo que o erro não tem variância constante. Essa constatação viola a a hipótese de homocedasticidade e devemos optar por um método de estimação das cargas fatoriais por um método alternativo. Nesse sentido, o modelo Mínimos Quadrados Ponderados serve para o caso em que os erros apresentem heterocedasticidade. Isso, viola a ideia de que os resíduos seguem uma distribuição normal (multivariada) conforme esperamos por Gauss-Markov. Matematicamente, procuramos um argumento que minimize a soma dos quadrados dos erros ponderados por um peso ω na diagonal principal da matriz dos resíduos, tal que

$$\widehat{\beta}_{WLS} = arg\ min \sum_{i=1}^{n} \epsilon_i^{*2}$$

com isso temos que a Soma do Quadrado dos Resíduos (Residual Sum of Squares - RSS), que em Mínimos Quadrados Ordinários era dado por

$$RSS(\beta) = \sum_{i=1}^{n} \left(y_i - x_i^* \beta\right)^2$$

ao passo que para Soma do Quadrado dos Resíduos por Mínimos Quadrados Ponderados (WSS) seria encontra por

$$WSS(\beta) = \sum_{i=1}^{n} w_i \left(y_i - x_i^* \beta\right)^2$$

Matricialmente, o formato é muito semelhante aos mínimos quadrados, exceto pela presença da matriz diagonal $\mathbf{\Omega}$ de pesos, isto é

$$\widehat{\beta}_{WLS} = \left(\mathbf{X}^T \mathbf{\Omega} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{\Omega} \mathbf{Y}.$$

tal que

$$\mathbf{\Omega} = \begin{bmatrix} \Omega_1 & 0 & \dots & 0 & 0 \\ 0 & \Omega_2 & \dots & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \Omega_{n-1} & \vdots \\ 0 & 0 & \dots & \dots & \Omega_n \end{bmatrix}$$

tal que cada elemento da diagonal principal de será o inverso multiplicativo da variância, isto é,

$$\Omega_i = \frac{1}{\sigma_i^2}, \qquad \forall\, i = 1, 2, \dots, nk.$$

tal que erros pequenos terão pesos grandes ao passo erros maiores deverão ter pesos menores. E a sua matriz de variâncias-covariâncias pode ser estimada por

$$\sigma^2_{p\times p}\left\{\widehat{\beta}_{WLS}\right\} = \left(\mathbf{X}^T\mathbf{\Omega}\mathbf{X}\right)^{-1}$$

2.2.2 Métodos da Máxima Verossimilhança

Uma alternativa, muito difundida para a estimação de parâmetros, posto um conjunto de dados e um modelo estatístico, é o método de máxima verossimilhança[5] estima os valores dos diferentes parâmetros do modelo estatístico de maneira a maximizar a probabilidade dos dados observados (isto é, busca parâmetros que maximizem a função de verossimilhança). O método de máxima verossimilhança apresenta-se como um método geral para estimação de parâmetros, principalmente no caso de variáveis com distribuição normal.

Esse método consiste na escolha de um conjunto de valores para os parâmetros que torne um máximo a função de verossimilhança. Nesse sentido, a função de máxima verossimilhança mostra quão próxima a amostra está dos prováveis valores para os parâmetros, isto é, ela traz os parâmetros que estão mais próximos da população dada uma amostra.

Por definição, temos que verossimilhança é "*the likelihood that any parameter (or set of parameters) should have any assigned value (or set of values) is porportional to the probability that if this were so, the totality of observations should be that observed*"(Fisher, 1922). Então, uma amostra caracterizada por um vetor de variáveis $X_1, X_2, \dots, X_k$ normalmente distribuídas, tem função densidade de probabilidade $f(x_i|\theta)$, com $i = 1, 2, \dots n$, tal que θ será o vetor $(n \times 1)$ de parâmetros que caracterizam $f(x_i|\theta)$. Se $X_i \sim N_k\left(\mu, \sigma^2\right)$ então

[5] Este item seguiu a abordagem em Zivot (2009).

$f(x_i|\theta) = \left(2\pi\sigma^2\right)^{-1/2} \exp^{\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right)}$ e $\theta = \left(\mu, \sigma^2\right)^T$. Assim, uma função densidade de probabilidade conjunta da amostra é dada por

$$\prod_{i=1}^{n} f(x_1, x_2, \ldots, x_n; \theta_1, \ldots, \theta_m)$$

A função de densidade de probabilidade conjunta[6] com dimensão n nada mais que é o produto das densidades de cada uma das observações,

$$f(x_1, \theta) \times f(x_2, \theta) \times \ldots \times f(x_n, \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

em que os parâmetros $\theta_1, \ldots, \theta_m$ são desconhecidos e X_i é a variável aleatória. Assim, a função de máxima verosimilhança é definida pela densidade conjunta como função dos parâmetros theta

$$L(\theta\,|\,x_1, \ldots, x_n) = f(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta).$$

O vetor de parâmetros theta maximizam a função sobre uma amostra com distribuição normal, com $\theta = \left(\mu, \sigma^2\right)$ e que $\vec{x} = (x_1, \ldots, x_n)$ denote um vetor da amostra observada, será

$$\begin{aligned} L(\theta\,|\,\vec{x}) &= \prod_{i=1}^{n} \left(2\pi\sigma^2\right)^{-1/2} \exp^{\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right)} \\ &= \left(2\pi\sigma^2\right)^{-n/2} \exp^{\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)} \end{aligned}$$

e para uma amostra multivariada com k variáveis podemos reapresentar por

$$L\left(\vec{\mu}, \boldsymbol{\Sigma}\right) = \prod_{i=1}^{n} q\left(x_i\,|\,\vec{\mu}, \boldsymbol{\Sigma}\right) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{\frac{k}{2}}\,|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp^{\left\{-\frac{1}{2}(x_i-\vec{\mu})^T \boldsymbol{\Sigma}^{-1}(x_i-\vec{\mu})\right\}}$$

k dimensão do vetor de médias, tal que $\boldsymbol{\Sigma}$ é uma matriz simétrica positiva definida e dimensão $k \times k$ e n número de casos

$$\ln L\left(\vec{\mu}, \boldsymbol{\Sigma}\right) = -\frac{nk}{2}\ln(2\pi) - \frac{n}{2}\ln(|\boldsymbol{\Sigma}|) - \frac{1}{2}\sum_{i=1}^{n}\left\{\left(x_i - \vec{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(x_i - \vec{\mu}\right)\right\}$$

Com um olhar voltado para modelos lineares de regressão, com erros normalmente

[6] Se $X_1, \ldots, X_n$ são variáveis aleatórias discretas, então $f(x_1, \ldots, x_n; \theta) = \text{Prob}(X_1 = x_1, \ldots, X_n = x_n)$.

distribuídos[7], partimos de

$$y_i = \underset{(1\times k)}{x_i^T}\underset{(1\times k)}{\beta} + \epsilon_i, \quad i = 1,\ldots,n$$

$$\epsilon_i|x_i \sim \ iid\ N\left(0,\sigma^2\right)$$

A função densidade de probabilidade $\epsilon_i|x_i$ é

$$f\left(\epsilon_i|x_i;\sigma^2\right) = \left(2\pi\sigma^2\right)^{-1/2} \exp^{\left(-\frac{1}{2\sigma^2}\epsilon_i^2\right)}$$

e sendo a função densidade de probabilidade de $y_i|x_i$ normal com média $x_i^T\beta$ e variância σ^2:

$$f\left(y_i|x_i;\theta\right) = \left(2\pi\sigma^2\right)^{-1/2} \exp^{\left(-\frac{1}{2\sigma^2}\left(y_i - x_i^T\beta\right)^2\right)}$$

tal que $\theta = \left(\beta^T, \sigma^2\right)$. Dada uma amostra iid com n observações, $\vec{x}$ e $\mathbf{X}$, a função densidade conjunta da amostra será

$$f\left(\vec{y}|\mathbf{X};\theta\right) = \left(2\pi\sigma^2\right)^{-n/2} \exp^{\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - x_i^T\beta\right)^2\right)}$$

$$= \left(2\pi\sigma^2\right)^{-n/2} \exp^{\left(-\frac{1}{2\sigma^2}\left(\vec{y}-\mathbf{X}\vec{\beta}\right)^T\left(\vec{y}-\mathbf{X}\vec{\beta}\right)\right)}$$

Para a função log-máxima verossimilhança, tem-se

$$\ln L\left(\theta|\vec{y},\mathbf{X}\right) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\left(\vec{y}-\mathbf{X}\vec{\beta}\right)^T\left(\vec{y}-\mathbf{X}\vec{\beta}\right)$$

O Estimador de Máxima Verossimilhança de θ irá satisfazer $S\left(\hat{\theta}|\vec{y},\mathbf{X}\right) = \frac{\partial}{\partial\theta}\ln L\left(\theta|\vec{y},\mathbf{X}\right)$ é o vetor de escores. Logo

$$\frac{\partial \ln L\left(\theta|\vec{y},\mathbf{X}\right)}{\partial\beta} = \frac{1}{2\sigma^2}\frac{\partial}{\partial\vec{\beta}}\left[\vec{y}^T\vec{y} - 2\vec{y}^T\mathbf{X}\vec{\beta} + \vec{\beta}^T\mathbf{X}^T\mathbf{X}\vec{\beta}\right]$$

$$= -\left(\sigma^2\right)^{-1}\left[-\mathbf{X}^T\vec{y} + \mathbf{X}^T\mathbf{X}\vec{\beta}\right]$$

$$\frac{\partial \ln L\left(\theta|\vec{y},\mathbf{X}\right)}{\partial\sigma^2} = -\frac{n}{2}\left(\sigma^2\right)^{-1} + \frac{1}{2}\left(\sigma^2\right)^{-2}\left(\vec{y}-\mathbf{X}\vec{\beta}\right)^T\left(\vec{y}-\mathbf{X}\vec{\beta}\right)$$

[7] Tal que *iid* significa independente e identicamente distribuída.

Referências

Bandalos, D. L. and Gerstner, J. J. (2016). Using factor analysis in test construction.

Briggs, N. E. and MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research*, 38(1):25–56.

Bruin, J. (2011). newtest: command to compute new test @online. UCLA: Statiscal Consulting Group. `https://stats.idre.ucla.edu/spss/seminars/introduction-to-factor-analysis/a-practical-introduction-to-factor-analysis/`.

DARTON, R. A. (1980). Rotation in factor analysis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 29(3):167–194.

de Winter, J. C., Dodou, D., and Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate behavioral research*, 44(2):147–181.

FABRIGAR, L. R. and WEGNER, D. T. (2009). *Structural Equation Modeling*, volume 5. Routledge.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222(594-604):309–368.

GRAZIANO DA SILVA, J. (1996). A nova dinâmica da agricultura brasileira: Campinas. *Editora da.*

Hair, J., Black, B., and Babin, B. (2006). *Multivariate Analysis.* Prentice Hall.

JOHNSON, R. A. and WICHERN, D. W. (1992). *Applied multivariate statistical analysis*, volume 3. Prentice hall Upper Saddle River, NJ.

Jolliffe, I. (2003). Principal component analysis. *Technometrics*, 45(3):276.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1):31–36.

Matsunaga, M. (2010). How to factor-analyze your data right: Do's, don'ts, and how-to's. *International journal of psychological research*, 3(1):97–110.

Osborne, J. W. and Banjanovic, E. S. (2016). *Exploratory factor analysis with SAS*. Sas Institute.

Park, H. S., Dailey, R., and Lemus, D. (2002). The use of exploratory factor analysis and principal components analysis in communication research. *Human Communication Research*, 28(4):562–577.

Sakaluk, J. K. and Short, S. D. (2017). A methodological review of exploratory factor analysis in sexuality research: Used practices, best practices, and data analysis resources. *The Journal of Sex Research*, 54(1):1–9.

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, 29(4):304–321.

Steiger, J. H. (1980). Statistically based tests for the number of common factors. In *the annual meeting of the Psychometric Society. Iowa City, IA. 1980*.

Tabachnick, B. G., Fidell, L. S., and Ullman, J. B. (2007). *Using multivariate statistics*, volume 5. Pearson Boston, MA.

Turner, N. E. (1998). The effect of common variance and structure pattern on random data eigenvalues: Implications for the accuracy of parallel analysis. *Educational and Psychological Measurement*, 58(4):541–568.

Watson, J. C. (2017). Establishing evidence for internal structure using exploratory factor analysis. *Measurement and Evaluation in Counseling and Development*, 50(4):232–238.

Zivot, E. (2009). Maximum likelihood estimation. *Lecture Notes on course"Econometric Theory I: Estimation and Inference (first quarter, second year PhD)", University of Washington, Seattle, Washington, USA*.

Zwick, W. R. and Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological bulletin*, 99(3):432.

$rafael$