Análise Fatorial

Análise Fatorial

- **Objetivo:** Descrever as relações de covariância entre muitas variáveis em termos de poucas quantidades aleatórias subjacentes e não observáveis

Análise Fatorial

- **Objetivo:** Descrever as relações de covariância entre muitas variáveis em termos de poucas quantidades aleatórias subjacentes e não observáveis
- **Motivação:** Variáveis de um grupo altamente correlacionadas entre si, mas com pequenas correlações de outros grupos.
  - É concebível que cada grupo de variáveis represente um fator (ou construto) que seja o responsável pelas correlações observadas

Análise Fatorial

Vamos supor cinco variáveis Y_1, Y_2, Y_3, Y_4 e Y_5 correspondentes às notas de estudantes em Espanhol, Inglês, Química, Física e Matemática, respectivamente.

Table 1: Matriz de correlações hipotética

	Y_1	Y_2	Y_3	Y_4	Y_5
Y_1	1	0,90	0,05	0,05	0,10
Y_2	0,90	1	0,05	0,10	0,05
Y_3	0,05	0,05	1	0,95	0,85
Y_4	0,05	0,10	0,95	1	0,90
Y_5	0,10	0,05	0,85	0,90	1

Análise Fatorial

Com base nas correlações, poderíamos imaginar dois fatores subjacentes, e não observáveis:

- **Fator 1:** Conhecimento em línguas - associado aos desempenhos em Espanhol e Inglês, fortemente correlacionados;

Análise Fatorial

Com base nas correlações, poderíamos imaginar dois fatores subjacentes, e não observáveis:

- **Fator 1:** Conhecimento em línguas - associado aos desempenhos em Espanhol e Inglês, fortemente correlacionados;
- **Fator 2:** Conhecimento em ciências exatas - associado aos desempenhos em Química, Física e Matemática.

Análise Fatorial

Com base nas correlações, poderíamos imaginar dois fatores subjacentes, e não observáveis:

- **Fator 1:** Conhecimento em línguas - associado aos desempenhos em Espanhol e Inglês, fortemente correlacionados;
- **Fator 2:** Conhecimento em ciências exatas - associado aos desempenhos em Química, Física e Matemática.

É importante compreender que "conhecimento" é algo que não se mede diretamente, mas que pode se manifestar através de variáveis observáveis, como as notas obtidas nos exames.

Análise Fatorial

Pode ser considerada uma extensão da Análise de Componentes Principais

Análise Fatorial

Pode ser considerada uma extensão da Análise de Componentes Principais

- Ambas são tentativas de aproximar S.

Análise Fatorial

Pode ser considerada uma extensão da Análise de Componentes Principais

- Ambas são tentativas de aproximar S.
- A aproximação baseada em Análise Fatorial é mais elaborada

Análise Fatorial

Pode ser considerada uma extensão da Análise de Componentes Principais

- Ambas são tentativas de aproximar S.
- A aproximação baseada em Análise Fatorial é mais elaborada

Questão principal:

- Dados são consistentes com a estrutura prescrita?

Tipos de Análise Fatorial

Análise Fatorial Exploratória

- Busca encontrar os fatores subjacentes às variáveis originais amostradas
- Em geral, efetuada quando não se tem noção clara da quantidade de fatores do modelo e nem do que representam

Análise Fatorial Confirmatória

- Tem-se em mãos um modelo fatorial pré-especificado (modelo hipotético) e deseja-se verificar se é aplicável ou consistente com os dados amostrais de que dispõe

Planejamento de uma análise fatorial

- **Verificar a adequabilidade da base de dados:** tamanho da amostra (desejável a razão de 5 indivíduos para cada variável), correlações significativas, KMO adequado.

Planejamento de uma análise fatorial

- **Verificar a adequabilidade da base de dados:** tamanho da amostra (desejável a razão de 5 indivíduos para cada variável), correlações significativas, KMO adequado.
- **Determinar o método de extração e o número de fatores a serem extraídos:** métodos de extração (componentes principais, fatores principais, máxima verossimilhança), número de fatores (critério de kaiser, screeplot, variância acumulada).

Planejamento de uma análise fatorial

- **Verificar a adequabilidade da base de dados:** tamanho da amostra (desejável a razão de 5 indivíduos para cada variável), correlações significativas, KMO adequado.
- **Determinar o método de extração e o número de fatores a serem extraídos:** métodos de extração (componentes principais, fatores principais, máxima verossimilhança), número de fatores (critério de kaiser, screeplot, variância acumulada).
- **Decidir o tipo de rotação:** Varimax, Quartimax, Orthomax...

Planejamento de uma análise fatorial

GIGO: Garbage In, Garbage Out!

Cuidado!

- A análise fatorial sempre produzirá fatores. Desta forma, ela é sempre um candidato para o fenômeno **"lixo dentro, lixo fora"**.
- Se o pesquisador incluir indiscriminadamente um grande número de variáveis e esperar que a análise fatorial "faça revelações", a possibilidade de resultados pobres será alta.
- A qualidade e o significado dos fatores determinados refletem **estruturas conceituais** das variáveis incluídas na análise (base conceitual).

Planejamento de uma análise fatorial

Questões conceituais...

- É responsabilidade do pesquisador garantir que os padrões observados sejam conceitualmente válidos. A AF é capaz somente de determinar a adequação das correlações entre variáveis.
- O pesquisador deve garantir que a amostra esteja adequada à estrutura fatorial inerente.
- Por exemplo, ao avaliar os fatores de risco de uma acidente vascular, se nenhuma variável relacionada aos fatores genéticos for incluída, a AF não será capaz de identificar esta dimensão.

Planejamento de uma análise fatorial

Questões conceituais...

- A qualidade e o significado dos fatores obtidos refletem as bases conceituais das variáveis incluídas na análise.
- O pesquisador não deve incluir indiscriminadamente inúmeras variáveis e esperar que a AF "arrume as coisas".
- No exemplo da avaliação dos fatores de risco de uma acidente vascular, incluir a variável, por exemplo, "cor dos olhos" interfere nos resultados da AF e não traz nenhuma informação relevante para o objeto de estudo.

Planejamento de uma análise fatorial

Questões estatísticas...

- Espera-se que haja correlações significativas, pois o objetivo é identificar conjuntos de variáveis inter-relacionadas.
- Se todas as correlações são pequenas ou iguais, o pesquisador deve questionar o uso da análise fatorial.
- Inspeção visual da matriz de correlação – coeficientes pelo menos em torno de 0,30.

Planejamento de uma análise fatorial

- **Teste de esfericidade de Bartlett (BTS):** O teste de esfericidade de Bartlett testa a hipótese de que as variáveis não sejam correlacionadas na população.

$$H_0 : P = I$$

A estatística do teste é dada por

$$\chi^2 = -\left[(n-1) - \frac{2p+5}{6}\right] \ln |R|$$

que tem uma distribuição qui-quadrado com ν graus de liberdade, sendo $\nu = \frac{p(p-1)}{2}$, n igual ao tamanho da amostra, p igual ao número de variáveis e $|R|$, o determinante da matriz de correlação amostral.

Planejamento de uma análise fatorial

- **Medida de adequação da amostra:** comumente utilizado o KMO (Kaiser-Meyer-Olkin). Rencher(2002) sugere que para um modelo de Análise Fatorial possa ser ajustado adequadamente aos dados é necessário que R^{-1} seja próxima da matriz diagonal. O coeficiente KMO baseia-se nesse princípio

$$KMO = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} r_{ij}^2}{\sum_{i=1}^{p}\sum_{j=1}^{p} r_{ij}^2 + \sum_{i=1}^{p}\sum_{j=1}^{p} v_{ij}^2},$$

Planejamento de uma análise fatorial

em que r_{ij} e v_{ij} são, respectivamente, os elementos na posição (i, j) da matriz de correlações amostral, R, e da matriz $V = UR^{-1}U$, na qual $U = \left[\text{diag}(R^{-1})^{\frac{1}{2}}\right]^{-1}$. Note que $\text{diag}(R^{-1})$ é a matriz cuja diagonal coincide com a diagonal de R^{-1} mas com os demais elementos nulos e $\text{diag}(R^{-1})^{\frac{1}{2}}$ é a matriz diagonal cujo i-ésimo elemento diagonal é a raiz quadrada do i-ésimo elemento diagonal de $\text{diag}(R^{-1})$. A matriz V é usualmente designada por matriz de correlação anti-imagem.

Planejamento de uma análise fatorial

- **Medida de adequação da amostra:** Para cada variável: MSA (Measure of sampling adequacy). Essa medida é bastante similar ao KMO. Novamente, desejamos verificar a possibilidade de existência de uma estrutura fatorial nos dados. A MSA deve ser calculada separadamente para cada variável, através de

$$MSA_i = \frac{\sum_{j=1}^{p} r_{ij}^2}{\sum_{j=1}^{p} r_{ij}^2 + \sum_{j=1}^{p} v_{ij}^2}.$$

Planejamento de uma análise fatorial

- O objetivo é verificar se uma dada variável pode ser explicada pelas demais, o que é esperado em um modelo fatorial.

Planejamento de uma análise fatorial

- O objetivo é verificar se uma dada variável pode ser explicada pelas demais, o que é esperado em um modelo fatorial.
- Valores baixos de MSA_i são indícios de que a respectiva variável pode ser retirada da análise, sem maiores prejuízos.

Planejamento de uma análise fatorial

- O objetivo é verificar se uma dada variável pode ser explicada pelas demais, o que é esperado em um modelo fatorial.
- Valores baixos de MSA_i são indícios de que a respectiva variável pode ser retirada da análise, sem maiores prejuízos.
- Note que esta medida é bastante similar ao KMO, apresentando uma lógica semelhante.

Planejamento de uma análise fatorial

Ambos variam de 0 a 1, sendo:

Table 2: Tabela de interpretação do índice KMO/MSA

KMO/MSA	Interpretação
$0,80 - 1,00$	Excelente
$0,70 - 0,80$	Ótimo
$0,60 - 0,70$	Bom
$0,50 - 0,60$	Regular
$0,00 - 0,50$	Insuficiente

O modelo fatorial ortogonal via matriz de correlações

- Seja o vetor aleatório $\mathbf{x} = [X_1, X_2, \cdots, X_p]^t$, com vetor de médias μ, matriz de covariâncias Σ e matriz de correlações P.

O modelo fatorial ortogonal via matriz de correlações

- Seja o vetor aleatório $\mathbf{x} = [X_1, X_2, \cdots, X_p]^t$, com vetor de médias μ, matriz de covariâncias Σ e matriz de correlações P.
- Sejam as variáveis originais padronizadas: $Z_i = \dfrac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}$

O modelo fatorial ortogonal via matriz de correlações

- Seja o vetor aleatório $\mathbf{x} = [X_1, X_2, \cdots, X_p]^t$, com vetor de médias μ, matriz de covariâncias Σ e matriz de correlações P.
- Sejam as variáveis originais padronizadas: $Z_i = \dfrac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}$
- P é a matriz de covariâncias do velor aleatório z, cujas componentes são as variáveis padronizadas.

O modelo fatorial ortogonal via matriz de correlações

- Modelo fatorial ortogonal
  - Construído via a matriz de correlação populacional
  - Relaciona linearmente as variáveis padronizadas e os m fatores comuns (que são desconhecidos)
  - Fatores são variáveis independentes

O modelo fatorial ortogonal via matriz de correlações

Na análise fatorial, vamos expressar cada uma das p variáveis padronizadas como combinação linear de $m < p$ fatores comuns $F_1, F_2, \cdots, F_m$ a menos de um termo correspondente ao erro $\epsilon_1, \epsilon_2, \cdots, \epsilon_p$:

$$
\begin{aligned}
Z_1 &= l_{11}F_1 + l_{12}F_2 + \cdots + l_{1m}F_m + \epsilon_1 \\
Z_2 &= l_{21}F_1 + l_{22}F_2 + \cdots + l_{2m}F_m + \epsilon_2 \\
&\vdots \\
Z_p &= l_{p1}F_1 + l_{p2}F_2 + \cdots + l_{pm}F_m + \epsilon_p
\end{aligned}
$$

O modelo fatorial ortogonal

- $F_1, F_2, \cdots, F_m$ são os fatores comuns **não observáveis**, ou (*latentes*);
- $\epsilon_1, \epsilon_2, \cdots, \epsilon_p$ são os fatores específicos associados a cada variável e,
- l_{ij}, $i = 1, 2, \cdots, p$, $j = 1, 2, \cdots, m$ são chamados de cargas fatoriais, ou (*loadings*).

O modelo fatorial ortogonal

Notação Matricial

$$\boxed{D^{-1}(\mathbf{x}-\mu) = LF + \epsilon}$$

$$F = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_p \end{bmatrix} \quad L = \begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1m} \\ l_{21} & l_{22} & \cdots & l_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \cdots & l_{pm} \end{bmatrix} \quad D = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p \end{bmatrix}$$

O modelo fatorial ortogonal

Suposições do modelo

Para construção do modelo fatorial ortogonal, algumas suposições se fazem necessárias:

a) Os fatores têm média igual a zero, variâncias iguais a 1 e não são correlacionados:

$$\boxed{E(F) = 0 \text{ e } \mathrm{Var(F)} = \mathrm{E(FF^t)} = \mathrm{I}}$$

O modelo fatorial ortogonal

Suposições do modelo

b) Os erros têm média zero, são não correlacionados e não necessariamente tem a mesma variância:

$$\boxed{E(\epsilon) = 0 e \mathrm{Var}(\epsilon) = \Psi = \mathrm{diag}(\ _1,\ _2, \cdots,\ _\mathrm{p}) = \mathrm{E}(\epsilon\epsilon^\mathrm{t})}$$

c) F e ϵ são independentes, isto é, fontes de variação distintas:

$$\boxed{\mathrm{Cov}(\mathrm{F}, \epsilon) = \mathrm{E}(\epsilon \mathrm{F}^\mathrm{t}) = 0}$$

Propriedades do modelo fatorial

Como consequência da definição do modelo e das suposições apresentadas, decorrem as seguintes propriedades:

$$P = LL^t + \Psi$$

Propriedades do modelo fatorial

Como consequência da definição do modelo e das suposições apresentadas, decorrem as seguintes propriedades:

$$P = LL^t + \Psi$$

- Esta decomposição implica que as variâncias das variáveis observadas podem ser escritas como:

$$\text{Var}(Z_i) = \underbrace{l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2}_{h_i^2} + \psi_i = h_i^2 + \psi_i, \qquad i = 1, \cdots, p$$

Propriedades do modelo fatorial

- h_i^2 é a **comunalidade:** variância comum de z - expressa o quanto da variabiliade de Z_i é explicada pelo modelo
- ψ_i é a **especificidade:** variância específica de cada Z_i - expressa o quanto da variabilidade de Z_i não é explicada pelo modelo.

Propriedades do modelo fatorial

- h_i^2 é a **comunalidade:** variância comum de z - expressa o quanto da variabiliade de Z_i é explicada pelo modelo

- ψ_i é a **especificidade:** variância específica de cada Z_i - expressa o quanto da variabilidade de Z_i não é explicada pelo modelo.

Além disso, as covariâncias das variáveis observadas podem ser escritas como:

$$\text{Cov}(Z_i, Z_k) = l_{i1}l_{k1} + l_{i2}l_{k2} + \cdots + l_{im}l_{km}, \quad i, k = 1, \cdots, p \ (i \neq k)$$

Propriedades do modelo fatorial

- h_i^2 é a **comunalidade:** variância comum de z - expressa o quanto da variabiliade de Z_i é explicada pelo modelo
- ψ_i é a **especificidade:** variância específica de cada Z_i - expressa o quanto da variabilidade de Z_i não é explicada pelo modelo.

Além disso, as covariâncias das variáveis observadas podem ser escritas como:

$$\text{Cov}(Z_i, Z_k) = l_{i1}l_{k1} + l_{i2}l_{k2} + \cdots + l_{im}l_{km}, \quad i, k = 1, \cdots, p \ (i \neq k)$$

e,

$$\text{Cov}(z, F) = \text{Cor}(z, F) = L$$

Métodos de Estimação

- **Método das componentes principais:** baseia-se na análise de componentes principais. Em geral, utilizado como um análise exploratória dos dados, em termos dos fatores subjacentes

Métodos de Estimação

- **Método das componentes principais:** baseia-se na análise de componentes principais. Em geral, utilizado como um análise exploratória dos dados, em termos dos fatores subjacentes
- **Método dos fatores principais:** refina a solução apresentada pelo método das componentes principais

Métodos de Estimação

- **Método das componentes principais:** baseia-se na análise de componentes principais. Em geral, utilizado como um análise exploratória dos dados, em termos dos fatores subjacentes
- **Método dos fatores principais:** refina a solução apresentada pelo método das componentes principais
- **Método da máxima verossimilhança:** supõe que os dados seguem uma distribuição normal multivariada

Métodos de estimação: Componentes Principais

Considerando $m < p$ o número de fatores comuns, a matriz de cargas fatoriais, estimada pelo método dos componentes principais, fica dada por:

$$\hat{L} = \left[\begin{array}{cccc} \sqrt{\hat{\lambda}}_1 \hat{e}_1 & \sqrt{\hat{\lambda}}_2 \hat{e}_2 & \cdots & \sqrt{\hat{\lambda}}_m \hat{e}_m \end{array} \right]$$

Métodos de estimação: Componentes Principais

Considerando $m < p$ o número de fatores comuns, a matriz de cargas fatoriais, estimada pelo método dos componentes principais, fica dada por:

$$\hat{L} = \left[\begin{array}{cccc} \sqrt{\hat{\lambda}}_1 \hat{e}_1 & \sqrt{\hat{\lambda}}_2 \hat{e}_2 & \cdots & \sqrt{\hat{\lambda}}_m \hat{e}_m \end{array} \right]$$

E a matriz de especificidades é dada por:

$$\hat{\Psi} = \mathrm{diag}(\mathrm{R} - \hat{\mathrm{L}}\hat{\mathrm{L}}^{\mathrm{t}})$$

Métodos de estimação: Componentes Principais

Considerando $m < p$ o número de fatores comuns, a matriz de cargas fatoriais, estimada pelo método dos componentes principais, fica dada por:

$$\hat{L} = \left[\begin{array}{cccc} \sqrt{\hat{\lambda}}_1 \hat{e}_1 & \sqrt{\hat{\lambda}}_2 \hat{e}_2 & \cdots & \sqrt{\hat{\lambda}}_m \hat{e}_m \end{array} \right]$$

E a matriz de especificidades é dada por:

$$\hat{\Psi} = \mathrm{diag}(\mathrm{R} - \hat{\mathrm{L}}\hat{\mathrm{L}}^{\mathrm{t}})$$

Aproximação de R

$$R \approx \hat{L}\hat{L}^{t} + \hat{\Psi}$$

Métodos de estimação: Componentes Principais

Matriz residual:

$$MRes = R - (\hat{L}\hat{L}^{t} + \hat{\Psi})$$

Métodos de estimação: Componentes Principais

Matriz residual:

$$MRes = R - (\hat{L}\hat{L}^{t} + \hat{\Psi})$$

- Pode servir como critério de avaliação do modelo
  - Seus valores deveriam ser próximos de zero
  - Matriz é nula somente quando o valor de m é igual a p

Métodos de estimação: Componentes Principais

Matriz residual:

$$MRes = R - (\hat{L}\hat{L}^{t} + \hat{\Psi})$$

- Pode servir como critério de avaliação do modelo
  - Seus valores deveriam ser próximos de zero
  - Matriz é nula somente quando o valor de m é igual a p
- Os elementos da diagonal da matriz R são reproduzidos exatamente pelo modelo
  - O mesmo não ocorre para os outros elementos da matriz R (covariâncias das variáveis Z_i e Z_j)

Métodos de estimação: Componentes Principais

Proporção da variabilidade total explicada pelo fator

$$\text{Proporção explicada}_{F_j} = \frac{\sum_{i=1}^{p} l_{ij}^2}{p}$$

Métodos de estimação: Componentes Principais

Proporção da variabilidade total explicada pelo fator

$$\text{Proporção explicada}_{F_j} = \frac{\sum_{i=1}^{p} l_{ij}^2}{p}$$

- Representa o quanto cada fator consegue captar da variabilidade original das variáveis Z_i.

Métodos de estimação: Fatores Principais

- Também chamado Método de Componentes Principais Iterativo

Métodos de estimação: Fatores Principais

- Também chamado Método de Componentes Principais Iterativo
- **Ideia básica:** Refinar as estimativas de $\hat{L}$ e $\hat{\Psi}$, encontradas a partir do método das componentes principais.

Métodos de estimação: Fatores Principais

- Também chamado Método de Componentes Principais Iterativo
- **Ideia básica:** Refinar as estimativas de $\hat{L}$ e $\hat{\Psi}$, encontradas a partir do método das componentes principais.
- **Procedimento:**
  - Estimativas iniciais pelo método das componentes principais
  - Troca dos elementos da diagonal de R pelas comunalidades estimadas
  - Novas estimações a partir da matriz R^*
  - Substituição dos elementos da diagonal principal pelas comunalidades estimadas
  - Procedimento é repetido até que as diferenças entres as comunalidades estimadas sejam desprezíveis

Métodos de estimação: Máxima Verossimilhança

Só pode ser utilizado quando a forma da distribuição de probabilidades é conhecida

- **Suposição:**
  - Vetor aleatório z tem distribuição normal p-variada

Métodos de estimação: Máxima Verossimilhança

Só pode ser utilizado quando a forma da distribuição de probabilidades é conhecida

- **Suposição:**
  - Vetor aleatório z tem distribuição normal p-variada
- **Consequência:**
  - Vetor das variáveis padronizadas é normal p-variado
  - Fatores tem distribuição normal multivariada com vetor de médias zero e matriz de covariâncias I_m
  - Erros tem distribuição normal p-variada com vetor de médias zero e matriz de covariâncias Ψ.

Métodos de estimação: Máxima Verossimilhança

A função de verossimilhança considerando uma amostra aleatória de tamanho n observada no vetor aleatório z é dada por:

$$L(0,P) = \frac{1}{(2\pi)^{\frac{np}{2}}|(LL^t + \Psi)|^{\frac{n}{2}}} \exp\left\{ -\frac{1}{2}\sum_{j=1}^{n} z_j^t (LL^t + \Psi)^{-1} z_j \right\}$$

Métodos de estimação: Máxima Verossimilhança

A função de verossimilhança considerando uma amostra aleatória de tamanho n observada no vetor aleatório z é dada por:

$$L(0, P) = \frac{1}{(2\pi)^{\frac{np}{2}} |(LL^t + \Psi)|^{\frac{n}{2}}} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{n} z_j^t (LL^t + \Psi)^{-1} z_j \right\}$$

- A função de verossimilhança depende da matrizes L e Ψ, através da matriz de correlação P.

Métodos de estimação: Máxima Verossimilhança

- Os estimadores de máxima verossimilhança de L e Ψ são as matrizes $\hat{L}$ e $\hat{\Psi}$ que **maximizam** a função de verossimilhança $L(0, P)$.

Métodos de estimação: Máxima Verossimilhança

- Os estimadores de máxima verossimilhança de L e Ψ são as matrizes $\hat{L}$ e $\hat{\Psi}$ que **maximizam** a função de verossimilhança $L(0, P)$.
- Maximização é feita por métodos numéricos.

Métodos de estimação: Máxima Verossimilhança

- Os estimadores de máxima verossimilhança de L e Ψ são as matrizes $\hat{L}$ e $\hat{\Psi}$ que **maximizam** a função de verossimilhança $L(0, P)$.
- Maximização é feita por métodos numéricos.
- Método mais sofisticado que os métodos de componentes e fatores principais: produz estimativas mais precisas

Métodos de estimação: Máxima Verossimilhança

Cuidado!

- Está fundamentado na suposição de normalidade multivariada dos vetores z, F e ϵ.
  - Apenas a normalidade do vetor z pode ser investigada a priori a partir dos dados amostrais
  - Fatores e erros são variáveis aleatórias não observáveis

Métodos de estimação: Máxima Verossimilhança

Cuidado!

- Está fundamentado na suposição de normalidade multivariada dos vetores z, F e ϵ.
  - Apenas a normalidade do vetor z pode ser investigada a priori a partir dos dados amostrais
  - Fatores e erros são variáveis aleatórias não observáveis

Observação Importante

- O método de estimação por Máxima Verossimilhança (ML) pode ser utilizado na Análise Fatorial mesmo quando os dados não seguem uma distribuição normal multivariada. A falta de normalidade afeta principalmente os testes de significância (χ^2, IC), mas não compromete a qualidade da extração dos fatores. ML é robusto para escalas Likert com 5 ou mais categorias.

Rotação dos fatores

Uma propriedade importante do modelo fatorial ortogonal é a **falta de unicidade das cargas fatoriais:** se L satisfaz à relação $P = LL^t + \Psi$, então $L^* = LT$, tal que $TT^t = T^tT = I$ (T ortogonal), também satisfaz:

$$\begin{aligned} L^*L^{*t} + \Psi &= (LT)(LT)^t + \Psi = LTT^tL^t + \Psi \\ &= LIL^t + \Psi = LL^t + \Psi = P \end{aligned}$$

Rotação dos fatores

Uma propriedade importante do modelo fatorial ortogonal é a **falta de unicidade das cargas fatoriais:** se L satisfaz à relação $P = LL^t + \Psi$, então $L^* = LT$, tal que $TT^t = T^tT = I$ (T ortogonal), também satisfaz:

$$\begin{aligned} L^*L^{*t} + \Psi &= (LT)(LT)^t + \Psi = LTT^tL^t + \Psi \\ &= LIL^t + \Psi = LL^t + \Psi = P \end{aligned}$$

A matriz de resíduos permanece a mesma ($\hat{h}_i^2$ e $\hat{\Psi}_i$)

$$R - \hat{L}\hat{L}^t - \hat{\Psi} = R - \hat{L}^*\hat{L}^{*t} - \hat{\Psi}$$

Rotação dos fatores

Uma propriedade importante do modelo fatorial ortogonal é a **falta de unicidade das cargas fatoriais:** se L satisfaz à relação $P = LL^t + \Psi$, então $L^* = LT$, tal que $TT^t = T^tT = I$ (T ortogonal), também satisfaz:

$$\begin{aligned} L^*L^{*t} + \Psi &= (LT)(LT)^t + \Psi = LTT^tL^t + \Psi \\ &= LIL^t + \Psi = LL^t + \Psi = P \end{aligned}$$

A matriz de resíduos permanece a mesma ($\hat{h}_i^2$ e $\hat{\Psi}_i$)

$$R - \hat{L}\hat{L}^t - \hat{\Psi} = R - \hat{L}^*\hat{L}^{*t} - \hat{\Psi}$$

- Do ponto de vista estatístico é irrelevante obter $\hat{L}$ ou $\hat{L}^*$

Rotação dos fatores

Comentários

- Uma vez extraídos os fatores, temos a chamada **matriz de cargas fatoriais**. Nesta matriz, cada elemento indica o grau de correspondência entre a variável e o fator;
- Com a rotação, busca-se uma estrutura mais simples: loadings originais podem não ter fácil interpretação
- **Ideal:** encontrar um padrão de loadings tais que cada variável carregue-se fortemente em um único fator (com loadings moderados nos outros fatores)
- Nem sempre é possível obter esta estrutura mais simples

Rotação dos fatores

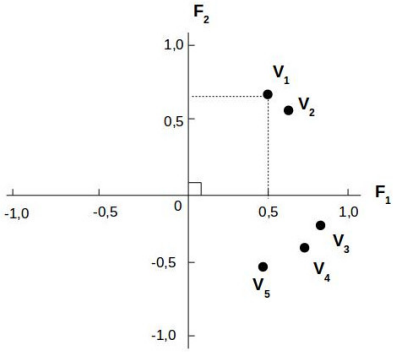
Ao rotacionar a matriz fatorial, busca-se redistribuir a variância dos primeiros fatores para os últimos, a fim de atingir um padrão fatorial mais simples e teoricamente mais significativo.

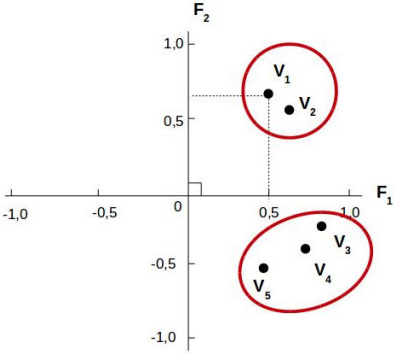
Table 3: Cargas fatoriais dos dois fatores extraídos

Variáveis	Fator 01	Fator 02
V_1	$0,50$	$0,82$
V_2	$0,60$	$0,61$
V_3	$0,91$	$-0,22$
V_4	$0,82$	$-0,30$
V_5	$0,49$	$-0,50$

Rotação dos fatores

2 agrupamentos: As cargas fatoriais, contudo, não parecem tão óbvias



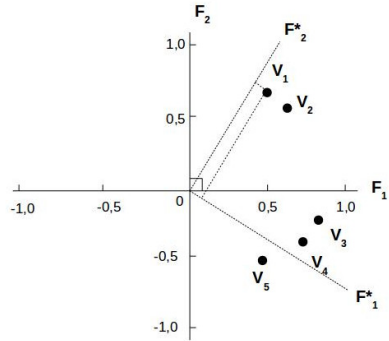


Rotação dos fatores

Rotação Ortogonal

Table 4: Cargas fatoriais estimadas para dois fatores

Variáveis	Fator 01	Fator 02
V_1	$0,03$	$0,90$
V_2	$0,19$	$0,82$
V_3	$0,95$	$0,22$
V_4	$0,89$	$0,12$
V_5	$0,78$	$-0,12$



Rotação dos fatores

- Os métodos de rotação têm como objetivo simplificar as linhas e colunas da matriz fatorial para facilitar a interpretação
  - Maximizar a carga de uma variável em um único fator
  - Reduzir ao máximo o número de variáveis com cargas altas por fator

Rotação dos fatores

- Os métodos de rotação têm como objetivo simplificar as linhas e colunas da matriz fatorial para facilitar a interpretação
  - Maximizar a carga de uma variável em um único fator
  - Reduzir ao máximo o número de variáveis com cargas altas por fator
- Alguns critérios para encontrar matriz ortogonal:
  - Varimax
  - Quartimax
  - Orthomax

Rotação dos fatores

- **Critério Varimax:** É um dos mais utilizados. Em geral, produz soluções mais simples

Rotação dos fatores

- **Critério Varimax:** É um dos mais utilizados. Em geral, produz soluções mais simples

- **Critério Quartimax:** Tem tendência de gerar fatores, onde todas as variáveis têm loadings elevados

Rotação dos fatores

- **Critério Varimax:** É um dos mais utilizados. Em geral, produz soluções mais simples
- **Critério Quartimax:** Tem tendência de gerar fatores, onde todas as variáveis têm loadings elevados
- **Critério Orthomax:** É uma média ponderada dos dois outros métodos

Rotação dos fatores

Comentários

- **Qualidade de ajuste:** A rotação não acrescenta nenhuma melhoria em relação ao ajuste original
  - Matriz residual original não é alterada pela transformação ortogonal
  - Valores estimados de comunalidade e variâncias específicas permanecem inalterados
- **Interpretação:** Novos fatores podem ser de mais fácil interpretação
- Quando a solução sem rotação já é de boa qualidade, não se recomenda rotação: Solução rotacionada pode ser pior

Quantos fatores usar?

Para determinar o valor de m...

- Analisar a proporção da variação total dos dados atribuída ao j-ésimo fator, dada por:

$$\frac{\hat{\lambda}_j}{p}, \quad j = 1, 2, \cdots, m, \text{ se usarmos } R$$

Quantos fatores usar?

Para determinar o valor de m...

- Analisar a proporção da variação total dos dados atribuída ao j-ésimo fator, dada por:

$$\frac{\hat{\lambda}_j}{p}, \quad j = 1, 2, \cdots, m, \text{ se usarmos } R$$

- **Critério de Kaiser:** retenção dos fatores com autovalor associado superior a 1;

Quantos fatores usar?

Para determinar o valor de m...

- Analisar a proporção da variação total dos dados atribuída ao j-ésimo fator, dada por:

$$\frac{\hat{\lambda}_j}{p}, \quad j = 1, 2, \cdots, m, \text{ se usarmos } R$$

- **Critério de Kaiser:** retenção dos fatores com autovalor associado superior a 1;

- Gráfico scree-plot

Quantos fatores usar?

Para determinar o valor de m...

- Analisar a proporção da variação total dos dados atribuída ao j-ésimo fator, dada por:

$$\frac{\hat{\lambda}_j}{p}, \quad j = 1, 2, \cdots, m, \text{ se usarmos } R$$

- **Critério de Kaiser:** retenção dos fatores com autovalor associado superior a 1;
- Gráfico scree-plot
- Análise paralela de Horn

Quantos fatores usar?

- Conjugar indicadores estatísticos a:
  - Interpretação prática dos resultados;
  - Busca de uma solução parcimoniosa;
  - Possível indicação do número de fatores segundo a teoria da área;

Quantos fatores usar?

- Conjugar indicadores estatísticos a:
  - Interpretação prática dos resultados;
  - Busca de uma solução parcimoniosa;
  - Possível indicação do número de fatores segundo a teoria da área;
- **Bom senso (sempre!)**.

Quantos fatores usar?

Comentários

- **Método de máxima verossimilhança:** Mudança de valor de m altera as estimativas dos loadings
- **Método de componentes principais:** Aumento no valor de m não altera os loadings para os fatores obtidos anteriormente
- Quando os dados provêm de distribuição normal multivariada
  - Usar método de componentes principais como análise exploratória dos fatores e estimação do valor provável de m
  - Posteriormente, qualidade da solução inicial poderá ser melhorada pelo uso do método de máxima verossimilhança

Interpretação dos fatores

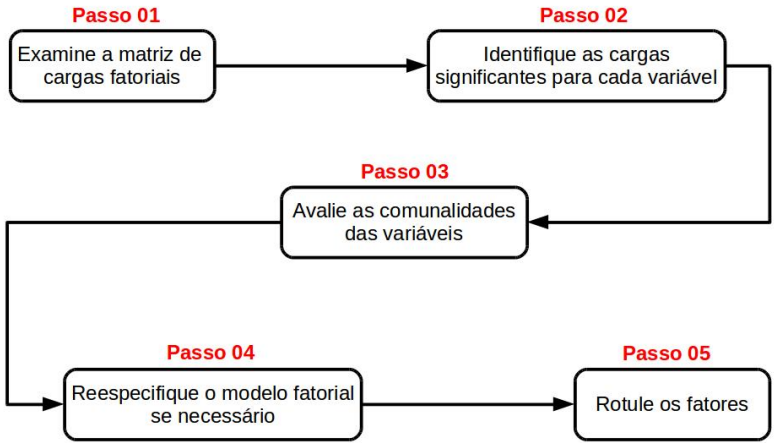
Avaliação da significância estatística baseada no tamanho da amostra

Table 5: Tamanho mínimo de amostra recomendado em função da carga fatorial

Carga Fatorial	Tamanho da amostra
0,30	350
0,35	250
0,40	200
0,45	150
0,50	120
0,55	100
0,60	85
0,65	70
0,70	60
0,75	50

Interpretação dos fatores

Interpretação de uma matriz de cargas fatoriais em 5 passos



Interpretação dos fatores

Algumas situações podem ser encontradas

A variável tem carga cruzada

Variável	Fator 01	Fator 02	Comunalidade
...	...	...	...
V_i	$-0,64$	$-0,56$	$0,72$
...	...	...	...

Interpretação dos fatores

Algumas situações podem ser encontradas

A variável não possui cargas significantes

Variável	**Fator 01**	**Fator 02**	**Comunalidade**
...	...	...	...
V_i	$-0,25$	$0,19$	$0,10$
...	...	...	...

Interpretação dos fatores

Algumas situações podem ser encontradas

Mesmo com cargas significantes, a comunalidade é baixa

Variável	**Fator 01**	**Fator 02**	**Comunalidade**
...	...	...	...
V_i	0,00	0,40	0,16
...	...	...	...

Interpretação dos fatores

Alternativas

- **Ignorar as variáveis problemáticas** e interpretar a solução fatorial como ela é, observando que as variáveis em questão são pobremente representadas na solução fatorial;

Interpretação dos fatores

Alternativas

- **Ignorar as variáveis problemáticas** e interpretar a solução fatorial como ela é, observando que as variáveis em questão são pobremente representadas na solução fatorial;

- Avaliar cada variável para uma **possível eliminação** e reespecificar o modelo após sua retirada;

Interpretação dos fatores

Alternativas

- **Ignorar as variáveis problemáticas** e interpretar a solução fatorial como ela é, observando que as variáveis em questão são pobremente representadas na solução fatorial;
- Avaliar cada variável para uma **possível eliminação** e reespecificar o modelo após sua retirada;
- Empregar um **método alternativo de rotação**;

Interpretação dos fatores

Alternativas

- **Ignorar as variáveis problemáticas** e interpretar a solução fatorial como ela é, observando que as variáveis em questão são pobremente representadas na solução fatorial;
- Avaliar cada variável para uma **possível eliminação** e reespecificar o modelo após sua retirada;
- Empregar um **método alternativo de rotação**;
- **Diminuir/aumentar o número de fatores** para avaliar se a nova estrutura fatorial representará as variáveis problemáticas;

Comentários finais sobre análise fatorial

- **Matriz de resíduos:** A observação da matriz de resíduos muitas vezes, pode indicar quando o número de fatores está superdimensionado
  - Ex.: Se m não for muito pequeno e a matriz de resíduos estiver próxima de zero, recomenda-se testar outras soluções para m menores que o valor já especificado

Comentários finais sobre análise fatorial

- **Matriz de resíduos:** A observação da matriz de resíduos muitas vezes, pode indicar quando o número de fatores está superdimensionado
  - Ex.: Se m não for muito pequeno e a matriz de resíduos estiver próxima de zero, recomenda-se testar outras soluções para m menores que o valor já especificado

- A análise fatorial deve ser utilizada apenas se utilizada em situações em que as variáveis originais são correlacionadas: Evitar soluções com m elevado tal que determinados fatores fiquem relacionados com uma única variável original

Comentários finais sobre análise fatorial

- **Matriz de resíduos:** A observação da matriz de resíduos muitas vezes, pode indicar quando o número de fatores está superdimensionado
  - Ex.: Se m não for muito pequeno e a matriz de resíduos estiver próxima de zero, recomenda-se testar outras soluções para m menores que o valor já especificado

- A análise fatorial deve ser utilizada apenas se utilizada em situações em que as variáveis originais são correlacionadas: Evitar soluções com m elevado tal que determinados fatores fiquem relacionados com uma única variável original

- Em situações em que aparecem fatores relacionados a uma única variável Z_i é recomendável retirar a variável Z_i e reestimar o modelo de análise fatorial

Usos adicionais dos resultados da AFE

Como posso usar os resultados de uma AFE em análises subsequentes?

- Seleção de variáveis substitutas para análise subseqüente
  - É selecionada a variável com maior carga no fator para atuar como **variável substituta representativa** (deve considerar o conhecimento que o pesquisador tem da teoria).

Usos adicionais dos resultados da AFE

Como posso usar os resultados de uma AFE em análises subsequentes?

- Seleção de variáveis substitutas para análise subseqüente
  - É selecionada a variável com maior carga no fator para atuar como **variável substituta representativa** (deve considerar o conhecimento que o pesquisador tem da teoria).

- **Escores Fatoriais:** representação de cada indivíduo no espaço fatorial. A estimação dos escores fatoriais necessita de métodos estatíscos elaborados, sendo os mais conhecidos os métodos da regressão e dos mínimos quadrados ponderados.

Escores fatoriais

- **Método dos Mínimos Quadrados Ponderados:**

$$\hat{f} = (\hat{L}^t \hat{\Psi}^{-1} \hat{L})^{-1} \hat{L}^t \hat{\Psi}^{-1} z$$

Escores fatoriais

- **Método dos Mínimos Quadrados Ponderados:**

$$\hat{f} = (\hat{L}^t \hat{\Psi}^{-1} \hat{L})^{-1} \hat{L}^t \hat{\Psi}^{-1} z$$

- **Método da Regressão:** suposição de normalidade multivariada para z, F e ϵ

$$\hat{f} = \hat{L}^t (\hat{L}\hat{L}^t + \hat{\Psi})^{-1} z$$

Contexto: empresa Telco

- Operadora de telefonia móvel em cidade do interior
- Pesquisa de satisfação com os clientes após um ano de operação
- Questionário com **10 itens (P1–P10)** em escala de 0 a 10

Contexto: empresa Telco

Os itens avaliam, por exemplo:

- Intensidade do sinal
- Qualidade do serviço e das informações
- Justiça nas tarifas
- Distribuição geográfica do sinal
- Plano de serviços
- Desconfiança em relação à empresa
- Tecnologia, atenção e atendimento

Itens e variáveis

Vamos representar os 10 itens por variáveis $X_1, \ldots, X_{10}$:

- X_1: intensidade do sinal
- X_2: distribuição geográfica do sinal
- X_3: tecnologia adotada
- X_4: atendimento como um todo
- X_5: injustiça nas tarifas
- X_6: plano de serviços
- X_7: qualidade das informações
- X_8: serviço prestado ao cliente
- X_9: atenção dedicada ao cliente
- X_{10}: atendimento especificamente ao cliente

Objetivo da análise

Objetivo: identificar dimensões latentes da qualidade percebida.

Objetivo da análise

Objetivo: identificar dimensões latentes da qualidade percebida.

```
# Instale o pacote pacman primeiro, se ainda não o tiver
if (!requireNamespace("pacman", quietly = TRUE)) {
  install.packages("pacman")
}

# Use p_load para instalar e carregar pacotes
pacman::p_load(tidyverse, psych, GPArotation, ggcorrplot, janitor, corrr, i
```

Lendo os dados Telco

```
path <- "https://raw.githubusercontent.com/tiagomartin/est014/refs/heads/ma
telco <- read_csv(path, show_col_types = FALSE) %>%
  clean_names()

glimpse(telco)
```

```
Rows: 875
Columns: 10
$ x1  <dbl> 2.021634, 2.634654, 2.161853, 2.611833, 2.296529, 2.265094, 2.1
$ x2  <dbl> 3.417002, 4.989637, 4.369748, 5.058855, 2.652786, 4.401388, 5.0
$ x3  <dbl> 3.203575, 3.548522, 3.423754, 3.438870, 3.394790, 3.446864, 3.4
$ x4  <dbl> 4.674381, 5.167249, 5.076795, 5.145240, 4.935702, 4.960152, 5.0
$ x5  <dbl> 7.845196, 7.527241, 7.432190, 7.381899, 7.572431, 7.401869, 7.3
$ x6  <dbl> 4.474652, 5.155799, 4.955571, 5.005566, 4.857968, 4.987190, 5.0
$ x7  <dbl> 1.652540, 2.303569, 3.000841, 3.084402, 2.834656, 4.363538, 4.4
$ x8  <dbl> 3.192175, 3.652853, 4.442697, 4.541976, 4.221692, 4.371027, 4.4
$ x9  <dbl> 7.092796, 7.518412, 6.690318, 6.761366, 6.098145, 6.766362, 6.8
```

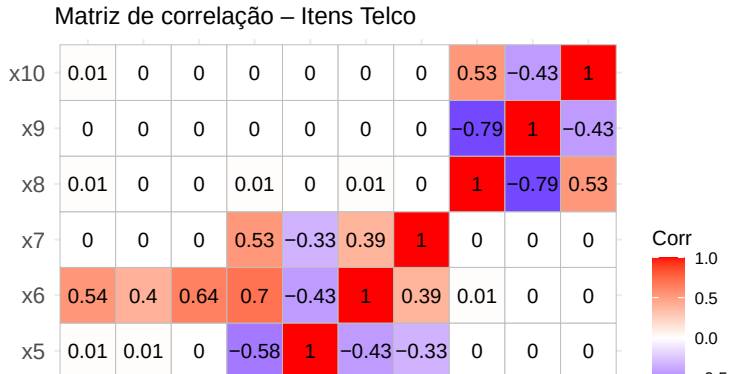
Matriz de correlação

```
R <- cor(telco, use = "pairwise.complete.obs")
R
```

```
              x1           x2           x3           x4           x5
x1   1.0000000000  0.4781493696  0.7589493789 -2.140846e-03  0.0053772224
x2   0.4781493696  1.0000000000  0.5673144741 -6.145747e-03  0.0072963197
x3   0.7589493789  0.5673144741  1.0000000000 -5.368382e-04  0.0047779539
x4  -0.0021408464 -0.0061457470 -0.0005368382  1.000000e+00 -0.5773156240
x5   0.0053772224  0.0072963197  0.0047779539 -5.773156e-01  1.0000000000
x6   0.5414366524  0.4011884586  0.6425555030  7.028403e-01 -0.4256035677
x7   0.0014174056 -0.0011068157  0.0020693251  5.251678e-01 -0.3252902914
x8   0.0075964582  0.0028010661  0.0011511906  7.496806e-03 -0.0041722500
x9   0.0001745702  0.0013298066  0.0008497782 -3.530898e-03 -0.0026125521
x10  0.0067917792  0.0001410227  0.0044244904  3.148538e-05  0.0004600118
              x6           x7           x8           x9           x10
x1   0.541436652  0.0014174056  0.007596458  0.0001745702  6.791779e-03
x2   0.401188459 -0.0011068157  0.002801066  0.0013298066  1.410227e-04
```

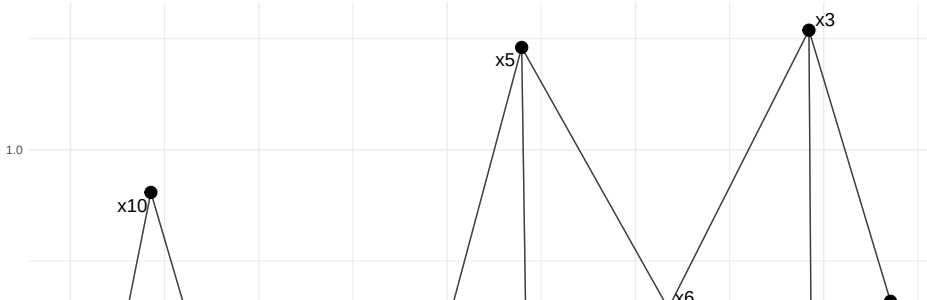
Visualização da matriz de correlação

```
ggcorrplot(
R,
type = "full",
lab = TRUE,
title = "Matriz de correlação - Itens Telco"
)
```



Visualização da matriz de correlação

```
## Variaveis aparentam se agrupar em tres grupos
tidy_cors <- telco %>% correlate() %>% stretch()
graph_cors <- tidy_cors %>%  filter(abs(r) > .3) %>% graph_from_data_frame
ggraph(graph_cors, layout = "stress") +
  geom_edge_link(alpha = 0.5) +
  geom_node_point(size = 4) +
  geom_node_text(aes(label = name), repel = TRUE, size = 5) +
  theme_minimal()
```



Medidas de adequação à AFE

KMO – Kaiser-Meyer-Olkin

```
KMO(R)
```

```
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = R)
Overall MSA =  0.63
MSA for each item =
  x1   x2   x3   x4   x5   x6   x7   x8   x9  x10
0.85 0.93 0.56 0.46 0.93 0.55 0.94 0.58 0.60 0.79
```

- KMO geral $>$ 0,60 $\rightarrow$ dados adequados para AFE
- MSA de $X_4 < 0,5 \rightarrow$ apresenta adequabilidade insuficiente, podendo ser retirada da analise

Medidas de adequação à AFE

Teste de esfericidade de Bartlett

```
cortest.bartlett(R, n = nrow(telco))
```

```
$chisq
[1] 5049.455

$p.value
[1] 0

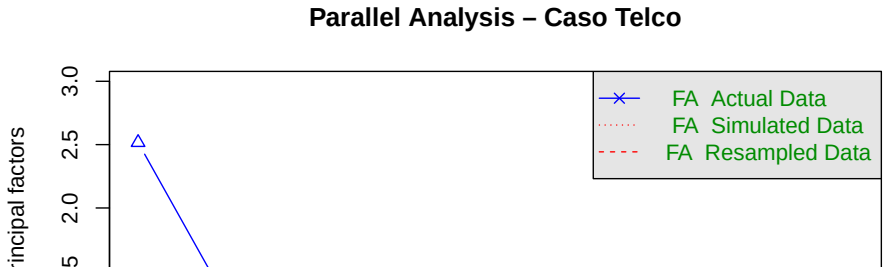
$df
[1] 45
```

- p-valor de Bartlett $< 0{,}05 \rightarrow$ correlações não são nulas

Escolha do número de fatores: Parallel Analysis

```
set.seed(123)

fa.parallel(
telco,
fm = "ml",
fa = "fa",
main = "Parallel Analysis - Caso Telco"
)
```



Escolha do número de fatores: Parallel Analysis

- Comparamos autovalores observados com autovalores obtidos ao acaso.
- Mantemos os fatores com autovalores acima da linha simulada.

```
n_fatores <- 3
```

Ajuste da AFE

```r
fa_telco <- factanal(
  x         = telco,
  factors   = n_fatores,
  rotation = "none",
  scores    = "none"
)

print(fa_telco, digits = 3, cut = 0.30) # n = 875: cargas acima de 0,30 sa
```

```
Call:
factanal(x = telco, factors = n_fatores, scores = "none", rotation = "none

Uniquenesses:
   x1    x2    x3    x4    x5    x6    x7    x8    x9   x10
0.359 0.642 0.101 0.056 0.647 0.016 0.707 0.019 0.360 0.711
```

Ajuste da AFE

Analisando as cargas significativas

- X_1 - Apresenta cargas cruzadas - fatores 01 e 03 - h^2: 0,641 - MSA: 0,85
- X_2 - Apresenta cargas cruzadas - fatores 01 e 03 - h^2: 0,358 - MSA: 0,93
- X_3 - Apresenta cargas cruzadas - fatores 01 e 03 - h^2: 0,899 - MSA: 0,56
- X_4 - Apresenta cargas cruzadas - fatores 01 e 03 - h^2: 0,944 - MSA: 0,46
- X_5 - Apresenta cargas cruzadas - fatores 01 e 03 - h^2: 0,353 - MSA: 0,93
- X_6 - fator 01 - h^2: 0,984 - MSA: 0,55
- X_7 - Apresenta cargas cruzadas - fatores 01 e 03 - h^2: 0,293 - MSA: 0,94
- X_8 - fator 02 - h^2: 0,981 - MSA: 0,58
- X_9 - fator 02 - h^2: 0,640 - MSA: 0,60
- X_{10} - fator 02 - h^2: 0,289 - MSA: 0,79

Vamos em busca de uma solucao melhor: rotacao Varimax()

```
varimax(fa_telco$loadings)
```

```
$loadings

Loadings:
    Factor1 Factor2 Factor3
x1                   0.801
x2                   0.598
x3                   0.948
x4   0.971
x5  -0.594
x6   0.729           0.673
x7   0.541
x8           0.991
x9          -0.800
x10          0.538
```

Rotação Varimax()

Analisando novamente as cargas significativas

- X_1 - fator 03
- X_2 - fator 03
- X_3 - fator 03
- X_4 - fator 01
- X_5 - fator 01
- X_6 - Apresenta cargas cruzadas - fatores 01 e 03 - h^2: 0,984 - MSA: 0,55
- X_7 - fator 01
- X_8 - fator 02
- X_9 - fator 02
- X_{10} - fator 02

Segunda tentativa: excluir da analise a variavel X_4 (MSA inadequado)

```
telco2 <- telco %>% select(-x4)

fa_telco2 <- factanal(
  x         = telco2,
  factors   = n_fatores,
  rotation = "varimax",
  scores    = "none"
)

print(fa_telco2, digits = 3, cut = 0.30) # n = 875: cargas acima de 0,30 sa
```

```
Call:
factanal(x = telco2, factors = n_fatores, scores = "none", rotation = "vari

Uniquenesses:
   x1    x2    x3    x5    x6    x7    x8    x9   x10
```

Terceira tentativa: voltar para a análise a variável X_4 e retirar a variável X_6

```r
telco3 <- telco %>% select(-x6)

fa_telco3 <- factanal(
  x         = telco3,
  factors   = n_fatores,
  rotation = "none",
  scores    = "Bartlett"
)

print(fa_telco3, digits = 3, cut = 0.30) # n = 875: cargas acima de 0,30 sa
```

```
Call:
factanal(x = telco3, factors = n_fatores, scores = "Bartlett",     rotatio

Uniquenesses:
```

Solução final excluindo X_6 da análise

- **Fator 01:** associado a X_8, X_9 e X_{10}
  - **Interpretação:** dimensão relacionada a atendimento / relacionamento com o cliente
- **Fator 02:** associado a X_1, X_2 e X_3
  - **Interpretação:** dimensão de qualidade técnica / infraestrutura (sinal, distribuição, tecnologia)
- **Fator 03:** associado a X_4, X_5 e X_7
  - **Interpretação:** dimensão de plano, informações e justiça tarifária

Solução final excluindo X_6 da análise

```
loadings_nr <- fa_telco3$loadings |>
unclass() |>
as.data.frame() |>
rownames_to_column("item")

ggplot(loadings_nr, aes(x = Factor1, y = Factor2, label = item)) +
geom_hline(yintercept = 0, linetype = "dashed", color = "grey70") +
geom_vline(xintercept = 0, linetype = "dashed", color = "grey70") +
geom_point(size = 3) +
geom_text_repel(size = 5) +
labs(
title = "Mapa das cargas não rotacionadas - Fator 1 × Fator 2",
x = "Factor 1",
y = "Factor 2"
) +
theme_minimal(base_size = 14)
```

Solução final excluindo X_6 da análise

Pontos a destacar:

- Proporção da variabilidade explicada por cada fator
  - **Fator 1:** cerca de 21%
  - **Fator 2:** cerca de 21%
  - **Fator 3:** cerca de 18%
- Proporção da variabilidade explicada pelo modelo 0.60:
- A solução de 3 fatores explica aproximadamente 60% da variância total das variáveis padronizadas.
- Para um estudo com itens de questionário (escala de 0 a 10), essa proporção é considerada muito boa.