Análise Fatorial Exploratória

Introdução

Dos livros A.D.M (D.Peña) e A.D.M. (J.Lattin, et al)

[1]EACH-USP
Universidade de São Paulo

Outline

Outline

Intuição

Exemplo:
Considere um teste psicológico com crianças (Holzinger e Swineford (1939)). Foram administrados vários testes em crianças da 7a e 8a séries (n=145). Focamos no exemplo em 5 testes:

- PARA (compreensão de parágrafo), X_1
- SENT (complementação de sentenças), X_2
- WORD (significado de palavras), X_3
- ADD (adição), X_4
- DOTS (contagem de pontos), X_5

As correlações entre os testes são apresentados abaixo.

Correlação entre testes

Tabela 5.3 Matriz de correlação para os dados do teste psicológico

	PARA	*SENT*	*WORD*	*ADD*	*DOTS*
PARA	1,000				
SENT	0,722	1,000			
WORD	0,714	0,685	1,000		
ADD	0,203	0,246	0,170	1,000	
DOTS	0,095	0,181	0,113	0,585	1,000

Tomaremos o mesmo exemplo dos testes psicológicos em crianças da 7a e 8a séries. A matriz de correlação baseada numa amostra de 145 é mostrada acima.

Fatores comuns e específicos

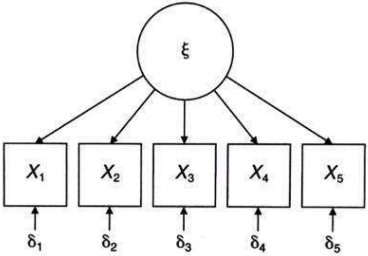

Figura 5.2 Diagrama de caminhos de um modelo de fator único com cinco variáveis.

Denotamos por ξ ao fator comum do teste e δ_i ao fator específico do teste i.
Para estimar os parâmetros do modelo de fator comum, estima-se apenas uma única variância (ou singularidade), que é a soma da variância desse fator específico e o erro de medida. Assumiremos que essa variância específica se deve inteiramente ao erro de medida, refletindo a incapacidade do teste de captar perfeitamente o fator comum subjacente.

suposições

- Os fatores δ_i são mutuamente não correlacionados
- os fatores δ_i são não correlacionados com o fator comum ξ
- suporemos também que tanto os X_i quanto ξ são variáveis padronizadas.
- Cada medida X_i possui duas fontes de variação (ambas não observáveis): o fator comúm e o fator específico δ_i

Podemos escrever os seguinte:

$$\begin{aligned} X_1 &= \lambda_1 \xi + \delta_1 \\ X_2 &= \lambda_2 \xi + \delta_2 \\ X_3 &= \lambda_3 \xi + \delta_3 \\ X_4 &= \lambda_4 \xi + \delta_4 \\ X_5 &= \lambda_5 \xi + \delta_5 \end{aligned} \tag{1}$$

a variância de X_i pode ser decomposta como:

$$Var(X_i) = var(\lambda_i \xi + \delta_i) = \lambda_i^2 + var(\delta_i) = 1$$

- λ_i é interpretável como um coeficiente de correlação
- λ_i^2 como a proporção de variação em X_i explicada pelo fator comum ξ (chamado de **Comunalidade** de X_i)
- a variância remanescente em X_i é explicada por δ_i
- Se $var(\delta_i) = \theta_{ii}^2$, representa a variância do fator específico, teremos que a comunalidade de X_i é:

$$\lambda_i^2 = 1 - \theta_{ii}^2$$

- a medida que $\lambda_i \to 1$ (ou $\theta_{ii}^2 \to 0$), X_i é quase uma medida perfeita de ξ
- se ξ não for captado pela medida de X_i, $\lambda \to 0$ e quase todas a variações em X_i podem ser explicadas pelo fator específico δ_i

Dois ou mais fatores comuns

Se o teste pode ser função de mais de uma capacidade subjacente. Por exemplo, poderiamos acreditar que o desempenho individual é determinado por duas fontes de capacidade inata: Aptidão verbal (ξ_1) e aptidão quantitativa (ξ_2).

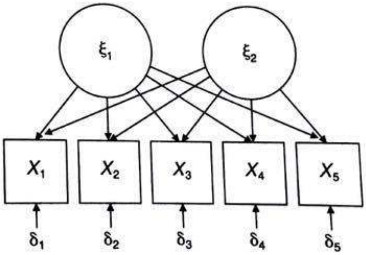

Figura 5.3 Diagrama dos caminhos de um modelo de dois fatores com cinco variáveis.

No modelo de dois fatores, três fontes contribuem para a variação em cada escore de medida do teste X_i: os dois fatores ξ_1 e ξ_2 e um fator específico δ_i:

$$\begin{aligned}
X_1 &= \lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \delta_1 \qquad (2)\\
X_2 &= \lambda_{21}\xi_1 + \lambda_{22}\xi_2 + \delta_2\\
X_3 &= \lambda_{31}\xi_1 + \lambda_{32}\xi_2 + \delta_3\\
X_4 &= \lambda_{41}\xi_1 + \lambda_{42}\xi_2 + \delta_4\\
X_5 &= \lambda_{51}\xi_1 + \lambda_{52}\xi_2 + \delta_5
\end{aligned}$$

- λ reflete a extensão com que cada fator comum contribui com a variância de cada escore de variável no teste.
- Se X e ξ forem padronizados, os parâmetros λ são interpretáveis como coeficientes de correlação.
- se assumirmos que os ξ_i são não correlacionados, a comunalidade de cada teste é dada pela soma das cargas do fator ao quadrado: a comunalidade de X_1 é $\lambda_{11}^2 + \lambda_{12}^2$
- um estudante com alto ξ_1 (verbal) e baixo ξ_2 (quantitativa), sairá melhor em testes que exigem mais capacidade verbal do que quantitativa.
- Se o desempenho na tarefa de completar uma sentença (X_1) depender somente da aptidão verbal, devemos esperar um valor de $\lambda_{11} \approx 1$ e $\lambda_{12} \approx 0$

Procedimentos para solução

Assim com em ACP, AF exploratória foca na decomposição da matriz X de covariância ou correlação. A diferença entre as duas abordagens são os fatores específicos que formam o modelo fatorial comum.
O elemento diagonal da matriz de covariância de qualquer medida X_i :

$$var(X_i) = var(\lambda_{i1}\xi_1 + \lambda_{i2}\xi_2 + \delta_i) \tag{3}$$

a partir dos presupostos de independência e de padronização temos:

$$var(X_i) = \lambda_{i1}^2 + \lambda_{i2}^2 + \theta_{ii}^2 = 1 \tag{4}$$

onde $\theta_{ii} = var(\delta_i)$ O unico lugar em que os fatores específicos aparecem na matriz de covariância de $\boldsymbol{X}$ é na diagonal, onde contribuem com uma parcela da variància de cada medida.

se conhecêssemos a variância de cada fator específico antecipadamente, poderiamos subtrair esses valores da diagonal da matriz de covariância. Restando nesse caso uma matriz na qual a única fonte de variâção e covariação são os fatores comuns.

- Na ACP decompomos a matriz de correlação $\boldsymbol{R}$ (1 na diagonal).
- para o modelo de fator comum, decompomos a matriz de correlação (1- θ_{ii} na diagonal), i.e, deixamos as comunalidades na diagonal .
- Este processo é chamado de fatoração do eixo principal ou abordagem de CP ao modelo Fator Comum.

Exemplo

suponha que $\theta_{ii} = 0,5\ \forall i$ (não é necessário que as variâncias de erro sejam iguais em todas as medidas). Substituindo os elementos diagonais de $\boldsymbol{R}$ pelos valores da comunalidade $1 - \theta_{ii} = 0,5\ \forall i$, efetuando a decomposição matricial da matriz modificada, os autovalores são:

$$\lambda_1 = 2,187 \quad \lambda_2 = 1,022 \quad \lambda_3 = -0,135 \quad \lambda_4 = -0,089 \quad \lambda_5 = 0,0,15$$

De quantos fatores necessitamos?, o critério não é o mesmo daquele usado em ACP, pois agora queremos inferir a maior variação restante possível com os fatores comuns. Nesse caso há dois ($c = 2$) autovalores sinificativamente diferentes de zero.

cargas fatoriais

i.e, a matriz de correlações entre as variáveis originais e os fatores comuns, é apresentada abaixo. O padrão da estrutura mostra que os primeiros três testes (Compreensão de parágrafo, Complementação de sentença e significado da palavra), possuem cargas mais altas no primeiro fator comum, enquanto os dois últimos testes (Adição e contagem de pontos), tem cargas mais altas no segundo fator comum.

Tabela 5.4 Matriz das cargas fatoriais para um modelo de dois fatores (com comunalidades aproximadas)

	Fator 1	Fator 2
PARA	0,7722	–0,2351
SENT	0,7838	–0,1576
WORD	0,7562	–0,2372
ADD	0,4293	0,6017
DOTS	0,3476	0,6506

Utilizando as cargas fatoriais, podemos calcular a proporção da variação em cada teste explicada pelos dois fatores comuns. Essa proporção é chamada de **Comunalidade**.
Para o teste de compreensão de parágrafo: X_1, os fatores comuns :

$$(0,77)^2 + (-0,24)^2 = 0,65$$

ou 65% da variação no teste.
O valor dessa comunalidade é mais alto que o valor de $1 - \theta_{ii}^2 = 0,5$.

Refinamos as estimativas substituindo na estimativa inicial, de comunalidade, os resultados a AF da primeira rodada e efetuamos novamente a análise. (Substituimos 0,5 na diagonal para X_1 por 0,65), e assim por diante para as variáveis restantes. Continuamos este processo até que haja convergência. Este processo iterativo é usado para estimar a comunalidade e é chamada de **método de fator principal.**

Correlação Múltipla Quadrática (Square Múltiple Correlation)

outra estimativa para a estimação inicial da comunalidade (quando não temos informações preliminares sobre o erro de medida em nossos dados) é a correlação múltipla quadrática (SMC) que é o montante de variação em uma variável explicada por todas as outras. Por exemplo, se quisermos usar a SMC para a estimativa inicial da comunalidade de X_1, podemos fazer a regressão de X_1 sobre as variáveis restantes $X_2, ...X_5$ e usar o valor de R^2.
Sabemos que a comunalidade é a proporção da variância em X explicada pelos ξ, sendo os ξ não observáveis. O que temos são as variáves X_i restantes e cada uma delas reflete os fatores ξ subjacentes (embora imperfeitamente). Como são medidos com erro, a sua capacidade de explicar a variância em X é atenuada. Assim, SMC serve como um limite inferior à verdadeira comunalidade.

Tabela 5.5 Análise fatorial usando SMCs como estimativas iniciais

Estimativa Prévia de Comunalidade: SMC

PARA	*SENT*	*WORD*	*ADD*	*DOTS*
0,6158	0,5914	0,5701	0,3672	0,3493

Autovalores Finais

	1	2	3	4	5
Autovalor	2,2826	1,0273	0,0252	−0,0010	−0,0247

Padrão de Fator

	Fator 1	Fator 2
PARA	0,8349	−0,2418
SENT	0,8253	−0,1398
WORD	0,7898	−0,2274
ADD	0,4146	0,6503
DOTS	0,3297	0,6890
Variância Explicada por Cada Fator	2,2826	1,0273

- A diferença entre as cargas de fator iniciais na tabela 4, e os da tabela 5, não são substanciais. O montante de variação explicado pelos fatores comuns é aproximadamente 66%.
- As comunalidades associadas aos dois últimos testes são mais baixas do que com os três primeiros, do que podemos concluir que os dois últimos testes são menos confiáveis do que os três primeiros.
- Se deixarmos de lado o teste de contagem de pontos (DOTS), sobrando ADD como única medidad de aptidão quantitativa, nessas circunstâncias, nossa estimativa inicial para a comunalidade de ADD baseada em SMC, será bastante baixa, e é provável concluirmos que ADD é um teste altamente não confiável, quando, de fato, é a única medida de um constructo subjacente importante.
- (Esse problema de estimativa de comunalidade, leva a alguns pesquisadores a preferir ACP em lugar de AF).

Rodando a solução fatorial

- Existe a indeterminação rotacional do modelo de fator comum (possui um número infinito de soluções)
- Seria vantajoso escolher uma orientação de modo que a matriz de cargas fosse simplificada (escolha da rotação).
- Para encontrar a matriz $\boldsymbol{T}$ de rotação (rotações ortogonais), que preservem a independência dos ξ, usamos o príncipio da estrutura simples (Thurstone 1947), que acredita que a maioria dos domínios de conteúdo envolveria diversos fatores latentes (não observados).

- Supõe-se que qualquer variável observada poderia estar correlacionada com poucos fatores subjacentes, e que qualquer fator individualmente poderia estar associado com apenas algumas variáveis.
- objetivo: encontrar agrupamentos de variáveis, com cada um definindo somente um fator.
- de modo geral: encontrar uma orientação dos eixos do fator de tal modo que cada teste tenha cargas relativamente altas em apenas alguns fatores e em outras sendo próximas de zero.

A matriz de cargas deve exibir um tipo particular de padrão:

- 1. A marioria das cargas de qualquer fator específico (coluna) deve ser pequena (proxima de zero), e somente algumas cargas devem possuir valor absoluto elevado
- 2. Uma linha específica de cargas, contendo as cargas de uma dada variável com cada fator deve exibir cargas diferentes de zero em apenas um ou não mais que em alguns poucos fatores.
- 3. Qualquer par de fatores (colunas) deve exibir diferentes padrões de cargas.

Exemplo

Imagine um estudo das percepções do consumidor de analgésicos, no qual cada sujeito é solicitado a classificar sua marca preferida de acordo com seis atributos:

Tabela 5.6 Cargas fatoriais para analgésicos

	Solução não rodada		Solução rodada	
Atributo	Fator 1	Fator 2	Fator 1	Fator 2
Não indispõe o estômago	0,579	−0,452	0,139	0,721
Sem efeitos colaterais indesejáveis	0,522	−0,572	0,017	0,774
Acaba com a dor	0,645	0,436	0,772	0,097
Funciona rapidamente	0,542	0,542	0,764	−0,051
Mantém-me acordado	−0,476	0,596	−0,034	−0,762
Alívio limitado	−0,613	−0,439	−0,750	−0,074
	Variância explicada por		**Variância explicada por**	
	1,921	1,562	1,765	1,718

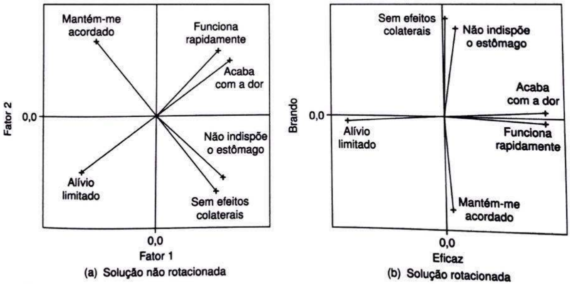

Figura 5.4 Gráfico das cargas fatoriais para atributos dos analgésicos: rodados e não rodados.

para a primeira parte da tabela acima, (solução não rodada), os 12 elementos da matriz são relativamente altos (maiores que $|0,4|$), sendo difícil obter uma interpretação clara dos fatores. Escolhendo um ângulo apropriado para a rotação, de modo que as projeções de cada atributo sobre os fatores comuns sejam ou elevados ou reduzidos. Observe na segunda parte da tabela, que os atributos 3 e 4 carregam-se positivamente e o atributo 6 negativamente. Do que nomeamos o primeiro fator como "eficácia"Ȯbserve também que os atributos 1 e 2 do segundo fator carregam-se positivamente e o atributo 5 negativamente, do que nomeamos o segundo fator como "brandura" . Estas observações podem também ser observadas na parte inferior da tabela.

Mecânica

Foi mostrado que resolver o problema de CP é equivalente a fazer uma decomposição do valor único da matriz $\boldsymbol{X}$ (padronizado):

$$\boldsymbol{X} = \boldsymbol{Z_s} \boldsymbol{D}^{1/2} \boldsymbol{U}^{‘} \tag{5}$$

onde $\boldsymbol{Z_s}$ representa os CP padronizados (mutuamente não correlacionados), $\boldsymbol{D}^{1/2}$ é uma matriz diagonal com o desvio padrão dos CP definindo a diagonal e $\boldsymbol{U}$ é a matriz de autovalores (ortogonais). Podemos reescrever a matriz de correlação $\boldsymbol{R}$ dada por:

$$\boldsymbol{R} = \frac{1}{n-1} \boldsymbol{X}^{‘} \boldsymbol{X} \tag{6}$$

temos então:

$$\begin{aligned} \boldsymbol{R} = & \frac{1}{n-1}(\boldsymbol{Z_s D^{1/2} U^{`}})^{`}(\boldsymbol{Z_s D^{1/2} U^{`}}) \\ = & \frac{1}{n-1}\boldsymbol{U D^{1/2}}(\boldsymbol{Z_s^{`} Z_s})\boldsymbol{D^{1/2} U^{`}} \\ = & (\boldsymbol{U D^{1/2}})(\boldsymbol{U D^{1/2}})^{`} \end{aligned} \quad (7)$$

porque $\frac{1}{n-1}\boldsymbol{Z_s^{`} Z_s}$ é apenas uma matriz identidade. Se lembrarmos que o produto $\boldsymbol{U D^{1/2}}$ é apenas a matriz de cargas fatoriais $\boldsymbol{F}$ (isto é a matriz de correlação entre a matriz original de dados $\boldsymbol{X}$ e a de CP $\boldsymbol{Z_s}$, podemos simplificar

$$\boldsymbol{R} = \boldsymbol{F F^{`}} \quad (8)$$

Assim com em CP um dos objetivos è a redução de dimensão, tentamos extrair um subconjunto de componentes c $(c < p)$ que se aproxime muito de $\boldsymbol{R}$. Assim, nos CP podemos determinar:

$$\boldsymbol{R} \approx \boldsymbol{F_c} \boldsymbol{F_c^{`}} \tag{9}$$

onde $\boldsymbol{F_c}$ representa somente as c primeiras colunas da matriz $\boldsymbol{F}$ de cargas fatoriais.

Em AF exploratória, tentamos aproximar $\boldsymbol{R}$: em lugar de substituir a decomposição de valor único para $\boldsymbol{X}$ na equação usamos o modelo de fator comum. O modelo geral com c fatores comuns pode ser escrito:

$$\begin{aligned} X_1 &= \lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \ldots + \lambda_{1c}\xi_c + \delta_1 \\ X_2 &= \lambda_{21}\xi_1 + \lambda_{22}\xi_2 + \ldots + \lambda_{2c}\xi_c + \delta_2 \\ &\vdots \qquad\qquad \vdots \\ X_p &= \lambda_{p1}\xi_1 + \lambda_{p2}\xi_2 + \ldots + \lambda_{pc}\xi_c + \delta_p \end{aligned} \tag{10}$$

na notação matricial:

$$\boldsymbol{X} = \boldsymbol{\Xi}\boldsymbol{\Lambda}_{\boldsymbol{c}}^{`} + \boldsymbol{\Delta} \tag{11}$$

onde $\boldsymbol{\Xi} = [\xi_1, \ldots \xi_c]$ e $\boldsymbol{\Delta} = [\delta_1, \ldots, \delta_p]$ é uma matriz de coeficientes $p \times c$

Pressupostos para o modelo de fator comum

- 1. Os fatores comuns ξ são mutuamentes não correlacionados com variância unitária:

$$\frac{1}{n-1}\boldsymbol{\Xi}`\boldsymbol{\Xi} = \boldsymbol{I}$$

- 2. Os fatores específicos δ são mutuamente não correlacionados com a matriz de covariância diagonal

$$\Theta = \frac{1}{n-1}\boldsymbol{\Delta}`\boldsymbol{\Delta} = diag(\theta_{11}^2, \theta_{22}^2, ..., \theta_{pp}^2)$$

- Os fatores comuns ξ e os fatores específicos δ são não correlacionados

$$\boldsymbol{\Xi}`\boldsymbol{\Delta} = \mathbf{0}$$

Substituimos o modelo de fator comum para obter uma aproximação à matriz $\boldsymbol{R}$

$$\begin{aligned} \boldsymbol{R} = & \frac{1}{n-1}(\boldsymbol{\Xi}\boldsymbol{\Lambda}_c^{`} + \boldsymbol{\Delta})^{`}(\boldsymbol{\Xi}\boldsymbol{\Lambda}_c^{`} + \boldsymbol{\Delta}) \\ = & \frac{1}{n-1}(\boldsymbol{\Lambda}_c\boldsymbol{\Xi}^{`}\boldsymbol{\Xi}\boldsymbol{\Lambda}_c^{`} + \boldsymbol{\Delta}^{`}\boldsymbol{\Xi}\boldsymbol{\Lambda}_c^{`} + \boldsymbol{\Lambda}_c\boldsymbol{\Xi}^{`}\boldsymbol{\Delta} + \boldsymbol{\Delta}^{`}\boldsymbol{\Delta}) \end{aligned} \tag{12}$$

Pelo presuposto 3 do modelo de fator comum, o segundo e o terceiro termos na expressão vão para zero. Pelo presuposto 1, podemos substituir a expressão $\frac{1}{n-1}\boldsymbol{\Xi}\boldsymbol{\Xi}^{`}$ do primeiro termo pela matriz $\boldsymbol{I}$. E, pelo pressuposto 2, o último termo na expressão torna-se $\boldsymbol{\Theta}$, do que chegamos a

$$\boldsymbol{R} = \boldsymbol{\Lambda}_c\boldsymbol{\Lambda}_c^{`} + \boldsymbol{\Theta} \tag{13}$$

ou

$$\boldsymbol{R} - \boldsymbol{\Theta} = \boldsymbol{\Lambda}_c\boldsymbol{\Lambda}_c^{`} \tag{14}$$

Uma comparação do modelo de CP com o modelo acima (fator comum) revela a semelhança entre as duas abordagens (e também a sua diferença). Em cada caso fazemos uma decomposição de matriz de uma matriz quadrada . A matriz $\boldsymbol{\Lambda_c}$ é análoga à matriz $\boldsymbol{F_c}$: é uma matriz de cargas fatoriais, cujos elementos são interpretáveis como as correlações entre as variáveis originais $\boldsymbol{X}$ e os fatores comuns extraídos (c).

Indeterminação Rotacional

Nos CP, escolhemos cada componente de modo sequencial para explicar o máximo possível de variação nos dados originais, com a limitação de não haver correlação entre componentes, o que garantia solução única. No Modelo de fator comum, não foi imposta tal limitação, havendo portanto, um número infinito de soluções que são idênticas no seu grau de aproximação da matriz $\boldsymbol{R} - \boldsymbol{\Theta}$. Referimo-nos a essa propriedade como indeterminação rotacional do modelo de fator comum.

Exemplo

Suponha que para os dados das tabelas acima, mudamos a orientação dos fatores rodando-os 30 graus em sentido horário. A rotação (que é executada pela multiplicação da matriz) preserva a ortogonalidade dos dois fatores. Uma matriz para realizar a rotação ortogonal (em duas dimensões) é:

$$\boldsymbol{T} = \begin{bmatrix} cos(\alpha) & -sen(\alpha) \\ sen(\alpha) & cos(\alpha) \end{bmatrix}$$

$$\mathbf{T} = \begin{bmatrix} \cos\alpha & -\operatorname{sen}\alpha \\ \operatorname{sen}\alpha & \cos\alpha \end{bmatrix}$$

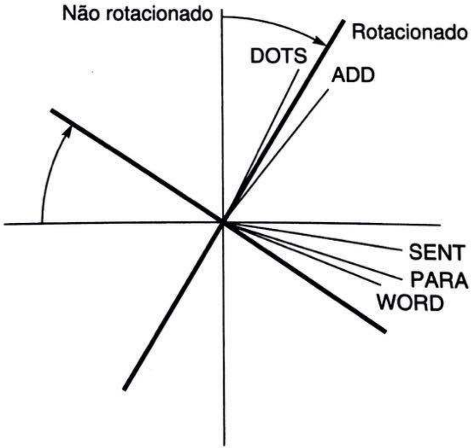

Figura 5.5 Diagrama mostrando uma rotação no sentido horário de 30 graus.

quando o ângulo de rotação $\alpha = -30$ graus:

$$\boldsymbol{T} = \begin{bmatrix} 0,866 & 0,500 \\ -0,500 & 0,866 \end{bmatrix} \tag{15}$$

Mudando a orientação dos eixos que representam os fatores comuns, também mudamos os valores das cargas fatoriais. As novas cargas:

$$\boldsymbol{\Lambda}_c^* = \frac{1}{n-1}(\boldsymbol{\Xi}\boldsymbol{\Lambda}_c^{`}\boldsymbol{\Delta})^{`}\boldsymbol{\Xi}\boldsymbol{T}$$

ou

$$\boldsymbol{\Lambda}_c^* = \boldsymbol{\Lambda}_c\boldsymbol{T}$$

porque $\frac{1}{n-1}\boldsymbol{\Xi}^{`}\boldsymbol{\Xi} = \boldsymbol{I}$ e $\boldsymbol{\Lambda}^{`}\boldsymbol{\Xi} = \boldsymbol{0}$

Tabela 5.7 Cargas fatoriais iniciais e cargas fatoriais rodadas –30 graus

	Cargas iniciais		Cargas rodadas –30 graus	
	Fator 1	Fator 2	Fator 1	Fator 2
PARA	0,811	–0,200	0,801	0,232
SENT	0,811	–0,106	0,754	0,313
WORD	0,779	–0,195	0,773	0,221
ADD	0,375	0,586	0,032	0,695
DOTS	0,295	0,614	–0,052	0,679

A tabela apresenta as cargas fatoriais rodadas Λ_c^*, ao lado das não rodadas. Observe que as principais cargas não se modificaram dramaticamente: Os três primeiros testes carregam sobre o primeiro fator e os dois últimos sobre o segundo. Mas, as cargas cruzadas mudaram na solução rodada. Os três primeiros testes agora carregam positivamente sobre o segundo fator e os dois últimos carregam quase exclusivamente sobre o segundo.

Propriedades da solução de fator rodado

- 1. Embora a variância explicada por cada um dos ξ se modifique, a variância total explicada por ambos permanece igual. Como a solução não rodada é obtida via decomposição da matriz $\boldsymbol{R} - \boldsymbol{\Theta}$, e orientada para que o primeiro fator comum explique o máximo de variação e o segundo a variação restante. A rotação muda a orientação, o que garante que a variação explicada pelo primeiro fator diminua. No entanto, a rotação somente muda a orientação dos eixos atravessando o espaço dos fatores comuns, portanto, não influencia o valor total da variação explicada

- 2. As comunalidades não são alteradas pela rotação. Observamos essa propriedade reconstruindo a matriz $\boldsymbol{R}-\boldsymbol{\Theta}$ a partir das cargas fatoriais. Temos:

$$\begin{aligned}\boldsymbol{R}-\boldsymbol{\Theta} &= \boldsymbol{\Delta}_c^{*}\boldsymbol{\Delta}_c^{*`} \\ &= \boldsymbol{\Delta}_c\boldsymbol{T}(\boldsymbol{\Delta}_c\boldsymbol{T})^{`} \\ &= \boldsymbol{\Delta}_c\boldsymbol{T}\boldsymbol{T}^{`}\boldsymbol{\Delta}_c^{`}\end{aligned} \tag{16}$$

multiplicando as matrizes, vemos que $\boldsymbol{T}\boldsymbol{T}^{`}=\boldsymbol{I}$

$$\begin{bmatrix} cos^2\alpha + sen^2\alpha = 1 & -cos\alpha sen\alpha + sen\alpha cos\alpha = 0 \\ sen\alpha cos\alpha - cos\alpha sen\alpha = 0 & cos^2\alpha + sen^2\alpha = 1 \end{bmatrix}$$

generiacamente, $\boldsymbol{T}\boldsymbol{T}^{`}=\boldsymbol{I}$ é sempre verdadeiro

Rotação Fatorial

Para encontrar a rotação apropriada, devemos determinar um meio de quantificar o que queremos dizer com estrutura simples na forma de uma função objetivo. Procuramos através de todos os ângulos possíveis de rotação e selecionamos a matriz de rotação $\boldsymbol{T}$, de tal forma que a matriz de cargas fatoriais rodada $\boldsymbol{A} = \boldsymbol{\Lambda_c T}$ exiba o maior valor da função objetivo para uma estrutura simples.

Rotação Varimax de Kaiser

Cada elemento da matriz de cargas rodada a_{ik} pode ser entendido como a correlação entre a X_i e o ξ_k. a_{ik}^2 é a proporção da variação em X_i atribuível a ξ_k. Como os ξ_i são não correlacionados, o valor total da variação explicada por todos os fatores comuns (**comunalidade**) é $h_i^2 = \sum_k a_{ik}^2$. Para obter uma estrutura simples, gostariamos de encontrar uma rotação que determine $a_{ik}^2 \to 1$ ou $a_{ik}^2 \to 0$
Focando nas colunas de $\boldsymbol{A}$ o procedimento Varimax escolhe a matriz de rotação $\boldsymbol{T}$ para maximizar a variância total da coluna a_{ik}^2. A k-ésima variância de coluna é

$$V_k = \frac{1}{p}\sum_{i=1}^{p}(a_{ik}^2)^2 - \frac{1}{p^2}\left(\sum_{i=1}^{p} a_{ik}^2\right)^2 \tag{17}$$

maximizar a soma dessas variâncias de coluna V_k para todos os k fatores equivale a maximizar:

$$V = \sum_{k=1}^{c} \sum_{i=1}^{p} a_{ik}^4 - \frac{1}{p} \sum_{k=1}^{c} \left(\sum_{i=1}^{p} a_{ik}^2 \right)^2 \quad (18)$$

Observe que V é maximizado quando $a_{ik}^2 \rightarrow 1$ ou $a_{ik}2 \rightarrow 0$. Por definição, quando $a_{ik}^2 \rightarrow 1$ para algum k, todos os outros elementos naquela linha tendem a zero.

É também possível construir a função objetivo para uma rotação varimax utilizando-se $\frac{a_{ik}^2}{h_i^2}$ (cargas ao quadrado normalizadas). Nesse caso, o elemento $\frac{a_{ik}^2}{h_i^2}$ é interpretado como a proporção da variância comum em X_i atribuível a ξ_k. Usar essa normalização garante que todas as variáveis recebam peso igual na escolha da rotação.

Rotação Quartimax

Diferente da rotação Varimax (que foca nas colunas de $\boldsymbol{A}$), a rotação quartimax foca nas linhas. A função objetivo depende de as comunalidades das variáveis não sejam alteradas pela rotação ortogonal, assim $\sum_k a_{ik}^2$ é constante e independente de $\boldsymbol{T}$. Também $\sum_i(\sum_k a_{ik}^2)^2$ é constante. Expandindo esta expressão:

$$\sum_{i=1}^{p}\sum_{k=1}^{c} a_{ik}^4 + \sum_{i=1}^{p}\left(\sum_{k=1}^{c}\sum_{j\neq k} a_{ik}^2 a_{ij}^2\right) \tag{19}$$

o segundo termo é a soma dos produtos cruzados das cargas ao quadrado. Quando uma matriz exibe uma estrutura simples, esse produto deve ser tão pequeno quanto possível (se uma variável carrega-se altamente sobre um ξ_k, a carga deve ser próxima de zero para os outros fatores). Uma vez que esta expressão é constante para todo $\boldsymbol{T}$, maximizar o primeiro termo é um modo de garantir que os termos do produto cruzado sejam pequenos.

Assim o quartimax escolhe uma matriz de rotação ortogonal $\boldsymbol{T}$ para maximizar:

$$Q = \sum_{i=1}^{p} \sum_{k=1}^{c} a_{ik}^4 \tag{20}$$

Assim como no varimax, é possível normalizar as cargas ao quadrado antes de realizar a rotação.

escores fatoriais

A AF pode não ser um fim em si mesma, senão um passo intermediário. Para realizar uma análise subsequente, precisamos da localização de cada uma das observações originais no reduzido espaço fatorial. Esses valores são chamados de **Escores fatoriais**. Na ACP os escores são combinações lineares das variáveis originais, que podem ser calculadas utilizando-se os coeficientes dos autovalores da matriz de correlação. Na análise de fator comum, os escores não podem ser calculados exatamente devido à indeterminação introduzida pelos fatores específicos. I.e, no modelo de fator comum, não podemos determinar ξ como função de X sem conhecer δ.

Para aproximar Ξ usamos

$$\Xi = X_s B \tag{21}$$

onde B é a matriz de coeficientes de escore de fator. Como os valores de Ξ não são observados, não podemos escolher B usando MQ. No entanto:

$$\frac{1}{n-1} X_s^{‘} \Xi = \frac{1}{n-1} X_s^{‘} X_s B \tag{22}$$

ou

$$\Lambda_c = RB$$

de onde premultiplicando por R^{-1} temos:

$$B = R^{-1} \Lambda_c \tag{23}$$

para os coeficientes do escore fatorial.

Substituindo essa expressão de $\boldsymbol{B}$ na equação 21, teremos:

$$\boldsymbol{\Xi} = \boldsymbol{X_s R^{-1} \Lambda_c} \tag{24}$$

como $\boldsymbol{\Lambda_c}$ está sujeita à indeterminação rotacional, os escores de fator não são únicos e são sujeitos à mesma indeterminação rotacional. O que não varia é o produto:

$$\hat{\boldsymbol{X}}_c = \boldsymbol{\Xi \Lambda_c^{`}}$$

em que $\hat{\boldsymbol{X}}_c$ é a parte apropriada de $\boldsymbol{X}$ explicada pelos c fatores comuns.

Questões relativas à aplicação da AF

- 1. Posso obter uma solução com fatores correlacionados? Em ACP forçamos a suposição de ortogonalidade, no entano, na AF esta suposição não é feita. Não limitar que a rotação seja ortogonal pode garantir melhor aproximação à estrutura simples na matriz de cargas transformadas, incrementando-se a capacidade de interpretação da solução. Nos referimos como **rotações oblíquas.** Cuja intenção é alinhar os eixos do fator o maximo possível a grupos de variáveis do conjunto original de dados.
  Numa rotação oblíqua faz-se necessária uma distinção entre **cargas de estrutura** e **cargas de padrão.**

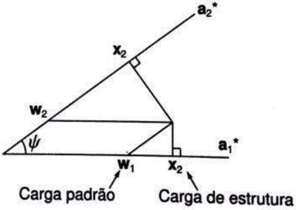

Figura 5.7 Diagrama representando a diferença entre cargas de estrutura e cargas de padrão.

Suponha que um ponto $\boldsymbol{p}$ no sistema oblíquo obtido pela rotação de $\boldsymbol{\Lambda_c}$, a solução fatorial original (ortogonal) da nova matriz $\boldsymbol{A}^* = \boldsymbol{\Lambda_c T}^*$ em que $\boldsymbol{T}^*$ representa a matriz de transformação oblíqua. O ângulo ψ reflete a correlação entre os fatores $\boldsymbol{a_1^*}$ e $\boldsymbol{a_2^*}$ no sistema oblíquo

observe que podemos registrar a posição de $\boldsymbol{p}$ como uma projeção nos eixos $\boldsymbol{a}_1^*$ e $\boldsymbol{a}_2^*$ (pontos x_1 e x_2), que são as cargas de estrutura. Podemos também registrar a posição de $\boldsymbol{p}$ no sistema de referência oblíquo (dadas por w_1 e w_2), que são as cargas padrão. No caso ortogonal, esta distinção desaparece.
Cargas de estrutura são correlações entre variáveis e fatores (entre -1 e 1), são os valores dados na matriz de cargas transformada: $\boldsymbol{A}^* = \boldsymbol{\Lambda}_c \boldsymbol{T}^*$. Em geral, não são úteis na interpretação porque a estrutura simples é disfarçada pelas correlações do fator subjacente (observe x_1 e x_2 relativamente amplas).

As cargas padrão são como coeficientes de regressão parcial, no sentido que levam em consideração a variação explicada por outros fatores. I.e, são coeficientes em uma combinação linear dos fatores oblíquos que nos permitem reconstruir a configuração original. Como resultado as cargas não são limitadas a -1 e 1. Na figura w_1 é maior que w_2 o que sugere que $\boldsymbol{p}$ está mais alinhado ao fator 1. As cargas são obtidas :

$$\boldsymbol{A}^{*}\boldsymbol{P} \tag{25}$$

onde $\boldsymbol{P}$ é uma matriz de direção dos cosenos (aprox. normalizados), que capta os ângulos entre os fatores correlacionados no sistema de referência oblíquo. No exemplo da figura:

$$[w_1 \quad w_2] = [x_1 \quad x_2] \begin{bmatrix} 1 & cos\psi \\ cos\psi & 1 \end{bmatrix} \tag{26}$$

Advertências:

- 1. como os fatores podem estar correlacionados, a comunalidade de uma variável não pode ser computada como a $\sum_k a_{ik}^2$.
- 2. A variância explicada por um fator não pode ser calculada como a $\sum_i a_{ik}^2$ na coluna de cargas fatoriais.

exemplo

No exemplo dos testes, podemos obter um retrato mais completo da estrutura simples com a aplicação de alguma rotação oblíqua? Usaremos a abordagem de ajuste de matriz alvo, que é uma representação específica da estrutura simples subjacente à solução do problema. Pode ser construida sobre base teórica ou pode representar a estrutura de uma matriz de cargas fatoriais interpretada. No nosso caso:

$$\boldsymbol{G} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

Nossa meta é encontrar uma matriz de transformação que rodará a matriz Λ_c (de cargas fatoriais) com a congruência mais proxima possível de $\boldsymbol{G}$. Se usarmos $\boldsymbol{T}^*$ como matriz de transformação, deve ser escolhida tal que $\boldsymbol{\Lambda_c T}^*$ seja o mais próxima possível de $\boldsymbol{G}$. A transformação que representa o melhor ajuste MQ a $\boldsymbol{G}$ é

$$\boldsymbol{T}^* = (\boldsymbol{\Lambda}_c^{`}\boldsymbol{\Lambda}_c)^{-1}\boldsymbol{\Lambda}_c^{`}\boldsymbol{G} \tag{27}$$

O procedimento consiste em regredir cada coluna de $\boldsymbol{G}$ sobre $\boldsymbol{\Lambda_c}$ e depois normalizar as colunas restantes de $\boldsymbol{T}^*$. Observe que as colunas de $\boldsymbol{T}^*$ não serão em geral ortogonais. Este tipo de rotação é conhecida como **rotação procrusteana** (Hurley e Cattel 1962)

Tabela 5.12 Resultados da rotação oblíqua dos dados do teste psicológico

	Cargas de padrão		Cargas de estrutura	
	Fator 1	Fator 2	Fator 1	Fator 2
PARA	0,8480	−0,0300	0,8687	0,1906
SENT	0,7905	0,0666	0,8344	0,2753
WORD	0,8016	−0,0271	0,8215	0,1815
ADD	0,0499	0,7312	0,2428	0,7696
DOTS	−0,0432	0,7488	0,1510	0,7626

Correlação interfatorial	
	Fator 2
Fator 1	0,2528

Os resultados mostram uma aproximação grande à estrutura simples de alvo: as cargas cruzadas estão entre 0,02 a 0,06 (em valor absoluto). E temos que o primeiro fator é medido pelos três primeiros testes e o segundo pelos dois últimos.
A transformação oblíqua resulta em uma correlação entre ξ_1 e ξ_2 de 0,25, o que sugere que os estudantes com alta aptidão verbal terão também uma aptidão maior do que a aptidão quantitativa média (e vice-versa).

Introdução

A Análise Fatorial (AF) tem por objetivo explicar um conjunto de variáveis observadas por meio de um pequeno conjunto de variáveis latentes ou não observadas, que são chamados de Fatores.
Suponha que tomamos 20 medidas físicas do corpo de um indivíduo: altura, tamanho das extremidades, largura dos ombros, peso, etc.
Supõe-se que todas as medidas não são independentes entre si. Uma explicação é que as dimensões do corpo humano dependem de fatores que caso sejam conhecidos, poderiamos prever as dimensões com erro pequeno.
A AF está relacionado com ACP, mas existem algumas diferenças: 1) as CP se constroem para explicar as variâncias, no entanto, os Fatores se constroem para explicar as covariâncias ou correlações entre variáveis. 2) a CP é uma ferramenta descritiva, enquanto a AF presupõe um modelo estatístico formal de geração dos dados.

Outline

Hipóteses

Supomos que observamos um vetor de variáveis **x**, de dimensões $(p \times 1)$ em elementos de uma população. O modelo de AF estabelece que este vetor de dados é gerado mediante a relação:

$$\mathbf{x} = \mu + \Delta\,\mathbf{f} + \mathbf{u} \tag{28}$$

Hipóteses

onde:

1. **f** $(m \times 1)$ é um vetor de variáveis latentes ou fatores não observados. Suporemos que **f**$\sim N_m(\mathbf{0}, \mathbf{I})$. I.e, os fatores são v. com média zero e independentes entre si e com distribuição normal.
2. Δ $(p \times m)$ é uma matriz de constantes desconhecidas $(m < p)$. Contém os coeficientes que descrevem como os fatores **f** afetam as variáveis observadas **x**. É denotada também de matriz de cargas.
3. **u** $(p \times 1)$ é um vetor de perturbações não observadas. Acumula o efeito de todas as variáveis que não são os fatores que exercem influência sobre **x**. Suporemos que **u**$\sim N_p(\mathbf{0}, \psi)$, onde ψ é diagonal e que as perturbações são não correlacionadas com os fatores **f**.

Hipóteses

deduzimos então que:

- μ é a média de **x**
- **x**$\sim N_p(\mu, \mathbf{V})$ (**V** é a matriz de covariâncias).

Isto implica que dada uma a.a.s. de n elementos gerada pelo modelo fatorial, cada x_{ij} pode ser escrito como :

$$x_{ij} = \mu_j + \lambda_{j1} f_{1i} + \ldots + \lambda_{jm} fmi + u_{ij}, i = 1, ..., n, j = 1, ..., p$$

Exemplo de padronização das componentes

Fig2_ACP.png