

Análise Fatorial Confirmatória

Introdução

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

¹EACH-USP
Universidade de São Paulo

- 1 Introdução a Análise Fatorial Confirmatória
 - Ideias Gerais
- 2 Mecânica
- 3 Questões relativas à aplicação de AFConfirmatória

Outline

- 1 Introdução a Análise Fatorial Confirmatória
 - Ideias Gerais
- 2 Mecânica
- 3 Questões relativas à aplicação de AFConfirmatória

Intuição

- Começamos com uma forte noção prévia da estrutura do modelo fatorial
- Em vez de exploração, a meta é confirmação
- começamos delineando uma abordagem para estimar os parâmetros baseada na verossimilhança máxima (ou MQ parciais), em vez de decompor a matriz.
- será possível determinar se uma versão limitada de um modelo mais geral fornece um ajuste aceitável dos dados.

Intuição

Usamos o mesmo exemplo da AFExploratória para apresentar a técnica e as diferenças entre AFexploratória e AFConfirmatória: É preferível utilizar a matriz de covariância S em vez da matriz de correlação R por dois motivos

- 1. A análise pode depender da escala de medidas
- 2. As matrizes **S** e **R** possuem propriedades diferentes, pode-se calcular incorretamente o Erro padrão (quando se analisa uma matriz **R** em lugar de **S**)

É preferível estimar os parâmetros da AFC utilizando a matriz S (não padronizada) e, depois verificar a solução padronizada.

Matriz de correlação

$$\mathbf{R} = \begin{bmatrix} 1,000 & 0,722 & 0,714 & 0,203 & 0,095 \\ 0,722 & 1,000 & 0,685 & 0,246 & 0,181 \\ 0,714 & 0,685 & 1,000 & 0,170 & 0,113 \\ 0,203 & 0,246 & 0,170 & 1,000 & 0,585 \\ 0,095 & 0,181 & 0,113 & 0,585 & 1,000 \end{bmatrix}$$

Na seção anterior especificamos um modelo como:

$$\begin{aligned} X_1 &= \lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \delta_1 \\ &\vdots \\ X_5 &= \lambda_{51}\xi_1 + \lambda_{52}\xi_2 + \delta_2 \end{aligned} \tag{1}$$

lembre que há mais de um conjunto de valores para os parâmetros em Λ que geram a mesma reprodução da matriz de correlação original.

Na abordagem Exploratória identificamos uma solução baseada em CP, escolhendo fatores mutuamente não correlacionados para explicar tanta variação quanto possível, depois selecionamos dentre as diferentes soluções possíveis, utilizando rotação para identificar uma única solução com estrutura de carga fatorial simples e interpretável. É possível que dois fatores possam estar correlacionados, i.e, no exemplo, um estudante com alta capacidade verbal possui também capacidade quantitativa acima da média. Especificamos portanto o modelo:

$$X_1 = \lambda_{11}\xi_1 + \delta_1 \quad (2)$$

$$X_1 = \lambda_{21}\xi_1 + \delta_2$$

$$X_3 = \lambda_{31}\xi_1 + \delta_3$$

$$X_4 = \lambda_{42}\xi_2 + \delta_4$$

$$X_5 = \lambda_{52}\xi_2 + \delta_5$$

$$Corr(\xi_1, \xi_2) = \phi_{12}$$

há duas diferenças entre AF exploratório e AF confirmatório:

- 1. Impomos limitações sobre a matriz de cargas fatoriais:

$$\Lambda = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \end{bmatrix}$$

o que reduz o número de parâmetros.

- 2. Permitimos correlação entre os fatores: A matriz de correlação fatorial é dada por

$$\begin{bmatrix} 1 & \phi_{12} \\ \phi_{21} & 1 \end{bmatrix}$$

Como a matriz é simétrica ($\phi_{12} = \phi_{21}$), permite que tenhamos um custo de um parâmetro extra para o modelo de dois fatores.

Não podemos estimar simultaneamente $var(\xi_i)$ e os λ_{ij} associados com ξ_i , e também observamos que: 1. os coeficientes λ aumentam, a $var(\xi_i)$ diminui e $var(X)$ permanece constante. Há duas convenções para tratar este problema:

- 1. padronizamos a solução, fixando $var(\xi_i) = 1$, o que torna S em R
- 2. Calculamos $var(\xi_i)$ como um parâmetro do modelo, fixando λ arbitrariamente (normalmente em 1), para uma das variáveis que carregam sobre cada fator. No exemplo de teste psicológico:

$$\Lambda = \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & 1 \\ 0 & \lambda_{52} \end{bmatrix} \quad \Phi = \begin{bmatrix} \phi_{11}^2 & \phi_{12} \\ \phi_{21} & \phi_{22}^2 \end{bmatrix}$$

a variância dos fatores subjacentes ($\phi_{11}^2 \phi_{22}^2$) não serão iguais a 1, mas é uma transformação direta para se obter uma solução padronizada

Qualquer abordagem produz o mesmo ajuste de modelo, no entanto, Gonzales e Griffin(2001) apontaram que eles podem produzir erros padrão diferentes (resultando em diferentes razões t para um parâmetro). Eles sugerem realizar um "análise de sensibilidade". O efeito dos pressupostos é que há informações suficientes em \mathbf{R} para que valores únicos de cada parâmetro sejam determinados.

No exemplo do teste psicológico:

Usamos 15 peças de informação para determinar os 11 parâmetros: 5 cargas: $(\lambda_{11}, \dots, \lambda_{52})$, 5 variâncias de fatores específicos: $(\theta_{11}^2, \dots, \theta_{55}^2)$ e a correlação entre dois fatores ϕ_{12} .

Os θ_{ii}^2 podem ser escolhidas para se ajustarem à diagonal de S , e fixando $\phi_{11} = \phi_{22} = 1$ teremos por exemplo: $\theta_{11} = 1 - \lambda_{11}^2$. Os parâmetros restantes são escolhidos de forma que as covariâncias ajustadas correspondam na medida do possível às atuais covariâncias.

para o exemplo, os 10 pares de covariâncias são:

$$\begin{aligned}
 \text{corr}(X_1, X_2) &= \text{corr}(\lambda_{11}\xi_1 + \delta_1, \lambda_{21}\xi_1 + \delta_2) = \lambda_{11}\lambda_{21} = 0,722 & (3) \\
 \text{corr}(X_1, X_3) &= \text{corr}(\lambda_{11}\xi_1 + \delta_1, \lambda_{31}\xi_1 + \delta_3) = \lambda_{11}\lambda_{31} = 0,714 \\
 \text{corr}(X_1, X_4) &= \text{corr}(\lambda_{11}\xi_1 + \delta_1, \lambda_{42}\xi_2 + \delta_4) = \lambda_{11}\phi_{12}\lambda_{42} = 0,203 \\
 \text{corr}(X_1, X_5) &= \text{corr}(\lambda_{11}\xi_1 + \delta_1, \lambda_{52}\xi_2 + \delta_5) = \lambda_{11}\phi_{12}\lambda_{52} = 0,095 \\
 \text{corr}(X_2, X_3) &= \text{corr}(\lambda_{21}\xi_1 + \delta_2, \lambda_{31}\xi_1 + \delta_3) = \lambda_{21}\lambda_{31} = 0,685 \\
 \text{corr}(X_2, X_4) &= \text{corr}(\lambda_{21}\xi_1 + \delta_2, \lambda_{42}\xi_2 + \delta_4) = \lambda_{21}\phi_{12}\lambda_{42} = 0,246 \\
 \text{corr}(X_2, X_5) &= \text{corr}(\lambda_{21}\xi_1 + \delta_2, \lambda_{52}\xi_2 + \delta_5) = \lambda_{21}\phi_{12}\lambda_{52} = 0,181 \\
 \text{corr}(X_3, X_4) &= \text{corr}(\lambda_{31}\xi_1 + \delta_3, \lambda_{42}\xi_2 + \delta_4) = \lambda_{31}\phi_{12}\lambda_{42} = 0,170 \\
 \text{corr}(X_3, X_5) &= \text{corr}(\lambda_{31}\xi_1 + \delta_3, \lambda_{52}\xi_2 + \delta_5) = \lambda_{31}\phi_{12}\lambda_{52} = 0,113 \\
 \text{corr}(X_4, X_5) &= \text{corr}(\lambda_{42}\xi_2 + \delta_4, \lambda_{52}\xi_2 + \delta_5) = \lambda_{42}\lambda_{52} = 0,585
 \end{aligned}$$

Este modelo é superdeterminado, há mais observações que parâmetros desconhecidos. O número de g.l. é a diferença entre o número de observações independentes e o número de parâmetros a serem estimados, neste caso há 4 g.l. associado ao modelo de dois

Propriedades desejáveis da solução de Máxima Verossimilhança

- 1. Somos capazes de obter erros padrão assintóticos das estimativas do parâmetro, o que nos permite avaliar a estabilidade das estimativas do parâmetro e até conduzir testes estatísticos dos valores.
- 2. Vantagem de um **teste de qualidade de ajuste do modelo**. Podemos comparar o ajuste de um modelo restrito com um modelo mais geral (ou completo), atentando para a razão de verossimilhança entre os modelos restrito e geral:

$$-2[\ln(L_R) - \ln(L_F)] \sim \chi^2_{g.l(R)-g.l(F)} \quad (4)$$

onde L_R indica o modelo restrito e L_F o modelo completo.

Comparação das abordagens nos testes de modelos

- em Análise de Regressão utilizamos R^2 para descrever o ajuste de um modelo, e é definido como a proporção de variância explicada relativa a um modelo restrito (com um intercepto).
- Utilizamos a estatística F para testar se o modelo proposto oferece ou não uma melhoria significativa no ajuste em relação ao modelo restrito.
- H_0 indicará que o modelo proposto não oferece melhoras no ajuste, portanto, buscamos rejeitá-la. Grandes valores de F e pequenos valores de p sustentam o modelo proposto (i.e, rejeitar H_0).

- Na AF Confirmatória, a abordagem é diferente. Testamos H_0 que o modelo proposto se ajusta tão bem quanto o modelo geral (ou perfeito). Se a diferença no ajuste for pequena, não rejeitamos H_0 . Como nosso objetivo é não rejeitar H_0 , utilizamos um p um pouco maior: aproximadamente 0,2.
- O teste χ^2 é sensível ao tamanho da amostra. I.e, quando n é grande, mesmo as menores discrepâncias são julgadas significativas, levando à rejeição de H_0
- são utilizadas duas medidas de ajuste alternativas: GFI (goodness-of-fit index) ou índice de qualidade do ajuste e o AGFI (adjusted goodness-of-fit index) ou índice de qualidade de ajuste ajustado.

GFI e AGFI

vantagens

- entre 0 e 1
- Não dependem do n

desvantagens:

- Não há um padrão objetivo que estipule o que representa um nível aceitável de GFI ou AGFI
- Regra de ouro para GFI: $GFI \geq 0,95$ para um bom ajuste e $GFI \geq 0,9$ para um ajuste aceitável
- Regra de ouro para AGFI: $AGFI \geq 0,9$ para um bom ajuste e $AGFI \geq 0,8$ para um ajuste aceitável

Tabela 6.3 Estatística da qualidade de ajuste para o modelo de dois fatores de desempenho de estudantes no teste

Índice de qualidade de ajuste (GFI)	0,9919	
GFI ajustado para graus de liberdade (AGFI)	0,9697	
Raiz quadrada média residual (RMR)	0,0218	
$\chi^2 = 2,9306$	$df = 4$	$p = 0,5695$
qui-quadrado do modelo nulo	$df = 10$	298,6480

Confiabilidade da medida

A confiabilidade nos diz se um particular X realiza um bom trabalho de medição do verdadeiro ξ que ela se propõe medir. Lord e Novick(1968) propõem:

$$\rho_{X\xi}^2 = \frac{\sigma_{\xi}^2}{\sigma_X^2} = 1 - \frac{\sigma_{\delta}^2}{\sigma_X^2} \quad (5)$$

Quanto $\rho_{X\xi}^2 \rightarrow 1$ menor o erro na variância na medida de X e mais próxima a correspondência entre a medida X e o valor verdadeiro de ξ .

Mecânica

A forma geral do modelo fatorial confirmatório é:

$$X = \Xi\Lambda' + \Delta \quad (6)$$

onde Λ é a matriz de cargas fatoriais $p \times c$, limitado para identificar os parâmetros do modelo, Ξ é a matriz $n \times c$ de escores fatoriais padronizados. A matriz de correlação fatorial Φ e a matriz de covariância Θ podem não ser diagonais.

A matriz de covariância de X (Σ) é modelada como

$$\text{var}(X) = \text{var}(\Xi\Lambda' + \Delta) = \Lambda\Phi\Lambda' + \Theta \quad (7)$$

Nossa meta é escolher as estimativas do parâmetro para que os valores ajustados de Σ sejam o mais próximos de S

Abordagem baseada na maximização da função de verossimilhança

Assumimos que X é composta de n extrações independentes de uma $N(\mathbf{0}, \Sigma)$. A probabilidade de extração de um $\mathbf{x}'_i = (x_{i1}, \dots, x_{ip})$ particular é

$$(2\pi|\Sigma|)^{-1/2} \exp\left(-\frac{1}{2}\mathbf{x}'_i \Sigma^{-1} \mathbf{x}_i\right) \quad (8)$$

para n observações temos a função de verossimilhança:

$$L = \prod_{i=1}^n (2\pi|\Sigma|)^{-1/2} \exp\left(-\frac{1}{2}\mathbf{x}'_i \Sigma^{-1} \mathbf{x}_i\right) \quad (9)$$

se tomarmos o log de L temos:

$$\begin{aligned} \ln(L) &= \sum_{i=1}^n \left[-\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(|\Sigma|) - \frac{1}{2} \mathbf{x}_i' \Sigma^{-1} \mathbf{x}_i \right] \\ &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(|\Sigma|) - \frac{1}{2} \sum_{i=1}^n \mathbf{x}_i' \Sigma^{-1} \mathbf{x}_i \end{aligned} \quad (10)$$

o último termo é igual ao traço da matriz $\mathbf{X}\Sigma^{-1}\mathbf{X}'$, assim

$$\ln(L) = -\frac{n}{2} \left[\ln(2\pi) + \ln(|\Sigma|) + \text{tr} \left(\frac{1}{n} \mathbf{X}\Sigma^{-1}\mathbf{X}' \right) \right] \quad (11)$$

como o traço não é afetado por permutas no produto de matrizes, temos que $\text{tr}(1/n \mathbf{X}\Sigma^{-1}\mathbf{X}') = \text{tr}(\mathbf{S}\Sigma^{-1})$, e também o resultado da maximização não é afetado pelo termo $\ln(2\pi)$, temos:

$$\ln(L) = -\frac{n}{2} \left[\ln(|\Sigma|) + \text{tr}(\mathbf{S}\Sigma^{-1}) \right] \quad (12)$$

A expressão acima é intratável de forma analítica (não linear) e pode ser resolvido utilizando métodos numéricos.

Testando modelos aninhados

Para um modelo restrito (superdeterminado, com parâmetros menores que observações independentes), esperamos diferenças entre Σ e S . Para um modelo geral (um parâmetro por observação independente), esperamos um ajuste perfeito entre Σ e S . Nesse caso:

$$\ln(L_F) = -\frac{n}{2} [\ln(|S|) + \text{tr}(SS^{-1})] \quad (13)$$

ou

$$\ln(L_F) = -\frac{n}{2} [\ln(|S|) + p] \quad (14)$$

onde p é o número de variáveis no modelo.

A expressão para o teste χ^2 de qualidade de ajuste é:

$$n [\ln(|\Sigma|) + \text{tr}(S\Sigma^{-1}) - \ln(|S|) - p] \sim \chi^2 \quad (15)$$

como não há g.l. com o modelo perfeitamente ajustado, o número de g.l. associado com a estatística χ^2 é igual ao número de g.l. para o modelo restrito.

Para o teste χ^2 acima, observamos que ela é diretamente proporcional a n , e quando n é grande, mesmo menores discrepâncias entre Σ e S são julgadas significativas (rejeitar H_0).

Dois índices de qualidade de ajuste são apresentados:

$$GFI = 1 - \frac{tr[(\Sigma^{-1}S - I)^2]}{tr[(\Sigma^{-1}S)^2]} \quad (16)$$

Se $\Sigma = S$ teremos $GFI = 1$ o que faz do GFI um análogo de R^2

$$GFI = 1 - \frac{p(p+1)}{2df} \frac{tr[(\Sigma^{-1}S - I)^2]}{tr[(\Sigma^{-1}S)^2]} \quad (17)$$

onde df é o número de g.l. associado com o modelo restrito (i.e, a diferença entre o número de observações independentes e o número estimado dos parâmetros).

A diferença entre GFI e AGFI é o termo $p(p + 1)/2df$ que é a razão entre o número de g.l. antes do cálculo dos parâmetros do modelo (n. de observações independentes) e o número de g.l. após o cálculo dos parâmetros do modelo. Quanto menos parcimonioso é o modelo (maior o número de parâmetros a serem estimados), maior o termo $p(p + 1)/2df$ e menor o valor de AGFI em relação a GFI. O AGFI é o análogo ao R^2 ajustado.

Avaliando a confiabilidade do modelo

Uma maneira de examinar a consistência interna de um índice é através do uso do **alpha de Cronbach**:

$$\alpha = \frac{k\bar{r}}{[1 + (k - 1)\bar{r}]} \quad (18)$$

onde k é o número de variáveis usado no índice e \bar{r} é a correlação média entre itens para os k itens. $0 \leq \alpha \leq 1$.

Outro índice de confiabilidade proposto (por Werts, Linn e Joreskog (1974)) é

$$\rho_C^2 = \frac{(\sum_{i=1}^k \lambda_i)^2}{(\sum_{i=1}^k \lambda_i)^2 + (\sum_{i=1}^k \theta_{ii})^2} \quad (19)$$

onde ρ_C^2 é a confiabilidade do índice composto e k é o número de itens que medem o construto subjacente ξ .

Testando parâmetros de modelo

Podemos usar o teste da razão de verossimilhança para testar uma diferença significativa no ajuste entre dois modelos: um com valor do parâmetro fixado em 0 e outro em que o valor é livre (estimado a partir dos dados). Se calcularmos o χ^2 para cada modelo, o teste da razão de verossimilhança será dado pela diferença entre esses dois valores. (com um g.l)

Para testar $H_0 : \phi_{12} = 0$, tomamos as diferenças entre a qualidade do ajuste χ^2 para o modelo restrito (10,78, com 5 g.l) e para o modelo não restrito (0,57 com 4 g.l). O resultado é um $\chi^2 = 10,21$ com 1 g.l. que é significativo em 0,005.

Observe que, em razão de o teste estatístico ser baseado na qualidade de ajuste, ele não depende da escolha da identificação do modelo.

Testando mais de um fator

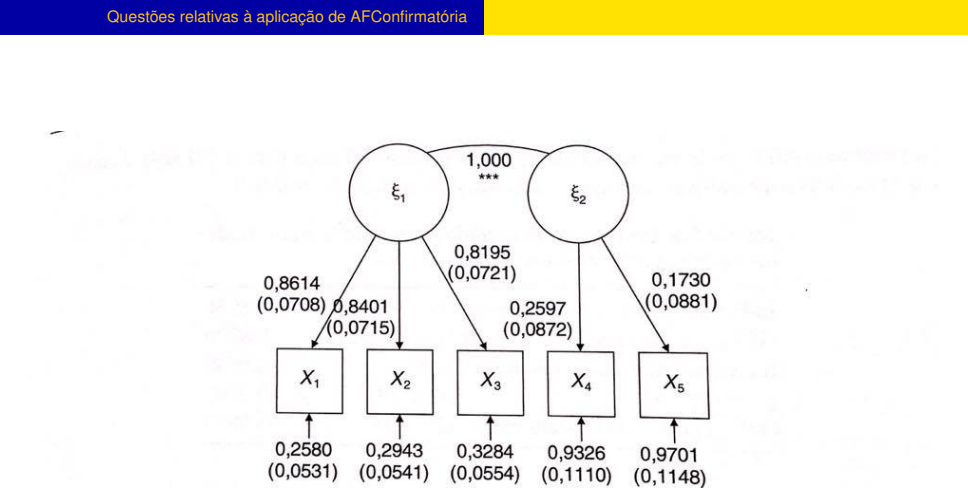
para os dados do teste psicológico, queremos saber se o modelo de fator único também fornece um bom ajuste.

Para criar o modelo de fator único, que possa se testar contra o modelo de dois fatores, basta estabelecer $\phi_{12} = 1$. Testamos então o modelo restrito (um fator com 5 g.l.) contra o modelo não restrito (2 fatores, 4 g.l.).

Na tabela abaixo é possível ver uma queda abrupta na qualidade do ajuste associada com o modelo de fator único. A estatística $\chi^2(4) = 2,93$ ($p > 0,5$) para o modelo com dois fatores, vai para $\chi^2(5) = 59,47$ com ($p > 0,001$). Tanto o GFI (0,88) quanto o AGFI(0,63) caem abaixo de níveis aceitáveis.

Tabela 6.5 Estatística de qualidade de ajuste para o modelo de fator único do desempenho de estudantes no teste

Índice da qualidade de ajuste (GFI)	0,8764
GFI ajustado para graus de liberdade (AGFI)	0,6291
Raiz quadrada média residual (RMR)	0,1414
$\chi^2 = 59,4732$	$df = 5$ $p = 0,0001$
Qui-quadrado do modelo nulo	$df = 10$ 298,6480



O diagrama ilustra um modelo de fator único com dois fatores latentes, ξ_1 e ξ_2 , representados por círculos no topo. Uma curva arqueada entre eles indica uma correlação de 1,000, marcada com três asteriscos (***). Abaixo, cinco variáveis observadas, X_1 a X_5 , são representadas por retângulos. As setas indicam as cargas fatoriais (estimativas de parâmetro) e os valores entre parênteses representam os erros padrão.

- ξ_1 influencia X_1 (0,8614 / 0,0708), X_2 (0,8401 / 0,0715) e X_3 (0,8195 / 0,0721).
- ξ_2 influencia X_3 (0,2597 / 0,0872) e X_5 (0,1730 / 0,0881).
- Além disso, há setas diretas para X_4 (0,9326 / 0,1110) e X_5 (0,9701 / 0,1148) que não são influenciadas diretamente pelos fatores latentes.