

Análise de Fatores

Jhessica Letícia Kirch
Universidade de São Paulo

Simpósio de Microbiologia Agrícola
11 de abril de 2023



O MODELO DE ANÁLISE DE FATORES - AF

Objetivo: Descrever um conjunto de p variáveis X_1, X_2, \dots, X_p em termos de poucos índices ou fatores (O mesmo da ACP).

Diferença da ACP: a AF é baseada em um modelo estatístico!

RELAÇÃO COM A ACP

As componentes principais são definidas como combinações lineares das variáveis originais.

Na análise fatorial, as variáveis originais são expressas como combinações lineares dos fatores.

A capacidade de **rotação**, que facilita a interpretação, é uma das vantagens da análise fatorial sobre a análise de componentes principais.

O MODELO DE ANÁLISE DE FATORES - AF

Modelo Geral:

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \cdots + a_{im}F_m + e_i$$

$a_{i1}, a_{i2} + \cdots + a_{im}$: são as cargas fatoriais;

F_1, F_2, \dots, F_m : são os m fatores comuns não correlacionados, F_i (média 0 e variância 1);

F_j e e_i são não correlacionados.

ANÁLISE DE FATORES POR COMPONENTES PRINCIPAIS

- Existe inúmeras soluções.
- Partindo-se da matriz X ($n \times p$), realiza-se uma ACP, obtendo-se p CP:

$$Z_i = b_{i1}X_1 + b_{i2}X_2 + \cdots + b_{ip}X_p \quad (i = 1, 2, \dots, p) \quad (5.1)$$

b_{ij} : são dados pelos autovetores da matriz de correlações.
Por se uma transformação ortogonal, o relacionamento inverso é:

$$X_i = b_{1i}Z_1 + b_{2i}Z_2 + \cdots + b_{pi}Z_p. \quad (i = 1, 2, \dots, p)$$

ANÁLISE DE FATORES POR COMPONENTES PRINCIPAIS

O modelo de $m < p$ fatores fica:

$$X_i = b_{1i}Z_1 + b_{2i}Z_2 + \cdots + b_{mi}Z_m + e_i \quad (i = 1, 2, \dots, p)$$

e_i : é uma combinação linear dos componentes de Z_{m+1} a Z_p .

Obs: o número de fatores pode ser determinado pelos critérios da análise de componentes principais (autovalores maiores que 1 ou percentual de explicação maior que 80%, por exemplo.)

ANÁLISE DE FATORES POR COMPONENTES PRINCIPAIS

Divide-se Z_i por seu desvio padrão $\sqrt{\lambda_i}$ (obtendo-se F_i), a raiz quadrada do correspondente autovalor na matriz de correlações.

As equações ficam:

$$X_i = \sqrt{\lambda_1}b_{1i}F_1 + \sqrt{\lambda_2}b_{2i}F_2 + \cdots + \sqrt{\lambda_m}b_{mi}F_m + e_i \quad (i = 1, 2, \dots, p)$$

Chamando $a_{ij} = \sqrt{\lambda_j}b_{ji}$ tem-se:

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \cdots + a_{im}F_m + e_i \quad (i = 1, 2, \dots, p) \quad (5.2)$$

ANÁLISE DE FATORES POR COMPONENTES PRINCIPAIS

Após qualquer rotação tem-se:

$$X_i = g_{i1}F_1^* + g_{i2}F_2^* + \cdots + g_{im}F_m^* + e_i \quad (5.3)$$

em que F_i^* representa o novo i-ésimo fator, $i = 1, 2, \dots, p$.

- A **rotação varimax** permite gerar novos fatores não correlacionados.

EXEMPLO

Emprego em países Europeus

Tabela 1: Porcentagem de força de trabalho de empregados em nove diferentes grupos de indústria em 30 países da Europa.

País	Grupo	AGR	MIN	FAB	FEA	CON	SER	FIN	SSP	TC
Bélgica	EU	2,6	0,2	20,8	0,8	6,3	16,9	8,7	36,9	6,8
Dinamarca	EU	5,6	0,1	20,4	0,7	6,4	14,5	9,1	36,3	7,0
França	EU	5,1	0,3	20,2	0,9	7,1	16,7	10,2	33,1	6,4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Albânia	Leste	55,5	19,4	0,0	0,0	3,4	3,3	15,3	0,0	3,0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Turquia	Outros	44,8	0,9	15,3	0,2	5,2	12,4	2,4	14,5	4,4

EXEMPLO

Emprego em países Europeus - legenda

AGR = agricultura, florestal e pesca;

MIN = mineração e exploração de pedreiras;

FAB = fabricação;

FEA = fornecimento de energia e água;

CON = construção;

SER = serviços;

FIN = finanças;

SSP = serviços sociais e pessoais;

TC = transportes e comunicações.

EXEMPLO

Emprego em países europeus

Uma ACP foi realizada (4 autovalores maiores que 1).

A matriz **R** é dada na Tabela 2 e os autovalores e autovetores de **R** são dados na Tabela 3.

EXEMPLO

Tabela 2. Matriz R do Exemplo dos países.

	AGR	MIN	FAB	FEA	CON	SER	FIN	SSP	TC
AGR	1								
MIN	0,316	1							
FAB	-0,254	-0,672	1						
FEA	-0,382	-0,387	0,388	1					
CON	-0,349	-0,129	-0,034	0,165	1				
SER	-0,605	-0,407	-0,033	0,155	0,473	1			
FIN	-0,176	-0,248	-0,274	0,094	-0,018	0,379	1		
SSP	-0,811	-0,316	0,050	0,238	0,072	0,388	0,166	1	
TC	-0,487	0,045	0,243	0,105	-0,055	-0,085	-0,391	0,475	1

EXEMPLO

Tabela 3. Autovalores e autovetores para dados de emprego europeus da Tabela 1.

Autovalores	Autovetores								
	AGR	MIN	FAB	FEA	CON	SER	FIN	SSP	TC
3,11	0,51	0,37	-0,25	-0,32	-0,22	-0,38	-0,13	-0,43	-0,21
1,81	0,02	0,00	-0,43	-0,11	0,24	0,41	0,55	-0,05	-0,52
1,50	0,28	-0,52	0,50	0,29	-0,07	-0,07	0,10	-0,36	-0,41
1,06	-0,02	-0,11	-0,06	-0,02	-0,78	-0,17	0,49	0,32	0,04
0,71	0,02	-0,35	0,23	-0,85	-0,06	0,27	-0,13	0,05	0,02
0,31	-0,04	0,20	-0,03	0,21	-0,50	0,67	-0,41	-0,16	-0,14
0,29	-0,16	0,21	0,24	-0,06	-0,02	0,17	0,46	-0,62	0,49
0,20	-0,54	0,45	0,43	-0,16	-0,03	-0,20	0,03	0,04	-0,50
0,000	0,58	0,42	0,45	0,03	0,13	0,25	0,19	0,41	0,06

EXEMPLO

Os autovetores na Tabela 3 fornecem os coeficientes das variáveis X para as componentes principais.

$$\begin{aligned} Z_1 &= 0,51(AGR) + 0,37(MIN) - 0,25(FAB) \\ &\quad - 0,32(FEA) - 0,22(CON) - 0,38(SER) \\ &\quad - 0,13(FIN) - 0,43(SSP) - 0,21(TC) \end{aligned}$$

$$var(Z_1) = 3,11$$

Autovale- lores	Autovetores								
	AGR	MIN	FAB	FEA	CON	SER	FIN	SSP	TC
3,11	0,51	0,37	-0,25	-0,32	-0,22	-0,38	-0,13	-0,43	-0,21
1,81	0,02	0,00	-0,43	-0,11	0,24	0,41	0,55	-0,05	-0,52
1,50	0,28	-0,52	0,50	0,29	-0,07	-0,07	0,10	-0,36	-0,41
1,06	-0,02	-0,11	-0,06	-0,02	-0,78	-0,17	0,49	0,32	0,04

Fazendo $a_{ij} = \sqrt{\lambda_j} b_{ji}$ e chamando $F_i = \frac{Z_i}{\sqrt{\lambda_i}}$ e considerando $m = 4$, teríamos:

$$\begin{aligned}
 a_{11} &= \sqrt{\lambda_1} b_{11} = \sqrt{3,11} * 0,51 = 0,90 \\
 a_{12} &= \sqrt{\lambda_2} b_{21} = \sqrt{1,81} * 0,02 = 0,03 \\
 a_{13} &= \sqrt{\lambda_3} b_{31} = \sqrt{1,50} * 0,28 = 0,34 \\
 a_{14} &= \sqrt{\lambda_4} b_{41} = \sqrt{1,06} * (-0,02) = -0,02
 \end{aligned}$$

$$X_1 = +\mathbf{0,90} F_1 + 0,03 F_2 + 0,34 F_3 - 0,02 F_4 + e_1 (0,93)$$

Estes são transformados em cargas de fator para 4 fatores usando Eq. 5.2 para dar o modelo:

$$X_1 = +\mathbf{0,90} F_1 + 0,03 F_2 + 0,34 F_3 - 0,02 F_4 + e_1 (0,93)$$

$$X_2 = +\mathbf{0,66} F_1 + 0,00 F_2 - \mathbf{0,63} F_3 - 0,12 F_4 + e_2 (0,85)$$

$$X_3 = -0,43 F_1 - \mathbf{0,58} F_2 + \mathbf{0,61} F_3 - 0,06 F_4 + e_3 (0,91)$$

$$X_4 = -\mathbf{0,56} F_1 - 0,15 F_2 + 0,36 F_3 - 0,02 F_4 + e_4 (0,46)$$

$$X_5 = -0,39 F_1 + 0,33 F_2 - 0,09 F_3 - \mathbf{0,81} F_4 + e_5 (0,92)$$

$$X_6 = -\mathbf{0,67} F_1 + \mathbf{0,55} F_2 - 0,08 F_3 - 0,17 F_4 + e_6 (0,79)$$

$$X_7 = -0,23 F_1 + \mathbf{0,74} F_2 + 0,12 F_3 + \mathbf{0,50} F_4 + e_7 (0,87)$$

$$X_8 = -\mathbf{0,76} F_1 - 0,07 F_2 - 0,44 F_3 + 0,33 F_4 + e_7 (0,88)$$

$$X_9 = -0,36 F_1 - \mathbf{0,69} F_2 - \mathbf{0,51} F_3 + 0,04 F_4 + e_9 (0,87)$$

EXEMPLO

Obs.:

- Os valores entre parênteses são as comunalidades, i.é. Para X_1 é:

$$(0,90)^2 + (0,03)^2 + (0,34)^2 + (-0,02)^2 = 0,93$$

- A comunalidade de X_i é a parte de sua variância que é relacionada aos fatores comuns.
- As comunalidades são bastante altas para as variáveis exceto X_4 (FEA, fornecimento de energia e água)
- Portanto grande parte da variância para as outras 8 variáveis originais está, portanto, contida nos quatro fatores comuns.

EXEMPLO

$$X_1 = +\mathbf{0,90} F_1 + 0,03 F_2 + 0,34 F_3 - 0,02 F_4 + e_1 (0,93)$$

$$X_2 = +\mathbf{0,66} F_1 + 0,00 F_2 - \mathbf{0,63} F_3 - 0,12 F_4 + e_2 (0,85)$$

- Cargas fatoriais $\geq |0,50|$ indicam como as variáveis estão relacionadas com os fatores.
- Pode ser visto que X_1 é quase inteiramente explicada pela fator 1 sozinho;
- X_2 é uma mistura do fator 1 com o fator 3.

EXEMPLO

Dificuldade de interpretação: Cinco das nove variáveis X são fortemente relacionadas a dois fatores.

Isso sugere que uma rotação pode oferecer um modelo mais simples para os dados.

Um rotação varimax foi realizada e produziu o modelo:

EXEMPLO

$$X_1 = +\mathbf{0,85} F_1 - 0,10 F_2 - 0,27 F_3 + 0,36 F_4 + e_1 (0,93)$$

$$X_2 = +0,11 F_1 - 0,30 F_2 - \mathbf{0,86} F_3 + 0,10 F_4 + e_2 (0,85)$$

$$X_3 = -0,03 F_1 - 0,32 F_2 + \mathbf{0,89} F_3 + 0,09 F_4 + e_3 (0,91)$$

$$X_4 = -0,19 F_1 + 0,04 F_2 + \mathbf{0,64} F_3 - 0,14 F_4 + e_4 (0,46)$$

$$X_5 = -0,02 F_1 - 0,08 F_2 + 0,04 F_3 - \mathbf{0,95} F_4 + e_5 (0,92)$$

$$X_6 = -0,35 F_1 + 0,48 F_2 + 0,15 F_3 - \mathbf{0,65} F_4 + e_6 (0,79)$$

$$X_7 = -0,08 F_1 + \mathbf{0,93} F_2 + 0,00 F_3 + 0,01 F_4 + e_7 (0,87)$$

$$X_8 = -\mathbf{0,91} F_1 + 0,17 F_2 + 0,12 F_3 - 0,04 F_4 + e_8 (0,88)$$

$$X_9 = -\mathbf{0,73} F_1 - \mathbf{0,57} F_2 + 0,30 F_3 + 0,14 F_4 + e_9 (0,87)$$

EXEMPLO

As comunalidades não mudam: Para X_1 é:
$$(0,85)^2 + (-0,10)^2 + (-0,27)^2 + (0,36)^2 = 0,93$$

Essa solução é apreciavelmente melhor, pois apenas X_9 é apreciavelmente dependente de mais do que um fator.

EXEMPLO

$$X_1 = +\mathbf{0,85} F_1 - 0,10 F_2 - 0,27 F_3 + 0,36 F_4 + e_1 (0,93)$$

$$X_2 = +0,11 F_1 - 0,30 F_2 - \mathbf{0,86} F_3 + 0,10 F_4 + e_2 (0,85)$$

$$X_3 = -0,03 F_1 - 0,32 F_2 + \mathbf{0,89} F_3 + 0,09 F_4 + e_3 (0,91)$$

$$X_4 = -0,19 F_1 + 0,04 F_2 + \mathbf{0,64} F_3 - 0,14 F_4 + e_4 (0,46)$$

$$X_5 = -0,02 F_1 - 0,08 F_2 + 0,04 F_3 - \mathbf{0,95} F_4 + e_5 (0,92)$$

$$X_6 = -0,35 F_1 + 0,48 F_2 + 0,15 F_3 - \mathbf{0,65} F_4 + e_6 (0,79)$$

$$X_7 = -0,08 F_1 + \mathbf{0,93} F_2 + 0,00 F_3 + 0,01 F_4 + e_7 (0,87)$$

$$X_8 = -\mathbf{0,91} F_1 + 0,17 F_2 + 0,12 F_3 - 0,04 F_4 + e_8 (0,88)$$

$$X_9 = -\mathbf{0,73} F_1 - \mathbf{0,57} F_2 + 0,30 F_3 + 0,14 F_4 + e_9 (0,87)$$

EXEMPLO

Nomeando os fatores:

- O fator 1 tem carga alta positiva para X_1 (agricultura, florestal e pesca) e cargas negativas altas para X_8 (serviços sociais e pessoais) e X_9 (transporte e comunicações). Ele mede o quanto de pessoas estão empregadas em agricultura em vez de em serviços e comunicações. Ele poderia ser chamado de “Indústrias rurais em contraste a serviços sociais e comunicação”.
- O fator 2 tem cargas positivas altas para X_7 (finanças) e um coeficiente negativo alto para X_9 (transporte e comunicações). Este pode ser chamado “indústria de finanças” em contraste com transporte e comunicações.

EXEMPLO

- O fator 3 tem uma carga negativa para X_2 (mineração e exploração de pedreiras), uma carga positiva alta para X_3 (fabricação) e uma carga moderadamente alta para X_4 (suprimento de energia). Este pode ser chamado “mineração em contraste a fabricação”
- O fator 4 tem uma carga negativa alta para X_5 (construção) e uma carga negativa moderadamente alta para X_6 (indústrias de serviços). Este poderia chamar-se “falta de Indústrias de construção e de serviços”

OPÇÕES EM ANÁLISES

- Há quatro ou cinco métodos para a extração de fatores e o mesmo número de métodos de rotação destes, incluindo não rotação. Isso fornece pelo menos 20 diferentes tipos de análises de fatores que podem ser até certo ponto diferentes.
- **Pressuposto:** Normalidade dos dados.

A IMPORTÂNCIA DA ANÁLISE DE FATORES

- É uma arte! Dependendo dos dados o modelo pode não se ajustar.
- Evitar executar AF em uma única amostra pequena.
- Chatfield and Collins (1980, p. 89) listam 6 problemas com AF, concluindo: “AF não deveria ser usada em muitas situações práticas”.
- Os resultados dessa análise são úteis! Por isso ela é muito utilizada!