



ANÁLISE FATORIAL

André Eiji Arahawa
Fernando Karchiloff G. de Amorim
Leonardo Victor Moreira de Souza
Lucas Pereira Castelo Branco
Priscila Shibata Mendes

Roteiro

- A técnica
- Utilizações
- Análise Fatorial Exploratória
- Análise Fatorial Confirmatória
- Estudo de caso

SOBRE A TÉCNICA

Uma técnica de interdependência, pois examina-se todo um conjunto de relações interdependentes.

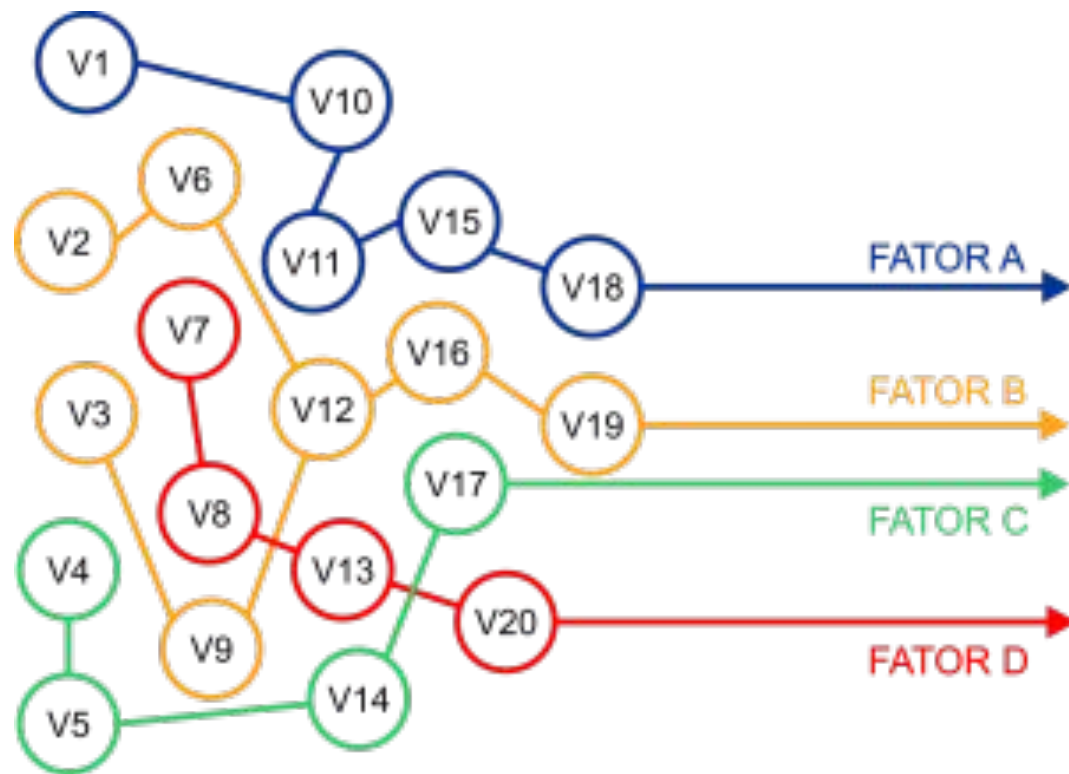
Uma classe de procedimentos utilizados para redução e resumo dos dados, de forma a identificar componentes que se relacionam aos dados, chamados de **fatores**.

ONDE É UTILIZADA?

Identificar um conjunto novo e menor de **características latentes e não-correlacionadas**, o que facilita na visualização e interpretação dos dados, para **identificar fatores que expliquem correlações** entre um conjunto de variáveis (perfis psicográficos, por exemplo);

Com esse conjunto menor, pode ser utilizado em uma análise multivariada subsequente.

ONDE É UTILIZADA?



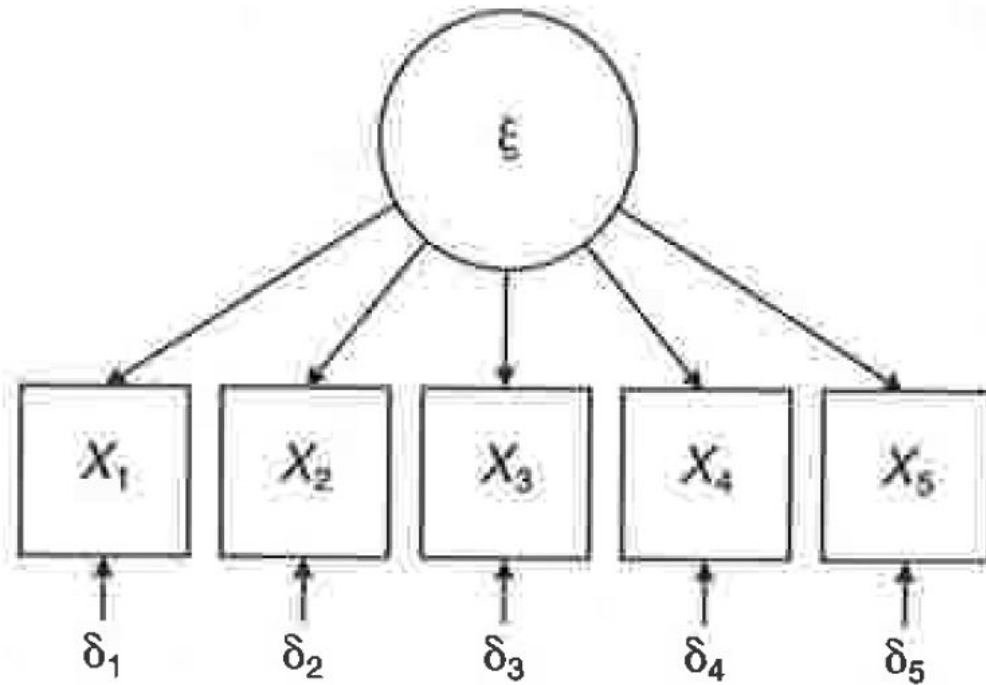
AF EXPLORATÓRIA

Como a análise de componentes principais, é um método que pode ser usado para **reduzir a dimensionalidade** dos dados multivariados.

Também é útil para compreensão dos **padrões de correlação** nos dados e para identificar os traços ou **características subjacentes**.

Os **fatores comuns** são assim chamados porque também **contribuem para a variação** em todas as variáveis X . Supondo que as variáveis são padronizadas (isto é, que $\text{var}(X_i)=1$ para todos os i), a **proporção de variância** em X_i atribuível aos fatores comuns é chamada de **comunalidade**.

INTUIÇÃO



$$X_1 = \lambda_1 \xi + \delta_1$$

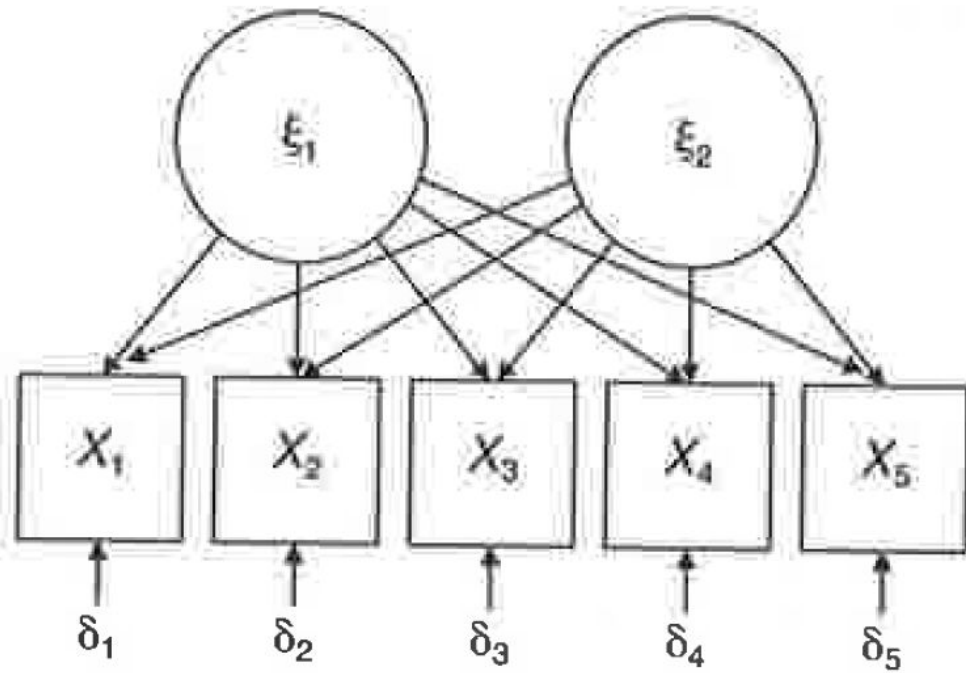
$$X_2 = \lambda_2 \xi + \delta_2$$

$$X_3 = \lambda_3 \xi + \delta_3$$

$$X_4 = \lambda_4 \xi + \delta_4$$

$$X_5 = \lambda_5 \xi + \delta_5$$

INTUIÇÃO – DOIS FATORES



$$X_1 = \lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \delta_1$$

$$X_2 = \lambda_{21}\xi_1 + \lambda_{22}\xi_2 + \delta_2$$

$$X_3 = \lambda_{31}\xi_1 + \lambda_{32}\xi_2 + \delta_3$$

$$X_4 = \lambda_{41}\xi_1 + \lambda_{42}\xi_2 + \delta_4$$

$$X_5 = \lambda_{51}\xi_1 + \lambda_{52}\xi_2 + \delta_5$$

PROCEDIMENTOS

Considerando que o problema já foi formulado, com base nos dados obtidos, constrói-se uma **matriz de correlação**. Porém, é preciso saber se a análise fatorial pode ser realizada adequadamente, com dois testes:

Teste de Esfericidade de Bartlett

Rejeitar ou não a hipótese nula (variáveis não correlacionadas)

Medida de adequação da amostra de Kaiser-Meyer-Olkin(KMO)

Valores abaixo de 0,5 indicam que as correlações entre pares de variáveis não são explicadas por outras variáveis

Matriz Anti-Imagem

Valores para cada variável individualmente, devendo-se excluir as que se encontram em domínio aceitável; quanto maior, melhor

RELAÇÕES KMO

KMO	Análise Fatorial
1 – 0,9	Muito boa
0,8 – 0,9	Boa
0,7 – 0,8	Média
0,6 – 0,7	Razoável
0,5 – 0,6	Má
< 0,5	Inaceitável

EXTRAINDO FATORES

Depois de visto que a análise fatorial pode ser utilizada, verifica-se **qual o método** será utilizado: *análise de componentes principais* ou a *análise fatorial comum*.

A diferença entre eles é que a **ACP** leva em conta a **variância total** dos dados para determinar o número mínimo de fatores que representam a máxima variância do conjunto de dados.

Já na **AFC**, os fatores são estimados pela **variância comum**, e a sua preocupação é **identificar as dimensões subjacentes**.

DETERMINANDO O NÚMERO DE FATORES

A priori: Já é sabido quantos fatores pode se esperar, por conhecimento prévio do pesquisador

Critério de Raiz Latente ou *Método de Kaiser*: Apenas os fatores com autovalores acima de 1,0 são representativos, onde cada autovalor **representa a quantidade de variância associada ao fator**. Valores menores que 1,0 não são melhores que uma variável isolada.

Critério do Gráfico de Declive (scree plot): O gráfico apresenta um grande declive entre os fatores de autovalores elevados e os fatores de autovalores reduzidos. Os **fatores anteriores ao declive** são os selecionados.

Porcentagem da variância: Baseado na porcentagem (resultado do autovalor da variável sobre o número de variáveis x 100), recolhe-se fatores até onde se alcance **no mínimo 60% da variância total**.

ROTACIONANDO OS FATORES

A matriz de fatores padronizados possui os coeficientes (**cargas fatoriais**) que representam as **correlações** entre os fatores e as variáveis. Uma carga fatorial com **valor absoluto elevado** indica que o fator e a variável estão **fortemente relacionados**.

Porém, na matriz original, não se pode interpretar os dados, pois alguns fatores podem se relacionar com muitas variáveis. Assim, com uma **rotação**, obtém-se uma matriz mais simples e que **pode levar a uma melhor visualização**.

ROTACIONANDO OS FATORES

Existem dois métodos diferentes de rotacionar a matriz de fatores:

- **Ortogonal** (onde os eixos são mantidos em ângulo reto)

Dentre eles, o varimax, o qual minimiza o número de variáveis com altas cargas sobre um fator, facilitando a interpretação.

- **Oblíqua** (os eixos não são mantidos em ângulo reto)

Quando os fatores tendem a ser fortemente correlacionados, pode ser utilizada uma rotação oblíqua, já que isso pode simplificar a matriz de fatores padronizados.

ROTAÇÃO ORTOGONAL

Existem diversas formas de realizar rotações. Comumente, usamos a **ortogonal**, na qual podemos citar entre as principais:

- **Varimax:** busca minimizar o número de variáveis que têm alta carga em um fator.
- **Quartimax:** busca minimizar o número de fatores que têm alta carga em uma variável;
- **Equamax:** tenta reduzir, ao mesmo tempo, o número de fatores e variáveis com valor alto de carga.

INDO À INTERPRETAÇÃO

Com os fatores resumidos, a interpretação dos dados fica muito mais simples. As variáveis que apresentam **altas cargas** em cima de um mesmo fator **determinam o que aquele fator representa**.

Uma solução fatorial exhibe **estrutura simples** quando qualquer **variável única** é **altamente correlacionada** com um ou alguns fatores comuns e qualquer **fator comum único** é altamente correlacionado **com somente algumas variáveis**.

AF CONFIRMATÓRIA

Na análise fatorial confirmatória, em lugar de se permitir que os dados sugiram a estrutura do modelo (como na exploratória), começamos com uma **forte noção prévia** da estrutura do modelo e procedemos ao **teste de sua adequação**.

Nessa análise, **impomos a estrutura** sobre a matriz de cargas fatoriais, permitindo que as variáveis carreguem somente sobre apenas um ou alguns fatores. Isso nos permite considerar os **fatores correlacionados** e os termos de **erro**.

AF CONFIRMATÓRIA

Uma **condição necessária** é que o número de parâmetros a ser calculados deve ser menor que o número de elementos sobre e abaixo da diagonal da matriz de covariância observada. A diferença entre essas duas quantidades é o número de **graus de liberdade** associados com o modelo.

Estritamente, um modelo fatorial confirmatório (aquele cuja estrutura do modelo é decidida antes da verificação dos dados) **não exige validação**. No entanto, **se o modelo é alterado ou ajustado** para melhorar sua qualidade de ajuste, então, a versão final **deve ser novamente testada** com dados de teste ou com dados de uma nova amostra.

AF CONFIRMATÓRIA

Muitos usada em **estudos de psicologia** em que se quer saber a **correlação** de determinada característica a certo perfil, por exemplo. Busca-se **testar a adequação** de fatores que podem ter sido colocados baseados na experiência.

Outra abordagem é de confirmar se determinado fator contribui para a pesquisa ou se está mudando o foco da mesma. Como, por exemplo, **confirmar a influência** da limpeza de uma loja com a satisfação dos clientes.

Estudo de caso

Foi feito um estudo de impressões de consumidores sobre cereais, patrocinado por uma marca muito conhecida, com o objetivo de caracterizar o comportamento de consumidores de cereal. Cada entrevistado precisou qualificar 3 de suas marcas preferidas, levando em consideração 25 atributos diferentes, utilizando uma escala de 5 pontos para a avaliação.

KMO e MSA

\$KMO							
0,85468199							
\$MSA							
Satisfaz	Natural	Fibra	Doce	Fácil			
0.8870484	0.9035202	0.8848490	0.7837295	0.8328754			
sal	Gratificante	Energia	Divertido	Crianças			
0.8172068	0.9111749	0.9057424	0.8536214	0.6773331			
Encharcado	Econômico	Saúde	Família	Calorias			
0.6282512	0.7325581	0.9169197	0.7305559	0.8574263			
Simple	Crocante	Regular	Açúcar	Fruta			
0.8151786	0.8302530	0.8655510	0.7849365	0.7735684			
Processo	Qualidade	Prazer	Chato	Nutritivo			
0.7925207	0.9066051	0.8776079	0.8753191	0.9230772			

Loadings:				
	Fator 1	Fator 2	Fator 3	Fator 4
Satisfaz	-0.744	0.111	0.193	-0.135
Natural	-0.786	-0.205		
Fibra	-0.840	-0.112	-0.133	
Doce		0.741		-0.341
Fácil	-0.264		0.414	
Sal		0.761		0.122
Gratificante	-0.651		0.440	-0.160
Energia	-0.703	0.109	0.191	-0.191
Divertido	-0.170	0.207	0.455	-0.505
Crianças			0.858	
Encharcado			0.148	0.647
Econômico		-0.333	0.525	0.258
Saúde	-0.835	-0.284		
Família			0.806	-0.105
Calorias	0.108	0.715		-0.102
Simples	0.150		0.112	0.729
Crocante		0.150	0.425	-0.515
Regular	-0.667			
Açúcar	0.191	0.822		-0.160
Fruta	-0.416	0.225	-0.338	-0.461
Processo	0.237	0.479		0.202
Qualidade	-0.680	-0.246	0.211	-0.162
Prazer	-0.258	0.266	0.344	-0.613
Chato	0.150	0.101	-0.239	0.612
Nutritivo	-0.846	-0.172		

Fatores

- Saudabilidade: Satisfaz, Natural, Fibra, Gratificante, Energia, Saúde, Regular, Qualidade, Nutritivo
- Energético: Doce, Sal, Calorias, Açúcar
- Público: Crianças, Família, Econômico
- Propriedades: Divertido, Crocante, Prazer; Encharcado, Simples, Chato