Análise Fatorial

ACH2036 – Métodos Quantitativos Aplicados à Adm. de Empresas I

Prof. Regis Rossi A. Faria

2º sem. 2020

EACH
Escola de Artes, Ciências e Humanidades
Universidade de São Paulo
USP

Créditos: Profa. Ana Amélia Benedito Silva (conteúdo parcial de slides)

Programa

- Introdução
- Conceitos
- Exemplos

Análise Fatorial

- Técnica de análise multivariada de interdependência
- Variáveis são agrupadas por meio de suas correlações - aquelas pertencentes a um mesmo grupo serão fortemente correlacionadas entre si, mas pouco correlacionadas com as variáveis de outro grupo.
- Todas as variáveis são simultaneamente consideradas
- Permite "explicar" o comportamento de um número grande de variáveis, em termos de um número relativamente pequeno de fatores.

Análise Fatorial

Suponha que um director de uma fábrica de automóveis pretende entender o que leva um consumidor a escolher um modelo específico de automóvel, isto é, quais os factores que levam os consumidores a escolher um modelo específico de automóvel. Para isso foram consideradas as opiniões de um conjunto de consumidores acerca da importância das seguintes variáveis para a escolha de um automóvel:

CRB - custos de reparação baixos
EIA - espaço interior amplo
FM - fácil de manejar
BM - bom motor
C - confortável
FC - fácil de conduzir
MG - mala grande
VC - variedade de cores à disposição
BC - bom consumo
DM - design moderno
PRA - preço de revenda alto
AS - aparência suave
MA - modelo atraente
FE - fácil de estacionar

Análise Fatorial

É difícil avaliar 14 variáveis separadamente ou desenvolver planos de acção tendo em conta tantas variáveis.
Em vez disso seria ideal saber como pensam os consumidores em termos de dimensões (factores) mais gerais.
Para identificar estas dimensões foi aplicada a análise factorial, cujos resultados sugerem que as 14 variáveis podem ser caracterizadas por

4 factores (**I**, **II**, **III** e **IV**) relacionados com

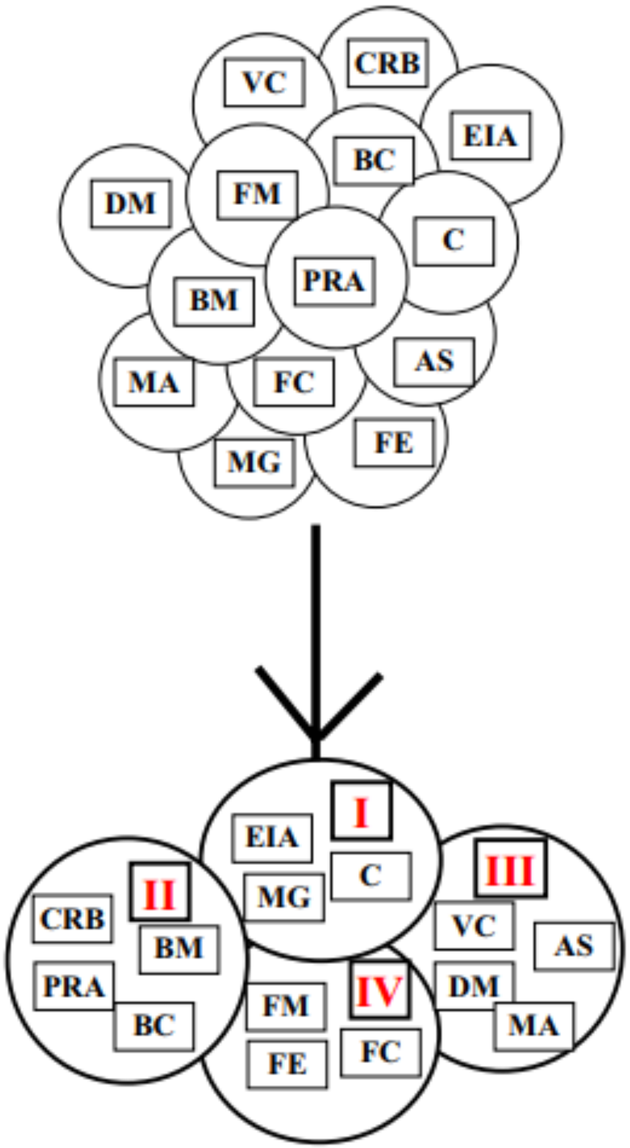
I ↔ **conforto**

II ↔ **custo/eficiência**

III ↔ **estilo**

IV ↔ **facilidade de manipulação**

Análise Fatorial



Referência: aula de análise fatorial exploratória, https://www.youtube.com/watch?v=LMOHQfOfFZg&t=13s

Exemplo

- Escala de Ansiedade-Traço do IDATE (Inventário de Ansiedade Traço-Estado)
  - Um pesquisador aplicou um questionário utilizado na mensuração de traços emocionais existentes em uma pessoa.
  - Deve-se avaliar cada frase, atribuindo-se uma nota entre 1 e 4, na qual 1 indica que aquilo que a frase descreve nunca ocorre e 4 indica que ocorre quase sempre.
  - A medida de ansiedade é obtida a partir da soma das notas de cada frase.
  - O questionário foi aplicado a uma amostra de 1110 estudantes universitários brasileiros

Exemplo

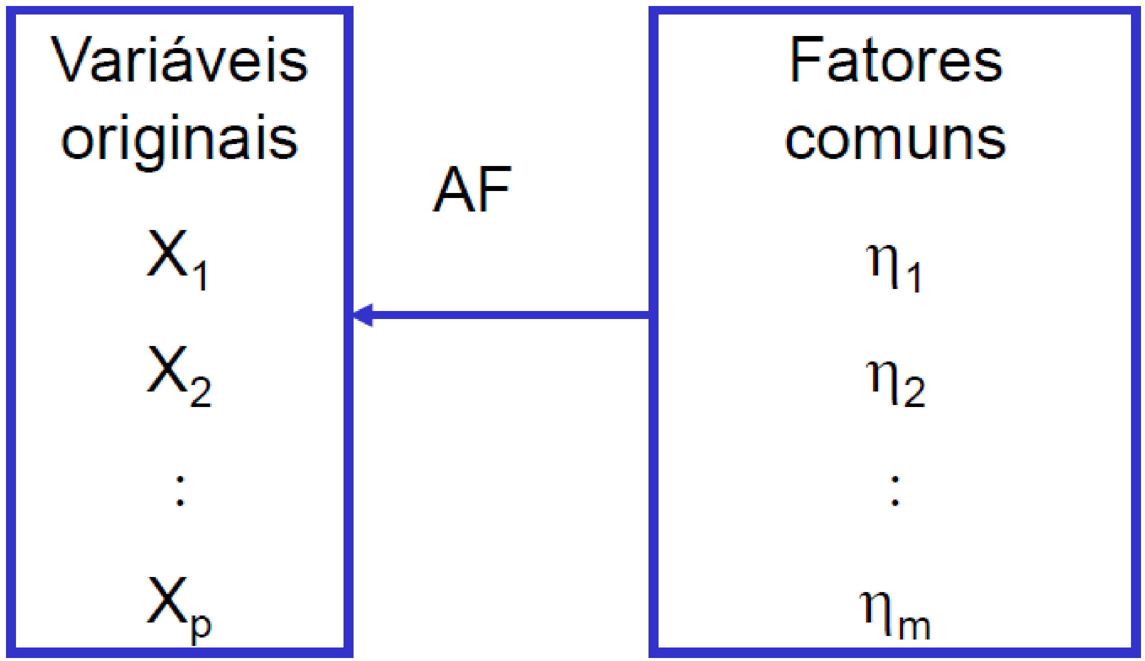
- Escala de Ansiedade-Traço do IDATE (Inventário de Ansiedade Traço-Estado)

X1	Sinto-me bem
X9	Preocupo-me demais com as coisas sem importância
X10	Sou feliz
X11	Deixo-me afetar muito pelas coisas
X13	Sinto-me seguro
X16	Estou satisfeito
X17	As vezes idéias sem importância me entram na cabeça e ficam me preocupando
X18	Levo os desapontamentos tão a sério que não consigo tirá-los da cabeça

Conceito de Análise Fatorial

- Nome genérico dado a uma classe de métodos estatísticos multivariados
- Objetivo principal - definir a estrutura subjacente em uma matriz de dados.
- Analisa a estrutura das correlações entre um grande número de variáveis, definindo um conjunto de dimensões latentes comuns chamadas <u>fatores</u>.

Modelo de Análise Fatorial



$m < p$

Modelo de Análise Fatorial

$$X_1 - \mu_1 = \lambda_{11}\,\eta_1 + \lambda_{12}\,\eta_2 + \ldots + \lambda_{1m}\,\eta_m + \varepsilon_1$$

$$X_2 - \mu_2 = \lambda_{21}\,\eta_1 + \lambda_{22}\,\eta_2 + \ldots + \lambda_{2m}\,\eta_m + \varepsilon_2$$

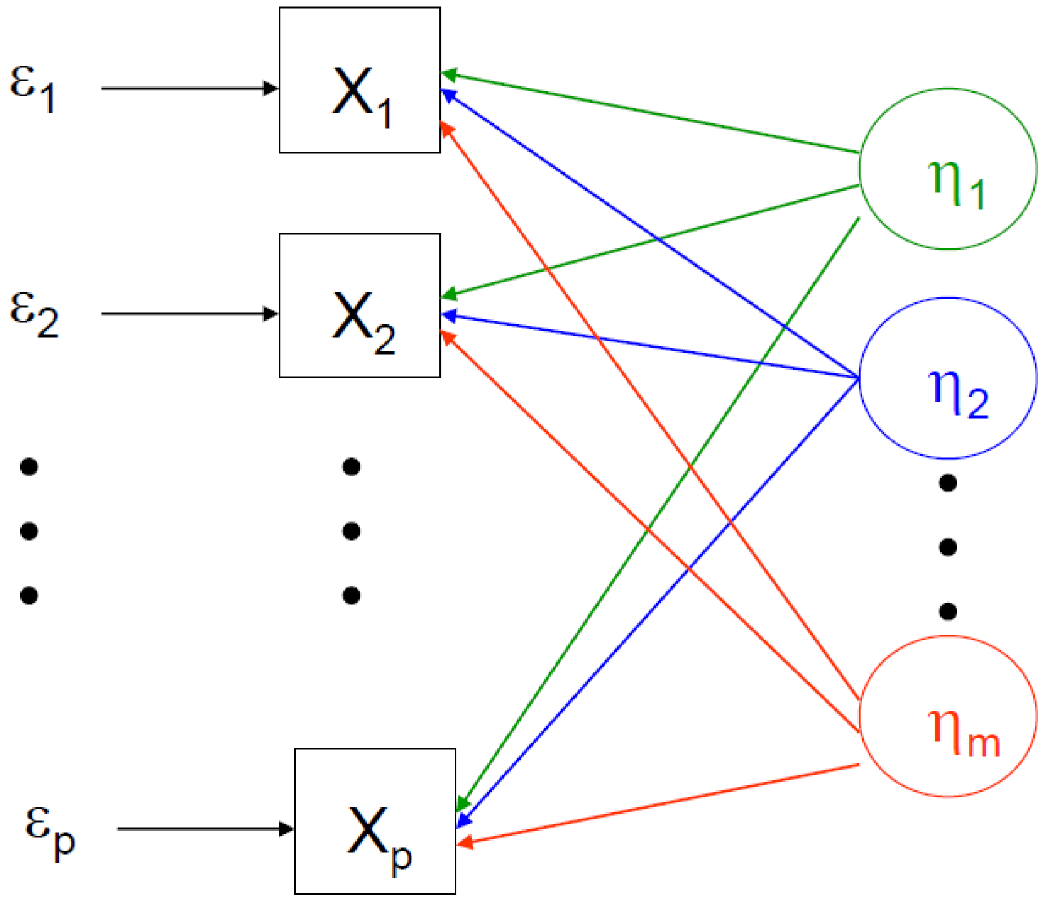
$$\ldots$$

$$X_p - \mu_p = \lambda_{p1}\,\eta_1 + \lambda_{p2}\,\eta_2 + \ldots + \lambda_{pm}\,\eta_m + \varepsilon_p$$

$\eta_1, \ldots, \eta_m$: fatores comuns

$\varepsilon_1, \ldots, \varepsilon_p$: fatores únicos ou específicos

Modelo esquematizado



Tipos de análise fatorial

- Exploratória
  - Não tenho ainda nenhum modelo, ainda estou desenvolvendo
  - Identifica a estrutura fatorial do instrumento e descreve o grau em que as variáveis são relacionadas nos fatores ao se agrupar os itens correlacionados entre si.

Tipos de análise fatorial

- Confimatória
  - Já tenho um modelo
  - Permite aceitar ou rejeitar se uma estrutura fatorial hipotética ou de um modelo previamente estabelecido é ajustada para os dados
  - Verifica o grau de correspondência entre os dados coletados e o modelo de medida proposto.

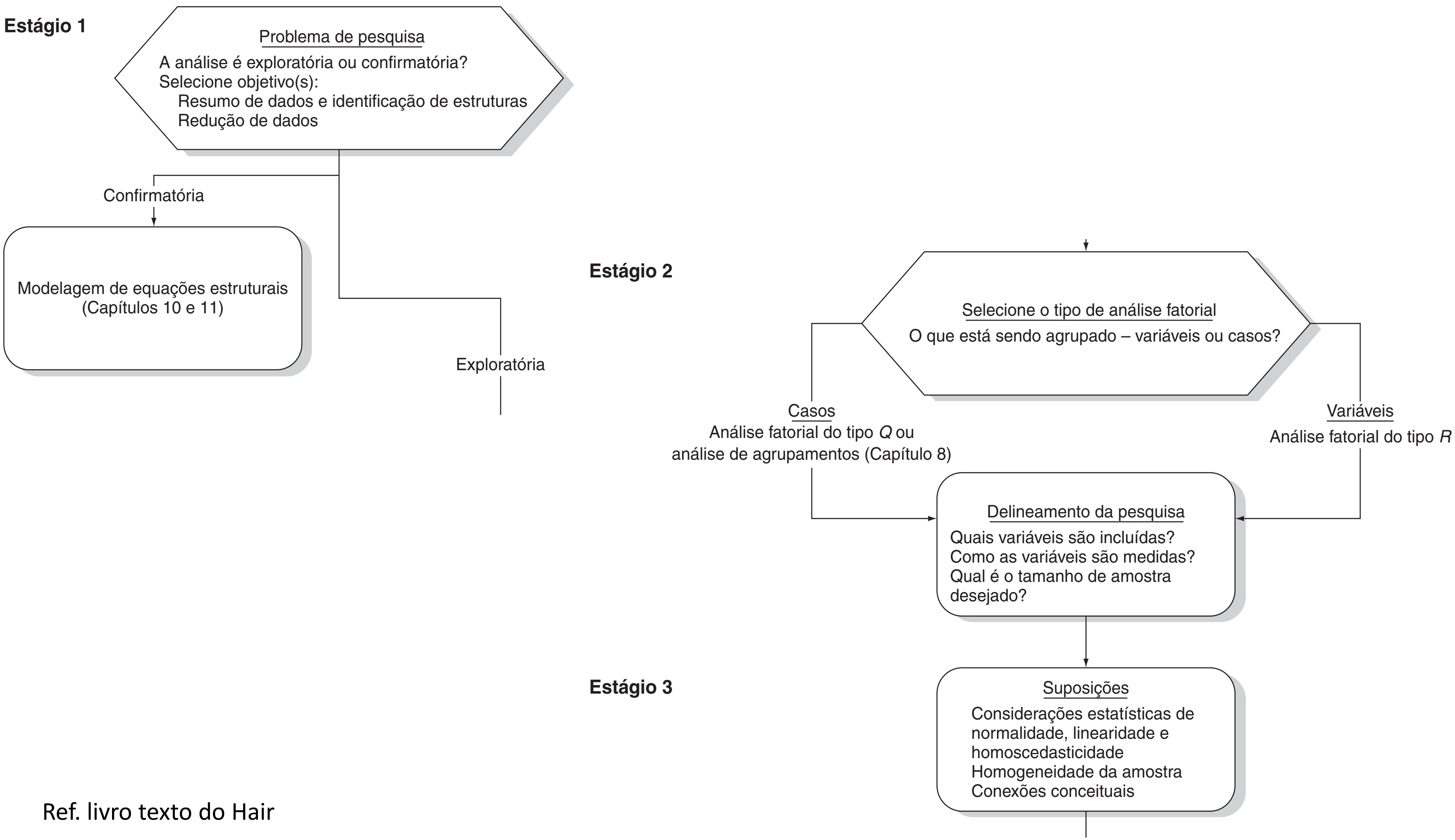
Tipos de análise fatorial

- **Análise R**: agrupar ou resumir características; aplicada a uma matriz correlações, analisa as variáveis para identificar dimensões latentes que as representem;
- **Análise Q**: condensa ou combina grande números de pessoas/ respondentes (casos) em grupos diferentes;
  - Analise de agrupamentos para agrupar respondentes, em muitos casos é preferivel à analise fatorial Q.

Planejamento de uma análise fatorial

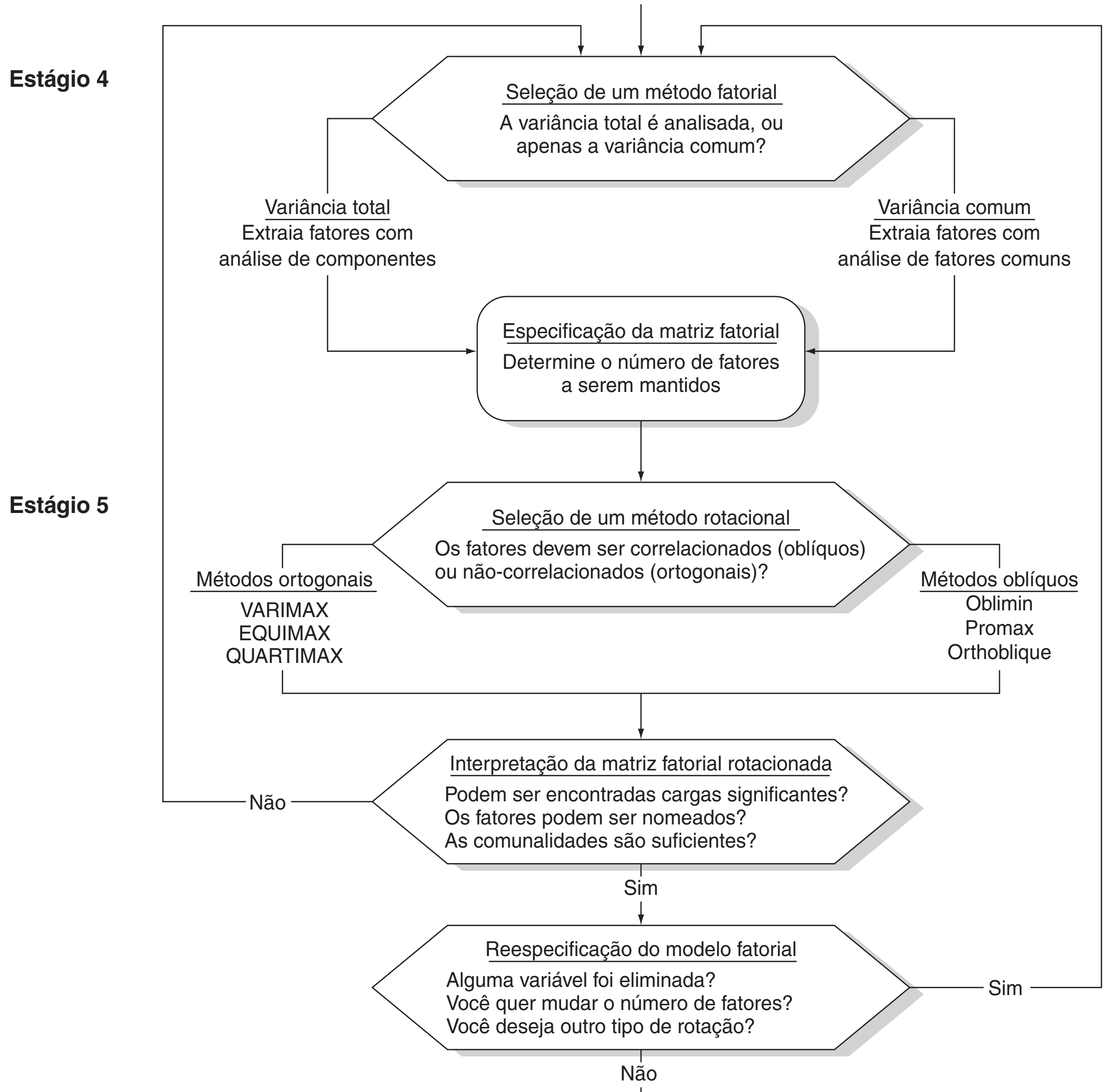
- Matriz de correlação -> Análise Fatorial
- Variáveis
  - Quantitativas contínuas
  - Número de variáveis por fator (5 ou +)
- Tamanho da amostra
  - Depende do número de variáveis por fator
  - Comunalidades -> % de variância comum entre as variáveis (o quanto que cada variável está sendo explicada pela análise)
  - N>= 300, mínimo 20 casos/variável, comunalidades >=0,5

Planejamento em estágios

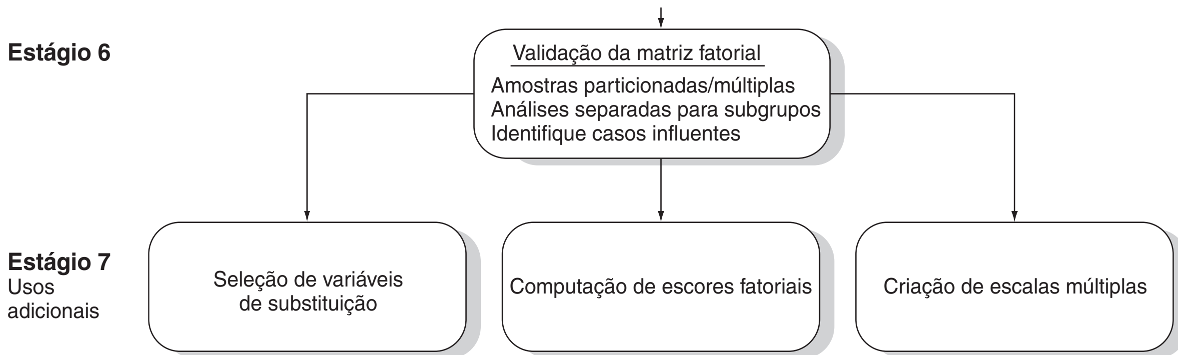


Ref. livro texto do Hair

Planejamento em estágios



Planejamento em estágios



Suposições sobre as variáveis

- **Normalidade**: necessária se testes estatísticos forem aplicados às variáveis ou para a significância dos fatores
- **Teste de Bartlett**: checa se há correlações significativas entre as variáveis
  - aumentar o tamanho da amostra faz o teste ficar mais sensível
- **Teste KMO**: compara relações entre as variáveis, como veremos a seguir

Viabilidade da Análise Fatorial

- Coeficiente **KMO**: Kaiser-Meyer-Olkin
- Proporção de variância que as variáveis apresentam em comum ou que são devidas a fatores comuns.

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

- r^2_{ij} é a correlação entre X_i e X_j
- a^2_{ij} é a correlação parcial entre X_i e X_j, eliminado o efeito das demais variáveis

Variâncias

- **Variância comum**: compartilhada com todas as variáveis na análise (comunalidade)
- **Variância específica**: única, da própria variável, não compartilhada
- **Variância do erro**: devido à erro de medida, presença de componente aleatória no fenômeno medido, ou devido à não-confiabilidade no processo de redução de dados

Interpretação da KMO

KMO	Interpretação
0.90 - 1.00	Excelente
0.80 - 0.90	Ótimo
0.70 - 0.80	Bom
0.60 - 0.70	Regular
0.50 - 0.60	Ruim
0.00 - 0.50	Inadequado
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

Para o exemplo da Escala IDATE:
KMO=0,841

Interpretação da 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

Para o exemplo da Escala IDATE:
KMO=0,841

Viabilidade da Análise Fatorial

- **Teste de Bartlett**
    - H0 : hipótese de que a matriz das correlações seja a matriz identidade com determinante igual a 1.
    - H1 : hipótese de que a matriz das correlações não seja a matriz identidade com determinante igual a 1.

  – Se a H0 não for rejeitada, significa que as variáveis são não-correlacionadas e que a AF não é uma análise adequada.

  – Serve para verificar a adequação em se aplicar a AF

Viabilidade da Análise Fatorial

- **MSA**: *Measure of Sampling Adequacy* (0 a 1; adequado quando MSA>0,5)
- Quantifica o grau de adequação para cada variável

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

- a^2_{ij} é a correlação parcial entre X_i e X_j, eliminado o efeito das demais variáveis
- r^2_{ij} é a correlação entre X_i e X_j

Método de extração de fatores

- Identifica os fatores que melhor caracterizam o conjunto de variáveis
  - **Análise de componentes principais** (ACP ou PCA – *Principal Component Analysis*)
    - Considera a variância total dos dados
    - Procura uma combinação linear das variáveis observadas de maneira a maximizar a variância total explicada
  - **Análise de fatores comuns** (AFC)
    - Permite que seus resultados possam ser utilizados como inputs de outras técnicas multivariadas

ACP ou AFC?

- **ACP** - Análise de componentes principais
  – Redução de dados para obtenção do número mínimo de fatores necessários para explicar o máximo de variância representada pelas variáveis originais
  – Útil quando a variância específica e do erro representam proporção pequena da variância total
- **AFC** - Análise de fatores comuns
  – Identificação de fatores que reflitam o que as variáveis têm em comum
  – Útil quando se desconhece o quanto a variância específica e do erro representam da variância total

Extração de fatores

- Fatores envolvem uma combinação linear das variáveis

$$y_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1q}x_q$$

- Um primeiro fator: deve ser o melhor resumo das relações lineares entre as variáveis → ex: critério da raiz latente (*eigenvalue*)
- Um segundo fator: melhor combinação linear entre as variáveis, mas sujeita à restrição de ser ***ortogonal*** ao primeiro fator → deve ser obtido da variância restante depois da extração do primeiro fator
- E assim por diante
- No caso de PCA, muitas vezes o primeiro fator pode representar um bom índice explicativo (em magnitude) para um critério (ex: severidade de sintomas) e os fatores subsequentes podem dar mais informações diferenciais (ex: sobre os padrões de sintomas)

Escolha do número de fatores

- 1. Critério de Kaiser
  - Escolhe-se o número de fatores a reter em função do número de autovalores > 1
  - Os autovalores mostram a variância explicada por cada fator

Eigenvalues and eigenvectors

From Wikipedia, the free encyclopedia

In linear algebra, an **eigenvector** or **characteristic vector** of a square matrix is a vector that does not change its direction under the associated linear transformation. In other words—if **v** is a vector that is not zero, then it is an eigenvector of a square matrix *A* if *A***v** is a scalar multiple of **v**. This condition could be written as the equation

$$A\mathbf{v} = \lambda\mathbf{v}, \qquad (1)$$

where λ is a number (also called a scalar) known as the **eigenvalue** or **characteristic value** associated with the eigenvector **v**. Geometrically, an eigenvector corresponding to a real, nonzero eigenvalue points in a direction that is stretched by the transformation and the eigenvalue is the factor by which it is stretched. If the eigenvalue is negative, the direction is reversed.[1]

Real matrices [edit]

See also: Euclidean vector and Matrix (mathematics)

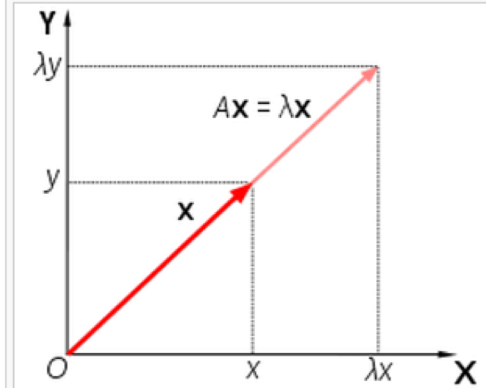
Consider *n*-dimensional vectors that are formed as a list of *n* real numbers, such as the three dimensional vectors,

$$\mathbf{u} = \begin{Bmatrix} 1 \\ 3 \\ 4 \end{Bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{Bmatrix} -20 \\ -60 \\ -80 \end{Bmatrix}.$$

These vectors are said to be scalar multiples of each other, also parallel or collinear, if there is a scalar λ, such that

$$\mathbf{u} = \lambda \mathbf{v}.$$

In this case $\lambda = -1/20$.



Matrix A acts by stretching the vector x, not changing its direction, so x is an eigenvector of A.

Now consider the linear transformation of *n*-dimensional vectors defined by an *n*×*n* matrix *A*, that is,

$$A\mathbf{v} = \mathbf{w},$$

or

$$\begin{bmatrix} A_{1,1} & A_{1,2} & \ldots & A_{1,n} \\ A_{2,1} & A_{2,2} & \ldots & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{n,2} & \ldots & A_{n,n} \end{bmatrix} \begin{Bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{Bmatrix} = \begin{Bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{Bmatrix}$$

where, for each index i,

$$w_i = A_{i,1}v_1 + A_{i,2}v_2 + \cdots + A_{i,n}v_n = \sum_{j=1}^{n} A_{i,j}v_j.$$

If it occurs that **w** and **v** are scalar multiples, that is if

$$A\mathbf{v} = \lambda\mathbf{v},$$

then **v** is an **eigenvector** of the linear transformation *A* and the scale factor λ is the **eigenvalue** corresponding to that eigenvector.

Two dimensional example [edit]

Consider the transformation matrix A, given by,

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

The figure on the right shows the effect of this transformation on point coordinates in the plane. The eigenvectors $\mathbf{v}$ of this transformation satisfy the equation,

$$A\mathbf{v} = \lambda\mathbf{v}.$$

Rearrange this equation to obtain

$$(A - \lambda I)\mathbf{v} = 0,$$

which has a solution only when its determinant $|A - \lambda I|$ equals zero.

Set the determinant to zero to obtain the polynomial equation,

$$p(\lambda) = |A - \lambda I| = 3 - 4\lambda + \lambda^2 = 0,$$

known as the characteristic polynomial of the matrix A. In this case, it has the roots $\lambda = 1$ and $\lambda = 3$.

For $\lambda = 1$, the equation becomes,

$$(A - I)\mathbf{v} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{Bmatrix} v_1 \\ v_2 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix},$$

which has the solution,

$$\mathbf{v} = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}.$$

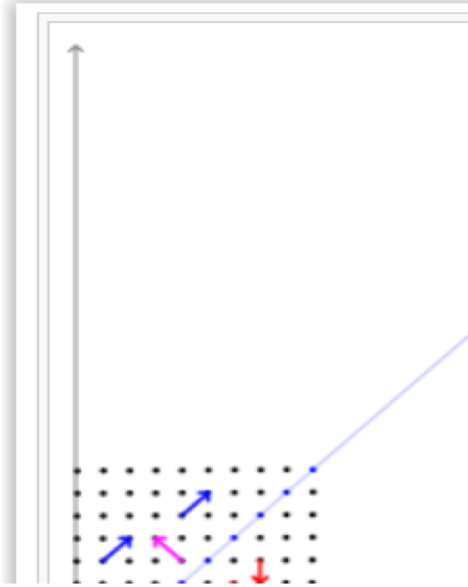
For $\lambda = 3$, the equation becomes,

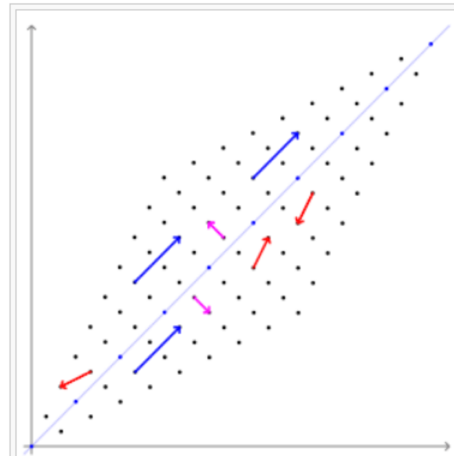
$$(A - 3I)\mathbf{w} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{Bmatrix} w_1 \\ w_2 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix},$$

which has the solution,

$$\mathbf{w} = \begin{Bmatrix} 1 \\ 1 \end{Bmatrix}.$$

Thus, the vectors $\mathbf{v}$ and $\mathbf{w}$ are eigenvectors of A associated with the eigenvalues $\lambda = 1$ and $\lambda = 3$, respectively.





The transformation matrix $A = \left[\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right]$ preserves the direction of vectors parallel to $\mathbf{v} = (1,-1)^T$ (in purple) and $\mathbf{w} = (1,1)^T$ (in blue). The vectors in red are not parallel to either eigenvector, so, their directions are changed by the transformation. See also: An extended version, showing all four quadrants.

Escolha do número de fatores

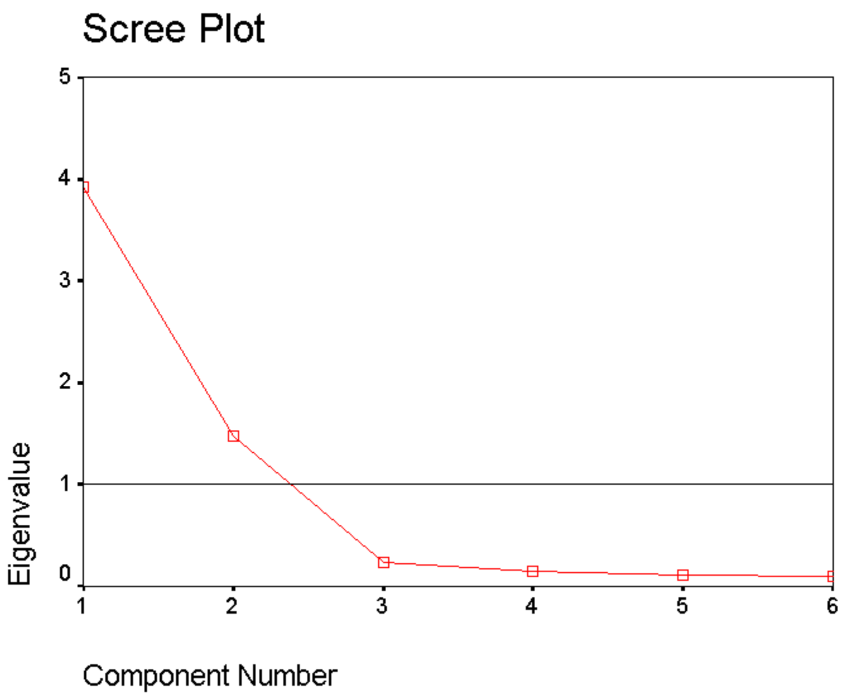
- 2. Critério *a priori*
  - O pesquisador já sabe quantos fatores extrair antes de iniciar a análise
  - Testa uma teoria ou hipótese sobre o numero de fatores a serem extraídos na tentativa de repetir o trabalho de outro pesquisador

Escolha do número de fatores

- 3. Critério da % da variância total explicada
  - O pesquisador fixa o nível satisfatório desejado (acima de 60%)

Escolha do número de fatores

- 4. Critério do gráfico de Scree
  - Identifica o número de fatores antes que a variância única comece a dominar a estrutura de variância comum (reta começa a ficar horizontal)



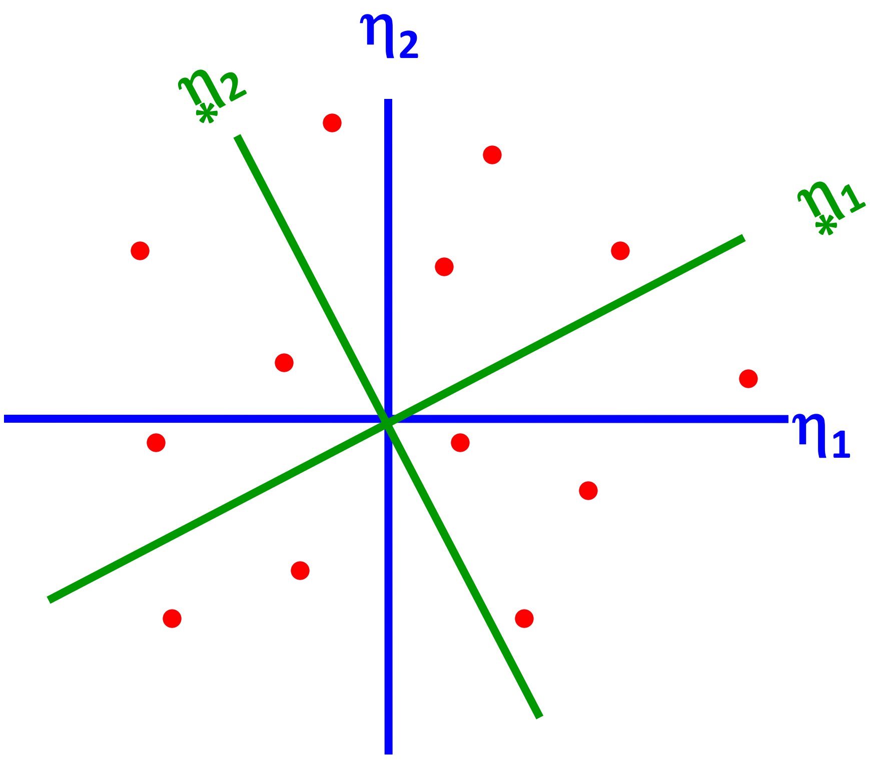
Rotação dos Fatores

- Nem sempre os fatores produzidos na fase de extração são facilmente interpretados
- O método de rotação tem por objetivo transformar os coeficientes dos componentes principais retidos em uma estrutura simplificada
- Como as cargas fatoriais são pontos entre eixos (fatores), podemos girar os eixos sem alterar a distância entre os pontos (relação entre fator e variável)
- A matriz de componentes, após a rotação ortogonal, visa maximizar os valores das cargas fatoriais (*loadings*), de modo que cada variável se associe a apenas um fator
- Variáveis com baixa carga fatorial devem ser eliminadas

Rotação dos Fatores

- Métodos de rotação ortogonais
  - Varimax: minimiza o número de variáveis que têm altas cargas em um fator, simplificando a interpretação dos fatores. Privilegia apenas alguns pesos significativos e todos os outros próximos de zero. É o mais utilizado
  - Quartimax: busca simplificar as linhas de uma matriz fatorial (número de fatores), tornando os pesos de cada variável elevados para um pequeno número de componentes, e próximos de zero todos os demais (minimiza o número de fatores para explicar uma variável)
  - Equamax: congrega características dos outros métodos, com objetivo de simplificar as linhas e colunas simultaneamente (fatores e variáveis)

Rotação - Interpretação geométrica



Exemplo: Solução com dois fatores

η_1 e η_2 definem um plano

η_1^* e η_2^* , obtidos através de uma rotação ortogonal dos eixos, definem o mesmo plano. Logo representam uma solução equivalente.

Exemplo

- Escala de Ansiedade-Traço do IDATE

X1	Sinto-me bem
X9	Preocupo-me demais com as coisas sem importância
X10	Sou feliz
X11	Deixo-me afetar muito pelas coisas
X13	Sinto-me seguro
X16	Estou satisfeito
X17	As vezes idéias sem importância me entram na cabeça e ficam me preocupando
X18	Levo os desapontamentos tão a sério que não consigo tirá-los da cabeça

Matriz de Correlação

- Escala de Ansiedade-Traço do IDATE

	X1	X10	X13	X16	X9	X11	X17	X18
X1	1.00							
X10	0.58	1.00						
X13	0.39	0.47	1.00					
X16	0.51	0.66	0.54	1.00				
X9	-0.14	-0.16	-0.31	-0.22	1.00			
X11	-0.20	-0.24	-0.38	-0.32	0.46	1.00		
X17	-0.18	-0.20	-0.33	-0.25	0.53	0.46	1.00	
X18	-0.32	-0.33	-0.37	-0.40	0.40	0.48	0.48	1.00

Matriz de Correlação

- Escala de Ansiedade-Traço do IDATE

	X1	X10	X13	X16	X9	X11	X17	X18
X1	1.00							
X10	0.58	1.00						
X13	0.39	0.47	1.00					
X16	0.51	0.66	0.54	1.00				
X9	-0.14	-0.16	-0.31	-0.22	1.00			
X11	-0.20	-0.24	-0.38	-0.32	0.46	1.00		
X17	-0.18	-0.20	-0.33	-0.25	0.53	0.46	1.00	
X18	-0.32	-0.33	-0.37	-0.40	0.40	0.48	0.48	1.00

Pergunta: O que sugeriria encontrar uma matriz com altas correlações entre todas as variáveis? E encontrar uma matriz identidade?

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

Para o exemplo da Escala IDATE:
KMO=0,841

Interpretação da MSA

- Para o exemplo da Escala do IDATE

Variável	MSA
X1	0.853
X9	0.818
X10	0.789
X11	0.865
X13	0.899
X16	0.820
X17	0.820
X18	0.878
Média	**0.843**

Não confundir o coeficiente MSA (*Measure of Sampling Adequacy)* com a sigla de Análise do Sistema de Medição - M.S.A (Measurement Systems Analysis) é uma metodologia estatística desenvolvida para estudar e analisar o comportamento do sistema de medição e, proporcionar o aumento de confiança e certeza na leitura obtida nos instrumentos.

Autovalores

- Autovalores e percentual explicado da variância pelos fatores (componentes, no caso de uma ACP) para o exemplo da Escala IDATE

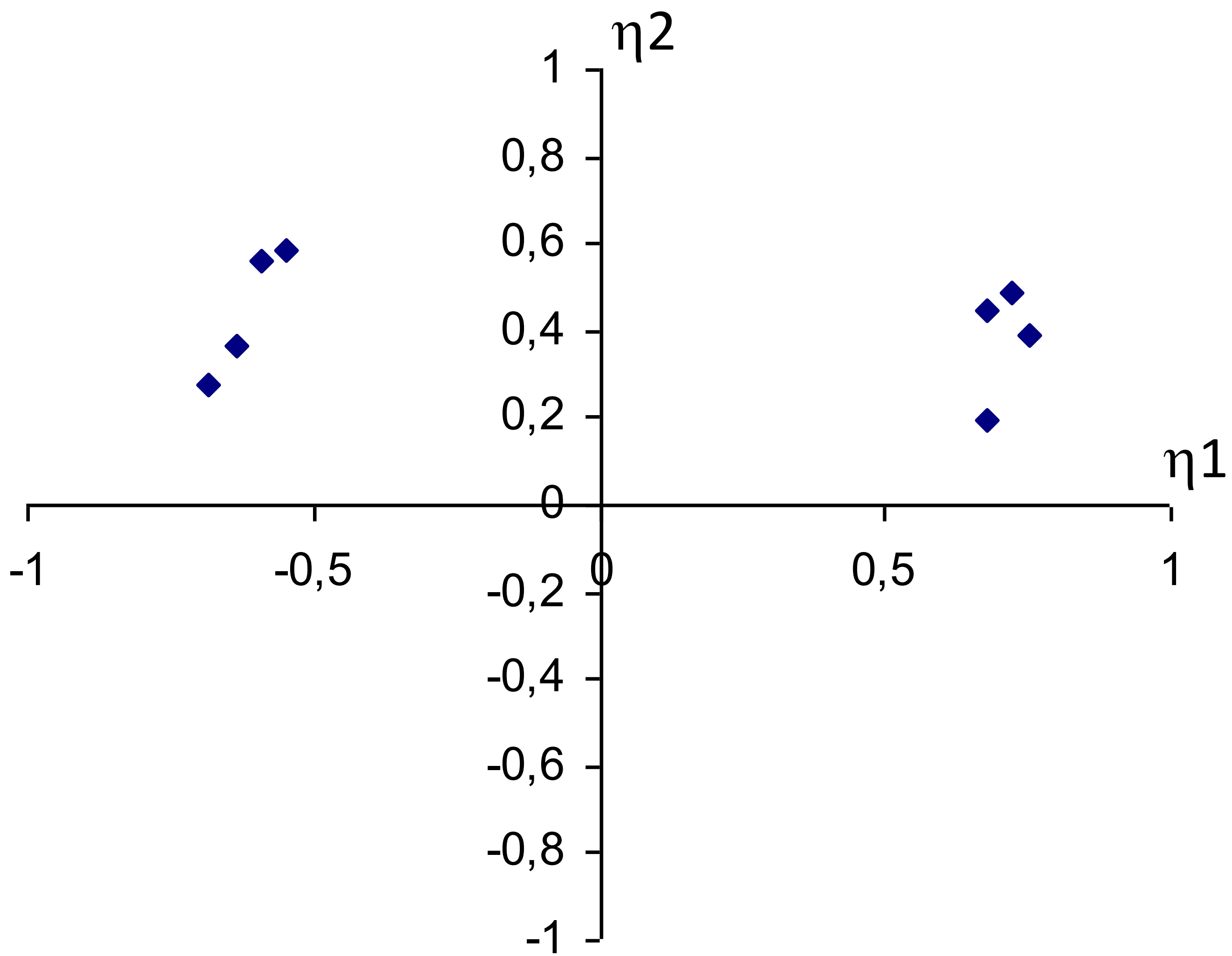
Componente	Autovalores	% da Variância	% Acumulada
1	3.525	44.06	44.06
2	1.504	18.80	62.86
3	0.665	8.31	71.17
4	0.614	7.68	78.85
5	0.512	6.40	85.25
6	0.444	5.55	90.80
7	0.425	5.31	96.11
8	0.311	3.89	100.00

Cargas Fatoriais

- Autovalores e percentual explicado da variância pelos fatores (componentes, no caso de uma ACP) para o exemplo da Escala IDATE

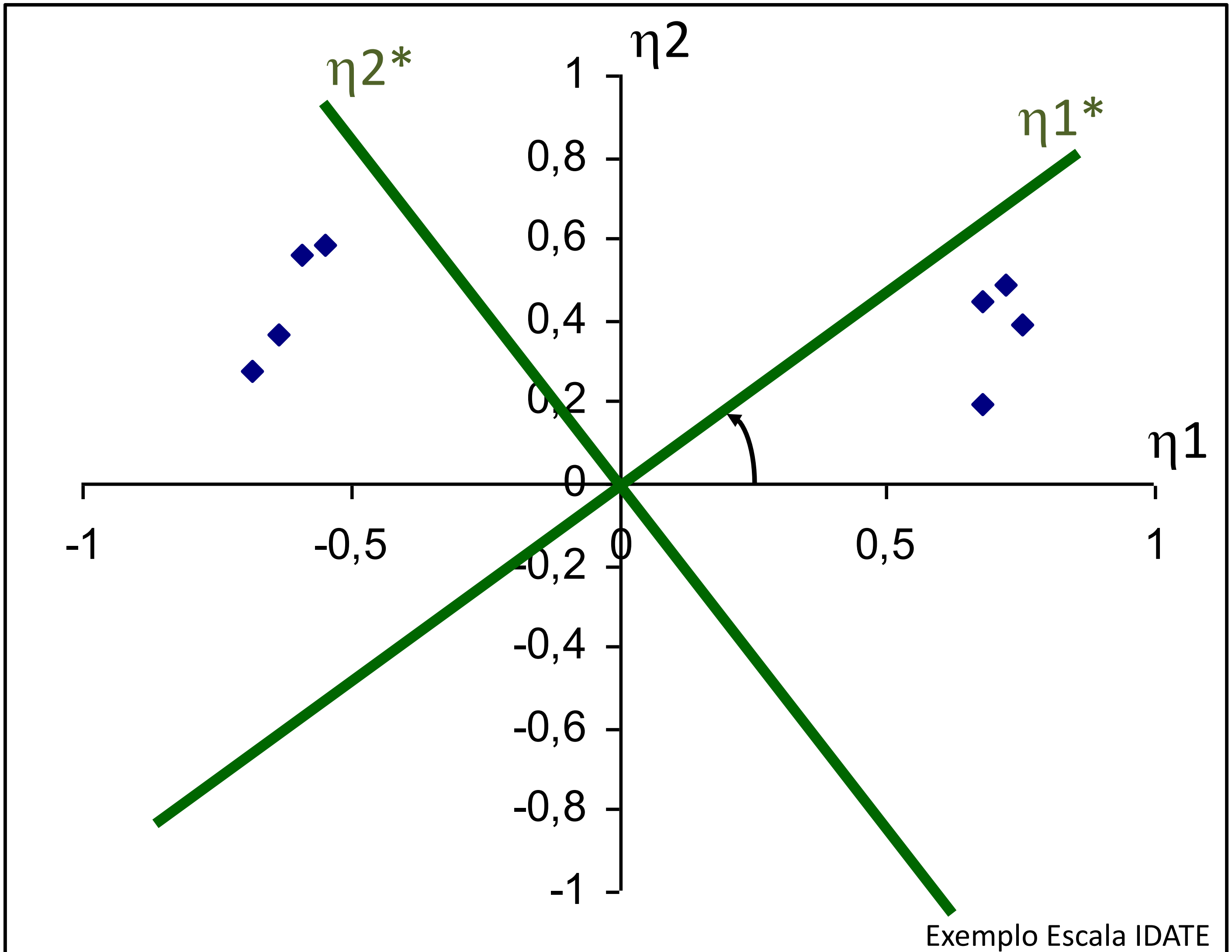
	η_1	η_2
X1	0.678	0.445
X9	-0.549	0.585
X10	0.719	0.492
X11	-0.633	0.367
X13	0.679	0.192
X16	0.751	0.392
X17	-0.593	0.564
X18	-0.686	0.279

Gráfico das Cargas Fatoriais



Exemplo Escala IDATE

Rotação



η2
η2*
η1*
η1
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1
-1
-0,5
0
0,5
1
Exemplo Escala IDATE

Cargas Fatoriais Rotacionadas

- Exemplo Escala IDATE

	η_1^*	η_2^*
X1	0.804	-0.101
X9	-0.038	0.802
X10	0.866	-0.092
X11	-0.244	0.690
X13	0.641	-0.294
X16	0.826	-0.189
X17	-0.086	0.814
X18	-0.341	0.657

O % de explicação da variância continua o mesmo, mas o % por fator foi redistribuído para os novos fatores.

Cargas Fatoriais Rotacionadas

- Note como as cargas fatoriais foram redistribuídas na rotação, concentrando-se mais a explicação das variáveis em um fator

	η_1	η_2
X1	0.678	0.445
X9	-0.549	0.585
X10	0.719	0.492
X11	-0.633	0.367
X13	0.679	0.192
X16	0.751	0.392
X17	-0.593	0.564
X18	-0.686	0.279

	η_1^*	η_2^*
X1	0.804	-0.101
X9	-0.038	0.802
X10	0.866	-0.092
X11	-0.244	0.690
X13	0.641	-0.294
X16	0.826	-0.189
X17	-0.086	0.814
X18	-0.341	0.657

Exemplo Escala IDATE

Cargas fatoriais relacionadas

Var. IDATE	η_1^*	η_2^*	Descrição
X1	0.804	-0.101	Sinto-me bem
X9	-0.038	0.802	Preocupo-me demais com as coisas sem importância
X10	0.866	-0.092	Sou feliz
X11	-0.244	0.690	Deixo-me afetar muito pelas coisas
X13	0.641	-0.294	Sinto-me seguro
X16	0.826	-0.189	Estou satisfeito
X17	-0.086	0.814	As vezes idéias sem importância me entram na cabeça e ficam me preocupando
X18	-0.341	0.657	Levo os desapontamentos tão a sério que não consigo tirá-los da cabeça

Interpretação dos fatores

- Fator 1: Satisfação pessoal
- Fator 2: Dificuldade em lidar com problemas

Interpretação dos fatores

- Devemos escolher quais cargas fatoriais devem ser consideradas
- Normalmente, considera-se apenas cargas fatorais acima de 0,30 (nível mínimo), cargas acima de 0,40 são consideradas mais importantes e, se forem maiores que 0,50 são consideradas estatisticamente significativas

FIM