

UCS (Universidade de Caxias do Sul)

3


UCS

Curso de Especialização em Estatística Aplicada

Prof. Lorí Viali, Dr.  
 viali@pucrs.br;  
 viali@mat.ufrgs.br;  
 http://www.pucrs.br/famat/viali;  
 http://www.mat.ufrgs.br/~viali/

Prof. Lorí Viali, Dr. - PUCRS/UFRGS - Departamento de Estatística

A teoria dos métodos estatísticos multivariados pode ser explicada razoavelmente bem somente com uso de alguma álgebra matricial. Por essa razão é útil, senão essencial ter pelo menos algum conhecimento nessa área (Bryan F. J. Manly).



Estatístico Ecologista com mais de 30 anos de experiência como pesquisador, consultor e professor de Estatística.

Prof. Lorí Viali, Dr. - PUCRS/UFRGS - Departamento de Estatística

# Análise de Fatores

Factor Analysis (FACAN)

Prof. Lorí Viali, Dr. - PUCRS/UFRGS - Departamento de Estatística

As principais aplicações da técnica da "Análise de Fatores" (Factor Analysis) são:

(1) Identificar dimensões latentes, isto é, fatores que justifiquem as correlações observadas entre as variáveis.

Prof. Lorí Viali, Dr. - PUCRS/UFRGS - Departamento de Estatística

(2) Substituir o conjunto original de variáveis (em geral grande) e correlacionadas por um conjunto menor de variáveis sem correlação ou com baixa correlação.

(3) Objetivo Global: parcimônia, isto é, redução da complexidade.

Prof. Lorí Viali, Dr. - PUCRS/UFRGS - Departamento de Estatística

## O que é Análise de Fatores

É uma classe de processos utilizados na redução e sumarização de dados (Malhotra, 2001);

É um nome genérico dado a uma classe de métodos estatísticos multivariados, cujo propósito principal é definir uma estrutura fundamental em uma matriz de dados (Hair et al., 1995).

Prof. Lorí Viali, Dr. - PUCRS/UFRGS - Departamento de Estatística

Além disso, a análise de fatores é aplicada como redutora de dados ou método (exploratório) de estruturas.

**L. L. Thurstone (1887-1955)**  
**Psicometrista**  
<http://www.indiana.edu/~intell/thurstone.html>

O termo "análise de fatores" foi introduzido por Thurstone em 1931.

UCS Prof. Lorí Viali, Dr. - PUCRS/UPFRGS - Departamento de Estatística

Quais variáveis explicam a mesma coisa?

$V_1$	$V_2$	$V_3$
$V_4$	$V_5$	$V_6$
$V_7$	$V_8$	$V_9$

Qualidade  $V_1, V_5$

Experiência  $V_3, V_6, V_8, V_9$

Utilidade  $V_2, V_4, V_7$

UCS Prof. Lorí Viali, Dr. - PUCRS/UPFRGS - Departamento de Estatística

Quais variáveis medem a mesma coisa?

O "quanto" estas variáveis medem a mesma coisa?

Qualidade  $V_1, V_5$

Experiência  $V_3, V_6, V_8, V_9$

Utilidade  $V_2, V_4, V_7$

Fatores

UCS Prof. Lorí Viali, Dr. - PUCRS/UPFRGS - Departamento de Estatística

### Objetivos

- Examinar a "interdependência" entre todas as variáveis (correlações).
- Reduzir diversas variáveis, provavelmente "correlacionadas", a uma quantidade menor e mais facilmente "gerenciável".
- Analisar a estrutura das correlações entre um grande número de variáveis, definindo um conjunto menor de dimensões básicas comuns, chamadas fatores.

UCS Prof. Lorí Viali, Dr. - PUCRS/UPFRGS - Departamento de Estatística

### Importância

- Reduzir "massas" de informação a um tamanho mais facilmente gerenciável;
- A ciência busca explicações mais simples (lei da parcimônia);
- AF propõe-se a reduzir a complexidade das variáveis a uma maior simplicidade.

UCS Prof. Lorí Viali, Dr. - PUCRS/UPFRGS - Departamento de Estatística

### Histórico

- Charles Spearman (1904), psicólogo americano
- Pesquisa sobre habilidades mentais
- Buscava identificar um "fator comum" para matemática, vocabulário, comunicação, arte, lógica, etc...
- Fator básico de inteligência geral - Fator G
- Desenvolveu a Análise de Fatores "Factor Analysis" ou FacAn

UCS Prof. Lorí Viali, Dr. - PUCRS/UPFRGS - Departamento de Estatística

## Aplicações

- Identificar fatores que expliquem as correlações entre um conjunto de variáveis;
- Identificar, em um conjunto maior de variáveis, um conjunto menor que se destaque para uso em uma futura análise multivariada;
- Sumarizar os dados, para obter uma melhor percepção do objeto de pesquisa.



Prof. Lorí Viali, Dr. - PUCRS/UPFRGS - Departamento de Estatística



## Termos básicos

**Comunalidade:** representada por  $h^2$ , é a proporção da variância de uma variável que é compartilhada com os fatores comuns na análise de fatores.

**Autovalor:** é a variância padronizada associada com um particular fator. A soma dos autovalores não pode exceder o número de variáveis (itens), uma vez que cada item contribui com a unidade na soma das variâncias.



Prof. Lorí Viali, Dr. - PUCRS/UPFRGS - Departamento de Estatística



**Fator:** uma combinação linear das variáveis (itens) no sentido de uma regressão, onde o escore total do teste é a variável dependente e os itens são as variáveis independentes.

**Carga do fator:** a carga de um fator expressa a correlação do fator com a variável. O quadrado da carga do fator indica a proporção da variância partilhada entre a variável e o fator.



Prof. Lorí Viali, Dr. - PUCRS/UPFRGS - Departamento de Estatística



**Escore do fator:** Medida composta criada para cada observação em cada fator da análise. Os pesos dos fatores são utilizados em conjunto com os valores originais da variável para calcular cada um dos escores. Os escores dos fatores são padronizados da mesma forma que o escore  $z$ .



Prof. Lorí Viali, Dr. - PUCRS/UPFRGS - Departamento de Estatística



**Matriz padrão dos fatores:** uma matriz contendo os coeficientes ou cargas utilizadas para expressar um item em termos do fator. Ela coincide com a matriz de estrutura se os fatores são ortogonais (não-correlacionados).

**Matriz estrutura dos fatores:** uma matriz contendo as correlações dos itens com cada um dos fatores.



Prof. Lorí Viali, Dr. - PUCRS/UPFRGS - Departamento de Estatística



**Solução rotada dos fatores:** uma solução, onde os eixos são girados com o propósito de mostrar um padrão mais visível das cargas dos fatores.

**Gráfico de declividade (Scree plot):** um diagrama mostrando os autovalores de cada fator.

**Teste de esfericidade de Bartlett:** Verifica se todas as correlações dentro da matriz de correlações são significativas.



Prof. Lorí Viali, Dr. - PUCRS/UPFRGS - Departamento de Estatística



**Matriz de correlação anti-imagem:** é a matriz das correlações parciais entre as variáveis (itens) após a análise de fatores. Representa o grau com que os fatores “explicam” um ao outro nos resultados.

**Análise de factores comum:** modelo de fatores na qual os fatores são baseados numa matriz de correlação reduzida, isto é, as comunalidades são inseridas na diagonal da matriz de correlação e a extração dos fatores é baseada somente na variância comum excluindo as variâncias específicas e do erro.

**Variância do erro:** variância de uma variável devido a erros na coleta ou medida dos valores.

**Medida de adequação da amostra (MAS - Measure of Sampling Adequacy):** medida calculada tanto para a matriz de correlação quanto para cada variável individualmente avaliando a adequação da aplicação da análise de fatores. Valores maiores do que 0,5 tanto para a matriz como um todo quanto para as variáveis individualmente indicam que o método é adequado.

## Etapas

1. Formular o problema
2. Montar matriz de correlações
3. Validar a Análise de Fatores
4. Determinar o método
5. Determinar Número de Fatores
6. Rotacionar Fatores
7. Interpretar Fatores
8. Atribuir Nomes aos Fatores

## 1. Formular o problema

Pesquisar e definir os constructos que fundamentarão as variáveis originais.

Altamente recomendado que as variáveis analisadas sejam especificadas com base em pesquisas anteriores.

Tamanho da amostra: 4 a 5 vezes o número de variáveis na pesquisa.

Observação: no mínimo, 100 unidades.

Determinar as variáveis a se pesquisar;

Definir o tamanho da amostra;

Aplicar o instrumento de pesquisa;

Preparar os dados em uma tabela  $n \times k$  (itens)  $\times$  (variáveis).

## Base de dados $n \times p$

Itens	Variáveis				
	$V_1$	$V_2$	$V_3$	...	$V_p$
$I_1$	$i_{11}$	$i_{12}$	$i_{13}$	...	$i_{1p}$
$I_2$	$i_{21}$	$i_{22}$	$i_{23}$	...	$i_{2p}$
$I_3$	$i_{31}$	$i_{32}$	$i_{33}$	...	$i_{3p}$
...	...	...	...	...	...
$I_n$	$i_{n1}$	$i_{n2}$	$i_{n3}$	...	$i_{np}$

## 2. Montar Matriz de Inter-correlações

Correlacionar todas variáveis entre si

Variáveis	$V_1$	$V_2$	$V_3$	...	$V_p$
$V_1$	1			...	
$V_2$	-0,3	1		...	
$V_3$	0,8	-0,6	1	...	
...	...	...	...	...	
$V_p$	0,8	0,3	0,9	...	1



Prof. Lorí Viali, Dr. - PUCRS/UPFRGS - Departamento de Estatística



## 3. Validar a Análise de Fatores

Confirmar a validade da AF para os dados coletados, através dos seguintes testes:

1. Teste de esfericidade de Bartlett;
2. Adequacidade da amostra de Kaiser-Meyer-Olkin (KMO).



Prof. Lorí Viali, Dr. - PUCRS/UPFRGS - Departamento de Estatística



Considere a matriz de inter-correlações abaixo conhecida como matriz **identidade**.

Variáveis	$V_1$	$V_2$	$V_3$	...	$V_p$
$V_1$	1			...	
$V_2$	0,00	1		...	
$V_3$	0,00	0,00	1	...	
...	...	...	...	...	
$V_p$	0,00	0,00	0,00	...	1



Prof. Lorí Viali, Dr. - PUCRS/UPFRGS - Departamento de Estatística



As variáveis são totalmente não-colineares.  
Se fosse aplicada a AF nesta matriz ...

- Seriam extraídos tantos fatores quanto variáveis, uma vez que cada variável seria seu próprio fator;
- Ela é totalmente não fatorável.



Prof. Lorí Viali, Dr. - PUCRS/UPFRGS - Departamento de Estatística



### 1. Teste de esfericidade de Bartlett

- Testa a hipótese nula, de que as variáveis não sejam correlacionadas na população;
- Fornece a probabilidade de que a matriz de correlações possua correlações significativas em algumas das variáveis.



Prof. Lorí Viali, Dr. - PUCRS/UPFRGS - Departamento de Estatística



- Matriz de correlações = matriz identidade;
- Valor elevado do determinante da matriz favorece a rejeição da hipótese nula;
- Executa uma transformação qui-quadrado do determinante da matriz de correlação.
- Se esta hipótese não puder ser rejeitada, então a conveniência da AF é questionável.



Prof. Lorí Viali, Dr. - PUCRS/UPFRGS - Departamento de Estatística



### 3. Adequacidade da amostra de Kaiser-Meyer-Olkin (KMO)

- É um índice que varia de 0 a 1.
- Compara as magnitudes dos coeficientes de correlação observados, com as dos coeficientes de correlação parciais;
- Índice  $< 0,50$ : as correlações entre os pares de variáveis não podem ser explicadas por outras variáveis - não se deve utilizar a AF.



Prof. Lorí Viali, Dr. - PUCRS/UFRRGS - Departamento de Estatística



### 4. Determinar o Método

#### Selecionar o método de extração de fatores

- Análise dos componentes principais (\*);
- Máxima verossimilhança (\*);
- Fatoração pelo eixo principal (\*);
- Mínimos quadrados generalizados (\*);
- Mínimos quadrados não-ponderados (\*);
- Fatoração alfa (\*);
- Mínimo residual;



Prof. Lorí Viali, Dr. - PUCRS/UFRRGS - Departamento de Estatística



### Componentes Principais

- Modelo no qual os fatores estão baseados na variância total;
- Tanto componentes quanto variáveis;
- O primeiro componente tem a maior variância comum, componentes sucessivos possuem mais variâncias específicas e variância do erro;
- É adequado para extrair a maior proporção da variância com o menor número de fatores.



Prof. Lorí Viali, Dr. - PUCRS/UFRRGS - Departamento de Estatística



### Máxima Verossimilhança

- Produz estimativas dos parâmetros que são as mais prováveis de terem produzidos as correlações observadas;
- A amostra deve ter se originado de uma população normal multivariada.



Prof. Lorí Viali, Dr. - PUCRS/UFRRGS - Departamento de Estatística



### Fatoração pelo eixo principal

- Utiliza o quadrado da correlação múltipla como estimativa das comunalidades;
- As comunalidades são colocadas na diagonal da matriz principal, antes da extração dos fatores.



Prof. Lorí Viali, Dr. - PUCRS/UFRRGS - Departamento de Estatística



### Mínimos quadrados não-ponderados

- Minimiza o quadrado da diferença entre as matrizes de correlações observadas e reproduzidas;



Prof. Lorí Viali, Dr. - PUCRS/UFRRGS - Departamento de Estatística



### Mínimos quadrados generalizados

- Também minimiza o quadrado da diferença entre as matrizes de correlações observadas e reproduzidas;
- Pondera o resultado pelos valores dos itens envolvidos na análise.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



### Fatoração alfa

- Trata os itens como uma amostra da população de todos os possíveis itens;
- Seleciona os fatores com o objetivo de maximizar o coeficiente alfa de confiabilidade.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



### Fatoração pela imagem

- É baseada no conceito de uma “imagem” de um item;
- O item seria a variável (dependente) de uma regressão múltipla onde os demais itens seriam as variáveis independentes.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



### Fatoração pelo mínimo residual

- Os fatores são extraídos da matriz de correlação ignorando os elementos da diagonal.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



## 5. Determinar Número de Fatores

Definir o critério de extração dos fatores.

Pode-se reter todos os fatores cujos autovalores excederem um valor especificado (por exemplo  $\geq 1$ ) ou reter um número específico de fatores.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



Critérios de extração dos fatores:

1. A priori;
2. Com base nos “autovalores”;
3. Com base no “gráfico de declive”;
4. Com base na “percentagem de variância”;
5. Com base na “confiabilidade meio a meio”;
6. Com base em “testes de Significância”.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



### 1. A Priori

- O pesquisador já sabe quantos fatores devem ser extraídos antes de começar;
- Os programas de computador permitem que se informe o número de fatores que se deseja extrair;
- Útil quando o pesquisador está testando uma teoria ou hipótese a respeito do número de fatores.

### 2. Com base nos autovalores

- É a técnica mais comum de extração de valores.
- Qualquer fator deve responder pela variância de pelo menos uma variável;
- Cada variável tem variância = 1 (padrão)
- Deve-se reter fatores com autovalor  $\geq 1$  (critério de Kaiser);
- Mais adequado para quantidade de variáveis entre 20 a 50.

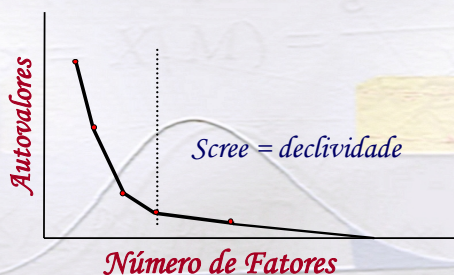
### Autovalores

- Variância extraída por um fator  
Escores padrão tem variância unitária e desta forma a variância total é igual ao número de variáveis na matriz de correlação;
- Um fator com autovalor maior ou igual a um explica mais a variabilidade do que uma única variável.

### 3. Com base na declividade

- Construir o gráfico da declividade
- Plotar os autovalores em ordem de magnitude decrescente.

### Gráfico da Declividade



### 4. Com base no % da variância

Adiciona fatores até que um determinado percentual da variância total seja obtido.



## 6. Rotação dos Fatores

- É o processo de ajustar os eixos dos fatores de modo a obter uma solução mais simples e teoricamente mais significativa;
- Eixo dos fatores: quadro de referência;
- Rotação: “mudança de perspectiva”.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



- O processamento (software) gera a “matriz de fatores”;
- A matriz contém os coeficientes que expressam as variáveis em termos de fatores;
- Os coeficientes são chamados de cargas dos fatores. São as correlações entre variáveis e fatores.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



- A matriz rotada de fatores será utilizada para se poder identificar mais claramente quais variáveis pertencem a quais fatores.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



### Métodos de rotação de fatores:

- Varimax
- Quartimax
- Oblíqua (Oblimin direta)



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



*Varimax* simplifica colunas. Fornece o contraste máximo entre as variáveis dentro de cada fator;

Existem dois métodos de rotação *varimax*;

- Critério de Thurstone e
- *Varimax* de Kaiser



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



### Critério de Thurstone :

Objetiva que cada linha da matriz das cargas dos fatores deve ter pelo menos um zero e que cada fator deve ter um conjunto de cargas próximas de zero. A rotação empregada pelo método era parcialmente subjetiva e não levava a soluções únicas.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



## Varimax de Kayser:

Rotações analíticas melhoram o critério de Thurstone (rotação gráfica). Entre eles o de Kayser é mais utilizado dos métodos de rotação analítica. Este método utiliza a maximização interativa das variâncias das colunas das cargas dos fatores.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



- **Quartimax:** simplifica linhas. Fornece contraste máximo nas cargas dos fatores dentro de cada variável;
- **Obliqua:** roda os eixos de forma que os vértices podem ter ângulos diferentes de 90 graus.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



## 7. Interpretar Fatores

- Construir um “Gráfico de Carregamento de Fatores” (Factor Loading);
- Utilizar a carga dos fatores como coordenadas para o gráfico;
- Preparar o gráfico a partir da matriz rotada de fatores.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



A matriz de inter-correlações  $k \times k$  é reduzida a matriz  $k \times f$  de carga dos fatores

Variáveis	Fatores				
	i	ii	iii	...	f
$V_1$	0,932	0,013	0,250	...	0,405
$V_2$	0,851	0,426	290	...	0,390
$V_3$	0,134	0,651	0,220	...	0,332
...	...	...	...	...	...
$V_p$	0,725	0,344	0,200	...	0,645



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



## Carga dos Fatores

- Correlação das variáveis originais (itens) com os fatores;
- Significância: regra prática (0,5);
- Comunalidades identificam itens “pobres”;
- Quanta variância de cada variável é explicada por todos os fatores;



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



A carga de um fator é a correlação entre a variável e o fator que foi extraído dos dados;

### Exemplo

Variáveis	Fatores				
	i	ii	iii	...	f
$V_1$	0,932	0,013	0,250	...	0,405

Considerando a tabela anterior observe-se a carga do fator “i”.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



## Interpretação

A variável  $V_1$  está altamente correlacionada com o fator “i”, praticamente sem correlação com o fator “ii”, com pequena correlação com o fator “iii” e uma média correlação com o fator “f”.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



## Comunalidade

O quanto da variância da variável  $V_1$  é medida ou explicada pelos “f” fatores que foram extraídos?

Simplemente tome a soma dos quadrados das cargas dos fatores:

$$0,932^2 + 0,013^2 + 0,250^2 + \dots + 0,405^2 =$$

Este resultado é denominado “Comunalidade” da variável.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



## Interpretação

- Os fatores são identificados em relação às variáveis que se manifestam – inspeção da carga dos fatores;
- A rotação simplifica a interpretação da matriz dos fatores;
- Estrutura simples: uma carga alta para cada variável em apenas um fator;



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



- Cada linha terá apenas uma carga significativa;
- Cada coluna terá várias cargas não significativas;
- Entre cada par de colunas cargas não significativas não se sobrepõem;
- Correlação entre fatores levanta a questão de fatores de segunda ordem.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



## Escores dos fatores

A matriz dos coeficientes dos fatores mostra os coeficientes pelo qual os itens são multiplicados para obter os escores dos fatores. Existem três métodos de determinação destes escores.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



**Escores de regressão:** os escores de regressão dos fatores possuem média zero e variância igual ao quadrado do coeficiente de regressão múltipla entre os escores estimados dos fatores e os valores verdadeiros do fator. Eles podem estar correlacionados mesmo quando os fatores são assumidamente ortogonais. A soma das discrepâncias ao quadrado entre os valores verdadeiro e estimado dos fatores sobre todos os itens é minimizada.



Prof. Lorí Viali, Dr. - PUCRS/UFPRGS - Departamento de Estatística



**Escores de Bartlett:** os escores dos fatores de Bartlett apresentam média zero. A soma dos quadrados dos fatores sobre todos os itens é minimizada.

**Escores de Anderson-Rubin:** os escores dos fatores de Anderson-Rubin são uma modificação dos escores de Bartlett para assegurar a ortogonalidade dos fatores estimados. Eles apresentam média zero e desvio padrão um.



## 8. Atribuir nomes aos Fatores

- Avaliar as variáveis de maior carga, de cada fator;
- Atribuir um nome descritivo aos fatores;
- Detectar a “essência” das variáveis individuais;
- Abstrair o objeto de pesquisa.



## Exemplo

Utilizar o arquivo “seven.var” e realizar uma análise de fatores utilizando o SPSS.



## Referências

DARLINGTON, R. B., WEINBERG, S., WALBERG, H. Canonical variate analysis and related techniques. *Review of Educational Research*, p. 453-454, 1973.

HAIR, J. F. Jr. et al. *Multivariate Data Analysis*, 5 ed., Prentice Hall, Upper Saddle River, NJ. 1998.

KACHIGAN, S. K.. *Statistical Analysis*. Radius Press, New York, NY. 1986.

KERLINGER, F. N., *Metodologia da Pesquisa em Ciências Sociais: um tratamento conceitual*. São Paulo: Edusp, 1979.

KIM, J. O.; MUELLER, C. W. *Factor Analysis: What it is and how to do it*. Sage Publications, 1978.

KIM, J. O.; MUELLER, C. W. *Introduction to Factor Analysis*. Sage Publications, 1978.

MALHOTRA, Naresh, K.: *Pesquisa de Marketing: Uma orientação aplicada*. Porto Alegre: Bookman, 2001 (3. ed.), 720 p.

