

Resumo Capítulo Análise Fatorial, HAIR (2009)

# Introdução

- Ao utilizar técnicas multivariadas podemos ter muitas variáveis e a medida que acrescentamos mais variáveis, cada vez mais a correlação pode acontecer entre as mesmas p102
- Quando as variáveis se tornam correlacionadas, o pesquisador precisa de caminhos para gerenciar essas variáveis p102:
  - a) agrupando variáveis altamente correlacionadas;
  - b) rotulando ou nomeando os grupos, e;
  - c) criando uma nova medida composta que possa representar cada grupo de variáveis.

# Análise fatorial

- técnica de interdependência métrica (todas as variáveis quantitativas são simultaneamente consideradas sem distinção quanto ao seu caráter de dependência ou independência p106)
- Seu propósito principal é definir a estrutura inerente entre as variáveis na análise p102.
- A análise fatorial ainda emprega o conceito de variável estatística, a composição linear de variáveis, mas em análise fatorial as variáveis estatísticas (fatores) são formadas para maximizar sua explicação do conjunto inteiro de variáveis, e não para prever uma ou mais variáveis dependentes p106

# Análise fatorial

- Fornece as ferramentas para analisar a estrutura das interrelações (correlações) em um grande número de variáveis (p. ex., escores de teste, itens de teste, respostas a questionários) definindo conjuntos de variáveis que são fortemente inter-relacionadas, conhecidos como fatores p102, com uma perda mínima de informação p104.
- Esses grupos de variáveis (fatores), que são por definição altamente intercorrelacionadas, são considerados como representantes de dimensões dentro dos dados p102

# Análise fatorial R x Análise fatorial Q

relações entre **variáveis** p104

análise fatorial R (matriz de correlação das variáveis)

relações entre **observações** p104

análise fatorial Q (matriz de correlação das observações baseada nas características. Têm-se uma matriz fatorial que identificaria indivíduos semelhantes p107)

# Resumo de dados X Redução de dados

- No resumo de dados, a análise fatorial obtém dimensões inerentes que, quando interpretadas e compreendidas, descrevem os dados em um número muito menor de conceitos do que as variáveis individuais originais p105.
- As variáveis individuais são agrupadas e então vistas não por aquilo que elas representam individualmente, mas por aquilo que representam coletivamente na expressão de um conceito p106
- definindo-se um pequeno número de fatores que adequadamente representam o conjunto original de variáveis p106
- O resumo de dados faz da identificação das dimensões ou fatores latentes um fim em si próprio. Assim, as estimativas dos fatores e as contribuições de cada variável aos fatores (chamadas de cargas) são tudo o que requer a análise p106

# Resumo de dados X Redução de dados

- Redução de dados estende esse processo derivando um valor empírico (escore fatorial) para cada dimensão (fator) e então substituindo o valor original por esse novo valor p105.
  - a) identificação de variáveis representativas a partir de um conjunto muito maior de variáveis para uso em análises multivariadas subsequentes, ou
  - b) pela criação de um conjunto inteiramente novo de variáveis, muito menor, para substituir parcial ou completamente o conjunto original de variáveis.

As cargas fatoriais também são necessárias, mas elas são usadas como a base para identificar variáveis para análises posteriores com outras técnicas ou para fazer estimativas dos próprios fatores (escores fatoriais ou escalas múltiplas), as quais substituem as variáveis originais em análises subsequentes p106.

# AF Exploratória X AF Confirmatória

- Muitos pesquisadores consideram-na apenas exploratória, útil na busca da estrutura em um conjunto de variáveis ou como um método de redução de dados. Sob essa perspectiva, as técnicas analíticas fatoriais “consideram o que os dados oferecem” e não estabelecem restrições a priori sobre a estimação de componentes nem sobre o número de componentes a serem extraídos p102.



# AF Exploratória X AF Confirmatória

- Em outras situações, o pesquisador tem idéias preconcebidas sobre a real estrutura dos dados, baseado em suporte teórico ou em pesquisas anteriores. Ele pode desejar testar hipóteses envolvendo questões sobre, por exemplo, quais variáveis deveriam ser agrupadas em um fator, ou o número exato de fatores p103.
- Nesses casos, o pesquisador espera que a análise fatorial desempenhe um papel confirmatório – ou seja, avalie o grau em que os dados satisfazem a estrutura esperada p103.

# Passo a passo

- Hair (2009) define sete estágios da análise fatorial p104-p128:
  1. Objetivos da análise p104
  2. Planejamento p107
  3. Suposições p109
  4. Determinação de fatores e avaliação do ajuste geral p110
  5. Interpretação dos fatores p115
  6. Validação p123
  7. Usos adicionais dos resultados p124

# Objetivos da análise p104

- Qual o Problema de pesquisa?
  - especificação da unidade de análise (variáveis ou observações?)
  - obtenção do resumo de dados e/ou redução dos mesmos?;
  - seleção de variáveis:
    - resumo de dados baseia-se em ter uma base conceitual para qualquer variável analisada p106
    - redução de dado dimensões conceitualmente definidas podem ser representadas pelos fatores obtidos p107
- uso de resultados da análise fatorial com outras técnicas multivariadas p107

# Planejamento p107

## REGRAS PRÁTICAS 3-1

- cálculo dos dados de entrada p107  
R ou Q?
- planejamento sobre número de variáveis, tipos e propriedades p107
  - métricas (quantitativas) p108
    - não métricas (qualitativas, de preferência dicotomizadas - 0 e 1) p108
- tamanho amostral x número de variáveis p107
  - cinco ou mais para representar cada fator proposto p108
  - o mínimo de cinco vezes mais observações do que o número de variáveis (aceitável teria uma proporção de dez para um) p108

# Suposições p109

## REGRAS PRÁTICAS 3-2

- Uma forte fundamentação conceitual é necessária para embasar a suposição de que existe uma estrutura antes que a análise fatorial seja realizada p110.
- as preocupações que se impõem se centram muito mais no caráter e na composição das variáveis incluídas na análise do que em suas qualidades estatísticas p109.
- Ou seja, a presença de variáveis correlacionadas e a subsequente definição de fatores não garantem relevância p109.
- Os padrões observados devem ser conceitualmente válidos e adequados para se estudar com análise fatorial p109.
- Se grupos diferentes são esperados na amostra, análises fatoriais separadas devem ser executadas, e os resultados devem ser comparados para identificar diferenças não refletidas nos resultados da amostra combinada p109.

# Suposições p109

- um pouco de multicolinearidade é desejável, pois o objetivo é identificar conjuntos de variáveis interrelacionadas e é necessário que sejam suficientemente correlacionadas umas com as outras para produzir fatores representativos p109
  - matriz de correlação: correlações parciais  $>.3$  e  $<.7$ ;
  - teste de esfericidade de Bartlett: teste para correlações significantes entre pelo menos algumas das variáveis;
  - medida de adequação da amostra (MSA): quantifica o grau de intercorrelações entre as variáveis e a adequação da análise fatorial (geral e específica)

$\geq .8$	Admirável		$\geq .5$	Ruim
$\geq .7$	Mediano		$< .5$	Inaceitável
$\geq .6$	Medíocre			

# Determinação de fatores e avaliação do ajuste geral p110

## REGRAS PRÁTICAS 3-3

- método de extração dos fatores (análise de fatores comuns x análise de componentes)

CrITÉRIOS: objetivos da análise fatorial e conhecimento prévio sobre a variância nas variáveis.

variância total (variâncias comum, específica e erro) p112

correlação  $\uparrow$  variância comum  $\uparrow$

variância do erro  $\uparrow$  variância comum  $\downarrow$

# Determinação de fatores e avaliação do ajuste geral p110

## REGRAS PRÁTICAS 3-3

Análise de fatores comuns	Análise de componentes principais
<p>o objetivo prioritário é identificar as dimensões ou construtos latentes;</p> <p>considera apenas variância em comum ou compartilhada;</p> <p>pouco conhecimento sobre a quantia de variância específica e de erro e, logo, não são de interesse na definição da estrutura das variáveis;</p>	<p>redução de dados é uma preocupação prioritária;</p> <p>considera a variância total;</p> <p>conhecimento prévio sugere que variância específica e do erro representam uma proporção relativamente pequena da variância total;</p> <p>deriva fatores que contêm pequenas proporções de variância única e, em alguns casos, variância de erro;</p>



# Determinação de fatores e avaliação do ajuste geral p110

## REGRAS PRÁTICAS 3-3

- Critérios para o número de fatores a extrair p 114
  - interesse na melhor combinação linear de variáveis;
  - fundamentação conceitual x evidência empírica;
  - há consequências negativas na seleção de fatores em excesso ou a menos para representar os dados;
  - Parcimônia é importante.

# Determinação de fatores e avaliação do ajuste geral p110

## REGRAS PRÁTICAS 3-3

- Critérios para o número de fatores a extrair p114
  - raiz latente: fatores que têm raízes latentes ou autovalores maiores que 1 são considerados significantes;
  - *priori*: o pesquisador já sabe quantos fatores extrair;
  - % de variância: percentual cumulativo especificado da variância total extraída por fatores sucessivos (60% ou mais);
  - teste *scree*: identificar o número ótimo de fatores que podem ser extraídos antes que a quantia de variância única comece a dominar a estrutura de variância comum (gráfico das raízes latentes em relação ao número de fatores em sua ordem de extração).
  - Heterogeneidade das observações: heterogeneidade da variância entre subgrupos da amostra.

# Interpretação dos fatores p115

REGRAS PRÁTICAS 3-4; 3-5; 3-6

- o pesquisador deve repetidamente fazer julgamentos subjetivos em decisões p115;
- decisões são melhor guiadas por bases conceituais do que por bases empíricas p115;
- três processos fundamentais:
  1. Estimativa da matriz fatorial: cargas fatoriais para cada variável sobre cada fator. Indicam o grau de correspondência entre a variável e o fator, com cargas maiores tornando a variável representativa do fator.

# Interpretação dos fatores p115

REGRAS PRÁTICAS 3-4; 3-5; 3-6

2. Rotação de fatores: os eixos de referência dos fatores são rotacionados em torno da origem até que alguma outra posição seja alcançada;
3. Interpretação e reespecificação de fatores: o pesquisador avalia as cargas fatoriais (rotacionadas) para cada variável a fim de determinar o papel da mesma e sua contribuição na determinação da estrutura fatorial. A partir da avaliação

# Interpretação dos fatores p115

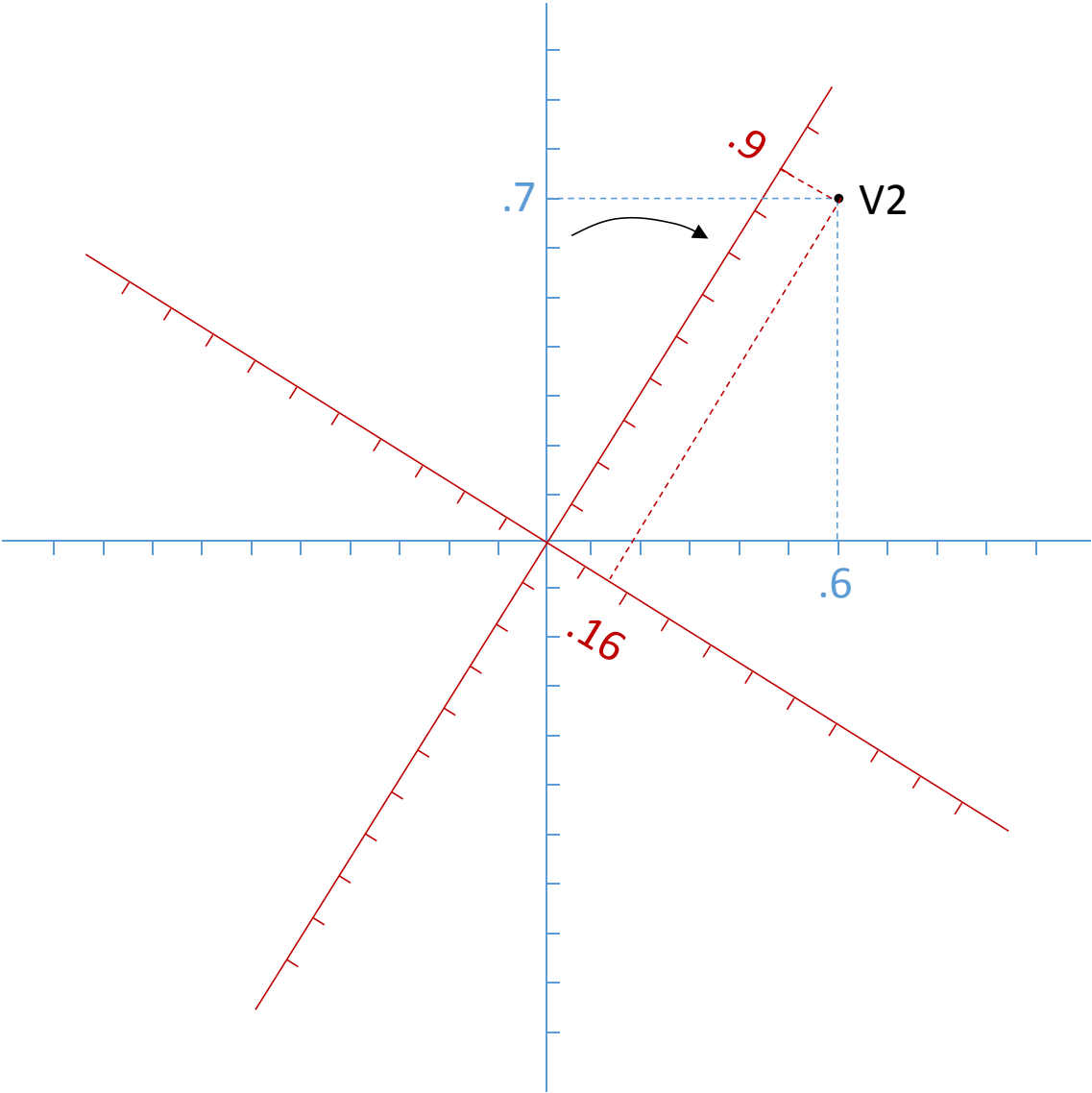
- Rotação de fatores p117
- simplificar as linhas (fatores) e colunas (cargas de uma variável ao longo dos fatores) da matriz fatorial para facilitar a interpretação.
- Rotação ortogonal; são os mais empregados e preferíveis quando o objetivo é redução de dados p120

quartmax: simplificar linhas da matriz fatorial (rotacionar o fator inicial de modo que uma variável tenha carga alta em um fator e cargas tão baixas quanto possível em todos os outros fatores) p119

varimax: simplificar colunas da matriz fatorial (maximiza a soma de variâncias de cargas exigidas da matriz fatorial) p119

equimax: “meio-termo” na simplificação de linhas e colunas p119

Variáveis	Cargas fatoriais não rotacionadas		Cargas fatoriais rotacionadas	
	Fator I	Fator II	Fator I	Fator II
V1	.50	.80	.03	.94
V2	.60	.70	.16	.90
V3	.90	-.25	.95	.24
V4	.80	-.30	.84	.15
V5	.60	-.50	.76	-.13



# Interpretação dos fatores p115

- Rotação oblíqua: São mais adequados ao objetivo de se obter diversos fatores ou construtos teoricamente relevantes, pois, na prática, poucos construtos no mundo são não-correlacionados p120
- permitem fatores correlacionados em vez de manterem independência entre os fatores rotacionados. O pesquisador deve ter o cuidado extra de validar fatores rotacionados obliquamente, uma vez que eles têm uma maneira adicional (não-ortogonalidade) de se tornarem específicos à amostra e não-generalizáveis, particularmente com pequenas amostras ou pequenas proporções de casos por variáveis p119

# Interpretação dos fatores p115

## REGRAS PRÁTICAS 3-4; 3-5; 3-6

- Apesar de cargas fatoriais de  $\pm 0,30$  a  $\pm 0,40$  serem minimamente aceitáveis, valores maiores que  $\pm 0,50$  são geralmente considerados necessários para significância prática p120;
- ser considerado significativo:
  - Uma carga menor com uma amostra maior ou um número maior de variáveis sob análise
  - Uma carga maior faz-se necessária com uma solução fatorial com um número maior de fatores, especialmente na avaliação de cargas em fatores posteriores
- O número de variáveis em análise também é importante na decisão sobre quais cargas são significantes. À medida que o número de variáveis em análise aumenta, o nível aceitável para considerar uma carga significativa diminui.



# Interpretação dos fatores p115

REGRAS PRÁTICAS 3-4; 3-5; 3-6

- Interpretação de uma matriz fatorial
  - interpretar as complexas relações representadas em uma matriz fatorial exige uma combinação da aplicação de critérios objetivos com julgamento gerencial (p121)
  - Simplificação do processo com 5 procedimentos (1. Examinar a matriz fatorial de cargas; 2. Identificar a carga significativa para cada variável; 3. Avaliação das comunalidades das variáveis; 4. Reespecificação do modelo fatorial (se necessário) e 5. Rotulação os fatores).

# Interpretação dos fatores p115

REGRAS PRÁTICAS 3-4; 3-5; 3-6

- Interpretação de uma matriz fatorial p121

1. Examinar a matriz fatorial de cargas

A matriz de cargas fatoriais contém a carga fatorial de cada variável em cada fator. Os fatores são dispostos como colunas representando as cargas de um único fator.

Se a rotação oblíqua foi usada, duas matrizes são fornecidas (matriz de padrão fatorial e matriz de estrutural fatorial)

# Interpretação dos fatores p115

REGRAS PRÁTICAS 3-4; 3-5; 3-6

- Interpretação de uma matriz fatorial p121

2. Identificar a carga significativa para cada variável

A interpretação deve começar com a primeira variável no primeiro fator e se mover horizontalmente da esquerda para a direita, procurando a carga mais alta para aquela variável em qualquer fator, identificando-a. O processo segue para as demais variáveis.

Variáveis com mais de uma carga significativa (conforme critérios) é chamada carga cruzada. Se uma variável persiste em ter cargas cruzadas, ela se torna candidata à eliminação.

# Interpretação dos fatores p115

REGRAS PRÁTICAS 3-4; 3-5; 3-6

- Interpretação de uma matriz fatorial p121

## 3. Avaliação das comunalidades das variáveis:

identificar variáveis nas quais faltam pelo menos uma carga significativa OU examinar a comunalidade de cada variável, representando a quantia de variância explicada pela solução fatorial para cada variável.

A comunalidade de uma variável é a estimativa de sua variância compartilhada, ou em comum, entre as variáveis como representadas pelos fatores obtidos p112.

# Interpretação dos fatores p115

REGRAS PRÁTICAS 3-4; 3-5; 3-6

- Interpretação de uma matriz fatorial p121
  4. Reespecificação do modelo fatorial (se necessário) problemas:
    - uma variável não tem cargas significantes;
    - mesmo com uma carga significativa, a comunalidade de uma variável é considerada muito baixa (melhor se  $>.5$ );
    - uma variável tem uma carga cruzada.

# Interpretação dos fatores p115

REGRAS PRÁTICAS 3-4; 3-5; 3-6

- Interpretação de uma matriz fatorial p121
  4. Reespecificação do modelo fatorial (se necessário) soluções:
    - Ignorar as variáveis problemáticas (redução de dados) p122;
    - Avaliar cada uma daquelas variáveis para possível eliminação p122;
    - Empregar um método alternativo de rotação p122;
    - Diminuir/aumentar o número de fatores p122;
    - Modificar o tipo de modelo fatorial usado p122.

# Interpretação dos fatores p115

REGRAS PRÁTICAS 3-4; 3-5; 3-6

- Interpretação de uma matriz fatorial

## 5. Rotulação os fatores

Designar algum significado para o padrão de cargas fatoriais. As variáveis com cargas mais altas são consideradas mais importantes e têm maior influência sobre o nome ou rótulo selecionado para representar um fator.

# Validação p123

- Avaliação do grau de generalidade dos resultados para a população e da influência potencial de casos ou respondentes individuais sobre os resultados gerais e é especialmente relevante nos métodos de interdependência, pois eles descrevem uma estrutura de dados que também deve ser representativa da população.
- Métodos:
  - Uso de uma perspectiva confirmatória;
  - Avaliação da estabilidade da estrutura fatorial;
  - Detecção de observações influentes.



# Usos adicionais dos resultados p124

- O pesquisador quer:

Identificar combinações lógicas de variáveis e entender melhor as interrelações entre variáveis (resumo dos dados);

ou

Identificar variáveis apropriadas para a aplicação subsequente em outras técnicas estatísticas (redução de dados);

As duas opções incluem:

# Usos adicionais dos resultados p124

- Selecionar a variável com a maior carga fatorial como uma representativa substituta para uma dimensão fatorial particular;
- Substituir o conjunto original de variáveis por um conjunto menor e inteiramente novo, criado a partir de escalas múltiplas ou escores fatoriais.

# Usos adicionais dos resultados p124

## **Seleção de variáveis de substituição**

- Selecionar a variável com a maior carga fatorial como uma representativa substituta para uma dimensão fatorial particular;
- Decisão pode ser baseada *a priori*.

# Usos adicionais dos resultados p124

## **Computação de escores fatoriais**

- medida composta de cada fator computadas para cada indivíduo;
- o escore fatorial representa o grau em que cada indivíduo tem escore elevado no grupo de itens que têm cargas elevadas em um fator;
- é computado com base nas cargas fatoriais de todas as variáveis no fator;
-

# Usos adicionais dos resultados p124

## **Criação de escalas múltiplas**

- combinação de diversas variáveis individuais em uma única medida composta;
- todas as variáveis com cargas elevadas em um fator são combinadas, e o escore médio é usado como uma variável de substituição;
- Supera o erro de medida;
- Representa múltiplos aspectos de um conceito com uma medida única;
- é calculada combinando-se apenas variáveis selecionadas.