



UNIVERSIDADE DE PASSO FUNDO
FACULDADE DE CIÊNCIAS ECONÔMICAS,
ADMINISTRATIVAS E CONTÁBEIS
CENTRO DE PESQUISA E EXTENSÃO DA FEAC
(www.upf.br/cepeac)

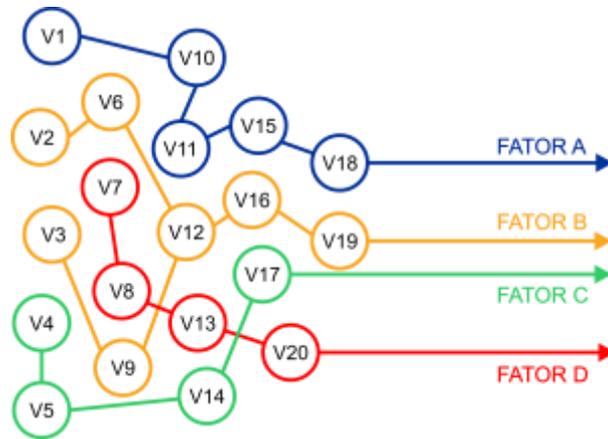
Texto para discussão

Texto para discussão Nº 02/2019

APOSTILA ANÁLISE FATORIAL

André da Silva Pereira
Berenice Paludo
Manoel Vieira
Rodolfo Henrique Cerbaro

UNIVERSIDADE DE PASSO FUNDO



APOSTILA ANÁLISE FATORIAL

André da Silva Pereira

Berenice Paludo

Manoel Vieira

Rodolfo Henrique Cerbaro

Passo Fundo,

2019

SUMÁRIO

TEORIA.....	3
<i>INTRODUÇÃO: O QUE É ANÁLISE FATORIAL?</i>	3
2. QUAL O PROCESSO DE UMA ANÁLISE FATORIAL?.....	5
3. QUAL É A UNIDADE DE ANÁLISE?.....	7
4. OBTER RESUMO VS REDUZIR DADOS.....	7
5. PLANEJAMENTO DE UMA ANÁLISE FATORIAL.....	8
6. SUPOSIÇÕES NA ANÁLISE FATORIAL.....	11
7. DETERMINAÇÃO DE FATORES E AVALIAÇÃO DO AJUSTE GERAL.....	13
8. ROTAÇÃO DOS FATORES.....	17
9. INTERPRETAÇÃO DA MATRIZ FATORIAL ROTACIONADA.....	19
10. VALIDAÇÃO DA ANÁLISE FATORIAL.....	21
Prática.....	25
Exercício com SPSS.....	25
Passo 1:.....	25
Passo 2:.....	26
Passo 3:.....	26
Passo 4:.....	27
Passo 5:.....	28
Passo 6:.....	28
Passo 7:.....	29
Interpretação dos Resultados.....	30
REFERÊNCIAS.....	36
ANEXOS.....	37

INTRODUÇÃO: O QUE É ANÁLISE FATORIAL?

Método estatístico usado para descrever variabilidade entre variáveis correlacionadas, a Análise Fatorial teve seus primeiros trabalhos realizados por Charles Spearman e Karl Pearson, em 1904. Nota-se que “no modelo da análise fatorial, há muitas variáveis observadas cujo objetivo é gerar fatores subjacentes não observados” (KING, 2001). Como precursor, Spearman testou a hipótese de que diferentes testes de habilidade mental, em matemática, verbais e de raciocínio lógico, poderiam ser explicados por um fator comum de inteligência a qual ele denominou “g”, o que é controverso, assim como a análise fatorial em si não é um tema livre de discussões acaloradas (BARTHOLOMEW, 1984; CUDECK; MACCALLUM, 2007). Esta apostila buscou tornar o tema mais acessível, e o importante é que se consiga interpretar os resultados gerados pela análise.

É notável que antigamente precisava semanas de trabalho para se fazer uma análise fatorial, porém hoje, com a computação moderna, a análise se tornou bem mais simples, sendo que pacotes estatísticos como Excel e IBM SPSS realizam a parte mecânica da análise fatorial quase sem esforço do pesquisador. A tarefa que cabe ao pesquisador, portanto, é a compreensão do que está sendo feito e a adequada interpretação, sendo o intuito desta apostila demonstrar através de uma breve estilização da teoria e de um exemplo prático como se dá o processo de realizar uma análise fatorial.

Em problemas que envolvam um grande número de variáveis observáveis são necessários muitos parâmetros para descrever toda a informação. Sempre que existam variáveis fortemente correlacionadas é concebível agrupá-las num grupo de modo que variáveis que estejam em grupos distintos apresentem fraca correlação (CARVALHO, 2013).

O objetivo da Análise Fatorial é gerar fatores subjacentes não observados, onde há muitas variáveis observadas. A principal função é diminuir uma grande quantidade de variáveis observadas em um número menor de fatores. Devido ao advento da tecnologia que gera uma quantidade gigantesca de informações, o interesse pelo uso da Análise Fatorial aumentou e ela tem sido usada nas mais diferentes áreas do conhecimento, e especialmente em ciências sociais. Um dos maiores desafios enfrentados pelos pesquisadores é operacionalizar conceitos abstratos em variáveis empiricamente observáveis. A análise fatorial tem se caracterizado como um dos métodos psicométricos mais poderosos para reduzir a complexidade de uma grande quantidade de variáveis a uma estrutura relativamente simples,

consistindo de um número menor de fatores. Os fatores são combinações lineares de variáveis observadas.

A aplicação da análise fatorial requer que as variáveis X_1, \dots, X_p sejam correlacionadas, pois caso contrário, tal aplicação não faz sentido. Se as variáveis não forem correlacionadas duas a duas, então a matriz de correlações é a matriz identidade (CARVALHO, 2013)

Apesar da popularidade da análise fatorial, tanto a complexidade matemática dos procedimentos analíticos quanto a flexibilidade da técnica contribuíram para que esta, hoje em dia, seja uma das técnicas que muitos profissionais e estudantes tem dificuldade para entender, sendo este o objetivo dessa apostila, facilitar o entendimento sobre Análise Fatorial para qualquer pessoa que esteja interessada, independente da formação. A Análise será exemplificada através do software: SPSS.

Análise fatorial é uma técnica de interdependência, cujo propósito principal é definir a estrutura inerente entre as variáveis na análise. As técnicas podem ser univariadas que se limita a uma única variável, como multivariadas, tendo dezenas, centenas e até milhares de variáveis. E como descrever e representar todas essas variáveis? Tornando as variáveis correlacionadas, o pesquisador precisa de passagem para gerenciar essas variáveis – agrupando variáveis altamente correlacionadas, rotulando ou nomeando os grupos, e quiçá até mesmo indicar uma nova medida composta que possa representar cada grupo de variáveis (HAIR et al., 2009).

Aí que entra a análise fatorial. Ela fornece as ferramentas para analisar a estrutura das inter-relações (correlações) em um grande número de variáveis definindo conjuntos de variáveis que são fortemente inter-relacionadas, conhecidos como fatores. Esses grupos de variáveis (fatores), que são por definição altamente inter-correlacionadas, são considerados como representantes de dimensões dentro dos dados.

A finalidade geral de técnicas de análise fatorial é encontrar um modo de sintetizar a informação contida em diversas variáveis originais em um conjunto menor de novas dimensões compostas ou variáveis estatísticas (fatores) com uma avaria mínima de informação – ou seja, buscar e definir os construtos fundamentais ou dimensões assumidas como essenciais às variáveis originais (HAIR et al., 2009).

Ao atingir seu propósito, a análise fatorial é adequada com quatro questões: (1) especificação da unidade de análise; (2) obtenção do resumo de dados e/ou redução dos

mesmos; (3) seleção de variáveis e (4) uso de resultados da análise fatorial com outras técnicas multivariadas.

Já para Field, 2009, nas ciências sociais, geralmente tentamos medir coisas que não podem ser medidas diretamente (as chamadas variáveis latentes ou construtos). Por exemplo, estudantes de administração (ou mesmo psicólogos) podem estar interessados em medir a exaustão (burnout), que é quando alguém que tem trabalhado muito num projeto (um livro, por exemplo) por um longo período de tempo subitamente se encontra destituído de motivação e inspiração e quer bater com a cabeça várias vezes no computador. Você não pode medir a exaustão diretamente: ela tem muitas facetas. Entretanto, pode medir vários aspectos da exaustão: a motivação, o nível de estresse, se a pessoa tem alguma ideia nova e assim por diante. Feito isso, seria útil saber se essas medidas realmente refletem uma única variável. Análise de fatores é uma técnica que tem três usos principais: (1) entender a estrutura de um conjunto de variáveis (por exemplo, pioneiros da inteligência como Spearman e Thurstone utilizaram a análise de fatores para tentar entender a estrutura da variável latente “inteligência”); (2) construir um questionário para medir uma variável subjacente (por exemplo, você pode fazer um questionário para medir a exaustão); e (3) reduzir um conjunto de dados a um tamanho mais manejável enquanto se retém o máximo da informação original possível.

1. QUAL O PROCESSO DE UMA ANÁLISE FATORIAL?

O processo inicial é estilizado por Hair (2009):

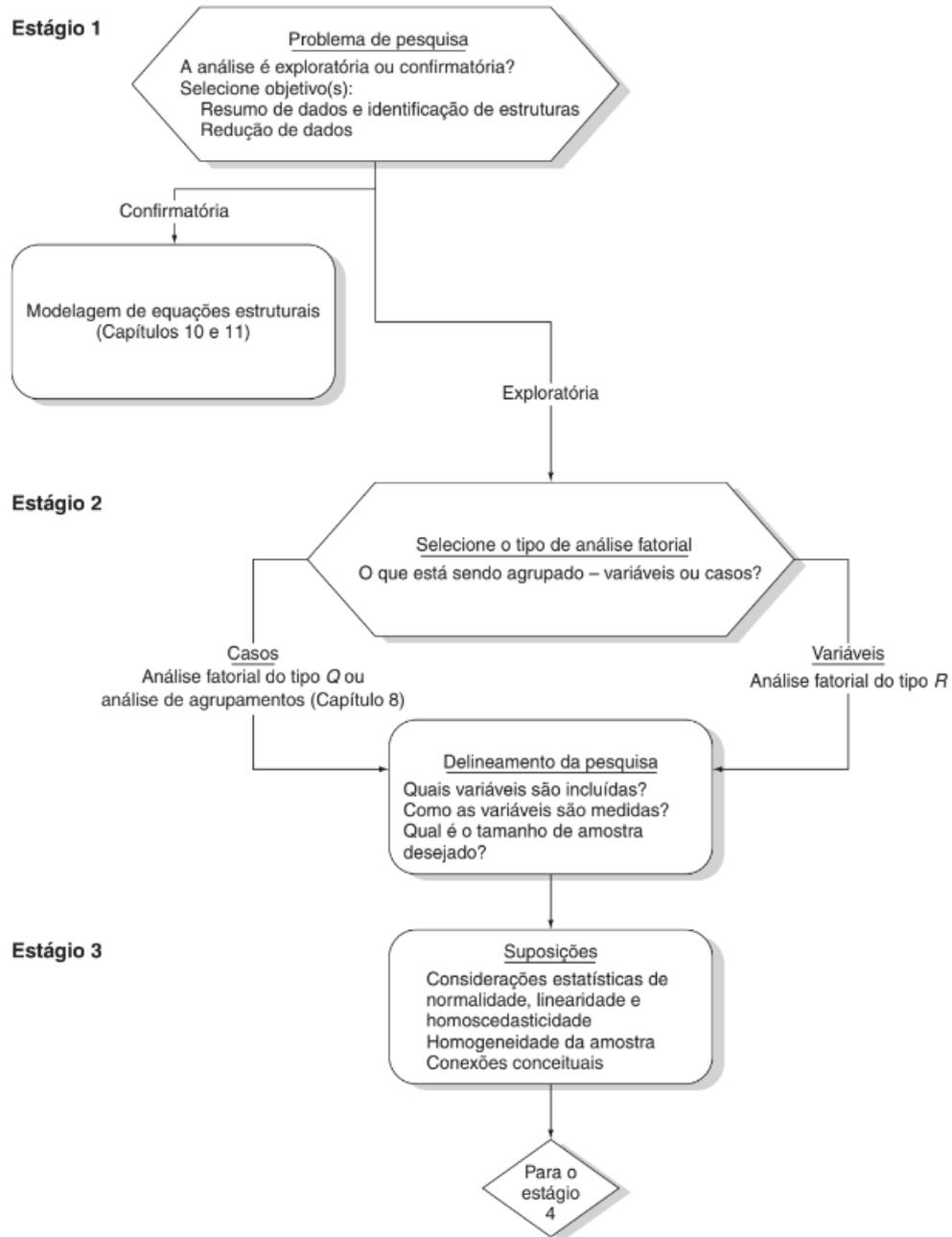


Figura 1: Processo Inicial Análise Fatorial

Fonte: Hair et al. (2009)

Nota-se que a utilização de análise fatorial exploratória ou confirmatória vai depender do problema de pesquisa. Esta apostila vai se focar na análise exploratória. O delineamento da pesquisa necessita se recorrer à teoria; uma vez estipulado, há necessidade de checar as suposições, tarefa muito importante que antecede a análise propriamente dita, porque de nada adianta utilizar o programa estatístico se os dados que formam a base da análise não satisfizerem as condições para análise.

2. QUAL É A UNIDADE DE ANÁLISE?

A especificação da unidade de análise se dá por R ou Q:

- ✚ Se o objetivo da pesquisa fosse resumir as características, a análise fatorial seria aplicada a uma matriz de correlação das variáveis. Esse é o tipo mais comum de análise fatorial e é chamado de **análise fatorial R**, que analisa um conjunto de variáveis para identificar as dimensões latentes (que não são fáceis de observar).
- ✚ A análise fatorial também pode ser aplicada a uma matriz de correlação dos respondentes individuais baseada nas características dos mesmos. Chamado de **análise fatorial Q**, este método combina ou condensa grandes números de pessoas em diferentes grupos de uma população maior. A análise fatorial Q não é utilizada muito frequentemente por causa das dificuldades computacionais. Em vez disso, a maioria dos pesquisadores utiliza algum tipo de análise de agrupamentos para agrupar respondentes individuais.

3. OBTER RESUMO VS REDUZIR DADOS

A análise fatorial equipa o pesquisador com duas saídas distintas, mas relacionadas: resumo de dados e 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; do outro lado, **redução de dados** estende esse processo derivando um valor empírico (score fatorial) para cada dimensão (fator) e então **substituindo** o valor original por esse novo valor (HAIR, 2009).

- ✚ Em ambos os usos da análise fatorial, o pesquisador implicitamente especifica as dimensões potenciais que podem ser identificadas por meio do caráter e da natureza das variáveis submetidas à análise fatorial. Por exemplo, ao avaliar as dimensões de imagem da loja, se nenhuma questão sobre pessoal da loja for incluída, a análise fatorial não será capaz de identificar tal dimensão.

- ✚ O pesquisador deve também lembrar que análise fatorial sempre produzirá fatores. Assim, a análise fatorial é sempre um candidato potencial para o fenômeno “lixo dentro, lixo fora”. Se o pesquisador incluir indiscriminadamente um grande número de variáveis e esperar que a análise fatorial “arrume as coisas”, então torna-se elevada a possibilidade de resultados pobres. A qualidade e o significado dos fatores obtidos refletem as bases conceituais das variáveis incluídas na análise.

4. PLANEJAMENTO DE UMA ANÁLISE FATORIAL

Para planejar uma análise fatorial abrange três decisões básicas: (1) cálculo dos dados de entrada (uma matriz de correlação) para atender os objetivos especificados de agrupamento de variáveis ou respondentes; (2) planejamento do estudo em termos de número de variáveis, propriedades de medida das variáveis e tipos de variáveis admissíveis; e (3) o tamanho necessário para a amostra em termos absolutos e como função do número de variáveis na análise.

No planejamento de uma análise fatorial a primeira decisão enfoca o cálculo dos dados de entrada para a análise. Com a análise fatorial do tipo R, o pesquisador usaria uma matriz tradicional de correlação (correlações entre variáveis) como entrada. Mas o pesquisador poderia também escolher a opção de obter a matriz de correlação a partir das correlações entre os respondentes individuais. Nessa análise fatorial de tipo Q, os resultados seriam uma matriz fatorial que identificaria indivíduos semelhantes. (HAIR, 2009)

DICA: O QUE É FATOR R?

Fator R: Se medirmos diversas variáveis, ou fizermos várias perguntas para alguém sobre eles mesmos, a correlação entre cada par de variáveis (ou perguntas) pode ser organizada em uma matriz-R. Uma matriz-R é uma matriz de correlações: uma tabela de coeficientes de correlações entre variáveis. Um exemplo é dado na figura a seguir.

	Habilidades					
	Conversa 1 (Talk 1)	Sociais (Social Skills)	Interessante (Interest)	Conversa 2 (Talk 2)	Egoísta (Selfish)	Mentiroso (Liar)
Conversa 1 (Talk 1)	1,000					
Habilidades Sociais (Social Skills)	0,772	1,000				
Interesse (Interest)	0,646	0,879	1,000			
Conversa 2 (Talk 2)	0,074	-0,120	0,054	1,000		
Egoísta (Selfish)	-0,131	0,031	-0,101	0,441	1,000	
Mentiroso (Liar)	0,068	0,012	0,110	0,361	0,277	1,000

Tabela 1: Uma matriz-R

Fonte: HAIR et al., 2009

Exemplo Gráfico Análise Fatorial do Tipo Q:

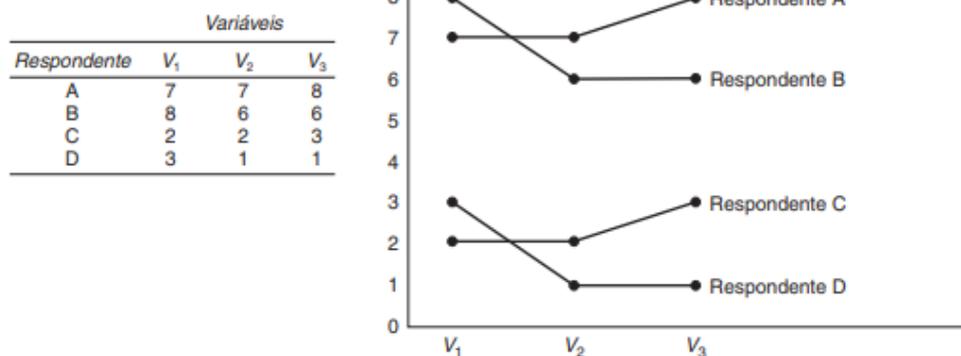


Figura 2: Comparações de perfis de escore para análise fatorial do tipo Q e análise de agrupamentos.

Fonte: HAIR et al., 2009

Duas questões específicas devem ser respondidas neste ponto:

- (1) Quais tipos de variáveis podem ser usados em análise fatorial?
- (2) Quantas variáveis devem ser incluídas?

Em termos dos tipos de variáveis incluídas, **o requisito principal é que um valor de correlação possa ser calculado entre todas as variáveis**. Variáveis métricas são facilmente medidas por vários tipos de correlações. Variáveis não-métricas, contudo, são mais problemáticas por não poderem usar os mesmos tipos de medida de correlação empregados em variáveis métricas. Apesar de alguns métodos especializados calcularem correlações entre variáveis não-métricas, a abordagem mais prudente é evitá-las. Se uma variável não-métrica deve ser incluída, um método é definir variáveis dicotômicas (codificadas como 0 e 1) para

representarem categorias de variáveis não-métricas. Se todas as variáveis são dicotômicas, então formas especializadas de análise fatorial, como análise fatorial booleana, são mais adequadas.

O pesquisador também deve **tentar tornar mínimo o número de variáveis incluídas**, mas sustentar um número razoável de variáveis por fator. Se um estudo está sendo planejado para avaliar uma estrutura proposta, o pesquisador deve certificar-se de incluir diversas variáveis (cinco ou mais) que possam representar cada fator escolhido. A eficácia da análise fatorial reside em encontrar padrões entre grupos de variáveis, e é de escasso uso na identificação de fatores compostos por uma única variável. Enfim, quando se delineia um estudo para ser ponderado por fatores, o pesquisador deve, se possível, identificar diferentes variáveis chave (algumas vezes chamadas de indicadores-chave ou variáveis de marcação) que profundamente reflitam os fatores latentes que foram previstos hipoteticamente. Isso valerá na validação dos fatores determinados e na avaliação da significância prática dos resultados. (HAIR et al., 2009)

Referindo-se à questão do **tamanho da amostra**, o pesquisador dificilmente realiza uma análise fatorial com uma amostra com menos de 50 observações, e de preferência o tamanho da amostra deve ser maior ou igual a 100. Como regra geral, o mínimo é ter pelo menos **cinco vezes mais observações do que o número de variáveis a serem analisadas**, e o tamanho mais aceitável teria uma proporção de dez para um. Alguns pesquisadores chegam a propor um mínimo de 20 casos para cada variável. Deve-se lembrar, contudo, que com 30 variáveis, por exemplo, há 435 correlações a serem calculadas na análise fatorial. Em um nível de significância de 0,05, talvez até mesmo 20 dessas correlações fossem consideradas significantes e aparecessem na análise fatorial somente por sorte. O pesquisador sempre deve tentar obter a maior proporção de casos por variável para minimizar as chances de super ajustar os dados. (HAIR et al., 2009)

Os coeficientes de correlação flutuam de amostra para amostra, mais em pequenas amostras do que em grandes. Portanto, a confiabilidade da análise dos fatores depende também do tamanho da amostra. Muito já foi escrito sobre o tamanho da amostra necessário para a análise de fatores, resultando em diversas “regras práticas”. A regra comum é sugerir que o pesquisador tenha pelo menos entre 10 e 15 participantes por variável (FIELD, 2009).

6 SUPOSIÇÕES NA ANÁLISE FATORIAL

As conjecturas críticas na análise fatorial são mais conceituais do que estatísticas. O pesquisador está sempre preocupado em atender a exigência estatística para qualquer técnica multivariada, mas em análise fatorial 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. (HAIR et al., 2009)

As suposições citadas no modelo de Hair (2009) são: normalidade, ou seja, segue uma curva de Bell, que é uma distribuição bastante comum e que pode ser checada tanto no Excel quanto no SPSS, linearidade, ou seja, adequa-se a uma reta; homocedasticidade, que é outro pressuposto do modelo linear e significa que a variância ao redor da linha de regressão é a mesma para todos os valores de variáveis dependentes e independentes e homogeneidade, ou seja, que as propriedades de uma parte do total dos dados são as mesmas que as outras partes. Uma vez testadas as suposições, sendo que tanto no SPSS e Excel há testes para fazê-los, passa-se aos próximos estágios de análise.

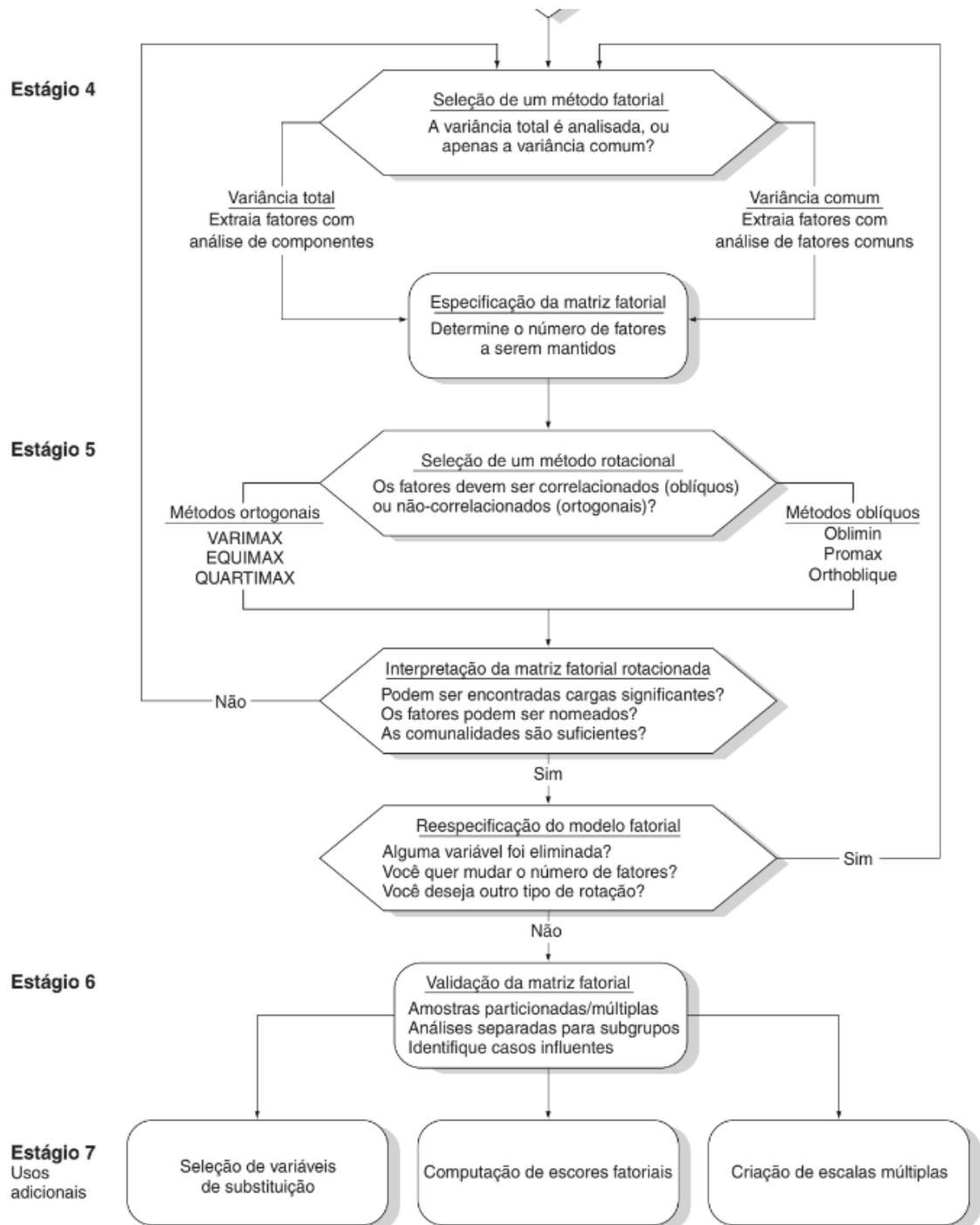


Figura 3: Estágios 4-7 no diagrama de decisão da análise fatorial.

Fonte: HAIR et al., 2009

7 DETERMINAÇÃO DE FATORES E AVALIAÇÃO DO AJUSTE GERAL

Uma vez que as variáveis sejam especificadas e a matriz de correlação seja preparada, o pesquisador está pronto para aplicar a análise fatorial para identificar a estrutura latente de relações. Nisso, as decisões devem ser tomadas com relação (1) ao método de extração dos fatores (análise de fatores comuns versus análise de componentes) e (2) ao número de fatores selecionados para explicar a estrutura latente dos dados.

Seleção do método de extração de fatores. O pesquisador pode propor a partir de dois métodos similares, ainda que singulares, para determinar os fatores que representem a estrutura das variáveis na análise. Essa determinação sobre o método a ser empregado deve combinar os objetivos da análise fatorial com o conhecimento sobre algumas características básicas das relações entre variáveis (HAIR et al., 2009).

Partição da variância de uma variável. Para optar entre os dois métodos de extração de fatores, o pesquisador deve primeiro ter certa compreensão da variância para uma variável e como ela é dividida ou particionada. Lembre-se que variância é um valor (i.e., o quadrado do desvio padrão) que representa a quantia total de dispersão de valores para uma única variável em torno de sua média. Em análise fatorial, agrupamos variáveis por suas correlações, de modo que variáveis em um grupo (fator) têm elevadas correlações umas com as outras. Assim, para os propósitos da análise fatorial, é importante entender o quanto da variância de uma variável é compartilhado com outras variáveis naquele fator versus o que não pode ser compartilhado (p.ex., inexplicado). A variância total de qualquer variável pode ser dividida (particionada) em três tipos de variância:

1. Variância comum é definida como aquela variância em uma variável que é compartilhada com todas as outras variáveis na análise.
2. Variância específica (também conhecida como variância única) é aquela associada com apenas uma variável específica.
3. Variância de erro é também variância que não pode ser explicada por correlações com outras variáveis, mas resulta da não confiabilidade no processo de coleta de dados, de erro de medida ou de componente aleatório no fenômeno medido.

Análise de fatores comuns versus análise de componentes. Com uma compreensão básica sobre como a variância pode ser particionada, o pesquisador está pronto para abordar as diferenças entre os dois métodos, conhecidos como análise de fatores comuns e análise de componentes. A escolha de um método em vez do outro é baseada em dois critérios: (1) os objetivos da análise fatorial e (2) o montante de conhecimento prévio sobre a variância nas variáveis. A análise de componentes é usada quando o objetivo é resumir a maior parte da informação original (variância) a um número mínimo de fatores para fins de previsão. Em contraste, análise de fatores comuns é usada prioritariamente para identificar fatores ou dimensões latentes que refletem o que as variáveis têm em comum. (HAIR et al., 2009)

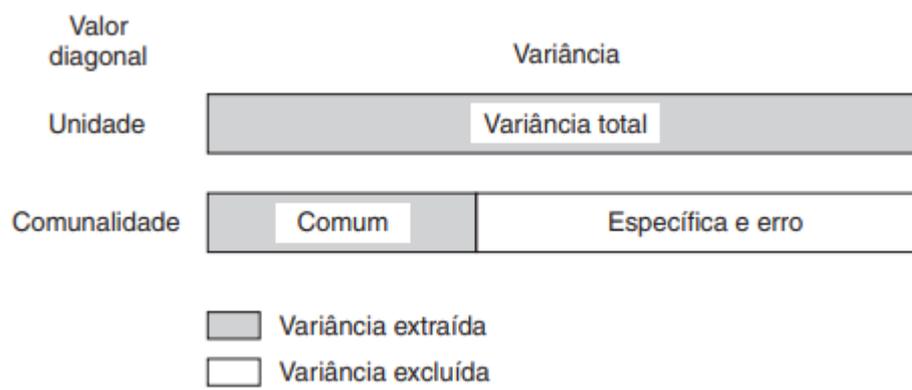


Figura 4: Tipos de variância considerados na matriz fatorial.

Fonte: HAIR et al., 2009

Crítérios para o número de fatores a extrair. Como decidimos sobre o número de fatores a serem extraídos? Ambos os métodos de análise fatorial estão interessados na melhor combinação linear de variáveis – melhor no sentido de que a combinação particular de variáveis originais explica a maior parte da variância nos dados como um todo comparada a qualquer outra combinação linear de variáveis. Logo, o primeiro fator pode ser visto como o melhor resumo de relações lineares exibidas nos dados. O segundo fator é definido como a segunda melhor combinação linear das variáveis, sujeita à restrição de que é ortogonal ao primeiro fator. Uma base quantitativa exata para decidir o número de fatores a extrair ainda não foi desenvolvida. No entanto, os seguintes critérios de parada têm sido utilizados (HAIR et al., 2009):

- (a) **Crítério da raiz latente.** A técnica mais comumente usada é o critério da raiz latente. Esta técnica é simples de aplicar na análise de componentes, bem como na análise de fatores comuns. O raciocínio para o critério da raiz latente é que qualquer fator

individual deve explicar a variância de pelo menos uma variável se o mesmo há de ser mantido para interpretação.

- (b) **Critério a priori.** O critério a priori é um critério simples, ainda que razoável sob certas circunstâncias. Quando aplicado, o pesquisador já sabe quantos fatores extrair antes de empreender a análise fatorial.
- (c) **Critério de percentagem de variância.** O critério de percentagem de variância é uma abordagem baseada na conquista de um percentual cumulativo especificado da variância total extraída por fatores sucessivos.
- (d) **Critério do teste *scree*.** Lembre-se que, no modelo fatorial de análise de componentes, os últimos fatores extraídos contêm tanto a variância comum quanto a única. Apesar de todos os fatores conterem pelo menos alguma variância única, a proporção de variância única é substancialmente maior nos últimos fatores. O teste *scree* é usado para 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.

Heterogeneidade dos respondentes. A variância compartilhada entre variáveis é a base para ambos os modelos fatoriais, de fator comum e de componentes. Uma suposição inerente é que a variância compartilhada se estende ao longo de toda a amostra.

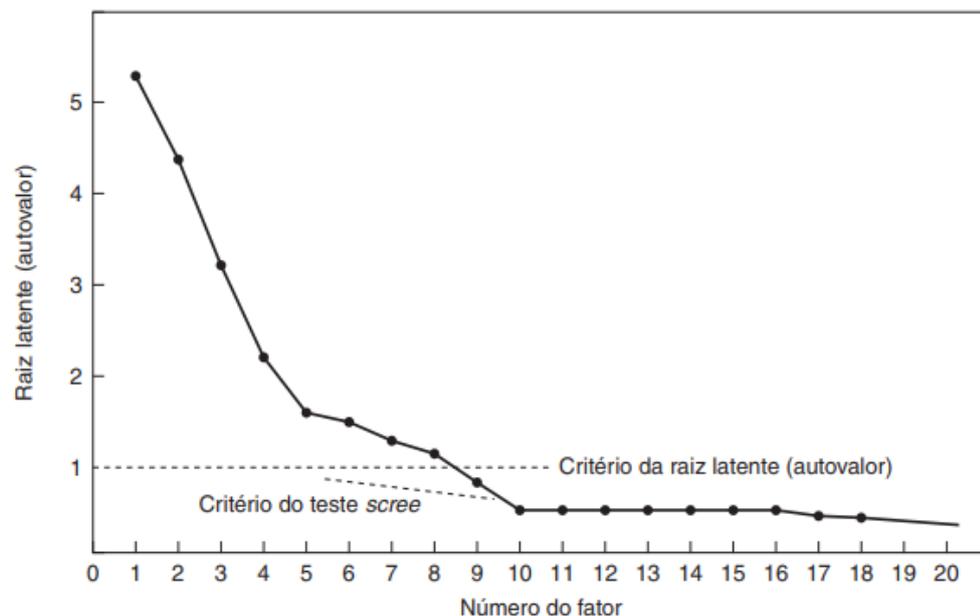


Figura 5: Gráfico de autovalor para o critério de teste *scree*.

Fonte: HAIR et al., 2009

Interpretação dos fatores. Apesar de não existirem processos ou orientações inequívocas para determinar a interpretação de fatores, o pesquisador com forte fundamentação conceitual para a estrutura antecipada e sua justificativa tem a maior chance de sucesso. Não podemos estabelecer de maneira suficientemente impactante a importância de uma forte fundamentação conceitual, seja ela vinda de pesquisa anterior, paradigmas teóricos ou princípios comumente aceitos. O pesquisador deve repetidamente fazer julgamentos subjetivos em decisões, como o número de fatores, quais são as relações suficientes para garantir variáveis que discriminam grupos, e como podem ser identificados esses grupos. Como pode atestar o pesquisador experiente, praticamente qualquer coisa pode ser descoberta se houver empenho suficientemente insistente (HAIR, 2009).

Os três processos de interpretação fatorial. A interpretação fatorial é circular por natureza. O pesquisador primeiramente avalia os resultados iniciais, em seguida faz alguns julgamentos vendo e refinando tais resultados, com a evidente possibilidade de que a análise seja reespecificada, exigindo-se uma volta ao passo avaliativo.

- (a) **Estimativa da matriz fatorial.** Primeiro, a matriz fatorial inicial não-rotacionada é computada, contendo as cargas fatoriais para cada variável sobre cada fator. Cargas fatoriais são a correlação de cada variável com o fator. Cargas indicam o grau de correspondência entre a variável e o fator, com cargas maiores tornando a variável representativa do fator. (HAIR et al., 2009)
- (b) **Rotação de fatores.** Soluções fatoriais não-rotacionadas atingem a meta de redução de dados, mas o pesquisador deve perguntar se a solução fatorial não-rotacionada (que preenche as exigências matemáticas desejáveis) fornecerá informação que oferece interpretação a mais adequada das variáveis sob exame.
- (c) **Interpretação e reespecificação de fatores.** Como um processo final, 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.

8 ROTAÇÃO DOS FATORES

Talvez a ferramenta mais importante na interpretação de fatores seja a rotação fatorial. O termo rotação significa exatamente o que sugere. Especificamente, os eixos de referência dos fatores são rotacionados em torno da origem até que alguma outra posição seja alcançada.

Métodos rotacionais ortogonais. Na prática, o objetivo de todos os métodos de rotação é simplificar as linhas e colunas da matriz fatorial para facilitar a interpretação. Três abordagens ortogonais principais foram desenvolvidas:

1. A meta final de uma rotação **QUARTIMAX** é simplificar as linhas de uma matriz fatorial; ou seja, QUARTIMAX se concentra em 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. Nessas rotações, muitas variáveis podem ter carga alta no mesmo fator, pois a técnica se concentra em simplificar as linhas. (HAIR et al.,2009)

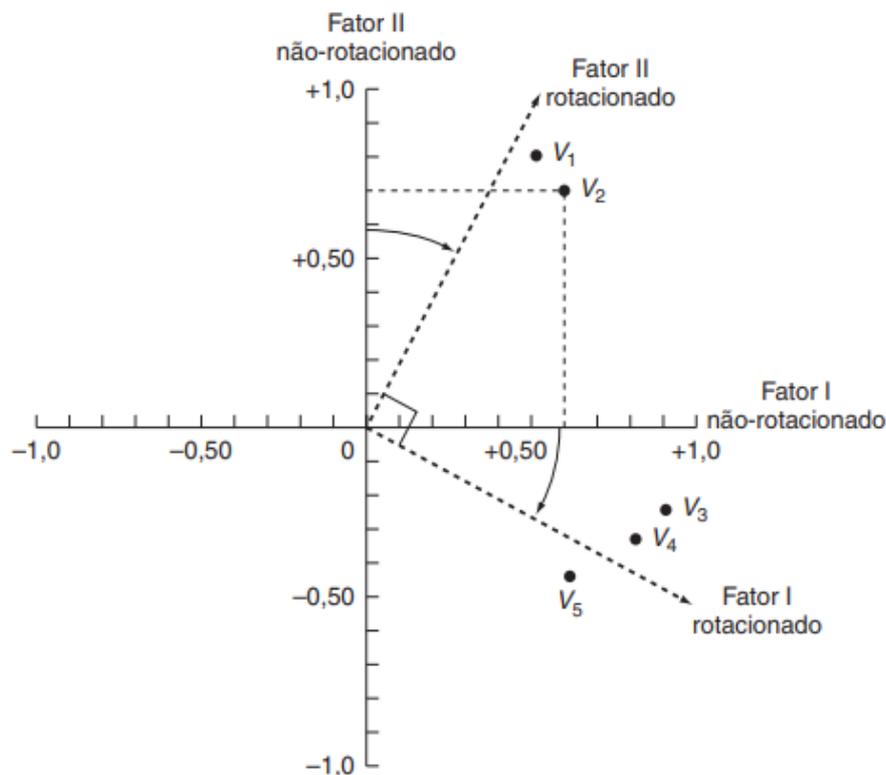


Figura 6: Rotação fatorial ortogonal.

Fonte: (HAIR et al., 2009)

2. Diferentemente de QUARTIMAX, o critério **VARIMAX** se concentra na simplificação das colunas da matriz fatorial. Com a abordagem rotacional VARIMAX, a simplificação máxima possível é conseguida se houver apenas 1s e 0s em uma coluna. Ou seja, o método VARIMAX maximiza a soma de variâncias de cargas exigidas da matriz fatorial. Lembre-se que, nas abordagens QUARTIMAX, muitas variáveis podem ter cargas altas ou próximas de altas no mesmo fator, pois a técnica se concentra em simplificar as linhas. Com a abordagem rotacional VARIMAX, há uma tendência para algumas cargas altas (isto é, próximas de -1 ou $+1$) e algumas cargas próximas de 0 em cada coluna da matriz. (HAIR et al., 2009)

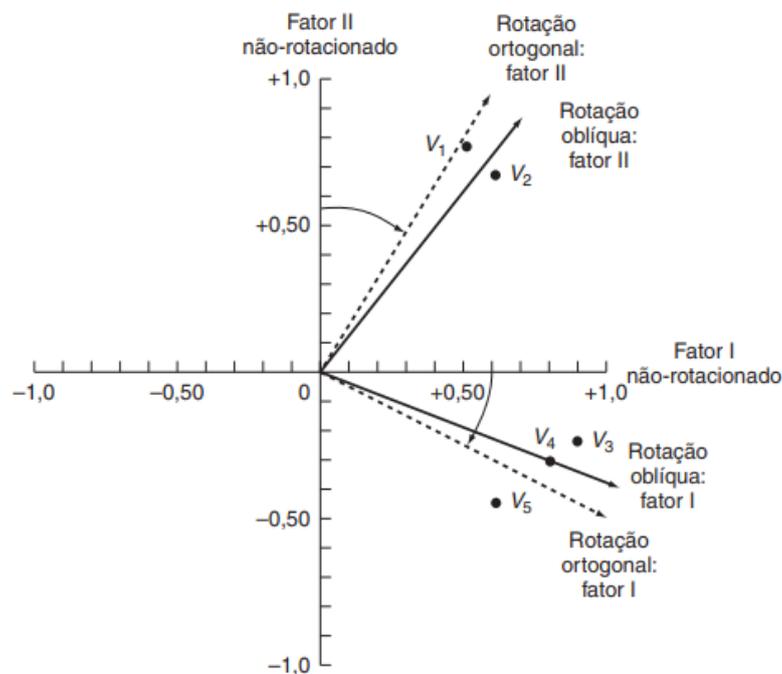


Figura 7: Rotação fatorial oblíqua.

Fonte: HAIR et al.,2009

3. O método **EQUIMAX** é uma espécie de meio-termo entre QUARTIMAX e VARIMAX. Em vez de se concentrar na simplificação de linhas ou de colunas, ele tenta atingir um pouco de cada. EQUIMAX não tem obtido ampla aceitação e é pouco usado.

Para análise, o SPSS tem três métodos de rotação ortogonal (varimax, quartimax e equamax) e dois métodos de rotação oblíqua (directoblmin e promax). Esses métodos diferem na forma de rodar os fatores e, portanto, a saída resultante dependerá do método selecionado. A rotação quartimax tenta maximizar a dispersão da carga dos fatores de uma variável por todos os fatores. Portanto, se torna mais fácil interpretar as variáveis. Entretanto,

isso geralmente resulta em muitas variáveis com altas cargas em um único fator. O varimax é o oposto disso, ele tenta maximizar a dispersão das cargas dentro dos fatores. Portanto, ele tenta agregar um menor número de variáveis sobre cada fator resultando em mais aglomerados de fatores interpretáveis. O equamax é um híbrido das outras duas abordagens e existem relatos de que ele apresenta um comportamento bastante errático

Métodos de rotação oblíqua. As rotações oblíquas são semelhantes às ortogonais, porém as oblíquas permitem fatores correlacionados em vez de manterem independência entre os fatores rotacionados. Porém, enquanto há várias escolhas entre abordagens ortogonais, há apenas escolhas limitadas na maioria dos pacotes estatísticos para rotações oblíquas. Por exemplo, SPSS fornece OBLIMIN; SAS tem PROMAX e ORTHOBLIQUE; e BMDP fornece DQUART, DOBLIMIN e ORTHOBLIQUE. Os objetivos de simplificação são comparáveis aos métodos ortogonais, com a característica extra de fatores correlacionados.

Seleção entre métodos rotacionais. Nenhuma regra específica foi desenvolvida para guiar o pesquisador na seleção de uma técnica rotacional ortogonal ou oblíqua em particular. Na maioria dos casos, o pesquisador simplesmente utiliza a técnica rotacional dada pelo programa de computador. A maioria dos programas tem como padrão de rotação o VARIMAX, mas todos os métodos rotacionais mais importantes estão amplamente disponíveis (HAIR et al., 2009).

9 INTERPRETAÇÃO DA MATRIZ FATORIAL ROTACIONADA

Julgamento da significância de cargas fatoriais. Ao interpretar fatores, é preciso tomar uma decisão sobre quais cargas fatoriais vale a pena considerar. A discussão a seguir detalha questões relativas à significância prática e estatística, bem como ao número de variáveis, que afetam a interpretação de cargas fatoriais.

Garantia de significância prática. A primeira orientação não é baseada em qualquer proposição matemática, mas se refere mais à significância prática ao fazer um exame preliminar da matriz fatorial em termos das cargas fatoriais.

Avaliação da significância estatística. Uma carga fatorial representa a correlação entre uma variável original e seu fator. Ao determinar um nível de significância para a interpretação de cargas, uma abordagem semelhante à determinação da significância estatística de coeficientes de correlação poderia ser usada.

Ajustes baseados no número de variáveis. Uma desvantagem das duas abordagens anteriores é que o número de variáveis analisadas e o fator específico em exame não são considerados.

Interpretação de uma matriz fatorial. O trabalho de interpretar uma matriz de cargas fatoriais para identificar a estrutura entre as variáveis pode parecer à primeira vista muito complicada. O pesquisador deve considerar todas as cargas fatoriais (lembre-se, cada variável tem uma carga sobre cada fator) para identificar as mais indicativas da estrutura latente. Mesmo uma análise relativamente simples de 15 variáveis sobre 4 fatores precisa de avaliação e interpretação de 60 cargas fatoriais. Usando os critérios para interpretação de cargas, o pesquisador descobre aquelas variáveis distintas para cada fator e procura uma correspondência com a fundamentação conceitual ou as expectativas administrativas depositadas na pesquisa para avaliar significância prática (HAIR et al., 2009). Passa-se a uma exemplificação por etapas da matriz rotacionada:

Etapa 1: Examine a matriz fatorial de cargas. A matriz de cargas fatoriais contém a carga fatorial de cada variável em cada fator. Elas podem ser cargas rotacionadas ou não rotacionadas, os fatores são dispostos como colunas; assim, cada coluna de números representa as cargas de um único fator. Se uma rotação oblíqua foi usada, duas matrizes fatoriais de cargas são fornecidas. A primeira é a matriz de padrão fatorial, a qual tem cargas que representam a contribuição única de cada variável ao fator. A segunda é a matriz de estrutura fatorial, a qual tem correlações simples entre variáveis e fatores, mas essas cargas contêm tanto a variância única entre variáveis e fatores quanto a correlação entre fatores.

Etapa 2: Identifique a(s) carga(s) significante(s) 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. Quando a maior carga (em valor absoluto) é identificada, deve ser sublinhada se for significativa como determinado pelos critérios anteriormente discutidos. A atenção agora se dirige para a segunda variável, e, novamente movendo-se horizontalmente da esquerda para a direita, procura-se a maior carga para aquela variável em qualquer fator, e a mesma deve ser sublinhada.

Etapa 3: Avalie as comunalidades das variáveis. Uma vez que todas as cargas significantes tenham sido identificadas, o pesquisador deve procurar por variáveis que não sejam adequadamente explicadas pela solução fatorial. Uma abordagem simples é identificar variáveis nas quais faltam pelo menos uma carga significativa. Outro método é examinar a

comunalidade de cada variável, representando a quantia de variância explicada pela solução fatorial para cada variável.

Etapa 4: Reespecifique o modelo fatorial se necessário. Uma vez que todas as cargas significantes tenham sido identificadas e as comunalidades, examinadas, o pesquisador pode encontrar diversos problemas: (a) uma variável não tem cargas significantes; (b) mesmo com uma carga significativa, a comunalidade de uma variável é considerada muito baixa, ou (c) uma variável tem uma carga cruzada.

Etapa 5: Rotule os fatores. Quando é obtida uma solução fatorial aceitável na qual todas as variáveis têm uma carga significativa em um fator, o pesquisador tenta designar algum significado para o padrão de cargas fatoriais (HAIR et al., 2009).

10 VALIDAÇÃO DA ANÁLISE FATORIAL

O sexto estágio envolve a avaliação do grau de generalidade dos resultados para a população e da influência potencial de casos ou respondentes individuais sobre os resultados gerais. A questão da generalidade é crítica para todo método multivariado, mas é especialmente relevante nos métodos de interdependência, pois eles descrevem uma estrutura de dados que também deve ser representativa da população. No processo de validação, o pesquisador deve abordar várias questões na área de delineamento de pesquisa e características de dados (HAIR et al., 2009)

Uso de uma perspectiva confirmatória. O método mais direto para validar os resultados é partir para uma perspectiva confirmatória e avaliar a repetitividade dos resultados, seja com uma amostra particionada no conjunto de dados originais, seja com uma amostra separada. A comparação de dois ou mais resultados de um modelo fatorial sempre é problemática. No entanto, existem várias opções para realizar uma comparação objetiva (HAIR et al., 2009).

Avaliação da estabilidade da estrutura fatorial. Outro aspecto da generalidade é a estabilidade dos resultados do modelo fatorial. A estabilidade fatorial depende principalmente do tamanho da amostra e do número de casos por variável. O pesquisador sempre é encorajado a obter a maior amostra possível e a desenvolver modelos parcimoniosos para aumentar a proporção casos-por-variáveis (HAIR et al., 2009).

Deteção de observações influentes. Além da generalidade, uma outra questão importante para a validação da análise fatorial é a deteção de observações influentes.

11 USOS ADICIONAIS DOS RESULTADOS DA ANÁLISE FATORIAL

Dependendo dos objetivos da aplicação da análise fatorial, o pesquisador pode parar com a interpretação fatorial ou utilizar-se de um dos métodos para redução de dados. Se o objetivo é simplesmente identificar combinações lógicas de variáveis e entender melhor as inter-relações entre variáveis, então a interpretação fatorial basta. Isso fornece uma base empírica para julgar a estrutura das variáveis e o impacto dessa estrutura quando se interpretam os resultados a partir de outras técnicas multivariadas. Se o objetivo, porém, é identificar variáveis apropriadas para a aplicação subsequente em outras técnicas estatísticas, então alguma forma de redução de dados será empregada (HAIR, 2009).

Seleção de variáveis substitutas para análise subsequente. Se a meta do pesquisador é simplesmente identificar variáveis apropriadas para a aplicação subsequente com outras técnicas estatísticas, o pesquisador tem a opção de examinar a matriz fatorial e selecionar a variável com a maior carga fatorial em cada fator para atuar como uma variável substituta representativa daquele fator. Essa é uma abordagem simples e direta somente quando uma variável tem uma carga fatorial bem maior do que todas as demais. Em muitos casos, porém, o processo de seleção é mais difícil porque duas ou mais variáveis têm cargas significantes e bastante próximas umas das outras, ainda que apenas uma seja escolhida como representativa de uma dimensão em particular (HAIR et al., 2009).

Criação de escalas múltiplas. Em termos simples, todas as variáveis com cargas elevadas em um fator são combinadas, e o total – ou, mais comumente, o escore médio das variáveis – é usado como uma variável de substituição. Uma escala múltipla apresenta dois benefícios específicos.

- ✚ Fornece um meio de superar consideravelmente o erro de medida inerente em todas as variáveis medidas. Erro de medida é o grau em que os valores observados não são representativos dos valores “reais” devido a diversas razões, que variam de erros reais (p.ex., erros na entrada de dados) à falta de habilidade de indivíduos fornecerem informações precisas. O impacto do erro de medida é mascarar parcialmente relações (p.ex., correlações ou comparações de médias de grupos) e dificultar a estimação de modelos multivariados. A escala múltipla reduz o erro de medida usando indicadores (variáveis) múltiplos para reduzir a dependência de uma única resposta.

Usando a resposta média ou típica de um conjunto de variáveis relacionadas, o erro de medida que poderia ocorrer em uma única questão será reduzido.

- ✚ Um segundo benefício da escala múltipla é sua habilidade de representar os múltiplos aspectos de um conceito com uma medida única. Muitas vezes, empregamos mais variáveis em nossos modelos multivariados como uma tentativa de representar as muitas facetas de um conceito que sabemos ser muito complexo. Entretanto, ao fazer isso, complicamos a interpretação dos resultados por causa da redundância nos itens associados ao conceito. Logo, gostaríamos de não apenas acomodar as descrições mais ricas de conceitos usando múltiplas variáveis, mas também de manter a parcimônia no número de variáveis em nossos modelos multivariados. A escala múltipla, quando corretamente construída, combina os múltiplos indicadores em uma só medida que representa o que acontece em comum no conjunto de medidas.

Definição conceitual. O ponto de partida para criar qualquer escala múltipla é sua definição conceitual. A definição conceitual especifica a base teórica para a escala múltipla definindo o conceito a ser representado em termos aplicáveis ao contexto de pesquisa. Na pesquisa acadêmica, as definições teóricas são baseadas em pesquisa prévia que define o caráter e a natureza de um conceito (HAIR et al., 2009).

A **validade de conteúdo** é a avaliação da correspondência das variáveis a serem incluídas em uma escala múltipla e sua definição conceitual. Essa forma de validade, também conhecida como validade de expressão, avalia subjetivamente a correspondência entre os itens individuais e o conceito por meio de avaliações de especialistas, pré-testes com múltiplas subpopulações ou outros meios (HAIR et al., 2009).

Dimensionalidade. Uma suposição inerente e exigência essencial para a criação de uma escala múltipla é que os itens sejam unidimensionais, significando que eles estão fortemente associados um com o outro e representam um só conceito.

Confiabilidade é uma avaliação do grau de consistência entre múltiplas medidas de uma variável. Uma forma de confiabilidade é teste/reteste, pelo qual a consistência é medida entre as respostas para um indivíduo em dois pontos no tempo. O objetivo é garantir que as respostas não sejam muito variadas durante períodos de tempo, de modo que uma medida tomada em qualquer instante seja confiável.

Validade. Após garantir que uma escala (1) está de acordo com sua definição conceitual, (2) é unidimensional e (3) atende aos níveis necessários de confiabilidade, o pesquisador deve fazer

uma avaliação final: validade da escala. Validade é o grau em que uma escala ou um conjunto de medidas representa com precisão o conceito de interesse.

Cálculo de escalas múltiplas. O cálculo de escalas múltiplas é um processo direto no qual os itens compreendendo a escala múltipla (i.e., os itens com cargas altas da análise fatorial) são somados ou têm suas médias calculadas. A abordagem mais comum é considerar a média dos itens na escala, o que fornece ao pesquisador um controle completo sobre o cálculo e facilita o uso em análises posteriores. Sempre que variáveis têm cargas positivas e negativas dentro do mesmo fator, ou as variáveis com cargas positivas, ou aquelas com cargas negativas devem ter seus dados revertidos. Tipicamente, as variáveis com as cargas negativas são revertidas no escore, de modo que correlações e cargas são agora todas positivas no mesmo fator (HAIR et al., 2009).

Cálculo de escores fatoriais para criar um conjunto menor de variáveis para substituir o conjunto original é o cálculo de escores fatoriais. Escores fatoriais também são medidas compostas de cada fator computadas para cada indivíduo. Conceitualmente, o escore fatorial representa o grau em que cada indivíduo tem escore elevado no grupo de itens que têm cargas elevadas em um fator. Assim, valores mais altos nas variáveis com cargas elevadas em um fator resultam em um escore fatorial superior.

Seleção entre os três métodos. Para escolher entre as três opções de redução de dados, o pesquisador deve tomar várias decisões, ponderando as vantagens e desvantagens de cada abordagem com os objetivos da pesquisa. A regra de decisão é, portanto, a seguinte:

- ✚ Se dados são usados somente na amostra original ou se ortogonalidade deve ser mantida, escores fatoriais são adequados.
- ✚ Se generalidade ou capacidade de transferência são desejáveis, então escalas múltiplas ou variáveis substitutas são mais apropriadas. Se a escala múltipla é um instrumento bem construído, válido e confiável, então é provavelmente a melhor alternativa.
- ✚ Se a escala múltipla não é testada e é exploratória, com pouca ou nenhuma evidência de confiabilidade ou validade, variáveis substitutas deverão ser consideradas caso uma análise adicional não seja possível para melhorar a escala múltipla.

Prática - Exemplo SPSS

Este exemplo se baseia em Andy Field, usando a mesma base e processo deste autor, com o SPSS versão 20. Mantenha duas noções em mente antes de iniciar a análise: (1) a confiabilidade da análise fatorial depende do tamanho da amostra (quanto maior, melhor) e (2) o SPSS quase sempre encontrará uma solução fatorial, mas se não tiver fundamentação teórica não haverá nenhum sentido a ser extraído. Uma vez que os dados estão postos no SPSS, siga as figuras para iniciar a análise:

Passo 1 – Acessar Análise Fatorial: Na seção analisar do SPSS, selecione redução de dimensão e depois fator:

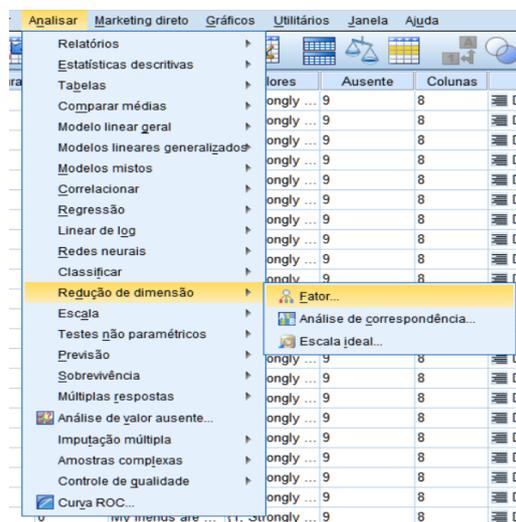


Figura 8: Acesso Análise Fatorial SPSS

Fonte: SPSS

Tal abrirá a seguinte caixa com opções:



Figura 9: Caixa de Análise Fatorial

Fonte: SPSS

As variáveis são listadas à esquerda, sendo que as flechas permitem selecionar as que se quer analisar. Os menus a direita são detalhados e explicados antes de se prosseguir com a análise.

Passo 2 – Descritivos. Ao clicar em descritivos, abre-se a seguinte caixa:

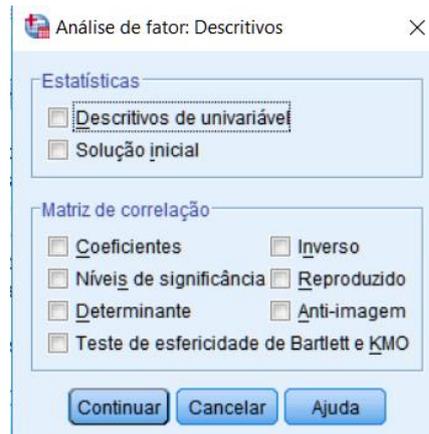


Figura 10: Descritivos

Fonte: SPSS

Lembra-se da matriz R da parte teórica desta apostila? Pois então, a opção de coeficiente produz a matriz-R, e o nível de significância torna possível saber para cada correlação da matriz-R a significância estatística. Este exemplo seguirá análise fatorial exploratória e pela matriz-R. Isto se dá pois a matriz-Q é mais raramente usada porque a análise confirmatória requer método de equação estrutural, conforme é possível visualizar no quadro do processo de análise fatorial exposto no referencial. Determinante é usado para checar Multicolinearidade, devendo ser maior que 0.00001 de acordo com Field (caso não seja, o ideal é ir eliminando variáveis que tenham uma correlação elevada, maior que 0.8). Finalmente, o teste de esfericidade mensura a adequação da amostra, devendo ser maior que 0.5, também segundo Field. Selecionamos, portanto, coeficientes, níveis de significância, determinante e teste de esfericidade. Clique em continuar, o que faz voltar para a caixa de análise de fator.

Passo 3 - Extração: Agora, clique em extração, o que abre a seguinte caixa:

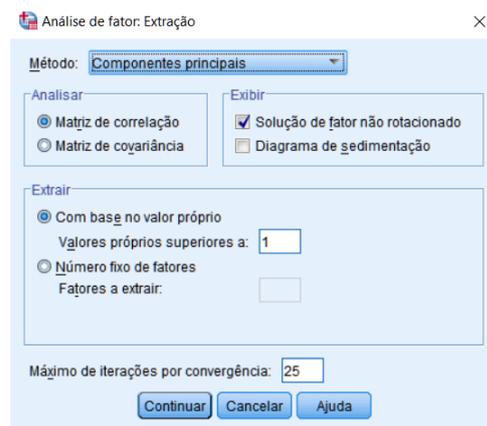


Figura 11: Método de Análise Fatorial

Fonte: SPSS

O método depende de muitos quesitos, sendo aconselhável antes de fazer sua análise fatorial estudar a teoria e averiguar qual o mais aconselhável. Neste exemplo, será utilizado o de componentes principais, também conhecido como “Principal Component Analysis” ou “PCA”. O screeplot, citado no referencial, é o diagrama de sedimentação na opção desta caixa. Relembrando, ele permite “identificar o número ótimo de fatores que podem ser extraídos” (HAIR et al., 2009), ou, em outros termos, “quantos fatores devem ser mantidos na análise” (FIELD, 2009). Quanto à solução de fator não rotacionado, este é “útil em perceber a melhora na interpretação devido á rotação” (FIELD, 2009). Finalmente, quanto à opção de extrair, o ideal é manter a opção “com base de valor próprio” em 1, segundo recomendações estatísticas, que é o padrão que se apresenta ao abrir a caixa. Field (2009) sugere para rodar a análise (1) com diagrama de sedimentação, (2) com base no valor próprio de 1. **Passo 4 - Rotação:** Clique em continuar e abra a opção de rotação, o que conduz à seguinte caixa:

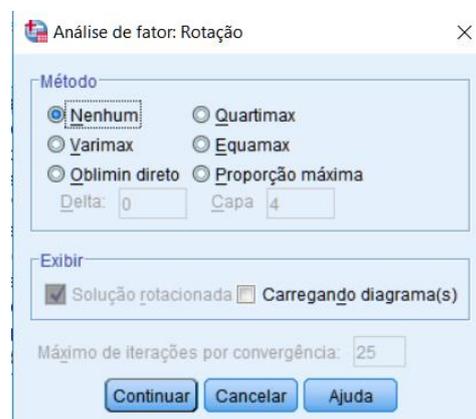


Figura 12: Rotação

Fonte: SPSS

Os métodos de rotação foram explicados no referencial, e cada análise dependerá do tipo de pesquisa que se está a fazer. Field (2009) em geral recomenda Varimax, usada neste exemplo. Note que a solução rotacionada é fundamental para interpretação. **Passo 5 - Pontuações:** Clique em continuar e abra “pontuações”, revelando a seguinte caixa:

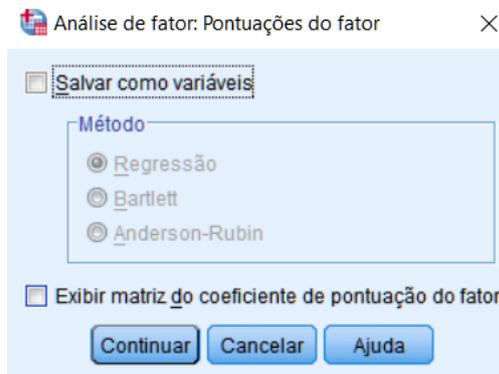


Figura 13: Pontuações

Fonte: SPSS

Salvar como variáveis permite maiores análises e é um tema que é de todo interessante se aprofundar, contudo vai além do escopo da presente apostila. **Passo 6 – Opções:** Finalmente, a caixa de opções permite algumas opções que ajudam na interpretação, sendo a parte teórica do porquê se encontra em Field (2009). Para este exemplo, repete-se o que Field seleciona, 0.4

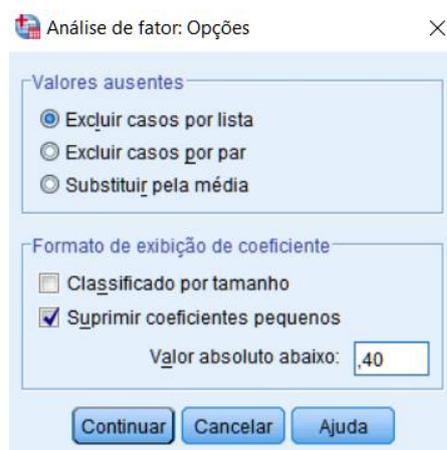


Figura 14: Opções

Fonte: SPSS

Passo 7 – Rodar a Análise: Selecione as variáveis e clique em ok para o SPSS rodar a análise.

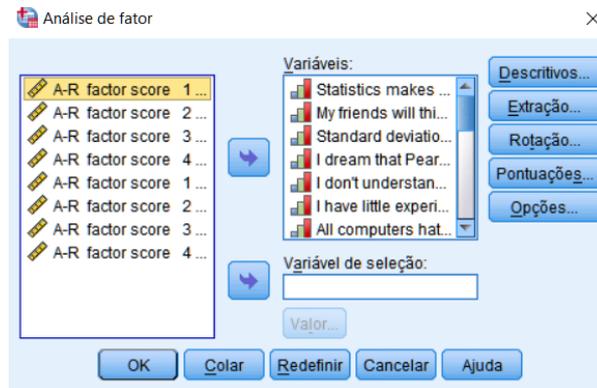


Figura 15: Rodando a Análise

Fonte: SPSS

Interpretação dos Resultados

O SPSS produz diversos resultados conforme o que se assinala na realização da análise, explanada anteriormente, os quais podem ser visualizados no próprio programa do SPSS: Algumas imagens são cortadas porque são extensas.

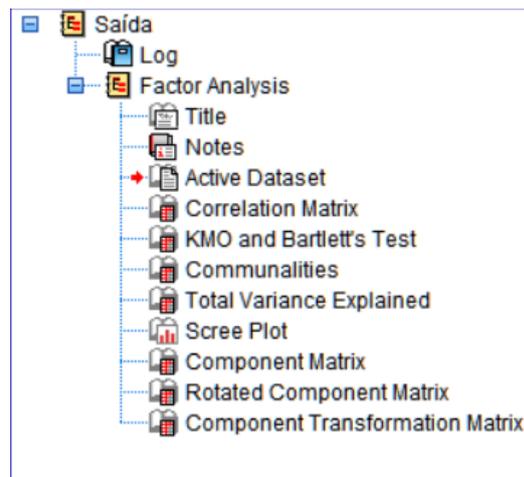


Figura 16: Saída de Resultados

Fonte: SPSS

Inicialmente, a matriz de correlação (“Correlation Matrix”), que produz a correlação na parte superior e a significância estatística na inferior.

		Statistics makes me cry	My friends will think I'm stupid for not being able to cope with SPSS	Standard deviations excite me
Correlation	Statistics makes me cry	1,000	-,099	-,337
	My friends will think I'm stupid for not being able to cope with SPSS	-,099	1,000	,318
	Standard deviations excite me	-,337	,318	1,000
	I dream that Pearson is attacking me with correlation coefficients	,436	-,112	-,380
	I don't understand statistics	,402	-,119	-,310
	I have little experience of computers	,217	-,074	-,227
	All computers hate me	,305	-,159	-,382
	I have never been good at mathematics	,331	-,050	-,259

Figura 17: Matriz de Correlação

Fonte: SPSS

Sig. (1-tailed)	Statistics makes me cry		,000	,000
	My friends will think I'm stupid for not being able to cope with SPSS	,000		,000
	Standard deviations excite me	,000	,000	
	I dream that Pearson is attacking me with correlation coefficients	,000	,000	,000
	I don't understand statistics	,000	,000	,000
	I have little experience of computers	,000	,000	,000
	All computers hate me	,000	,000	,000
	I have never been good at mathematics	,000	,006	,000

Figura 18: Matriz de Correlação 2

Fonte: SPSS

É importante usar a tabela para notar relações não estatisticamente significantes (geralmente maiores que sig 0,05). Correlações acima de 0,9 também devem ser checadas, por causa de singularidade nos dados (FIELD, 2009). A seguir, o teste de Kaiser:

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,930
Bartlett's Test of Sphericity	Approx. Chi-Square	19334,492
	df	253
	Sig.	,000

Figura 19: Teste Bartlett

Fonte: SPSS

Quanto mais perto de 1, significa que os padrões de correlações são relativamente compactos e, portanto, os resultados tendem a ser confiáveis. No caso, note o valor de 0,93, sendo que o mínimo para ser tolerável é 0,5 (FIELD, 2009; KAISER, 1974). Quanto ao teste de Bartlett, a importância reside especialmente na significância estatística, que mostra se a análise fatorial é apropriada ou não.

A análise de componentes principais tem como suposição de que a variância é comum, portanto antes da extração o valor da “communalities” é 1 para todos os valores.

Communalities		
	Initial	Extraction
Statistics makes me cry	1,000	,435
My friends will think I'm stupid for not being able to cope with SPSS	1,000	,414
Standard deviations excite me	1,000	,530
I dream that Pearson is attacking me with correlation coefficients	1,000	,469
I don't understand statistics	1,000	,343
I have little experience of computers	1,000	,654
All computers hate me	1,000	,545
I have never been good at mathematics	1,000	,739

Figura 20: Comunalidades

Fonte: SPSS

Na parte “extração”, significa que, por exemplo, na primeira pergunta, 43,5% da variância é variância comum ou compartilhada.

A seguir, a tabela da variância explicada:

Component	Total Variance Explained								
	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7,290	31,696	31,696	7,290	31,696	31,696	3,730	16,219	16,219
2	1,739	7,560	39,256	1,739	7,560	39,256	3,340	14,523	30,742
3	1,317	5,725	44,981	1,317	5,725	44,981	2,553	11,099	41,841
4	1,227	5,336	50,317	1,227	5,336	50,317	1,950	8,476	50,317
5	,988	4,295	54,612						
6	,895	3,893	58,504						
7	,806	3,502	62,007						
8	,783	3,404	65,410						
9	,751	3,265	68,676						
10	,717	3,117	71,793						
11	,684	2,972	74,765						
12	,670	2,911	77,676						
13	,612	2,661	80,337						
14	,578	2,512	82,849						
15	,549	2,388	85,236						
16	,523	2,275	87,511						
17	,508	2,210	89,721						
18	,456	1,982	91,704						
19	,424	1,843	93,546						
20	,408	1,773	95,319						
21	,379	1,650	96,969						
22	,364	1,583	98,552						
23	,333	1,448	100,000						

Extraction Method: Principal Component Analysis.

Figura 21: Variância

Fonte: SPSS

Esta tabela apresenta os autovalores (eigenvalues) antes da extração, depois da extração e depois da rotação. Os autovalores associados com cada fator mostram o quanto de variância é explicada. Note que os primeiros valores explicam uma parte elevada da variância neste exemplo. Os autovalores maiores que 1 são extraídos pelo software, perfazendo 4 fatores. A soma extraída de cargas elevadas ao quadrado (“extracted sum of squared

loadings”) perfaz tais quatro fatores, com os mesmos valores, mas excluindo os autovalores menores que 1. Finalmente, na parte final da tabela, as somas rotacionais de cargas elevadas ao quadrado (“rotation sum of squared loadings”) mostra os autovalores dos fatores pós-rotação. Note que os valores se aproximam mais uns dos outros do que anteriormente, sendo que a rotação faz com que haja uma aproximação ou equalização.

A matriz de componentes, segundo Field (2009) não é de muita significância para interpretação, sendo que os espaços em branco é porque foi delimitado que os “loadings” menores de 0.4 fossem suprimidos.

Component Matrix ^a				
	Component			
	1	2	3	4
Statistics makes me cry	,586			
My friends will think I'm stupid for not being able to cope with SPSS		,548		
Standard deviations excite me	-,629			
I dream that Pearson is attacking me with correlation coefficients	,634			
I don't understand statistics	,556			
I have little experience of computers	,562		,571	
All computers hate me	,685			
I have never been good at mathematics	,549	,401		-,417

Figura 22: Matriz de Componente

Fonte: SPSS

O número de fatores a extrair depende do pesquisador, sendo o software incapaz de saber. Neste caso, foram 4 fatores extraídos. Nota-se que tal se dá quando existem (1) menos de 30 variáveis, (2) comunalidades maiores que 0.7.

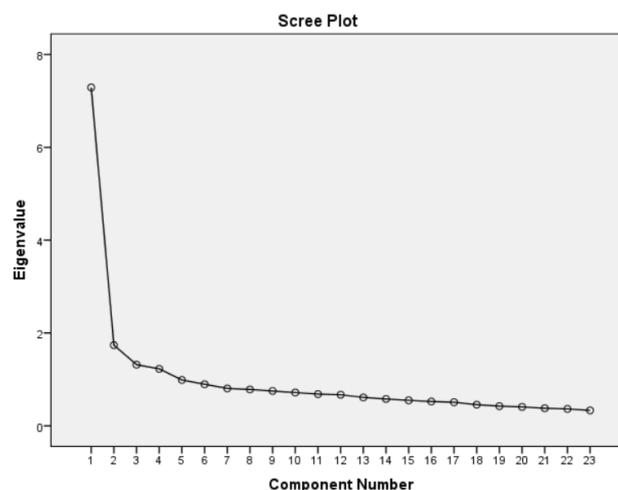


Figura 23: Scree Plot

Fonte: SPSS

O scree plot indica o ponto de inflexão da curva (FIELD, 2009) e este autor também nota que se os critérios de Kaiser não se aplicarem a sua análise, o scree plot pode ser usado quando a amostra é elevada (mais de 300 casos).

A matriz rotacionada é dada a seguir:

Rotated Component Matrix^a

	Component			
	1	2	3	4
Statistics makes me cry		,496		
My friends will think I'm stupid for not being able to cope with SPSS				,543
Standard deviations excite me		-,567		
I dream that Pearson is attacking me with correlation coefficients		,516		
I don't understand statistics		,429		
I have little experience of computers	,800			
All computers hate me	,638			
I have never been good at mathematics			,833	
My friends are better at statistics than me				,648
Computers are useful only for playing games	,550			
I did badly at mathematics at school			,747	
People try to tell you that SPSS makes statistics easier to understand but it doesn't	,473	,523		
I worry that I will cause irreparable damage because of my incompetence with computers	,647			
Computers have minds of their own and deliberately go wrong whenever I use them	,579			
Computers are out to get me	,459			
I weep openly at the mention of central tendency		,514		
I slip into a coma whenever I see an equation			,747	
SPSS always crashes when I try to use it	,684			
Everybody looks at me when I use SPSS				,428
I can't sleep for thoughts of eigenvectors		,677		
I wake up under my duvet thinking that I am trapped under a normal distribution		,661		
My friends are better at SPSS than I am				,645
If I'm good at statistics my friends will think I'm a nerd				,586

Extraction Method: Principal Component Analysis.
 Rotation Method: Varimax with Kaiser Normalization.
 a. Rotation converged in 8 iterations.

Figura 24: Matriz Rotacionada

Fonte: SPSS

Field (2009) nota que esta matriz deve ser comparada com a solução não-rotacionada. Em geral, após a rotação há clarificação, conforme o que se está estudando. Usa-se rotação ortogonal quando os fatores são teoricamente diferentes e rotação oblíqua quando os fatores se relacionam entre si. A parte final de interpretação é identificar temas comuns dos construtos analisados. Neste exemplo, nota-se que se pode classificar os fatores rotacionados como “medo de computadores”, “medo de estatística”, “medo de matemática” e “avaliação por pares”; se conseguiu extrair construtos relacionados ou talvez sub-componentes da ansiedade frente SPSS: isto cabe a literatura, sendo que o Software não explica por si só. Usando o embasamento teórico e sabendo como rodar a análise no SPSS, nota-se que a parte mais complexa é se interpretar conforme a teoria.

REFERENCIAS

BARTHOLOMEW, David J. The foundations of factor analysis. *Biometrika*, v. 71, n. 2, p. 221-232, 1984.

CARVALHO, Francisco Ricardo Duarte. **Análise Fatorial**. Dissertação para a obtenção do Grau de Mestre em Matemática. Universidade de Coimbra, 2013.

CUDECK, Robert; MACCALLUM, Robert C. (Ed.). *Factor analysis at 100: Historical developments and future directions*. Routledge, 2007.

FIELD, Andy. **Descobrimo a estatística usando o SPSS**. Porto Alegre :Artmed, 2009.

HAIR, Joseph F. et al. *Análise multivariada de dados*. Bookman Editora, 2009..

KING, G. How not to lie with statistics [Online] Disponível em:

<<http://gking.harvard.edu/files/mist.pdf>> Acesso em: 28/04/2019

ANEXOS

A base de dados utilizada nesta apostila pode ser encontrada em:
https://www.discoveringstatistics.com/repository/ds_data_files.zip: arquivo SAQ.sav

Os resultados completos do SPSS utilizando esta base podem ser encontrados em:
<https://www.dropbox.com/s/3x44rp68vrukdw/dados%20rodados%20apostila.spv?dl=0>