

GIA5005 – Estatística Aplicada a Administração

ANÁLISE FATORIAL (AF) *Factor Analysis*

“A análise fatorial (AF) busca descrever a estrutura de dependência existente em um conjunto de variáveis por meio da criação de fatores (não observados) que, eventualmente, expressam construtos subjacentes aos dados”. (Mingoti, 2007)

Construto (*s. m.*): Construção mental ou síntese feita a partir da combinação de vários elementos. Resultado de um processo de construção. (*Psicologia*). Percepção ou pensamento formado a partir da combinação de lembranças com acontecimentos atuais.

Exemplos: Nível de ansiedade, satisfação, bem-estar, comprometimento emocional, autoestima, aptidão etc.

Subjacente (*adj.*): que se encontra por baixo; que não se expressa com clareza; que está oculto; implícito.

Reis (1997): Afirma que Análise Fatorial (AF) é um conjunto de técnicas estatísticas cujo objetivo é representar ou descrever um número de variáveis iniciais por um número menor de variáveis hipotéticas.

Mingoti (2007): Se as variáveis originais do estudo estão agrupadas em subconjuntos de novas variáveis não correlacionadas, a AF tem o objetivo de encontrar estes fatores (latentes) de agrupamentos.

A técnica de Análise Fatorial foi introduzida por Spearman (1904). Lawley (1940) reapresentou a técnica com maior rigor matemático, melhorando a sua aceitação no meio científico.

Exemplo: Suponha que sejam coletadas informações sobre dimensões de uma pessoa (comprimento do tórax, das pernas, dos braços, além do peso e altura) e outras variáveis que representem a sua preferência por roupas, alimentos e formas de entretenimento.

- É inútil criar um índice geral que combine linearmente todas essas variáveis...
- É mais útil criar um índice que represente o dimensionamento corporal global do indivíduo e outros índices que representem a sua preferência por roupas, por alimentos e por formas de entretenimento.

Com esses índices diferenciados pela natureza das variáveis envolvidas na sua construção, pode ser possível identificar os vários grupos/segmentos da população amostrada.

Exemplos de aplicação (Mingoti, 2008):

- Psicologia e Ciências Sociais: Identificar fatores relacionados com a inteligência humana e associá-los à etnia.
- Marketing: Identificar o perfil de consumidores ou os fatores que os levam a comprar ou preferir certos produtos ou marcas.
- Analisar o posicionamento de produtos e serviços de uma empresa perante os seus concorrentes de mercado.
- Descobrir os fatores relacionados à satisfação do indivíduo com a sua vida pessoal ou com a sua atividade profissional.
- Elaborar índices diferenciados de qualidade de serviços.
- Identificar fatores relacionados à aceitação de um novo produto comestível.

Existem dois tipos de análise fatorial:

- **Exploratória:** Busca encontrar os fatores subjacentes às variáveis originais do estudo. Quando o pesquisador realiza esta análise ele não tem noção clara de quantos fatores fazem parte do modelo e nem o que eles representam.
- **Confirmatória:** O pesquisador já tem em mãos um modelo fatorial pré-estabelecido (teórico) e deseja verificar se ele é aplicável ou se ele é consistente com os dados amostrais de que dispõe.

Neste caso, mesmo que o pesquisador já tenha definido um modelo hipotético para o seu problema, é recomendável que ele também faça uma análise fatorial exploratória, inicial, para ter uma noção prévia se o número de fatores e a sua interpretação estão coerentes com seus dados atuais.

Importante: Na AF é possível saber quanto e como cada fator está associado a cada variável original do estudo e quanto o conjunto de fatores explica da variabilidade geral dos dados originais.

Problema: As variáveis com maior variabilidade (maior variância) podem ser predominantes na construção dos fatores e podem mascarar a presença de variáveis com menor variabilidade.

Solução: Nesses casos, sugere-se trabalhar com as variáveis padronizadas, cujas variâncias são iguais a um e cujas covariâncias correspondem às correlações entre as variáveis originais.

Lembre-se: Se os dados de uma variável X tem média \bar{x} e desvio padrão s , os dados da variável padronizada $Z = \frac{X - \bar{x}}{s}$ terão média zero e desvio padrão igual a um.

PRINCIPAIS MÉTODOS USADOS NA OBTENÇÃO DOS FATORES

1) MÉTODO DA MÁXIMA VEROSSIMILHANÇA

É um método pouco usado, porque pressupõe que as variáveis envolvidas no estudo tenham distribuição normal multivariada, o que é difícil de ser comprovado.

Não é indicado para os casos em que a suposição de normalidade das variáveis envolvidas não esteja satisfeita, o que limita a sua utilização como, por exemplo, no estudo de escalas psicológicas de avaliação.

Estas escalas, em sua maioria, são compostas por itens (questões) cujas respostas não podem ser modeladas pela distribuição normal (escalas de Likert, Thurstone *etc.*).

2) MÉTODO BASEADO NA ANÁLISE DE COMPONENTES PRINCIPAIS

- Seu uso é mais frequente porque as variáveis envolvidas não precisam ter distribuição normal.
- Tecnicamente os fatores são obtidos a partir de uma decomposição espectral da matriz de correlação ou de covariâncias.
- Com a decomposição espectral obtêm-se as cargas fatoriais, que indicam o quanto cada variável está associada a cada fator.
- Os autovalores dessa decomposição refletem a importância de cada fator.
- Quando o número de fatores é igual ao número de variáveis, a soma dos autovalores é igual à soma das variâncias das variáveis originais (não padronizadas), ou ao número de variáveis padronizadas.

- A razão entre um autovalor e a soma das variâncias indica a proporção da variabilidade total que é explicada pelo fator.
- A **comunalidade** reflete a proporção da variância de cada variável que é explicada pelos fatores. Geralmente é expressa como h^2 e é calculada como a soma dos quadrados das cargas fatoriais.

COMO INTERPRETAR OS FATORES?

- A interpretação dos fatores de uma AF é baseada nas cargas fatoriais, que expressam as covariâncias (ou correlações) entre cada fator e as variáveis originais.

Problema: Para um mesmo conjunto de variáveis é possível encontrar diversas soluções, ou seja, um grande conjunto de fatores que explica muito bem o comportamento dos dados.

Dica: Para facilitar a interpretação dos fatores devemos marcar em negrito as cargas fatoriais mais altas (superiores a 0,5), evidenciando quais variáveis predominam na definição de cada fator.

- **Problema:** Em geral, a primeira solução fornecida pelos programas estatísticos não gera fatores que tenham uma interpretação simples ou adequada.

Dica: Soluções equivalentes, mas com maior potencial de interpretabilidade podem ser obtidas realizando procedimentos chamados rotação dos fatores. A rotação *varimax* é uma das mais utilizadas.

MODELO GERAL DE ANÁLISE DE FATORES

Cada uma das p variáveis observadas Z_i pode ser escrita como uma combinação de m fatores escolhidos por algum critério:

$$Z_i = a_{i1}F_1 + a_{i2}F_2 + \cdots + a_{im}F_m + e_i$$

Em que:

Z_i é o i -ésimo escore observado (padronizado) do teste;

a_{i1}, \dots, a_{im} são as cargas dos fatores para o i -ésimo teste;

F_1, \dots, F_m são os m fatores comuns não correlacionados, cada um com média zero e variância unitária;

e_i é um fator específico para o i -ésimo teste, tem média zero e não está correlacionado com qualquer um dos fatores comuns.

Das suposições deste modelo, segue que:

$$\text{var}(Z_i) = a_{i1}^2 + a_{i2}^2 + \dots + a_{im}^2 + \text{var}(e_i)$$

- A primeira parte, $a_{i1}^2 + a_{i2}^2 + \dots + a_{im}^2$ é chamada **comunalidade** da variável Z_i , ou seja, é a parte de $\text{var}(Z_i)$ que está compartilhada com todas as outras variáveis ou fatores comuns;
- $\text{var}(e_i)$, chamada **especificidade** de Z_i , é a parte da variância que não está relacionada com os fatores comuns.

As cargas fatoriais assumem valores entre -1 e 1 , porque a comunalidade não pode exceder um.

Pode ser mostrado que duas variáveis podem ser altamente correlacionadas se elas tiverem altas cargas fatoriais nos mesmos fatores.

ANÁLISE FATORIAL POR COMPONENTES PRINCIPAIS

É mais comum usar a ACP para definir o número de fatores:

Passo 1: Com p variáveis (X_1, X_2, \dots, X_p) medidas em n indivíduos realizamos uma análise de componentes principais (ACP) com os dados originais ou com os dados padronizados e escolhemos um número (m) adequado de componentes principais (que explique uma boa proporção da variabilidade total).

Passo 2: O modelo fatorial não rotacionado pode ser escrito como:

$$X_1 = a_{11}F_1 + a_{21}F_2 + \dots + a_{m1}F_m + e_1$$

$$X_2 = a_{12}F_1 + a_{22}F_2 + \dots + a_{m2}F_m + e_2$$

⋮

$$X_p = a_{1p}F_1 + a_{2p}F_2 + \dots + a_{mp}F_m + e_p$$

Após o ajuste do modelo, numa situação ideal, espera-se que as comunalidades de cada variável sejam altas (próximas de um).

Passo 3: Se a interpretação dos fatores estiver complicada porque as variáveis estão correlacionadas com mais de um fator, podemos realizar uma rotação dos eixos, que pode facilitar a interpretação.

Com a rotação dos fatores espera-se que as cargas fatoriais das variáveis sejam maiores com um único fator.

Métodos de rotação ortogonal: *varimax*, *equimax*, *quartimax*.

A rotação dos eixos não altera a sua comunalidade.

Passo 4: Calcular os escores fatoriais que podem indicar agrupamentos interessantes de indivíduos, similar ao que foi feito na ACP.

Um exemplo inicial vai ajudar a interpretar os resultados de uma análise fatorial.

Exemplo 1: Uma AF foi aplicada aos itens da forma traço do Inventário de Ansiedade Traço-Estado (IDATE-T) (Reis, 1997).

- Os itens (questões) desse inventário constituem-se de frases e são avaliadas por meio da atribuição de um escore de 1 a 4, onde 1 indica que o estado descrito na frase quase nunca ocorre e 4 indica que ocorre quase sempre.
- Na análise desses itens os escores foram convertidos de modo que um escore alto represente sempre um estado negativo.
- A AF foi feita utilizando a ACP e das 20 variáveis (itens) do estudo original foram utilizadas somente as oito variáveis que são descritas a seguir:

Descrição dos itens da escala (parcial) IDATE-T

X1: Sinto-me bem.

X9: Preocupo-me demais com as coisas sem importância.

X10: Sou feliz.

X11: Deixo-me afetar muito pelas coisas.

X13: Sinto-me seguro.

X16: Estou satisfeito.

X17: Às vezes ideias sem importância entram na cabeça e ficam me preocupando.

X18: Levo os desapontamentos tão a sério que não consigo tirá-los da cabeça.

Lembre-se: Um escore alto indica um estado negativo.

Matriz de correlação entre os itens

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

As correlações entre as variáveis não são muito altas. Percebe-se um grupo de correlações positivas e outro grupo de correlações negativas, indicando a possível existência de dois fatores que avaliam características distintas.

1º Passo: Análise de componentes principais das variáveis originais (sem padronização)

Autovalores e % de variância explicada

CP	Autovalor	% Var	%Var Acum
1	3.525	44.07	44.07
2	1.504	18.80	62.86
3	0.665	8.31	71.17
4	0.614	7.68	78.85
5	0.512	6.40	85.25
6	0.444	5.55	90.80
7	0.425	5.31	96.11
8	0.311	3.89	100.00

Vamos escolher somente dois fatores, que apresentam autovalores superiores a um e que juntos explicam 63% da variabilidade total.

2º passo: Realização da análise fatorial (sem rotação):

Cargas fatoriais e comunalidade (Solução inicial)

Variável	F_1	F_2	Comunalidade
X1	0.678	0.445	0.657
X9	-0.549	0.585	0.644
X10	0.719	0.492	0.758
X11	-0.633	0.367	0.536
X13	0.679	0.192	0.497
X16	0.751	0.392	0.719
X17	-0.593	0.564	0.670
X18	-0.686	0.279	0.548

Na coluna Comunalidade percebe-se que os dois fatores explicam juntos 50% ou mais da variabilidade que é comum a todas as variáveis.

A interpretação das cargas fatoriais não é clara. Por exemplo: o primeiro fator está altamente correlacionado (positiva ou negativamente) com quase todas as variáveis em estudo!

3º Passo: Realizando uma rotação *varimax* dos fatores temos:

Variável	F_1	F_2	Comunalidade ^(*)
X1	0.804	-0.101	0.657
X9	-0.038	0.802	0.644
X10	0.866	-0.092	0.758
X11	-0.244	0.690	0.536
X13	0.641	-0.294	0.497
X16	0.826	-0.189	0.719
X17	-0.086	0.814	0.670
X18	-0.341	0.657	0.548

^(*) As comunalidades não são alteradas pela rotação do fator.

Após a rotação, os dois fatores passam a ter correlações altas com apenas um dos fatores, o que facilita a interpretação.

Cargas fatoriais rotacionadas

	F_1	F_2	Variável
X1	0.804	-0.101	Sinto-me bem
X9	-0.038	0.802	Preocupo-me demais com as coisas sem importância
X10	0.866	-0.092	Sou feliz
X11	-0.244	0.690	Deixo-me afetar muito pelas coisas
X13	0.641	-0.294	Sinto-me seguro
X16	0.826	-0.189	Estou satisfeito
X17	-0.086	0.814	Às vezes ideias sem importância me entram na cabeça e ficam me preocupando
X18	-0.341	0.657	Levo os desapontamentos tão a sério que não consigo tirá-los da cabeça

IMPORTANTE: Para interpretar e dar nomes aos fatores devemos avaliar com quais variáveis do estudo cada fator apresenta correlações mais altas:

F_1 : Satisfação pessoal, pois está altamente relacionado com as variáveis **X1** (Sinto-me bem), **X10** (Sou feliz), **X13** (Sinto-me seguro) e **X16** (Estou satisfeito).

F_2 : Dificuldade em lidar com problemas, pois está altamente relacionado com **X9** (Preocupo-me demais com as coisas sem importância), **X11** (Deixo-me afetar muito pelas coisas), **X17** (Às vezes ideias sem importância entram na cabeça e ficam me preocupando) e **X18** (Levo os desapontamentos tão a sério que não consigo tirá-los da cabeça).

4º Passo: Se tivermos as respostas dos indivíduos a cada um dos itens (ou variáveis) podemos calcular os escores fatoriais dos indivíduos em estudo e identificar como eles expressam essas características (construtos) relacionadas à ansiedade.

COMO ESCOLHER O NÚMERO DE FATORES?

- A escolha do número de fatores é muito importante em uma AF.
- Segundo Hair et al. (1995) se o pesquisador opta por um número muito reduzido, ele pode não identificar estruturas importantes existentes nos dados e, por outro lado, se o número é excessivo, ele pode ter problemas de interpretabilidade dos fatores.
- Existem na literatura vários critérios que auxiliam na determinação do número de fatores que, invariavelmente, quando empregados em um mesmo conjunto de dados, podem conduzir a resultados diferentes.

Regra geral: o pesquisador deve buscar um compromisso entre o número de fatores, que deve ser o menor possível, e a interpretabilidade dos fatores.

- É comum comparar soluções com diferentes números de fatores e fazer a escolha com base no **bom senso** do pesquisador.
- Os métodos de escolha têm **caráter** apenas **indicativo**. Não existe uma hierarquia de importância entre eles.

-

Critério de Kaiser (1958)

O número de fatores deve ser igual ao número de autovalores maiores ou iguais à média das variâncias das variáveis analisadas.

Quando a AF está baseada na matriz de correlação (variáveis padronizadas), esse critério indica a exclusão de fatores com autovalores inferiores a um, ou seja, descarta os fatores que tenham um grau de explicação inferior ao de uma variável isolada.

Critério *scree test*

O número ótimo de fatores é obtido quando a variação da explicação entre fatores consecutivos passa a ser pequena.

Critério da porcentagem da variância explicada

O número é determinado de modo que o conjunto de fatores explique uma porcentagem pré-definida da variabilidade global.

Métodos inferenciais

Referem-se aos casos em que cada uma das variáveis originais segue uma distribuição normal, o que não é comum em diversas pesquisas. Consistem no desenvolvimento de testes estatísticos que indicam a melhor escolha do número de fatores.

TAMANHO IDEAL DE AMOSTRA PARA REALIZAR UMA BOA ANÁLISE FATORIAL?

- Uma AF envolve a estimação de um grande número de parâmetros e para que isso seja feito com um mínimo de qualidade, é necessário um tamanho amostral relativamente grande em comparação ao número de variáveis envolvidas.
- Hair et al. (1995) indica que a AF não deve ser utilizada em amostras com menos de 50 observações.
- Reis (1997, p. 274) e Hair et al. (1995, p. 373) sugerem que o número de observações seja de no mínimo 5 vezes o número de variáveis, e que, preferencialmente, a análise seja feita com, pelo menos, 100 observações.

O QUE É UMA ANÁLISE FATORIAL “BEM-SUCEDIDA”?

- O sucesso de uma AF está diretamente ligado aos objetivos iniciais do pesquisador.
- Se a intenção é a simples redução do número de variáveis, ela será bem-sucedida se for possível determinar um pequeno conjunto de fatores interpretáveis, que consigam explicar uma parte considerável da variabilidade do conjunto original de variáveis.
- Encontrar um número pequeno de fatores que possuam um alto grau de explicação da variabilidade original dos dados.
- Encontrar fatores que sejam interpretáveis.

RAZÕES QUE PODEM EXPLICAR O INSUCESSO DE UMA AF

- **Tamanho insuficiente da amostra.** Uma amostra pequena pode não refletir de maneira precisa a estrutura de interdependência dos dados;
- **Variáveis com uma fraca interdependência.** Hair et al. (1995) discutem que para o sucesso de uma AF é necessário que exista um número razoável de correlações superiores (em módulo) a 0,30, caso contrário a estrutura de interdependência será muito tênue para produzir resultados satisfatórios
- **A estrutura de dependência pode não ser homogênea em toda a amostra.** Exemplo: alguns itens possuem estruturas de dependência diferentes para homens e mulheres. Neste caso, o ideal é realizar uma AF para cada um dos gêneros.

PROCEDIMENTO GERAL PARA UMA ANÁLISE DE FATORES

1ª Fase: Obtenção de cargas fatoriais provisórias a_{ij}

- Usar a ACP para escolher o número (m) de fatores importantes a serem utilizados na análise fatorial. Os outros CP 's são negligenciados e vão compor o erro específico (e_i).
- Os m fatores escolhidos não são correlacionados entre si e nem com os fatores específicos, mas têm variâncias diferentes de um.
- Os fatores específicos são correlacionados entre si, o que é um problema. Deixa de ser um problema quando as comunalidades forem altas.
- A primeira solução da AF apresenta fatores de difícil interpretação.

2ª Fase: Rotação de fator

- As cargas dos fatores provisórios (a_{ij}) não são únicas e, geralmente, dificultam a interpretação dos fatores.
- O ideal é buscar soluções equivalentes à inicial, mas com maior potencial de interpretabilidade, utilizando a **rotação dos fatores**.
- As rotações podem ser ortogonais ou oblíquas. Hair et al. (2009) afirmaram que as duas formas de rotação produzem resultados semelhantes.
- As rotações ortogonais são mais utilizadas, são mais fáceis de interpretar. No entanto, o pesquisador deve assumir que os construtos são independentes, o que, na prática é um pressuposto mais difícil de ser respeitado, principalmente nas Ciências Sociais.

- Dentre os métodos de rotação ortogonais mais utilizados podemos citar os métodos: *equimax*, *varimax* e *quartimax*.

Rotação *varimax*: maximiza a dispersão das cargas fatoriais dentro dos fatores e busca minimizar o número de variáveis que apresentam altas cargas fatoriais em diferentes fatores. É o método mais utilizado, por facilitar a interpretação dos fatores.

Rotação *quartimax*: maximiza a dispersão das cargas fatoriais de uma variável em todos os fatores, o que pode ocasionar muitas variáveis com cargas altas em um único fator.

Rotação *equimax*: é um híbrido dos métodos anteriores e pouco utilizado.

- Após uma rotação dos fatores espera-se que as cargas (g_{ij}) dos novos fatores sejam mais próximas de zero ou de um.
 - $g_{ij} \cong 0$: a variável X_i é pouco relacionada com o fator F_j .
 - $g_{ij} \cong 1$ ou $g_{ij} \cong -1$: a variável X_i é praticamente determinada pelo fator F_j .

Neste estágio é comum colocar **rótulos** (nomes) nos fatores, o que não é uma atividade fácil, porque requer muita criatividade, imaginação e conhecimento das variáveis envolvidas no estudo.

3ª Fase:

- Calcular os escores fatoriais, ou seja, os valores dos fatores rotacionados ($F_1^*, F_2^*, \dots, F_m^*$) para cada um dos n indivíduos.

Comentários gerais:

- A escolha do número de fatores (m) depende de cada analista.
- Se os fatores provisórios forem obtidos pela ACP com os dados padronizados, podemos escolher $m =$ número de autovalores maiores que um.
- Um fator associado a um autovalor $\lambda_i < 1$, responde por menos variação nos dados do que os escores de teste originais.
- Aumentar o número de fatores \Rightarrow aumenta a comunalidade das variáveis (o que é *bom!*) \Rightarrow aumenta também a dificuldade de interpretação dos fatores (o que é *ruim!*).
- **As comunalidades não são alteradas pela rotação do fator.**

USANDO O MINITAB PARA REALIZAR A ANÁLISE

Exemplo 2: Os candidatos a um emprego foram avaliados com base em 12 características julgadas importantes pelo responsável por Recursos Humanos de uma empresa: Histórico escolar, Aparência, Comunicação, Adequação à empresa, Experiência, Adequação ao cargo, Carta de interesse, Simpatia, Organização, Potencial, Currículo e Autoconfiança.

Dados: As notas dos 50 candidatos em cada um desses requisitos encontram-se na planilha **AF - Emprego**, na pasta Dados Amostrais, criada junto com a instalação do Minitab 17.

Uma análise de componentes principais com base nos dados originais (matriz de covariância) indicou a escolha de cinco CP's, que juntos explicam 90% da variabilidade original dos dados.

Em **Stat > Multivariado > Análise fatorial > Variáveis** incluímos todas as colunas de dados; em **Números de fatores para extrair:** digitamos 5; em **Método de extração** marcamos **Análise de Componentes principais**; em **Opções > Matriz para Fator** marcar **Correlação**; em **Resultados** marcar **Somente cargas fatoriais** e **Ordenar cargas fatoriais**.

Na coluna **Comum** percebe-se que os 5 fatores explicam 80% ou mais da variabilidade comum de cada variável (muito bom!). Juntos, os 5 fatores explicam 88,9% da variabilidade inicial dos dados.

Algumas variáveis (Organização, Experiência, Carta) apresentam cargas fatoriais altas com mais de um fator (não é bom!), indicando a necessidade de realizar uma rotação dos fatores.

Cargas Fatoriais Não Rotacionadas e Itens Comuns Ordenados

Variável	Fator1	Fator2	Fator3	Fator4	Fator5	Comum
Potencial	0.814	0.290	-0.326	0.167	-0.068	0.884
Adeq Cargo	0.813	0.078	-0.029	0.365	0.368	0.937
Adeq Empresa	0.802	-0.060	0.048	0.428	0.306	0.925
Simpatia	0.739	-0.295	-0.117	-0.346	0.249	0.828
Hist Escolar	0.726	0.336	-0.326	0.104	-0.354	0.883
Aparência	0.719	-0.271	-0.163	-0.400	-0.148	0.799
Autoconfiança	0.719	-0.262	-0.294	-0.409	0.175	0.870
Comunicação	0.712	-0.446	0.255	0.229	-0.319	0.925
Currículo	0.709	0.298	0.465	-0.343	-0.022	0.926
Organização	0.706	-0.540	0.140	0.247	-0.217	0.918
Experiência	0.644	0.605	-0.182	-0.037	-0.092	0.823
Carta	0.625	0.327	0.654	-0.134	0.031	0.945
Variância	6.3876	1.4885	1.1045	1.0516	0.6325	10.6647
% de Var	0.532	0.124	0.092	0.088	0.053	0.889

Repetindo a análise marcando em **Tipo de Rotação** a opção **Varimax** obtemos:

Cargas Fatoriais Rotacionadas e Itens Comuns Ordenados

Variável	Fator1	Fator2	Fator3	Fator4	Fator5	Comum
Autoconfiança	0.870	0.218	0.141	0.083	0.197	0.870
Simpatia	0.809	0.101	0.198	0.195	0.292	0.828
Aparência	0.750	0.286	0.355	0.170	-0.025	0.799
Hist Escolar	0.191	0.870	0.243	0.116	0.129	0.883
Experiência	0.150	0.791	-0.103	0.353	0.200	0.823
Potencial	0.278	0.774	0.177	0.118	0.403	0.884
Comunicação	0.223	0.130	0.878	0.213	0.206	0.925
Organização	0.307	0.086	0.854	0.082	0.284	0.918
Carta	0.063	0.158	0.173	0.917	0.216	0.945
Currículo	0.304	0.270	0.130	0.858	0.096	0.926
Adeq Cargo	0.232	0.352	0.219	0.211	0.817	0.937
Adeq Empresa	0.199	0.262	0.371	0.191	0.802	0.925
Variância	2.4442	2.4355	2.0175	1.9343	1.8332	10.6647
% de Var	0.204	0.203	0.168	0.161	0.153	0.889

Nota: As variáveis aparecem em uma outra ordem, mas suas communalidades não se alteram

Avaliando as cargas mais altas das variáveis com cada um dos fatores o Minitab sugere nomear os fatores como:

- **Qualidades Pessoais:** as variáveis Autoconfiança (0,870), Simpatia (0,809) e Aparência (0,750) têm cargas fatoriais altas e positivas no **Fator 1**.
- **Qualificações:** as variáveis Histórico escolar (0,870), Experiência (0,791) e Potencial (0,774) têm cargas fatoriais altas e positivas no **Fator 2**.
- **Habilidades Profissionais:** as variáveis Comunicação (0,878) e Organização (0,854) têm cargas fatoriais altas e positivas no **Fator 3**.
- **Habilidades de Escrita:** as variáveis Carta (0,917) e Currículo (0,858) têm cargas fatoriais altas e positivas no **Fator 4**.

- **Adequação à Empresa e ao Cargo:** Adequação ao Cargo (0,817) e Adequação à Empresa (0,802) têm cargas fatoriais altas e positivas no **Fator 5**.

Realizando a análise novamente e em **Resultados > Exibição de resultados** marcar **Coefficientes de escore de fator e cargas fatoriais**, obtemos também, para cada fator, os coeficientes que podem ser usados para calcular os escores fatoriais de cada indivíduo em cada fator.

Coefficientes de Escore do Fator

Variável	Fator1	Fator2	Fator3	Fator4	Fator5
Hist Escolar	-0.127	0.574	0.169	-0.141	-0.283
Aparência	0.373	0.070	0.106	-0.025	-0.334
Comunicação	-0.162	0.007	0.628	0.021	-0.187
Adeq Empresa	-0.091	-0.134	-0.025	-0.053	0.624
Experiência	-0.064	0.422	-0.178	0.078	-0.072
Adeq Cargo	-0.042	-0.094	-0.171	-0.049	0.661
Carta	-0.139	-0.142	-0.007	0.623	0.004
Simpatia	0.477	-0.213	-0.187	0.001	0.136
Organização	-0.078	-0.046	0.559	-0.086	-0.064
Potencial	-0.043	0.381	-0.026	-0.157	0.074
Currículo	0.043	-0.054	-0.060	0.550	-0.162
Autoconfiança	0.533	-0.084	-0.202	-0.098	0.020

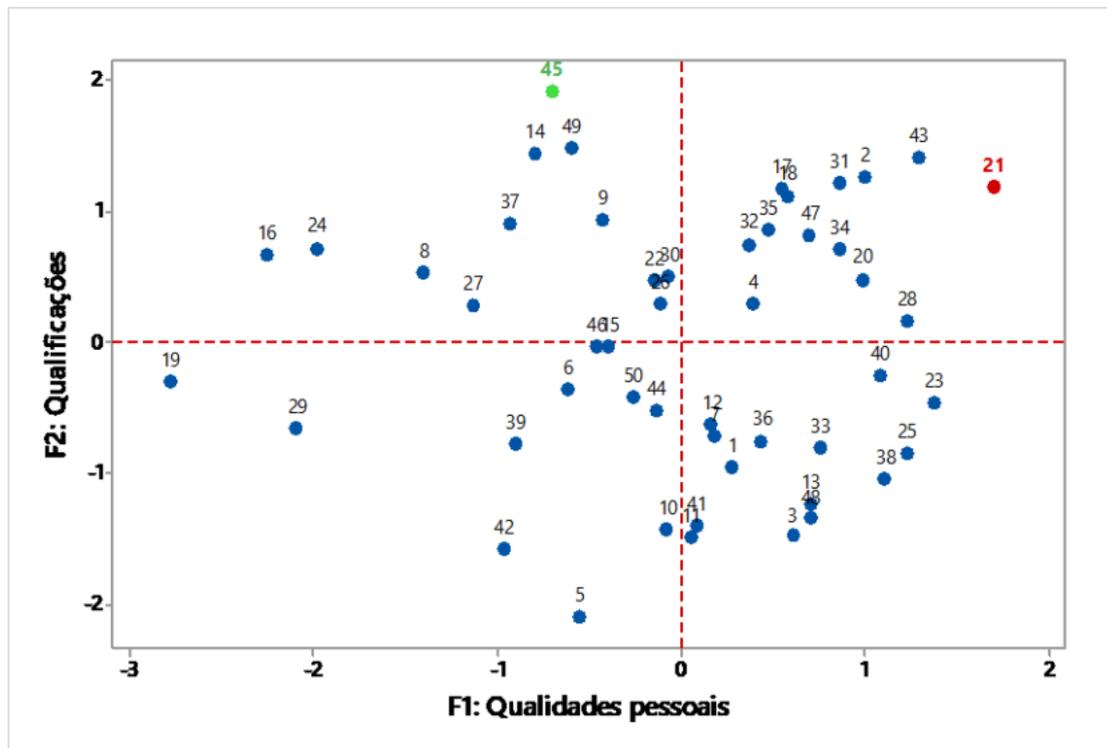
O Minitab pode calcular os escores dos candidatos em cada um dos fatores. Repita a análise e em **Armazenamento > Escores** digite:

Escore1 Escore2 Escore3 Escore4 Escore5.

Exemplo de uso dos escores fatoriais: Para selecionar algum indivíduo, identificado pelo número de linha, com as melhores qualidades pessoais (Fator 1) e mais qualificado (Fator 2), podemos identificá-lo no gráfico de dispersão com os escores dos fatores 1 e 2, lembrando que escores mais altos e positivos indicam indivíduos mais qualificados.

O indivíduo 21 apresenta as melhores qualidades pessoais (Fator 1) e é bem qualificado. O candidato 45 é o mais bem qualificado, mas tem um baixo escore em qualidades pessoais.

Nota: Os escores fatoriais podem ser usados em outros estudos, por exemplo, como variável resposta ou regressora em estudos de regressão



Exemplo 3. (Mingoti, S.A. 2005 – pág. 108). Pesquisa de mercado para avaliar a aceitação pelo consumidor de um novo produto comestível.

Amostra de 200 consumidores: 103 do Sexo feminino e 97 do Sexo masculino.

Notas de 1 a 5 em 7 atributos: Sabor, Aroma, Cor, Textura, Utilidade, Locais de compra e Embalagem

Dados na planilha **AF - Aceitação**.

Correlação linear de Pearson:

	Sabor	Aroma	Cor	Textura	Utilidade	Locais
Aroma	0.969					
Cor	0.801	0.711				
Textura	0.960	0.991	0.696			
Utilidade	0.050	0.037	0.138	0.029		
Locais	0.056	0.046	0.096	0.045	0.833	
Embalagem	0.103	0.090	0.163	0.087	0.693	0.530

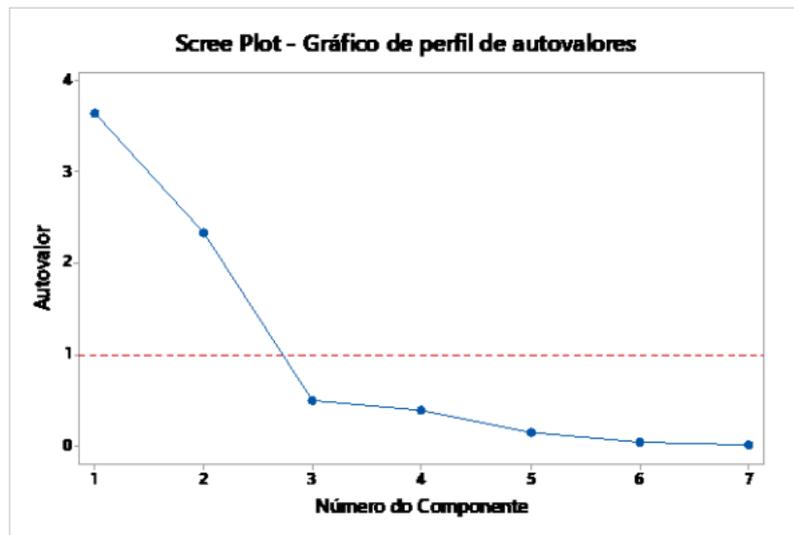
Avaliando os valores numéricos das correlações, podemos visualizar dois grupos de variáveis:

(1) Sabor, Aroma, Cor e Textura

(2) Utilidade, Locais de compra e Embalagem

Autoanálise (Autovalores e Autovetores) da Matriz de Correlação

Autovalores	3.6338	2.3322	0.4864	0.3780	0.1337	0.0276	0.0083
Proporção	0.519	0.333	0.069	0.054	0.019	0.004	0.001
Acumulado	0.519	0.852	0.922	0.976	0.995	0.999	1.000



O quadro com os autovalores da matriz de correlação calculados com os dados padronizados, indica que para explicar mais de 90% da variabilidade inicial dos dados são necessários 3 componentes.

Já o *scree-plot* sugere a necessidade de somente 2 componentes, que juntos explicam 85,2% da variabilidade inicial dos dados.

Decisão: Vamos realizar a análise fatorial com 2 fatores, utilizando o método dos componentes principais para estimação das cargas fatoriais.

Em **Multivariado > Análise fatorial** selecionamos as variáveis Sabor, Aroma, Cor, Textura, Utilidade, Locais de compra e Embalagem; em **Número de fatores para extrair** indicamos **2**; em **Método de extração** marcamos **Análise de componentes principais** e em **Tipo de rotação**, marcamos **Nenhuma**.

Cargas Fatoriais Não Rotacionadas e Itens Comuns

Variável	Fator1	Fator2	Comum
Sabor	0.974	-0.174	0.979
Aroma	0.959	-0.188	0.955
Cor	0.841	-0.051	0.710
Textura	0.952	-0.192	0.944
Utilidade	0.217	0.928	0.908
Locais	0.204	0.868	0.794
Embalagem	0.252	0.783	0.677
Variância	3.6338	2.3322	5.9660
% de Var	0.519	0.333	0.852

Na coluna **Comum** percebe-se que os 2 fatores explicam 71% ou mais da variabilidade comum de cada variável (muito bom!) e que, juntos, os 2 fatores explicam 85,2% da variabilidade inicial dos dados.

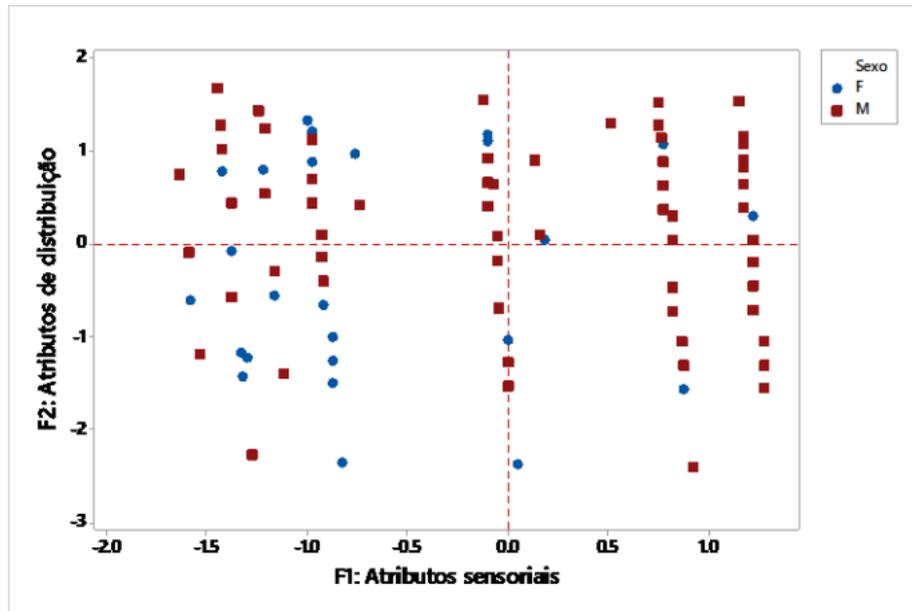
Como todas as variáveis apresentam cargas fatoriais maiores com um único fator, não existe a necessidade de tentar uma rotação de eixos para facilitar a interpretação dos fatores.

Fator 1: Índice de opinião do consumidor sobre os atributos sensoriais do produto.

Como as correlações desses atributos com o Fator 1 são altamente positivas, consumidores que deram notas altas para esses atributos terão os maiores escores neste fator.

Fator 2: Índice de opinião do consumidor sobre os atributos associados à distribuição do produto.

Note que ambos os fatores representam critérios de julgamento utilizados pelos consumidores na avaliação do novo produto.



No gráfico dos escores dos dois primeiros fatores percebe-se maior quantidade de homens com escores positivos no primeiro quadrante (valores positivos de F1 e F2).

Exemplo 4. (Manly, 2008) Força de trabalho de empregados (%) em nove diferentes grupos de indústrias em países da Europa, com o intuito de ajudar a entender o relacionamento entre os países.

Variáveis: AGR: Agricultura, MIN: Mineração, FAB: Fabricação, FEA: Fornecimento de energia e água, CON: Construção, SER: Serviços, FIN: Finanças, SSP: Serviços sociais e pessoais e TC: Transportes e comunicações.

Amostra: 30 países da Europa.

1) Carregamos a planilha **Cluster Empregos**. Para obter a matriz de correlações entre as 9 variáveis, escolhemos: **Estat > Estatísticas Básicas > Correlação**; selecionamos todas as variáveis e desmarcamos a opção **Exibir valor-p**.

Correlações								
	AGR	MIN	FAB	FEA	CON	SER	FIN	SSP
MIN	0.316							
FAB	-0.254	-0.672						
FEA	-0.382	-0.387	0.388					
CON	-0.349	-0.129	-0.034	0.165				
SER	-0.605	-0.407	-0.033	0.155	0.473			
FIN	-0.176	-0.248	-0.274	0.094	-0.018	0.379		
SSP	-0.811	-0.316	0.050	0.238	0.072	0.388	0.166	
TC	-0.487	0.045	0.243	0.105	-0.055	-0.085	-0.391	0.475

Conteúdo da Célula
Correlação de Pearson

Percebe-se a existência de poucas correlações altas entre as variáveis, indicando que a AF pode não ser efetiva.

2) Realizamos a análise de componentes principais (ACP) para decidir sobre o número de CP's (fatores) escolhendo: **Estat > Multivariado > Componentes Principais...**; selecionamos todas as **Variáveis** e marcamos a opção **Tipo de Matriz: Correlação**.

Autoanálise (Autovalores e Autovetores) da Matriz de Correlação

Autovalor	3.1123	1.8092	1.4962	1.0634	0.7103	0.3113	0.2934	0.2038	0.0000
Proporção	0.346	0.201	0.166	0.118	0.079	0.035	0.033	0.023	0.000
Acumulado	0.346	0.547	0.713	0.831	0.910	0.945	0.977	1.000	1.000

Escolhemos quatro componentes (autovalores maiores que 1) que explicam, juntos, 83,1% da variabilidade total dos dados padronizados.

3) Para realizar a Análise Fatorial inicial com 4 fatores (sem rotação) escolhemos: **Estat > Multivariado > Análise Fatorial**; selecionamos todas as variáveis; em **Número de fatores para extrair = 4**; em **Método de Extração > Análise de componentes principais** e o **Tipo de Rotação: Nenhum**.

Cargas Fatoriais Não Rotacionadas e Itens Comuns

Variável	Fator1	Fator2	Fator3	Fator4	Comum
AGR	-0.902	0.032	-0.341	-0.017	0.932
MIN	-0.662	-0.001	0.630	-0.117	0.848
FAB	0.434	-0.581	-0.614	-0.060	0.907
FEA	0.558	-0.147	-0.359	-0.024	0.462
CON	0.391	0.326	0.087	-0.807	0.918
SER	0.673	0.549	0.080	-0.174	0.791
FIN	0.231	0.744	-0.117	0.504	0.875
SSP	0.755	-0.074	0.441	0.327	0.877
TC	0.362	-0.695	0.505	0.043	0.871
Variância	3.1123	1.8092	1.4962	1.0634	7.4812
% de Var	0.346	0.201	0.166	0.118	0.831

- As cargas fatoriais iguais ou superiores a 0,50 (valor absoluto) estão marcadas em negrito.

- As comunalidades (coluna **Comum**) são relativamente altas, com exceção da variável FEA (Fornecimento de energia e água), que apresenta comunalidade menor que 0,50, indicando que os quatro fatores explicam pouco da sua variabilidade.

Conclusão: Grande parte da variância de cada uma das variáveis está contida nos quatro fatores comuns (*bom!*).

- Uma propriedade indesejável da análise inicial: cinco das nove variáveis são fortemente relacionadas com mais de um fator, o que dificulta a interpretação.

Sugestão: Realizar uma rotação de fatores, como a **varimax**.

4) Para realizar a rotação *varimax* voltamos na janela: **Estat > Multivariado > Análise Fatorial** e marcamos **Tipo de Rotação: Varimax**.

Para facilitar a interpretação dos fatores, é comum marcarmos as cargas fatoriais mais altas (superiores a 0.5)

Cargas Fatoriais Rotacionadas e Itens Comuns Ordenados

Variável	Fator1	Fator2	Fator3	Fator4	Comum
SSP	0.911	-0.124	0.174	-0.040	0.877
AGR	-0.851	0.266	-0.097	0.357	0.932
TC	0.726	-0.034	-0.568	0.142	0.871
FAB	0.032	-0.892	-0.319	0.093	0.907
MIN	-0.108	0.860	-0.296	0.099	0.848
FEA	0.192	-0.637	0.036	-0.137	0.462
FIN	0.078	-0.004	0.932	0.012	0.875
CON	0.018	-0.042	-0.080	-0.954	0.918
SER	0.349	-0.153	0.479	-0.645	0.791
Variância	2.2593	2.0521	1.6571	1.5126	7.4812
% de Var	0.251	0.228	0.184	0.168	0.831

Mesmo com a marcação das cargas fatoriais mais altas, fica difícil dar nomes aos fatores.

- **Fator 1:** carga negativa alta com AGR (Agricultura, florestal e pesca) e cargas positivas e altas com SSP (Serviços sociais e pessoais) e TC (Transporte e comunicações).
- **Fator 2:** carga positiva alta com MIN (Mineração e exploração de pedreiras) e cargas negativas com FAB (Fabricação) e FEA (Fornecimento de energia e água).
- **Fator 3:** carga positiva alta com FIN (Finanças) e carga negativa alta com TC (Transporte e comunicações).
- **Fator 4:** cargas negativas altas com CON (Construção) e SER (Serviços).

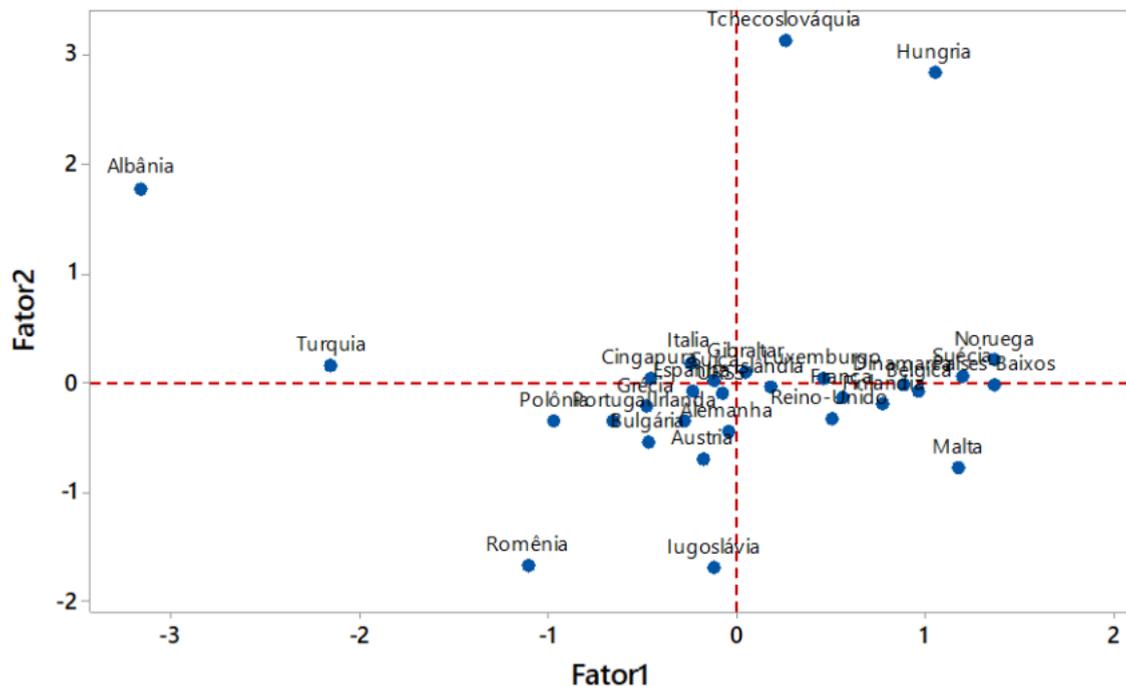
Mesmo sem nomear os fatores, podemos comparar os países quanto aos seus escores positivos ou negativos em cada um dos fatores. Os sinais das cargas são importantes na hora de tirar conclusões.

5) Para calcular os escores dos países em cada um dos fatores, na janela **Stat > Multivariado > Análise Fatorial** selecionamos **Armazenamento** e em **Escores** digitamos os nomes Fator1, Fator2, Fator3 e Fator 4.

Podemos construir gráficos de dispersão com o intuito de identificar os países com escores mais altos e mais baixos nesses fatores.

Para desenhar o gráfico do Fator2 \times Fator1, por exemplo, selecionamos **Gráfico > Gráfico de Dispersão > Simples**; em **Variáveis Y** selecionamos a coluna **Fator2** e em **Variáveis X**, a coluna **Fator1**.

Gráfico dos escores do Fator2 versus Fator1

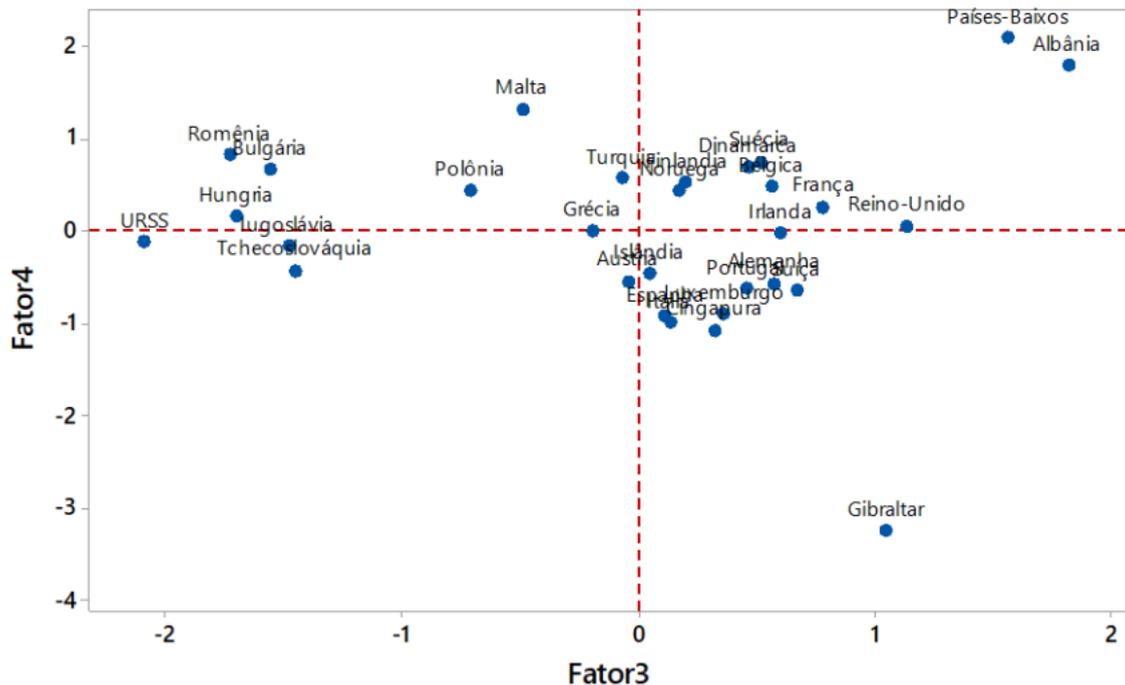


Comentários:

- Os escores negativos e extremos de Albânia e Turquia no **Fator1** enfatizam que esses países tinham mais pessoas empregadas na Agricultura do que em Serviços, Transporte e Comunicação.
- Os escores positivos e altos da Tchecoslováquia, Hungria e Albânia no **Fator2** indicam níveis bem maiores de emprego em Mineração do que Fabricação, em contraste com a Romênia e a Iugoslávia, que apresentam escores negativos e extremos no Fator 2.

Podemos repetir a operação para construir um gráfico do Fator4 \times Fator3.

Gráfico dos escores do Fator4 versus Fator3



Comentários:

- Os escores negativos e extremos do **Fator 3** indicam que URSS, Bulgária, Tchécoslováquia, Hungria, Romênia e Iugoslávia tinham mais pessoas empregadas em Transporte e Comunicação do que em Finanças, em contraste com a Albânia e Países-Baixos, onde acontece o contrário.
- Os escores do **Fator 4** contrastam Gibraltar, que apresenta os mais baixos índices de emprego na Construção e Serviços, com os Países-Baixos, Albânia e Malta, que apresentam escores altos e positivos.

Bibliografia consultada ou referenciada:

Artes, R. Aspectos estatísticos da análise fatorial de escalas de avaliação. **Rev. Psiq. Clin.** 25 (5) Edição especial: 223-228, 1998.

Hair, J.F.Jr.; Anderson, R.E.; Tatham, R.L; Black, W.C. **Multivariate Data Analysis** (with readings), 4 ed. Prentice Hall, Englewood Cliffs, 1995.

Johnson, R.A.; Wichern, D.W. **Applied Multivariate Statistical Analysis**. Prentice Hall, Englewood Cliffs, 1992.

Mingoti, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2005.

Reis, E. **Estatística Multivariada Aplicada**. Edições Sílabo, Lisboa, 1997.

Spearman, C. General Intelligence Objectively Determined and Measured. **American Journal of Psychology** 15:201-293, 1904.