SANTANA, A. C. **Elementos de economia, agronegócio e desenvolvimento local.** Belém, PA: UFRA, 2005. 206p. Cap. 6.

CAPÍTULO 6
APÊNDICE – ANÁLISE FATORIAL

1. INTRODUÇÃO

Antes de entrar no estudo da análise fatorial, alguns pontos devem ser referidos, a título de chamar a atenção do leitor para que não cometa erros ou venha a ter problemas que podem ser evitados com relativa facilidade.

Toda análise multivariada necessita de um exame meticuloso da massa de dados, pois o poder analítico dos instrumentos de análise multivariada que podem ser postos a disposição do pesquisador depende diretamente da qualidade e consistência dos dados.

O exame dos dados antes da aplicação de uma das técnicas da análise multivariada (componentes principais ou análise fatorial) permite que se tenha uma visão crítica dos dados. Este exame envereda por alguns caminhos para saber a forma da distribuição dos dados (histogramas), a relação entre as variáveis (gráfico de dispersão, correlação), identificação de autovalores, erros de medida e assim por diante.

É importante, também, antecipar algumas termos-chave para facilitar o entendimento da nomenclatura técnica que envolve a análise multivariada. A lista de termos-chave foi extraída de Hair et al. (2005, p.90-91).

a) **Análise de componente principal:** modelo fatorial em que os fatores são baseados na variância total da nuvem de dados.

b) **Análise fatorial:** analisa relações entre variáveis para identificar grupos de variáveis que formam dimensões latentes (fatores).

c) **Autovalor:** soma em coluna de cargas fatoriais ao quadrado para um dado fator; também conhecido como **raiz latente** e representa a magnitude de variância explicada por um fator específico.

d) **Cargas fatoriais:** correlação entre as variáveis originais e os fatores, considerado a chave para o entendimento da natureza de um fator; as cargas fatoriais ao quadrado indicam o porcentual da variância de uma variável original que é explicado por um fator.

e) **Comunalidade:** magnitude total da variância que uma variável original compartilha com todas as outras variáveis incluídas na análise.

f) **Homocedasticidade:** quando a variância dos termos de erro aparece constante ao longo de um domínio de variáveis explicativas, diz-se que os dados são homocedásticos.

g) **Linearidade:** usada para expressar o conceito de que o modelo possui as propriedades de aditividade e homogeneidade. Em geral, os modelos lineares prevêem valores que recaem em uma linha reta que tem uma mudança com unidade constante (coeficiente angular) da variável dependente em relação a uma mudança com unidade constante da variável independente.

h) **Matriz fatorial:** tabela de cargas fatoriais de todas as variáveis sobre cada fator.

i) **Multicolinearidade:** grau em que uma variável pode ser explicada pelas outras variáveis na análise.

j) **Ortogonal:** independência matemática (sem correlação) de eixos fatoriais, um em relação ao outro (ângulos retos 90 graus).

k) **Rotação fatorial:** processo de ajuste dos eixos fatoriais para conseguir uma solução fatorial mais simples e pragmaticamente mais significativa.

l) **Varimax:** é um dos métodos de rotação fatorial ortogonal mais popular.

2. O QUE É ANÁLISE FATORIAL

Análise fatorial é um nome genérico dado a uma classe de métodos estatísticos multivariados cujo propósito fundamental é definir a estrutura subjacente em uma matriz de dados. De modo geral, a análise fatorial é utilizada para analisar a estrutura das inter-relações (correlações) entre um grande número de variáveis, definindo um conjunto de dimensões latentes comuns que facilitam a compreensão da estrutura da nuvem de dados, chamadas de **fatores**. Com a análise fatorial, inicialmente pode-se identificar as dimensões isoladas da estrutura dos dados e então determinar o grau em que cada variável é explicada por cada dimensão ou fator. Depois dessa etapa, a análise fatorial pode ser empregada para resumir ou para reduzir a massa de dados.

Com o resumo dos dados, a análise fatorial gera dimensões latentes ou fatores que quando interpretadas e compreendidas, descrevem os dados a partir de um número muito menor de conceitos do que as variáveis individuais originais.

A redução de dados, por ser conseguida calculando escores para cada fator e substituindo as variáveis originais pelos referidos fatores.

2.1 OBJETIVO DA ANÁLISE FATORIAL

O objetivo da análise fatorial é encontrar uma maneira de condensar (resumir) a informação contida em diversas variáveis originais em um conjunto menor de novas dimensões compostas ou variáveis estatísticas (fatores) com uma perda mínima de informação, ou seja, buscar e definir os constructos fundamentais ou dimensões assumidas como inerentes às variáveis originais. De modo geral, a análise fatorial pode satisfazer a um dos objetivos: identificação da estrutura por meio do resumo de dados ou redução de dados.

O primeiro objetivo diz respeito à identificação da estrutura de relações entre as variáveis, examinando-se as correlações entre elas. O segundo objetivo implica em identificar variáveis representativas de um conjunto muito maior de variáveis para uso em análises multivariadas subseqüentes.

2.2 TAMANHO DA AMOSTRA

No que tange ao tamanho da amostra, dificilmente se faz uma análise fatorial com menos de 50 observações. De preferência, o tamanho da amostra deve ser igual ou superior a 100. Como regra geral, deve ter-se pelo menos cinco vezes mais observações do que o número de variáveis a serem analisadas e o número mais seguro deveria atingir uma proporção de 10 para um.

2.3 SUPOSIÇÕES NA ANÁLISE FATORIAL

Do ponto de vista estatístico, os desvios da normalidade, da homocedasticidade e da linearidade aplicam-se apenas no nível em que elas diminuem as correlações observadas. A normalidade é necessária somente se um teste estatístico for aplicado para a significância dos fatores, mas estes testes raramente são usados. Em geral, um pouco de **multicolinearidade** é desejável, pois o objetivo é identificar conjuntos de variáveis inter-relacionadas.

Além da base estatística para as correlações da matriz de dados, deve-se garantir que a matriz de dados tenha correlações suficientes para justificar a aplicação da análise fatorial. Se a análise não revela um número substancial de correlações superiores a 0,30, então a análise fatorial provavelmente não é adequada aos dados. Se existem fatores “verdadeiros” nos dados, a correlação parcial deverá ser pequena ou não significativa, pois a variável pode ser explicada pelos fatores (variáveis estatísticas com cargas para cada variável). Se as correlações parciais são altas, então não há fatores latentes "verdadeiros" e a análise fatorial é inadequada.

2.4 ANÁLISE DE FATORES VERSUS ANÁLISE DE COMPONENTES

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 propósito de previsão. Em contraste, a análise de fatores comuns é usada principalmente para identificar fatores ou dimensões latentes que reflitam o que as variáveis têm em comum.

A variância comum é definida como a variância de uma variável que é compartilhada com todas as outras variáveis da análise. As **comunalidades** são estimativas da variância compartilhada, ou comum, entre as variáveis. A análise de componentes considera a variância total e determina fatores que contêm pequenas proporções de variância única e, em alguns casos, variância do erro.

A escolha entre um modelo e outro é baseada no objetivo da análise fatorial ou no montante de conhecimento prévio sobre a variância das variáveis.

O modelo fatorial de componentes é apropriado quando a preocupação principal é a previsão ou o número mínimo de fatores necessários para explicar a parte máxima da variância representada no conjunto original de variáveis ou quando conhecimento anterior sugere que as variâncias específicas e de erro representam uma proporção relativamente pequena da variância total.

Por outro lado, quando o objetivo principal é identificar as dimensões ou constructos latentes representados nas variáveis originais e o pesquisador tem pouco conhecimento sobre o montante da variância específica e do erro e, portanto, deseja eliminar essa variância, o modelo de fatores comuns é mais adequado. Em função de suas suposições mais restritivas e uso apenas de dimensões latentes (variância compartilhada), é visto como algo teoricamente mais fundamentado.

2.5 CRITÉRIOS PARA A EXTRAÇÃO DE FATORES

Qual o número de fatores a serem extraídos? Alguns critérios são utilizados para a extração de fatores. Os principais são: raiz latente, *a priori* e porcentagem da variância.

a) **Critério da raiz latente:** a técnica mais comumente utilizada é de raiz latente. Esta técnica parte do princípio de que qualquer fator individual deve explicar a variância de pelo menos uma variável para que seja mantido para interpretação. Cada variável contribui com um valor 1 do autovalor total. Com efeito, apenas os fatores que têm raízes latentes ou autovalores maiores que 1 são considerados significantes e os demais fatores com autovalores menores do que 1 são considerados insignificantes e descartados. Este critério é mais confiável quando o número de variáveis está entre 20 e 50.

b) **Critério *a priori***: o pesquisador já sabe quantos fatores extrair antes de empreender a análise fatorial. Este tratamento é útil quando se testa uma teoria ou hipótese sobre o número de fatores a serem extraídos. Também se justifica quando se trate de uma tentativa de repetir o trabalho de outro pesquisador.

c) **Critério da porcentagem da variância:** é baseado na conquista de um porcentual cumulativo especificado da variância total extraída por fatores sucessivos. O objetivo é garantir é garantir significância prática para os fatores determinados, garantindo que expliquem pelo menos um montante especificado de variância. Em ciências naturais, o procedimento de obtenção de fatores não deve parar até que os fatores extraídos explicarem pelo menos 95% da variância. Por outro lado, nas ciências sociais, é comum considerar uma solução que explique 60% da variância total como satisfatória.

2.6 ROTAÇÃO DE FATORES

A solução de fatores não-rotacionados extraem fatores na ordem de sua importância. O primeiro fator explica a maior parcela da variância total. O segundo e os seguintes são extraídos baseados na variância residual.

A rotação fatorial é um processo em que os eixos de referência dos fatores são rotacionados em torno da origem até que alguma outra posição seja alcançada. O efeito final de rotacionar a matriz fatorial é redistribuir a variância dos primeiros fatores para os últimos com o objetivo de atingir um padrão fatorial mais simples e teoricamente mais significativo.

2.7 MODELO BÁSICO DE ANÁLISE FATORIAL

O modelo de análise fatorial guarda grande semelhança com o modelo de componentes principais. O modelo de componentes principais com **m** componentes e **p** variáveis (**q** < **p**), pode ser escrito como na equação 1 (DILLON; GOLDSTEIN, 1984).

$$\begin{aligned} CP_1 &= \gamma_{11}X_1 + \gamma_{12}X_2 + \ldots + \gamma_{1p}X_p \\ CP_2 &= \gamma_{21}X_1 + \gamma_{22}X_2 + \ldots + \gamma_{2p}X_p \\ &\vdots \\ CP_q &= \gamma_{m1}X_1 + \gamma_{m2}X_2 + \ldots + \gamma_{qp}X_p \end{aligned} \tag{1}$$

Em que:

CP_i = são as i-esimas componentes principais (i = 1, 2, ..., q);

γ_{ij} = são os coeficientes relacionados a cada variável;

X_j = são as j-ésimas variáveis (j = 1, 2, ..., p).

Por outro lado, o modelo básico de análise fatorial, expressa cada variável em termos dos fatores latentes comuns e de um fator único ou fator específico. A representação algébrica do modelo é dada pela equação 2.

$$\begin{aligned} X_1 &= \lambda_{11}FC_1 + \lambda_{12}FC_2 + \ldots + \lambda_{1q}FC_q + \varepsilon_1 \\ X_2 &= \lambda_{21}FC_1 + \lambda_{22}FC_2 + \ldots + \lambda_{2q}FC_q + \varepsilon_2 \\ &\vdots \\ X_p &= \lambda_{p1}FC_1 + \lambda_{p2}FC_2 + \ldots + \lambda_{pq}FC_q + \varepsilon_{mp} \end{aligned} \tag{2}$$

Em que:

X_i = são as i-esimas variáveis (i = 1, 2, ..., p);

λ_{ij} = são os coeficientes relacionados a cada fator comum;

FC_j = são os j-ésimos fatores comuns (j = 1, 2, ..., q);

ε_i = são os i-ésimos fatores específicos.

O modelo básico de fatores comuns é usualmente expresso na forma matricial como na equação 3.

$$X = \Lambda F + \mathrm{E} \tag{3}$$

Em que:

X = é o p-dimensional vetor de variáveis originais, X' = (x_1, x_2, ..., x_p);

F = é o q-dimensional de variáveis não-observáveis ou fatores comuns, F' = (f_1, f_2, ..., f_q);

E = é o p-dimensional vetor de variáveis não observáveis ou fatores únicos, E' = (e_1, e_2, ..., e_p);

Λ = é a matriz (p,q) de constantes desconhecidas, ou cargas fatoriais.

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}; \Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1q} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2q} \\ \vdots & \vdots & & \vdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pq} \end{bmatrix}; F = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_q \end{bmatrix}; \mathrm{E} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{bmatrix}$$

No modelo de análise fatorial pressupõe-se que os fatores específicos são ortogonais entre si e com todos os fatores comuns, ou seja, a matriz de variância é dada por:

$$E(ee') = \Psi = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix},$$

e a covariância é dada por:

$$\operatorname{cov}(e, f') = 0$$

A matriz de covariância de resposta ao vetor *X*, associada ao modelo (3), denotada por Σ_{xx}, pode ser expressa como:

$$\Sigma_{xx} = \Lambda\Omega\Lambda' + \Psi \tag{4}$$

em que Λ e Ψ são foram definidas acima e Ω é dada por:

$$\Omega = \begin{bmatrix} 1 & 0 & \cdots & \\ \omega_{21} & 1 & \cdots & \\ \vdots & \vdots & & \vdots \\ \omega_{q1} & \omega_{q2} & \cdots & 1 \end{bmatrix}$$

Os elementos de Ω são as covariâncias (correlações) entre os fatores comuns. Nota-se que cada coluna de Λ pode ser selecionada arbitrariamente, assume-se, sem perda de generalidade, que cada fator comum apresenta variância unitária (diagonal principal da matriz Ω). Como os fatores são ortogonais, ou seja, não são correlacionados entre si, Ω = I e a equação 4, torna-se:

$$\Sigma_{xx} = \Lambda\Lambda' + \Psi \tag{5}$$

Alternativamente, o modelo 3 pode ser reescrito da seguinte forma:

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1q} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2q} \\ \vdots & \vdots & & \vdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pq} \end{bmatrix} \cdot \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_q \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{bmatrix} \tag{6}$$

ou

$$X_i = \sum_{j=1}^{q} \lambda_{ij} f_j + e_i \tag{7}$$

O conjunto de equações de 4 é chamado de fator padrão. Para simplicidade do fator padrão é tipicamente mostrado na forma tabular apenas as cargas fatoriais listados.

Cada equação do modelo 7 pode ser particionado em duas partes não correlacionadas, dadas por:

$$X_i = c_i + e_i \tag{7}$$

em que $c_i = \lambda_{i1}f_1 + \lambda_{i2}f_2 + \ldots + \lambda_{iq}f_q$ é a parte de cada variável que é comum às outros *p-1* variáveis, e e_i é a parte de cada variável que é única ou específica.

Dado que as partes comuns e únicas de das variáveis são não-correlacionadas e os fatores comuns têm variância unitária, pode-se particionar a variância de X_i em:

$$\operatorname{var}(X_i) = \operatorname{var}(c_i) + \operatorname{var}(e_i) \tag{8}$$

em que a $var(c_i)$ e $var(e_i)$ representam a variância comum e a variância única de X_i, respectivamente. A variância comum de uma variável é também chamada de comunalidade da variável. A comunalidade de uma variável é a porção da variância total das variáveis associadas aos fatores comuns. Chamando de h_i^2 a comunalidade da i-ésima variável, a variância de X_i pode ser escrita da seguinte forma:

$$\mathrm{var}(X_i) = h^2{}_i + \Psi_i \tag{9}$$

em que var(e_i) = Ψ_i do modelo básico (3). Assim, tem-se que:

$$\mathrm{var}(c_i) = \sum_{j=1}^{q} \lambda^2{}_{ij} = h^2{}_i \tag{10}$$

É, pois, apenas a soma do quadrado dos elementos da i-ésima linha de Λ.

A contribuição total do fator f_j para a variância total do conjunto de variáveis é dada pelo autovalor do fator f_j, computado da seguinte maneira:

$$V_j = \sum_{i=1}^{p} \lambda^2{}_{ij} = \lambda'{}_j \lambda_j \tag{11}$$

em que λ_j é a i-ésima coluna da matriz Λ. A contribuição total de todos os fatores comuns para o total da variância entre todas as variáveis originais é a comunalidade total, definida por:

$$V = \sum_{j=1}^{q} V_j \tag{12}$$

A variância entre todas as variáveis associadas ao fator f_j como porcentagem do total de todos os fatores é dado por:

$$V_c = \frac{V_j}{V} \tag{13}$$

Finalmente, a variância total pode ser escrita da seguinte forma:

$$VT = tr(\Sigma_{xx}) = \sum_{j=1}^{q} Vj + \sum_{i=1}^{p} \Psi_i = \sum_{i=1}^{p} \sum_{j=1}^{q} \lambda^2{}_{ij} + \sum_{i=1}^{p} \Psi_i \tag{14}$$

2.8 EXEMPLO DE APLICAÇÃO

Como exemplo de aplicação, apresenta-se o resultado da análise fatorial de unidades de produção em Tomé-Açu com sistemas agroflorestais e com cultivos tradicionais. Conta-se com um conjunto de sete variáveis: valor bruto da produção (VB Produção), área cultivada (Área), mão-de-obra contratada utilizada na produção (Mão-de-obra), valor das máquinas, equipamentos, ferramentas e implementos utilizados na produção (capital), valor dos insumos variáveis utilizados na produção (insumo), depreciação das máquinas, equipamentos e instalações utilizados na produção (depreciação) e a produtividade das unidades de produção (tecnologia).

Essas observações foram obtidas para se estimar uma função de produção e determinar a eficiência técnica e econômica das unidades de produção. Se as variáveis formarem apenas um bloco, pode-se especificar uma função de produção do tipo Cobb-Douglas. Porém, se tais variáveis definirem mais de um fator, pode-se optar por uma função de produção do tipo transcendental, para combinar os dois ou mais fatores.

2.8.1 Adequação dos dados

Inicialmente, avalia-se a viabilidade da adequação da análise fatorial a partir da análise fatorial e do teste Bartlett. A matriz de correlação simples entre as variáveis é apresentada na Tabela 2.8i. Observa-se que 17 das 21 correlações (81%) são significantes ao nível de 0,01. Isto fornece uma base adequada para avançar ao próximo passo, o exame empírico da adequação dos dados para a análise fatorial tanto para a massa de dados como para cada variável considerada na análise.

A adequação da análise é feita por meio do teste Bartlett, que avalia a significância geral da matriz de correlação. Neste caso, as correlações, em geral, são significantes ao nível 0,0001.

Tabela 8.2i
Matriz de correlação simples de Pearson e significância estatística.

Variáveis	**VB Produção**	**Área**	**Mão-de-obra**	**Capital**	**Insumo**	**Depreciação**	**Tecnol.**
VB Produção	1,000						
Área	0,846*	1,000					
Mão-de-obra	0,730*	0,933*	1,000				
Capital	0,735*	0,699*	0,551*	1,000			
Insumo	0,872*	0,777*	0,662*	0,744*	1,000		
Depreciação	-0,125n	-0,262*	-0,254*	-0,062n	0,104n	1,000	
Tecnologia	0,735*	0,690*	0,543*	0,989*	0,759*	-0,013n	1,000

(*) correlação significante a 0,01. (n) correlação não significante.
Teste Bartlett (aproximação qui-quadrado) = 880,96 (0,0001).

2.8.2 Análise fatorial de componentes

A matriz de correlação, como descrito na metodologia, é transformada por meio de um modelo fatorial para gerar a matriz fatorial. As cargas de cada variável associadas aos fatores são interpretadas para identificar a estrutura latente das variáveis, que são a função de produção. O primeiro passo dessa tarefa é selecionar o número de componentes a serem mantidos para a análise seguinte.

A Tabela 2.8ii contém a informação sobre os sete fatores possíveis e seu poder explanatório relativo expresso pelos autovalores. Aplicando o critério da **raiz latente**, duas componentes serão mantidas ou extraídas. Observa-se que os dois fatores explicam 84,86% da variância total da nuvem de dados, o que é satisfatório pelo critério da **porcentagem da variância**. A terceira componente poderia ser extraída se houvesse conveniência teórica para isto, mesmo com um autovalor considerado baixo.

Tabela 8.2ii
Resultados dos autovalores para a extração de fatores componentes.

Componente	Autovalores iniciais (λ_i)		
	Variância total ou **autovalor**	% da Variância total	% da Variância acumulada
1	**4,778**	**68,260**	**68,260**
2	**1,162**	**16,599**	**84,859**
3	0,638	9,118	93,977
4	0,276	3,949	97,926
5	0,098	1,397	99,323
6	0,038	0,539	99,862
7	0,010	0,138	100,000

Eleitos então os fatores, passa-se à sua interpretação. Observa-se, portanto, que até este ponto da análise, não se faz diferença entre as técnicas de análise de componentes e análise de fatores.

2.8.3 Interpretação de fatores

As duas primeiras colunas são os resultados para os dois fatores extraídos, ou seja, as cargas fatoriais para cada variável em cada fator. A terceira coluna fornece a estatística resumo, detalhando o grau em que cada variável é "explicada" pelas duas componentes, denominada de comunalidade. Das duas últimas linhas, a primeira é a soma da coluna de cargas fatoriais ao quadrado (autovalores) e indica a importância relativa de cada fator na explicação da variância associada ao conjunto de variáveis analisado. A soma dos dois fatores são 4,778 e 1,162, respectivamente. Como esperado, a solução fatorial extraiu os fatores na ordem de sua importância, com o fator 1 explicando a maior parcela da variância (68,26%) e o fator 2 muito menos (16,6%). O número 5,94 representa a soma total de cargas fatoriais ao quadrado e indica a parcela total de variância extraída pela solução fatorial.

Tabela 8.2.iii
Matriz fatorial rotacionada Varimax da análise de componentes.

Variáveis	Fatores		Comunalidade
	1	2	
VB Produção	0,908	-0,165	0,851
Área	0,874	-0,392	0,917
Mão-de-obra	0,762	-0,451	0,784
Capital	0,902	0,037	0,816
Insumo	0,926	0,083	0,864
Depreciação	0,022	0,936	0,876
Tecnologia	0,909	0,083	0,833
Soma de quadrado do autovalor	4,778	1,162	5,940
Porcentual do traço (%)	68,26	16,599	84,859

A parcela total da variância explicada pela solução fatorial (5,94) pode ser comparada com a variação total do conjunto de variáveis que é representada pelo traço da matriz fatorial. O traço é a variância total a ser explicada e é igual à soma dos autovalores do conjunto de variáveis (soma da primeira coluna da Tabela 8.2ii), que é igual a 7,0, dado que cada variável tem um autovalor possível igual a 1,0. Os porcentuais de traço explicados por cada um dos dois fatores (68,26% e 16,6%, respectivamente) são mostrados na última linha da Tabela 8.2iii. A soma total dos porcentuais de traço extraído para a solução fatorial, serve como índice para determinar o grau de adequação da solução fatorial em relação ao que todas as variáveis representam. O índice para esta solução mostra que 84,859% da variância total são representados pela informação contida na matriz fatorial da solução em termos dos dois fatores. O índice é considerado alto, e as variáveis estão, como esperado, estreitamente relacionadas umas com as outras.

A soma em linha de cargas fatoriais ao quadrado gera a comunalidade, última coluna da Tabela 2.8iii. Estes números mostram a magnitude da variância em uma variável que é explicada pelos dois fatores tomados juntos. O tamanho da comunalidade é um índice útil para avaliar o quanto de variância em uma dada variável é explicado pela solução fatorial. Comunalidades grandes indicam que uma grande parcela da variância em uma variável foi extraída pela solução fatorial. Uma comunalidade pequena mostra que uma boa parte da variância contida em uma variável não é explicada pelos fatores. Neste caso, as variáveis que apresentam comunalidades inferiores a 0,50 não têm explicação suficiente, podendo ser deixadas fora da solução fatorial.

O passo seguinte é a nomeação dos fatores para a definitiva análise da solução fatorial.

2.8.4 Nomeação de fatores

Sempre que uma solução fatorial satisfatória é obtida, é importante atribuir um significado a ela. O processo envolve substantiva interpretação do padrão das cargas fatoriais para as variáveis, incluindo seus sinais, como esforço para nomear cada fator. Em geral, todas as cargas fatoriais significantes são utilizadas no processo de interpretação, porém, as variáveis com maior carga influenciam mais na seleção de nomes ou rótulo para representar um fator.

A seleção das variáveis significativas que devem fazer parte de um fator é eleita com base na magnitude da carga fatorial. Inicialmente, pode-se adotar um ponto de corte, considerando apenas as cargas acima de 0,30 na Tabela 8.2iii. Depois disso, as variáveis significativas podem ser eleitas, olhando-se da esquerda para a direita ao longo de cada linha e selecionando-se as cargas de maior valor. Adotando este processo, o fator tem seis cargas significativas e o fator 2, uma.

No primeiro fator estão VB Produção, Área, Mão-de-obra, Capital, Insumo e Tecnologia, ambas com sinais positivos e altos, demonstrando que todas as variáveis variam juntas, estando coerente com um processo de produção racional. Isto indica que indica que à medida que essas variáveis aumentam a produção agregada da unidade produtiva também aumenta. Este fator pode ser chamado de **produção racional**, uma vez que pode representar o estádio racional de produção especificado em uma função do tipo Cobb-Douglas.

No fator 2, apenas a variável depreciação está relacionada à componente de capital ou patrimônio da unidade de produção. Esta variável está ligada à gestão patrimonial e seus efeitos se manifestam em longo prazo, de modo que no período da análise sua contribuição para a produtividade dos fatores ainda não esteja plenamente incorporada ao processo produtivo, ou seja, possa está sendo empregada no primeiro estádio da produção, uma vez que o sinal. Este fator pode então ser rotulado de **gestão patrimonial**.

2.8.5 Equações do modelo

A solução fatorial permite apresentar o peso do fator 1 para explicas seis das sete variáveis, de modo que o fator1 explica apenas uma variável.

Variáveis explicadas pelo Fator 1:

$$VBP = 0{,}908 f_1 + e_1$$
$$Area = 0{,}874 f_1 + e_2$$
$$Mão-de-obra = 0{,}762 f_1 + e_3$$
$$Capital = 0{,}902 f_1 + e_4$$
$$Insumo = 0{,}926 f_1 + e_5$$
$$Tecno\log ia = 0{,}901 f_1 + e_6$$

Variável explicada pelo Fator 2: $Depreciação = 0{,}936 f_2 + e_7$

Portanto, a visualização gráfica dos dois fatores está na Figura 2.8i, que indica claramente a disjunção entre os dois fatores.

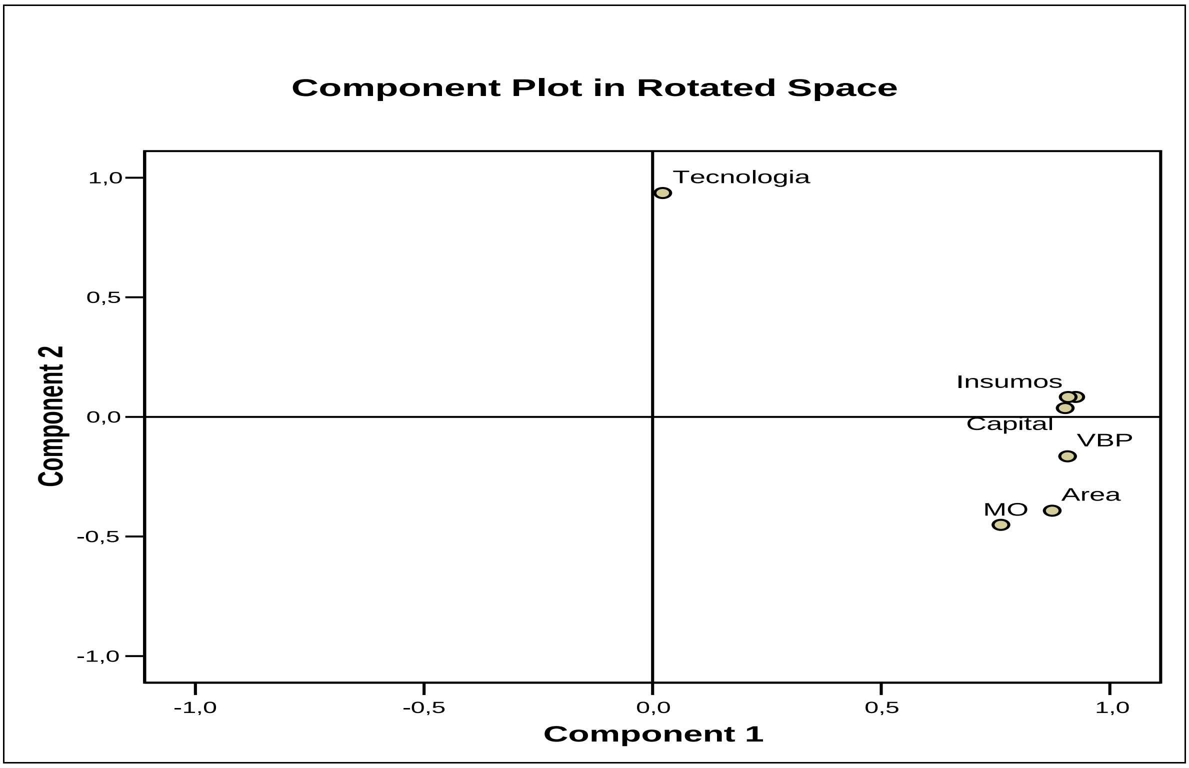

Figura 8.2i – Visualização gráfica dos dois fatores extraídos da solução fatorial.

2.9 CONSIDERAÕES FINAIS

Pelo que se observou nesta análise, há pouca diferença em relação à análise de componentes principais. Nesta, portanto, não há preocupação em rotular as componentes, o que é uma praxe na análise fatorial.

Na análise fatorial, a preocupação está em determinar a explicação de cada variável original pelos fatores comuns e pelos fatores específicos, enquanto que na análise de componentes o interesse está voltado para se expressar cada componente em função do conjunto de variáveis relevantes.

Além destas diferenças, deve-se reforçar que a análise de componentes principais toma por base o total da variância dos dados e a análise fatorial apenas a parcela da variância associada aos fatores selecionados.

2.10 REFERÊNCIAS

DILLON, W.R.; GOLDSTEIN, M. **Multivariate analysis: methods and applications.** New York: John Wiley & Sons, 1984.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis.** Prentice-Hall, 1992.

HAIR JR, J.F., ANDERSON, R.E., TATHAM, R.L., BLACK, W.C. **Análise multivariada de dados.** 5. ed. Porto Alegre: Bookman, 2005.

2.11 EXERCÍCIOSDE APRENDIZAGEM

E1. Qual a diferença entre os métodos da análise de componentes principais e da análise fatorial quanto à variância explicada da massa dos dados que se deseja descrever e analisar?

E2. Com base nos resultados da tabela abaixo sobre as empresas de demoveis da Região Metropolitana de Belém (RMB), indique quais variáveis compõem cada fator e o rotule. Escreva a equação linear atribuída a cada fator (análise de componentes) e a cada variável (análise de fator). Calcule a comunalidade relativa a cada variável e a contribuição total de cada fator para a variância total do conjunto dos fatores. Calcule o porcentual do traço e o índice total de adequação da solução fatorial.

Variáveis	Fator 1	Fator 2	Fator 3	Fator 4
X1. Emprego	0,5098	0,0045	-0,3028	0,0461
X2. Salário médio	0,8662	0,0683	0,0088	0,0644
X3. Subcontratação de empresas	-0,4384	0,2085	0,0256	0,1523
X4. Nível de capacidade ociosa	-0,1708	0,8185	0,0315	-0,0692
X5. Organização social da produção	0,2139	0,6671	0,1862	-0,0006
X6. Crédito de curto prazo (custeio)	0,0603	0,1347	0,8282	-0,0082
X7. Crédito de longo prazo (investimento)	-0,1279	0,0802	0,7964	0,0359
X8. Número de fornecedores	0,8159	-0,0370	-0,0061	0,0822
X9. Mercado local	-0,0752	0,1427	0,1015	-0,7469
X10. Mercado regional	0,1105	-0,0516	0,0974	0,8309
X11. Controle de qualidade dos produtos	-0,1172	0,1582	-0,1245	0,5304
Soma de quadrado do autovalor				
Porcentual de traço (%)				