



Unidade IV – Análise Fatorial

4.1. Introdução

A ideia fundamental da Análise Fatorial ("Factor Analysis") foi proposta por Sperman e por Pearson, no início do século XX, para entender problemas relacionados à psicologia educacional, na tentativa de definir inteligência (MARRIOTT, 1974). Sendo assim essa técnica é muito utilizada na psicometria.

O propósito da técnica é descrever, se possível, a estrutura de covariância entre muitas variáveis em termos de uma poucas, subjacentes, mas não observáveis quantidades aleatórias chamadas *fatores*.

Desse modo na Análise Fatorial (AF) um número relativamente grande de variáveis observadas, são representadas como combinações lineares de um número relativamente pequeno de *variáveis latentes*, *construtos* ou *fatores*. Os fatores podem ser não correlacionados (fatores ortogonais) ou correlacionados (fatores oblíquos). As variáveis são agrupadas por meio de suas correlações, ou seja, aquelas pertencentes a um mesmo grupo serão fortemente correlacionadas entre si, mas pouco correlacionadas com as variáveis de outro grupo. Cada grupo de variáveis representará um fator (JOHNSON & WICHERN, 1988).

A AF é uma boa estratégia se o número de variáveis estudadas é grande, e se deseja simplificar ou melhor estruturar o conjunto de dados, a partir das inter-relações entre tais variáveis. Tais inter-relações podem ser medidas pelas covariâncias ou pelos coeficientes de correlação entre as variáveis.

4.2. Modelo Fatorial ortogonal

Considerando um conjunto de p variáveis, com n observações para cada variável, extraídos de uma população π que possui vetor de médias μ e matriz de covariâncias, Σ , obtém-se a matriz de dados multivariados:

$$\mathbf{X}_{n \times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jp} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots & x_{np} \end{bmatrix}$$

O modelo da análise de fatores supõe que cada variável X_k é linearmente dependente de poucas variáveis aleatórias não observadas F_1, F_2, \dots, F_m ($m < p$) chamadas fatores comuns, e p fontes adicionais de variação $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$, chamadas erros ou, algumas vezes, fatores específicos (JOHNSON & WICHERN, 1988). Em particular, o modelo da análise de fatores pode ser escrito como:

$$\begin{aligned} X_1 - \mu_1 &= a_{11}F_1 + a_{12}F_2 + \cdots + a_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= a_{21}F_1 + a_{22}F_2 + \cdots + a_{2m}F_m + \varepsilon_2 \\ &\vdots \\ X_p - \mu_p &= a_{p1}F_1 + a_{p2}F_2 + \cdots + a_{pm}F_m + \varepsilon_p \end{aligned}$$

ou seja,

$$X_k - \mu_k = a_{k1}F_1 + a_{k2}F_2 + \cdots + a_{km}F_m + \varepsilon_k \quad (4.1)$$

onde X_k é a k -ésima variável, $a_{k1}, a_{k2}, \dots, a_{km}$ são as cargas dos fatores (loadings) para a k -ésima variável, F_1, F_2, \dots, F_m são m fatores comuns não correlacionados, com m menor que p e ε_k são os fatores específicos.

Os p valores observados X_p são expressos em termos de $p + m$ variáveis aleatórias não observáveis ($F_1, F_2, \dots, F_m; \varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$). Isso distingue o modelo fatorial do modelo de regressão múltipla, no qual as variáveis independentes podem ser observadas, e cujas posições são ocupadas por F no modelo fatorial.

Matricialmente teríamos

$$\underset{(p \times 1)}{\mathbf{X}} - \underset{(p \times 1)}{\boldsymbol{\mu}} = \underset{(p \times m)}{\mathbf{A}} \underset{(m \times 1)}{\mathbf{F}} + \underset{(p \times 1)}{\boldsymbol{\varepsilon}} \quad (4.2)$$

Uma verificação direta do modelo fatorial, a partir das observações X_1, X_2, \dots, X_p , é impossibilitada por tantas quantidades não observáveis. Entretanto, com algumas pressuposições impostas aos vetores aleatórios, F e ε , o modelo fatorial implica em certas relações de covariância, que podem ser verificadas (JOHNSON & WICHERN, 1988). Assim os vetores F e ε devem satisfazer as seguintes condições:

$$E(\mathbf{F}) = \mathbf{0}_{(m \times 1)}, \quad \text{Cov}(\mathbf{F}) = E(\mathbf{F}\mathbf{F}') = \mathbf{I}_{(m \times m)}, \text{ logo } \text{Var}(F_j) = 1, j = 1, 2, \dots, m$$

$$E(\boldsymbol{\varepsilon}) = \mathbf{0}_{(p \times 1)}, \quad \text{Cov}(\boldsymbol{\varepsilon}) = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \boldsymbol{\Psi}_{(p \times p)} \quad \text{onde } \boldsymbol{\Psi} \text{ é uma matriz diagonal.}$$

$$\text{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}$$

e que \mathbf{F} e $\boldsymbol{\varepsilon}$ são independentes. Assim,

$$\text{Cov}(\boldsymbol{\varepsilon}, \mathbf{F}) = E(\boldsymbol{\varepsilon}\mathbf{F}') = \mathbf{0}_{(p \times m)} \quad (4.3)$$

Essas pressuposições e a relação em (4.2) constituem o chamado modelo de fatores ortogonal.

Das pressuposições acima, obtemos a estrutura da matriz de covariâncias de \mathbf{X} , que será representada por $\boldsymbol{\Sigma}$.

$$\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X}) = \text{Cov}(\boldsymbol{\Lambda}\mathbf{F} + \boldsymbol{\varepsilon}).$$

Mas, por (4.3), $\boldsymbol{\Lambda}\mathbf{F}$ e $\boldsymbol{\varepsilon}$ não são correlacionadas, portanto, a matriz de covariância da soma será a soma das suas matrizes de covariâncias:

$$\begin{aligned} \boldsymbol{\Sigma} &= \text{cov}(\boldsymbol{\Lambda}\mathbf{F} + \boldsymbol{\varepsilon}) = \text{Cov}(\boldsymbol{\Lambda}\mathbf{F}) + \text{Cov}(\boldsymbol{\varepsilon}) \\ &= \boldsymbol{\Lambda}\text{Cov}(\mathbf{F})\boldsymbol{\Lambda}' + \boldsymbol{\Psi} \\ &= \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi} \\ &= \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi} \end{aligned} \quad (4.4)$$

De uma maneira mais fácil de entender e usando apenas propriedades da variância, poderíamos, a partir da relação (4.1), chegar ao mesmo resultado:

$$X_k - \mu_k = a_{k1}F_1 + a_{k2}F_2 + \cdots + a_{km}F_m + \varepsilon_k$$

aplicando as propriedades da variância, e com base nas pressuposições em (4.3), teremos:

$$X_k - \mu_k = a_{k1}F_1 + a_{k2}F_2 + \cdots + a_{km}F_m + \varepsilon_k$$

$$\text{Var}(X_k - \mu_k) = \text{Var}(X_k) = a_{k1}^2 \text{Var}(F_1) + a_{k2}^2 \text{Var}(F_2) + \cdots + a_{km}^2 \text{Var}(F_m) + \text{Var}(\varepsilon_k)$$

Como $\text{Var}(F_j) = 1, j = 1, 2, \dots, m$, e $\text{Var}(\varepsilon_k) = \psi_k$ tem-se:

$$\text{Var}(X_k) = a_{k1}^2 + a_{k2}^2 + \cdots + a_{km}^2 + \psi_k$$

onde $a_{k1}^2 + a_{k2}^2 + \cdots + a_{km}^2$, ou seja $\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$, é chamada de comunalidade da variável X_k (a parte da sua variância que está relacionada com os fatores comuns) enquanto $\boldsymbol{\Psi}_k$ é chamada especificidade de X_k (a parte da sua variância que não está relacionada com os fatores comuns).

Tem-se, portanto, a estrutura de covariância:

$$\begin{aligned} a) \quad \text{Cov}(\mathbf{X}) &= \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi} \quad \text{ou} \\ \text{Var}(X_k) &= a_{k1}^2 + a_{k2}^2 + \cdots + a_{km}^2 + \boldsymbol{\Psi}_k \\ \text{Cov}(X_i, X_j) &= a_{i1}a_{j1} + \cdots + a_{im}a_{jm} \\ b) \quad \text{Cov}(\mathbf{X}, \mathbf{F}) &= \boldsymbol{\Lambda} \\ \text{Cov}(X_i, F_j) &= a_{ij} \end{aligned}$$

O modelo fatorial pressupõe efeitos aditivos; fatores, variáveis e resíduos normalmente distribuídos, resíduos independentes e relações lineares entre as variáveis (JOHNSON & WICHERN, 1988)

O modelo fatorial assume que as $p + p(p-1)/2 = p(p+1)/2$ variâncias e covariâncias para \mathbf{X} podem ser reproduzidas a partir de pm cargas fatoriais (a_{kj}) e p variâncias específicas (ψ_k). Quando $m = p$, qualquer matriz de covariância ($\boldsymbol{\Sigma}$) pode ser reproduzida exatamente como $\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$, e $\boldsymbol{\Psi}$ pode ser nula. Contudo, a análise fatorial será mais eficiente e útil quando m for pequeno em relação a p , proporcionando uma explicação mais simples da covariação das variáveis em \mathbf{X} , com base num número de parâmetros menor do que os $p(p+1)/2$ parâmetros de $\boldsymbol{\Sigma}$ (JOHNSON & WICHERN, 1988).

Em resumo o modelo fatorial implica na imposição de condições que permitem obter estimativas únicas de $\boldsymbol{\Lambda}$ e $\boldsymbol{\Psi}$. Posteriormente, a matriz de cargas fatoriais é submetida à rotação (multiplicação por uma matriz ortogonal), a qual é

determinada por critérios de facilidade de interpretação. Obtidas as cargas e as variâncias específicas, os fatores são identificados e comumente calcula-se os valores dos escores fatoriais.

4.3. Procedimentos para a Análise Fatorial

Existem três estágios na AF. O problema inicial é a determinação das cargas dos fatores (a_{kj}) com $j = 1$ a p , e $k = 1$ a m (sendo $m < p$). Há vários métodos para a estimação de parâmetros na análise fatorial, dos quais os mais comuns são o Método do Componente Principal e o Método da Máxima Verossimilhança.

Não importando qual método seja usado para a determinação das cargas dos fatores, é possível mostrar que tais cargas não são únicas. Se F_1, F_2, \dots, F_m são os fatores provisórios, então, combinações lineares desses fatores podem ser construídas, e "explicar" os dados tão bem quanto os anteriores. Existe um número infinito de soluções alternativas para os modelos de análise fatorial, e isto nos conduz ao segundo estágio da análise, o qual é chamado rotação dos fatores. Dessa forma, os fatores provisórios são transformados, com o objetivo de se encontrar novos fatores, mais fáceis de serem interpretados.

A rotação dos fatores pode ser ortogonal ou oblíqua. Com a rotação ortogonal, os novos fatores serão não-correlacionados, enquanto com a rotação oblíqua os novos fatores serão correlacionados. Para qualquer tipo de rotação usada, é desejável que as cargas para os novos fatores sejam ou próximas de zero ou muito diferentes de zero. Uma carga a_{kj} próxima de zero indica que a variável X_k não é muito relacionada com o fator F_j . Um grande (positivo ou negativo) valor de a_{kj} (carga dos fatores) indica que X_k (a variável k) é determinada por F_j (fator j) em grande extensão. Se cada variável for fortemente relacionada com alguns fatores, mas não de todo relacionada com outros, então isto faz com que os fatores sejam mais facilmente identificados. Um método de rotação dos fatores ortogonal, que é frequentemente usado, é chamado Rotação Varimax (KARSON, 1982).

O último estágio da análise corresponde à estimação dos valores dos fatores comuns F_1, F_2, \dots, F_m para cada indivíduo, como função das variáveis observadas.

4.4. Métodos de estimação das cargas fatoriais

4.4.1. Método do Componente Principal

O nome desse método ajuda a aumentar a confusão entre as técnicas de Análise de componentes principais e Análise fatorial. No método de componentes principais para estimação de cargas fatoriais não são calculados CPs, o nome se dá porque as cargas fatoriais são os coeficientes padronizados dos primeiros CPs amostrais, sendo assim, os fatores poderiam ser interpretados da mesma maneira que os CPs, entretanto, após a rotação das cargas, a interpretação usualmente fica diferente.

A matriz de covariância Σ pode ser fatorada pelo processo de expansão de matrizes simétricas, denominado decomposição espectral (visto em Álgebra). Dessa forma, Σ será formada por pares de autovalores e autovetores normalizados ($\lambda_k \mathbf{e}_k$), com $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. Ou seja:

$$\Sigma = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \dots + \lambda_p \mathbf{e}_p \mathbf{e}_p' = \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1 & \sqrt{\lambda_2} \mathbf{e}_2 & \dots & \sqrt{\lambda_p} \mathbf{e}_p \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1' \\ \sqrt{\lambda_2} \mathbf{e}_2' \\ \vdots \\ \sqrt{\lambda_p} \mathbf{e}_p' \end{bmatrix}$$

Se tomarmos o modelo da análise fatorial tendo tantos fatores quanto variáveis (ou seja, $m = p$) e as variâncias específicas $\Psi_k = 0$ para todo k , então a matriz de cargas fatoriais (Λ) teria $m = p$ colunas e seria dada por $\sqrt{\lambda_k} \mathbf{e}_k$. Ou seja, nós poderíamos escrever:

$$\underset{(p \times p)}{\Sigma} = \underset{(p \times p)}{\Lambda} \underset{(p \times p)}{\Lambda'} + \underset{(p \times p)}{\mathbf{0}} = \Lambda \Lambda', \quad (4.5)$$

mas apesar dessa representação da análise de fatores ser exata, ela não é particularmente útil porque ela usa tantos fatores comuns quantos são o número de variáveis, além de não permitir nenhuma variação nos fatores específicos ε_k . Portanto, são preferíveis modelos que expliquem a estrutura de covariância levando em conta apenas alguns fatores comuns. Uma aproximação, quando os últimos $p-m$ autovalores são pequenos, é omitir a contribuição dos termos $\lambda_{m+1} \mathbf{e}_{m+1} \mathbf{e}_{m+1}' + \dots + \lambda_p \mathbf{e}_p \mathbf{e}_p'$ para a matriz Σ obtendo:

$$\Sigma = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \dots + \lambda_m \mathbf{e}_m \mathbf{e}_m' = \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1 & \sqrt{\lambda_2} \mathbf{e}_2 & \dots & \sqrt{\lambda_m} \mathbf{e}_m \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1' \\ \sqrt{\lambda_2} \mathbf{e}_2' \\ \vdots \\ \sqrt{\lambda_m} \mathbf{e}_m' \end{bmatrix} = \begin{matrix} \mathbf{\Lambda} \\ (p \times m) \end{matrix} \begin{matrix} \mathbf{\Lambda}' \\ (m \times p) \end{matrix} \quad (4.6)$$

Essa representação aproximada assume que os fatores específicos $\boldsymbol{\varepsilon}$ são de menor importância e podem também ser ignorados na fatoração de Σ . Se os fatores específicos forem incluídos no modelo, suas variâncias poderiam ser tomadas como sendo os elementos diagonais de $\Sigma - \mathbf{\Lambda}\mathbf{\Lambda}'$ e a aproximação torna-se:

$$\Sigma = \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1 & \sqrt{\lambda_2} \mathbf{e}_2 & \dots & \sqrt{\lambda_m} \mathbf{e}_m \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1' \\ \sqrt{\lambda_2} \mathbf{e}_2' \\ \vdots \\ \sqrt{\lambda_m} \mathbf{e}_m' \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{bmatrix} = \begin{matrix} \mathbf{\Lambda} \\ (p \times m) \end{matrix} \begin{matrix} \mathbf{\Lambda}' \\ (m \times p) \end{matrix} + \begin{matrix} \mathbf{\Psi} \\ (p \times p) \end{matrix} \quad (4.7)$$

Para aplicar esta aproximação para um conjunto de dados $x_{1k}, x_{2k}, \dots, x_{nk}$, normalmente, as observações são centradas pela subtração da média amostral \bar{x}_k :

$$(x_{ik} - \bar{x}_k) = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \dots \\ x_{ip} \end{bmatrix} - \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \dots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} x_{i1} - \bar{x}_1 \\ x_{i2} - \bar{x}_2 \\ \dots \\ x_{ip} - \bar{x}_p \end{bmatrix}, \quad i = 1, 2, \dots, n.$$

As observações centradas terão a mesma matriz de covariância amostral \mathbf{S} , que os dados originais.

Nos casos em que as unidades das variáveis em estudo não são comensuráveis, é preferível trabalhar com as variáveis padronizadas para média 0 e variância 1. Com esta transformação, a matriz de covariância corresponde à matriz de correlação das observações originais. A padronização evita o problema de variáveis com grande variância influenciarem indevidamente a determinação das cargas fatoriais.

A representação em (4.7), levando-se em consideração os fatores específicos, quando aplicada à matriz de covariância amostral \mathbf{S} , ou à matriz de correlação amostral \mathbf{R} , é conhecida como solução por componentes principais. O nome deriva do fato de que as cargas fatoriais são os coeficientes escalares dos primeiros componentes principais amostrais (JOHNSON & WICHERN, 1988).

A solução da AF por componentes principais é especificada pelos pares de autovalores-autovetores $(\hat{\lambda}_1, \hat{\mathbf{e}}_1), (\hat{\lambda}_2, \hat{\mathbf{e}}_2), \dots, (\hat{\lambda}_p, \hat{\mathbf{e}}_p)$, onde $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$. Considerando o número de fatores comuns $m < p$, a matriz de cargas fatoriais estimadas $\{\hat{a}_{kj}\}$ é dada por:

$$\hat{\mathbf{\Lambda}} = \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1 & \sqrt{\hat{\lambda}_2} \hat{\mathbf{e}}_2 & \dots & \sqrt{\hat{\lambda}_m} \hat{\mathbf{e}}_m \end{bmatrix} \quad (4.8)$$

As variâncias específicas estimadas são dadas pelos elementos da diagonal principal da matriz $\mathbf{S} - \hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}'$, ou seja,

$$\hat{\mathbf{\Psi}} = \begin{bmatrix} \hat{\psi}_1 & 0 & \dots & 0 \\ 0 & \hat{\psi}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{\psi}_p \end{bmatrix} \quad \text{com} \quad \hat{\psi}_k = s_{kk} - \sum_{j=1}^m \hat{a}_{kj}^2 \quad (4.9)$$

As comunalidades são estimadas por:

$$\hat{h}_k^2 = \hat{a}_{k1}^2 + \hat{a}_{k2}^2 + \dots + \hat{a}_{km}^2, \quad (4.10)$$

que é a soma de quadrados das cargas na k -ésima linha de $\hat{\mathbf{\Lambda}}$.

Por (4.9) e (4.10), a variância da k -ésima variável é particionada em uma parte que é devida aos fatores e outra parte que é devida unicamente à variável:

$$s_{kk} = \hat{h}_k^2 + \hat{\psi}_k = \hat{a}_{k1}^2 + \hat{a}_{k2}^2 + \dots + \hat{a}_{km}^2 + \hat{\psi}_k.$$

Então o j -ésimo fator contribui com \hat{a}_{kj}^2 para a variância s_{kk} . A contribuição do j -ésimo fator para a variância amostral total, dada por $\text{tr}(\mathbf{S}) = s_{11} + s_{22} + \dots + s_{pp}$, é dada por:

$$\text{Contribuição do } j\text{-ésimo fator para a variância total} = \sum_{k=1}^p \hat{a}_{kj}^2 = \hat{a}_{1j}^2 + \hat{a}_{2j}^2 + \dots + \hat{a}_{pj}^2$$

que é a soma de quadrados das cargas na j -ésima coluna de $\hat{\mathbf{A}}$.

Mas usando (4.8) temos:

$$\sum_{k=1}^p \hat{a}_{kj}^2 = \sum_{k=1}^p \left(\sqrt{\hat{\lambda}_j} \hat{\mathbf{e}}_{kj} \right)^2 = \hat{\lambda}_j \sum_{k=1}^p \hat{\mathbf{e}}_{kj}^2 = \hat{\lambda}_j, \text{ pois os autovetores normalizados têm norma 1.}$$

Logo, a Variância devido ao j -ésimo fator é igual ao j -ésimo autovalor, $\hat{\lambda}_j$. A proporção da variância amostral total explicada pelo j -ésimo fator é portanto:

$$\frac{\sum_{k=1}^p \hat{a}_{kj}^2}{\text{tr}(\mathbf{S})} = \frac{\hat{\lambda}_j}{\text{tr}(\mathbf{S})}.$$

Se for usada a matriz de correlação \mathbf{R} , ao invés de \mathbf{S} , a proporção da variância amostral total devido ao j -ésimo fator será dada por:

$$\frac{\sum_{k=1}^p \hat{a}_{kj}^2}{\text{tr}(\mathbf{R})} = \frac{\hat{\lambda}_j}{p}$$

onde p é o número de variáveis.

Por esta técnica de estimação, as estimativas das cargas dos fatores para um dado fator não mudam à medida que o número de fatores é aumentado. Por exemplo, se $m = 1$, $\hat{\mathbf{A}}$ será dado por $\left[\sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1 \right]$ e se $m = 2$, $\hat{\mathbf{A}}$ será dado por $\left[\sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1 \quad \sqrt{\hat{\lambda}_2} \hat{\mathbf{e}}_2 \right]$ onde $(\hat{\lambda}_1, \hat{\mathbf{e}}_1)$ e $(\hat{\lambda}_2, \hat{\mathbf{e}}_2)$ são os primeiros dois pares de autovalores-autovetores normalizados para a matriz de covariância amostral \mathbf{S} , ou para a matriz de correlação amostral \mathbf{R} .

Exemplo 4.1. (Brown; Williams; Barlow, 1984). Uma menina avaliou seus colegas, com uma escala de 0 a 9, em relação a 5 adjetivos: Gentil, inteligente, feliz, confiável e justo. Os resultados são apresentados na Tabela 1:

Tabela 1. Dados de percepção: Avaliação de cinco adjetivos para sete pessoas

Colegas	Gentil	Inteligente	Feliz	Confiável	Justo
C1	1	5	5	1	1
C2	8	9	7	9	8
C3	9	8	9	9	8
C4	9	9	9	9	9
C5	1	9	1	1	9
C6	9	7	7	9	9
C7	9	7	9	9	7

A matriz de correlação é:

$$\mathbf{R} = \begin{bmatrix} 1.000 & 0.296 & \mathbf{0.881} & \mathbf{0.995} & 0.545 \\ 0.296 & 1.000 & -0.022 & 0.326 & \mathbf{0.837} \\ \mathbf{0.881} & -0.022 & 1.000 & \mathbf{0.867} & 0.130 \\ \mathbf{0.995} & 0.326 & \mathbf{0.867} & 1.000 & 0.544 \\ 0.545 & \mathbf{0.837} & 0.130 & 0.544 & 1.000 \end{bmatrix}$$

Os valores em negrito indicam dois grupos de variáveis: {1, 3, 4} e {2, 5}. Espera-se, portanto, que as correlações entre as variáveis possam ser bem explicadas por dois fatores.

Utilizando o R são obtidos os autovalores: 3.263, 1.538, 0.168, 0.031 e 0.000. Observa-se que R é uma matriz singular, pois o último autovalor é aproximadamente zero.

Os primeiros dois fatores são responsáveis $(3.263+1.538)/5 = 0.96$ (96%) da variância amostral. Portanto dois fatores são suficientes para explicar a variabilidade amostral. Os dois primeiros autovetores são:

$$\mathbf{e}_1 = \begin{bmatrix} 0.537 \\ 0.288 \\ 0.434 \\ 0.537 \\ 0.390 \end{bmatrix} \text{ e } \mathbf{e}_2 = \begin{bmatrix} -0.186 \\ 0.621 \\ -0.473 \\ -0.169 \\ 0.538 \end{bmatrix}$$

As cargas fatoriais são obtidas multiplicando os autovetores pela raiz quadrada dos seus respectivos autovalores. As cargas estimadas para os dois primeiros fatores são dadas por:

$$\hat{\mathbf{A}}_2 = \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1 & \sqrt{\hat{\lambda}_2} \hat{\mathbf{e}}_2 \end{bmatrix} = \begin{bmatrix} \sqrt{3.263} \times \begin{bmatrix} 0.537 \\ 0.288 \\ 0.434 \\ 0.537 \\ -0.390 \end{bmatrix} & \sqrt{1.538} \times \begin{bmatrix} -0.186 \\ 0.621 \\ -0.473 \\ -0.169 \\ 0.538 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 0.969 & -0.231 \\ 0.519 & 0.807 \\ 0.785 & -0.587 \\ 0.971 & -0.210 \\ 0.704 & 0.667 \end{bmatrix}$$

As communalidades são obtidas somando-se os quadrados dos resultados das linhas de $\hat{\mathbf{A}}_2$:

$$\hat{h}_1^2 = \hat{a}_{11}^2 + \hat{a}_{12}^2 = 0.969^2 + (-0.231)^2 = 0.993$$

$$\hat{h}_2^2 = \hat{a}_{21}^2 + \hat{a}_{22}^2 = 0.519^2 + 0.807^2 = 0.921$$

⋮

$$\hat{h}_5^2 = \hat{a}_{51}^2 + \hat{a}_{52}^2 = 0.704^2 + 0.667^2 = 0.940$$

Para se obter as variâncias específicas usa-se (4.9) mas subtraindo-se a communalidade de 1 (porque estamos fatorando a matriz \mathbf{R} e não a \mathbf{S}): $\hat{\psi}_k = 1 - \hat{h}_k^2$

$$\hat{\psi}_1 = 1 - \hat{h}_1^2 = 1 - 0.993 = 0.007$$

⋮

$$\hat{\psi}_5 = 1 - \hat{h}_5^2 = 1 - 0.940 = 0.060$$

A variância explicada por cada fator, que é a contribuição de cada fator para a variância total é a soma dos quadrados das colunas correspondentes na matriz de cargas fatoriais e é também igual ao valor do autovetor correspondente.

Contribuição do j -ésimo fator para a variância total = $\sum_{k=1}^p \hat{a}_{kj}^2$

Variância explicada pelo Fator 1: $0.969^2 + 0.519^2 + \dots + 0.704^2 = 3.263$

Variância explicada pelo Fator 2: $(-0.231)^2 + 0.807^2 + \dots + 0.667^2 = 1.538$

A proporção da variância total devida a cada fator é dada pela razão entre a variância explicada pelo fator dividido por 5 ($\text{tr}(\mathbf{R}) = p$):

$$\text{proporção da variância explicada pelo fator 1} = \frac{\sum_{k=1}^p \hat{a}_{k1}^2}{\text{tr}(\mathbf{R})} = \frac{\hat{\lambda}_1}{p} = \frac{3.263}{5} = 0.653$$

$$\text{proporção da variância explicada pelo fator 2} = \frac{\sum_{k=1}^p \hat{a}_{k2}^2}{\text{tr}(\mathbf{R})} = \frac{\hat{\lambda}_2}{p} = \frac{1.538}{5} = 0.308$$

A proporção da variância acumulada pelos dois fatores (0.653+0.308) explica 96% da variância total e, portanto, representa as cinco variáveis muito bem.

Os resultados podem ser organizados na Tabela de AF:

Tabela 2. AF com cargas estimadas pelo método de componentes principais

Variáveis	Cargas fatoriais		Comunalidades	Variâncias específicas
	F ₁	F ₂		
Gentil	0.969	-0.231	0.993	0.007
Inteligente	0.519	0.807	0.921	0.079
Feliz	0.785	-0.587	0.960	0.040
Confiável	0.971	-0.210	0.987	0.013
Justo	0.704	0.667	0.940	0.060
Variância devido ao fator	3.263	1.538		
Proporção da variância total	0.653	0.308		
Proporção da variância acumulada	0.653	0.960		

4.4.2. Método do Fator Principal

Para estimação das cargas fatoriais pelo método do componente principal, visto anteriormente, a matriz de covariâncias Ψ é desprezada e a fatoração é feita na matriz S ou R . No método do fator principal é usada uma estimativa inicial $\hat{\Psi}$ e fatora-se $S - \hat{\Psi}$ ou $R - \hat{\Psi}$ para obter:

$$S - \hat{\Psi} \cong \hat{\Lambda}\hat{\Lambda}',$$

$$R - \hat{\Psi} \cong \hat{\Lambda}\hat{\Lambda}',$$

onde $\hat{\Lambda}$ é uma matriz de dimensão $p \times m$ calculada como em (5.8), mas usando autovalores e autovetores obtidos de $S - \hat{\Psi}$ ou $R - \hat{\Psi}$.

O k -ésimo elemento da diagonal de $S - \hat{\Psi}$ é dado por $s_{kk} - \hat{\psi}_k$, que é a k -ésima comunalidade, $\hat{h}_k^2 = s_{kk} - \hat{\psi}_k$ (observe que os valores serão diferentes se usar S ou R). Substituindo esses valores na diagonal principal, $S - \hat{\Psi}$ e $R - \hat{\Psi}$ tomam as formas:

$$S - \hat{\Psi} = \begin{bmatrix} \hat{h}_1^2 & s_{12} & \cdots & s_{1p} \\ s_{21} & \hat{h}_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & \hat{h}_p^2 \end{bmatrix}$$

$$R - \hat{\Psi} = \begin{bmatrix} \hat{h}_1^2 & r_{12} & \cdots & r_{1p} \\ r_{21} & \hat{h}_2^2 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & \hat{h}_p^2 \end{bmatrix}$$

Existem métodos para estimar a comunalidade e um dos mais populares é usar a correlação múltipla entre X_k e as outras $p - 1$ variáveis como estimativa inicial em $R - \hat{\Psi}$, ou seja,

$$\hat{h}_k^2 = R_k^2 = 1 - \frac{1}{r^{kk}}$$

onde r^{kk} é o k -ésimo elemento da diagonal de R^{-1} .

Para $S - \hat{\Psi}$, uma estimativa inicial da comunalidade, análoga à anterior é:

$$\hat{h}_k^2 = s_{kk} - \frac{1}{s^{kk}}$$

onde s_{kk} é o k -ésimo elemento da diagonal de S e s^{kk} é o k -ésimo elemento da diagonal de S^{-1} .

Para usar essas estimativas, as matrizes R ou S precisam ser não singulares para a estimação. Porém, se R é singular, pode ser usado o valor absoluto ou o quadrado das maiores correlações em cada k -ésima linha de R como uma estimativa para a comunalidade.

Depois de obter as estimativas para a comunalidade, são calculados os autovalores e autovetores de $S - \hat{\Psi}$ ou $R - \hat{\Psi}$ e usado (5.8), como na técnica de componentes principais, para obter as cargas fatoriais, $\hat{\Lambda}$. Posteriormente as colunas e linhas de $\hat{\Lambda}$ podem ser usadas para obter novos autovalores (variância explicada) e comunalidades, respectivamente. A soma de quadrados da k -ésima linha de $\hat{\Lambda}$ é a comunalidade de X_k . A proporção da variância explicada pelo j -ésimo fator é dada por:

$$\frac{\hat{\lambda}_j}{tr(S - \hat{\Psi})} = \frac{\hat{\lambda}_j}{\sum_{k=1}^p \hat{\lambda}_k} \quad \text{ou} \quad \frac{\hat{\lambda}_j}{tr(R - \hat{\Psi})} = \frac{\hat{\lambda}_j}{\sum_{k=1}^p \hat{\lambda}_k}$$

onde $\hat{\lambda}_j$ é o j -ésimo autovalor de $S - \hat{\Psi}$ ou $R - \hat{\Psi}$. As matrizes $S - \hat{\Psi}$ e $R - \hat{\Psi}$ não são necessariamente semidefinidas positivas e muitas vezes têm alguns pequenos autovalores negativos. Nesse caso, a proporção acumulada da variância excederá 1 e depois diminuirá para 1 conforme os autovalores negativos são adicionados.

Exemplo 4.2. Utilizaremos o método do fator principal para estimar as cargas fatoriais dos dados de percepção apresentados na Tabela 1. Como a matriz de correlação calculada para os dados é singular, as correlações múltiplas não serão usadas como estimativas das comunalidades, ao invés disso será utilizado o valor absoluto da maior correlação de cada linha de R . A matriz $R - \hat{\Psi}$ será dada por:

$$\mathbf{R} - \mathbf{\Psi} = \begin{bmatrix} 0.995 & 0.296 & 0.881 & 0.995 & 0.545 \\ 0.296 & 0.837 & -0.022 & 0.326 & 0.837 \\ 0.881 & -0.022 & 0.881 & 0.867 & 0.130 \\ 0.995 & 0.326 & 0.867 & 0.995 & 0.544 \\ 0.545 & 0.837 & 0.130 & 0.544 & 0.837 \end{bmatrix}$$

Utilizando o R são obtidos os autovalores: 3.202, 1.395, 0.030, -0.0002 e -0.080. Os dois primeiros autovetores são:

$$\mathbf{e}_1 = \begin{bmatrix} 0.548 \\ 0.272 \\ 0.431 \\ 0.549 \\ 0.373 \end{bmatrix} \text{ e } \mathbf{e}_2 = \begin{bmatrix} -0.178 \\ 0.656 \\ -0.460 \\ -0.159 \\ 0.549 \end{bmatrix}$$

Quando esses autovalores são multiplicados pela raiz quadrada dos seus respectivos autovetores, são obtidas as cargas fatoriais para os dois primeiros fatores $\hat{\Lambda}_2$. Somando em cada linha da matriz $\hat{\Lambda}_2$, as cargas elevadas ao quadrado, obtém-se as communalidades. A proporção da variância é obtida dividindo-se a variância explicada pelo fator (autovalor) pela soma dos autovalores, que é 4.546. Para o primeiro fator, por exemplo, a proporção da variância é: $3.202/4.546 = 0.704$. Os resultados comparando as duas técnicas de estimação das cargas fatoriais estão apresentados na Tabela 3.

Tabela 3. AF com cargas estimadas pelos métodos de componentes principais e fator principal

Variáveis	Cargas Componentes principais			Cargas Fator principal		
	F ₁	F ₂	Comunalidades	F ₁	F ₂	Comunalidades
Gentil	0.969	-0.231	0.993	0.981	-0.210	0.995
Inteligente	0.519	0.807	0.921	0.487	0.774	0.837
Feliz	0.785	-0.587	0.960	0.771	-0.544	0.881
Confiável	0.971	-0.210	0.987	0.982	-0.188	0.995
Justo	0.704	0.667	0.940	0.667	0.648	0.837
Variância devido ao fator	3.263	1.538		3.202	1.395	
Proporção da variância	0.653	0.308		0.704	0.307	
Proporção da variância acumulada	0.653	0.960		0.704	1.010	

4.4.3. Método da Máxima Verossimilhança

O método da Máxima Verossimilhança para Análise Fatorial foi introduzido por Lawley, em 1940, e não era muito usado anteriormente, por suas dificuldades computacionais. Atualmente, no entanto, devido aos trabalhos de Jöreskog, já se encontram procedimentos rápidos e eficientes para a obtenção dos estimadores por este método (FACHEL, 1976). Mas a principal vantagem de utilizar o Método da Máxima Verossimilhança para Análise Fatorial é, talvez, a possibilidade de se desenvolverem testes de hipóteses, com o objetivo de testar a adequação do modelo, o que não é possível através dos outros dois métodos apresentados anteriormente, devido às suas características semi-paramétricas.

Além das suposições habituais do modelo fatorial, suporemos que os vetores aleatórios \mathbf{F} (fatores comuns) e $\mathbf{\varepsilon}$ (fatores específicos) têm distribuição normal multivariada com vetores de média zero e com matrizes de covariância \mathbf{I} (pois estamos considerando os fatores não correlacionados) e $\mathbf{\Psi}$, respectivamente. Das suposições habituais (5.3) e da suposição de normalidade, segue que \mathbf{F} e $\mathbf{\varepsilon}$ são mutuamente independentes. Como \mathbf{X} é expresso em termos de \mathbf{F} e $\mathbf{\varepsilon}$, conforme a equação (5.2), temos que o vetor das variáveis observáveis é também normal, com média zero e matriz de covariância $\mathbf{\Sigma}$. Portanto, as condições para a aplicação do Método da Máxima Verossimilhança na Análise Fatorial são:

- \mathbf{X} tem uma distribuição normal multivariada, com vetor de médias zero e matriz de covariância $\mathbf{\Sigma}$;
- $\mathbf{X} = \mathbf{\Lambda F} + \mathbf{\varepsilon}$, onde $E(\mathbf{F}) = E(\mathbf{\varepsilon}) = \mathbf{0}$, $\text{Var}(\mathbf{F}) = \mathbf{I}$, $\text{Var}(\mathbf{\varepsilon}) = \mathbf{\Psi} = \text{diag}(\psi_k^2)$, $\text{Cov}(\mathbf{F}, \mathbf{\varepsilon}) = \mathbf{0}$;
- $\mathbf{\Sigma} = \mathbf{\Lambda \Lambda}' + \mathbf{\Psi}$;
- As colunas de \mathbf{X} são amostras aleatórias de tamanho n das variáveis em questão e, portanto, a matriz de covariância amostral \mathbf{S} tem distribuição de Wishart.

O caminho a ser tomado será o de maximizar a função de verossimilhança de \mathbf{X} , com respeito aos pm elementos de $\mathbf{\Lambda}$ e os p elementos de $\mathbf{\Psi}$.

De acordo com o modelo fatorial, toda a informação da amostra está contida em \mathbf{S} . Portanto, segundo KARSON (1982), ignorando os termos que não envolvem parâmetros e usando a distribuição de Wishart, a qual nos dá a distribuição de probabilidade multivariada para os elementos de \mathbf{S} e tem a matriz de parâmetros $\mathbf{\Sigma}$ na função de verossimilhança para os elementos de \mathbf{S} , a função de verossimilhança pode ser escrita como

$$L = |\mathbf{\Sigma}|^{-(n-1)/2} e^{-[(n-1)/2]tr(\mathbf{S}\mathbf{\Sigma}^{-1})}$$

Conforme feito usualmente, utiliza-se $\ln L$ para maximizar

$$\ln L = -[(n-1)/2][\ln|\mathbf{\Sigma}| + tr(\mathbf{S}\mathbf{\Sigma}^{-1})] \quad (4.11)$$

Substituindo $\mathbf{\Sigma}$ por $\mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$ em (4.11) tem-se,

$$\ln L = -[(n-1)/2][\ln|\mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}| + tr(\mathbf{S}(\mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi})^{-1})] \quad (4.12)$$

O método da máxima verossimilhança envolve determinar os pm elementos de $\mathbf{\Lambda}$ e os p elementos de $\mathbf{\Psi}$ que maximizem (5.12). Para evitar a indeterminação de $\mathbf{\Lambda}$, será feita a imposição da condição de que $\mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda}$ deverá ser diagonal. Se $\hat{\mathbf{\Lambda}}$ denotar a matriz dos estimadores de máxima verossimilhança de $\mathbf{\Lambda}$ e $\hat{\mathbf{\Psi}}$ denotar o estimador de máxima verossimilhança de $\mathbf{\Psi}$, então $\hat{\mathbf{\Lambda}}$ e $\hat{\mathbf{\Psi}}$ serão obtidos diferenciando (5.12) em relação a $\mathbf{\Lambda}$ e $\mathbf{\Psi}$; fazendo as $p(m+1)$ derivadas iguais a zero; e resolvendo o sistema das $p(m+1)$ equações. Pode ser demonstrado que as estimativas $\hat{\mathbf{\Lambda}}$ e $\hat{\mathbf{\Psi}}$ devem satisfazer:

$$\hat{\mathbf{\Lambda}}(\mathbf{I} + \hat{\mathbf{\Lambda}}'\hat{\mathbf{\Psi}}^{-1}\hat{\mathbf{\Lambda}}) = \mathbf{S}\hat{\mathbf{\Psi}}^{-1}\hat{\mathbf{\Lambda}} \quad (4.13)$$

$$\hat{\mathbf{\Psi}} = \text{diag}(\mathbf{S} - \hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}') \quad (4.14)$$

$$\hat{\mathbf{\Lambda}}'\hat{\mathbf{\Psi}}^{-1}\hat{\mathbf{\Lambda}} \text{ é diagonal} \quad (4.15)$$

A solução pelo método da máxima verossimilhança usa a matriz de covariância amostral \mathbf{S} , não a matriz de correlação amostral \mathbf{R} , e, portanto, não assume implicitamente, que os valores das variáveis estão padronizados.

Uma vantagem do método da máxima verossimilhança é que ele permite ao analista verificar se o modelo proposto é adequado, a partir de um teste de hipótese formal. Independente de qual método seja usado, o analista deveria observar as magnitudes dos elementos da matriz residual $\mathbf{\Sigma} - (\mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi})$ para um dado número de fatores m . Quanto menores estes elementos, melhor a solução obtida reproduz $\mathbf{\Sigma}$, sendo também melhor a estrutura proposta para \mathbf{X}_k .

No teste de adequação do modelo fatorial são testadas as seguintes hipóteses:

$H_0: \mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$

$H_1: \mathbf{\Sigma} = \text{qualquer outra matriz positiva definida.}$

A estatística de teste segue a distribuição Qui-quadrado e é dada por:

$$\chi^2_{\text{calc.}} = \left((n-1) - \frac{2p+4m+5}{6} \right) \left\{ \ln \left(\frac{|\hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}' + \hat{\mathbf{\Psi}}|}{|\mathbf{S}|} \right) + tr \left[(\hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}' + \hat{\mathbf{\Psi}})^{-1} \mathbf{S} \right] - p \right\}$$

com $\nu = \frac{1}{2}[(p-m)^2 - (p+m)]$ graus de liberdade

A hipótese H_0 será rejeitada se $\chi^2_{\text{calc.}} \geq \chi^2_{(\nu)} \alpha$, para um determinado nível de significância α .

Se a hipótese nula é rejeitada, então o modelo fatorial escolhido é inadequado. O número de fatores m deve, então, ser aumentado de 1 e o modelo testado novamente, sendo o processo repetido até que um valor $\chi^2_{\text{calc.}}$ não-significativo seja obtido (SRIVASTAVA & CARTER, 1983).

JOHNSON & WICHERN (1988) comentam que se n é grande e m é pequeno (relativamente a p), a hipótese H_0 será, geralmente, rejeitada, ocasionando a retenção de mais fatores comuns. No entanto, $\mathbf{\Sigma} = \hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}' + \hat{\mathbf{\Psi}}$ pode estar suficientemente próxima de \mathbf{S} , de modo que mais fatores não acrescentariam muita informação, apesar deles serem "significativos". Portanto, um julgamento pessoal por parte do analista deve ser levado em consideração na escolha do número de fatores m .

4.5. Rotação dos fatores

Para procurar uma melhor interpretação dos fatores, é prática comum fazer uma rotação ou uma transformação dos fatores. Pode ser mostrado que o conjunto de cargas fatoriais, obtidas por qualquer método de solução fatorial, quando o número de fatores comuns é maior do que um, não é único, pois outros conjuntos equivalentes podem ser encontrados,

por transformações ortogonais de cargas. As cargas rotacionadas preservam as propriedades essenciais das cargas originais. Elas reproduzem a matriz de covariância e satisfazem todas as suposições básicas.

Geometricamente, as cargas na k -ésima linha de $\hat{\Lambda}$ constituem as coordenadas de um ponto no espaço de cargas correspondentes a X_k . A rotação dos p pontos dá as coordenadas deles em relação a novos eixos (fatores), mas deixa sua configuração geométrica básica intacta. Espera-se encontrar um novo quadro de referência no qual os fatores são mais interpretáveis. Para tanto, o objetivo da rotação é colocar os eixos perto de tantos pontos quanto possível. Se houver grupos de pontos (correspondendo a agrupamentos de X 's), procura-se mover os eixos para passar pelos pontos ou perto destes clusters. Isso associaria cada grupo de variáveis a um fator (eixo) e faria a interpretação mais objetiva. Os eixos resultantes representam os fatores naturais.

Se for alcançada uma rotação em que cada ponto esteja próximo a um eixo, então cada variáveis terá altas cargas no fator correspondente ao eixo e pequenas cargas nos demais fatores. Nesse caso, não há ambiguidade. E a configuração é chamado de estrutura simples, que tem a interpretação bastante simplificada. N qual meramente se observa quais variáveis estão associadas a cada fator, e o fator é definido ou nomeado em conformidade com essas variáveis.

A fim de identificar os agrupamentos naturais de variáveis, buscamos uma rotação para um padrão interpretável para as cargas, em que as variáveis têm altas cargas em apenas um fator. O número de fatores em que uma variável tem cargas moderadas ou altas é chamada de *complexidade* da variável. Na situação ideal referida anteriormente como estrutura simples, todas as variáveis têm uma complexidade de 1. Neste caso, as variáveis foram claramente agrupadas em grupos correspondentes aos fatores.

As rotações são de dois tipos básicos: ortogonal e oblíqua. A rotação ortogonal envolve uma matriz ortogonal na qual os eixos perpendiculares originais são rotacionados rigidamente e permanecem perpendiculares. Os ângulos e distâncias são preservados, as comunalidades não se alteram e a configuração básica dos pontos permanece a mesma. Somente o eixo de referência difere. A rotação oblíqua promove uma transformação na qual os eixos não são requeridos permanecerem perpendiculares e, portanto, são livres para passar mais próximo de grupos de pontos.

Apesar de estarmos livres para escolher qual rotação fazer, de modo a termos uma melhor interpretação dos fatores, não é aconselhável fazermos isto subjetivamente, porque poderíamos estar forçando o ajuste das cargas dos fatores com um padrão preconcebido. Uma escolha conveniente e mais utilizada é o chamado método Varimax, que será descrito de maneira resumida.

Método Varimax

Este método de rotação ortogonal foi proposto por Kaiser (1958). A ideia do método consiste no seguinte: Para cada rotação dos fatores que ocorre, há o aparecimento de altas cargas para poucas variáveis, enquanto as demais cargas ficarão próximas de zero. No início Kaiser definiu a simplicidade de um fator j como a variância do quadrado de suas cargas, isto é:

$$s_j^2 = \frac{1}{p} \sum_{k=1}^p (a_{kj}^2)^2 - \frac{1}{p^2} \left(\sum_{j=1}^p a_{kj}^2 \right)^2$$

onde a_{kj} é a nova carga para a variável k no fator j , $k = 1, 2, \dots, p$ e $j = 1, 2, \dots, m$.

Quando a variância atinge o máximo, o fator tem maior interpretabilidade ou simplicidade, no sentido de que as cargas deste fator tendem ou à unidade, ou à zero. O critério de máxima simplicidade de uma matriz fatorial completa é definido como a maximização da soma destas simplicidades.

Como este critério dá igual peso às variáveis com comunalidades grandes ou pequenas, Kaiser sugeriu que antes de iniciar o processo de maximização, as cargas fossem divididas pela raiz quadrada da comunalidade correspondente, o que é equivalente a normalizar os vetores a_k . Após a obtenção de uma matriz ortogonal M , as cargas finais deveriam ser multiplicadas novamente pela raiz quadrada da comunalidade. A este critério varimax modificado, Kaiser denominou critério varimax normal, e é o mais utilizado.

Portanto, este método requer que as cargas dos fatores finais sejam tais que maximizem a função:

$$V = \sum_{j=1}^m s_j^2 = \frac{1}{p} \sum_{j=1}^m \sum_{k=1}^p \left(\frac{a_{kj}^2}{h_k^2} \right)^2 - \frac{1}{p^2} \sum_{j=1}^m \left(\sum_{k=1}^p \frac{a_{kj}^2}{h_k^2} \right)^2$$

onde h_k^2 é a comunalidade da variável k .

Ou, de uma maneira mais simples, após a multiplicação da expressão anterior por p^2 , já que a multiplicação por uma constante não afeta o processo de maximização:

$$V = p \sum_{j=1}^m s_j^2 = \frac{1}{p} \sum_{j=1}^m \sum_{k=1}^p \left(\frac{a_{kj}^2}{h_k^2} \right)^2 - \frac{1}{p} \sum_{j=1}^m \left(\sum_{k=1}^p \frac{a_{kj}^2}{h_k^2} \right)^2$$

Esta expressão foi chamada por Kaiser como critério varimax normal ou simplesmente critério varimax.

As rotações varimax de cargas fatoriais, obtidas a partir de diferentes métodos de estimação (componentes principais, máxima verossimilhança, etc.), em geral, não são coincidentes. Da mesma forma, se fatores comuns adicionais são incluídos no modelo, o padrão de carregamento rotacionado pode mudar consideravelmente. Se existe um único (geral) fator dominante (no qual todas as variáveis apresentam cargas altas), geralmente, ele é obscurecido por qualquer rotação ortogonal. Entretanto, este fator pode ser mantido fixo e os demais rotacionados (MENEZES et al., 1978).

A rotação de fatores comuns é particularmente recomendada para cargas estimadas pelo método da máxima verossimilhança, porque as cargas iniciais são constrangidas a satisfazer a condição de unicidade (matriz diagonal). Esta condição, geralmente, facilita a maximização da função de verossimilhança, mas pode gerar fatores não facilmente interpretáveis (JOHNSON & WICHERN, 1988).

4.6. Estimação dos escores fatoriais

Em muitas aplicações, principalmente quando a Análise Fatorial é preliminar a algum outro tipo de análise multivariada, ou quando o seu uso principal é para construção de índices, é conveniente, além de estimar os parâmetros do modelo fatorial, procurar descrever os fatores em termos das variáveis observadas. Para isto, estimam-se os valores de cada fator para cada indivíduo. Estes valores são denominados escores fatoriais.

Serão considerados dois métodos de estimação dos escores dos fatores, o método dos mínimos quadrados ponderados e o método da regressão. Estes métodos serão descritos supondo-se que os fatores sejam correlacionados, por ser mais geral. No caso de fatores não correlacionados e padronizados, os resultados são obtidos do mesmo modo, considerando-se $Cov(\mathbf{F}) = \Phi = \mathbf{I}$.

Ambos os métodos apresentam os seguintes pressupostos (JOHNSON & WICHERN, 1988):

- As estimativas das cargas fatoriais (a_{kj}) e das variâncias específicas (ϵ_k) são tomadas como valores paramétricos;
- Os métodos envolvem transformações lineares dos dados originais padronizados, e geralmente, utilizam as cargas estimadas nos cálculos dos escores.

Em análise de Componentes Principais, as componentes são definidas como funções lineares das variáveis observadas e então, os valores de cada componente, para cada indivíduo, podem ser facilmente encontrados. Em Análise Fatorial, os fatores não são funções lineares das variáveis observadas apenas e os escores de um indivíduo sobre eles não podem ser encontrados da mesma maneira. É necessário então estimar os fatores a partir das variáveis observadas X_k .

4.6.1. Método da Regressão (Método de Thomson)

Os fatores F_m podem ser relacionados com as variáveis observadas por meio de uma regressão centrada, lembrando que $E(\mathbf{F}) = \mathbf{0}$:

$$\begin{aligned} F_1 &= \beta_{11}(X_1 - \bar{X}_1) + \beta_{12}(X_2 - \bar{X}_2) + \cdots + \beta_{1p}(X_p - \bar{X}_p) + \epsilon_1 \\ F_2 &= \beta_{21}(X_1 - \bar{X}_1) + \beta_{22}(X_2 - \bar{X}_2) + \cdots + \beta_{2p}(X_p - \bar{X}_p) + \epsilon_2 \\ &\vdots \\ F_m &= \beta_{m1}(X_1 - \bar{X}_1) + \beta_{m2}(X_2 - \bar{X}_2) + \cdots + \beta_{mp}(X_p - \bar{X}_p) + \epsilon_m. \end{aligned}$$

Que pode ser escrito na forma matricial como:

$$\mathbf{F} = \mathbf{B}'(\mathbf{X} - \bar{\mathbf{X}}) + \boldsymbol{\epsilon} \quad (4.16)$$

Na regressão será estimada a matriz \mathbf{B} , que será usada para estimar \mathbf{F} por meio de $\hat{\mathbf{F}} = \hat{\mathbf{B}}'(\mathbf{X} - \bar{\mathbf{X}})$.

O modelo (4.16) assegura para cada observação possa ser escrita a relação:

$$\mathbf{F}_i = \mathbf{B}'(\mathbf{X}_i - \bar{\mathbf{X}}) + \boldsymbol{\epsilon}_i, \quad i = 1, 2, \dots, n.$$

Na forma transposta tem-se:

$$\mathbf{F}_i' = (\mathbf{X}_i - \bar{\mathbf{X}})' \mathbf{B} + \boldsymbol{\epsilon}_i' \quad i = 1, 2, \dots, n.$$

Essas n equações podem ser combinadas em um modelo da seguinte forma:

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_1' \\ \mathbf{F}_2' \\ \vdots \\ \mathbf{F}_n' \end{bmatrix} = \begin{bmatrix} (\mathbf{X}_1 - \bar{\mathbf{X}})' \mathbf{B} \\ (\mathbf{X}_2 - \bar{\mathbf{X}})' \mathbf{B} \\ \vdots \\ (\mathbf{X}_n - \bar{\mathbf{X}})' \mathbf{B} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1' \\ \boldsymbol{\epsilon}_2' \\ \vdots \\ \boldsymbol{\epsilon}_n' \end{bmatrix} = \begin{bmatrix} (\mathbf{X}_1 - \bar{\mathbf{X}})' \\ (\mathbf{X}_2 - \bar{\mathbf{X}})' \\ \vdots \\ (\mathbf{X}_n - \bar{\mathbf{X}})' \end{bmatrix} \mathbf{B} + \mathbf{E} = \mathbf{X}_c \mathbf{B} + \mathbf{E} \quad (4.17)$$

onde \mathbf{X}_c é a matriz das variáveis \mathbf{X} centradas.

Para um modelo de regressão centrado na média, as estimativas dos parâmetros podem ser representadas matricialmente por:

$$\hat{\mathbf{B}} = (\mathbf{X}'_c \mathbf{X}_c)^{-1} \mathbf{X}'_c \mathbf{F} \quad (4.18)$$

Entretanto, \mathbf{F} é não observado. Para contornar esse problema, o modelo (4.18) pode ser reescrito como função das matrizes de covariâncias:

$$\hat{\mathbf{B}} = \mathbf{S}_{XX}^{-1} \mathbf{S}_{XF} \quad (4.19)$$

Substituindo \mathbf{S}_{XX} por \mathbf{S} e \mathbf{S}_{XF} por $\hat{\mathbf{\Lambda}}$ (pois $\hat{\mathbf{\Lambda}}$ estima $Cov(\mathbf{X}, \mathbf{F}) = \mathbf{\Lambda}$), o modelo (4.19) pode ser reescrito como:

$$\hat{\mathbf{B}} = \mathbf{S}^{-1} \hat{\mathbf{\Lambda}} \quad (4.20)$$

Do modelo (5.17) os valores de \mathbf{F}_i estimados são dados por

$$\hat{\mathbf{F}} = \begin{bmatrix} \hat{\mathbf{F}}'_1 \\ \hat{\mathbf{F}}'_2 \\ \vdots \\ \hat{\mathbf{F}}'_n \end{bmatrix} = \mathbf{X}_c \hat{\mathbf{B}} = \mathbf{X}_c \mathbf{S}^{-1} \hat{\mathbf{\Lambda}} \quad (4.21)$$

Se a matriz usada na fatoração for \mathbf{R} , ao invés de \mathbf{S} , (4.20) e (4.21) tornam-se:

$$\begin{aligned} \hat{\mathbf{B}} &= \mathbf{R}^{-1} \hat{\mathbf{\Lambda}} \\ \hat{\mathbf{F}} &= \mathbf{X}_s \mathbf{R}^{-1} \hat{\mathbf{\Lambda}} \end{aligned}$$

onde \mathbf{X}_s é a matriz de variáveis observadas padronizadas: $(x_{jk} - \bar{X}_k)/s_k$.

4.6.2. Método dos Mínimos Quadrados Ponderados (Método de Bartlett)

Este método, desenvolvido por Bartlett (1938) usa o método de mínimos quadrados na estimação. Os escores são obtidos de tal forma que a soma de quadrados dos resíduos padronizados seja mínima, em relação aos elementos de \mathbf{F} . Assim

$$\sum_{k=1}^p \frac{\varepsilon_k^2}{\psi_k} = \varepsilon' \psi^{-1} \varepsilon = (\mathbf{X}_c - \mathbf{\Lambda} \mathbf{F})' \Psi^{-1} (\mathbf{X}_c - \mathbf{\Lambda} \mathbf{F})$$

é derivado em relação ao vetor \mathbf{F} e obtém-se:

$$\hat{\mathbf{F}}^* = (\mathbf{\Lambda} \Psi^{-1} \mathbf{\Lambda})^{-1} \mathbf{\Lambda} \Psi^{-1} \mathbf{X}$$

que é a expressão para se obter os estimadores dos escores fatoriais, tanto no caso de fatores correlacionados, como no caso de fatores não correlacionados.

Para o caso de fatores não correlacionados e quando a matriz $\mathbf{\Lambda} \Psi^{-1} \mathbf{\Lambda}$ é diagonal, observa-se que os estimadores de \mathbf{F} , pelos dois métodos descritos, diferem apenas por um fator escala.

Lawley & Maxwell (1971) mostraram que os estimadores obtidos pelo método de regressão são viesados, enquanto os obtidos pelo método de Bartlett são não viesados. No entanto, os estimadores do método de regressão têm menor variância do que os de Bartlett

4.7. Validade do modelo fatorial

Após ter verificado a normalidade e linearidade dos dados, deve-se utilizar os seguintes passos para a aplicação da AF:

(i) analisar a matriz de correlações, (ii) verificar a estatística KMO e o teste de esfericidade de Bartlett e (iii) a analisar a matriz anti-imagem.

Análise da matriz de Correlação

De acordo com Hair *et al.* (2005), se a inspeção visual da matriz de correlação não revela um número substancial de correlações maiores que 0,30, então a AF provavelmente é inapropriada.

KMO e Teste de Esfericidade de Bartlett

Para Fávero *et al.* (2009) utiliza-se o teste de esfericidade de Bartlett com intuito de avaliar a hipótese de que a matriz das correlações pode ser a matriz identidade com determinante igual a 1, conforme apresentado a seguir:

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Se a matriz de correlação for igual à matriz identidade, isso significa que as inter-relações entre as variáveis são iguais a 0 e, nesse caso, deve-se reconsiderar a utilização de análise fatorial. Vale destacar que este teste requer que as variáveis apresentem normalidade multivariada (PESTANA e GAGEIRO, 2005).

Quando se quer comparar as correlações simples com as correlações parciais se usa a estatística de Kaiser-Meyer-Olkin (KMO). Essa medida é dada pela seguinte expressão:

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} a_{ij}^2}$$

onde:

r_{ij} = coeficiente de correlação entre variáveis;

a_{ij} = coeficiente de correlação parcial.

A estatística KMO, assume valores que variam de 0 e 1, onde mais próximo de 0 indica que a análise fatorial pode não ser adequada, pois existe uma correlação fraca entre as variáveis. Em contrapartida, quanto mais próximo de 1 o seu valor, mas adequada é a utilização da técnica. A Tabela 4 mostra os intervalos de análise dos valores de KMO.

Tabela 4: Classificação da Estatística KMO para Aplicação da Análise Fatorial.

KMO	Análise Fatorial
0,90 1,00	Muito boa
0,80 0,90	Boa
0,70 0,80	Média
0,60 0,70	Razoável
0,50 0,60	Má
0,00 0,50	Inaceitável

Assim, valores para a estatística KMO inferior a 0,60 indicam que a análise fatorial pode ser inadequada.

Pode ser calculada também uma Medida de Adequação da Amostra, ou Measure of Sampling Adequacy (MSA), para cada variável de forma similar à estatística KMO a partir de:

$$MSA = \frac{\sum_{j=1}^p r_{ij}^2}{\sum_{j=1}^p r_{ij}^2 + \sum_{j=1}^p a_{ij}^2}$$

O pesquisador deve analisar primeiramente os valores de MSA para cada variável individualmente e excluir as que se encontram no domínio inaceitável (HAIR *et al.* 2005). Vale ressaltar que, por vezes, a baixa correlação de determinada variável com as demais não necessariamente implica sua eliminação, uma vez que esta variável pode representar um fator isoladamente.

Análise da matriz anti-imagem

A principal premissa de uma Análise Fatorial é que exista uma estrutura de dependência clara entre as variáveis envolvidas. Essa estrutura pode ser vista inicialmente na matriz de covariância ou de correlação. Mas se tal estrutura existe, isso implica que uma variável pode, dentro de certos limites, ser prevista pelas demais. Para isso avaliar isso são calculados os coeficientes de correlação parcial entre os pares de variáveis. A matriz resultante desse processo é chamada

de matriz anti-imagem, e ela é construída com esses coeficientes com sinais invertidos. Ela é uma forma de obter indícios sobre a necessidade de eliminação de alguma variável do modelo fatorial. A diagonal da matriz de correlação anti-imagem contém os valores KMO para as variáveis individuais, valores pequenos sugerem a retirada da variável da análise.

Exemplo 4.3. Exemplificando o uso da análise fatorial no campo florestal, selecionou-se parte dos dados publicados no trabalho de HOOGH & DIETRICH (1979). O trabalho envolve a avaliação de sítios para *Araucaria angustifolia* em povoamentos artificiais, localizados em diversos municípios dos Estados do Rio Grande do Sul, Santa Catarina, Paraná, São Paulo e Minas Gerais (Análise realizada pelo Prof. Agostinho Lopes (UFV))

Tomaremos como objetivo verificar a relação entre a produtividade dos sítios, expressa pelo Índice de Sítio, e as características do solo. Para tanto, foram selecionados 21 sítios (Quadro 1) e 10 variáveis (Quadro 2), dentre as apresentadas por HOOGH & DIETRICH (1979), cujas correlações lineares com o índice de Sítio foram maiores do que 0,3.

Quadro 1 – Relação dos 21 pontos de amostragem estudados (extraído de HOOGH & DIETRICH, 1979)

Parc.	Perfil solo	Localidade	IS*
1	SC-28	Três Barras - SC	15,7
2	SC-30	Três Barras - SC	10,2
3	PR-32	Teixeira Soares - PR	13,6
4	PR-37	Teixeira Soares - PR	19,5
5	PR-44	Ponta Grossa - PR	8,8
6	PR-68	Telêmaco Borba - PR	16,6
7	PR-73	Telêmaco Borba - PR	18,9
8	PR-83	Jussara - PR	19,8
9	RS-85	São Francisco de Paula - RS	16,3
10	RS-87	São Francisco de Paula - RS	17,9
11	RS-107	Passo Fundo - RS	16,6
12	RS-113	Passo Fundo - RS	14,5
13	PR-123	Cascavel - PR	18,7
14	SC-142	Caçador - SC	15,8
15	SC-146	Chapecó - SC	19,1
16	SP-152	Capão Bonito - SP	10,7
17	SP-153	Capão Bonito - SP	9,3
18	MG-187	Passa Quatro - MG	13,5
19	SP-216	Caieiras - SP	21,9
20	PR-240	Renascença - PR	20,8
21	SC-246	Rio Negrinho - SC	11,2

* Índice de Sítio = altura dominante aos 25 anos.

Quadro 2 – Relação das 10 variáveis utilizados no estudo (extraído de HOOGH & DIETRICH, 1979)

Parc.	Variáveis*									
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	4,3	4,1	20,1	9,80	11,60	2,00	6	6	5,5	5,5
2	3,9	4,0	17,5	11,20	19,40	1,33	9	9	4,3	3,2
3	3,8	3,9	19,8	6,80	6,80	10,00	7	13	2,7	3,3
4	4,0	5,2	21,6	9,00	9,80	25,00	9	12	6,2	6,2
5	3,8	4,0	10,7	3,80	5,40	4,44	5	15	1,4	1,4
6	3,8	4,2	6,6	2,80	4,80	10,00	16	12	1,0	1,5
7	4,0	4,2	18,6	9,40	10,40	15,00	7	9	3,8	5,2
8	3,8	4,0	6,1	3,80	4,40	15,00	9	18	2,8	25,0
9	4,0	4,0	11,4	9,80	9,80	6,67	27	8	2,3	2,2
10	4,5	4,0	16,3	18,00	18,00	2,00	77	20	2,6	2,6
11	3,7	3,9	17,0	10,40	11,40	0,01	6	8	4,1	10,2
12	4,0	3,9	16,3	13,00	13,00	17,50	30	18	3,8	7,8

Continua...

Quadro 2, Cont.

Parc.	Variáveis*									
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
13	4,2	4,0	24,2	28,20	27,80	6,67	9	5	7,4	5,2
14	3,8	3,9	19,8	21,00	20,20	2,50	21	9	4,9	4,7
15	4,0	4,2	18,6	17,50	10,10	3,33	6	9	5,4	5,7
16	4,0	4,2	6,6	2,60	3,30	5,00	10	12	2,7	1,8
17	3,8	4,0	12,0	6,40	7,90	5,00	16	5	3,2	4,8
18	4,0	4,3	13,7	4,00	5,20	10,00	8	8	11,0	7,0
19	6,7	6,0	14,7	6,60	9,80	60,00	88	67	4,2	9,3
20	3,7	3,9	19,1	21,60	22,70	3,33	4	18	19,0	8,2
21	3,7	3,8	11,8	4,90	4,10	2,50	7	9	2,4	1,6

* V1 = pH KCl horiz.A; V2 = pH KCl horiz.B; V3 = Al₂O₃ total horiz.A; V4 = Fe₂O₃ total horiz.A; V5 = Fe₂O₃ total horiz.B; V6 = Mg/K trocáveis horiz.B; V7 = saturação de bases horiz. A; V8 = saturação de bases horiz.B; V9 = argila/silte horiz.A; V10 = argila/silte horiz.B.

O Quadro 3 apresenta a matriz de correlações (R) entre as 10 variáveis estudadas. Os fatores foram extraídos utilizando-se os três métodos expostos anteriormente, ou seja, Componente Principal, Máxima Verossimilhança e Fator Principal, tendo em vista verificar qual deles apresenta o melhor ajuste do modelo fatorial.

Quadro 3 – Matriz de correlações entre as 10 variáveis estudadas

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
V1	1,00000									
V2	0,82020	1,00000								
V3	0,06980	0,06033	1,00000							
V4	-0,01637	-0,20272	0,72569	1,00000						
V5	0,05063	-0,14588	0,70816	0,92705	1,00000					
V6	0,83291	0,89914	-0,01912	-0,22396	-0,16482	1,00000				
V7	0,80509	0,53532	-0,04946	0,08025	0,12672	0,59442	1,00000			
V8	0,89243	0,76280	-0,10896	-0,12760	-0,05588	0,86571	0,77032	1,00000		
V9	-0,05391	0,01495	0,42672	0,47071	0,48363	-0,05085	-0,19875	-0,00891	1,00000	
V10	0,09076	0,12015	-0,13210	-0,04682	-0,05767	0,28618	-0,00715	0,24326	0,17153	1,00000

Os resultados iniciais de comunalidades, autovalores e percentuais de variância explicada por cada fator são mostrados no Quadro 4. Nesse caso, o Método do Componente Principal foi usado, devido a maior facilidade de obtenção dos resultados.

Quadro 4 – Comunalidades e fatores iniciais, extraídos pelo Método do Componente Principal, com os seus respectivos autovalores e percentuais de variância explicada

Variáveis	Comunal.*	Fator	Autovalor	Var. Expl. (%)	Var. Acum. (%)
V1	0,90357	1	4,21030	42,1	42,1
V2	0,87504	2	2,89058	28,9	71,0
V3	0,70916	3	1,19116	11,9	82,9
V4	0,88693	4	0,74910	7,5	90,4
V5	0,87344	5	0,49495	4,9	95,4
V6	0,91351	6	0,14695	1,5	96,8
V7	0,79563	7	0,11797	1,2	98,0
V8	0,90775	8	0,08866	0,9	98,9
V9	0,50310	9	0,06701	0,7	99,6
V10	0,32338	10	0,04333	0,4	100,0

* Comunalidades iniciais estimadas por R² e usadas nos Métodos da Máxima Verossimilhança e do Fator Principal.

No modelo inicial completo com 10 fatores, somente os três primeiros fatores apresentaram autovalores maiores do que um. Esses três fatores explicaram, individualmente, 42,1; 28,9 e 11,9% da variância, acumulando 82,9% da variação

total. Diante disso, optou-se por prosseguir a análise considerando apenas os três primeiros fatores comuns. As estruturas fatoriais iniciais ficaram compostas pelas matrizes apresentadas no Quadro 5, considerando os três métodos de estimação.

As comunalidades iniciais foram consideradas iguais a um, quando do uso do Método do Componente Principal, ou foram estimadas pelos quadrados dos coeficientes de correlação múltipla (R^2) de equações de regressão entre a variável considerada e todas as demais variáveis.

Com apenas três fatores, o Método do Componente Principal mostrou o melhor ajuste, explicando 82,9% da variância e apresentando comunalidades finais que oscilaram entre 0,68 e 0,94 (Quadro 6).

Os Métodos da Máxima Verossimilhança e do Fator Principal explicaram 75% da variância total. As comunalidades foram baixas para algumas variáveis, ou seja, pelo Método da Máxima Verossimilhança elas variaram entre 0,08 e 0,95 e pelo Método do Fator Principal, entre 0,11 e 0,96. Nesse caso, para ambos os métodos, as variáveis V9 e V10 (argila/silte horiz. A e B, respectivamente) apresentaram pouca contribuição no modelo fatorial ajustado, principalmente V10.

A estatística Qui-Quadrado (χ^2), calculada para o modelo com três fatores ajustado pelo Método da Máxima Verossimilhança, com 18 graus de liberdade, atingiu o valor de 12,85 ($P=0,8006$), indicando uma boa estimativa da matriz R.

As matrizes de correlação estimadas por cada um dos três métodos, com os respectivos resíduos em relação à matriz de correlações original R (Quadro 3), são apresentadas no Quadro 7. Nesse caso, as melhores estimativas de R foram obtidas pelos Métodos da Máxima Verossimilhança e do Fator Principal. Para o Método da Máxima Verossimilhança, obteve-se 1,16732 de soma de resíduos absolutos (média = 0,02594), sendo que 20% deles foram maiores do que 0,05. Pelo Método do Fator Principal, obteve-se 1,3764 de soma de resíduos absolutos (média = 0,03058), sendo 17% deles maiores do que 0,05.

Quadro 5 – Cargas iniciais dos três primeiros fatores extraídos pelos três métodos de estimação

Variáveis	Fator 1	Fator 2	Fator 3
Método do Componente Principal			
V1	0,92992	0,23728	-0,13261
V2	0,88350	0,09561	0,06094
V3	-0,14197	0,83565	-0,07549
V4	-0,25661	0,91073	-0,08451
V5	-0,19138	0,92349	-0,09395
V6	0,93733	0,05297	0,15171
V7	0,77744	0,20603	-0,35831
V8	0,94301	0,11726	0,05190
V9	-0,15080	0,62078	0,52063
V10	0,21760	-0,01300	0,85031
Método da Máxima Verossimilhança			
V1	0,93223	0,18797	0,02913
V2	0,85514	-0,04189	0,36392
V3	-0,10684	0,71718	0,37997
V4	-0,20624	0,94985	0,04710
V5	-0,13962	0,94355	0,05464
V6	0,91043	-0,05701	0,31597
V7	0,82132	0,28273	-0,43674
V8	0,93133	0,06768	0,02461
V9	-0,14827	0,44386	0,42670
V10	0,15831	-0,05260	0,24914
Método do Fator Principal			
V1	0,92710	0,22627	-0,08641
V2	0,85219	0,06824	0,24398
V3	-0,12136	0,74242	0,13428
V4	-0,24697	0,93703	-0,10111
V5	-0,18049	0,94226	-0,09822
V6	0,94104	0,03191	0,28123
V7	0,78498	0,21523	-0,54895
V8	0,92666	0,10161	0,01653
V9	-0,12961	0,51149	0,39712

V10	0,17314	-0,01838	0,27849
-----	---------	----------	---------

Quadro 6 – Comunalidades finais, autovalores e percentuais de variância explicada pelos três primeiros fatores, extraídos pelos três métodos

Variáveis	Comunalid.	Fator	Autovalor	% da Var.	% Acum.
Método do Componente Principal					
V1	0,93864	1	4,21030	42,1	42,1
V2	0,79342	2	2,89058	28,9	71,0
V3	0,72417	3	1,19116	11,9	82,9
V4	0,90243				
V5	0,89829				
V6	0,90440				
V7	0,77525				
V8	0,90571				
V9	0,67917				
V10	0,77055				
Método da Máxima Verossimilhança					
V1	0,90524	1	4,09163	40,9	40,9
V2	0,86546	2	2,63149	26,3	67,2
V3	0,67014	3	0,81820	8,2	75,4
V4	0,94697				
V5	0,91277				
V6	0,93197				
V7	0,94525				
V8	0,87257				
V9	0,40107				
V10	0,08990				
Método do Fator Principal					
V1	0,91818	1	4,10126	41,0	41,0
V2	0,79041	2	2,69256	26,9	67,9
V3	0,58395	3	0,72087	7,2	75,1
V4	0,94925				
V5	0,93008				
V6	0,96566				
V7	0,96387				
V8	0,86929				
V9	0,43613				
V10	0,10787				

Já o Método do Componente Principal apresentou 2,54337 de soma de resíduos absolutos (média = 0,05652), com 51% deles acima de 0,05 (Quadro 7). Em termos percentuais, considerando o valor de soma dos resíduos pelo Componente Principal como 100%, o Método da Máxima Verossimilhança atingiu 46% e o Método do Fator Principal alcançou 54%. Esse resultado já era esperado, pois os Métodos da Máxima Verossimilhança e do Fator Principal envolvem iterações e cálculos mais elaborados do que o Componente Principal, e por isso, oferecem melhor estimativa da estrutura de correlações original.

Buscando obter-se uma estrutura fatorial mais simples para cada modelo, procedeu-se a rotação ortogonal Varimax dos três fatores. Os resultados são mostrados no Quadro 8. As estruturas fatoriais obtidas, após a rotação dos eixos, pouco se alteraram em relação às estruturas originais (Quadro 5).

Quadro 7 – Matrizes de correlação estimadas e resíduos obtidos por cada um dos três métodos de estimação*

	Método do Componente Principal									
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
V1	0,93864	-0,01599	-0,00648	-0,00504	-0,00299	-0,03117	-0,01426	-0,00543	0,00807	0,00425
V2	0,83619	0,79342	0,11046	-0,05793	-0,05937	0,05671	-0,14941	-0,08472	0,05710	-0,12267
V3	0,07627	-0,05013	0,72417	-0,07817	-0,09782	0,08114	-0,13831	-0,06915	-0,07414	-0,02616
V4	-0,01133	-0,14479	0,80387	0,90243	0,02894	-0,01885	0,06183	0,01198	-0,08935	0,09272
V5	0,05361	-0,08651	0,80598	0,89811	0,89829	-0,02010	0,05158	0,02118	-0,06961	0,07586
V6	0,86409	0,84243	-0,10026	-0,20511	-0,14472	0,90440	-0,09084	-0,03228	-0,02137	-0,04610
V7	0,81936	0,68473	0,08885	0,01842	0,07515	0,68526	0,77525	0,03163	-0,02286	0,13103
V8	0,89786	0,84752	-0,03981	-0,13958	-0,07706	0,89799	0,73869	0,90571	0,03348	-0,00455
V9	-0,06198	-0,04215	0,50086	0,56007	0,55324	-0,02948	-0,17589	-0,04239	0,67917	-0,23029
V10	0,08651	0,24282	-0,10595	-0,13953	-0,13353	0,33228	-0,13819	0,24781	0,40182	0,77055
	Método da Máxima Verossimilhança									
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
V1	0,90523	0,02029	0,02352	-0,00401	0,00184	-0,01430	-0,00099	0,01078	-0,01155	-0,05419
V2	0,79991	0,86545	0,04345	-0,00370	-0,00684	0,00322	0,00376	-0,03973	0,00505	-0,10810
V3	0,04628	0,01687	0,67014	0,00455	-0,00422	-0,00102	0,00147	-0,06735	-0,06958	-0,17214
V4	-0,01236	-0,19902	0,72115	0,94697	-0,00055	0,00308	0,00166	-0,00097	-0,00157	0,02406
V5	0,04879	-0,13904	0,71237	0,92760	0,91277	-0,00118	-0,00151	0,00895	0,02081	0,00045
V6	0,84722	0,89592	-0,01810	-0,22704	-0,16365	0,93197	0,00078	0,01388	-0,02537	0,06033
V7	0,80608	0,53156	-0,05093	0,07859	0,12824	0,59364	0,94525	-0,00299	-0,01610	-0,01350
V8	0,88165	0,80254	-0,04161	-0,12664	-0,06483	0,85183	0,77331	0,87256	0,08864	0,09325
V9	-0,04236	0,00990	0,49630	0,47228	0,46282	-0,02547	-0,18265	-0,09755	0,40107	0,11204
V10	0,14495	0,22825	0,04003	-0,07087	-0,05812	0,22585	0,00634	0,15001	0,05949	0,08990
	Método do Fator Principal									
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
V1	0,91818	0,03578	0,02593	-0,00816	-0,00374	-0,02244	-0,01880	0,01177	-0,01517	-0,04154
V2	0,78442	0,79041	0,08033	-0,03154	-0,03241	0,02641	-0,01439	-0,03785	-0,00639	-0,09409
V3	0,04387	-0,02000	0,58395	0,01362	-0,00011	0,03364	-0,04027	-0,07415	-0,02207	-0,13484
V4	-0,00821	-0,17119	0,71207	0,94925	-0,01039	0,00699	0,01693	0,00771	-0,00042	0,04133
V5	0,05436	-0,11347	0,70827	0,93744	0,93008	0,00258	0,01168	0,01725	0,01729	0,01826
V6	0,85535	0,87273	-0,05276	-0,23095	-0,16741	0,96566	0,00323	-0,01421	-0,05688	0,04551
V7	0,82389	0,54971	-0,00919	0,06332	0,11505	0,59119	0,96387	0,03012	0,01091	0,01376
V8	0,88067	0,80066	-0,03480	-0,13532	-0,07313	0,87991	0,74020	0,86929	0,05265	0,08008
V9	-0,03874	0,02134	0,44880	0,47114	0,46634	0,00603	-0,20965	-0,06157	0,43613	0,09278
V10	0,13230	0,21424	0,00274	-0,08814	-0,07592	0,24067	-0,02092	0,16318	0,07875	0,10787

* O triângulo inferior esquerdo de cada matriz contém a matriz estimada de R; a diagonal, as communalidades finais, conforme o Quadro 6; e o triângulo superior direito, os resíduos entre as correlações observadas e as estimadas.

Quadro 8 – Estruturas fatoriais após a rotação ortogonal Varimax dos eixos coordenados

Variáveis	Fator 1	Fator 2	Fator 3
Método do Componente Principal			
V1	0,96586	0,06335	-0,04170
V2	0,87671	-0,07674	0,13750
V3	0,02191	0,85017	-0,02983
V4	-0,07563	0,94598	-0,04285
V5	-0,00857	0,94662	-0,04617
V6	0,91328	-0,13343	0,22919
V7	0,83141	0,07324	-0,28043
V8	0,93977	-0,06632	0,13473
V9	-0,08056	0,61042	0,54778
V10	0,13320	-0,09780	0,86212
Método da Máxima Verossimilhança			
V1	0,90949	0,04093	0,27639
V2	0,69515	-0,05898	0,61542
V3	-0,02630	0,80707	0,13448
V4	0,02313	0,93756	-0,25967
V5	0,08197	0,92366	-0,23000
V6	0,75569	-0,09659	0,59293
V7	0,94959	0,00261	-0,20864
V8	0,88091	-0,07317	0,30199
V9	-0,14288	0,57159	0,23226
V10	0,07068	0,00336	0,29136
Método do Fator Principal			
V1	0,95291	0,04812	0,08843
V2	0,78860	-0,06108	0,40595
V3	0,00184	0,76166	0,06176
V4	-0,03760	0,95164	-0,20548
V5	0,02691	0,94521	-0,18959
V6	0,86020	-0,10884	0,46246
V7	0,90007	0,01752	-0,39171
V8	0,90917	-0,06383	0,19656
V9	-0,09979	0,56263	0,33109
V10	0,11133	-0,02167	0,30824

Considerando-se o conhecimento teórico da matéria em pauta, a racionalidade dos resultados e a proporção da variância amostral total explicada por cada um dos três modelos fatoriais finais, pode-se, seguramente, optar pelo modelo gerado a partir do Método do Componente Principal.

Nesse caso, o **Fator 1** envolve as variáveis relacionadas com a parte **química do solo** (fertilidade, acidez, disponibilidade de nutrientes etc.), ou seja, V1 e V2 (pH KCl horiz. A e B), V6 (Mg/K trocáveis horiz. B), V7 e V8 (saturação de bases horiz. A e B) (Quadro 15). O **Fator 2** abriga as variáveis relacionadas com a **mineralogia do solo** (V3 = Al₂O₃ total horiz. A; V4 e V5 = Fe₂O₃ total horiz. A e B). Já o **Fator 3** traz a contribuição da **física do solo** (textura e indiretamente, estrutura, densidade, porosidade, aeração, drenagem, etc.) através das variáveis V9 e V10 (argila/silte horiz. A e B). A variável V9 também apresentou parcela de contribuição significativa no Fator 2, o que não deixa de ser lógico, pois sempre há relação entre a mineralogia e as frações texturais do solo, notadamente as mais finas. Os modelos finais, gerados pelos Métodos da Máxima Verossimilhança e do Fator Principal (Quadro 8), apresentaram estruturas similares à do Método do Componente Principal, para os Fatores 1 e 2. Porém, no Fator 3, voltaram a incluir as variáveis V2 e V6, enquanto a variável V10 não foi "carregada" em nenhum dos três fatores extraídos. Nesse caso, os modelos iniciais, antes da rotação (Quadro 5), mostraram uma estrutura teórica mais lógica do que após a rotação (Quadro 8).

Foi eleito o modelo fatorial ajustado pelo Método do Componente Principal como o melhor.

No presente caso, gerou-se, então, a matriz de coeficientes dos escores dos fatores (Quadro 9) e finalmente, calculou-se os escores de cada um dos três fatores, pelo método da regressão que, no caso do Método do Componente Principal, dá resultados coincidentes aos obtidos pelo Método de Bartlett. Isso foi feito para cada uma das 21 unidades de amostra (Quadro 10). Assim, cada unidade de amostra passou a ser caracterizada não mais por 10 variáveis, mas, agora, por apenas três.

Quadro 9 – Matriz de coeficientes dos escores dos fatores rotacionados, extraídos pelo Método do Componente Principal

Variáveis	Fator 1	Fator 2	Fator 3
V1	0,24141	0,04439	-0,08756
V2	0,20680	-0,00995	0,06984
V3	0,02604	0,29312	-0,04615
V4	0,00486	0,32414	-0,05409
V5	0,02156	0,32594	-0,06043
V6	0,20966	-0,03077	0,14565
V7	0,22121	0,05042	-0,27962
V8	0,22271	-0,00490	0,06393
V9	-0,03525	0,19483	0,44637
V10	-0,01520	-0,05101	0,71375

Quadro 10 – Matriz de escores dos fatores, extraídos pelo Método do Componente Principal

Unid. Amostra	Fator 1	Fator 2	Fator 3
1	-0,30147	0,30655	-0,02706
2	-0,39386	0,54345	-0,52960
3	-0,30241	-0,24660	-0,43030
4	0,52647	0,20411	0,59964
5	-0,37114	-1,01120	-0,73891
6	-0,16489	-1,33242	-0,78151
7	-0,10108	0,01128	-0,03215
8	-0,22096	-1,50556	2,84876
9	-0,15365	-0,38631	-0,92570
10	0,70846	0,82952	-1,69472
11	-0,62582	-0,00270	0,53879
12	0,18691	0,19270	0,04780
13	-0,22436	2,25605	-0,16686
14	-0,32563	1,19761	-0,49028
15	-0,30860	0,46846	0,04205
16	-0,24847	-1,32062	-0,54128
17	-0,43633	-0,61489	-0,29308
18	-0,26252	-0,41754	1,10724
19	4,12529	-0,18436	0,31015
20	-0,49336	1,91736	1,85108
21	-0,61257	-0,90488	-0,69407

De posse dos escores dos fatores, e a título de ilustrar como poderia ser feito um posterior estudo baseado nos novos fatores, procedeu-se o ajuste de uma regressão linear múltipla, considerando os três fatores como as variáveis independentes (Quadro 10) e o índice de Sítio (Quadro 1) como a variável dependente do modelo (Quadro 11). Nesse caso, tem-se a certeza absoluta de que o modelo linear ajustado não apresenta problemas de multicolinearidade, o que certamente não ocorreria, caso fossem utilizadas as variáveis originais, devido às altas correlações entre algumas delas (Quadro 3).

A equação final ajustada, considerando o nível de 5% de significância, foi a seguinte:

$$\hat{IS} = 15,671429 + 1,769236 (\text{Fator 1}) + 1,735753 (\text{Fator 3}) + 1,603322 (\text{Fator 2})$$

$$R^2 = 56,62\%$$

Quadro 11 - Ajuste de modelo linear múltiplo pelo método Stepwise conforme saída das análises do SPSS

Equação Número 1 - Variável Incluída no Passo Número 1. FATOR 1 REGR FATOR ESCORE 1

R Múltiplo .45099
 R^2 .20339
 R^2 Ajustado .16146
 Erro Padrão (SE) 3.59238
 F = 4.85106 Signif. F = .0402

----- Variável na Equação ----- Variáveis não incluídas na Equação -----

Variável	B	SE B	Beta	F	Sig F	Variável	Beta In	Parcial	Min Toler	F	Sig F
FATOR 1	1.7692	.803281	.450987	4.851	.0402	FATOR 2	.408695	.457906	1.000000	4.776	.0423
Constante	15.6714	.783922		399.642	.0000	FATOR 3	.442452	.495728	1.000000	5.865	.0262

* * * * *

Equação número 2 - Variáveis Incluídas no Passo Número 2. FATOR 3 REGR FATOR ESCORE 3

R Múltiplo .63179
 R^2 .39915
 R^2 Ajustado .33239
 Erro Padrão (SE) 3.20540
 F = 5.97887 Signif. F = .0102

----- Variáveis na Equação ----- Variável não incluída na Equação -----

Variável	B	SE B	Beta	F	Sig F	Variável	Beta In	Parcial	Min Toler	F	Sig F
FATOR 1	1.769236	.716748	.450987	6.093	.0238	FATOR 2	.408695	.527251	1.000000	6.546	.0204
FATOR 3	1.735753	.716748	.442452	5.865	.0262						
(Constante)	15.671429	.699475		501.964	.0000						

* * * * *

Equação número 3 - Variável Incluída no Passo Número 3. FATOR 2 REGR FATOR ESCORE 2

R Múltiplo .75245
 R^2 .56619
 R^2 Ajustado .48963
 Erro Padrão (SE) 2.80262
 F = 7.39575 Signif. F = .0022

----- Variáveis na Equação -----

Variável	B	SE B	Beta	F	Sig F
FATOR 1	1.769236	.626685	.450987	7.970	.0117
FATOR 3	1.735753	.626685	.442452	7.671	.0131
FATOR 2	1.603322	.626685	.408695	6.546	.0204
(Constante)	15.671429	.611581		656.612	.0000

O modelo linear obtido explicou 56.62% da variação do índice de Sítio, apresentando ainda um R^2 ajustado de 48,96%. As contribuições individuais dos fatores foram: 20,34% para o **Fator 1 (fertilidade)**; 19,57% para o **Fator 3 (textura)**; e 16,71 para o **Fator 2 (mineralogia)**, sendo todos significativos ao nível de 5%, pelo teste F (Quadro 11).

Esse resultado, aparentemente, indica um modelo com baixo poder preditivo. Entretanto, a maioria dos trabalhos de relação solo-sítio apresenta níveis de explicação comumente não superiores a 70%. As grandes diferenças de localização das parcelas (Quadro 1), certamente, implicam em grandes variações climáticas e fisiográficas, que não foram contempladas no presente estudo.

~~~~~  
Por outro lado, os dados de solo foram coletados após o crescimento da floresta e em diferentes idades (de 7 a 48 anos). Isso compromete a fidedignidade dos dados, principalmente com relação à fertilidade do solo, pois as árvores absorvem nutrientes e conseqüentemente, alteram as características químicas do solo. Numa condição ideal, dever-se-ia trabalhar com dados de análise de solo, coletados antes do plantio da floresta.

Diante disso, pode-se concluir que o modelo fatorial apresentou bons resultados e ofereceu subsídios para a melhor compreensão dos fatores de solo que determinam a produtividade de plantios de *Araucaria angustifolia*, bem como permitiu o ajuste de um modelo linear relativamente simples para explicar, quantitativamente, a contribuição relativa de cada um dos fatores extraídos.