

ESTATÍSTICA MULTIVARIADA: APLICAÇÃO DA ANÁLISE FATORIAL NA ENGENHARIA DE PRODUÇÃO

MULTIVARIATE STATISTICS: FACTORIAL ANALYSIS APPLICATION IN PRODUCTION ENGINEERING

Hanne Alves Bakke¹, Alexandre Santos de Moura Leite², Luiz Bueno da Silva³

¹Universidade Federal da Paraíba – UFPB – João Pessoa – Brasil

h_bakke@yahoo.com.br

²Universidade Federal da Paraíba – UFPB – João Pessoa – Brasil

asmleite@gmail.com

³Programa de Pós-Graduação em Engenharia de Produção, UFPB, João Pessoa – Brasil

bueno@ct.ufpb.br

Resumo

A Estatística Multivariada permite o estudo de fenômenos complexos, pois realiza o tratamento de diversas variáveis simultaneamente, mesmo quando não se conhece o modelo teórico das relações entre as variáveis. A Análise Fatorial, uma das técnicas da estatística multivariada, pode ser aplicada quando há uma grande quantidade de dados e busca-se a sua sumarização com fatores latentes, que são definidos através de comportamento semelhantes entre as variáveis. Este artigo tem como objetivo fazer uma revisão da literatura sobre a Estatística Multivariada de dados e mostrar a aplicação dessa técnica de análise na Engenharia de Produção. O Inventário de Estresse em Enfermeiros, composto por 44 variáveis, foi aplicado a 391 indivíduos. A aplicação da Análise Fatorial extraiu oito fatores determinantes para o estresse, que pode servir como subsídio para seu gerenciamento.

Palavras-chave: estatística multivariada; análise fatorial; engenharia de produção.

1. Introdução

Atualmente, há uma vasta quantidade de dados que cobrem vários aspectos das atividades humanas e que dizem respeito às diferentes áreas de conhecimento. A complexidade de vários fenômenos requer uma análise de muitas variáveis diferentes. Além disso, o avanço da tecnologia computacional tem proporcionado avanços e facilidades no que diz respeito à análise destes dados em diversas áreas. Assim, percebe-se uma necessidade de procedimentos de análise e interpretação adequados para a extração das informações, principalmente para tomada correta de decisões de forma a possibilitar um gerenciamento mais eficiente e um crescimento na produtividade.

De acordo com Escofier e Pages (1992), os métodos de análise de dados multivariados têm comprovado amplamente sua eficácia no estudo de grandes massas de informação complexas. Tratam-se de métodos chamados de multidimensionais que permitem a confrontação entre duas ou mais variáveis. Pode-se, então, extrair-se as tendências mais sobressalentes e hierarquizá-las, eliminando os efeitos que perturbam a percepção global.

A estatística multivariada pode ser aplicada com diversas finalidades, mesmo nos casos em que não se dispõe de antemão de um modelo teórico rigorosamente estruturado a respeito das relações entre as variáveis. A finalidade de sua aplicação pode ser de reduzir dados ou de simplificação estrutural, de classificar e agrupar, de investigar a dependência entre variáveis, de predição e de elaborar hipóteses e testá-las (JOHNSON; WICHERN, 1992). Técnicas analíticas multivariadas têm sido utilizadas amplamente em indústrias, governos e em centros de pesquisas acadêmicas de diversas áreas (Psicologia, Educação, Geologia, Ciências Sociais, Engenharias, Ergonomia, etc.).

A estatística multivariada pode ser aplicada com diversas finalidades, mesmo quando não se dispõe de antemão de um modelo teórico rigorosamente estruturado das relações entre as variáveis. A finalidade de sua aplicação pode ser de reduzir dados ou de simplificar a sua estrutura, de classificar e agrupar, de investigar a dependência entre variáveis, de predizer e de elaborar hipóteses e testá-las (JOHNSON; WICHERN, 1992). Técnicas analíticas multivariadas têm sido utilizadas amplamente em indústrias, governos e em centros de pesquisas acadêmicas de diversas áreas.

Isso em mente, este artigo tem como objetivo fazer uma revisão da literatura sobre a Estatística Multivariada de dados e mostrar a aplicação da técnica de análise na Engenharia de Produção.

2. Análise de Dados Multivariados

A Estatística Multivariada, segundo Míngoti (2005), consiste em um conjunto de métodos aplicados em situações onde várias variáveis são medidas simultaneamente em cada elemento amostral. Uma observação multivariada de dimensão p , ou p -variada, é um vetor:

$$X = (X_1, X_2 \dots X_p)^t$$

onde t = transposto e cujas coordenadas X_1 a X_p são variáveis aleatórias oriundas de várias medidas de um mesmo elemento amostral.

Assim, uma matriz de observações p -variadas, ou matriz de dados, é uma tabela $X: n \times p$. Dessa forma, Johnson e Wichern (1992), afirmam que n medidas em p variáveis podem ser dispostas da seguinte forma:

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1j} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2j} & \dots & X_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{i1} & X_{i2} & \dots & X_{ij} & \dots & X_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{p1} & X_{p2} & \dots & X_{pj} & \dots & X_{pn} \end{bmatrix}$$

onde X é a matriz que contém todas as observações de todas as variáveis. Quanto maior o número de variáveis, mais complexa é a análise através de técnicas univariadas.

Um vetor aleatório é aquele cujos elementos integrantes são variáveis aleatórias. De forma semelhante, uma matriz aleatória é composta por variáveis aleatórias. O valor esperado de uma matriz (ou vetor) aleatória, segundo Johnson e Wichern (1992), é a matriz que consiste dos valores esperados de cada um de seus elementos. Assim, sendo X um vetor aleatório, o vetor de médias de X pode ser representado por:

$$\mu = E(X) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

onde $\mu_i = E(X_i)$ denota a média, ou esperança, da variável aleatória X_i , $i=1, 2, \dots, p$. Esta medida é bastante utilizada para sintetizar a tendência central da distribuição de valores da variável X_i (MINGOTI, 2005).

Cada variável, segundo Mingoti (2005), pode ser analisada separadamente no vetor aleatório através do comportamento de sua distribuição de probabilidades. Ressalva-se, no entanto, a importância da análise do vetor como um todo devido à possibilidade da existência de relacionamentos entre as p -variáveis.

A variância do i -ésimo componente do vetor X é obtida por $\text{Var}(X_i) = \sigma_i^2 = \sigma_{ii}$, assim como a variância entre os valores da i -ésima e j -ésima variáveis do vetor X é definida por $\text{Cov}(X_i, X_j) = \sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)]$. A independência estatística tem uma implicação importante na covariância. Se as variáveis estão distribuídas de forma independente, a covariância e, portanto, a correlação, é nula. No entanto, de acordo com Morrison (1976), o inverso geralmente não é verdadeiro, pois é possível construir exemplos de variáveis aleatórias altamente dependentes cuja correlação é zero.

Quando a matriz X é p -dimensional multivariada, então as covariâncias do vetor aleatório são dispostas em uma matriz simétrica de variância-covariância $\Sigma = E(X - \mu)(X - \mu)'$. Assim, segundo Johnson e Wichern (1992), a covariância das variáveis do vetor X é definida por:

$$\text{Cov}(\mathbf{X}) = \Sigma = E(\mathbf{X} - \mu)(\mathbf{X} - \mu)' = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

Ademais, de acordo com Härdle e Simar (2007), a correlação é uma medida mais adequada para avaliar o grau de relacionamento linear entre variáveis estatísticas. A vantagem de sua utilização inclui a ausência de escalas, o que significa que a mudança nas escalas de medidas não altera o valor da correlação, além de estar relacionada com a independência de variáveis aleatórias. Assim, para um vetor p-dimensional ($X_1 \dots X_p$), a matriz de correlação é denotada por:

$$\mathbf{P} = \begin{bmatrix} \mathbf{1} & \rho_{12} & \rho_{13} & \cdots & \rho_{1p} \\ \rho_{21} & \mathbf{1} & \rho_{23} & \cdots & \rho_{2p} \\ \rho_{31} & \rho_{32} & \mathbf{1} & \cdots & \rho_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \rho_{p3} & \cdots & \mathbf{1} \end{bmatrix}$$

Johnson (1998) *apud* Mingoti (2005) afirma que, para que sejam estimadas as matrizes de covariâncias e correlações, é necessário ter-se um tamanho de amostra n maior que p . Desta forma, a estimação das matrizes ocorre com $p+1$ observações amostrais. Todavia, deve-se evitar trabalhar com este número mínimo de n .

O constructo da análise multivariada é a variável estatística, uma combinação linear de variáveis às quais são atribuídos pesos empiricamente determinados. As variáveis são determinadas pelo pesquisador, e os pesos pela técnica multivariada para atingir o objetivo almejado. Assim, de acordo com Hair *et al.* (2005), uma variável estatística de n variáveis (X_1 à X_n) pode ser definida da seguinte forma:

$$\text{Valor da variável estatística} = w_1 X_1 + w_2 X_2 + w_3 X_3 \dots + w_n X_n$$

onde X_n é a variável observada e w_n é o peso que foi determinado através da técnica multivariada. Com isso, obtém-se um único valor que representa uma combinação de todo o conjunto das variáveis que melhor representa o objetivo da análise multivariada específica. Combinações lineares têm um importante papel na análise de dados multivariados, pois, assim, pode-se sumarizar as informações das p-variáveis originais (MINGOTI, 2005).

2.1 Tipos de técnicas Multivariadas

A análise de dados multivariados conta com diversas técnicas. Dentre elas, pode-se citar a (a) análise fatorial, (b) a regressão múltipla e correlação múltipla, (c) a análise discriminante

múltipla, (d) a análise multivariada de variância e covariância, (e) a análise conjunta, (f) a correlação canônica, (g) a análise de agrupamentos e (h) escalonamento.

- a) **Análise fatorial:** inclui a análise de componentes principais e análise dos fatores comuns. É aplicada quando há um número grande de variáveis e correlacionadas entre si, com o objetivo de identificar um número menor de novas variáveis alternativas, não correlacionadas e que, de algum modo, resumem as informações principais das variáveis originais encontrando os fatores ou variáveis latentes (MINGOTI, 2005);
- b) **Regressão múltipla:** é o método de análise adequado quando há uma única variável dependente métrica relacionada a duas ou mais variáveis independentes. O principal objetivo de sua aplicação é prever as mudanças na variável dependente de acordo com as variações nas variáveis independentes (HAIR *et al.*, 2005);
- c) **Análise discriminante múltipla:** este tipo de análise é utilizado quando os grupos são conhecidos *a priori*. A análise discriminante é composta por um conjunto de métodos e ferramentas utilizados para distinguir grupos de populações e classificar as novas observações nos grupos determinados (HÄRDLE; SIMAR, 2007);
- d) **Análise multivariada de variância e covariância:** também conhecidas como MANOVA (análise multivariada de variância) e MANCOVA (análise multivariada de covariância), têm o objetivo de verificar a semelhança entre grupos multivariados explorando simultaneamente as relações entre diversas variáveis independentes e duas ou mais variáveis dependentes métricas (HAIR *et al.*, 2005);
- e) **Análise conjunta:** segundo Hair *et al.* (2005), esta é uma técnica de dependência que vem sendo utilizada na avaliação de objetos, tais como produtos novos, serviços ou idéias. A aplicação mais direta é no desenvolvimento de novos produtos e serviços, permitindo a avaliação de produtos complexos e mantendo um contexto realista de decisão para o respondente;
- f) **Correlação canônica:** tem como objetivo principal o estudo das relações lineares existentes entre dois conjuntos de variáveis. A aplicação desta análise resume a informação de cada conjunto de variáveis-resposta em combinações lineares buscando-se maximizar a correlação entre os dois conjuntos (MINGOTI, 2005);
- g) **Análise de agrupamento:** trata-se de uma análise que identifica grupos em objetos de dados multivariados. O objetivo é formar grupos com propriedades homogêneas de amostras heterogêneas grandes. Devem-se buscar grupos mais homogêneos possíveis e que as diferenças entre os eles sejam as maiores possíveis (HÄRDLE e SIMAR, 2007);

fatorial assume que as variáveis estão relacionadas linearmente com novas variáveis aleatórias F_j , $j=1, 2, \dots, m$, que deverão ser identificados. O vetor ε_{pxl} representa os erros aleatórios de medida e à variação de Z_i , que não é explicada pelos fatores comuns F_j do modelo. O coeficiente l_{ij} , denominado de *loading*, corresponde ao grau de relacionamento linear entre Z_i e F_j . Assim, pode-se dizer que a informação das p -variáveis originais padronizadas (Z_1, Z_2, \dots, Z_p) é representada por $(p+m)$ variáveis aleatórias não observáveis, isto é, $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p)$ e (F_1, F_2, \dots, F_p) .

Para a continuação da análise fatorial é necessário estimar a matriz de correlação teórica P_{pp} , através de uma matriz de correlação amostral R_{pp} . Para estimar m , basta extrair os autovalores da matriz R_{pp} e ordená-los em ordem decrescente. Assim, segundo Mingoti (2005) para a seleção do número de fatores pode-se utilizar os critérios:

- a) **Análise da proporção da variância total:** permanecem aqueles que representam maiores porções da variância total e, portanto, m será igual ao número de autovalores retidos. Malhotra (2001) recomenda que o número de fatores escolhidos corresponda a, no mínimo, 60% da variância;
- b) **Autovalores:** o valor de m será igual ao número de autovalores maiores ou igual a 1. Esse critério mantém no sistema dimensões que representam pelo menos a informação de variância de uma variável original;
- c) **Scree-plot:** representação gráfica onde prure-se um “ponto de salto”, que represente um decréscimo de importância em relação à variância total. O valor de m seria, então, igual ao número de autovalores anteriores ao “ponto de salto”.

A matriz de componentes inicial indica a relação entre os fatores e as variáveis individuais, no entanto, esta raramente resulta em fatores que possam ser interpretados. A Análise Fatorial é mais ou menos útil em função da sua capacidade de produzir tais fatores, e isto é feito através da rotação, que transforma a matriz de fatores em uma matriz rotacionada mais simples e fácil de interpretar (MALHOTRA, 2001). De acordo com Hair *et al.* (2005), 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 mais significativo. Há diferentes tipos de rotações possíveis de aplicação. A rotação ortogonal inclui a QUARTIMAX, EQUIMAX, e a VARIMAX, que é a mais utilizada por se concentrar na máxima simplificação das colunas da matriz fatorial através da maximização da soma de variâncias de cargas exigidas da matriz fatorial. Os métodos de rotação oblíqua permitem fatores correlacionados em vez de manterem independência entre os fatores rotacionados, mas há apenas escolhas limitadas nos pacotes estatísticos.

De acordo com Hair *et al.* (2005) o pesquisador deve buscar minimizar o número de variáveis incluídas, mas manter um número razoável de variáveis por fator. A força da análise fatorial consiste na definição de padrões entre grupos de variáveis, tornando-se de pouco utilidade

na identificação de fatores compostos por uma única variável. Além disso, deve-se dar cuidado especial ao tamanho da amostra. Como regra geral, o mínimo é ter pelo menos cinco vezes mais observações do que o número de variáveis a serem analisadas, e a proporção mais aceitável é de dez para um.

As principais estatísticas-chaves associadas à análise fatorial estão listados no Quadro 1.

Quadro 1: Estatísticas da Análise Fatorial

Teste de Esfericidade de Bartlett	Estatística de teste usada para examinar a hipótese de que as variáveis não sejam correlacionadas na população, ou seja, a matriz de correlação da população é uma matriz identidade, onde cada variável se correlaciona perfeitamente com ela própria ($r=1$), mas não apresenta correlação com as outras variáveis ($r=0$). A significância para o teste não deve ultrapassar 0,05.
Matriz de Correlação	O triângulo inferior da matriz que exibe as correlações simples entre todos os pares possíveis de variáveis incluídos na análise. Os elementos da diagonal, que são todos igual a 1, em geral são omitidos
Comunalidade	Porção da variância que uma variável compartilha com todas as outras variáveis consideradas. É também a proporção de variância explicada pelos fatores comuns
Autovalor (Eigenvalue)	Representa a variância total explicada por cada fator.
Cargas dos fatores	Correlações simples entre as variáveis e os fatores
Gráfico de cargas dos fatores	Gráfico das variáveis originais, utilizando as cargas dos fatores como coordenadas
Matriz de Fatores	Cargas dos fatores de todas as variáveis em todos os fatores extraídos
Escores fatoriais	Escores compostos estimados para cada entrevistado nos fatores derivados
Medida de Adequação de Kaiser-Meyer-Olkin (KMO)	Índice usado para avaliar a adequação da análise fatorial. Os valores obtidos através deste teste variam entre 0 e 1, onde 0,80 ou acima é admirável; 0,70 ou acima, mediano; 0,60 ou acima, medíocre; 0,50 ou acima, ruim; e abaixo de 0,50, inaceitável
Porcentagem de variância	Porcentagem da variância total atribuída a cada fator
Resíduos	Diferenças entre as correlações observadas na matriz de correlação de entrada e as correlações reproduzidas, conforme estimadas na matriz de fatores
Scree plot	Gráfico dos autovalores versus o número de fatores, por ordem de extração

Fonte: Adaptado de Malhotra (2001)

Todavia, um fator crucial que deve anteceder a aplicação da análise fatorial é a análise de confiabilidade do constructo utilizado para obtenção dos dados. A confiabilidade é o grau em que uma escala produz resultados consistentes entre medidas repetidas ou equivalentes de um mesmo objeto, relevando-se a ausência do erro. Trata-se da análise das escalas de mensuração, que permite determinar a extensão em que os itens estão relacionados com os demais e a fidedignidade do constructo. De acordo com Corrar *et al.* (2007), um modelo bastante utilizado é o Alfa de Cronbach, que trata da consistência interna baseada na correlação média entre os itens. Este modelo pode ser obtido através da seguinte fórmula:

$$\alpha = \frac{k \left(\frac{cov}{var} \right)}{1 + (k - 1) \left(\frac{cov}{var} \right)}$$

onde k = número de variáveis consideradas; cov é a média das covariâncias e var é a média das variâncias. Alfa pode assumir valores entre 0 e 1, e quanto mais próximo ao 1, maior a fidedignidade das dimensões do constructo. Há divergências sobre qual o valor mínimo aceitável para o Alfa de Cronbach, no entanto Hair (1998) *apud* Corrar *et al.* (2007) sugere o mínimo ideal de 0,7 para pesquisas aplicadas e 0,6 para pesquisas exploratórias.

3.1 Aplicação da Técnica Multivariada Análise Fatorial

O Inventário de Estresse em Enfermeiros proposto por Stacciarini and Trocoli (2000), composto por 44 questões de múltipla escolha e sete de identificação, foi aplicado a 391 profissionais médicos que trabalham em hospitais na cidade de João Pessoa (PB). As questões fechadas apresentavam como alternativas de respostas: nunca, raramente, algumas vezes, muitas vezes e sempre, às quais foram atribuídas os valores, 1, 2, 3, 4, 5 respectivamente.

Analisando-se a estrutura do Inventário utilizado para obtenção dos dados através do modelo de Alfa de Cronbach, percebe-se que o constructo e a escala utilizada apresentam uma boa confiabilidade interna e, portanto, são adequados para o propósito para o qual foram designados (Quadro 2).

Quadro 2 – Teste Alfa de Cronbach para as 44 variáveis

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N Items
0,914	0,921	44

Fonte: Pesquisa de campo (2008)

Para verificar a aplicação da Análise Fatorial (AF) utilizou-se o teste KMO de medida de adequação da amostra. O resultado expresso neste teste (0,881) demonstra um bom grau de ajuste para aplicação da técnica multivariada Análise Fatorial, sendo ratificado pelo teste Bartlett de Esfericidade ao nível de significância 0,000 (Quadro 3).

Quadro 3 – Teste KMO e Bartlett

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0,881
Bartlett's Test of Sphericity	Approx. Chi-Square	6240,701
	df	946
	Sig.	0,000

Fonte: Pesquisa de campo (2008)

Através da Análise de Componentes Principais, foi observado que 60,8% do total da variância pode ser explicada por 10 fatores. O mesmo resultado foi obtido ao se fazer a análise dos autovalores maiores que 1,0. Estes resultados podem ser observados na Tabela 1.

Tabela 1 – Autovalores e porcentagem da variância explicada

Componentes	Autovalores Iniciais		
	Total	% da Variância	% Var Acumulada
1	10,276	23,355	23,355
2	4,822	10,960	34,315
3	2,075	4,717	39,032
4	1,820	4,136	43,168
5	1,592	3,619	46,787
6	1,529	3,475	50,262
7	1,374	3,123	53,385
8	1,145	2,602	55,987
9	1,070	2,431	58,419
10	1,049	2,384	60,803
11	0,998	2,268	63,071
12	0,941	2,138	65,209

Fonte: Pesquisa de campo (2008)

Os fatores extraídos na análise, extraídos através da Rotação Varimax, foram submetidos ao teste de confiabilidade Alfa de Cronbach, onde o valor inferior mínimo é de 0,6 por se tratar de uma pesquisa exploratória. Dois dos fatores apontados pela aplicação da técnica apresentaram valores abaixo daquele estipulado como limite. Um dos fatores, com $\alpha=0,529$, é composto pelas variáveis: conciliar as questões profissionais com as familiares, atender familiares de pacientes, atender um grande número de pessoas; o outro fator obteve $\alpha=0,126$ e é composto pelas variáveis: fazer um trabalho repetitivo, sentir-se impotente diante das tarefas a serem realizadas e receber este salário.

Sendo assim, as variáveis que faziam parte destes dois fatores foram desconsideradas e uma nova extração foi realizada. O Quadro 4 mostra as variáveis que formam cada fator, com suas respectivas cargas, bem como o valor do Alfa de Cronbach para cada um dos fatores.

Quadro 4 – Composição dos fatores com respectivas cargas

FATOR	VARIÁVEIS	CARGA	ALFA DE CRONBACH (α)
1	Relacionamento com a equipe médica (REM)	0,881	0,886
	Relacionamento com os colegas médicos (RCM)	0,865	
	Trabalhar em equipe (TE)	0,819	
	Relacionamento com a chefia (RC)	0,806	
	Prestar assistência ao paciente (PAP)	0,711	
	Manter-se atualizado (MA)	0,518	
	A especialidade em que trabalha (EQT)	0,517	
	Dedicação exclusiva à profissão (DEP)	0,488	
2	Trabalhar em instalações físicas inadequadas (TII)	0,816	0,822
	Falta de material necessário ao trabalho (FMT)	0,808	
	Falta de recursos humanos (FRH)	0,703	
	Trabalhar com pessoas despreparadas (TPD)	0,659	
	Falta de espaço no trabalho para discutir as experiências, tanto as positivas como as negativas (FED)	0,560	
	Trabalhar em ambiente insalubre (TAI)	0,485	
	Distanciamento entre a teoria e a prática (DTP)	0,425	
	3	Restrição da autonomia profissional (RAP)	
Indefinição do papel do médico (IPM)		0,684	
Interferência da política institucional no trabalho (IPI)		0,674	
Impossibilidade de prestar assistência direta ao paciente (IAD)		0,619	
Ter um prazo curto para cumprir ordens (PCO)		0,545	
Trabalhar em clima de competitividade (TCC)		0,479	
Responsabilizar-se pela qualidade de serviço que a instituição presta (RQS)		0,455	
4	Responder por mais de uma função neste emprego (RMF)	0,711	0,747
	Levar serviço para fazer em casa (LSC)	0,692	
	Administrar ou supervisionar o trabalho de outras pessoas (ATP)	0,664	
	Desenvolver atividades além da minha função ocupacional (AAF)	0,571	
	Cumprir na prática uma carga horária maior (CHM)	0,465	
5	Desenvolver pesquisa (DP)	0,668	0,710
	Executar procedimentos rápidos (EPR)	0,626	
	Ensinar o aluno (EA)	0,595	
	Prestar assistência a pacientes graves (APG)	0,465	
6	Fazer esforço físico para cumprir o trabalho (EFT)	0,702	0,645
	Sentir desgaste emocional com o trabalho (DET)	0,689	
7	Trabalhar em horário noturno (THN)	0,808	0,611
	Fazer turnos alternados de trabalho (TAT)	0,618	
8	Começar em uma função nova (CFN)	0,680	0,606
	Executar tarefas distintas simultaneamente (TDS)	0,664	
	Resolver imprevistos que acontecem no local de trabalho (RIT)	0,514	

Fonte: Pesquisa de campo (2008)

Tendo em vista que os fatores podem ser representados pelo relacionamento linear das variáveis, os fatores determinantes para o surgimento do estresse que resultaram deste estudo podem ser expressos pelas seguintes equações:

$$F_1 = 0,881REM + 0,865RCM + 0,819TE + 0,806RC + 0,711PAP + 0,518MA + 0,517EQT + 0,488DEP$$

$$F_2 = 0,816TII + 0,808FMT + 0,703FRH + 0,659TPD + 0,560FED + 0,485TAI + 0,425DTP$$

$$F_3 = 0,796RAP + 0,684IPM + 0,674IPI + 0,619IAD + 0,545PCO + 0,479TCC + 0,455RQS$$

$$F_4 = 0,711RMF + 0,692LSC + 0,664ATP + 0,571AAF + 0,460CHM$$

$$F_5 = 0,668DP + 0,626EPR + 0,595EA + 0,465APG$$

$$F_6 = 0,702EFT + 0,689DET$$

$$F_7 = 0,808THN + 0,618TAT$$

$$F_8 = 0,680CFN + 0,664TDS + 0,514RIT$$

4. Conclusão

A Estatística Multivariada é uma ferramenta poderosa na análise de dados, pois consiste em um conjunto de métodos estatísticos que permite confrontar diversas variáveis simultaneamente de cada elemento amostral. As técnicas de estatística multivariada têm o propósito de simplificar ou facilitar a interpretação do fenômeno estudado e o seu desenvolvimento tem possibilitado o estudo acurado de fenômenos cada vez mais complexos. Estas podem ser empregadas com o intuito de se construir índices ou variáveis alternativas e grupos de elementos amostrais, analisar as relações de dependência das variáveis e comparar as populações, facilitado atualmente pelo avanço tecnológico.

Isto em vista, a Análise Fatorial é uma técnica de Estatística Multivariada que tem como objetivo descrever a variabilidade de um vetor aleatório X , através de um número menor m de variáveis aleatórias, denominadas de fatores, *i.e.*, sumariza um conjunto de variáveis, das quais não se observa tendências explícitas, através da descoberta de fatores latentes que expliquem melhor este conjunto. A aplicação desta ferramenta é comprovada mediante a realização dos testes KMO e Bartlett de Esfericidade. O teste Alfa de Cronbach pode ser aplicada para verificar a confiabilidade interna do constructo (instrumento de coleta de dados) e dos fatores latentes encontrados.

No estudo conduzido com profissionais médicos da cidade de João Pessoa (PB), esta técnica demonstrou ser bem sucedida e adequada, pois atingiu o objetivo de reduzir o número inicial de 44 variáveis, com estrutura complexa e com relações não observáveis diretamente, para oito dimensões determinantes para o estresse. Além disso, a análise dos resultados possibilitou apontar as variáveis que possuem maior impacto dentro dos fatores (REM – 88,10%, TII – 81.60%, RAP – 79,60%, dentre outros), e estabelecer relações lineares entre todas as variáveis de um mesmo fator. Isto facilita a interpretação e a compreensão das variáveis iniciais, podendo servir como subsídio para tomadas de decisão no gerenciamento de estresse de forma pontual, promovendo uma melhora na produtividade e na qualidade de vida.

Pode-se concluir, então, que a estatística multivariada tem grande valia para a área de Engenharia de Produção por permitir uma maior clareza e melhor interpretabilidade dos fenômenos complexos com que lida; e nesse estudo de caso apresentado, a técnica utilizada proporcionou

verificar com exatidão o quanto uma variável terá influência em fatores associados ao estresse em médicos.

Abstract

Multivariate Statistics allows the study of complex phenomena due to the possibility of treatment of many variables simultaneously, even when the theoretic model of intervariable relations isn't known. Factorial Analysis, one of the techniques of multivariate statistics, can be used when there is a large number of data and the summarization is sought through latent factors that are defined according to the behavior within the variables. This article has the objective of making a literature review about Multivariate Statistics and demonstrating the Factorial Analysis technique application in Production Engineering. The Inventário de Estresse em Enfermeiros is composed of 44 variables and was applied to 391 individuals. The Factorial Analysis extracted eight determining stress factors that can assist its management.

Keywords: multivariate statistics, factorial analysis, production engineering.

Referências

CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M. **Análise multivariada**. FIPECAFI: Atlas, 2007.

CRIVISQUI, E. M. **Análisis factorial de correspondencias: um instrumento de investigación em ciências sociais**. Universidad Católica de Assunción: Ed. Laboratório de Informática Social, 1993.

ESCOFIER, B.; PAGÈS, J. **Análisis factorialies simples y multiples: objetctivo, métodos y interpretación**. Bilbao. Ed. Universidad del Pais Vasco, 1992.

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

HÄRDLE, W.; SIMAR, L. **Applied multivariate statistical analysis**. 2. ed. Berlin: Springer, 2007.

JOHSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 3. ed. Englewood Cliffs: Prentice Hall, 1992.

MALHOTRA, N. K. **Pesquisa de marketing: uma orientação aplicada**. 3ª ed. Porto Alegre: Ed. Bookman, 2001.

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

MORRISON, D. F. **Multivariate statistical methods**. New York: McGraw-Hill, 1976.

STACCIARINI, J. M. R.; TROCOLI, B. T. Instrumento para mensurar o estresse ocupacional: inventário de estresse em enfermeiros (IEE). **Revista Latino-am. Enfermagem**. 2000; 8(6): 40-49.



Dados dos autores:

Nome completo: Hanne Alves Bakke

Filiação institucional: UFPB

Departamento: Programa de Pós-Graduação em Engenharia de Produção

Função ou cargo ocupado: Mestranda

Endereço completo para correspondência: Rua José Gomes de Sá Filho, 85. Apt. 203 A. Bairro: Bessa. João Pessoa (PB). CEP: 58.037-580.

Telefones para contato: 083 9971-5216

e-mail: h_bakke@yahoo.com.br

Nome completo: Alexandre Santos de Moura Leite

Filiação institucional: UFPB

Departamento: Programa de Pós-Graduação em Engenharia de Produção

Função ou cargo ocupado: Mestrando

Endereço completo para correspondência: Avenida Doutor Francisco Pinto, s/n, Bodocongó. CEP: 58.109-783.

Telefones para contato: 083 9372-1690

e-mail: asmleite@gmail.com

Nome completo: Luiz Bueno da Silva

Filiação institucional: UFPB

Departamento: Programa de Pós-Graduação em Engenharia de Produção

Função ou cargo ocupado: Professor Adjunto IV

Endereço completo para correspondência: Av. Acre, 130, Apto 201, Ed. Rosa Leal, Bairro dos Estados, João Pessoa, Paraíba, CEP: 58030-230

Telefones para contato: 83 32444166

e-mail: bueno@ct.ufpb.br