

**Antonio Faria Neto**

*Universidade Estadual Paulista Júlio  
de Mesquita Filho – UNESP,  
Guaratinguetá, SP, Brasil*  
*Universidade de Taubaté – UNITAU,  
Taubaté, SP, Brasil*

**Giovanna Faria**

*Universidade Estadual Paulista Júlio  
de Mesquita Filho – UNESP,  
Guaratinguetá, SP, Brasil*

**Correspondência/Contato**

UNIVERSIDADE DE TAUBATÉ  
Departamento de Engenharia Mecânica

Rua Daniel Danelli, s/n, Jd. Morumbi  
Taubaté - SP  
CEP 12060-440  
Fone (12) 3625-4193

**Editores responsáveis**

Prof. Dr. Luiz Eduardo Nicolini do P. Nunes  
luiz.nunes@unitau.com.br

Prof. Dra. Valesca Alves Correa  
valesca.correa@unitau.com.br

# PROPOSTA DE UM MÉTODO PARA ORDENAÇÃO DE VARIÁVEIS QUANTO À SUA IMPORTÂNCIA

## RESUMO

A pesquisa do tipo *Survey* aplicada à identificação dos fatores que impactam um determinado fenômeno é realizada em duas etapas: uma exploratória, que visa à identificação de variáveis que afetam o fenômeno e uma quantitativa, que objetiva sintetizar as variáveis em um número menor de fatores subjacentes. A segunda etapa normalmente se utiliza de um questionário fechado para quantificar a importância de cada uma das variáveis aplicando-se uma escala psicométrica, como por exemplo a escala de Likert. Contudo, essa pode não ser a melhor maneira de se avaliar a importância de uma variável, ou fator, sobretudo se os respondentes atribuírem o mesmo nível de importância a mais de um deles. Outro inconveniente ocorre quando as variáveis apontadas pelos respondentes na etapa exploratória se correlacionam fracamente, inviabilizando a aplicação de técnicas fatoriais para a extração dos fatores. A análise de agrupamentos aplicadas às variáveis também pode não apresentar bons resultados. Nesses casos, a análise das variáveis deveria ficar restrita ao campo qualitativo. Como alternativa, este artigo propõe um método, que ordena, por importância, as variáveis, ou fatores, obtidos na etapa exploratória, além de permitir uma avaliação absoluta de sua importância, dando um caráter quantitativo à análise. Este método lança mão da instrumentação das técnicas clássicas de agrupamento, procurando associar as variáveis a postos de importância a que se assemelhem mais. O método proposto pode ser aplicado concomitantemente às técnicas multivariadas clássicas, uma vez que a tabulação inicial dos dados pode ser utilizada, como entrada para elas.

**Palavras-chave:** Análise de agrupamentos, análise multivariada, análise fatorial, ordenação.

## ABSTRACT

A Survey applied to determine the factors that most affect a particular phenomenon is usually performed in two stages: an exploratory, which aims to identify variables that affect the phenomenon and a quantitative one, that describes variability among the observed, correlated variables in terms of a lower number of unobserved variables called factors. The further step usually uses a questionnaire to quantify the importance of each variable applying a psychometric scale, such as the Likert scale. However, this may not be the best way to evaluate the importance of a variable, or factor, especially if the respondents attach the same level of importance to more than one of them. Another drawback occurs when the variables identified by the respondents in the exploratory stage are weakly correlated, preventing the application of factorial techniques for extracting factors. The cluster analysis applied to the variables also may not deliver results. In these cases, the analysis of the variables should be restricted to the qualitative field. Alternatively, this paper proposes a method that ranks, by relevance, variables, or factors, obtained from the exploratory stage, and allows an absolute assessment of its importance, giving a quantitative approach to the analysis. This method makes use of the instrumentation of classic clustering techniques, as it associates the variables to the ranks they are closer. The proposed method can be applied simultaneously to the classical multivariate techniques because its initial data table can be used as input to them.

**Keywords:** Cluster analysis, multivariate data analysis, factor analysis, ranking.

## 1 INTRODUÇÃO

A pesquisa do tipo *Survey* (MIGUEL, 2011), aplicada à identificação dos fatores que afetam um determinado fenômeno pode ser realizada em duas etapas: uma aberta visando à identificação das variáveis e outra fechada, onde essas variáveis podem ser sintetizadas em um número, potencialmente menor, de fatores subjacentes, não diretamente mensuráveis.

A primeira etapa é uma pesquisa exploratória, em que o pesquisador envia um questionário aberto a uma amostra da população de interesse, pedindo que se indique livremente as possíveis variáveis que impactam o fenômeno sob análise. Uma análise de conteúdo desses questionários permite elencar uma relação de variáveis influentes sobre o fenômeno.

A segunda etapa consiste em uma pesquisa fechada em que se procura quantificar a importância de cada uma das variáveis aplicando-se em geral uma escala psicométrica, como por exemplo a escala de *Likert* (MIGUEL, 2011). Se os respondentes atribuírem o mesmo nível de importância à maioria das variáveis, ou fatores, a avaliação da importância dessas variáveis pode não ser conclusiva.

Outro inconveniente ocorre quando as variáveis apontadas pelos respondentes na etapa exploratória são fracamente correlacionadas. Nesse caso, não será possível a aplicação de técnicas fatoriais para a extração de fatores subjacentes comuns, e tampouco o agrupamento das variáveis por meio dos algoritmos de *clustering* apresentará bons resultados (FREI, 2006; LATTIN et al, 2011). Por conta desses obstáculos, a análise das variáveis, muitas vezes, se limita aos aspectos qualitativos.

Como alternativa, este artigo propõe um método, para ordenar, por importância, as variáveis obtidas na etapa exploratória, além de permitir uma avaliação absoluta de sua importância, dando um caráter quantitativo à análise. Este método se utiliza da instrumentação das técnicas clássicas de agrupamento, procurando associar as variáveis a postos de importância.

Além disso, o método proposto e as técnicas multivariadas clássicas não são mutuamente excludentes.

## 2. O MÉTODO PROPOSTO

Considere que  $m$  indivíduos sejam chamados a classificar  $n$  variáveis segundo a sua percepção da importância de cada uma delas. Esses  $m$  indivíduos produziram  $m$  classificações para as  $n$  variáveis. Como a percepção de importância das variáveis é de cunho subjetivo, essa ordenação deverá ser diferente para cada um dos entrevistados, embora classificações coincidentes sejam possíveis.

O resultado final da aplicação deste método é a ordenação das variáveis segundo a opinião de todos os entrevistados.

Cada entrevistado deve atribuir um valor de 1 a  $n$  para cada uma das variáveis de acordo com a importância atribuída a ela. Por convenção, atribui-se 1 à variável mais importante e  $n$  à variável menos importante, sem repetição. Desta forma cada entrevistado ordenará as variáveis em ordem decrescente de importância. A Tabela 1 ilustra como essas classificações podem ser organizadas em uma matriz, que será denominada matriz de importância das variáveis.

**Tabela 1.** Matriz de importância das variáveis.

	$V^1$	$V^2$	...	$V^j$	...	$V^n$
Entrevistado $_1$						
Entrevistado $_2$						
⋮						
Entrevistado $_i$						
⋮						
Entrevistado $_m$						

Assim, cada uma dessas variáveis, passa a ser um ponto de um espaço  $m$ -dimensional, em que suas  $m$  coordenadas são a importância a elas atribuídas segundo os  $m$  entrevistados.

O método proposto, parte da premissa de que para  $n$  variáveis a serem ordenadas haverá  $n$  postos de importância. Considera-se que cada posto de importância seja ocupado por uma variável virtual, onde todas as suas coordenadas são numericamente iguais ao posto. As variáveis virtuais podem ser entendidas como sendo variáveis para as quais os respondentes teriam atribuído, unanimemente, a importância do posto que elas ocupam. Para maior compreensão, a Tabela 2 ilustra uma matriz de importância preenchida apenas com variáveis virtuais.

**Tabela 2.** Matriz de importância das variáveis virtuais

	VarVirt <sup>1</sup>	VarVirt <sup>2</sup>	...	VarVirt <sup>k</sup>	...	VarVirt <sup>n</sup>
Entrevistado $_1$	1	2		$k$		$n$
Entrevistado $_2$	1	2		$k$		$n$
⋮	⋮	⋮		⋮		⋮
Entrevistado $_i$	1	2		$k$		$n$
⋮	⋮	⋮		⋮		⋮
Entrevistado $_m$	1	2		$k$		$n$

Assim, o espaço das variáveis é composto por  $n$  variáveis reais e mais  $n$  variáveis virtuais.

Assim, o princípio de ordenação proposto consiste na formação de aglomerados em torno dos postos de importância, o que equivale a dizer, em torno das variáveis fictícias. Desta forma, em cada posto de importância, devem estar agrupadas as variáveis que mais se assemelham à variável fictícia ocupante do posto em questão. Isto pode fazer com que alguns postos de importância não sejam ocupados por nenhuma variável, enquanto outros sejam ocupados por mais de uma variável.

A similaridade entre as variáveis reais e os postos de importância pode ser avaliada por uma medida de distância entre a variável real e a variável fictícia ocupante dos postos, isto é, uma variável será mais semelhante à um determinado posto quanto menor for a distância entre ela e a variável fictícia que ocupa esse posto. Desta forma, cada variável será agrupada em torno do posto mais semelhante, ou seja, mais próximo.

A medida de distância utilizada para a formação dos agrupamentos será a distância euclidiana. Assim, a distância entre a variável  $j$  e o posto de importância  $k$ , é calculada conforme a equação (1).

$$d_{jk} = \sqrt{\sum_{i=1}^m (V_i^j - k)^2} \quad (1)$$

Essas distâncias podem ser resumidas na matriz de distâncias, em que as colunas representam as variáveis e as linhas, os postos.

### 3. CASO EXEMPLO

A fim de ilustrar a aplicação do método proposto, considere o caso apresentado por Rambo (2011), em que se estuda os fatores que impactam na eficiência das auditorias de certificação da qualidade no setor aeroespacial brasileiro do ponto de vista das organizações certificadas. Na fase exploratória deste estudo identificou-se as dezoito variáveis apresentadas no Quadro 1.

**Quadro 1.** Variáveis que impactam na eficiência das auditorias de certificação da qualidade no setor aeroespacial brasileiro.

VARIÁVEL	DESCRIÇÃO
Variável 1	Competência técnica do auditor
Variável 2	Atributos pessoais do auditor
Variável 3	Apresentação das constatações e conclusões da auditoria
Variável 4	Análise prévia do SGQ da organização
Variável 5	Comunicação durante a auditoria
Variável 6	Pró-atividade do auditado
Variável 7	Comprometimento da alta direção com o desenvolvimento, implantação e melhoria contínua do SGQ
Variável 8	Facilitação da interação do auditor com o nível operacional
Variável 9	Análise crítica dos resultados
Variável 10	Divulgar os resultados da auditoria
Variável 11	Planejamento para receber a auditoria
Variável 12	Preparação da documentação a ser auditada
Variável 13	Treinamento dos colaboradores
Variável 14	Disciplina no cumprimento de normas e procedimentos
Variável 15	Responder com transparência e honestidade
Variável 16	Comprometimento da equipe auditada
Variável 17	Tomar conhecimento dos resultados durante e ao final da auditoria
Variável 18	O gerente não deve relegar suas atribuições

As variáveis do Quadro 1 foram apresentadas a 25 organizações para que fossem ordenadas segundo a sua importância, resultando na matriz apresentada na Tabela 4.

**Tabela 4.** Classificação das 18 Variáveis, por ordem de importância, segundo as organizações respondentes da pesquisa.

ORG	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18
A	2	3	8	5	6	17	1	13	18	9	4	14	15	16	10	12	7	11
B	1	6	2	16	17	14	8	13	7	9	15	12	11	3	4	5	10	18
C	3	2	16	4	9	13	1	12	14	15	6	5	7	8	11	10	17	18
D	5	11	1	12	6	15	3	17	2	4	14	18	10	7	8	16	13	9
E	9	10	7	8	11	12	1	13	2	6	3	4	15	14	16	5	18	17
F	8	5	2	7	11	16	1	15	13	12	3	4	6	17	10	9	18	14
G	1	6	3	4	2	7	5	13	8	10	9	14	18	15	11	16	17	12
H	6	15	5	1	16	17	4	14	13	12	2	3	18	10	7	9	11	8
I	13	18	16	2	12	9	1	15	10	17	3	4	6	5	11	8	14	7
J	7	8	18	11	16	15	1	9	14	6	17	4	12	2	5	13	10	3
K	11	12	2	13	3	4	1	5	15	16	14	6	17	18	7	9	8	10
L	5	6	7	16	9	8	1	18	10	15	4	2	3	11	12	13	14	17
M	1	2	4	3	5	15	6	16	12	13	14	7	8	9	10	11	17	18
N	8	9	10	11	5	6	1	12	17	16	7	14	15	2	4	3	13	18
O	6	10	12	13	1	2	4	14	5	9	8	16	18	15	3	7	11	17
P	1	6	14	3	15	5	4	16	17	18	2	10	7	8	9	12	13	11
Q	15	16	17	8	6	5	1	13	7	18	2	3	4	9	10	11	12	14
R	5	10	4	12	6	8	1	13	14	9	17	16	15	2	7	11	18	3
S	7	8	16	13	9	15	1	14	17	18	6	2	3	4	11	10	12	5
T	5	6	7	16	8	9	1	10	18	15	4	3	2	12	13	14	17	11
U	7	18	14	8	16	17	2	9	6	5	1	3	11	4	10	15	12	13
V	2	10	7	11	3	6	1	5	12	14	13	15	16	4	9	8	17	18
W	2	3	4	5	6	7	1	8	9	10	11	12	13	14	15	16	17	18
X	8	10	15	4	9	7	1	11	16	17	5	6	12	2	13	3	18	14
Y	3	2	17	4	9	13	1	12	14	16	6	5	7	11	8	10	15	18

Na Tabela 4, as organizações respondentes foram identificadas de A até Y e as variáveis a serem ordenadas de V1 a V18. A título de ilustração, vê-se que a organização A classificou V7 como a variável mais importante e V9 como a menos importante, e a organização B apontou V1 como a mais importante e V18 como a menos importante.

Como são dezoito as variáveis a serem classificadas, os postos de importância também serão dezoito. A Tabela 5 apresenta a matriz de distâncias para esse estudo, calculada conforme (2). Por exemplo, a aplicação de (2) para o cálculo da distância entre V1 e o Posto 6 é feita conforme:

$$d_{V1-Post6} = \sqrt{(2-6)^2 + (1-6)^2 + (3-6)^2 + (5-6)^2 + (9-6)^2 + (8-6)^2 + \dots + (3-6)^2} = 18,68 \quad (2)$$

As demais distâncias são obtidas de forma análoga.

**Tabela 5.** Matriz de distâncias entre as variáveis e os postos de importância.

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18
<b>Posto 1</b>	29.73	44.03	49.55	43.69	44.53	52.74	11.05	59.25	57.99	60.75	41.73	44.01	54.99	47.06	44.80	49.79	66.84	64.16
<b>Posto 2</b>	26.02	39.87	45.54	39.55	40.32	48.29	9.54	54.46	53.46	56.10	37.89	40.10	50.61	42.97	40.17	45.19	62.01	59.57
<b>Posto 3</b>	22.80	35.93	41.75	35.62	36.32	43.97	10.49	49.71	49.02	51.52	34.37	36.46	46.35	39.08	35.65	40.68	57.20	55.04
<b>Posto 4</b>	20.32	32.28	38.24	32.00	32.59	39.80	13.38	45.01	44.70	47.02	31.24	33.17	42.25	35.47	31.27	36.30	52.43	50.60
<b>Posto 5</b>	18.87	29.03	35.09	28.79	29.24	35.85	17.26	40.39	40.53	42.64	28.65	30.35	38.37	32.23	27.11	32.11	47.71	46.27
<b>Posto 6</b>	18.68	26.34	32.40	26.15	26.42	32.19	21.61	35.86	36.58	38.41	26.76	28.14	34.77	29.50	23.28	28.18	43.05	42.10
<b>Posto 7</b>	19.80	24.39	30.32	24.27	24.31	28.93	26.19	31.48	32.91	34.38	25.71	26.70	31.56	27.40	19.97	24.64	38.47	38.12
<b>Posto 8</b>	22.02	23.37	28.95	23.32	23.11	26.23	30.90	27.31	29.63	30.64	25.61	26.15	28.86	26.12	17.49	21.68	34.01	34.41
<b>Posto 9</b>	25.06	23.39	28.41	23.43	22.96	24.27	35.69	23.47	26.89	27.31	26.48	26.55	26.83	25.75	16.22	19.57	29.73	31.06
<b>Posto 10</b>	28.65	24.45	28.74	24.58	23.87	23.24	40.53	20.15	24.86	24.56	28.21	27.86	25.63	26.34	16.43	18.60	25.71	28.21
<b>Posto 11</b>	32.62	26.44	29.92	26.63	25.75	23.26	45.41	17.64	23.73	22.58	30.68	29.95	25.38	27.84	18.08	18.95	22.09	26.02
<b>Posto 12</b>	36.84	29.15	31.84	29.39	28.39	24.33	50.31	16.31	23.62	21.61	33.70	32.68	26.10	30.10	20.83	20.54	19.10	24.66
<b>Posto 13</b>	41.23	32.42	34.39	32.70	31.61	26.32	55.23	16.46	24.56	21.77	37.16	35.90	27.71	32.97	24.31	23.13	17.09	24.27
<b>Posto 14</b>	45.75	36.08	37.44	36.39	35.24	29.05	60.16	18.06	26.42	23.04	40.94	39.50	30.08	36.30	28.25	26.42	16.40	24.90
<b>Posto 15</b>	50.36	40.04	40.88	40.36	39.18	32.33	65.10	20.76	29.03	25.26	44.96	43.37	33.05	39.99	32.48	30.18	17.20	26.48
<b>Posto 16</b>	55.04	44.20	44.61	44.54	43.34	36.00	70.05	24.21	32.22	28.20	49.15	47.46	36.46	43.93	36.91	34.26	19.31	28.84
<b>Posto 17</b>	59.77	48.53	48.57	48.88	47.66	39.96	75.01	28.12	35.82	31.65	53.49	51.70	40.20	48.07	41.46	38.56	22.36	31.83
<b>Posto 18</b>	64.54	52.97	52.71	53.33	52.10	44.14	79.97	32.34	39.72	35.48	57.93	56.07	44.19	52.36	46.11	43.01	26.02	35.27

Observa-se na Tabela 5 que o posto mais próximo de V1 é o posto 6, analogamente, o posto mais próximo de V18 é o posto 13. As menores distâncias entre as variáveis e os postos encontram-se destacadas na Tabela.

Além disso, observa-se a formação de agrupamentos em torno de alguns postos. Em torno do posto 8 estão aglomeradas V2, V4, V11 e V12; em torno do posto 9 estão agrupadas as variáveis V3, V5, V14 e V15. Nesses casos deve-se considerar que essas variáveis têm a mesma importância, não sendo possível distinguir a mais importante entre elas sem que se conduza uma nova pesquisa centrada apenas nessas variáveis.

A Tabela 6 sintetiza a ordenação de todas as variáveis, na qual se observa que V7 é a variável mais importante e que V17 é a menos importante.

**Tabela 6.** Agrupamentos de variáveis em torno dos postos de importância

<b>POSTO</b>	<b>VARIÁVEIS</b>
<b>Posto 1</b>	
<b>Posto 2</b>	V7
<b>Posto 3</b>	
<b>Posto 4</b>	
<b>Posto 5</b>	
<b>Posto 6</b>	V1
<b>Posto 7</b>	
<b>Posto 8</b>	V2, V4, V11, V12
<b>Posto 9</b>	V3, V5, V14, V15
<b>Posto 10</b>	V6, V16
<b>Posto 11</b>	V13
<b>Posto 12</b>	V8, V9, V10
<b>Posto 13</b>	V18
<b>Posto 14</b>	V17
<b>Posto 15</b>	
<b>Posto 16</b>	
<b>Posto 17</b>	
<b>Posto 18</b>	

Além da importância relativa entre as variáveis, também é possível atribuir a elas uma importância absoluta. Observa-se que V7 é uma variável cuja importância é muito alta, porque ela está muito próxima ao posto 1, que é o posto ocupado pela variável que teoricamente seria a mais relevante para todos os entrevistados. Também pode-se notar que a variável menos importante, V17, ocupa o posto 14, ou seja, está a quatro postos do posto de menor importância, portanto sua importância não pode ser negligenciada.

#### **4. CONCLUSÃO**

Foi proposto um método para ordenação de variáveis, ou fatores, por importância, segundo a percepção de um conjunto de entrevistados. A aplicação deste método parte da criação de uma estrutura de dados denominada matriz de importância, a qual pode ser utilizada como entrada para diversos métodos multivariados, tais como: análise de agrupamentos, análise fatorial e análise de componentes principais.

Esta ferramenta mostra-se particularmente útil nos casos em que se deseja priorizar ações em função da relevância das variáveis, ou fatores, levantados na fase exploratória de uma pesquisa, na medida em que permite a classificação desses elementos do ponto de vista relativo, ao mesmo tempo em que fornece uma estimativa de sua importância absoluta.

Por conseguinte, esse método é uma alternativa, ou um complemento, para as técnicas clássicas de análise de dados multivariados, nos casos em que a aplicação das tais ferramentas não se mostre adequada.

#### **5. AGRADECIMENTOS**

Os autores agradecem o suporte financeiro da FAPESP, por intermédio do processo nº 2014/18628-9, Fundação de Amparo à Pesquisa do Estado de São Paulo.

## REFERÊNCIAS

- Brock, G.; Pihur, V.; Datta, S. & Datta, S. (2008). clValid: An R package for Cluster Validation. *Journal of Statistical Software*, 25(4), 1-22.
- Chicco, G.; Napoli, R. & Piglion, F. (2006). Comparisons Among Clustering Techniques for Electricity Customer Classification, *IEEE Transactions On Power Systems*, 21(2), 933-940.
- Frei, F. (2006). *Introdução à análise de agrupamentos: teoria e prática*, São Paulo: Editora UNESP.
- Hartigan, J.A. & Wong, M.A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Applied Statistics*, 28(1), 100-108.
- Kaufman, L. & Rousseeuw, P.J. (1986). Clustering large datasets. In: Gelsema, E.S.; Kanal, L.N. (Eds.). *Pattern Recognition in Practice II*. Amsterdam: Elsevier/North-Holland 425-437.
- Lattin, J.; Carrol, J.D. & Green, P.E. (2011). *Análise de Dados Multivariados*. São Paulo: Cengage Learning.
- Miguel, P.A.C. (2010). *Metodologia de Pesquisa em Engenharia de Produção e Gestão de Operações*. Rio de Janeiro: Elsevier.
- Oliveira, T. M. V. (2001). Escalas de Mensuração de Atitudes: Thurstone, Osgood, Stapel, Likert, Guttman, Alpert. *Administração On Line: Prática – Pesquisa – Ensino*, 2(2).
- Pessanha, J.F.M. & Laurencel, L.C. (2009). Clustering Electric Load Curves: The Brazilian Experience, In: *Workshop Franco-Brasílien sur la Fouille des Données*, Recife.
- Pessanha, J.F.M.; Xavier, V.L.; Amaral, M.R. S. & Laurencel, L. C. (2015). Construindo Tipologias de Curvas de Carga com o Programa R. *Podes - Revista Eletrônica Pesquisa Operacional para o Desenvolvimento*, 7(1), 29-54.
- Rambo, A. (2011). *O Impacto da Certificação na Qualidade do Setor Aeroespacial Brasileiro na Perspectiva das Organizações Certificadas*. Dissertação de Mestrado. Universidade de Taubaté. Programa de Mestrado Profissional em Engenharia Mecânica: Área de concentração Produção Mecânica. Taubaté, São Paulo, Brasil.