

## PROCEDIMENTO PARA FORMAÇÃO DE GRUPOS DE EMPRESAS E PARA CONSTRUÇÃO DE ÍNDICE DE AVALIAÇÃO DOS AGRUPAMENTOS

**Enio Júnior Seidel<sup>a</sup>**

**Marcelo Silva de Oliveira<sup>a</sup>**

**Luiz Marcelo Antonialli<sup>a</sup>**

<sup>a</sup> Universidade Federal de Lavras (UFLA), Lavras, MG, Brasil

**Bruno Tavares<sup>b</sup>**

<sup>b</sup> Universidade Federal de Viçosa (UFV), Viçosa, MG, Brasil

---

### Resumo

*A investigação de padrões de semelhança entre empresas pode ser realizada através da construção de grupos por similaridade. Contudo, geralmente se tem dúvidas quanto à melhor forma de construir os agrupamentos. Desse modo, tem-se o objetivo de estudar alguns métodos de agrupamentos e alguns possíveis tamanhos de grupos, de maneira a se obter um bom agrupamento. Além disso, se propõe um índice de avaliação dos agrupamentos formados. Consideraram-se seis métodos de agrupamentos e três tamanhos de grupos. Após a aplicação da análise de agrupamentos, foi realizada a avaliação dos agrupamentos formados através da análise discriminante, do coeficiente de concordância Kappa e de um índice desenvolvido para avaliar os agrupamentos iniciais. De forma geral, verificou-se que a melhor possibilidade seria aplicar o método MW com três grupos e, após, aplicar a análise discriminante até obter um número final adequado de empresas por grupo. Com este procedimento, se obtêm grupos relativamente semelhantes em termos de número de elementos por grupo. Além disso, uma alternativa interessante é tomar uso do índice de avaliação dos agrupamentos iniciais para a escolha do método de agrupamento a ser utilizado, permitindo fazer a opção pelo melhor procedimento.*

**Palavras-Chave:** Análise de Agrupamentos, Análise Discriminante, Tamanhos de Grupos, Índice de Avaliação dos Agrupamentos.

## 1. INTRODUÇÃO

A investigação de padrões de semelhança entre empresas constitui-se numa poderosa maneira para a construção do conhecimento na gestão bem sucedida. Em pesquisas na área de Administração, cujo objeto de estudo dificulta a identificação de relações de determinação entre as variáveis, a análise multivariada permite a compreensão da complexidade dos sistemas sociais de modo mais amplo na medida em que a relação de causa e efeito provocada por agentes externos pode ser substituída pelas relações internas dos próprios sistemas sociais. Nestes casos, as técnicas multivariadas, como a análise de agrupamentos, permitem investigar os padrões de comportamentos das empresas em relação a determinados aspectos de interesse. Nessas situações, a investigação de padrões pode ser realizada com a construção de grupos de empresas por similaridade, para que se possa ter um maior entendimento do comportamento das empresas em termos de semelhanças e diferenças em relação aos aspectos em estudo. A análise de agrupamentos tem em sua essência essa construção de grupos por similaridade, fazendo emergir, no modelo estatístico inerente, a explicitação de relações entre variáveis que são o próprio conhecimento revelador e, quase sempre, inovador. Contudo, geralmente se tem dúvidas quanto à melhor forma de construir os agrupamentos.

Rohrich e Cunha (2004), em seu estudo, utilizaram a análise de *clusters* para formar grupos de empresas e, após, para verificar se existia diferença entre os grupos encontrados, aplicaram o teste de Kruskal-Wallis. Já no estudo de Stallivieri, Campos e Britto (2009), foi utilizado o método *k-means* para agrupar empresas e, num segundo momento, numa tentativa de confirmar os resultados obtidos, utilizaram o método de agrupamentos da ligação completa.

Joia e Malheiros (2010) utilizaram a análise de *clusters* em estudo sobre influência de alianças estratégicas sobre a geração de ativos intangíveis de firmas participantes dessas alianças. Para decidir qual número de grupos seria o adequado, os autores tomaram por base o aumento da distância entre os grupos em cada estágio de construção dos agrupamentos.

Outro exemplo pode ser o estudo de Toledo e Simões (2010), com o objetivo de caracterizar e analisar os perfis e a maturidade da gestão do Processo de Desenvolvimento de Produto (PDP) em empresas de pequeno e médio porte de capital nacional do segmento de máquinas e implementos agrícolas do Estado de São Paulo, onde foi utilizada a análise de *clusters*, através do método de Ward e do método *k-means*.

Verifica-se, pelo exposto acima, que existem muitas tentativas de realizar agrupamentos que, em geral, utilizam procedimentos distintos. Ou seja, não se tem unanimidade em um procedimento único de agrupamentos.

Desse modo, uma questão vem à tona: Qual a melhor forma de construir agrupamentos de empresas?

Com vistas a sanar esta questão é que se propõe este estudo, em que se tem o objetivo de estudar alguns métodos de agrupamentos e alguns possíveis tamanhos de grupos, de maneira a se obter um bom agrupamento com aplicação em um conjunto de dados referente a uma pesquisa realizada com empresas do ramo de confecções. Além disso, se propõe um índice de avaliação dos agrupamentos formados.

A análise de *clusters*, ou análise de agrupamentos (MINGOTI, 2005; FERREIRA, 2008) tem caráter exploratório, em que se busca uma visão do padrão dos dados. Assim, espera-se que este estudo contribua com as investigações sobre a aplicabilidade da análise de agrupamentos no sentido de que se apresentam alguns procedimentos para investigar quais as melhores formas de construir agrupamentos.

## 2. METODOLOGIA

Para a realização deste estudo, foram considerados dados referentes à pesquisa de Tavares (2011). As variáveis *competência gerencial*, *competência inovativa* e *competência humana* foram elaboradas a partir da média dos escores de outras variáveis estudadas por Tavares (2011).

As variáveis foram medidas em 127 empresas do ramo de confecções das cidades de Divinópolis-MG e Muriaé-MG. Porém, algumas empresas apresentaram dados faltantes e foram excluídas da análise para possibilitar a aplicação da análise de agrupamentos. Desse modo, passou-se a compor o conjunto de dados com 117 empresas sem apresentarem dados faltantes.

As primeiras investigações realizadas foram relativas a uma descrição inicial dos dados por meio de medidas descritivas.

Como a análise de agrupamentos é uma análise multivariada exploratória, optou-se por avaliar vários métodos de agrupamentos e verificar a melhor adequação aos dados em estudo.

Para a aplicação da análise de agrupamentos seguiram-se os seguintes procedimentos:

- 1° - Utilizou-se a distância euclidiana para medir a dissimilaridade entre as empresas;
- 2° - Optou-se por utilizar os seguintes métodos de agrupamento hierárquicos: Método da Ligação Simples (MLS), Método da Ligação Completa (MLC), Método da Ligação Média (MM), Método da Mediana (MMd), Método Centroide (MC) e Método de Ward (MW);
- 3° - Considerou-se a possibilidade de formar de 3 a 5 grupos em cada método de agrupamento.

Assim, têm-se seis métodos de agrupamentos e três tamanhos de grupos para serem avaliados em conjunto, isto é, para cada método de agrupamento utilizado, serão considerados três tamanhos de grupos.

Após a aplicação da análise de agrupamentos, foi realizada a avaliação dos agrupamentos formados e sua coerência. Para avaliar a qualidade dos agrupamentos obtidos, foram utilizadas a análise discriminante e o coeficiente de concordância Kappa. A ideia é de que seja possível verificar a existência, ou não, de concordância entre os agrupamentos formados pela análise de agrupamentos e as classificações realizadas com a análise discriminante. Além disso, foi construído um índice para avaliar os agrupamentos iniciais com base no número de aplicações necessárias da análise discriminante.

A aplicação da metodologia foi realizada através da utilização do *software* R (R DEVELOPMENT CORE TEAM, 2007).

## 2.1. Procedimentos para construção dos agrupamentos

A análise de agrupamentos é utilizada para classificar elementos em grupos relativamente homogêneos, chamados de *clusters* (MALHOTRA, 2006). A ideia é de que elementos semelhantes pertençam ao mesmo grupo e elementos distintos pertençam a grupos diferentes. Desse modo, a composição dos grupos terá característica de homogeneidade dentro dos grupos e heterogeneidade entre os grupos.

Para atribuir uma medida de dissimilaridade entre os elementos, utiliza-se a distância euclidiana entre os elementos. A distância euclidiana entre dois elementos  $X_l$  e  $X_k$ , com  $l \neq k$ , pode ser definida por (MINGOTI, 2005):

$$d(X_l, X_k) = \left[ \sum_{i=1}^p (X_{il} - X_{ik})^2 \right]^{\frac{1}{2}} \quad (1)$$

Porém, o *software* R tem implementada a distância euclidiana como definida pelo quadrado da distância (1), isto é, para o *software* R distância euclidiana é o quadrado de (1).

Neste trabalho, utilizam-se métodos hierárquicos de agrupamentos que permitem a construção de um dendrograma que mostra as ligações entre os elementos. Mais especificamente, serão utilizados métodos hierárquicos aglomerativos nos quais, inicialmente, considera-se que cada elemento é um grupo e, por fim, ao concluir o procedimento, ter-se-á apenas um grupo composto por todos os elementos. Objetos mais similares

serão agrupados primeiro e os agrupamentos irão surgindo de acordo com a similaridade dos elementos (JOHNSON; WICHERN, 1992).

Os métodos aglomerativos utilizados são o Método de Ligação Simples (MLS), o Método de Ligação Completa (MLC), o Método da Ligação Média (MM), o Método da Mediana (MMd), o Método Centróide (MC) e o Método Ward (MW). Para mais detalhes destes métodos, ver Mingoti (2005) e Ferreira (2008).

Estes métodos foram usados por serem mais populares e pela facilidade de serem encontrados em vários softwares, conforme também destacam Albuquerque *et al.* (2006).

## 2.2. Procedimentos para avaliação dos agrupamentos

Para realizar a avaliação dos agrupamentos formados, foram utilizadas a análise discriminante, o cálculo do coeficiente de concordância Kappa e o índice de avaliação dos agrupamentos. Em seu trabalho sobre pressões cambiais, Moreira, Pinto e Souza (2004), utilizaram a análise de agrupamentos e, como forma de validar os agrupamentos formados, tomaram uso da análise discriminante.

A análise discriminante envolve a determinação de uma variável discriminante como uma combinação linear de duas ou mais variáveis independentes que melhor discriminam entre os elementos nos grupos definidos *a priori* (HAIR JR. *et al.*, 2009). A ideia foi baseada em Santos *et al.* (2004), que também utilizaram análise de agrupamentos e análise discriminante em sua pesquisa. A primeira análise serviu para construir grupos e a segunda foi usada para verificar a pertinência dos grupos. Para mais detalhes sobre a análise discriminante, ver Mingoti (2005) e Ferreira (2008).

O coeficiente Kappa foi usado para verificar a concordância entre os grupos formados pela análise de agrupamentos e as classificações realizadas com a análise discriminante. Para mais detalhes sobre o coeficiente Kappa, ver Siegel e Castellan Jr. (2006).

Por fim, foi construído um índice de avaliação dos agrupamentos. Este índice é calculado da seguinte forma:

$$i = 1 / nc \quad (2)$$

em que  $nc$  é o número de aplicações da análise discriminante, após o agrupamento inicial, para que se obtenha uma tabela do tipo:

Penúltima aplicação da classificação	Última aplicação da classificação			
	Grupo 1	Grupo 2	...	Grupo k
Grupo 1	n1	0	0	0
Grupo 2	0	n2	0	0
...	0	0	...	0
Grupo k	0	0	0	nk

em que, na diagonal, têm-se n1 elementos no Grupo 1, n2 elementos no Grupo 2, ... e nk elementos no Grupo k. Fora da diagonal, têm-se somente zeros.

O índice  $i$  varia no intervalo  $(0,1]$ . Quanto mais próximo de 1, melhor é o agrupamento inicial. Porém, quanto mais próximo de zero for o índice, pior é o desempenho do agrupamento inicial.

Na prática, isso significa que, quanto mais aplicações da análise discriminante sejam necessárias até se obter uma tabela do tipo acima, pior foi o agrupamento inicial. Espera-se que, para bons agrupamentos iniciais, poucas aplicações da análise discriminante sejam necessárias. Dessa forma, o índice pode ser utilizado para avaliar os agrupamentos iniciais pelos métodos da análise de agrupamentos.

### 3. RESULTADOS E DISCUSSÃO

Primeiramente, realizou-se uma análise descritiva dos dados em estudo para avaliar o comportamento das variáveis. Tal análise está apresentada na Tabela 1.

Tabela 1 – Análise descritiva das variáveis em estudo.

Variável	Mínimo	Máximo	Média	Desvio-padrão
Competência Gerencial	1,1670	3,8330	2,8080	0,4945
Competência Inovativa	1,3750	3,8750	2,7090	0,4921
Competência Humana	1,0000	4,0000	2,5640	0,5725

Observando a Tabela 1, verifica-se que as variáveis apresentam comportamentos semelhantes em termos de média e desvio-padrão.

Após a primeira investigação, o segundo passo foi a aplicação da análise de agrupamentos (lembrando que foi utilizada a distância euclidiana como medida de dissimilaridade entre as empresas para a construção dos agrupamentos, conforme implementada no *software R*).

A Tabela 2 apresenta as formações dos grupos de empresas para 3 grupos, 4 grupos e 5 grupos, pelos métodos de agrupamentos considerados.

Tabela 2 – Número de empresas por grupo para cada método de agrupamento, para 3 grupos, 4 grupos e 5 grupos.

Método	3 Grupos			4 Grupos				5 Grupos				
	G. 1	G. 2	G. 3	G. 1	G. 2	G. 3	G. 4	G. 1	G. 2	G. 3	G. 4	G. 5
MLS	115	1	1	114	1	1	1	111	1	1	3	1
MLC	58	40	19	54	40	19	4	54	40	16	4	3
MM	95	19	3	95	18	3	1	95	15	3	1	3
MMd	115	1	1	113	1	2	1	113	1	1	1	1
MC	115	1	1	114	1	1	1	110	1	1	4	1
MW	41	26	50	41	26	27	23	18	26	27	23	23

Percebe-se que os métodos MLS, MM, MMd e MC formaram grupos um tanto discrepantes, em que o grupo 1 teve muitas empresas e os demais grupos tiveram poucas empresas. Já os métodos MLC e MW, foram, até certo ponto, regulares em termos de número de empresas por grupo. Para Ferreira (2008), as técnicas de agrupamentos podem revelar padrões existentes no conjunto de dados que seriam difíceis de serem detectadas com inspeções simples visuais e de suas estatísticas descritivas.

A validação dos grupos formados poderia ser feita utilizando métodos estatísticos univariados ou multivariados. Segundo Valgas *et al.* (2010), a validação dos agrupamentos poderia ser feita utilizando-se desde testes univariados, para comparação de médias, até testes multivariados. Em sua pesquisa, Valgas *et al.* (2010) utilizaram a análise de variância multivariada para realizar a validação dos agrupamentos. Considerando métodos multivariados, poderia ser usada, além da análise de variância multivariada, a análise discriminante (FERREIRA, 2008).

Segundo Ferreira (2008), a ideia da validação é que se deve analisar a consistência do agrupamento obtido *a posteriori*. Para Valgas *et al.* (2010), validar os agrupamentos significa garantir que realmente os grupos diferem entre si.

Neste trabalho, foi utilizada a análise discriminante para validar os agrupamentos. Para cada um dos agrupamentos realizados, foi aplicada a análise discriminante para verificar a existência ou não de concordância

entre o agrupamento e a classificação. Como os dados não apresentaram normalidade multivariada e as matrizes de covariâncias dos grupos não apresentaram homocedasticidade, foi utilizada a função discriminante quadrática (KHOURY JR *et al.*, 2005).

Para os métodos MLS, MM, MMd e MC, não foi possível aplicar a validação, pois a função discriminante quadrática não se aplicou com grupos com menos de 4 elementos. Assim, nem a análise discriminante, nem o coeficiente Kappa foram aplicados para estes métodos.

A seguir, tem-se a Tabela 3 com a concordância entre os agrupamentos com o método MLC e as classificações para 3 grupos. O agrupamento pelo método MLC e a classificação com três grupos apresentaram um percentual de 97,44% (114 empresas de um total de 117) de empresas alocadas nos mesmos grupos, com coeficiente de concordância Kappa de 0,9582. Pelo método MLC para quatro grupos, 96,58% (113 empresas de um total de 117) foram agrupadas nos mesmos grupos, com coeficiente de concordância Kappa de 0,9473.

Tabela 3 – Concordância para 3 grupos para o método MLC.

Agrupamento	Classificação			Total
	1	2	3	
1	56	1	1	58
2	1	39	0	40
3	0	0	19	19
Total	57	40	20	117

A Tabela 4 apresenta a concordância para 3 grupos para o método MW. O agrupamento pelo método MW e a classificação com três grupos apresentaram 94,02% (110 das 117 empresas) de concordância, coeficiente de concordância Kappa de 0,9071. Com quatro grupos, 93,16% (109 das 117 empresas) das empresas foram alocadas da mesma forma, e o coeficiente de concordância Kappa foi de 0,9074. E com cinco grupos, ocorreu 92,31% (108 das 117 empresas) de concordância nos agrupamentos, com coeficiente de concordância Kappa de 0,9035.

Tabela 4 – Concordância para 3 grupos para o método MW.

Agrupamento	Classificação			Total
	1	2	3	
1	38	1	2	41
2	2	24	0	26
3	2	0	48	50
Total	42	25	50	117

Realizando uma análise geral sobre a utilização das técnicas de agrupamentos e análise discriminante, verifica-se que as alocações coincidentes pelos agrupamentos e pelas classificações variaram de 92,31% a 97,44%, que são percentuais relativamente altos. Assim, verifica-se que a concordância entre os agrupamentos e as classificações foi boa.

Além disso, considerando o coeficiente de concordância Kappa, verifica-se o mesmo comportamento, com coeficientes variando entre 0,9035 e 0,9582, que são consideradas altas concordâncias.

Desse modo, a melhor alternativa seria utilizar o resultado do método MLC ou do método MW, com 3 grupos. O método MW apresentou tamanhos de grupos relativamente mais semelhantes do que o método MLC.

Uma possível justificativa para o fraco desempenho dos métodos MLS, MM, MC e MMd pode ser o fato de que as variáveis utilizadas na construção dos grupos têm pouca variabilidade e possuem pequena amplitude.

Outra possibilidade para obter melhores resultados é reaplicar a análise discriminante até que não se tenham empresas alocadas em grupos distintos pelas duas discriminações. Esta ideia de aplicar, primeiramente,

a análise de agrupamentos e, depois, aplicar a análise discriminante, tem objetivo de realocar os elementos, que estão classificados incorretamente, aos grupos cuja probabilidade *a posteriori* de classificação for maior (FERREIRA, 2008).

Considerando inicialmente o agrupamento com o método MLC com 3 grupos, após 4 aplicações da análise discriminante se tem os seguintes grupos formados pela análise discriminante: 57 empresas no grupo 1, 38 empresas no grupo 2 e 22 empresas no grupo 3. A Tabela 5 apresenta as duas últimas reaplicações.

Tabela 5 – Concordância entre as penúltima e última aplicações da análise discriminante com base no agrupamento inicial pelo método MLC com 3 grupos.

Penúltima aplicação da classificação	Última aplicação da classificação			Total
	1	2	3	
1	57	0	0	57
2	0	38	0	38
3	0	0	22	22
Total	57	38	22	117

Como os grupos foram alocados adequadamente pela análise discriminante após a quarta alocação, pode-se calcular o índice de avaliação de desempenho dos agrupamentos. Pelo índice expresso por (2), verifica-se que o método MLC com 3 grupos teve valor  $i = 1/4 = 0,25$

Para o método MW com 3 grupos, obtém-se, para o índice, valor  $i = 1/3 = 0,33$ , ou seja, foram necessárias 3 aplicações para classificar as empresas adequadamente. A Tabela 6 apresenta as duas últimas reaplicações.

Tabela 6 – Concordância entre as penúltima e última aplicações da análise discriminante com base no agrupamento inicial pelo método MLC com 3 grupos.

Penúltima reaplicação da classificação	Última reaplicação da classificação			Total
	1	2	3	
1	39	0	0	39
2	0	25	0	25
3	0	0	53	53
Total	39	25	53	117

Desse modo, considerando o índice  $i$  como critério para escolha do agrupamento, optar-se-ia pelo método MW com 3 grupos. Esta escolha se baseia na comparação dos índices de avaliação, onde o índice para o método MW foi de 0,33 e para o método MLC foi de 0,25. Quanto mais próximo de 1, melhor.

A configuração dos grupos pelo método MW seria de 41 empresas no grupo 1, 26 no grupo 2, e 50 empresas no grupo 3. Porém, se fosse considerada a configuração de grupos após as 3 aplicações da análise discriminante, ter-se-iam 39 empresas no grupo 1, 25 empresas no grupo 2 e 53 empresas no grupo 3.

#### 4. CONCLUSÃO

Este artigo teve por objetivo avaliar diferentes métodos estatísticos de agrupamentos e configurações de grupo, visando um maior entendimento do comportamento das empresas em termos de semelhanças e diferenças, fazendo emergir, no modelo estatístico inerente, a explicitação de relações entre variáveis. Para tanto, foram avaliadas várias configurações de grupos, tomando uso de alguns métodos de agrupamentos e possíveis tamanhos de grupos. Porém, alguns resultados não foram muito satisfatórios. Apenas os métodos MLC e MW tiveram resultados regulares.

De forma geral, verificou-se que a melhor possibilidade seria aplicar o método MW com três grupos e, depois, aplicar a análise discriminante até obter um número final adequado de empresas por grupo. Com este procedimento, obtêm-se grupos relativamente semelhantes em termos de número de elementos por grupo.

Além disso, espera-se que a construção do índice de avaliação dos agrupamentos iniciais possa servir como um critério para a escolha do método de agrupamento a ser utilizado.

A proposta desta forma de avaliação vai ao encontro da necessidade gerada pela utilização cada vez mais intensa das técnicas multivariadas para finalidades distintas no âmbito da Administração. A amplitude de possibilidades tem abrangido desde conceitos como corrupção (BORINI; GRISI, 2009), alianças estratégicas (JOIA; MALHEIROS, 2010), desenvolvimento de produtos (TOLEDO; SIMÕES, 2010) e muitos outros. Diante da disseminação destas técnicas, este artigo contribui para atender, ainda que pontualmente, a necessidade de se aprofundar o conhecimento em torno da utilização de técnicas multivariadas para a compreensão do comportamento das empresas e para o campo da administração como um todo.

Por fim, uma observação importante a se fazer é a de que em cada situação se devem testar várias formas de agrupamento, pois não se pode dizer que um método ou um procedimento é o melhor sempre. Longe disso, em uma situação um método pode ser melhor, mas em outra situação ele pode não ser. Uma alternativa interessante, que é um método proposto neste artigo, é tomar uso do índice de avaliação dos agrupamentos iniciais para a escolha do método de agrupamento a ser utilizado, permitindo fazer a opção pelo melhor procedimento.

## 5. REFERÊNCIAS

- ALBUQUERQUE, M. A. et al. Estabilidade em análise de agrupamento: estudo de caso em ciência florestal. **R. Árvore**, Viçosa, v. 30, n. 2, pp. 257-265, 2006.
- BORINI, F. M.; GRISI, F. C. A corrupção no ambiente de negócios: survey com as micro e pequenas empresas da cidade de São Paulo. **RAUSP**, São Paulo, v. 44, n. 2, pp. 102-117, 2009.
- FERREIRA, D. F. **Estatística multivariada**. 1 ed. Lavras: Ed. UFLA, 2008. 661 p.
- HAIR JR., J. F. et al. **Análise multivariada de dados**. 6 ed. Porto Alegre: Bookman, 2009. 688 p.
- JOHNSON, R. A; WICHERN, D. W. **Applied multivariate statistical analysis**. 3 ed. Englewood Cliffs: Prentice-Hall, 1992. 642 p.
- JOIA, L. A.; MALHEIROS, R. Evidências empíricas da influência de alianças estratégicas no capital intelectual de empresas. **BASE – Revista de Administração e Contabilidade da Unisinos**, São Leopoldo, v. 7, n. 2, pp. 162-177, 2010.
- KHOURY JR., J. K. et al. Análise discriminante paramétrica para reconhecimento de defeitos em tábuas de eucalipto utilizando imagens digitais. **R. Árvore**, Viçosa, v. 29, n. 2, pp. 299-309, 2005.
- MALHOTRA, N. K. **Pesquisa de marketing: uma orientação aplicada**. 4 ed. Porto Alegre: Bookman, 2006. 720 p.
- MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Ed. UFMG, 2005. 295 p.
- MOREIRA, T. B. S.; PINTO, M. B. P.; SOUZA, G. S. Uma metodologia alternativa para mensuração de pressão sobre o mercado de câmbio. **Est. Econ.**, São Paulo, v. 34, n. 1, pp. 73-99, 2004.
- R Development Core Team (2007). **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing. Disponível em: <<http://www.r-project.org/>>. Version 2.6.1.
- ROHRICH, S. S.; CUNHA, J. C. A proposição de uma taxonomia para a análise da gestão ambiental no Brasil. **RAC**, v. 8, n. 4, pp. 81-97, 2004.

SANTOS, J. H. S. et al. Distinção de grupos ecológicos de espécies florestais por meio de técnicas multivariadas. **R. Árvore**, Viçosa, v. 28, n. 3, pp. 387-396, 2004.

SIEGEL, S.; CASTELLAN JR., N. J. **Estatística não-paramétrica para ciências do comportamento**. 2 ed. Porto Alegre: Artmed, 2006. 448 p.

STALLIVIERI, F.; CAMPOS, R. R.; BRITTO, J. N. P. Indicadores para a análise da dinâmica inovativa em arranjos produtivos locais: uma análise exploratória aplicada ao arranjo eletrometal-mecânico de Joinville/SC. **Estudos Econômicos**, v. 39, n. 1, pp. 185-219, 2009.

TAVARES, B. **A estrutura das aglomerações produtivas e capacitações das micro e pequenas empresas**: um estudo comparativo no setor de vestuário em Minas Gerais. 2011. Tese (Doutorado em Administração) – UFLA, Lavras, 2011.

TOLEDO, J. C.; SIMÕES, J. M. S. Gestão do desenvolvimento de produto em empresas de pequeno e médio porte do setor de máquinas e implementos agrícolas do Estado de SP. **Gestão da Produção**, São Carlos, v. 17, n. 2, pp. 257-269, 2010.

VALGAS, R. A. et al. Técnicas de agrupamento aplicadas no mapeamento da divergência genética de subpopulações de *Araucária Angustifolia* (Bert.) O. Ktze em Irati, PR e Caçador, SC por marcadores isoenzimáticos. **Ciência e Natura**, Santa Maria, v. 32, n. 2, pp. 35-50, 2010.

## PROCEDURE FOR THE FORMATION OF GROUPS OF COMPANIES AND FOR BUILDING THE INDEX FOR CLUSTER ASSESSMENT

### Abstract

*The investigation of similarity patterns between companies can be accomplished by the formation of similarity groups. However, there is always doubt about the best way to build clusters. Thus, the objective of this work was to study some clustering methods and some possible group sizes, so as to obtain a good clustering. In addition, an index for evaluating the clusters formed was proposed. Six clustering methods and three group sizes were considered. After the application of cluster analysis, the groups formed were assessed by the discriminant analysis, the Kappa coefficient and an index developed to assess the initial clusters. Overall, it was found that the best option would be to apply the MW method with three groups and after, to apply the discriminant analysis to obtain an appropriate final number of companies per group. This procedure allowed obtaining groups relatively similar in terms of number of elements per group. Furthermore, an interesting alternative is to make use of the index for assessing initial clusters to select the clustering method to be used, allowing selecting the best procedure.*

**Keywords:** Cluster Analysis. Discriminant Analysis. Group Sizes. Index for Evaluating Clusters.