

Avaliação da adequação de aplicação de técnicas multivariadas em estudos do comportamento do consumidor em teses e dissertações de duas instituições de ensino superior

Maria Aparecida Gouvêa
Leandro Campi Prearo
Maria do Carmo Romeiro

Recebido em 12/novembro/2010
Aprovado em 21/novembro/2011

Sistema de Avaliação: *Double Blind Review*
Editor Científico: Nicolau Reinhard

DOI: 10.5700/rausp1043

RESUMO

Este artigo faz parte de um amplo estudo de avaliação da adequação no uso de técnicas estatísticas multivariadas em teses e dissertações de duas instituições de ensino superior na área de *marketing* na temática do comportamento do consumidor, entre 1997 e 2006. Neste artigo são focalizadas onze técnicas multivariadas (análise de regressão, análise discriminante, análise de regressão logística, correlação canônica, análise multivariada de variância, análise conjunta, modelagem de equações estruturais, análise fatorial, análise de conglomerados, análise de correspondência, escalonamento multidimensional), as quais têm apresentado grande potencial de uso em estudos de *marketing*. Foi objetivo no trabalho relatado a análise da adequação do emprego dessas técnicas às necessidades dos problemas de pesquisa apresentados nas teses e dissertações e, também, a aferição do nível de adequação no atendimento de suas premissas. De forma geral, os resultados sugerem a necessidade de um aumento do comprometimento dos pesquisadores na verificação de todos os preceitos teóricos de aplicação das técnicas multivariadas.

Palavras-chave: análise multivariada, adequação do uso de técnicas estatísticas, comportamento do consumidor.

1. INTRODUÇÃO

Os aspectos metodológicos têm grande relevância no desenvolvimento de trabalhos científicos. Para Gamboa (1987, p.17), a importância da análise da

Maria Aparecida Gouvêa é Professora Associada do Departamento de Administração da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo (CEP 05508-010 – São Paulo/SP, Brasil).
E-mail: magouvea@usp.br
Endereço:
Universidade de São Paulo
FEA – Departamento de Administração
Avenida Professor Luciano Gualberto, 908 – FEA1
Cidade Universitária
Butantã
05508-010 – São Paulo – SP

Leandro Campi Prearo, Graduado em Matemática, Mestre e Doutorando em Métodos Quantitativos na Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, é Professor na Universidade Municipal de São Caetano do Sul (CEP 09550-051 – São Caetano do Sul/SP, Brasil).
E-mail: leandro.prearo@uscs.edu.br

Maria do Carmo Romeiro, Mestre e Doutora em Administração e Marketing pela Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, é Professora na Universidade Municipal de São Caetano do Sul (CEP 09550-051 – São Caetano do Sul/SP, Brasil).
E-mail: mromeiro@uscs.edu.br

produção científica, amparada nas competências metodológicas, justifica-se, visto que sua questão é de fundamental e decisiva importância para o desenvolvimento e os resultados da pesquisa acadêmica.

Diversas obras a partir da década de 1990 têm tratado da avaliação dos aspectos metodológicos por meio de pesquisas que focam a questão da qualidade da produção científica em Administração, destacando-se: administração pública (MACHADO-DA-SILVA, 1989); finanças (LEAL, OLIVEIRA e SOLURI, 2003); *marketing* (POWERS, SWAN e PATTON, 1998; VIEIRA, 1998; BOTELHO e MACERA, 2001); métodos quantitativos (BREI e LIBERALI, 2004); operações (ARKADER, 2003); organizações (MACHADO-DA-SILVA, 1990; BERTERO e KEINERT, 1994; VERGARA e CARVALHO, 1995; BERTERO e KEINERT, 1999; RODRIGUES FILHO, 2002); pesquisa em administração (MARTINS, 1994; TORRES, 2000; PERIN, 2002); recursos humanos (ROESCH, ANTUNES e SILVA, 1997; CALDAS, TONELLI e LACOMBE, 2002; TONELLI *et al.*, 2003); sistemas de informação (HOPPEN, 1998).

A preocupação dos autores dessas obras na avaliação da produção científica tem se dirigido para a discussão sobre aspectos epistemológicos e sobre aspectos metodológicos, mas no caso dessa segunda vertente o número de estudos disponíveis na literatura pesquisada é mais reduzido.

Na literatura internacional pesquisada encontram-se diversas abordagens que tratam da inadequação no entendimento de escalas básicas de medidas (GAITO, 1980; TOWNSEND e ASHBY, 1984; VELLEMAN e WILKINSON, 1993) e inadequação do uso de testes paramétricos (FRICK, 1998; YU *et al.*, 2000). No entanto, estudos que avaliam especificamente a adequação do uso dos métodos quantitativos são praticamente inexistentes.

Os métodos quantitativos foram largamente empregados nas dissertações e teses apresentadas na Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo (FEA-USP), na Faculdade de Economia e Administração da Pontifícia Universidade Católica (FEA-PUC) e na Escola de Administração de Empresas de São Paulo da Fundação Getúlio Vargas (EAESP-FGV), entre os anos de 1980 e 1993, segundo Martins (1994, p.65-66), que ressaltou na avaliação dessas obras o uso superficial das técnicas, sem consideração de seu nível de complexidade e sofisticação. Tal superficialidade também é ressaltada pelo autor na aparente despreocupação com as premissas teóricas de aplicação da maior parte dos métodos apresentados nesses trabalhos.

No tratamento quantitativo dos dados para a produção científica, tem sido crescente o emprego de técnicas de análise multivariada pelo fato de que, nas últimas décadas, vários pacotes computacionais estatísticos foram aperfeiçoados sobremaneira no sentido de facilitar o acesso do usuário a seus comandos na área de trabalho, sem a necessidade de serem escritas sintaxes para o processamento das técnicas estatísticas.

Essa evolução tecnológica notória nos *software* estatísticos proporcionou a diminuição de barreiras para os usuários, atendendo adequadamente à demanda dos cientistas das ciências sociais aplicadas, entre elas a Administração. Entretanto, tem-se constatado, a partir da experiência de orientação e docência dos autores deste estudo, a ocorrência de erros de aplicação, seja na inadequação dos objetivos de uso das ferramentas com os objetivos propostos na pesquisa, seja na violação de premissas de aplicação das técnicas. Há situações em que o pesquisador apenas exercita o emprego de uma técnica e distancia-se de seu problema de pesquisa e do alcance dos objetivos inicialmente traçados.

Martins (1994, p.66) afirma que os métodos quantitativos são, em Administração, mais aplicados pelos autores de pesquisa nas áreas de *marketing*, produção e finanças.

A área de *marketing* apresenta campo fértil para utilização de técnicas quantitativas, especialmente a análise multivariada, principalmente pela necessidade de conhecer-se o mercado consumidor a partir da mensuração das opiniões, atitudes, preferências, perfil e outras características dos consumidores (MALHOTRA, 2001).

Nesse sentido, as ferramentas de análise multivariada têm sido intensamente aplicadas em *marketing*, especialmente na temática do comportamento do consumidor. Para Milagre (2001, p.74), o uso da técnica multivariada tornou-se mais comum a partir do momento que os acadêmicos e profissionais de *marketing* passaram a aplicá-la em estudos sobre a preferência e satisfação do consumidor, bem como seu perfil e comportamento de compras.

Assim, neste estudo pretende-se aprofundar a discussão sobre aspectos metodológicos no uso de técnicas estatísticas multivariadas, especialmente sob o recorte de suas aplicações nos estudos em *marketing* sobre o comportamento do consumidor. Para tanto, serão focalizadas as técnicas: análise de regressão, análise discriminante, análise de regressão logística, correlação canônica, análise multivariada de variância, análise conjunta, modelagem de equações estruturais, análise fatorial, análise de conglomerados, análise de correspondência e escalonamento multidimensional.

Nesse contexto, os objetivos a serem alcançados neste estudo são:

- identificar a intensidade de uso das técnicas multivariadas na área de *marketing*, temática do comportamento do consumidor, tendo como unidade de análise dissertações e teses de duas universidades públicas no período de 1997 a 2006: Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo e Escola de Administração da Universidade Federal do Rio Grande do Sul (EA-UFRGS).
- identificar o nível de adequação do uso das técnicas multivariadas aos problemas de pesquisa apresentados nessas dissertações e teses selecionadas;
- identificar fontes de erros da aplicação das técnicas multivariadas, a partir do não atendimento a suas premissas básicas, no conjunto da produção científica examinada.

A seleção desses programas de pós-graduação como público-alvo deste estudo deveu-se à necessidade de delimitação, dada a dificuldade operacional de um levantamento amostral representativo da produção nacional, principalmente pela indisponibilidade de material em base de dados *on-line* por uma parcela importante das instituições do País.

Nesse sentido, optou-se inicialmente por um recorte focado nos programas de pós-graduação de alto desempenho na última avaliação da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), que estava disponível quando este estudo teve início. Registre-se que a opção por esse critério não se motivou pela suposição de que os programas com avaliação menos positiva, abaixo da nota seis, tratariam as técnicas em estudo de forma mais ou menos corretas. A avaliação da Capes apontava, na Avaliação Trienal 2007, três Instituições com nota seis: a FEA-USP, a EA-UFRGS e a EAESP-FGV. Observação: essas instituições permanecem bem avaliadas mesmo com sua atual pontuação pela Capes, referente ao período de avaliação de 2007 a 2009 (Avaliação Trienal 2010), a saber: FEA-USP, nota sete; EA-UFRGS, nota seis; EAESP-FGV, nota seis.

Um segundo critério adveio da necessidade de garantir a maior homogeneidade possível da população-alvo, já que não é objetivo específico neste estudo a comparação entre instituições. Nesse sentido, a seleção levou em conta a natureza da instituição: pública ou privada. Considerando que, entre as três citadas, duas são públicas, optou-se pela análise das dissertações e teses das instituições públicas: FEA-USP e EA-UFRGS.

Destacam-se alguns fatores restritivos na abrangência dessa investigação:

- delimitação do público-alvo: os resultados deste estudo são apenas válidos para as dissertações e teses da FEA-USP e da EA-UFRGS, de 1997 a 2006;
- as informações sobre o atendimento às premissas das técnicas multivariadas são baseadas nos relatos dos autores de cada dissertação ou tese sob análise, que evidentemente podem conter uma descrição subestimada ou superestimada do que foi realizado.

2. FUNDAMENTAÇÃO TEÓRICA

Nesta seção haverá uma introdução sobre análise multivariada, seguida por uma apresentação geral das técnicas multivariadas e suas premissas.

2.1. Análise multivariada

Segundo Steiner (1995), a necessidade de entender a relação entre diversas variáveis aleatórias faz da análise multivariada uma metodologia com grande potencial de uso.

Para Lourenço e Matias (2001), por um lado as técnicas estatísticas multivariadas são mais complexas do que aquelas da estatística univariada. Por outro lado, apesar de uma

razoável complexidade teórica fundamentada na matemática, as técnicas multivariadas, por permitirem o tratamento de diversas variáveis ao mesmo tempo, podem oferecer ao pesquisador um material bastante robusto para a análise dos dados da pesquisa.

Conforme Hair Jr. *et al.* (2009), a análise multivariada auxilia na formulação de questões relativamente complexas de forma específica e precisa, possibilitando a condução de pesquisas teoricamente significativas.

2.2. Técnicas estatísticas de análise multivariada

A escolha dos métodos e tipos de análises empregados nos trabalhos científicos deve ser determinada pelo problema de pesquisa. Nesse sentido, Johnson e Wichern (1998, p.2) propõem uma classificação dos objetivos para atendimento do problema em cinco categorias, como poder ser visto no quadro 1.

O passo seguinte para a escolha da técnica de análise multivariada, após se considerarem os objetivos do problema de pesquisa, é verificar o tipo de relação examinada, número de variáveis dependentes e tipo de escala utilizada.

Sobre o tipo de relação, as técnicas são classificadas como de dependência ou de interdependência. Na primeira situação, uma ou mais variáveis (variáveis dependentes) podem ser explicadas ou preditas por outras (variáveis independentes). Na segunda, todas as variáveis são analisadas simultaneamente, sem a orientação de dependência ou independência.

Sobre o tipo de escala utilizada, pode-se generalizar a classificação teórica de mensuração em dois grandes grupos: variáveis métricas e variáveis não métricas.

Do cruzamento entre o tipo de relação examinada, número de variáveis dependentes e tipo de escala utilizada, tem-se um esquema de classificação para o conjunto de técnicas, exibido nos quadros 2 e 3.

2.3. Premissas das técnicas multivariadas

No quadro 4, apresenta-se a relação de premissas teóricas na aplicação das técnicas multivariadas, bem como a indicação, por meio de “X”, de cada técnica que apresenta associação com determinada condição. Nesse quadro, não foi incluída a análise conjunta pelo fato de nenhuma das premissas teóricas se aplicar a essa técnica.

Todas as técnicas estatísticas de análise multivariada têm suposições intrínsecas, matemáticas e estatísticas que, caso não satisfeitas, podem influenciar de forma bastante negativa os resultados da análise.

Nesse ponto, deve-se registrar que essas premissas ou suposições são inerentemente integrantes ao processo de construção teórica da técnica e, sem seu atendimento, não se pode garantir que o algoritmo inerente à técnica realmente terá o comportamento que se espera.

Quadro 1

Categorias dos Objetivos das Técnicas Estatísticas de Análise Multivariada

Classificação	Técnicas Relacionadas
<p>Investigação da dependência entre as variáveis Todas as variáveis são mutuamente independentes ou uma ou mais variáveis são dependentes de outras.</p>	Análise Discriminante Análise de Regressão Correlação Canônica Regressão Logística Análise Conjunta Análise da Variância Multivariada (MANOVA)
<p>Predição As relações entre as variáveis devem ser determinadas com o objetivo de prever o valor de uma ou mais variáveis com base nas observações de outras variáveis.</p>	Análise Discriminante Análise de Regressão Análise de Regressão Logística
<p>Construção de hipóteses e testes Hipóteses estatísticas específicas, formuladas em termos de parâmetros da população multivariada, são testadas. Isso pode ser feito para validar premissas ou para reforçar convicções prévias.</p>	Modelagem de Equações Estruturais Análise Fatorial Confirmatória
<p>Redução dos dados ou simplificação estrutural O fenômeno em estudo é representado de um modo tão simples quanto possível, sem sacrificar informações importantes.</p>	Análise Fatorial Exploratória
<p>Agrupamento de objetos ou variáveis Grupos de objetivos ou variáveis “similares” são criados com base nas medidas características.</p>	Análise de Conglomerados Análise de Correspondência Escalonamento Multidimensional

Fonte: Adaptado de Johnson e Wichern (1998, p.2).

Quadro 2

Métodos de Dependência

		Variável Dependente ou Predita				
		Uma Variável		Duas ou Mais Variáveis		
		Métrica	Não Métrica	Métrica	Não Métrica	
Variável Independente ou Preditor	Uma Variável	Métrica	Regressão Simples	Análise Discriminante Regressão Logística	Correlação Canônica	Correlação Canônica Não Paramétrica
		Não Métrica	Regressão Simples com Variável <i>Dummy</i>	Regressão Logística ANOVA(*)	Análise da Variância Multivariada (MANOVA)	Correlação Canônica Não Paramétrica
	Duas ou Mais Variáveis	Métrica	Regressão Múltipla	Análise Discriminante Regressão Logística	Correlação Canônica Modelagem de Equações Estruturais (SEM)	Correlação Canônica Não Paramétrica
		Não Métrica	Análise Conjunta Regressão Múltipla com Variável <i>Dummy</i>	Regressão Logística Análise Conjunta	Análise da Variância Multivariada (MANOVA) Modelagem de Equações Estruturais (SEM)	Correlação Canônica Não Paramétrica

Nota: (*) A ANOVA é considerada uma técnica estatística de análise univariada, não fazendo parte, portanto, do escopo deste estudo.

Fonte: Adaptado de Sharma (1996, p.6).

Quadro 3

Métodos de Interdependência

Variáveis Métricas	Variáveis Não Métricas
Análise Fatorial Exploratória Análise Fatorial Confirmatória Análise de Conglomerados Escalonamento Multidimensional	Análise de Correspondência Modelo Loglinear Escalonamento Multidimensional Análise de Conglomerados

Fonte: Adaptado de Sharma (1996, p.6).

Quadro 4

Premissas Subjacentes ao Uso das Técnicas Multivariadas

Premissas \ Técnicas (*)	1	2	3	4	5	6	7	8	9	10
Baixa multicolinearidade	X	X	X	X	X	X	X	X		
Ausência de erros correlacionados	X			X						
Ausência de observações atípicas	X	X	X	X	X	X		X		
Correlação entre as variáveis preditas					X					
Homoscedasticidade	X	X		X	X					
Linearidade	X	X		X	X	X	X			
Normalidade multivariada		X		X	X	X	X			
Normalidade univariada	X									
Número de variáveis maior ou igual a 3										X
Número de categorias por variável maior ou igual a 3									X	
Número de dimensões menor do que o número de casos										X
Padronização das variáveis							X	X		X
Tamanho da amostra	X	X	X	X	X	X	X	X	X	X
Tamanho dos grupos		X								

Nota: (*) Técnicas Multivariadas = 1 – Análise de Regressão; 2 – Análise Discriminante; 3 – Análise de Regressão Logística; 4 – Correlação Canônica; 5 – Análise Multivariada de Variância; 6 – Modelagem de Equações Estruturais; 7 – Análise Fatorial; 8 – Análise de Conglomerados; 9 – Análise de Correspondência; e 10 – Escalonamento Multidimensional.

Fontes: Peter (1979); Dillon e Goldstein (1984); Hoyle (1995); Peduzzi *et al.* (1996); Schumacker e Lomax (1996); Sharma (1996); Stevens (1996); Tabachnik e Fidell (1996); Kline (1998); Gujarati (2000); Pestana e Gageiro (2000); Aaker (2001); Malhotra (2001); Hair Jr. *et al.* (2009); Garson (2011a); Garson (2011b); Garson (2011c); Garson (2011d); Garson (2011e); Garson (2011f); Garson (2011g); Garson (2011h).

2.3.1. Multicolinearidade

A abordagem algébrica da multicolinearidade pode ser retratada sucintamente pelo diagrama de Ballentine. Na figura 1, os círculos de Y, X₂ e X₃ representam, respectivamente, as variações em Y (a variável dependente) e X₂ e X₃ (as variáveis

explicativas). O grau de multi(colinearidade) pode ser medido pelo nível de sobreposição (área sombreada) dos círculos X₂ e X₃ (GUJARATI, 2000, p.319-320).

Os efeitos da multicolinearidade podem ser classificados em termos de explicação e estimação. Com relação à explicação, na situação de multicolinearidade, o processo de se-

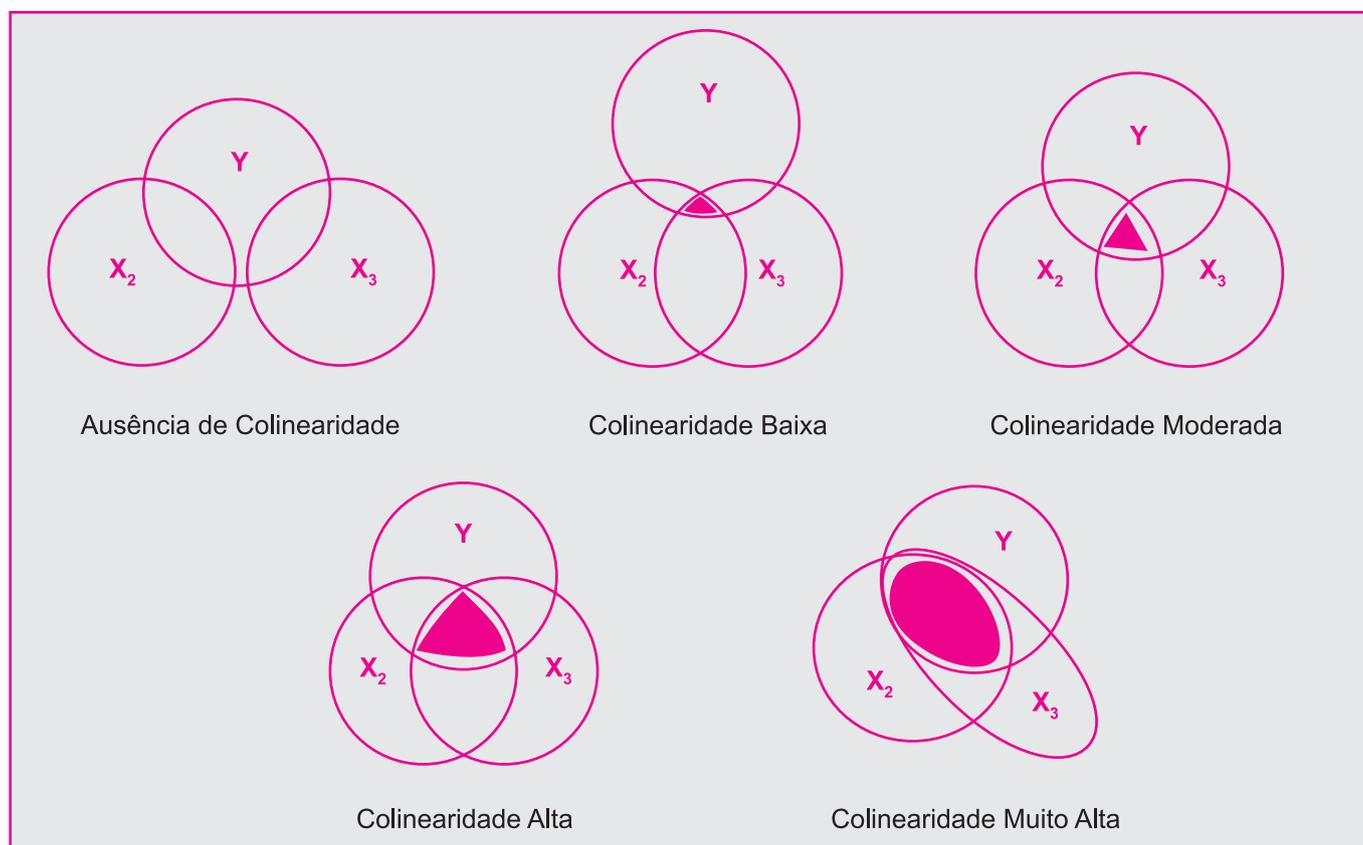


Figura 1: Diagrama de Ballentine

Fonte: Gujarati (2000, p.320).

paração de efeitos individuais de cada variável independente sobre uma dependente torna-se mais difícil, ficando mais complicada a determinação da contribuição de cada variável independente, porque os efeitos das variáveis independentes estão **misturados**.

- “A multicolinearidade resulta em partes maiores de variância compartilhada e níveis mais baixos de variância única da qual os efeitos das variáveis independentes individuais podem ser determinados” (HAIR JR. *et al.*, 2009, p.191).

Com relação à estimação, graus elevados de multicolinearidade podem fazer com que os coeficientes das variáveis independentes sejam incorretamente estimados e tenham até mesmo os sinais errados.

Em regressão múltipla, essa premissa é **relaxada** quando o objetivo for predição (GUJARATI, 2000).

2.3.2. Ausência de erros correlacionados (autocorrelação dos resíduos)

Segundo Kendall e Buckland (1971, p.8), o significado de autocorrelação pode ser entendido como a “correlação entre

membros de séries de observações ordenadas no tempo (séries temporais) ou no espaço (*crosssection*)”.

No escopo das técnicas, análise de regressão e correlação canônica pressupõe-se que não exista essa correlação, ou seja, admite-se que o erro referente a uma previsão qualquer não seja influenciado pelo erro de outra previsão qualquer.

Se houver correlação entre os erros, não se poderá afirmar a respeito de sua independência nos níveis de significância em que se está tentando fazer previsão da variável dependente. Provavelmente, alguma outra variável independente esteja afetando os resultados, sem ter sido incluída na análise (HAIR JR. *et al.*, 2009, p.86).

A literatura consultada apresenta alguns testes para detecção da presença de autocorrelação, entre eles: teste de Geary (ou teste das carreiras), teste de Breush-Godfrey, teste M de Durbin e teste de Durbin-Watson, o mais largamente utilizado (GUJARATI, 2000; HAIR JR. *et al.*, 2009).

2.3.3. Observações atípicas

As observações atípicas ou extremas (*outliers*) são aquelas substancialmente diferentes das outras.

Sobre os *outliers* multivariados, esses são casos com valores extremos no conjunto de múltiplas variáveis e operacionalmente definidos por meio de três medidas mais comumente usadas e disponíveis na maioria dos pacotes estatísticos comerciais: distância de Cook, distância de Mahalanobis e distância de Leverage (HAIR JR. *et al.*, 2009). De um ponto de vista prático, segundo Hair Jr. *et al.* (2009, p.77), observações atípicas podem ter um efeito sensível sobre qualquer tipo de análise empírica, como o cálculo da renda média familiar de 21 pessoas, havendo para 20 delas variações entre R\$ 20.000,00 e R\$ 100.000,00 ao ano, para a 21ª uma renda de R\$ 1 milhão ao ano. A observação atípica deve ser vista sob a ótica do quão representativa ela é na população, devendo o pesquisador avaliar se o valor marginal deve ser mantido ou eliminado, no caso de exercer influência indevida sobre os resultados.

2.3.4. Correlação entre as variáveis preditas

As variáveis dependentes na MANOVA devem apresentar certo grau de correlação, embora não muito acentuado; caso contrário, seria recomendável empregar a análise univariada de variância para cada uma das variáveis dependentes.

Testes individuais por variável dependente realizados na análise univariada de variância (ANOVA) ignoram as correlações entre as variáveis dependentes e usam menos do que a informação total disponível para avaliar as diferenças globais dos grupos. Quando as variáveis dependentes apresentam correlação entre elas, a MANOVA tem maior poder de detectar diferenças significantes nos grupos do que a ANOVA.

2.3.5. Homoscedasticidade

Diz-se que, quando a variância dos termos de erro parece constante ao longo do domínio da variável preditora, tem-se homoscedasticidade (HAIR JR. *et al.*, 2009). Essa é uma propriedade fundamental que deve ser garantida, sob pena de invalidar toda a análise estatística. Quando a homoscedasticidade não é atendida, podem ser observadas melhores previsões em alguns níveis das variáveis independentes do que em outros. Essa variabilidade afeta os erros padrões e torna os testes de hipóteses muito restritos ou insensíveis (HAIR JR. *et al.*, 2009, p.85).

A homoscedasticidade pode inicialmente ser verificada por meio de gráficos de resíduos (erros). Os gráficos dos erros contra os valores reais e contra os valores previstos são importantes. Se os pontos estão distribuídos aleatoriamente, sem demonstrar um comportamento definido, há homoscedasticidade.

A literatura consultada apresenta uma série de testes estatísticos para avaliação da homoscedasticidade, entre eles (GARSON, 2007; HAIR JR. *et al.*, 2009): teste de Goldfeld-Quandt; teste de Park; teste de Breusch-Pagan-Godfrey; teste de White's; teste de Levene para homogeneidade das variâncias; teste de Bartlett's para homogeneidade das variâncias; teste F-max; teste Box's M.

2.3.6. Linearidade

Conforme Hair Jr. *et al.* (2009), de forma geral os modelos lineares preveem valores que se ajustam a uma linha reta, que tem uma mudança com unidade constante da variável dependente em relação a uma mudança constante na variável independente.

Essa suposição está implícita em todas as técnicas multivariadas baseadas em medidas correlacionais de associação. Como as correlações representam apenas a associação linear entre variáveis, os efeitos não lineares não serão representados no valor da correlação, resultando em uma subestimação da força real da relação (HAIR JR. *et al.*, 2009, p.86).

A inspeção simples de gráficos como o diagrama de dispersão (*scatterplot*), por exemplo, é um método simples e comum para a verificação de linearidade. Além disso, a análise de correlação e outros testes mais complexos são disponíveis para detecção dessa premissa.

Huberty (1994) e Eisenbeis (1997) admitem o relaxamento dessa premissa quando a normalidade multivariada é atendida e quando a amostra é razoavelmente grande.

2.3.7. Normalidade

Nas técnicas estatísticas de análise multivariada que utilizam variáveis métricas e testes estatísticos para análise dos resultados, a normalidade multivariada é a condição mais fundamental de aplicação. Entretanto, no caso da análise de regressão, há a premissa de normalidade univariada, considerando-se apenas a variável referente aos resíduos.

Johnson e Wichern (1998) alertam que, para dados reais, a presença de variáveis com distribuição normal multivariada exata dificilmente ocorre. Hair Jr. *et al.* (2009, p.83) alertam que, se a variação em relação à distribuição normal é suficientemente grande, todos os testes estatísticos resultantes são inválidos, uma vez que a normalidade é exigida no emprego das estatísticas *F* e *t*.

Para Sharma (1996, p.380), há poucos métodos disponíveis para testar a normalidade multivariada. Baseado nas funções de Skewness e Kurtosis, o índice de Mardia encontra-se disponível no pacote estatístico LISREL e no pacote estatístico EQS.

2.3.8. Número de variáveis

Para a adequada aplicação da técnica escalonamento multidimensional, o número de variáveis deve ser maior ou igual a 3, segundo Hair Jr. *et al.* (2009).

2.3.9. Número de categorias por variável

Uma exigência da técnica análise de correspondência é que o número mínimo de categorias por variável deve ser três, porém, sem ser muito elevado, de acordo com Hair Jr. *et al.* (2009). A exigência de no mínimo três categorias por variável

ocorre para a análise de correspondência binária, tendo em vista que a fórmula do cálculo do número máximo de dimensões do mapeamento gerado pela técnica, a partir de uma tabela com r linhas e c colunas, é igual ao número mínimo entre r e c subtraído de 1 (HAIR JR. *et al.*, 2009, p.513).

2.3.10. Número de dimensões

Para o emprego de escalonamento multidimensional, o número de dimensões não pode superar o número de casos. Se houver mais dimensões que objetos, a solução da técnica será instável. Se houver muito poucos objetos em relação a dimensões, as medidas de ajuste serão **falsamente** boas (SPSS, 2003).

2.3.11. Padronização das variáveis

As técnicas análise de conglomerados e escalonamento multidimensional requerem uma padronização de variáveis métricas que apresentem diferentes magnitudes. No caso da análise fatorial, a padronização não é uma premissa; entretanto pode ser aplicada aos dados antes do processamento desta técnica (HAIR JR. *et al.*, 2009). A necessidade de padronização deve-se ao fato de a maioria das medidas de distância ser bastante sensível a diferentes escalas ou magnitudes entre as variáveis (HAIR JR. *et al.*, 2009, p.445).

2.3.12. Tamanho da amostra

Aliado a um correto procedimento de amostragem, o tamanho da amostra (número de casos, indivíduos, observações, entrevistas) deve ser adequado para permitir a generalização dos resultados, os quais podem ser verificados quanto à significância estatística dos testes.

As facilidades oferecidas pela maioria dos pacotes computacionais de aplicação estatística podem comprometer a observância do tamanho mínimo da amostra pelo pesquisador, fato que pode causar uma série de problemas importantes. Para Hair Jr. *et al.* (2009), essa omissão pode resultar em: baixíssimo poder estatístico dos testes de significância; um ajuste muito adequado dos dados, tornando os resultados artificialmente bons.

2.3.13. Tamanho dos grupos

Uma das exigências para o emprego adequado da técnica de análise discriminante é o dimensionamento de cada grupo estabelecido pela variável dependente. Segundo Hair Jr. *et al.* (2009), cada grupo deve ter no mínimo 20 observações para estabelecer-se a função discriminante.

3. METODOLOGIA DE PESQUISA

Nesta seção, serão apresentados os tópicos referentes a: caracterização da população, desenho metodológico da pesqui-

sa empírica, construção do critério de adequação da aplicação das técnicas multivariadas, instrumento de coleta de dados e variáveis de resultados.

3.1. Caracterização da população-alvo

Retomando as considerações iniciais sobre o público-alvo, a população sob análise contempla as dissertações e teses em Administração, área de *marketing*, temática do comportamento do consumidor, apresentadas nos programas de pós-graduação da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo e da Escola de Administração da Universidade Federal do Rio Grande do Sul, no período entre 1997 e 2006.

O universo de interesse foi definido como a dissertação ou tese com foco em *marketing*, especificamente comportamento do consumidor, e utilização de técnica estatística de análise multivariada como instrumento de solução do problema de pesquisa, apresentada aos programas de pós-graduação em Administração das instituições públicas com mais alta avaliação pela Capes na área de Administração.

A unidade populacional, no entanto, refere-se à aplicação de técnica estatística multivariada nesse universo de interesse, podendo ser mais do que uma aplicação em cada estudo integrante do universo de interesse.

A identificação da população-alvo foi realizada em três etapas:

- seleção dos trabalhos que apresentavam o termo comportamento do consumidor como uma das palavras-chave;
- leitura dos resumos dos trabalhos não selecionados na primeira etapa, com o objetivo de identificar aqueles que tratavam do comportamento do consumidor, ainda que não apresentassem o termo como palavra-chave e, assim, selecioná-los;
- exame do conteúdo dos trabalhos selecionados na primeira e na segunda etapa, identificando aqueles com aplicação de alguma técnica de análise multivariada.

Na tabela 1 apresenta-se o ambiente de identificação da população-alvo deste estudo, registrando 196 dissertações e teses sobre comportamento do consumidor (universo – U), 56 dissertações e teses com aplicação de técnicas estatísticas multivariadas (universo de interesse – UI) e 99 aplicações de técnicas estatísticas de análise multivariada (população-alvo – PA).

3.2. Desenho metodológico da pesquisa

A construção metodológica da pesquisa empírica foi orientada, inicialmente, pela avaliação do processo de solução do problema gerador de dissertações e teses com aplicação de técnica estatística de análise multivariada, na temática aqui selecionada.

Considerou-se, ainda, o fato de que essa avaliação exige ser orientada por um critério, o qual foi construído dentro deste

Tabela 1

Dissertações e Teses por Ano de Defesa

Ano	FEA-USP									EA-UFRGS									Total		
	Dissertações			Teses			Total			Dissertações			Teses			Total					
	U	UI	PA	U	UI	PA	U	UI	PA	U	UI	PA	U	UI	PA	U	UI	PA	U	UI	PA
2006	2	0	0	6	5	10	8	5	10	8	2	3	1	0	0	9	2	3	17	7	13
2005	12	6	11	3	1	2	15	7	13	4	3	5	5	1	2	9	4	7	24	11	20
2004	10	5	7	4	2	5	14	7	12	10	4	8	1	0	0	11	4	8	25	11	20
2003	9	0	0	3	1	2	12	1	2	9	3	4	0	0	0	9	3	4	21	4	6
2002	2	1	4	4	3	5	6	4	9	17	3	6	1	0	0	18	3	6	24	7	15
2001	8	1	1	1	1	3	9	2	4	19	5	8	2	1	1	21	6	9	30	8	13
2000	12	0	0	5	1	1	17	1	1	11	5	9	0	0	0	11	5	9	28	6	10
1999	3	1	1	1	0	0	4	1	1	5	1	1	1	0	0	6	1	1	10	2	2
1998	2	0	0	2	0	0	4	0	0	4	0	0	0	0	0	4	0	0	8	0	0
1997	5	0	0	0	0	0	5	0	0	3	0	0	1	0	0	4	0	0	9	0	0
Total	65	14	24	29	14	28	94	28	52	90	26	44	12	2	3	102	28	44	196	56	99

Nota: U = Universo; UI = Universo de Interesse; PA = População-Alvo.

próprio estudo, não tendo sido submetido a um processo de validação anterior.

Essas duas condições, por si sós, remetem este estudo ao âmbito da pesquisa exploratória, visto que o processo de aprofundamento do entendimento do problema é uma etapa aqui cumprida para subsidiar a construção do critério de avaliação da adequação da aplicação da estatística multivariada aos trabalhos selecionados do período de 1997 a 2006.

Portanto, este estudo insere-se na abordagem quantitativa, utilizando a análise de conteúdo. O delineamento exploratório para este estudo vai ao encontro da abordagem de Selltiz (1974, p.60), que evidencia ser uma das finalidades desse método a apresentação de um recenseamento de problemas considerados urgentes por pessoas que trabalham em determinado campo de relações sociais. Exemplifica essa convergência o alerta feito em outros estudos de que a rigidez teórica, explicitada nas premissas para aplicações das técnicas estatísticas de análise multivariada, parece não ser acompanhada de rigidez empírica, explicitada nas concessões feitas pelos pesquisadores, o que, por vezes, pode resultar em conclusões não precisas sobre a solução dos problemas.

A opção pelo uso da técnica de análise de conteúdo foi orientada especialmente pela necessidade primária deste estudo de interpretar a situação problema das dissertações e teses selecionadas com o propósito de identificar o processo de sua solução por meio da seleção de uma técnica estatística de análise multivariada pertinente.

Da mesma forma como encontrado na abordagem de vários autores, entre eles Richardson (1999, p.221-222) e Rocha e Deusdará (2005, p.309), também aqui a identificação precisa da natureza da técnica de análise oscila entre a discussão quantitativa e a qualitativa.

Assim, de um lado, busca-se uma objetividade bastante intensa (RICHARDSON, 1999, p.221) para categorização dos problemas das dissertações e teses, mediante a definição de critérios rígidos de julgamento da solução desses problemas. De outro, o processo geral de avaliações a ser implementado às unidades de análise (dissertações e teses) está contaminado por julgamentos, na medida em que a análise do conteúdo supõe também o exame das características ausentes ou registros parciais do atendimento às premissas da técnica estatística utilizada.

Nesse sentido, algumas definições de análise de conteúdo parecem contemplar o caminho metodológico da coleta de dados, de acordo com propostas encontradas em Janis (1986, *apud* RICHARDSON, 1999, p.222):

- “Assim, a análise de conteúdo pode ser definida como qualquer técnica: na base de regras explicitamente formuladas e sempre quando os juízos do analista sejam considerados como relatórios de um observador científico”.

Entretanto, a definição encontrada em Bardin (1979, p.31) propicia a convivência com as duas abordagens, na medida em que inclui em sua definição de análise de conteúdo, a geração

de indicadores quantitativos ou não no processo analítico:

- “Análise de conteúdo é um conjunto de técnicas de análise das comunicações, visando obter, através de procedimentos sistemáticos e objetivos de descrição dos conteúdos das mensagens, indicadores (quantitativos ou não), que permitam inferir conhecimentos relativos às condições de produção dessas mensagens”.

3.3. A construção do critério de avaliação da aplicação das técnicas multivariadas

Dois critérios são apresentados a seguir com a finalidade de avaliar o grau de acerto na aplicação das técnicas multivariadas.

3.3.1. Procedimento de categorização do problema de pesquisa da unidade de análise – critério 1

Neste artigo destacam-se os resultados correspondentes à pertinência e à adequação do emprego das técnicas multivariadas. A orientação para categorizar o problema de pesquisa de cada unidade de análise (unidade *i* de análise) foi dada pela finalidade ou objetivo teórico de aplicação de diferentes técnicas estatísticas de análise multivariada. Segundo Johnson e Wichern (1998, p.2), as categorias são sintetizadas em cinco modalidades: redução dos dados ou simplificação estrutural; agrupamento; dependência entre variáveis; predição; formulação de hipóteses e testes.

A categoria identificada na unidade *i* de análise foi confrontada com a categoria objetivo das técnicas multivariadas, na condição de tratamento estatístico aplicado para solução

do problema. A avaliação feita por meio desse confronto tem um caráter dicotômico, ou seja, foi considerada como uma aplicação adequada da técnica multivariada quando a categoria objetivo de aplicação da técnica ajustou-se à categoria do problema de pesquisa da unidade *i* de análise.

3.3.2. Procedimento de avaliação do atendimento às premissas das técnicas multivariadas – critério 2

O atendimento às premissas das técnicas multivariadas foi decorrente da confirmação de que todas as premissas postuladas pela teoria foram atendidas. A violação de pelo menos uma das premissas, explicitada no conteúdo apresentado na unidade *i* de análise, bem como a não explicitação da situação de cada premissa (atendimento ou não atendimento) resultou na categoria **não atendimento às premissas**. Dessa forma, o critério de avaliação do atendimento às premissas é dicotômico: atendimento a todas as premissas das técnicas e não atendimento a pelo menos uma das premissas dessas técnicas.

3.3.3. Avaliação final do nível de adequação do uso das técnicas multivariadas

O registro do nível de adequação do uso das técnicas multivariadas foi feito em três níveis, como pode ser visto na figura 2, que estão descritos a seguir.

- **Nível 1** (não adequação do uso da técnica) – não ajuste da categoria objetivo de aplicação das técnicas à categoria do problema de pesquisa da unidade *i* de análise, independen-

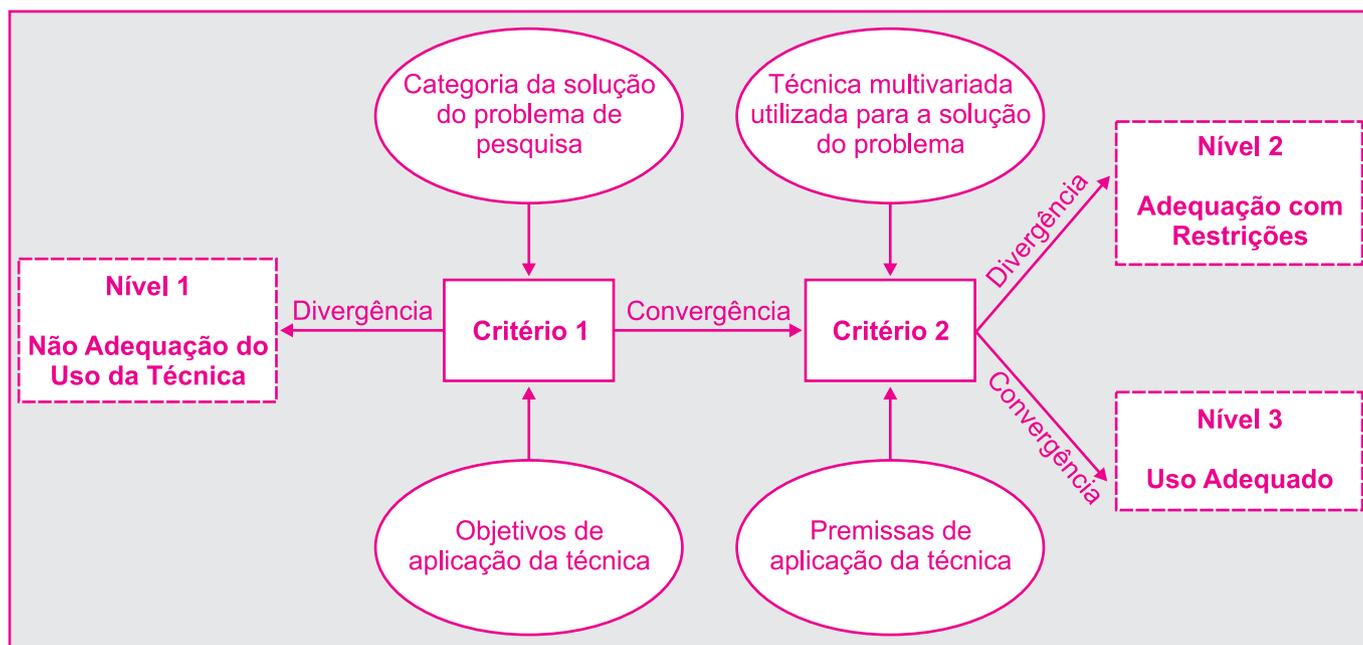


Figura 2: Avaliação do Nível de Adequação das Técnicas

temente do atendimento ou não atendimento às premissas dessa técnica (não atendimento ao **critério 1**).

- **Nível 2** – ajuste da categoria objetivo de aplicação das técnicas à categoria do problema de pesquisa da unidade **i** de análise e não atendimento a pelo menos uma das premissas das técnicas ou à não explicitação da situação de cada premissa (atendimento ou não atendimento) no documento da unidade **i** (atendimento ao **critério 1** e não atendimento ao **critério 2**).
- **Nível 3** – ajuste da categoria objetivo de aplicação das técnicas à categoria do problema de pesquisa da unidade **i** de análise e atendimento de todas as premissas da técnica, conforme documento da unidade **i** (atendimento ao **critério 1** e ao **critério 2**).

3.4. Instrumento de coleta de dados e resultados

Para a coleta de dados foi desenvolvido um instrumento para o registro das ocorrências encontradas em cada unidade **i** de análise. Esses registros referem-se a:

- identificação numérica das unidades de análise;
- ano de defesa do trabalho;
- descrição do problema de pesquisa;
- categorização do problema de pesquisa a partir dos objetivos de aplicação das técnicas de análise multivariada;
- identificação da técnica de análise multivariada utilizada no trabalho;
- aplicação do critério 1 – categorização do problema de pesquisa da unidade de análise;
- identificação e contagem das premissas atendidas na aplicação das técnicas;
- aplicação do critério 2 – procedimento de avaliação do atendimento às premissas das técnicas multivariadas;
- avaliação final do nível de adequação do uso das técnicas multivariadas.

Os resultados são identificados por meio da ocorrência evidenciada em cada critério. A partir desses resultados, a complementação do plano analítico pressupõe:

- evidenciar quais técnicas estatísticas de análise multivariada foram mais utilizadas no período de estudo, independentemente da categoria dos problemas de pesquisa;
- evidenciar as categorias dos problemas de pesquisa que estariam apresentando maior número de ocorrências de aplicação;
- evidenciar quais premissas são mais negligenciadas no processo de aplicação das técnicas multivariadas.

4. ANÁLISE DOS RESULTADOS

A primeira etapa de análise dos resultados tratou de apresentar as evidências quanto à intensidade de uso das técnicas multivariadas de modo geral, conforme o primeiro objetivo.

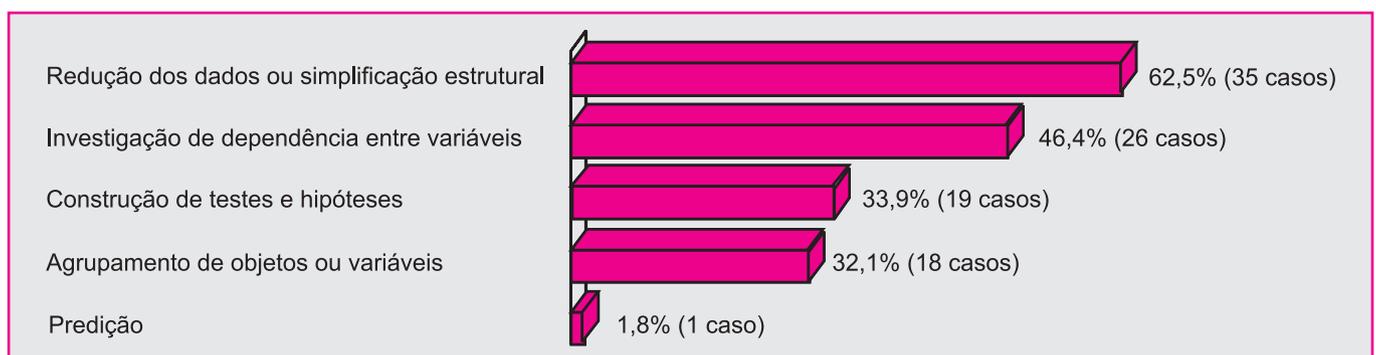
A segunda parte, conforme o segundo objetivo, tratou do nível de adequação do uso das técnicas multivariadas aos problemas de pesquisa das dissertações e teses sob análise.

A terceira parte identificou, a partir do não atendimento às premissas básicas de aplicação dessas técnicas, fontes potenciais de erro.

4.1. Resultados relativos ao primeiro objetivo

Entre 1997 e 2006, as dissertações e teses sobre a temática comportamento do consumidor, defendidas nos programas de pós-graduação em Administração da FEA-USP e da EA-UFRGS, apresentaram maior intensidade de uso de técnicas estatísticas multivariadas para atender ao objetivo de reduzir ou simplificar a estrutura de dados coletada (62,5%), conforme apresentado no gráfico a seguir, com a aplicação da técnica de análise fatorial exploratória, única representante dessa categoria.

A categoria de investigação de dependência entre variáveis foi a segunda mais utilizada (46,4%). Nesse contexto, seis téc-



**Categoria dos Objetivos de Aplicação das Técnicas Estatísticas Multivariadas
(Base: Universo de Interesse)**

nicas de análise multivariada apresentam essa categoria como um dos objetivos de aplicação: análise de regressão (15 casos), análise discriminante (3 casos), análise multivariada da variância (3 casos), análise conjunta (3 casos), regressão logística (1 caso) e correlação canônica (1 caso).

A construção de hipóteses e testes, exclusivamente representada neste estudo pela técnica estatística multivariada de modelagem de equações estruturais somou 19 casos (33,9%).

Já a categoria de agrupamento de objetos ou variáveis foi utilizada em 18 estudos (33,9%), com maior número de casos no uso da técnica de análise de conglomerados (12 casos), seguido da técnica de análise de correspondência (5 casos) e da técnica de escalonamento multidimensional (1 caso).

Constatou-se que apenas um dos estudos selecionados utilizou técnica de análise multivariada com o objetivo de predição de algum fenômeno ou fato (regressão logística), o que possivelmente ocorra mais pela ausência de interesse por esse problema de pesquisa na temática de comportamento do consumidor do que pela restrição do número de técnicas disponíveis, visto que essa categoria é representada, neste estudo, por três técnicas de análise multivariada: análise de regressão, análise discriminante e regressão logística.

Ainda, a fim de ilustrar o ambiente de uso das técnicas estatísticas multivariadas, registre-se que, em 64,3% dos estudos selecionados, pelo menos duas técnicas são utilizadas para atendimento à solução dos problemas de pesquisa. O uso mais intenso das técnicas ocorre, em termos relativos, nas teses, com 81,2% dos casos com aplicação de pelo menos duas técnicas, contra 57,5% nas dissertações, sugerindo maior sofisticação dos estudos no primeiro grupo. Na tabela 2 consta a quantidade de técnicas de análise multivariada diferentes usadas nas dissertações e teses.

De 1997 a 1998 não se detectou aplicação de técnica estatística multivariada nas instituições de ensino alvo deste estudo. Quanto à categoria de aplicação das técnicas multivariadas ao longo do período de 1997 a 2006, devem-se registrar as seguintes considerações:

- a categoria de redução dos dados ou simplificação estrutural apresentou ligeira queda no período, porém mantém pelo menos 30% de aplicação desde 1999;
- as técnicas integrantes de categoria de investigação de dependência entre variáveis apresentaram tendência de queda no período de 1999 a 2003, voltando a crescer a partir de 2004;
- a categoria de construção de hipóteses e testes apresentou, de forma geral, comportamento relativamente estável, registrando pico de uso em 2003 (67%);
- a categoria de agrupamento de objetos ou variáveis apresentou tendência de crescimento no período estudado;
- as técnicas integrantes de categoria predição só apresentaram uso em 2004.

Quanto à intensidade de uso de cada técnica ao longo do tempo, nada se pode afirmar sobre possíveis tendências de aplicação. Ainda assim, alguns aspectos merecem ser destacados:

- somente a partir de 2005 foram encontrados estudos que utilizaram a técnica estatística multivariada de análise multivariada de variância;
- houve aplicação de análise fatorial exploratória em pelo menos 30% dos estudos, em todos os anos selecionados;
- a técnica de modelagem de equações estruturais apresentou incremento de uso, relativamente a outras técnicas, no período de 2000 a 2003.

A tabela 3 ilustra esses resultados.

4.2. Resultados relativos ao segundo objetivo

Um resultado positivo deste estudo é a coerência, nas dissertações e teses selecionadas, quanto à adequação das técnicas multivariadas ao problema de pesquisa. Assim, pelo critério 1, todas as aplicações das técnicas multivariadas são **adequadas**.

Tabela 2

Quantidade de Técnicas de Análise Multivariada Diferentes Usadas nas Dissertações e Teses

Técnica Estatística de Análise Multivariada	Nível do Estudo		Total de Casos %
	Dissertação %	Tese %	
Uma técnica	42,5 (17 casos)	18,8 (3 casos)	35,7 (20 casos)
Duas técnicas	47,5 (19 casos)	68,7 (11 casos)	53,6 (30 casos)
Três técnicas	7,5 (3 casos)	12,5 (2 casos)	8,9 (5 casos)
Quatro técnicas	2,5 (1 caso)	Nenhum caso	1,8 (1 caso)

Tabela 3

Distribuição do Uso das Técnicas Multivariadas ao Longo do Tempo (Período de 1997 a 2006)

Técnica	1999	2000	2001	2002	2003	2004	2005	2006
Análise de Regressão	-	3	5	1	-	2	3	1
Análise Discriminante	-	-	-	1	-	1	-	1
Análise de Regressão Logística	-	-	-	-	-	1	1	-
Correlação Canônica	-	-	-	1	-	-	-	-
Análise Multivariada de Variância	-	-	-	-	-	-	1	2
Análise Conjunta	1	-	-	-	-	2	-	-
Modelagem de Equações Estruturais	-	1	3	4	4	3	4	-
Análise Fatorial Exploratória	1	5	4	5	2	7	6	5
Análise de Conglomerados	-	-	1	2	-	3	3	3
Análise de Correspondência	-	1	-	1	-	1	2	-
Escalonamento Multidimensional	-	-	-	-	-	-	-	1
Total	2	10	13	15	6	20	20	13

4.3. Resultados relativos ao terceiro objetivo

Na tabela 4 apresentam-se os índices de atendimento às premissas das técnicas multivariadas que foram registrados nas análises das dissertações e teses estudadas neste trabalho.

Convém lembrar que, analogamente ao quadro 4, não foi incluída a análise conjunta na tabela 4 pelo fato de nenhuma das premissas teóricas aplicar-se a essa técnica. O total de aplicações de análise multivariada é 96, porque não foram incluídas as três análises conjuntas.

Antes de prosseguir com os comentários a respeito da tabela 4, cumpre esclarecer que o cálculo dos valores em porcentagem sobre o número de aplicações de cada técnica tem a finalidade apenas de simplificar a comparação entre os resultados por técnica, apesar de elas apresentarem números absolutos de uso muito baixos.

Primeiramente serão destacados os resultados da avaliação das premissas das técnicas de dependência.

As técnicas de análise de regressão, análise discriminante e correlação canônica apresentaram baixos níveis de atendimento às premissas subjacentes a seu uso nessas dissertações e teses. Apenas 2 num total de 15 estudos atenderam a mais da metade das premissas de análise de regressão e nenhum dos poucos casos de aplicação de análise discriminante e correlação canônica atendeu a mais da metade das premissas. Quanto a MANOVA e modelagem de equações estruturais, foram altos os índices de atendimento de mais da metade de

suas condições: todos os 3 casos e 15 entre os 19 casos estudados, respectivamente.

A técnica de análise de regressão foi a que apresentou o menor número de premissas atendidas. Mais da metade das dissertações e teses (8, num total de 15) não atendeu a nenhuma das premissas exigidas. Com exceção do tamanho da amostra, nenhuma das premissas apresentou mais de um quarto de atendimento, com destaque negativo para homoscedasticidade e linearidade, com atendimento por 1 caso entre os 19 estudados, em cada uma dessas premissas.

O atendimento à multicolinearidade verificado em duas aplicações da análise discriminante foi resultante do uso do método *stepwise*.

Quanto à regressão logística, com uma análise limitada em função do pequeno número de aplicações encontradas, a aplicação foi menos prejudicada até em função do pequeno número de premissas que necessitam ser atendidas (multicolinearidade, tratamento de observações atípicas – *outliers* e sensibilidade ao tamanho da amostra).

Quanto à correlação canônica, sobre a multicolinearidade registre-se que sua verificação deu-se por meio de um processo de inércia, visto que as variáveis de entrada da técnica foram escores fatoriais já construídos com baixa correlação pela análise fatorial.

Com três casos de aplicação da análise multivariada de variância, é reforçada a ausência da prática de verificação da linearidade e multicolinearidade (sob o conceito de verificação direta) no uso das técnicas estatísticas multivariadas. Em

Tabela 4

Atendimento às Premissas Subjacentes ao Uso das Técnicas Multivariadas – Porcentagem sobre o Total de Aplicações de cada Técnica

Premissas \ Técnicas ^(*)	1	2	3	4	5	6	7	8	9	10
Baixa multicolinearidade	3	2	2	1	1	8	23	3		
Ausência de erros correlacionados	2			0						
Ausência de observações atípicas	0	0	1	0	2	13		3		
Correlação entre as variáveis preditas					3					
Homoscedasticidade	1	1		0	3					
Linearidade	1	0		0	0	8	6			
Normalidade multivariada		0		0	0	5	1			
Normalidade univariada	2									
Número de variáveis										1
Número de categorias por variável									5	
Número de dimensões										1
Padronização das variáveis							31	11		0
Tamanho da amostra	12	1	2	1	3	16	33	8	3	1
Tamanho dos grupos		3								
Total de aplicações	15	3	2	1	3	19	35	12	5	1
Atendimento a nenhuma premissa	8	0	0	0	0	1	3	0	0	0
Atendimento a até 50% das premissas	10	3	0	1	0	4	26	3	0	0
Atendimento a mais de 50% das premissas	2	0	1	0	3	15	9	9	5	1
Atendimento a todas as premissas	1	0	1	0	0	3	4	1	0	0

Nota: (*) Técnicas Multivariadas = 1 – Análise de Regressão; 2 – Análise Discriminante; 3 – Análise de Regressão Logística; 4 – Correlação Canônica; 5 – Análise Multivariada de Variância; 6 – Modelagem de Equações Estruturais; 7 – Análise Fatorial; 8 – Análise de Conglomerados; 9 – Análise de Correspondência; e 10 – Escalonamento Multidimensional.

nenhum dos casos verificou-se a premissa de linearidade e em apenas um foi verificada a premissa de multicolinearidade.

A técnica estatística de análise multivariada de modelagem de equações estruturais apresentou a maior quantidade de premissas atendidas nos estudos selecionados. Assim, 15 entre os 19 trabalhos atenderam a mais da metade das premissas exigidas.

Com relação às técnicas de interdependência, é também possível detectar quais são as premissas com menor índice de atendimento.

A análise fatorial apresentou baixos níveis de atendimento às premissas subjacentes a seu uso nessas dissertações e teses.

Apenas 9 entre os 35 estudos atenderam a mais da metade das premissas e só 4 atenderam a todas as premissas dessa técnica, conforme a tabela 4. Contribuiu para esse baixo nível de atendimento a não verificação da premissa de linearidade, com menos de 20% de estudos que declararam seu uso.

A técnica de análise de conglomerados apresentou um número expressivo de premissas atendidas (9 casos entre os 12 estudados de atendimento a mais de 50% das premissas). Registre-se que as duas premissas que exigiram maior participação do pesquisador quando da aplicação da técnica (multicolinearidade e tratamento de observações atípicas – *outliers*) são as que obtiveram menor taxa de uso (3 casos em cada).

Nos 5 casos em que se detectou a aplicação da análise de correspondência, foram respeitadas duas das três premissas encontradas na literatura pesquisada. A condição não atendida em alguns trabalhos ocorreu para a premissa de tamanho suficiente da amostra para garantia de frequências esperadas maiores ou iguais a 5.

No único trabalho em que se detectou a aplicação do escalonamento multidimensional, foram respeitados três dos quatro preceitos encontrados no referencial teórico consultado. A exceção ocorreu para a premissa de padronização de variáveis que deveria, mas não foi aplicada.

4.4. Critério 2 de avaliação dos níveis de adequação no uso das técnicas multivariadas

O resultado final da avaliação das aplicações encontradas na população-alvo sugere um frágil ambiente operacional de aplicação das técnicas multivariadas no que se refere a seu uso e especificamente ao atendimento às respectivas premissas.

4.5. Avaliação final do nível de adequação do uso das técnicas multivariadas

A avaliação final do nível de adequação do uso das técnicas multivariadas pode ser sintetizada pelos índices percentuais obtidos nos três níveis retratados na figura 2.

- **Nível 1** – todas as aplicações das 11 técnicas multivariadas (inclusive análise conjunta) apresentaram convergência entre a categoria-objetivo de aplicação de cada técnica com a categoria do problema da pesquisa.
- **Nível 2** – 14 de 15 estudos (regressão linear), 1 de 1 estudo (discriminante), 1 de 2 estudos (logística), 1 de 1 estudos (correlação canônica), 5 de 5 estudos (MANOVA), 16 de 19 estudos (modelagem de equações estruturais), 31 de 35 estudos (fatorial), 11 de 12 estudos (conglomerados), 5 de 5 estudos (análise correspondência) e 1 de 1 estudo (escalonamento) das aplicações das técnicas atenderam parcialmente a suas premissas.
- **Nível 3** – 1 de 15 estudos (regressão linear), 0 de 1 estudo (discriminante), 1 de 2 estudos (logística), 0 de 1 estudo (correlação canônica), 0 de 5 estudos (MANOVA), 3 de 19 estudos (modelagem de equações estruturais), 4 de 35 estudos (análise fatorial), 1 de 12 estudos (análise de conglomerados), 0 de 5 estudos (análise de correspondência) e 0 de 1 estudo (escalonamento multidimensional) das aplicações das técnicas apresentaram uso adequado quanto ao atendimento de suas premissas.

5. CONCLUSÕES

Nos trabalhos avaliados, as soluções para os problemas de pesquisa concentraram-se no emprego de técnicas de redução ou simplificação estrutural dos dados (35 estudos), seguido de técnicas de investigação de dependência entre variáveis (26 estudos).

Embora não seja possível fazer afirmações sobre tendência de intensidade de uso da análise multivariada, este estudo permitiu observar as técnicas que se destacaram nos trabalhos focalizados. Houve aplicação de análise fatorial exploratória em pelo menos 30% dos estudos, em todos os anos selecionados neste trabalho. A técnica de modelagem de equações estruturais apresentou incremento de uso, relativamente a outras técnicas, no período de 2000 a 2003.

Cumprir destacar o acerto dos analistas na identificação das situações favoráveis de uso das técnicas multivariadas na resolução dos problemas de pesquisa das dissertações e teses focalizadas neste estudo.

Quanto à qualidade de aplicação dessas técnicas, destacou-se que a verificação de todas as suas premissas só foi constatada em 6,7% (regressão linear), 0% (discriminante), 50% (logística), 0% (correlação canônica), 0% (MANOVA), 15,8% (modelagem de equações estruturais), 11,4% (análise fatorial), 8,3% (análise de conglomerados), 0% (análise de correspondência) e 0% (escalonamento multidimensional).

Assim, ainda que não possam ser extrapolados para a produção acadêmica de outras instituições ou de outros períodos de tempo, os achados deste estudo sugerem a necessidade de maior cuidado conceitual nas aplicações das técnicas multivariadas.

A partir dos resultados aqui obtidos, pode-se ressaltar a importância das bancas de qualificação no sentido de alertar os pesquisadores sobre a necessidade desses cuidados, incentivando a prática do atendimento às premissas inerentes a essas técnicas.

Sobre isso parece pertinente a sugestão da disponibilização pelos programas de pós-graduação em Administração de infraestrutura de suporte ao pesquisador quando da utilização da pesquisa quantitativa. Além disso, o oferecimento de cursos de capacitação a esses pesquisadores, bem como aos docentes desses programas, pode auxiliar, internamente, na melhor qualidade da orientação e, externamente, na melhor qualidade das observações em bancas de qualificação das quais participam.

Espera-se que o resultado apurado neste estudo, relativo ao elevado nível de não adequação do uso das técnicas multivariadas na população-alvo, em função da não verificação de todas as suas premissas, seja um estímulo aos pesquisadores para a busca de um entendimento mais detalhado e aprofundado das técnicas a serem empregadas como parte do processo de solução do problema de pesquisa, de forma a minimizar eventuais erros decorrentes de aplicação não adequada. ♦

- AAKER, D. *Pesquisa de marketing*. São Paulo: Atlas, 2001.
- ARKADER, R. A pesquisa científica em gerência de operações no Brasil. *Revista de Administração de Empresas (RAE)*, São Paulo, v.43, n.1, p.70-79, jan./mar. 2003.
- BARDIN, L. *Análise de conteúdo*. Lisboa: Edições 70, 1979.
- BERTERO, C.O.; KEINERT, T.M.M. A evolução da análise organizacional no Brasil (1961-93). *Revista de Administração de Empresas (RAE)*, São Paulo, v.34, n.3, p.81-90, maio/jun. 1994.
- BERTERO, C.O.; KEINERT, T.M.M. Produção científica em administração de empresas: provocações, insinuações e contribuições para um debate local. *Revista de Administração Contemporânea (RAC)*, Curitiba, v.3, n.1, p.147-178, jan./mar./abr. 1999.
- BOTELHO, D.; MACERA, A. Análise meteórica de teses e dissertações da área de marketing apresentadas na FGV-EAESP (1974-1999). In: ENCONTRO DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 25., 2001, Campinas. *Anais...* Campinas: Anpad, 2001.
- BREI, V.A.; LIBERALI, G. O uso de modelagem em equações estruturais na área de marketing no Brasil. In: ENCONTRO DE MARKETING DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO (EMA), 1., Porto Alegre, 2004. *Anais...* Porto Alegre, RS: Anpad, 2004.
- CALDAS, M.P.; TONELLI, M.J.; LACOMBE, B.M.B. Espelho, espelho meu: meta-estudo da produção científica em Recursos Humanos nos EnANPADs da década de 90. In: ENCONTRO DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 26., 2002, Salvador. *Anais...* Salvador, BA: Anpad, 2002.
- DILLON, W.R.; GOLDSTEIN, M. *Multivariate analysis: methods and applications*. USA: John Wiley & Sons, 1984.
- EISENBEIS, R. Pitfalls in the application of discriminant analysis in business, finance and economics. *The Journal of Finance*, Berkeley, v.XXXII, n.3, p.875-900, June 1997.
- FRICK, R.W. Interpreting statistical testing: process and propensity, not population and random sampling. *Behavior Research Methods, Instruments & Computers*, New York, v.30, n.3, p.527-535, Aug. 1998.
- GAITO, J. Measurement scales and statistics: resurgence of an old misconception. *Psychological Bulletin*, Washington, v.87, n.3, p.564-567, May 1980.
- GAMBOA, S.S. *Epistemologia da pesquisa em educação*. 1987. Tese (Doutorado) – Universidade Estadual de Campinas, Campinas, São Paulo, Brasil, 1987.
- GARSON, G.D. *Testing of assumptions, from statnotes: topics in multivariate analysis*. 2011a. Disponível em: <www2.chass.ncsu.edu/garson/pa765.statnote.html>. Acesso em: 18 jul. 2011.
- GARSON, G.D. *Cluster analysis, from statnotes: topics in multivariate analysis*. 2011b. Disponível em: <www2.chass.ncsu.edu/garson/pa765.statnote.html>. Acesso em: 20 jul. 2011.
- GARSON, G.D. *Fatorial analysis, from statnotes: topics in multivariate analysis*. 2011c. Disponível em: <www2.chass.ncsu.edu/garson/pa765.statnote.html>. Acesso em: 20 jul. 2011.
- GARSON, G.D. *Discriminant analysis, from statnotes: topics in multivariate analysis*. 2011d. Disponível em: <www2.chass.ncsu.edu/garson/pa765.statnote.html>. Acesso em: 20 jul. 2011.
- GARSON, G.D. *Canonical correlation, from statnotes: topics in multivariate analysis*. 2011e. Disponível em: <www2.chass.ncsu.edu/garson/pa765.statnote.html>. Acesso em: 20 jul. 2011.
- GARSON, G.D. *Regression analysis, from statnotes: topics in multivariate analysis*. 2011f. Disponível em: <www2.chass.ncsu.edu/garson/pa765.statnote.html>. Acesso em: 20 jul. 2011.
- GARSON, G.D. *Logistic regression, from statnotes: topics in multivariate analysis*. 2011g. Disponível em: <www2.chass.ncsu.edu/garson/pa765.statnote.html>. Acesso em: 20 jul. 2011.
- GARSON, G.D. *Structural equation modeling, from statnotes: topics in multivariate analysis*. 2011h. Disponível em: <www2.chass.ncsu.edu/garson/pa765.statnote.html>. Acesso em: 20 jul. 2011.
- GARSON, G.D. *Testing of assumptions, from statnotes: topics in multivariate analysis*. 2007. Disponível em: <www2.chass.ncsu.edu/garson/pa765.statnote.html>. Acesso em: 20 jul. 2007.
- GUJARATI, D.N. *Econometria básica*. São Paulo: Makron Books, 2000.
- HAIR JR., J.F.; WILLIAM, B.; BABIN, B.; ANDERSON, R.E. *Análise multivariada de dados*. 6.ed. Porto Alegre: Bookman, 2009.
- HOPPEN, N. Sistemas de informação no Brasil: uma análise dos artigos científicos dos anos 90. In: ENCONTRO DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 22., 1998, Foz do Iguaçu. *Anais...* Foz do Iguaçu, PR: Anpad, 1998.
- HOYLE, R.H. *Structural equation modeling: concepts, issues, and applications*. Thousand Oaks, CA: Sage, 1995.
- HUBERTY, C.J. *Applied discriminant analysis*. New York: John Wiley, 1994.

- JOHNSON, R.A.; WICHERN, D.W. *Applied multivariate statistical analysis*. 4th ed. New Jersey: Prentice Hall, 1998.
- KENDALL, M.G.; BUCKLAND, W.R. *A dictionary of statistical terms*. New York: Hafner, 1971.
- KLINE, R.B. *Principles and practices of structural equation modeling*. New York: Guilford, 1998.
- LEAL, R.P.C.; OLIVEIRA, J.; SOLURI, A. Perfil da pesquisa em finanças no Brasil. *Revista de Administração de Empresas*, São Paulo, v.43, n.1, p.91-103, jan./mar. 2003.
- LOURENÇO, A.; MATIAS, R.P. *Estatística multivariada*. Porto: Instituto Superior de Engenharia do Porto, 2001.
- MACHADO-DA-SILVA, C. Produção acadêmica em administração pública: período 1983-1988. In: ENCONTRO DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 13., 1989, Florianópolis. *Anais...* Florianópolis, SC: Anpad, 1989. p.1599-1618.
- MACHADO-DA-SILVA, C. Organizações: o estado da arte da produção acadêmica no Brasil. In: ENCONTRO DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 14., 1990, Rio de Janeiro. *Anais...* Rio de Janeiro: Anpad, 1990.
- MALHOTRA, N.K. *Pesquisa de marketing*. Porto Alegre: Bookman, 2001.
- MARTINS, G.A. *Epistemologia da pesquisa em administração*. 1994. Tese de (Livre-Docência) – Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, São Paulo, São Paulo, Brasil, 1994.
- MILAGRE, R.A. *Estatística: uma proposta de ensino para os cursos de administração de empresas*. 2001. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brasil, 2001.
- PEDUZZI, P.; CONCATO, J.; KEMPER, E.; HOLFORD, T.R.; FEINSTEIN, A.R. A simulation of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, Maryland, v.99, n.12, p.1373-1379, Dec. 1996.
- PERIN, M.G. A pesquisa survey em artigos de marketing nos Enanpads da década de 90. *Revista Interdisciplinar de Marketing*, Maringá, v.1, n.1, p.44-59, jan./mar./abr. 2002.
- PESTANA, M.H.; GAGEIRO, J.N. *Análise de dados para ciências sociais: a complementaridade do SPSS*. 2.ed. Lisboa: Silabo, 2000.
- PETER, J.P. Reliability: a review of psychometric basics and recent marketing practices. *Journal of Marketing Research*, Chicago, v.16, n.1, p.6-17, 1979.
- POWERS, T.; SWAN, J.; PATTOM, J.F. Career research productivity patterns of marketing academicians. *Journal of Business Research*, New York, v.42, n.1, p.75-86, May 1998.
- RICHARDSON, R.J. *Pesquisa social*. 3.ed. São Paulo: Atlas, 1999.
- ROCHA, D.; DEUSDARÁ, B. Análise de conteúdo e análise do discurso: aproximações e afastamentos na re(construção) de uma trajetória. *ALEA*, Rio de Janeiro, v.7, n.2, p.305-322, July/Dec. 2005.
- RODRIGUES FILHO, J. *Estudos críticos em administração no Brasil: classificação da produção de conhecimento sob a ótica da teoria crítica de Jurgen Habermas*. 2002. Tese (Concurso Professor Titular) – Universidade Federal da Paraíba, João Pessoa, Paraíba, Brasil, 2002.
- ROESCH, S.; ANTUNES, E.; SILVA, L.V. Tendências da pesquisa em recursos humanos e organizações – uma análise das dissertações de mestrado. In: ENCONTRO DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 21., 1997, Rio das Pedras. *Anais...* Rio de Janeiro: Anpad, 1997.
- SCHUMACKER, R.E.; LOMAX, R.G. *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates, 1996.
- SELLTIZ, C. *Métodos de pesquisa nas relações sociais*. São Paulo: Herder, 1974.
- SHARMA, S. *Applied multivariate techniques*. New York: Wiley, 1996.
- STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES (SPSS). *Regression models 12.0*. Chicago: SPSS, 2003.
- STEINER, M.T.A. *Uma metodologia para o reconhecimento de padrões multivariados com resposta dicotômica*. 1995. Tese (Doutorado em Engenharia) – Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brasil, 1995. Disponível em: <www.eps.ufsc.br/teses/steiner/capit_2/cap2_ste.htm>. Acesso em: 28 out. 2007.
- STEVENS, J. *Applied multivariate for the social sciences*. 3rded. Mahwah, NJ: Lawrence Erlbaum, 1996.
- TABACHNICK, B.G.; FIDELL, L.S. *Using multivariate statistics*. 3rded. New York: Harper Collins, 1996.
- TONELLI, M.; CALDAS, M.P.; LACOMBE, B.M.B.; TINOCO, T. Produção acadêmica em recursos humanos no Brasil: 1991-2000. *Revista de Administração de Empresas (RAE)*, São Paulo, v.43, n.1, p.105-122, jan./mar. 2003.
- TORRES, R.R. *Estudo sobre os planos amostrais das dissertações e teses em administração da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo e da Escola de Administração da Universidade Federal do Rio Grande do Sul: uma contribuição crítica*.

REFERÊNCIAS

2000. Dissertação (Mestrado) – Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, São Paulo, São Paulo, Brasil, 2000.
- TOWNSEND, J.T.; ASHBY, F.G.; Measurement scales and statistics: the misconception misconceived. *Psychological Bulletin*, v.96, p.394-402, Jan. 1984.
- VELLEMAN, P.F.; WILKINSON, L. Nominal, ordinal, interval and ratio typologies are misleading. *The American Statistician*, v.47, p.65-72, Jan. 1993.
- VERGARA, S.C.; CARVALHO JR., D.S. Nacionalidade dos autores referenciados na literatura brasileira sobre organizações. In: ENCONTRO DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 19., 1995, João Pessoa. *Anais... Paraíba: Anpad*, 1995.
- VIEIRA, G.D. Por quem os sinos doam? Uma análise da publicação científica na área de marketing do EnANPAD. In: ENCONTRO DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 22., 1998, Foz do Iguaçu. *Anais... Paraná: Anpad*, 1998.
- YU, C.H.; OHLUND, S.; DiGANGI, S.; JANNASCH-PENELL, A. Incoherence and the parametric test framework: misconceived relationships among sample, sampling distribution, and population. *American Statistical Association 1999 Proceedings of the Section on Statistical Education*, Arizona, 2000. p.225-230.

ABSTRACT

Evaluating the appropriateness of using multivariate techniques in studies of consumer behavior in the theses and dissertations of two higher education institutions

This paper is part of a large study to assess the appropriateness of using multivariate statistical techniques in the theses and dissertations of two institutions of higher education in the field of marketing and as regards consumer behavior, from 1997 to 2006. This paper focuses on 11 multivariate techniques (regression analysis, discriminant analysis, logistic regression analysis, canonical correlation, multivariate analysis of variance, conjoint analysis, structural equation modeling, factor analysis, cluster analysis, correspondence analysis, and multidimensional scaling) that show great potential for use in marketing studies. The objective of this study was to analyze whether applying these techniques suits the needs of the research problem at the core of these theses and dissertations, as well as to evaluate to what extent they fulfill their premises. Overall, the results suggest that researchers should be more committed to checking all the theoretical precepts that underlie the application of these multivariate techniques.

Keywords: multivariate analysis, appropriateness of using statistical techniques, consumer behavior.

RESUMEN

Evaluación de la conveniencia del uso de técnicas multivariadas en estudios de comportamiento del consumidor en tesis y disertaciones de dos instituciones de educación superior

Este artículo forma parte de un amplio estudio para evaluar la adecuación del uso de técnicas estadísticas multivariadas en tesis y disertaciones de dos instituciones de educación superior en el área de *marketing*, en el tema comportamiento del consumidor, entre 1997 y 2006. Se enfocan once técnicas multivariadas (análisis de regresión, análisis discriminante, análisis de regresión logística, correlación canónica, análisis multivariado de varianza, análisis conjunto, modelado de ecuaciones estructurales, análisis factorial, análisis de conglomerados, análisis de correspondencia, escalamiento multidimensional) que han presentado gran potencial de uso en estudios de mercado. El objetivo de este estudio es analizar si el empleo de dichas técnicas es adecuado a las necesidades de los problemas de investigación presentados en las tesis y disertaciones, así como evaluar el nivel de atención a sus supuestos. En general, los resultados sugieren la necesidad de un mayor compromiso por parte de los investigadores en la verificación de todos los preceptos teóricos de aplicación de las técnicas multivariadas.

Palabras clave: análisis multivariado, adecuación del uso de técnicas estadísticas, comportamiento del consumidor.