

Presidente da República
Luíz Inácio Lula da Silva

Ministro do Planejamento, Orçamento e Gestão
Paulo Bernardo Silva

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE

Presidente
Eduardo Pereira Nunes

Diretor-Executivo
Sérgio da Costa Côrtes

ÓRGÃOS ESPECÍFICOS SINGULARES

Diretoria de Pesquisas
Wasmália Socorro Barata Bivar

Diretoria de Geociências
Guido Gelli

Diretoria de Informática
Luiz Fernando Pinto Mariano

Centro de Documentação e Disseminação de Informações
David Wu Tai

Escola Nacional de Ciências Estatísticas
Sérgio da Costa Côrtes (interino)

Ministério do Planejamento, Orçamento e Gestão
Instituto Brasileiro de Geografia e Estatística - IBGE

REVISTA BRASILEIRA DE ESTATÍSTICA

volume 67 número 227 julho/dezembro 2006

ISSN 0034-7175

R. bras. Estat., Rio de Janeiro, v. 67, n. 227, p. 1-125, jul./dez. 2006

Instituto Brasileiro de Geografia e Estatística - IBGE

Av. Franklin Roosevelt, 166 - Centro - 20021-120 - Rio de Janeiro - RJ - Brasil

© IBGE. 2007

Revista Brasileira de Estatística, ISSN 0034-7175

Órgão oficial do IBGE e da Associação Brasileira de Estatística – ABE.

Publicação semestral que se destina a promover e ampliar o uso de métodos estatísticos (quantitativos) na área das ciências econômicas e sociais, através de divulgação de artigos inéditos.

Temas abordando aspectos do desenvolvimento metodológico serão aceitos, desde que relevantes para os órgãos produtores de estatísticas.

Os originais para publicação deverão ser submetidos em três vias (que não serão devolvidas) para:

Francisco Louzada-Neto
Editor responsável – RBEs – IBGE.
Av. República do Chile, 500 – Centro
20031-170 – Rio de Janeiro, RJ.

Os artigos submetidos às RBEs não devem ter sido publicados ou estar sendo considerados para publicação em outros periódicos.

A Revista não se responsabiliza pelos conceitos emitidos em matéria assinada.

Editor Responsável

Francisco Louzada-Neto (UFSCAR)

Editor de Estatísticas Oficiais

Denise Britz do Nascimento Silva (GAB/IBGE)

Editor de Metodologia

Enrico Antonio Colosimo (UFMG)

Editores Associados

Gilberto Alvarenga Paula (USP)

Dalton Francisco de Andrade (UFSC)
Ismeria Blavatsky de Magalhães (DPE/IBGE)
Helio dos Santos Migon (UFRJ)
Francisco Cribari-Neto (UFPE)

Editoração

Helem Ortega da Silva - Coordenação de Métodos e Qualidade - DPE/COMEQ/IBGE

Impressão

Gráfica Digital/Centro de Documentação e Disseminação de Informações - CDDI/IBGE, em 2004.

Capa

Renato J. Aguiar – Coordenação de Marketing/CDDI/IBGE

Ilustração da Capa

Marcos Balster – Coordenação de Marketing/CDDI/IBGE

Revista brasileira de estatística/IBGE, - v.1, n.1 (jan./mar.1940), - Rio de Janeiro:IBGE, 1940-

v.

Trimestral (1940-1986), semestral (1987-).

Continuação de: Revista de economia e estatística.

Índices acumulados de autor e assunto publicados no v.43 (1940-1979) e v. 50 (1980-1989).

Co-edição com a Associação Brasileira de Estatística a partir do v.58.

ISSN 0034-7175 = Revista brasileira de estatística.

I. Estatística – Periódicos. I. IBGE. II. Associação Brasileira de Estatística.

IBGE. CDDI. Div. de Biblioteca e Acervos Especiais

CDU 31 (05)

RJ-IBGE/88-05 (rev.98)

PERIÓDICO

Impresso no Brasil/Printed in Brazil

Sumário

Nota do Editor.....5

Artigos

Previsão da arrecadação do ICMS no Brasil através de Redes Neurais7

Juan Camilo Santana Contreras

Francisco Cribari-Neto

Diagnóstico em análise discriminante.....45

Sueli Maria Beltrame Reigada

Silvia Nagib Elian

Intervalos de confiança Bootstrap baseados na amostragem por conjuntos ordenados para a variância populacional de uma distribuição normal.....75

Clécia Aparecida Garcia

Maria Cecília Mendes Barreto

Estimação de parâmetros no ajuste de modelos teóricos a semivariogramas experimentais97

Miguel Angel Uribe-Opazo

Célia Rupp Kavanagh

Política editorial.....123

Nota do Editor

Neste segundo volume da RBEs do ano de 2006 estamos publicando mais quatro artigos. O primeiro artigo, de autoria de Juan Camilo Santana Contreras e Francisco Cribari-Neto, avalia a utilidade de métodos de previsão baseados em redes neurais na previsão de valores futuros de arrecadações tributárias. O segundo artigo, de autoria de Sueli Maria Beltrame Reigada e Silvia Nagib Elian, apresenta medidas de diagnóstico em análise discriminante em duas situações, para duas populações com matriz de covariância constante e para duas populações com matrizes de covariância desiguais. O terceiro artigo, de autoria de Clécia Aparecida Garcia e Maria Cecilia Mendes Barreto, propõe intervalos de confiança *Bootstrap* para a variância populacional de uma distribuição normal baseados na amostragem por conjuntos ordenados. O quarto artigo, de autoria de Miguel Angel Uribe-Opazo e Célia Rupp Kavanagh, estuda o ajuste de modelos teóricos a semivariogramas experimentais, importante etapa na análise geostatística de dados espacialmente referenciados.

Gostaria de aproveitar a oportunidade para agradecer a grande colaboração de todos os Editores Associados e revisores do periódico, bem como à equipe do IBGE.

Uma excelente leitura.

Francisco Louzada-Neto
Editor Responsável

Previsão da arrecadação do ICMS no Brasil através de Redes Neurais ¹

Juan Camilo Santana Contreras*
Francisco Cribari-Neto**

Resumo

A produção de previsões de valores futuros de arrecadações tributárias é a temática principal deste artigo. Em particular, procura-se avaliar a utilidade de métodos de previsão baseados em redes neurais. As séries do ICMS nacional e de três estados da união (São Paulo, Rio de Janeiro e Pernambuco) serão prognosticados e avaliados através das metodologias de redes neurais, SARIMA e alisamento exponencial de Holt-Winters. Estas séries financeiras freqüentemente exibem *outliers*, fazendo necessário sua correção em uma fase preliminar ao modelamento; as redes neurais podem resultar afetadas por este evento, motivo pelo qual a identificação e correção de *outliers* é incluída na análise. A incorporação das componentes observáveis como a tendência e sazonalidade são examinadas na avaliação da capacidade preditiva das redes neurais. Igualmente, a combinação de previsões é aplicada neste sentido. Os resultados revelam que as redes neurais proporcionam melhores previsões comparados com as metodologias tradicionais; além disso, a inclusão das componentes não observáveis mostram-se eficientes no desempenho preditivo das redes neurais, entanto que a combinação de previsões que incluem redes neurais tendem a alcançar maior precisão preditiva do que combinações que não a incluem.

¹ Agradecimentos à CNPq pelo apoio financeiro no desenvolvimento deste artigo.

* Endereços para correspondência: Dept^o de Estatística da Universidade Federal de Pernambuco - Recife/PE, 50740-540, e-mail: jcsc161534@cox.de.ufpe.br.

** Dept^o de Estatística da Univ. Federal de Pernambuco - Recife/PE, 50740-540, e-mail: cribari@de.ufpe.br.

1. Introdução

As três esferas de governo - união, estados e municípios - possuem competência impositiva para cobrar tributos. O *ICMS*, Imposto sobre Operações Relativas à Circulação de Mercadorias e sobre Prestações de Serviços de Transporte Interestadual e Intermunicipal e de Comunicação, compete principalmente aos estados e é um dos impostos mais importantes, juntamente com o Imposto de Renda - IR e com o Imposto sobre Produtos Industrializados - *IPI*. O conhecimento sobre o padrão evolutivo das arrecadações desses tributos, dentro do contexto da atividade econômica brasileira, é útil para o planejamento e para a tomada de decisões por parte dos governantes e dos formuladores de políticas econômicas. Assim, a geração de previsões sobre arrecadação de impostos desempenha papel importante dentro deste contexto.

Uma estratégia alternativa de previsão utilizando redes neurais será considerada com vistas à elaboração de prognósticos relativos à arrecadação do *ICMS* agregado em nível nacional, como também, relativas às arrecadações do *ICMS* em São Paulo, Rio de Janeiro e Pernambuco. Estas serão contrastadas com as previsões proporcionadas pelas metodologias de Box e Jenkins, alisamento exponencial e combinação de previsões.

As investigações dos últimos anos têm mostrado um desempenho satisfatório das redes neurais na detecção de padrões, na classificação, como também na previsão através da aprendizagem baseada na experiência. As redes neurais são classificadas estatisticamente como um método não-paramétrico de fácil aplicabilidade, que se tem exibido superior a várias metodologias estatísticas da análise univariado em diferentes casos na prática; por exemplo, a competição de *Santa fé* sob modelos de prognóstico deixaram as redes neurais como os melhores métodos, veja Casdagli & Eubank(1992) e Weigend & Gershenfeld(1993).

Uma das maiores áreas de aplicação das redes neurais é a previsão, veja Sharda(1994), daí que é uma ferramenta atrativa pelos pesquisadores em comparação com as metodologias tradicionais baseadas em modelos. As redes neurais tentam emular o comportamento biológico do cérebro. Como sabemos, o cérebro humano é um conjunto completo de interconexões de elementos simples chamados neurônios. Cada

neurônio recebe um sinal de entrada proveniente de outros neurônios através de estímulos externos. Localmente, o neurônio processa a informação recebida através de uma função de ativação ou transferência que produz um sinal de saída, transformado que vai até outros neurônios ou como resposta, resultado de um estímulo. Enquanto cada neurônio individualmente não proporciona informação útil, em conjunto, pode fazer um surpreendente número de tarefas de forma eficiente. Esta característica faz das redes neurais um mecanismo poderoso computacionalmente que aprende através de exemplos e depois generaliza para casos que não têm sido vistos.

Ainda que diferentes arquiteturas de redes neurais têm sido propostas na literatura, Haykin(1994), nosso interesse é descrever especialmente a rede *perceptron multicamadas (MLP)*. Uma rede *MLP* está composta por várias camadas de neurônios. Os neurônios da primeira camada estão encarregados de receber a informação do exterior, entanto que a última camada proporciona a resposta associada a esta informação. Entre as duas camadas podem ter inumeráveis camadas e neurônios. Além disso, os neurônios das camadas adjacentes estão completamente conectados. A Figura 1 exibe uma *MLP* com todas as conexões entre os neurônios de cada camada.

Diferentes autores, em geral, Hornik, Stinchcombe & White(1989), Cybenko(1989), mostraram que o *perceptron multicamadas* é um aproximador universal, além de ser na atualidade uma das arquiteturas mais utilizadas na solução de problemas, devido ao seu fácil uso e aplicabilidade; veja Cohen, Franco, Morgan, Rumelhart & Abrash(1993), Narendra & Parthasaranty(1990) e Wiegand, Huberman & Rumelhart(1990).

Num problema de previsão de séries temporais com redes neurais, as entradas da rede são variáveis defasadas da original, entanto que a saída da rede representa seu valor futuro, assemelhando-se a um modelo auto regressivo. No processo de obter o valor futuro de interesse, primeiramente a rede deve ser treinada com a finalidade de obter os pesos associados à rede. O conhecimento aprendido pela rede é armazenado em cada vínculo que representa a conexão entre os neurônios. É através destas conexões que a rede pode captar as dinâmicas não lineares da série temporal. O treinamento da rede *MLP* é supervisionado, caso no qual a resposta desejada para cada padrão de entrada é sempre disponível.

A informação de treinamento é ingressada na rede como vetores de variáveis ou padrões de entrada. Cada elemento do vetor de entrada é associado com cada neurônio na camada de entrada, daí que o número de entradas à rede é igual à dimensão do vetor da entrada. Na previsão de séries temporais, o número de entradas é difícil de estabelecer, no entanto, uma quantidade fixa de variáveis defasadas são freqüentemente estabelecidas na análise, Zhang, Patuwo & Hu(1998). O conjunto de dados disponível é geralmente dividido em um conjunto de treinamento e um conjunto de validação. O conjunto de treinamento é utilizado na estimação dos pesos, enquanto que o conjunto de validação é utilizado para avaliar a capacidade de generalização da rede. O algoritmo conhecido como a regra delta generalizada é utilizado para encontrar os pesos que minimizem uma medida de erro global, como, por exemplo, a soma de quadrados do erro, Rumelhart, Hilton & Williams(1986a).

A arquitetura de uma rede neural concentra-se em definir os seguintes elementos:

- i) o número de camadas de entrada;
- ii) o número de camadas ocultas e neurônios por camada; e
- iii) o número de neurônios da camada de saída.

A seleção destes parâmetros é um problema dependente. Ainda que, existem muitos métodos para encontrar a melhor arquitetura, citando, por exemplo, o *pruning algorithm*, Sietsma & Dow(1988); a decomposição canônica, Wang, Massino, Tham & Morris(1994) e os critérios de informação, Murata, Yoshizama & Amari(1994); nenhum pode garantir uma solução ótima para qualquer problema de previsão. Freqüentemente os pesquisadores determinam a melhor arquitetura através da experimentação, motivo pelo qual, o trabalho com redes neurais é considerado mais uma arte que uma ciência.

Os neurônios e camadas ocultas permitem à rede detectar todas as características não lineares associada com os dados. Vários pesquisadores, entre eles, Zhang(1994), Cybenko(1988), Lippmann(1987), têm coincidido que as redes neurais com no máximo duas camadas ocultas são suficientes para resolver qualquer problema relacionado com a previsão. Determinar o número de neurônios ocultos é uma tarefa ainda mais complexa; geralmente, as redes com poucos neurônios são preferidas devido àquelas que proporcionam boa capacidade de generalização e menos problemas de sobreajuste,

no entanto, uma quantidade reduzida de neurônios poderia afetar o processo da modelagem e aprendizagem da rede. Até o momento não têm regras teóricas para selecionar este número, ainda que existam propostas sob alguns procedimentos sistemáticos, Gorr & Szczypula(1994).

As funções de ativação das que falamos com anterioridade introduz alguma não linearidade, que é avaliada constantemente através da rede neural. Qualquer função não pode ser uma função de ativação, senão unicamente um conjunto de funções bem comportadas, limitadas, monótonas crescentes e diferenciáveis, são utilizadas. Entre as mais freqüentes temos:

- i) função logística (sigmoidal);
- ii) função tangente hiperbólica;
- iii) função seno e coseno; e
- iv) função linear.

Entre estas funções, a mais utilizada é a sigmoidal, especialmente nas tarefas de classificação, entanto que a função tangente hiperbólica na previsão; Kaastra & Boyd(1996).

Em procura de melhorar o processo de otimização que permita obter o melhor conjunto de pesos, é comum entre os pesquisadores transformar os dados na entrada da rede antes de iniciar o treinamento. Entre as funções freqüentemente utilizadas estão: i) transformação linear; ii) standardização estatística; e iii) standardização simples, uma descrição detalhada sob estas funções é dada por Zhang et al.(1998).

Na área do prognóstico, a idéia da combinação de previsões individuais sob um mesmo evento tem sido amplamente pesquisada, aproximadamente meio século depois do reconhecido trabalho precursor de Bates & Granger(1969). O interesse na combinação de previsões é devido à redução na variância do erro de prognóstico em comparação com o erro de prognóstico das previsões individuais; além disso, se dois modelos proporcionam previsões viesadas é possível que através de alguma combinação este viesamento possa ser eliminado, obtendo-se uma estimacão melhorada. Pesquisas sob o efeito dos erros de estimacão paramétrica na avaliação dos procedimentos de previsão têm sido muito estudados na prática, Marcellino(2002), Guidolin & Na(2006),

como também na área teórica, Hendry & Clements(2004), Smith & Wallis (2005).

Outro tópico de interesse no modelamento de séries temporais, faz referência à correção de *outliers*, fenômeno freqüente nas séries econômicas, cuja não correção poderia ter efeitos no modelamento e compressão do processo subjacente gerador dos dados, mesmo que na previsão, Peña(2001). Estas conseqüências podem ter-se, também, na modelagem com redes neurais, Pernía-Espinoza,Ordieres,de Pinsón & Marcos(2005). Uma vez tendo sido corrigidos os *outliers*, a extração de sinais (componentes de tendência e sazonalidade), como é descrito por Gómez & Maravall(2001b), são importantes para entender o comportamento presente da série e ajustar prognósticos.

O presente artigo encontra-se organizado da seguinte forma. Na seção 2, conceitos gerais sobre as metodologias de alisamento exponencial, Box e Jenkins e combinação de previsões são considerados. Além disso, são feitas referências, relacionadas com a decomposição das séries em as componentes não-observáveis, também como a identificação e correção de *outliers*. Na seção 3, as principais características da rede *neural perceptron multicamadas* são discutidas. Na seção 4, previsões sobre arrecadações do Imposto sobre Operações Relativas à Circulação de Mercadorias e sobre Prestações de Serviços de Transporte Interestadual e Intermunicipal e de Comunicação agregado do Brasil e de três estados da união são produzidos e avaliados. Comparações com as metodologias clássicas e combinação de previsões são realizadas. Por último, na seção 5, fornece-se um resumo das conclusões.

2. Estratégias Usuais de Previsão

2.1. Alisamento exponencial de Holt-Winters

Através desta estratégia, objetiva-se extrair da série algum tipo de padrão, que constituirá o elemento principal para explicar o comportamento histórico e para prever o comportamento futuro da série. Para detalhes sobre algoritmos de alisamento exponencial, Morettin & Toloï(2004) e Montgomery & Johnson(1976).

Se Z_t é uma série sazonal de período s , esta pode ser modelada de forma aditiva como uma decomposição em quatro fatores, sendo, F_t e T_t as componentes que descrevem a sazonalidade e a tendência, respectivamente, entanto que μ_t é o nível médio da série e a_t o erro no instante t ; isto é,

$$Z_t = \mu_t + T_t + F_t + a_t, t = 1, 2, \dots, N. \quad (1)$$

As três equações de suavização são dadas por

$$\begin{aligned} \hat{F}_t &= \gamma(Z_t - \bar{Z}_t) + (1-\gamma)\hat{F}_{t-s}, \quad 0 < \gamma < 1, \\ \bar{Z}_t &= \alpha(Z_t - \hat{F}_{t-s}) + (1-\alpha)(\bar{Z}_{t-1} + \hat{T}_{t-1}), \quad 0 < \alpha < 1, \\ \hat{T}_t &= \beta(\bar{Z}_t - \bar{Z}_{t-1}) + (1-\beta)\hat{T}_{t-1}, \quad 0 < \beta < 1, \end{aligned} \quad (2)$$

com $t = s+1, \dots, N$, sendo α , β e γ as constantes de suavização. Em geral, a determinação das constantes é feita minimizando a soma dos quadrados dos erros de previsão um passo à frente; Granger & Newbold(1977).

2.2. Modelo SARIMA

Se $\{Z_t\}$ é o processo original que não é estacionário devido a uma tendência não-determinística e possui estrutura sazonal, é muitas vezes possível obter um processo estacionário $\{W_t\}$, fazendo $W_t = \Delta^d \Delta_s^D Z_t$, em que $\Delta_s = (1-B^s)$ e $\Delta_s^D = (1-B^s)^D$. Este processo estacionário pode ser modelado através do que é chamado de modelo autorregressivo integrado de médias móveis sazonal, denotado por SARIMA $(p, d, q) \times (P, D, Q)_s$ e definido como

$$\phi(B)\Phi(B^s)\Delta^d \Delta_s^D Z_t = \theta(B)\Theta(B^s)a_t, \quad (3)$$

em que d , D e s representam o número de diferenças simples e sazonais feitas sobre Z_t e o período de sazonalidade, respectivamente. Os polinômios associados à componente sazonal em (3) são dados por

$$\begin{aligned}\Phi(B^s) &= 1 - \Phi_1 B^s - \dots - \Phi_p B^{sp}, \\ \Theta(B^s) &= 1 - \Theta_1 B^s - \dots - \Theta_q B^{sq}.\end{aligned}\tag{4}$$

Os demais polinômios são os definidos por

$$\begin{aligned}\phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q.\end{aligned}\tag{5}$$

O processo de identificação, estimação e diagnóstico associado com a modelagem de séries temporais encontra-se descrito em Box & Jenkins (1994), Box, Jenkins & Reisen (1994) e Morettin & Toloi (2004), entre outros.

2.3. Correção por *outliers*

Os valores de uma série temporal podem ser afetados por eventos externos e atípicos, que podem chegar a ter efeito notável sobre a identificação e previsão do modelo, motivo pelo qual é muitas vezes preciso utilizar uma metodologia que permita identificá-los. As observações que são inconsistentes com o comportamento geral da série histórica são denominadas *valores atípicos* ou *outliers*. Para controlar o efeito que essas observações espúrias possam exercer sobre o modelo, utiliza-se a análise de intervenção; para uma discussão mais detalhada sobre o tema Guerrero (1991) e Morettin & Toloi (2004). Note que há uma distinção entre variáveis de intervenção e *outliers*; Box & Tiao (1975).

Quatro tipos de *outliers* têm sido considerados na literatura estatística, Lung & Box (1986), Chen & Liu (1993), Peña (2001): atípico aditivo (*AO - additive outlier*), atípico de inovação (*IO - innovation outlier*), mudança de nível (*LS - level shift*) e mudança transitória (*TC - transitory change*). Assumindo que a série observada tem k *outliers*,

seu efeito combinado pode ser expresso como

$$Z_t^* = \sum_{j=1}^k \xi_j(B) \omega_j I_t^{(\tau_j)} + Z_t,$$

em que Z_t^* é a série contaminada, Z_t a série que segue, supondo por simplicidade o modelo (3) quando não existe estrutura sazonal. Aqui, ω_j representa o impacto inicial do *outlier* no tempo $t = \tau_j$, $I_t^{(\tau_j)}$ é uma variável *dummy* que é igual a 1 se $t = \tau_j$ e igual a 0 caso contrário, e $\xi_j(B)$ determina a dinâmica do *outlier* que acontece em $t = \tau_j$ de acordo com os seguintes esquemas:

$$\begin{aligned} AO: \xi_j(B) &= 1, \\ LS: \xi_j(B) &= 1/(1-B), \\ TS: \xi_j(B) &= 1/(1-\delta B), 0 < \delta < 1, \\ IO: \xi_j(B) &= \hat{\theta}(B)/\hat{\phi}(B). \end{aligned} \tag{6}$$

Estes quatro tipos de *outliers* afetam a série observada de maneiras diferentes; Kaiser & Maravall(2001). Chen & Liu(1993) sugerem uma aproximação para detecção e correção automática de *outliers* que conduz a procedimentos confiáveis e eficientes; Gomez & Maravall(2001a).

2.4. Extração das componentes não-observáveis

Suponha que a série de interesse foi modificada através de um esquema de pré-processamento de forma a minorar os efeitos exercidos pelos *outliers* e suponha que há interesse em componentes não-observáveis, em geral, tendência e sazonalidade. A suposição preliminar é que o sinal de interesse pode ser extraído de Z_t de forma aditiva:

$$Z_t = s_t + n_t,$$

em que n_t representa a componente de não-sinais e s_t é o sinal de interesse; se o sinal for a série ajustada sazonalmente, então n_t será a componente sazonal, ou se o sinal for a tendência, um ruído branco ou componente transitório pode ser incluído em n_t .

Além disso, as duas componentes, s_t e n_t , seguem os seguintes processos estocásticos lineares:

$$\begin{aligned}\phi_s(B)s_t &= \theta_s(B)a_{st}, \\ \phi_n(B)n_t &= \theta_n(B)a_{nt}.\end{aligned}\tag{7}$$

Com relação a (7) as seguintes suposições serão feitas:

- Os processos a_{st} e a_{nt} são mutuamente independentes, ruídos brancos de média zero e variâncias σ_s^2 e σ_n^2 , respectivamente;
- Os polinômios $\phi_s(B)$ e $\phi_n(B)$ são primos; e
- Os polinômios $\theta_s(B)$ e $\theta_n(B)$ não têm raízes unitárias, simultaneamente. Esta suposição estabelece uma condição suficiente de invertibilidade sobre a série Z_t .

Assumindo o modelo ARIMA, simplificando o definido em (3) para a série observada Z_t , os polinômios das partes auto-regressiva e de médias móveis (invertível) podem ser expressos em função de (7) como

$$\begin{aligned}\phi(B) &= \phi_s(B)\phi_n(B), \\ \theta(B)a_t &= \phi_n(B)\theta_s(B)a_{st} + \phi_s(B)\theta_n(B)a_{nt}.\end{aligned}$$

Tendo, então, o conjunto de observações Z_1, Z_2, \dots, Z_t , a idéia geral é obter estimadores de mínimo erro quadrático (MMSE) e previsões de s_t e n_t ; uma discussão sobre o tema pode ser encontrada em Maravall & Kaiser(2000). Neste sentido, a obtenção destas componentes será muito importante no processo da modelagem através de redes neurais.

2.5. Combinação de previsões

Tem sido sugerido na literatura que combinações lineares de duas ou mais previsões proporcionadas por diferentes metodologias podem apresentar melhores resultados em comparação com as previsões individuais; Barnard (1963) e Hendry & Clements (2004).

Se Z_t é a série a ser predita no instante t e f_1, f_2, \dots, f_k são as previsões individuais através de k diferentes metodologias, então a combinação das previsões no prognóstico da série no tempo t será dada por:

$$\sum_{j=1}^k \gamma_j f_j,$$

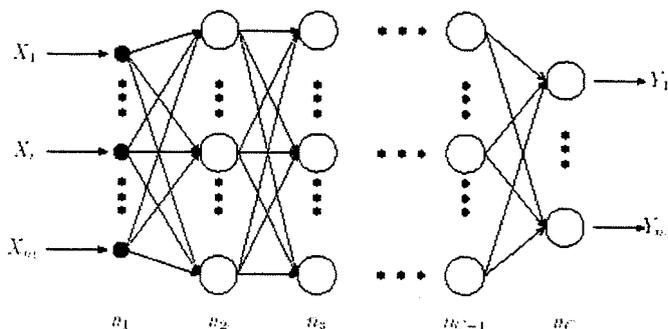
em que $0 \leq \gamma_j \leq 1$ e $\sum_{j=1}^k \gamma_j = 1$. Cada γ_j é definido como sendo inversamente proporcional ao erro de previsão fornecido pelo j -ésimo método e padronizado para cumprir as suposições estabelecidas.

3. Redes neurais artificiais

A arquitetura do *perceptron multicamadas* é caracterizada pelo fato de seus neurônios serem agrupados em camadas de diferentes níveis. Cada uma destas camadas é formada por um conjunto de neurônios. Há três tipos de camadas diferentes: a camada de entrada, as camadas ocultas e a camada de saída, como se observa na Figura 1. Os neurônios da camada de entrada se encarregam unicamente de receber sinais ou padrões que vêm do exterior e propagam tais sinais a todos os neurônios da camada seguinte. A última camada atua como saída da rede, proporcionando ao exterior a resposta da rede para cada um dos padrões de entrada. Os neurônios das camadas ocultas realizam um processamento não-linear dos padrões recebidos. Como pode ser observado na Figura 1, as conexões do *perceptron multicamadas* estão sempre dirigidas para frente, isto é, os neurônios de uma camada conectam-se com os neurônios da

camada seguinte e por isso recebem o nome de redes alimentadas para frente ou redes *feedforward*. Às conexões é associado um número real, chamado peso da conexão, e aos neurônios da rede um umbral ou bias que, no caso do *perceptron multicamadas*, é tratado como uma conexão adicional ao neurônio.

Figura 1 - Rede neural *feedforward*



3.1. Propagação dos padrões da entrada

Uma vez descrita a forma como a informação percorre as camadas e neurônios ao longo do *perceptron multicamadas*, apresentaremos, a seguir, as expressões para o cálculo das ativações dos neurônios da rede.

Considere um *perceptron multicamadas* com C camadas ($C-2$ camadas ocultas) e n_c neurônios na camada c , para $c=1,2,\dots,C$. Seja $W^c = (w_{ij}^c)$ a matriz de pesos associada às conexões da camada c à camada $c+1$, para $c=1,2,\dots,C-1$, em que w_{ij}^c representa o peso da conexão do neurônio i da camada c ao neurônio j da camada $c+1$; além disso, seja $U^c = (u_i^c)$ o vetor de umbrais dos neurônios da camada c para $c=2,\dots,C$. Denota-se por a_i^c a ativação do neurônio i da camada c ; estas ativações são calculadas da seguinte forma:

- Ativação dos neurônios da camada de entrada (a_i^1). Os neurônios da camada de entrada encarregam-se de transmitir para a rede os sinais recebidos do exterior;

Portanto,

$$a_i^1 = x_i, \quad (8)$$

para $i=1,2,\dots,n_1$, em que $X=(x_1,x_2,\dots,x_{n_1})$ representa o vetor ou padrão de entrada à rede.

• Ativação dos neurônios da camada oculta c (a_i^c). Os neurônios ocultos da rede processam a informação recebida, aplicando à função de ativação f a soma dos produtos das ativações que recebe pelos seus correspondentes pesos, isto é:

$$a_i^c = f\left(\sum_{j=1}^{n_{c-1}} w_{ij}^{c-1} a_j^{c-1} + u_i^c\right), \quad (9)$$

para $i=1,2,\dots,n_c$ e $c=2,3,\dots,C-1$.

• Ativação dos neurônios da camada de saída (a_i^C). Como no caso anterior, a ativação destes neurônios é dada pela função de ativação f aplicada sob a soma dos produtos das entradas que recebe pelos seus correspondentes pesos:

$$y_i = a_i^C = f\left(\sum_{j=1}^{n_{C-1}} w_{ij}^{C-1} a_j^{C-1} + u_i^C\right), \quad (10)$$

para $i=1,2,\dots,n_C$, em que $Y=(y_1,y_2,\dots,y_{n_C})$ é o vetor de saída da rede.

A função f é chamada *função de ativação*. Para o *perceptron multicamadas*, as funções de ativação mais utilizadas são a sigmóide e a tangente hiperbólica; no entanto, várias outras funções de ativação são utilizadas; Gately(1996). O propósito da função de ativação ou transferência é não permitir saídas com valores muito grandes, os quais podem *paralisar* a rede neural e inibir o treinamento.

3.2. Algoritmo de retropropagação

O algoritmo de aprendizado é o mecanismo mediante o qual se vão adaptando e modificando todos os parâmetros da rede. O problema de aprendizagem da rede é um problema de minimização da seguinte forma:

$$\min_W E,$$

sendo W o conjunto de parâmetros da rede (pesos e umbrals ou bias) e E uma função do erro que avalia a diferença entre a saída da rede e a saída desejada. Na maioria dos casos, a função de erro é definida como

$$E = \frac{1}{N} \sum_{n=1}^N e(n), \quad (11)$$

em que N é o número de observações ou padrões e $e(n)$ é o erro cometido pela rede para o n -ésimo padrão, que é dado por

$$e(n) = \frac{1}{2} \sum_{i=1}^{n_C} (s_i(n) - y_i(n))^2, \quad (12)$$

sendo $Y_n = (y_1(n), y_2(n), \dots, y_{n_C}(n))$ e $S_n = (s_1(n), s_2(n), \dots, s_{n_C}(n))$ os vetores de saída da rede e saída desejada para o n -ésimo padrão, respectivamente.

Desta forma, se W^* é um mínimo da função (11), isto é, o ponto onde o erro é o menor possível e a saída da rede encontra-se o mais próxima possível da desejada, obtém-se o fim do processo de aprendizado.

Para o *perceptron multicamadas*, o método de otimização não-linear mais utilizado é o *steepest descent* sobre a função E . Desta forma, cada parâmetro w da rede é modificado para cada padrão de entrada n de acordo com a seguinte lei de aprendizado:

$$w(n) = w(n-1) - \eta \frac{\partial e(n)}{\partial w}, \quad (13)$$

em que $e(n)$ é como definido em (12) e η é a taxa de aprendizado que influi na magnitude de deslocamento sobre a superfície do erro. O método do gradiente pode ser aplicado de forma eficiente, resultando no conhecido algoritmo de retropropagação ou *regra delta generalizada*; Rumelhart, Hilton & Williams(1986b).

3.3. A regra delta generalizada

Os resultados obtidos por Isasietal, relativos ao desenvolvimento matemático da regra delta generalizada, são apresentados a seguir; dois casos são considerados na obtenção dos pesos e umbrais da rede.

- Pesos da camada oculta $C-1$ à camada de saída C e umbrais da camada de saída.

Pesos:

$$w_{ji}^{C-1}(n) = w_{ji}^{C-1}(n-1) + \eta \delta_i^C(n) a_j^{C-1}(n). \quad (14)$$

Umbrais:

$$u_i^C(n) = u_i^C(n-1) + \eta \delta_i^C(n), \quad (15)$$

em que

$$\delta_i^C(n) = -(s_i(n) - y_i(n)) f' \left(\sum_{j=1}^{n_{C-1}} w_{ji}^{C-1} a_j^{C-1} + u_i^C \right) \quad (16)$$

e $f'(\cdot)$ é a primeira derivada da função de ativação, $j=1,2,\dots,n_{C-1}$, $i=1,2,\dots,n_C$;

- Pesos da camada c à camada $c+1$ e umbrais dos neurônios da camada $c+1$ para $c=1,2,\dots,C-2$;

Pesos:

$$w_{kj}^c(n) = w_{kj}^c(n-1) + \eta \delta_j^{c+1}(n) a_k^c(n). \quad (17)$$

Umbrais:

$$u_j^{c+1}(n) = u_j^{c+1}(n-1) + \eta \delta_j^{c+1}(n), \quad (18)$$

em que

$$\delta_j^{c+1}(n) = f' \left(\sum_{k=1}^{n_c} w_{kj}^c a_k^c + u_j^{c+1} \right) \sum_{i=1}^{n_{c+2}} \delta_i^{c+2}(n) w_{ji}^{c+1} \quad (19)$$

e $f'(\cdot)$ é a primeira derivada da função de ativação, $k=1,2,\dots,n_c$, $j=1,2,\dots,n_{c+1}$ e $c=1,2,\dots,C-2$.

A mudança em um peso, como pode ser observado em (13), é proporcional ao gradiente do erro, com a proporcionalidade sendo dada pelo parâmetro η . Valores altos da taxa de aprendizado, em princípio, poderiam favorecer uma convergência mais rápida, pois permitem avançar rapidamente na superfície do erro. No entanto, taxas de aprendizado altas podem ter conseqüências negativas sobre a aprendizagem, fazendo com que o método pule um mínimo ou oscile ao redor do mínimo. Valores pequenos das taxas de aprendizagem podem evitar estes problemas, ainda que possivelmente acarretem convergência mais lenta do algoritmo de aprendizagem, pois a magnitude de deslocamento na superfície do erro é menor.

Um método simples para evitar a instabilidade no algoritmo de aprendizagem devido à taxa de aprendizado é modificar (13) através da inclusão de um segundo termo, chamado *momento*, obtendo desta forma a seguinte lei:

$$w(n) = w(n-1) - \eta \frac{\partial e(n)}{\partial w} + \alpha \Delta w(n-1), \quad (20)$$

em que α é um número positivo ponderando a quantidade anterior. Esta regra foi proposta por Rumelhart et al.(1986a) e preserva as propriedades da regra definida em (13), no sentido que modifica os parâmetros da rede para minimizar a função do erro

(11). O novo termo, $\alpha\Delta w(n-1)$, incorpora ao método alguma inércia, fazendo com que a modificação atual do parâmetro dependa só da direção da modificação anterior, podendo evitar oscilações. Fazendo cálculos sucessivos sobre $\Delta w(n-1)$, Isasi & Galván(2004) apresentam uma outra forma de escrever a equação (20):

$$w(n) = w(n-1) - \eta \sum_{t=0}^n \alpha^{n-t} \frac{\partial e(t)}{\partial w}. \quad (21)$$

3.4. Processo de aprendizagem do *perceptron multicamadas*

Seja $\{(X(n), S(n))\}$ o conjunto de padrões que representam o problema a ser resolvido, em que $X(n) = (x_1(n), \dots, x_{n_1}(n))$ são os padrões de entrada à rede, $S(n) = (s_1(n), \dots, s_{n_1}(n))$ são as saídas desejadas para tais entradas e N é o número de padrões disponíveis. Frequentemente os padrões de entrada e saída são padronizados mediante uma transformação linear nos intervalos $[0,1]$ e $[-1,1]$, dependendo da função de ativação utilizada. O processo de aprendizagem do *perceptron multicamadas* é composto pelos seguintes passos:

- Os pesos e umbrais são inicializados ao acaso no intervalo $(0,1)$;
- O n -ésimo padrão $(X(n), S(n))$ é utilizado no treinamento da rede, em que o vetor $X(n)$ é propagado até a saída da rede utilizando as equações (8), (9) e (10), obtendo-se a resposta $Y(n)$;
- Avalia-se o erro quadrático médio para o n -ésimo padrão utilizando a equação (12).
- Utiliza-se a regra delta generalizada para modificar os pesos e umbrais da rede. Para tanto, são utilizadas as equações (14) a (19).
- Os passos 2, 3, 4 são repetidos para todos os padrões de treinamento, completando assim uma iteração ou ciclo de aprendizado. Este ciclo é chamado de *epoch*.
- Avalia-se o erro total E cometido pela rede, utilizando os padrões de treinamento.

- Os passos 2, 3, 4, 5 e 6 são repetidos até se obter o menor erro de treinamento, para o qual são realizados m épocas.

O processo de aprendizado do *perceptron* multicamadas deve ser finalizado quando $\frac{\partial E}{\partial w} \approx 0$, momento no qual os parâmetros da rede não mudam de forma perceptível entre iterações consecutivas. Duas formas de decidir a maneira como deve ser detido o treinamento são freqüentemente utilizadas; na primeira, o processo de treinamento é realizado até atingir um global mínimo; entanto que na segunda, é feito um processo treinamento-validação, no qual a partir de um conjunto de redes neurais pre-estabelecidas, o treinamento é calculado sob um número fixo de *épocas* e depois é avaliada a capacidade de generalização da rede, e o treinamento geral resumido. A generalização é uma propriedade que faz de um modelo baseado em uma amostra ter a capacidade de fazer previsões sob a população de interesse. O segundo esquema será utilizado na simulação das redes neurais na previsão do *ICMS*, para uma descrição mais detalhada sob estas duas escolas do pensamento Kaastra & Boyd(1996).

3.5. Capacidade de generalização

Na hora de avaliar o comportamento da rede e, em particular, do *perceptron* multicamadas, não é necessário saber apenas se a rede aprendeu com sucesso os padrões utilizados durante a aprendizagem, mas também conhecer o comportamento da rede frente a padrões que não foram utilizados durante o treinamento.

Para isso, é necessário dispor de dois conjuntos de padrões: um *conjunto de treinamento*, que treina e modifica os pesos e umbrais da rede, e um *conjunto de validação* ou *teste*, que mede a capacidade da rede para responder corretamente a padrões que não foram ingressados durante o treinamento. Quando a rede aproxima corretamente os padrões de aprendizagem, mas não responde bem aos padrões de validação, diz-se que houve subaprendizagem, o que pode ocorrer devido a vários fatores, como o uso de um número excessivo de neurônios ou camadas ocultas; Isasi & Galván(2004).

4. Avaliação relativa de previsões de arrecadação tributária

A análise de modelagem e previsão apresentada, a seguir, basear-se-á em séries temporais de arrecadação do Imposto sobre Circulação de Mercadorias e sobre Prestações de Serviços de Transporte Interestadual e Intermunicipal e de Comunicação agregado, e dos Estados de Pernambuco, Rio de Janeiro e São Paulo, os quais serão abreviados por *ICMS*, *ICMSPE*, *ICMSRJ* e *ICMSSP*, respectivamente. Os dados foram obtidos do banco de dados *IPEADATA* (www.ipeadata.gov.br), com atualizações em 1º de julho de 2005. As séries a serem analisadas foram deflacionadas pelo Índice de Preços ao Consumidor - *IPC*, ficando a preços de novembro de 2004. Os valores são expressos em milhões de reais. O período entre julho de 1994 e dezembro de 2004 é utilizado na modelagem e previsão de cada uma das séries temporais. Os logaritmos naturais das séries deflacionadas foram utilizados na análise.

Em uma primeira análise (análise I), utilizamos o período de julho de 1994 a novembro de 2004 para a modelagem de cada série, na obtenção da previsão do mês de dezembro de 2004; o período de janeiro a março de 2005 também sendo prognosticado. O interesse principal reside na obtenção de uma boa previsão do valor arrecadado em dezembro de 2004 e avaliar, posteriormente, as previsões para o período de janeiro a março de 2005. Depois, em uma segunda análise (análise II), o período entre julho de 1994 e junho de 2004 é utilizado na modelagem com vistas à previsão das observações entre julho e dezembro de 2004. Em cada uma das análises, deseja-se observar como cada modelo captura as estruturas dinâmicas das séries e as reflete nas previsões de cada um dos períodos já definidos.

Duas formas de medição do erro de previsão serão adotadas, para estabelecer as comparações das previsões realizadas por cada metodologia:

• **Erro percentual absoluto médio (MAPE).** Se Z_1, Z_2, \dots, Z_t são as observações da série de tempo e $Z_{t+1}, Z_{t+2}, \dots, Z_{t+h}$ são as previsões h passos à frente, então

$$MAPE(\%) = \frac{1}{h} \sum_{k=1}^h \frac{|e_{t+k}|}{Z_{t+k}} \times 100,$$

em que e'_{t+k} é o erro de previsão de Z_{t+k} , o valor da série no instante $t+k$.

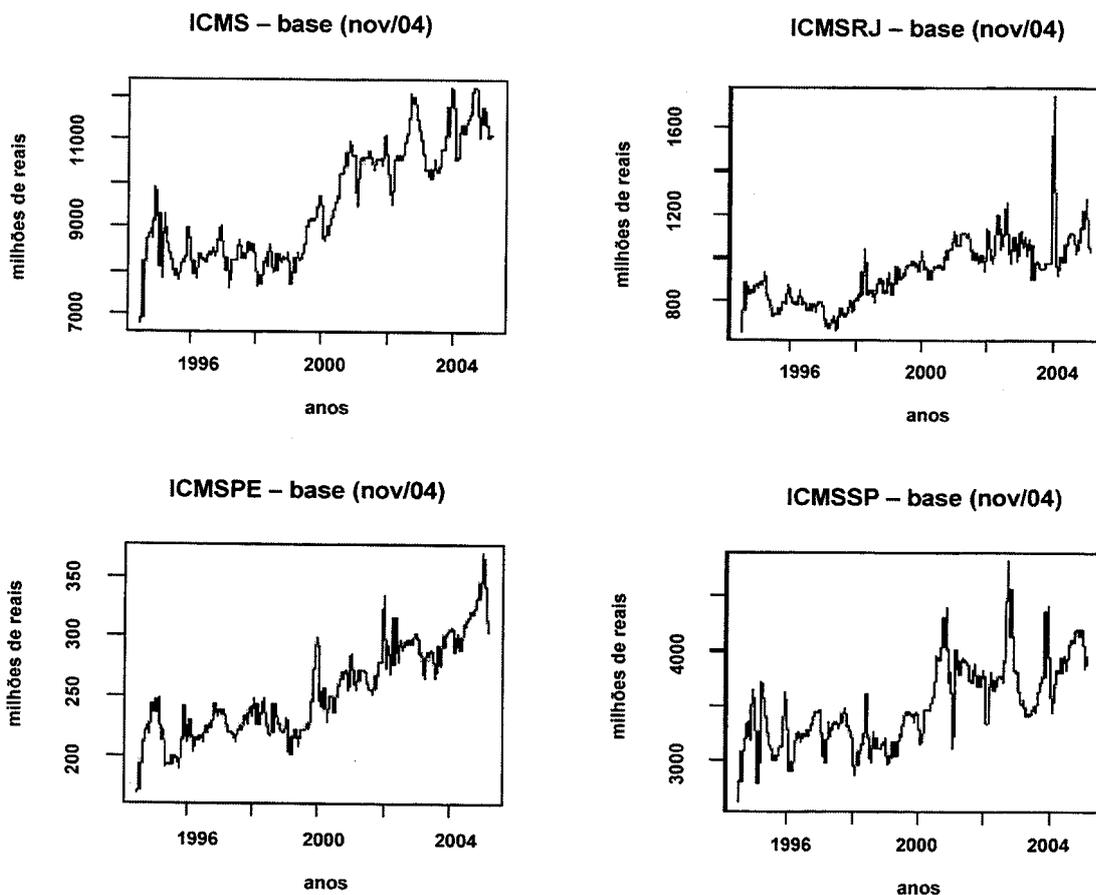
- **Desvio absoluto da mediana (MAD).** Aqui,

$$MAD = med(|e_{t+1} - med_e|, \dots, |e_{t+h} - med_e|),$$

em que e_{t+k} é o erro de previsão no instante $t+k$, $k=1,2,\dots,h$, e med_e é a mediana dos erros de previsão. Esta medida foi sugerida por Dijk(1999) e é robusta a *outliers*.

As quatro séries de arrecadação deflacionadas, para o período entre julho de 1994 e março de 2005, são apresentadas na Figura 2. Para cada uma das séries na Figura 2, é possível observar fortes indícios de não-estacionariedade.

Figura 2 - Séries de arrecadação nacional e de três estados da união



4.1. Modelagem e Previsão: Análise I

Os algoritmos aditivo e multiplicativo de Holt Winters são utilizados para obter, inicialmente, previsões referentes ao período que se estende de dezembro de 2004 a março de 2005. Os valores otimizados das constantes para cada uma das séries na escala logarítmica foram semelhantes para os modelos aditivo e multiplicativo; isto implica que as previsões obtidas através dos modelos aditivo e multiplicativo são semelhantes. Utilizaremos o modelo aditivo na geração das previsões.

Como é exibido na Tabela 1, o *MAPE* e o *MAD* são calculados para as previsões um passo (prev-o) e quatro passos (prev-m) à frente, para cada uma das séries. É possível observar através desta tabela que os erros de previsão quatro passos à frente proporcionados pelo alisamento exponencial, em geral, são menores do que os obtidos para um passo à frente. As séries corrigidas por *outliers* também foram utilizadas para gerar as previsões de dezembro de 2004 até março de 2005, porém não houve muita diferença com relação aos resultados exibidos na Tabela 1.

Tabela 1 - Medidas de erro das previsões de alisamento exponencial: um e quatro passos à frente

Séries	MAPE (%)		MAD	
	prev-o	prev-m	prev-o	prev-m
<i>ICMS</i>	2.2165	1.7816	243.4697	113.2117
<i>ICMSPE</i>	3.3489	3.3569	8.9322	8.3283
<i>ICMSRJ</i>	6.5078	5.4974	48.6304	44.9598
<i>ICMSSP</i>	3.9701	5.2460	85.4566	34.0477

Para o processo de identificação do modelo através da metodologia de Box e Jenkins, foi considerado o critério bayesiano de informação BIC, tendo sido escolhido o modelo que apresentou o menor BIC. Igualmente, correção por *outliers* é tipicamente importante na análise de séries temporais, como descrito por Box & Tiao(1975), Chen & Liu(1993), Kaiser & Maravall(2001), Maravall & Kaiser(2000), Peña(2001), entre outros. O programa *TRAMO-SEATS* foi utilizado no processo de correção de *outliers*.

O modelo $SARIMA(0,1,1) \times (0,1,1)_{12}$ foi escolhido para a série do *ICMS* uma vez que apresentou o menor critério de informação, $BIC = -6.6462$. No que tange à série do *ICMSPE*, o modelo $SARIMA(1,1,0) \times (0,1,1)_{12}$ com $BIC = -5.6916$ foi escolhido; para

ICMSRJ, o modelo selecionado foi o *ARIMA(0,1,1)*, com $BIC = -5.5901$. Por último, para o *ICMSSP* o modelo *SARIMA(0,1,1) × (0,1,1)₁₂*, com $BIC = -6.0643$, foi escolhido. Uma vez os modelos foram selecionados, análises de diagnóstico revelaram que não há evidência contra a hipótese de ausência de autocorrelação dos resíduos, como também, contra a hipótese de normalidade dos resíduos no nível de significância de 5%, excetuando o teste de normalidade feito sobre os resíduos do modelo de *ICMSSP*, onde não houve rejeição da hipótese nula no nível de significância de 1%.

Por último, a Tabela 2 exibe as medidas de erro de previsão, *MAPE* e *MAD*, para cada uma das séries na escala original. As previsões de *ICMS*, *ICMSPE* e *ICMSSP* apresentam valores do *MAPE* menores que 5% na previsão um e quatro passos à frente, enquanto a previsão de *ICMSRJ* exibe *MAPE* na previsão um passo à frente de 10.05%. O *MAD* das previsões um passo à frente de *ICMS* e de *ICMSSP* são menores que o *MAD* das previsões quatro passos à frente; no entanto, para o *ICMSRJ* e o *ICMSPE*, acontece o caso oposto.

Tabela 2 - Medidas de erro das previsões *SARIMA*: um e quatro passos à frente

Séries	MAPE (%)		MAD	
	prev-o	prev-m	prev-o	prev-m
<i>ICMS</i>	2.6824	3.0101	219.1853	315.8002
<i>ICMSPE</i>	4.8448	4.0248	12.6355	9.6259
<i>ICMSRJ</i>	10.0483	6.8847	104.0241	60.9193
<i>ICMSSP</i>	1.1906	2.2296	15.1958	34.3546

Na modelagem de cada uma das séries através de redes neurais, a proposta de Varfis & Versino(1990) foi utilizada. O procedimento é descrito a seguir. Primeiro, cada série transformada é reescalada no intervalo $[-1,1]$ antes de incluir variáveis defasadas como neurônios de entrada à rede. O objetivo principal é obter uma rede com bom ajuste e a melhor previsão possível. Além disso, duas camadas ocultas serão consideradas, dois parâmetros de aprendizagem, dois parâmetros de momento, 10 000 épocas e um máximo de seis neurônios por camada oculta. A função de transferência

para as camadas ocultas é a função hiperbólica anti-simétrica $f(x)=1.716\tanh\left(\frac{2}{3}x\right)$

descrita em Haykin, entanto que $f(x)=x$ é utilizada na camada da saída. Três classes de arquitetura de redes neurais são definidas:

- Rede 1. Somente uma camada oculta é considerada, com um número de até seis neurônios;
- Rede 2. Duas camadas ocultas são consideradas, com igual número de neurônios em cada, de até seis neurônios; e
- Rede 3. Duas camadas ocultas são consideradas, com um neurônio adicional na segunda camada, com cada camada tendo até seis neurônios.

Para esta primeira análise, o interesse recai sobre a previsão da arrecadação do mês de dezembro de 2004 fazendo uso do conjunto de treinamento (jul. 1994 – nov. 2004). Ainda que a previsão de dezembro de 2004 seja a prioridade nesta primeira parte, as previsões de janeiro até março de 2005 também serão avaliadas. A fim de designar a rede neural identificada, será utilizada a notação de Souza & Zandonade (1993); dada por $ANN(n_1, n_2, \dots, n_C)$, em que $n_c, c=1, 2, \dots, C$, é o número de neurônios na camada c .

Nesta primeira parte, as componentes de tendência e sazonalidade são utilizadas na simulação das redes neurais. A componente sazonal definida como doce variáveis *dummies* que representam cada mês será utilizada, devido às simulações prévias que apresentaram melhores ajustes quando estas variáveis foram incluídas nas entradas das redes neurais, do que quando foi utilizada a componente de sazonalidade extraída do modelo *SARIMA*. Os cenários das variáveis defasadas a serem incluídas na entrada da rede neural foram definidas segundo as características de cada série temporal e nosso parecer. Assim, o *ICMS* utiliza as defasagens de ordem 1; 1,2 e 1,2,3; o *ICMSPE* utiliza as defasagens 1,2; 1,2,12 e 1,2,3,12; o *ICMRJ* as defasagens 1; 1,2 e 1,2,3; por último, o *ICMSSP* utiliza as defasagens de ordem 1 e 1,2. A rede com o melhor ajuste e generalização é obtida.

Tabela 3 - Medidas de erro das previsões de *ANW* com tendência: um e quatro passos à frente

Séries	MAPE (%)		MAD	
	prev-o	prev-m	prev-o	prev-m
<i>ICMS</i>	2.6619	2.4225	210.8636	197.4972
<i>ICMSPE</i>	4.6070	3.6205	14.3465	8.8613
<i>ICMSRJ</i>	7.2851	6.9249	42.1014	58.2231
<i>ICMSSP</i>	1.7806	1.6440	51.2257	43.7689

Para o *ICMS*, tem-se uma rede neural com 16 neurônios de entrada, definida pelas três primeiras defasagens da variável, a tendência e 12 variáveis *dummy* que identificam cada mês do ano. A rede da classe 2, com cinco neurônios, $\eta=0.01$ e $\alpha=0.1$ é a que fornece a melhor previsão. Para o *ICMSPE*, a rede com 16 neurônios de entrada é definida pelas defasagens de primeira, segunda e décima segunda ordem, pela tendência e por 12 variáveis *dummy* como antes. Esta corresponde com a rede da classe 3, com dois neurônios, $\eta=0.01$ e $\alpha=0.5$. A rede neural com quatro neurônios de entrada, definida pelas primeiras três defasagens da série *ICMSRJ* e pela tendência, exibe os melhores resultados. Esta rede corresponde à classe 3, com dois neurônios, $\eta=0.01$ e $\alpha=0.5$. A rede neural do *ICMSSP* é identificada com 14 neurônios de entrada, determinados pela primeira defasagem, pela tendência e por 12 variáveis *dummy*, como antes; ele pertence à rede da classe 2, com um neurônio, $\eta=0.01$ e $\alpha=0.5$. A Tabela 3 exibe as medidas de erro da previsão, *MAPE* e *MAD*, para as previsões obtidas por redes neurais. As tendências foram obtidas através da decomposição do modelo *SARIMA*.

Em particular, a previsão do mês de dezembro de 2004 proporcionada pelas redes neurais exibe o menor erro relativo de previsão, excetuando possivelmente a obtida para o *ICMSPE* Tabela 4. Além disso, a medida de ajuste calculada sobre os resíduos do modelo estimado e associado às redes neurais é menor que a proporcionada pelo alisamento exponencial e pelo método de Box e Jenkins. Os resultados comparativos entre as três metodologias individuais, alisamento exponencial, *SARIMA* e redes neurais, exibe ao alisamento exponencial como o melhor na previsão um e quatro passos à frente do *ICMS*, *ICMSPE*, *ICMSRJ* segundo o *MAPE*; o modelo *SARIMA* tem resultado

aceitáveis para o *ICMSSP* na previsão um passo à frente, entanto que as redes neurais o faz na previsão múltiplos passos, segundo o *MAPE*. No que respeita ao *MAD*, a maioria dos melhores resultados estão associados com o alisamento exponencial.

Tabela 4 - *er(%)* da previsão de dezembro de 2004

Série	SARIMA	HW	ANN
<i>ICMS</i>	1.6909	0.3554	-0.0722
<i>ICMSPE</i>	-7.3896	-9.0362	-6.3962
<i>ICMSRJ</i>	-1.8382	-0.5905	-0.6203
<i>ICMSSP</i>	-1.2673	0.2169	-0.0284

Tendo escolhido as redes neurais sobre este fato (minimizar o erro de previsão do mês de dezembro de 2004), as previsões dos outros meses (janeiro até março de 2005) são melhor descritas, segundo o *MAPE*, pelo alisamento exponencial na previsão um e múltiplos passos à frente; no entanto, o *MAPE* das previsões um passo à frente do *ICMSSP* sobre o modelo *SARIMA* é menor que o proveniente de outros métodos. As previsões múltiplos passos fornecidas pela rede neural para o *ICMSSP* são mais precisas que as proporcionadas pelas outras metodologias. Por outro lado, o *MAD* das previsões múltiplos passos do alisamento exponencial para todas as séries é menor que os proporcionados pelos outros métodos. As previsões da rede neural apresentam os menores *MAD* para o *ICMS* e o *ICMSRJ* na previsão um passo à frente. O *MAD* das previsões um passo à frente proporcionadas pelo modelo *SARIMA* são menores para o *ICMSSP*. O *MAD* para as previsões um e quatro passos à frente, determinadas pelo alisamento exponencial do *ICMSPE* são as menores.

A média ponderada das previsões individuais proporcionadas pelo alisamento exponencial, pelo método de Box e Jenkins e por redes neurais será utilizada para se obter uma previsão melhorada. A Tabela 5 exhibe as previsões um e quatro passos à frente, considerando a combinação das previsões dos três métodos conjuntamente, e do qual são obtidas as melhores medidas de precisão, comparado com as outras possíveis combinações entre os três métodos.

Tabela 5 - Combinação de previsões *HW-SARIMA-ANN*: um e quatro passos à frente

Séries	MAPE (%)		MAD	
	prev-o	prev-m	prev-o	prev-m
<i>ICMS</i>	1.7714	1.7988	179.7358	121.9652
<i>ICMSPE</i>	3.1962	3.1784	10.4816	9.0475
<i>ICMSRJ</i>	3.5286	2.1943	21.4649	6.5727
<i>ICMSSP</i>	1.0777	1.9176	17.7254	45.8953

Da Tabela 5 é possível destacar a notável melhoria das previsões do *ICMSRJ* a partir da combinação de previsões. Comparado com as outras combinações possíveis, a combinação *HW-SARIMA-ANN* exibe os melhores resultados na previsão um passo e múltiplos passos à frente. Em comparação com as previsões individuais, a combinação *HW-SARIMA-ANN* melhora as previsões ainda mais, apresentando menores valores do *MAPE* para todas as séries na previsão um e quatro passos à frente, no entanto, os melhores resultados baseados no *MAD* são dados pelo modelo de alisamento exponencial e *SARIMA*, para o *ICMSPE* e o *ICMSRJ*, respectivamente. A combinação *HW-SARIMA* exibe resultados favoráveis unicamente para o *ICMSRJ* na previsão um passo à frente.

4.2. Modelagem e previsão: análise II

Nesta segunda parte, as previsões referentes ao período que se estende de julho a dezembro de 2004 são obtidas através de alisamento exponencial, da metodologia de Box e Jenkins e via redes neurais.

As formas aditiva e multiplicativa do algoritmo de Holt-Winters conduziram a medidas de precisão de previsões semelhantes. O método aditivo será o considerado nos resultados. A Tabela 6 fornece as medidas de precisão, *MAPE* e *MAD*, referentes às previsões de cada série.

Tabela 6 - Medidas de erro das previsões de alisamento exponencial: um e seis passos à frente

Séries	MAPE (%)		MAD	
	prev-o	prev-m	prev-o	prev-m
ICMS	4.2235	5.3950	269.3703	503.0721
ICMSPE	3.1478	6.1928	6.3404	3.5893
ICMSRJ	4.1621	7.0321	33.0275	12.2830
ICMSSP	2.5089	2.2243	91.9084	60.6565

Notamos na Tabela 6 que as previsões um passo à frente apresentam os menores *MAPE* para o *ICMS*, o *ICMSPE* e o *ICMSRJ*, enquanto o *MAPE* é menor nas previsões múltiplos passos para o *ICMSSP*. Com relação ao *MAD*, os menores desvios ocorrem nas previsões múltiplos passos das séries *ICMSPE*, *ICMSRJ* e *ICMSSP*; no entanto, a previsão um passo à frente para o *ICMS* exibe o menor desvio.

As séries deflacionadas são utilizadas para ajustar o modelo *SARIMA* e produzir as previsões um e seis passos à frente. As séries são transformadas através do logaritmo natural para evitar possíveis flutuações.

O modelo escolhido para o *ICMS* foi o *SARIMA*(2,1,0)×(0,1,1)₁₂, que apresentou o menor valor do critério de informação, BIC=−6.6831. O modelo identificado para o *ICMSPE* foi o *SARIMA*(1,1,0)×(0,1,1)₁₂ com BIC=−5.6648; o *MAPE* das previsões fornecidas por este modelo foi o menor. O modelo *SARIMA*(0,1,1)×(0,1,1)₁₂ foi identificado para o *ICMSSP*, com BIC=−6.0260. Por último, o melhor modelo identificado para o *ICMSRJ* foi o *ARIMA*(0,1,2), com BIC=−5.6041. Identificação, correção de *outliers* e estimação dos parâmetros do modelo *SARIMA*, para cada uma das séries, foi realizada utilizando o programa *TRAMO-SEATS*. Através da análise de diagnóstico, pode-se concluir no nível de significância de 5% que cada conjunto de resíduos é um ruído branco normal, excetuando os resíduos do *ICMSSP*, para os quais a hipótese de normalidade não é rejeitada no nível significância de 1%.

A Tabela 7 contém medidas de precisão sobre as previsões dos modelos *SARIMA*, identificados para cada série. Notamos da mesma tabela que as previsões um passo à frente apresentam o menor *MAPE* para o *ICMS*, o *ICMSPE* e o *ICMSRJ*, enquanto para o *ICMSSP* as previsões seis passos à frente exibem o menor *MAPE*. Com relação ao *MAD*,

os menores desvios são obtidos pelas previsões um passo à frente do *ICMS*, e múltiplos passos do *ICMSPE*, *ICMSRJ* e *ICMSSP*.

Tabela 7 - Medidas de erro das previsões *SARIMA*: um e seis passos à frente

Séries	MAPE (%)		MAD	
	prev-o	prev-m	prev-o	prev-m
<i>ICMS</i>	3.3665	4.3965	123.4046	383.1682
<i>ICMSPE</i>	3.0418	6.1477	9.8625	2.7045
<i>ICMSRJ</i>	5.3159	6.0694	50.5311	22.1103
<i>ICMSSP</i>	2.3915	1.6860	73.6681	34.9270

O processo de identificação das redes neurais foi realizado de forma semelhante à adotada na análise I. Neste caso, o conjunto de treinamento, para cada série, foi o período de julho de 1994 a junho de 2004 e o conjunto de teste foi o período de julho a dezembro de 2004. As três classes de arquiteturas de redes neurais, funções de ativação e transformações dos dados, são definidas como na análise I. Os valores dos parâmetros de aprendizagem e momento dados por 0.01, 0.1 e 0.1, 0.5, respectivamente, foram utilizados. Além disso, 10 000 épocas foram consideradas no treinamento de cada rede neural. Dois cenários são simulados neste análise, um inclui e outro exclui a componente de tendência. No primeiro, os cenários das variáveis defasadas utilizadas na simulação são: *ICMS* com as defasagens 1,2, 1,2,3, 1,2,3,4,5; *ICMSPE* com 1,2,12; *ICMSRJ* e *ICMSSP* com 1, 1,2, 1,2,3 e 1,2,3,4. No segundo cenário, *ICMS* utiliza as defasagens 1, 1,2, 1,2,3; *ICMSPE* utiliza 1, 1,2, 1,2,3, 1,2,3,4 e 1,2,12; entanto que *ICMSRJ* utiliza 1, 1,2, 1,2,3. As simulações das redes neurais com a componente de tendência do *ICMSSP* proporcionaram os melhores resultados na previsão, impossíveis de se melhorar ao momento de simular as redes do segundo cenário, motivo pelo qual não fazemos uma descrição deste cenário. A rede com o melhor ajuste e capacidade de generalização é escolhida.

Tabela 8 - Medidas de erro das previsões através de ANN com tendência: um e seis passos à frente

Séries	MAPE (%)		MAD	
	prev-o	prev-m	prev-o	prev-m
<i>ICMS</i>	4.2351	3.9608	361.2909	367.1399
<i>ICMSPE</i>	6.2879	5.5972	3.4639	2.5865
<i>ICMSRJ</i>	7.4286	6.0430	29.7292	22.8573
<i>ICMSSP</i>	1.6654	1.7574	49.3134	53.8354

As redes identificadas inicialmente através deste processo apresentam menores medidas de erro na previsão múltiplos passos à frente em comparação com as proporcionadas pelos métodos tradicionais. Este é o caso da rede *ANN(15,1,1)* com $\eta=0.1$ e $\alpha=0.1$ da série *ICMS*; as primeiras duas defasagens foram utilizadas, juntamente com a tendência e com 12 variáveis *dummy* identificando cada mês do ano. Para o *ICMSPE*, a rede *ANN(16,1,1)* com $\eta=0.1$, $\alpha=0.5$ e as três primeiras defasagens exibiram as menores medidas de erro nas previsões múltiplos passos, em comparação com as dadas pelas outras duas metodologias. Para o *ICMSRJ*, a rede *ANN(3,1,1,1)* com as duas primeiras defasagens, $\eta=0.01$ e $\alpha=0.5$, apresenta menores medidas de erro na previsão múltiplos passos; quanto ao *ICMSSP*, a rede *ANN(15,1,1,1)* com $\eta=0.01$, $\alpha=0.5$ e as duas primeiras defasagens apresentou um ligeiro aumento no *MAPE* da previsão múltiplos passos, mas uma diminuição considerável na medida do erro das previsões um passo à frente, em comparação com as demais metodologias.

A Tabela 8 exibe *MAPE* e *MAD* para cada uma das redes descritas anteriormente, podendo ser comparadas com as respectivas medidas referenciadas nas Tabelas 6 e 7. É possível observar que o *MAPE* das previsões múltiplos passos à frente das redes neurais é menor que o proporcionado pelo *SARIMA* e pelo alisamento exponencial de Holt-Winters. Através do *MAD*, notamos que os desvios obtidos com as redes neurais foram menores para o *ICMS* na previsão múltiplos passos à frente em comparação com as outras metodologias. No que tange à rede neural do *ICMSPE*, foram obtidos menores desvios na previsão um e múltiplos passos à frente; para as redes neurais do *ICMSRJ* e do *ICMSSP* isto acontece na previsão um passo à frente. O melhor desempenho das

previsões múltiplos passos à frente através das redes neurais relativo a metodologias alternativas é coerente com as conclusões obtidas por Fernandes(1995).

No segundo cenário como foi definido antes, a tendência foi excluída da entrada da rede para observar se é possível melhorar as previsões. Foram escolhidas as redes com os melhores ajustes e os menores erros de previsão sobre os conjuntos de teste.

A Tabela 9 exhibe as medidas de precisão relativas às previsões um passo e múltiplos passos à frente, apresentando melhores resultados na previsão múltiplos passos à frente comparado com os métodos tradicionais; Tabelas 6 e 7. Note que o *MAD* das previsões de redes neurais sem tendência foram maiores do que obtido através das redes neurais com tendência.

Tabela 9 - Medidas de erro das previsões através de ANN sem tendência: um e seis passos à frente

Séries	MAPE (%)		MAD	
	prev-o	prev-m	prev-o	prev-m
<i>ICMS</i>	3.1836	4.3451	215.8590	392.4455
<i>ICMSPE</i>	3.5231	5.6878	143.8273	290.1802
<i>ICMSRJ</i>	4.4673	5.8166	100.5846	245.3943
<i>ICMSSP</i>	1.6654	1.7574	116.2060	267.7710

Na segunda análise, em resumo, examinou-se a capacidade das redes neurais em produzir boas previsões um e múltiplos passos à frente. Utilizando as redes neurais e a tendência como neurônio de entrada, é possível obter redes capazes de fornecer boas previsões múltiplos passos à frente, com relação ao *MAPE* e *MAD*, como aconteceu com as previsões múltiplos passos do *ICMS*, *ICMSPE* e *ICMSRJ*, Tabela 8; no que tange às previsões múltiplos passos do *ICMSSP*, o modelo *SARIMA* exhibe os melhores resultados. Nas previsões um passo à frente, as redes neurais são as melhores na previsão do *ICMSSP*, enquanto o modelo *SARIMA* exhibe os melhores resultados na previsão um passo à frente para o *ICMS* e para o *ICMSPE*, mesmo que o alisamento exponencial de Holt-Winters (*HW*) apresenta os melhores resultados na previsão um passo à frente para o *ICMSRJ*; as Tabelas 6, 7 e 8 para fazer as comparações. Com relação ao *MAD*, as redes neurais apresentam, em geral, os menores desvios, tanto na previsão de um e quanto na de múltiplos passos à frente.

Considerando-se as redes neurais sem o neurônio de tendência, as previsões múltiplos passos à frente seguem sendo melhores, em comparação com as proporcionadas pelas outras duas metodologias; as Tabelas 6, 7 e 9. No caso do *ICMS*, através das redes neurais, os melhores resultados são obtidos na previsão um e múltiplos passos à frente. O *ICMSSP* também é previsto de forma apurada, usando-se redes neurais quando o horizonte de previsão é de um passo; adicionalmente, as redes neurais conseguem proporcionar resultados aceitáveis na previsão múltiplos passos do *ICMSRJ* e *ICMSPE*.

Semelhantemente ao que foi realizado na análise I, as previsões proporcionadas por cada método foram combinadas com vistas a se produzir uma previsão melhorada.

As previsões resultantes da combinação das previsões da rede neural com o neurônio de tendência (ANN_{ct}) e das fornecidas pelos outros dois métodos foram as mais precisas. Através desta combinação, os valores *MAPE* das previsões um e seis passos à frente para cada série foram menores do que aqueles fornecidos por cada um dos métodos individualmente, com a exceção da previsão seis passos à frente do *ICMSPE*, que é superada pela combinação *HW-ANN_{ct}*. No caso do *MAD*, as previsões individuais apresentaram menores desvios em comparação com a combinação entre eles, com exceção de alguns poucos casos. Os melhores resultados na previsão, segundo o *MAPE*, foram as proporcionados pela combinação *HW-SARIMA-ANN_{ct}*; Tabela 10.

Tabela 10 - Combinação de previsões *HW-SARIMA-ANN_{ct}*: um e seis passos à frente

Séries	MAPE (%)		MAD	
	prev-o	prev-m	prev-o	prev-m
<i>ICMS</i>	2.7433	3.5245	134.7323	431.5789
<i>ICMSPE</i>	2.4658	5.7267	4.8750	2.2824
<i>ICMSRJ</i>	3.6037	6.0000	29.4758	16.2487
<i>ICMSSP</i>	1.2964	1.1329	52.2552	44.3115

Tabela 11 - Combinação de previsões HW-SARIMA- ANN_{st} : um seis passos à frente

Séries	MAPE (%)		MAD	
	prev-o	prev-m	prev-o	prev-m
ICMS	2.8790	3.5081	106.9346	399.7609
ICMSPE	2.1194	5.4827	5.1981	2.1534
ICMSRJ	3.5770	5.8405	35.2078	14.9543
ICMSSP	1.2964	1.1329	52.2552	44.3115

Quando a rede neural sem o neurônio de tendência (ANN_{st}) foi utilizada, as combinações das previsões proporcionadas pelas metodologias HW-SARIMA- ANN_{st} , HW- ANN_{st} e SARIMA- ANN_{st} exibem os menores valores do MAPE na previsão um passo à frente, em comparação com as previsões dadas por cada metodologia individualmente, Tabela 11; no entanto, os valores do MAPE proporcionados pela combinação HW-SARIMA- ANN_{st} , nos casos do ICMSPE e ICMSSP, ainda foram menores na previsão um passo à frente do que os fornecidos pelas combinações HW- ANN_{st} e SARIMA- ANN_{st} . No caso das previsões seis passos à frente, a combinação HW-SARIMA- ANN_{st} apresenta as melhores previsões para o ICMS e o ICMSSP, segundo o MAPE; a combinação HW- ANN_{st} proporciona as melhores previsões para o ICMSPE e a combinação SARIMA- ANN_{st} , para o ICMSRJ. Com relação ao MAD, os menores desvios, na previsão um e seis passos à frente, foram distribuídos entre as combinações HW- ANN_{st} e SARIMA- ANN_{st} . Com relação à combinação HW-SARIMA, resultados destacáveis não foram obtidos em comparação com a combinação das outras metodologias; só alguns resultados foram melhores do que as previsões fornecidas por cada método individualmente.

5. Conclusões

O presente artigo abordou a geração de previsões de receitas tributárias. O interesse particular residiu na produção e avaliação de previsões da arrecadação do ICMS, tanto em nível nacional quanto em nível dos Estados de Pernambuco, Rio de Janeiro e São Paulo. Duas análises avaliaram o desempenho das redes neurais sob

diferentes esquemas preditivos. A rede *Perceptron multicamadas* foi utilizada no modelamento; três arquiteturas da classe *feedforward* foram propostas, entanto que a escolha do número de neurônios, número de camadas, funções de ativação e técnicas de estandarização estiveram baseadas em pesquisas preliminares sob previsão com redes neurais. A seleção das redes neurais com o melhor ajuste e prognóstico foram obtidas através de um esquema de análise treinamento-generalização sob um conjunto de redes preestabelecidas. A identificação e correção de *outliers* foi realizada sob as séries do *ICMS*, como, também, as componentes de tendência e sazonalidade utilizadas nas duas análises.

A primeira parte da análise comparou a capacidade preditiva das redes neurais, os modelos *SARIMA* e o alisamento exponencial na previsão. As redes neurais foram escolhidas minimizando o erro da previsão do *ICMS* em dezembro de 2004. Também, foram avaliadas as previsões no período entre janeiro e março de 2005. A capacidade preditiva das redes neurais foi evidente na previsão em dezembro de 2004 comparada com as metodologias tradicionais, ainda que, nas previsões do período seguinte, o alisamento exponencial apresentou os melhores resultados. O fato que as redes neurais não tenham conseguido generalizar completamente, foi consequência da escolha da melhor rede baseada unicamente na previsão em dezembro de 2004. Além disso, a combinação de previsões individuais mostrou-se efetiva na melhora dos prognósticos.

Uma segunda análise foi efetuada, onde as três metodologias foram avaliadas na previsão do período entre julho e dezembro de 2004. Adicionalmente, foi examinada a vantagem de incluir as componentes de tendência e sazonalidade em procura de melhorar as previsões. A evidência empírica exibiu que incluir nas redes neurais a componente sazonal definida como indicadores que representam cada mês do ano, foi efetiva na previsão, ao contrário de se utilizar a componente não observável extraída da metodologia de Box e Jenkins. Além disso, quando a componente de tendência não foi incorporada na rede neural, as previsões segundo o *MAPE* foram melhores às proporcionadas pelas redes neurais que incluíam a tendência, entanto que segundo o *MAD* aconteceu o caso contrário. Em geral, as redes neurais proporcionaram os melhores resultados na previsão múltiplos passos comparadas com as subministradas

pelas metodologias tradicionais. A combinação das previsões individuais dos modelos *SARIMA*, Holt-Winters e redes neurais (com e sem o neurônio de tendência) mostrou-se satisfatória na previsão um e múltiplos passos à frente, superando notavelmente as previsões dos modelos individuais segundo as medidas de erro utilizadas. Note que, as combinações nas que foram incluídas as previsões com redes neurais apresentaram resultados ainda melhores.

Em resumo, os resultados apresentados revelaram que é possível obter previsões mais precisas do que aquelas oriundas de metodologias tradicionais, como alisamento exponencial e Box e Jenkins recorrendo ao uso de redes neurais. Adicionalmente, mostrou-se que a inclusão de previsões obtidas através de redes neurais em esquemas de combinação de previsões é claramente benéfica.

Referências bibliográficas

- BARNARD, G. A. (1963), 'New Methods of Quality Control', *Journal of the Royal Statistical Society A* 126, 255--259.
- BATES, J. & GRANGER, W. J. (1969), 'The Combination of Forecast', *Operational Research Quarterly* 20, 451--468.
- BOX, G. E. P. & JENKINS, G. M. (1976), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.
- BOX, G. E. P., JENKINS, G. M. & REINSEL, G. (1994), *Time Series Analysis: Forecasting and Control*, 3rd edn, Englewood Cliffs: Prentice Hall.
- BOX, G. E. P. & TIAO, G. C. (1975), 'Intervention Analysis with Applications to Economic and Environmental problems', *Journal of the American Statistical Association* 70, 71--79.
- CASDAGLI, M. & EUBANK, S. (1992), *Nonlinear Modeling and Forecasting*, Addison-Wesley, chapter XII of SFI Studies in the Sciences of Complexity.
- CHEN, C. & LIU, L. (1993), 'Joint Estimation of Model Parameters and Outliers Effects in Time Series', *Journal of the American Statistical Association* 88, 284-- 297.
- COHEN, M., FRANCO, H., MORGAN, N., RUMELHART, D. & ABRASH, V. (1993), *Advances in Neural Information Processing Systems*, Morgan Kaufmann, chapter Context-dependent multiple distribution phonetic model ing with MLPs, pp. 649--657.
- CYBENKO, G. (1988), Continuous Valued Neural Networks with Two Hidden Layers are Sufficient, Technical report, Tuf University.
- CYBENKO, M. (1989), 'Aproximation by Superposition of a Sigmoidal Function', *Mathematics of Control, Signals and Systems* 2, 303--314.

- DIJK, D. V. (1999), *Smooth Transition Models: Extensions and Outlier Robust Inference*, PhD thesis, Tinbergen Institute. Rotterdam, The Netherlands. www.few.eur.nl/few/people/djvandijk/thesis.
- FERNANDES, L. L. (1995), 'Utilização de Redes Neurais na Análise e Previsão de Séries Temporais'. Dissertação de Mestrado. U.F.G.R.S. - Pós-Graduação da computação.
- GATELY, E. (1996), *Neural Networks for Financial Forecasting*, New York: John Wiley and Sons.
- GÓMEZ, V. & MARAVALL, A. (2001a), *Automatic Modelling Methods for Univariate Series*, New York: John Wiley and Sons, chapter 7, in Peña D., Tiao G.C. & Tsay R.S, A course in Advanced Time Series Analysis.
- GOMÉZ, V. & MARAVALL, A. (2001b), *Seasonal Adjustment and Signal Extraction in Economic Time Series*, New York: John Wiley and Sons, chapter 8, in Peña D., Tiao G.C.& Tsay R.S, A Course in Advanced Time series Analysis.
- GORR, L. & SZCZYPULA, J. (1994), 'Comparative Study of Artificial Neural Networks and Statistical Models Predicting Student Grade Point Averages', *International Journal of Forecasting* 10, 17--34.
- GRANGER, C. W. & NEWBOLD, J. P. (1977), *Forecasting Economic Time Series*, New York: Academic Press.
- GUERRERO, V. (1991), *Análisis de Series de Tiempo Económicas*, Universidad Autónoma Metropolitana de México D.F.
- GUIDOLIN, M. & NA, C. F. (2006), The economic and statistical value of forecast combinations under regime switching: An application to predictable US returns, Technical report, Research Division-Federal Reserve Bank of St Louis.
- HAYKIN, S. (1994), *Neural Networks*, New York: McMillan College Publishing Company.
- HENDRY, D. F. & CLEMENTS, M. P. (2004), 'Pooling of Forecasts', *Econometrics Journal* 7, 1--31.
- HORNIK, K., STINCHCOMBE, M. & WHITE, H. (1989), 'Multilayer Feedforward Networks and Universal Approximations', *Neural Networks* 2, 359--366.
- ISASI, P. & GALVÁN, I. (2004), *Redes Neuronales Artificiales -- Un Enfoque Práctico*, Madrid: Pearson-Prentice Hall.
- KAASTRA, I. & BOYD, M. (1996), 'Design a Neural Network for Forecasting Financial and Economic Time Series', *Neurocomputing* 10, 215--236.
- KAISER, R. & MARAVALL, A. (2001), Seasonal Outliers in Time Series. www.bde.es/servicio/software/trabajos.htm.
- LIPPMANN, R. P. (1987), 'A Introduction to Computing with Neural Nets', *IEEE ASSP Magazine* pp. 4--22.
- LJUNG, G. M. & BOX, G. E. P. (1986), 'Time Series Model Specification in Presence of Outliers', *Journal of the American Statistical Association* 81, 132--141.
- MARAVALL, A. & KAISER, R. (2000), Notes on Time Series Analysis, ARIMA models and Signal Extraction. www.bde.es/servicio/software/trabajos.htm.
- MARCELLINO, M. (2002), Forecast Pooling for Short Time Series of Macroeconomic Variables, Technical report, Innocenzo Gasparini Institute for Economic Research.

- MONTGOMERY, D. C. & JOHNSON, L. A. (1976), *Forecasting and Time Series Analysis*, New York: McGraw Hill.
- MORETTIN, P. A. & TOLOI, C. M. (2004), *Análise de Séries Temporais*, ABE - Projeto Fisher. São Paulo: Edgard Blücher.
- MURATA, N., YOSHIKAWA, S. & AMARI, S. (1994), 'Network Information Criterion-Determining the Number of Hidden Units for an Artificial Neural Network Model', *Proceedings of the IEEE Transactions on Neural Networks* 5(6), 865--872.
- NARENDRA, K. & PARTHASARANTY, K. (1990), 'Identification and Control of Dynamical Systems Using Neural Networks', *IEEE Transactions on Neural Networks* 1, 4--27.
- PEÑA, D. (2001), *Outliers Influential Observations and Missing data*, New York: John Wiley and Sons, chapter 6, in Peña D., Tiao G.C. & Tsay R.S., *A Course in Advanced Time series Analysis*.
- PERNÍA-ESPIÑOZA, A., ORDIERES, J., DE PISÓN, F. & MARCOS, A. (2005), 'TAO-Robust Backpropagation Learning Algorithm', *Neural Networks* 18, 191--204.
- RUMELHART, D., HILTON, G. & WILLIAMS, R. (1986a), 'Learning Representations by Backpropagating Errors', *Nature* 323, 533--536.
- RUMELHART, D., HILTON, G. & WILLIAMS, R. (1986b), *Parallel Distributed Processing*, Cambridge: The MIT Press, chapter Learning representations by backpropagating errors.
- SHARDA, R. (1994), 'Neural networks for the MS/OR analyst: An application bibliography', *Interfaces* 24(2), 116--130.
- SIETSMAN, J. & DOW, R. (1988), 'Neural Net Pruning-Why and How?', *Proceedings of the IEEE Transactions on Neural Networks* 1, 325--333.
- SMITH, J. & WALLIS, K. (2005), *Combining Point Forecast: The Simple Average Rules Ok?*, Technical report, Department of Economics-University of Warwick.
- SOUZA, R. C. & ZANDONADE, E. (1993), *Forecasting Via Neural Networks: Comparative Study*, Technical report, Department of Electrical Engineering, Catholic University of Rio de Janeiro.
- VARFIS, A. & VERSINO, C. (1990), *Univariate Economic Time Series Forecasting*, Cambridge University Press.
- WANG, Z., MASSINO, C., THAM, M. & MORRIS, A. (1994), 'A Procedure for Determining the Topology of Multilayer Feedforward Neural Networks', *Neural Networks* 7(2), 291--300.
- WEIGEND, A. S. & GERSHENFELD, N. A. (1993), *Time Series Prediction: Forecasting the Future and Understanding the Past*, Addison-Wesley, chapter XV of *SFI Studies in the Sciences of Complexity*.
- WIEGGEND, A., HUBERMAN, B. & RUMELHART, D. (1990), *Predicting the Future: a Connectionist Approach*, Technical report, PARC.
- ZHANG, G., PATUWO, B. & HU, Y. (1998), 'Forecasting with Artificial Neural Networks: The State of Art', *International Journal of Forecasting* 14, 35--62.
- ZHANG, X. (1994), 'Time Series Analysis and Prediction by Neural Networks', *Optimization Methods and Software* 4, 151--170.

Abstract

The production of Brazilian sale taxes revenues forecasts is the principal one thematic of this article. Especially, we want to evaluate the utility of forecasting with neural networks methods. The national *ICMS* and of three states (São Paulo, Rio de Janeiro and Pernambuco) will be forecast and evaluated through neural networks methodologies, *SARIMA* methods and exponential smoothing of Holt-Winters. These financial series often show *outliers*, making necessary its correction in a preliminary phase to the fit; the neural networks can result affected for this event, reason for which, the identification and correction of *outliers* is enclosed in the analysis. The incorporation of the unobservable components as trend and seasonality is examined in the neural networks to evaluate their predictive capacity. Also, the combination of forecasts is applied in this direction. The results show that neural networks provide better forecasts compared with the traditional methodologies; moreover, the inclusion of unobservable components improve the neural networks capacity of forecasting, meanwhile, combination of forecasts that include neural networks predictions leads to higher forecasting accuracy.

Diagnóstico em análise discriminante

*Sueli Maria Beltrame Reigada**
*Silvia Nagib Elian***

Resumo

Este trabalho apresenta medidas de diagnóstico em análise discriminante em duas situações principais: para duas populações com matriz de covariância constante e duas populações com matrizes de covariância desiguais (análise discriminante quadrática). É estudada ainda a abordagem Bayesiana para a detecção de observações influentes. No final do trabalho, examina-se também as medidas de diagnóstico para a análise discriminante múltipla, ou seja, quando se tem mais que duas populações. O estudo é complementado com a aplicação das principais medidas estudadas a um conjunto de dados reais.

1. Introdução

A literatura sobre análise discriminante bem como sobre diagnóstico é muito vasta, mas a pesquisa na determinação de pontos influentes em análise discriminante ainda é pouco explorada.

Neste trabalho, serão apresentadas as medidas de diagnóstico em análise discriminante, definidas pelos principais pesquisadores do assunto nos últimos 27 anos.

* Endereço para correspondência: smbreigada@uol.com.br.

** (IME-USP) - selian@ime.usp.br.

Tais medidas indicam pontos que podem influenciar na regra de alocação, na probabilidade de classificação incorreta, ou em ambos.

O presente artigo está esquematizado da seguinte forma: na seção 2, serão apresentadas as medidas de diagnóstico em análise discriminante para duas populações e matriz de covariância constante. A seção 3 trata da detecção de observações influentes no contexto Bayesiano. Sob essa abordagem, são definidas medidas de influência globais, locais e individuais. Na seção 4, são apresentadas as medidas de diagnóstico utilizadas na análise discriminante quadrática. A seção 5 se destina às medidas que detectam observações influentes em análise discriminante múltipla. Uma aplicação das medidas de diagnóstico a um conjunto de dados reais encontra-se na seção 6. A seção 7 apresenta um estudo de simulação e a seção 8 finaliza o trabalho apresentando algumas conclusões.

2. Diagnóstico em análise discriminante com duas populações e matriz de covariância constante

2.1 Função de influência em análise discriminante

Na análise de diagnóstico, quando o procedimento estatístico adotado envolve estimação, uma estratégia comum na determinação de pontos influentes é proceder ao ajuste com e sem o ponto suspeito e comparar as estimativas obtidas. Nessa linha de estudo, Campbell (1978) sugeriu o uso da função de influência, proposta por Hampel (1974).

O objetivo principal da função de influência teórica é calcular a influência de um particular ponto \mathbf{x} no parâmetro de interesse, de modo que valores altos para essa função indicariam que \mathbf{x} tem grande influência no parâmetro. Em análise de grupos multivariados, a função de influência é determinada excluindo uma observação de apenas um grupo.

Campbell (1978) considera o caso em que o vetor aleatório \mathbf{X} tem distribuição normal p -variada com média μ_j e matriz de covariância Σ , para \mathbf{X} pertencendo à

população $\pi_j, j = 1, 2$, e utiliza a função de influência definida a seguir.

Dadas g populações com funções distribuição $F_1, F_2, \dots, F_k, \dots, F_g$, seja $\theta = T(F_1, F_2, \dots, F_k, \dots, F_g)$, um parâmetro geral, obtido a partir das funções de distribuição $F_k, k = 1, \dots, g$. Para um valor \mathbf{x} do vetor aleatório \mathbf{X} , a função de influência $I(\mathbf{x}; \theta)$ é definida como:

$$I(\mathbf{x}; \theta) = \lim_{\varepsilon \rightarrow 0} \left(\frac{\tilde{\theta} - \theta}{\varepsilon} \right) \quad \text{em que} \quad 0 < \varepsilon < 1, \quad \tilde{\theta} = T(\tilde{F}) \quad , \quad \tilde{F} = (F_1, \dots, \tilde{F}_k, \dots, F_g)$$

e $\tilde{F}_k = (1 - \varepsilon)F_k + \varepsilon \delta_{\mathbf{x}}$ é a função distribuição F_k após sofrer perturbação no ponto \mathbf{x} , sendo $\delta_{\mathbf{x}}$ uma função de distribuição que assume probabilidade 1 no ponto \mathbf{x} .

Considera-se inicialmente perturbações em Σ^{-1} e μ_j , sendo que $\Sigma = w_1 \Sigma_{F_1} + w_2 \Sigma_{F_2}$, Σ_{F_1} e Σ_{F_2} são as matrizes de covariâncias para as populações π_1 e π_2 com $w_1 + w_2 = 1$ e $w_j > 0$. Usualmente os pesos w_1 e w_2 são adotados como

$$w_1 = \frac{n_1}{n_1 + n_2} \quad \text{e} \quad w_2 = \frac{n_2}{n_1 + n_2} .$$

Tomando a função linear discriminante $y = \mathbf{l}'\mathbf{x} = \delta' \Sigma^{-1} \mathbf{x}$ em que $\delta = \mu_1 - \mu_2$, e a distância de Mahalanobis $\Delta^2 = (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2)$, calcula-se os parâmetros após a exclusão de um elemento da população π_1 , (Reigada 2005, pg. 8), e usando a notação * sobrescrito para os parâmetros após perturbação tem-se:

$$\begin{aligned} \mu_1^* &= \mu_1 + \varepsilon(\mathbf{x} - \mu_1) = \mu_1 + \varepsilon \mathbf{z} \quad \text{com} \quad \mathbf{z} = \mathbf{x} - \mu_1 \\ \delta^* &= \delta + \varepsilon \mathbf{z} \quad , \quad \Sigma^* = (1 - \varepsilon w_1) \Sigma + \varepsilon w_1 \mathbf{z} \mathbf{z}' \quad \text{e} \quad \Sigma^{-1*} \cong (1 + \varepsilon w_1) \Sigma^{-1} - \varepsilon w_1 \Sigma^{-1} \mathbf{z} \mathbf{z}' \Sigma^{-1} . \end{aligned}$$

Para construir a função de influência para a distância de Mahalanobis $\Delta^2 = \delta' \Sigma^{-1} \delta$, obtém-se $\Delta^{2*} \cong (1 + \varepsilon w_1) \Delta^2 + 2\varepsilon \phi - \varepsilon w_1 \phi^2$ em que $\phi = \delta' \Sigma^{-1} \mathbf{z}$, e conseqüentemente a função de influência $I(\mathbf{x}; \Delta^2) = w_1 \Delta^2 + 2\phi - w_1 \phi^2$.

Se \mathbf{x} pertence à população π_1 que foi perturbada, tem-se: $E(\phi) = E(\delta' \Sigma^{-1} \mathbf{z}) = 0$ e $Var(\phi) = \Delta^2$, portanto $\phi \sim N(0, \Delta^2)$, e $\phi_p = \frac{\phi}{\Delta}$ tem distribuição normal padrão.

Tomando $\phi = \phi_p \Delta$, a função de influência para Δ^2 pode ser escrita em termos de ϕ_p como:

$$I_p(\mathbf{x}; \Delta^2) = w_1 \Delta^2 + 2\Delta \phi_p - w_1 \Delta^2 \phi_p^2.$$

A função de influência pode ser considerada como uma variável aleatória, uma vez que é uma transformação do vetor aleatório \mathbf{X} . Se \mathbf{X} tem distribuição normal multivariada, verifica-se (Reigada, 2005, pg. 144) que a distribuição de probabilidades da função de influência para a distância de Mahalanobis (Δ^2) tem assimetria negativa.

Tomando a função de influência para a distância de Mahalanobis com ϕ não padronizado, $I(\mathbf{x}; \Delta^2) = w_1 \Delta^2 + 2\phi - w_1 \phi^2$ e derivando em relação a ϕ , obtém-se o ponto de máximo para $\phi = \frac{1}{w_1}$, assim, $I_{max}(\mathbf{x}; \Delta^2) = w_1 \Delta^2 + \frac{1}{w_1}$.

Campbell (1978) sugere o uso de $I_m(\mathbf{x}; \Delta^2) = I_{max}(\mathbf{x}; \Delta^2) - I(\mathbf{x}; \Delta^2)$, dada por $I_m(\mathbf{x}; \Delta^2) = w_1^{-1} (1 - 2w_1 \Delta \phi_p + w_1^2 \Delta^2 \phi_p^2)$. Esta variável aleatória é sempre positiva e, como $I(\mathbf{x}; \Delta^2)$ tem assimetria negativa, segue que $I_m(\mathbf{x}; \Delta^2)$ tem distribuição positivamente assimétrica. Multiplicando e dividindo $I_m(\mathbf{x}; \Delta^2)$ por ϕ_p^2 tem-se

$$I_m(\mathbf{x}; \Delta^2) = w_1 \Delta^2 (\phi_p - w_1^{-1} \Delta^{-1})^2.$$

Como $\phi_p \sim N(0,1)$ então

$$(\phi_p - w_1^{-1} \Delta^{-1}) \sim N_\phi(-w_1^{-1} \Delta^{-1}, \mathbf{1})$$

e dessa forma, $(\phi_p - w_1^{-1} \Delta^{-1})^2 = (\phi_p - w_1^{-1} \Delta^{-1})' \mathbf{1}^{-1} (\phi_p - w_1^{-1} \Delta^{-1})$ tem distribuição qui-quadrado não central com 1 grau de liberdade e parâmetro de não centralidade $(w_1^{-1} \Delta^{-1})^2$.

Portanto $\frac{I_m(\mathbf{x}; \Delta^2)}{w_1 \Delta^2}$ tem uma distribuição qui-quadrado não central com um grau de

liberdade e parâmetro de não centralidade $(w_1^2 \Delta^2)^{-1}$. Observa-se então que a distribuição nula de $I_m(\mathbf{x}; \Delta^2)$ é chi-quadrado com um grau de liberdade, sendo assim a distribuição esperada na inexistência de pontos discrepantes.

2.2. Principais estatísticas no diagnóstico em análise discriminante

Fung (1992) apresentou duas estatísticas fundamentais no diagnóstico em análise discriminante. Uma delas é a medida ϕ , apresentada na seção anterior. Tomando a regra discriminante de Fisher em que uma observação x é alocada na população π_1 se $(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} \mathbf{x} \geq \frac{1}{2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} \bar{\mathbf{x}}_1 + \frac{1}{2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} \bar{\mathbf{x}}_2$, chega-se a $\hat{\phi} \geq -\frac{1}{2} D^2$, em que $\hat{\phi} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1)$ e D é a estimativa para a distância de Mahalanobis.

Portanto, a medida ϕ é importante tanto na discriminação quanto na análise de influência.

A segunda medida é $d_{ji}^2 = (\mathbf{x}_{ji} - \bar{\mathbf{x}}_j)' \mathbf{S}^{-1} (\mathbf{x}_{ji} - \bar{\mathbf{x}}_j)$ $i=1, \dots, n_j$, $j=1, 2$. Healy (1968) utiliza essa estatística em testes de normalidade e também com o objetivo de detectar *outliers* sob normalidade.

Assim como o resíduo é o ponto de alavanca em regressão, Fung (1992) aponta d_{ji}^2 e $\hat{\phi}_{ji} = \hat{\mathbf{I}}' (\mathbf{x}_{ji} - \bar{\mathbf{x}}_j)$ como estatísticas básicas para detectar *outliers* e observações influentes em análise discriminante, uma vez que várias medidas de diagnóstico são escritas em função dessas duas estatísticas.

Por facilidade de notação, serão usados os termos d_i^2 e ϕ_i nas próximas seções.

Fung (1995-a) mostra que, assintoticamente, d_i^2 é distribuída como χ_p^2 ($p =$ dimensão do vetor X) e $\frac{\hat{\phi}_i}{D}$ tem distribuição $N(0, 1)$. Tais resultados assintóticos são

fundamentais uma vez que calculando-se estas estatísticas para os elementos amostrais, valores maiores que os respectivos valores críticos da distribuição chi-

quadrado e da distribuição normal são indicadores de pontos influentes, Reigada ,2005, pg. 25.

2.3. Medidas relacionadas com a probabilidade de classificação incorreta

Em análise discriminante com duas populações, a probabilidade de classificação incorreta (PCI) é definida como $PCI = P(1/2)(1-q) + P(2/1)q$ em que

$P(i/j)$ = probabilidade de classificar uma observação em π_i quando ela é de π_j ;

q = probabilidade *a priori* de π_1 ; e

$1-q$ = probabilidade *a priori* de π_2 .

Quando as probabilidades *a priori* são iguais, $PCI = \frac{1}{2}P(2/1) + \frac{1}{2}P(1/2)$. Sabe-se

que $PCI = \Phi\left(\frac{-\Delta}{2}\right)$, em que $\Phi(\cdot)$ é a função de distribuição acumulada da distribuição normal padrão.

Fung(1992) propõe uma medida de diagnóstico com base na diferença das estimativas das probabilidades de classificação incorreta definida como:

$$DPCI_i = \left[\Phi\left(-\frac{D_{(i)}}{2}\right) \right] - \left[\Phi\left(-\frac{D}{2}\right) \right] \text{ em que } D_{(i)} \text{ é a estimativa da distância de}$$

Mahalanobis sem a observação i do grupo 1.

$$\text{Após vários cálculos, obtém-se } DPCI_i \cong \frac{\Phi\left(-\frac{1}{2}D\right)}{4D(n_1-1)^2} \left[(1-w_1\hat{\phi}_i)^2 (d_i^2 - \hat{\phi}_i^2 / D^2) + \frac{1}{4}\hat{\phi}_i^2 \right],$$

que depende dos valores de $\hat{\phi}_i$ e d_i^2 .

Critchley e Vitiello (1991), independentemente de Fung (1992), determinaram duas estatísticas principais que compõe as medidas de influência. São elas a "atipicalidade", que exerce o papel da medida d_i^2 e a diferença entre o escore discriminante linear da observação e o escore discriminante linear de sua média amostral, que corresponde à medida ϕ_i . Os autores determinaram a medida de influência

para avaliar o efeito em Δ^2 produzido pela eliminação de uma única observação i da amostra proveniente de π_1 ,

$$\hat{\Delta}_{1(i)}^2 = \left(\frac{n_1 + n_2 - 1}{n_1 + n_2} \right) \left[\hat{\Delta}^2 - \frac{1}{w_1(n_1 - 1)} + \frac{\{e(\mathbf{x}_{1i}) - w_1^{-1}\}^2}{w_1^{-1}(n_1 - 1) - \hat{\alpha}_1^2(\mathbf{x}_{1i})} \right],$$

que está relacionada com a probabilidade de classificação incorreta visto que $PCI = \Phi\left(\frac{-\Delta}{2}\right)$.

2.4. Medidas baseadas na média quadrática

Na análise discriminante com duas populações, uma regra de alocação eficiente é tal que minimiza o custo médio de classificação incorreta, ou seja, consiste em alocar \mathbf{x}_0 (de origem desconhecida) em π_1 se $o(x_0) = \frac{\Pr(x_0 \in \pi_1 / d)}{\Pr(x_0 \in \pi_2 / d)} > \frac{c(1/2)}{c(2/1)} \cdot \frac{1-q}{q}$, e em π_2 caso contrário, sendo que $c(i/j)$ = custo de classificar uma observação em π_i quando de fato ela pertence a $\pi_j, i, j = 1, 2, i \neq j$. Aplicando a função logaritmo obtém-se

$$(\mu_1 - \mu_2)' \Sigma^{-1} \mathbf{x}_0 - \frac{1}{2} (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 + \mu_2) \geq \ln \left[\frac{c(1/2)}{c(2/1)} \cdot \frac{1-q}{q} \right].$$

Tomando essa regra discriminante de Fisher, assumindo os custos de classificação incorreta iguais e usando estimativas das quantidades desconhecidas, aloca-se \mathbf{x}_0 em

$$\pi_i \text{ se : } \hat{\mathbf{I}}' \left[\mathbf{x}_0 - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right] > \ln \left(\frac{1-q}{q} \right).$$

A quantidade $\hat{\mathbf{I}}' \left[\mathbf{x}_0 - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right] - \ln \left(\frac{1-q}{q} \right)$ é denominada logaritmo da razão das

probabilidades estimadas (log-odds). Na maioria dos casos, admite-se $1 - q = q$, e essa quantidade reduz-se à função discriminante para a observação \mathbf{x}_0 .

Fung (1995-a) propôs a medida $E(DLO_i)^2$, que é a média do quadrado da diferença entre os log-odds da amostra toda e da amostra sem a observação i :

$$E(\hat{\gamma}'\mathbf{y} - \hat{\gamma}'_{(i)}\mathbf{y})^2 \text{ em que } \hat{\gamma}'\mathbf{y} - \hat{\gamma}'_{(i)}\mathbf{y} = \left(-\hat{\mathbf{l}}' \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} + \hat{\mathbf{l}}'\mathbf{x}\right) - \left(-\hat{\mathbf{l}}'_{(i)} \frac{(\bar{\mathbf{x}}_{1(i)} + \bar{\mathbf{x}}_2)}{2} + \hat{\mathbf{l}}'_{(i)}\mathbf{x}\right).$$

Para esse cálculo foi utilizado $E(DLO_i)^2 = Var(DLO_i) + E^2(DLO_i)$, obtendo-se

$$E(DLO_i)^2 = (w_1B_1 + w_2B_2)^2 + V$$

em que $B_1 = \frac{A_1}{2} - \frac{A_2}{2}$ e $B_2 = \frac{-A_1}{2} - \frac{A_2}{2}$, com

$$A_1 = \frac{D^2}{n-2} + cf_i(\hat{\phi}_i)^2 + ch\hat{\phi}_i - chf_id_i^2\hat{\phi}_i$$

$$A_2 = c \cdot h \cdot (\hat{\phi}_i - h \cdot d_i^2)g_i$$

$$V = \frac{D^2}{(n-2)^2} + c(f_i\hat{\phi}_i + hg_i) \left[\frac{2\hat{\phi}_i}{n-2} + (f_i\hat{\phi}_i + hg_i)kd_i^2 \right]$$

$$c = \frac{n-3}{n-2}, \quad f_i = \frac{b}{1+bd_i^2}, \quad b = \frac{-n_1}{(n_1-1)(n-2)}, \quad h = \frac{1}{n_1-1}, \quad g_i = \frac{1}{1+bd_i^2}.$$

Portanto, essa medida também é função de $\hat{\phi}_i$ e d_i^2 , e detecta observações com grande influência no logaritmo da razão das probabilidades estimadas (escore discriminante quando $1 - q = q$).

Adicionalmente, Fung (1996-b) obteve medidas de diagnóstico condicionais ao fato que a observação \mathbf{x} está próxima do plano discriminante, uma vez que essas observações são de difícil classificação. Fung (1998) mostra que a medida condicional tomada para probabilidades *a priori* iguais é assintoticamente proporcional à medida $DPCI_i$, proposta em 1992 pelo próprio autor.

3. Detecção de observações influentes no contexto Bayesiano

3.1. Densidades preditivas e divergência de Kullback-Leibler

Utilizando densidades preditivas, que são a base da análise discriminante Bayesiana, será apresentado nesta seção o estudo da influência das observações sobre aspectos inferenciais, tais como: alocação de futuras observações, separação entre populações e determinação da probabilidade de uma observação pertencer a certa população. Para esse estudo, são analisadas estatísticas de influências definidas por Johnson (1987). Em adição à notação já utilizada, sejam

o vetor de parâmetros (μ_1, μ_2, Σ) , que será denominado por θ ;

X_{ji} com distribuição normal p -variada condicional aos parâmetros;

\mathbf{d} , definido como o conjunto de todos os dados observados $X_{ji}, j = 1, 2, i = 1, 2 \dots n_j$;

$p(\theta/\pi_j), j = 1, 2$, a densidade *a priori* para θ sob π_j ;

$p(\theta/\mathbf{d}, \pi_j)$, densidade *a posteriori* para θ sob π_j ; e

\mathbf{Y} vetor aleatório de observações futuras, independente do conjunto de dados, com distribuição condicional aos parâmetros $N_p(\mu_j, \Sigma)$ sob π_j .

No enfoque *Bayesiano*, o problema de previsão de observações futuras é resolvido utilizando-se $f(\mathbf{y}/\mathbf{d})$, que é a densidade *a posteriori* da quantidade a ser prevista, \mathbf{Y} , dada a amostra observada \mathbf{d} . Essa densidade, denominada densidade preditiva *a posteriori* de \mathbf{Y} dado \mathbf{d} , é definida como sendo a esperança *a posteriori* da densidade amostral para \mathbf{Y} :

$$f_j(\mathbf{y}) = f(\mathbf{y}/\mathbf{d}, \pi_j) = \int N(\mathbf{y}/\theta, \pi_j) p(\theta/\mathbf{d}, \pi_j) d\theta, \quad j = 1, 2, \dots$$

esperança essa tomada com relação à distribuição *a posteriori* de θ dada a amostra \mathbf{d} .

Se o vetor de parâmetros θ é conhecido, a densidade preditiva para \mathbf{Y} (sob π_j) será a densidade amostral $N(\cdot/\theta, \pi_j)$.

Quando θ é desconhecido, Johnson (1987) obtém densidades preditivas aproximadas, substituindo os parâmetros da densidade normal pelas correspondentes estimativas. Outra possibilidade seria o uso da *priori* de referência

$$p(\theta / \pi_j) \propto |\Sigma|^{-\frac{1}{2}(p+1)} \quad j=1,2 \quad (3.1)$$

sendo que, com base nessa distribuição, Geisser (1964, 1966) obteve a densidade preditiva:

$$Y / \mathbf{x}, \pi_j \sim St(\nu, \bar{\mathbf{x}}_j, (1+n_j^{-1})\tilde{\mathbf{S}})$$

que é a distribuição *t-Student* p -variada com ν graus de liberdade, vetor de locação $\bar{\mathbf{x}}_j$ e matriz de dispersão $(1+n_j^{-1})\tilde{\mathbf{S}}$, em que

$$\nu = n - p - 1, \quad \bar{\mathbf{x}}_j = \sum_{i=1}^{n_j} \frac{\mathbf{x}_{ji}}{n_j} \quad \text{e} \quad \tilde{\mathbf{S}} = \nu^{-1} \sum_{j=1}^2 \sum_{i=1}^{n_j} (\mathbf{x}_{ji} - \bar{\mathbf{x}}_j)(\mathbf{x}_{ji} - \bar{\mathbf{x}}_j)' = \nu^{-1}(n-2)\mathbf{S}.$$

Com o objetivo de avaliar a discrepância entre duas densidades preditivas, Johnson (1987) usou a medida de divergência de Kullback-Leibler (1951) :

$$I(f_i, f_j) = E_{f_i} \ln(f_i f_j^{-1}) = \int_{\mathcal{R}^p} f_i(\mathbf{y}) \ln \left(\frac{f_i(\mathbf{y})}{f_j(\mathbf{y})} \right) d\mathbf{y}.$$

Se f_i é a distribuição $N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, verifica-se que :

$$2I(f_i, f_j) = (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}_j^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) + \text{tr} \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j^{-1} - \ln |\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j^{-1}| - p.$$

Quando f_i é a densidade *t-Student* com η_j graus de liberdade, vetor de locação $\boldsymbol{\alpha}_j$ e matriz de dispersão $\boldsymbol{\lambda}_j$, não é possível calcular $I(f_i, f_j)$ explicitamente. Em face da complexidade na solução das integrais envolvidas no cálculo, aproxima-se tal densidade por uma densidade normal multivariada apropriada.

Johnson (1987) propõe aproximar a densidade *t-Student* por uma densidade normal com parâmetros $\boldsymbol{\mu}_j$ e $\boldsymbol{\Sigma}_j$ de modo a minimizar $I(St(\eta_j, \boldsymbol{\alpha}_j, \boldsymbol{\lambda}_j), N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j))$. Verifica-

se que o ínfimo dessa expressão é atingido para $\boldsymbol{\mu}_j = \boldsymbol{\alpha}_j$ e $\boldsymbol{\Sigma}_j = \frac{\eta_j}{\eta_j - 2} \boldsymbol{\lambda}_j$. Dessa forma,

a estimativa normal ótima para a densidade *t-Student* $St(\eta_j, \alpha_j, \lambda_j)$ é

$$\tilde{f}_j = N\left(\alpha_j, \left(\frac{\eta_j}{\eta_j - 2}\right)\lambda_j\right).$$

Quando o objetivo é alocar \mathbf{y} em π_1 ou π_2 , a regra ótima consiste em alocar \mathbf{y} em π_1 se : $o(\mathbf{y}) = \frac{\Pr(\mathbf{y} \in \pi_1 / d)}{\Pr(\mathbf{y} \in \pi_2 / d)} > \frac{c(1/2)}{c(2/1)} \cdot \frac{1-q}{q}$. Verifica-se que $o(\mathbf{y}) = \frac{q}{(1-q)} \cdot \frac{f_1(\mathbf{y})}{f_2(\mathbf{y})}$.

Quando os parâmetros são conhecidos, utiliza-se

$$\ln(o(\mathbf{y})) = \ln\left(\frac{q}{1-q}\right) + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} \mathbf{y} - \frac{1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2),$$

que é a função discriminante de *Fisher* quando as probabilidades q e $1-q$ forem iguais.

Se θ é desconhecido, os parâmetros são substituídos por suas estimativas. Utilizada *a priori* de referência (3.1) calcula-se o logaritmo da razão das probabilidades estimadas (*log odds*) *a posteriori* como:

$$\ln(o(\mathbf{y})) = \ln\left(\frac{q}{1-q}\right) + \ln\left(\frac{St(\mathbf{y}/v, \bar{\mathbf{x}}_1, \kappa_1 \tilde{\mathbf{S}})}{St(\mathbf{y}/v, \bar{\mathbf{x}}_2, \kappa_2 \tilde{\mathbf{S}})}\right) \text{ com } \kappa_1 = 1 + n_1^{-1} \text{ e } \kappa_2 = 1 + n_2^{-1}.$$

Medida de influência global, local e individual

Os conceitos de medida de influência global, local e individual foram definidos por Johnson (1987). Uma medida de influência é dita global quando avalia o quanto uma observação afeta dois ou mais aspectos da análise. Uma vez que a densidade preditiva

$$f_j(\mathbf{y} / d, \pi_j) = \int N(\mathbf{y} / \boldsymbol{\theta}, \pi_j) p(\boldsymbol{\theta} / d, \pi_j) d\boldsymbol{\theta}$$

envolve vários elementos – funções: densidade, probabilidades, dados, parâmetros –, analisando o efeito que as observações têm sobre a densidade preditiva, obtêm-se medidas de influência globais.

Conforme descrito na seção 3.1, para medir o efeito de eliminar a observação x_i , $i=1,2,\dots,n_j$, na j -ésima densidade preditiva, Johnson (1987) emprega a medida de divergência simétrica de Kullback-Leibler. Na impossibilidade de calcular $J(f_j, f_{j(i)})$ explicitamente, é necessário usar aproximações.

Se utilizada *a priori* de referência (3.1), a densidade preditiva obtida é $St(\eta_j, \alpha_j, \lambda_j)$, em que $\eta_j = n - p - 1$, $\lambda_j = (1 + n_j^{-1})\tilde{S}$ e $\alpha_j = \bar{x}_j$. A aproximação normal ótima para esta densidade é $N\left(\alpha_j, \frac{\eta_j}{(\eta_j - 2)}\lambda_j\right)$, sendo que $\alpha_j = \bar{x}_j$, e a matriz de covariância é aproximada por $\frac{n-2}{n-p-3}S$.

A medida de influência é local quando mede o efeito de uma observação em um objetivo inferencial específico, tal como, por exemplo, alguma estatística de separação. Para medir o efeito de uma observação no logaritmo da razão das probabilidades estimadas, Johnson (1987) considerou a esperança do quociente dessas medidas na presença e ausência da observação. Tal medida, denominada esperança do logaritmo da razão das probabilidades estimadas é definida como:

$ELO_i = E \ln\left\{\frac{o(\mathbf{y})}{o_{(i)}(\mathbf{y})}\right\}$ sendo que a esperança é calculada em relação à densidade

preditiva com todos os dados, $f(\mathbf{y}) = qf_1(\mathbf{y}) + (1-q)f_2(\mathbf{y})$ e $o(\mathbf{y}) = \frac{q}{(1-q)} \frac{f_1(\mathbf{y})}{f_2(\mathbf{y})}$.

Se a estatística ELO_i for positiva, indica que a propensão para alocar uma futura observação y em π_1 é menor depois de eliminada a observação i .

Para medir o efeito da observação i no logaritmo da razão das probabilidades estimadas amostral, o autor considera o logaritmo da razão das probabilidades estimadas relativo amostral, definido como :

$$LORAM_i = n^{-1} \sum_{j=1}^2 \sum_{k=1}^{n_j} \ln\left\{\frac{o(\mathbf{x}_{jk})}{o_{(i)}(\mathbf{x}_{jk})}\right\}, \quad i = 1, \dots, n_1.$$

4. Observações influentes em análise discriminante quadrática

Partindo de duas populações normais p -variadas com vetor de médias μ_i e matriz de covariância Σ_j , e considerando amostras aleatórias dessas duas populações, sejam \bar{x}_j , $\hat{\Sigma}_j$, $j=1,2$, os estimadores não viciados dos vetores de médias e matrizes de covariância.

Assumindo probabilidades *a priori* e custos de classificação incorreta iguais para as duas populações, a regra discriminante quadrática aloca uma observação x_0 (de origem desconhecida) para o grupo com maior estimativa da densidade, ou seja aloca x_0 em π_1 se

$$(\mathbf{x}_0 - \bar{\mathbf{x}}_1)' \hat{\Sigma}_1^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_1) + \log |\hat{\Sigma}_1| < (\mathbf{x}_0 - \bar{\mathbf{x}}_2)' \hat{\Sigma}_2^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_2) + \log |\hat{\Sigma}_2| \quad (4.1)$$

e caso contrário, aloca x_0 em π_2 .

Uma vez que a regra discriminante quadrática (4.1) depende da função densidade, analisa-se o efeito das observações nas densidades estimadas.

Considerando-se a eliminação da observação i do grupo 1, têm-se as estimativas $\bar{x}_{1(i)}$ e $\hat{\Sigma}_{1(i)}$, sem essa observação, produzindo a densidade $\hat{f}_{1(i)}$, normal multivariada com média $\bar{x}_{1(i)}$ e matriz de covariância $\hat{\Sigma}_{1(i)}$. Nesta análise, as estimativas de parâmetros da população 2 não são afetadas.

Fung (1996-a) propõe o uso de medidas que levem em conta a estrutura da análise discriminante quadrática, ou seja, a probabilidade estimada que uma observação

x pertença ao grupo j determinada por $\hat{p}_j(\mathbf{x}) = \frac{\hat{f}_j(\mathbf{x})}{\hat{f}_1(\mathbf{x}) + \hat{f}_2(\mathbf{x})}$, para $j=1,2$.

Estuda-se o efeito da observação i nessa probabilidade através do cálculo de

$$E_{\mathbf{x}} \left[\hat{p}_j(\mathbf{x}) - \hat{p}_{j(i)}(\mathbf{x}) \right]^2 \quad j = 1, 2,$$

em que $\hat{p}_{j(i)}(\mathbf{x})$ é a probabilidade estimada que uma observação \mathbf{x} pertença ao grupo j depois de eliminada a observação i . Calculando-se essa esperança com base na distribuição empírica, define-se a medida de diagnóstico de probabilidade quadrática relativa:

$$PQR_i = \frac{1}{n} \sum_{j=1}^2 \sum_{k=1}^{n_j} \left[\hat{p}_1(\mathbf{x}_{jk}) - \hat{p}_{1(i)}(\mathbf{x}_{jk}) \right]^2, \quad i = 1, 2, \dots, n_1.$$

Para medir a influência da i -ésima observação no logaritmo da razão das probabilidades estimadas, a medida proposta por Fung (1996-a) é:

$$LOQR_i = \frac{1}{n} \sum_{j=1}^2 \sum_{k=1}^{n_j} \left[\log \left\{ \frac{\hat{p}_1(\mathbf{x}_{jk})}{\hat{p}_2(\mathbf{x}_{jk})} \right\} - \log \left\{ \frac{\hat{p}_{1(i)}(\mathbf{x}_{jk})}{\hat{p}_{2(i)}(\mathbf{x}_{jk})} \right\} \right]^2$$

que pode ser escrita como

$$LOQR_i = \frac{1}{n} \sum_{j=1}^2 \sum_{k=1}^{n_j} R_j^2(\mathbf{x}_{jk})$$

em que

$$2R_j(\mathbf{x}_{jk}) = -\log |\hat{\Sigma}_1| - (\mathbf{x}_{jk} - \bar{\mathbf{x}}_1)' \hat{\Sigma}_1^{-1} (\mathbf{x}_{jk} - \bar{\mathbf{x}}_1) + \log |\hat{\Sigma}_{1(i)}| + (\mathbf{x}_{jk} - \bar{\mathbf{x}}_{1(i)})' \hat{\Sigma}_{1(i)}^{-1} (\mathbf{x}_{jk} - \bar{\mathbf{x}}_{1(i)})$$

5. Observações influentes em análise discriminante para g grupos

Na análise discriminante múltipla, Fung (1995-b) sugeriu quatro medidas de diagnóstico para avaliar o quanto as estimativas da probabilidade de uma observação de origem desconhecida pertencer a um certo grupo podem ser afetadas pelas observações.

Seja \mathbf{x}_{jk} , $j = 1, 2, \dots, g$, $k = 1, \dots, n_j$ uma amostra de tamanho $n = n_1 + n_2 + \dots + n_g$ selecionada aleatoriamente de g populações, de modo que \mathbf{x}_{jk} tem distribuição normal

p -variada com média μ_i e matriz de covariância Σ . A média amostral do grupo j é denotada por \bar{x}_j e a estimativa da matriz de covariância é

$$\hat{\Sigma} = \sum_j \sum_k (x_{jk} - \bar{x}_j)(x_{jk} - \bar{x}_j)' / (n - g).$$

Se q_j é a probabilidade *a priori* de que uma observação x_0 de origem desconhecida venha da população j , $j = 1, 2, \dots, g$, a estimativa da probabilidade de que x_0 pertença ao grupo j é:

$$\hat{p}_j(x_0) = \frac{q_j \hat{f}_j(x_0)}{\sum_j q_j \hat{f}_j(x_0)} \quad j = 1, \dots, g$$

ou seja

$$\hat{p}_j(x_0) = \frac{q_j \exp[-(x_0 - \bar{x}_j)' \hat{\Sigma}^{-1} (x_0 - \bar{x}_j) / 2]}{\sum_j q_j \exp[-(x_0 - \bar{x}_j)' \hat{\Sigma}^{-1} (x_0 - \bar{x}_j) / 2]} \quad j = 1, 2, \dots, g. \quad (5.1)$$

A importância dessa quantidade deve-se ao fato que, quando os custos de classificação incorreta são iguais, a regra de classificação é obtida alocando-se x_0 no grupo j tal que $\hat{p}_j(x_0)$ é máxima. Quando os custos não são iguais, aloca-se x_0 na população π_j com $\sum_{\substack{i=1 \\ i \neq j}}^g q_i \hat{p}_i(x) c(j/i)$ mínima.

Para determinar elementos influentes no cálculo de $\hat{p}_j(x)$, utiliza-se novamente o método da omissão de observações. Se x_{ri} , i -ésima observação do grupo r , $r = 1, 2, \dots, g$,

$i = 1, 2, \dots, n_r$ é a observação eliminada da amostra, a média do r -ésimo grupo fica:

$$\bar{x}_{r(ri)} = \sum_{\substack{j=1 \\ j \neq i}}^{n_r} \frac{x_{rj}}{n_r - 1}$$

e as demais médias não são afetadas. A matriz de covariância

resultante é denotada como $\hat{\Sigma}_{(ri)}$.

Após a retirada de x_{ri} , a probabilidade estimada, $\hat{p}_j(x_0)$, é calculada através da

expressão (5.1) usando as estimativas para a amostra reduzida. Indicando esta estimativa por $\hat{p}_{j(r)}(\mathbf{x})$, a influência da observação retirada é calculada como

$\hat{p}_j(\mathbf{x}_0) - \hat{p}_{j(r)}(\mathbf{x}_0)$ e com base nela, Fung (1995-b) propõe as seguintes medidas de diagnóstico:

a) diferença absoluta das probabilidades estimadas

$$DAP_{ri} = \sum_{j=1}^g \sum_{k=1}^{n_j} \frac{|\hat{p}_j(\mathbf{x}_{jk}) - \hat{p}_{j(r)}(\mathbf{x}_{jk})|}{n}, \quad r = 1, 2, \dots, g, \quad i = 1, 2, \dots, n_r;$$

b) diferença quadrática das probabilidades estimadas

$$DQP_{ri} = \sum_{j=1}^g \sum_{k=1}^{n_j} \frac{[\hat{p}_j(\mathbf{x}_{jk}) - \hat{p}_{j(r)}(\mathbf{x}_{jk})]^2}{n}, \quad r = 1, 2, \dots, g, \quad i = 1, 2, \dots, n_r;$$

c) diferença absoluta dos logaritmos das probabilidades estimadas

$$DALP_{ri} = \sum_{j=1}^g \sum_{k=1}^{n_j} \frac{|\log \hat{p}_j(\mathbf{x}_{jk}) - \log \hat{p}_{j(r)}(\mathbf{x}_{jk})|}{n}, \quad r = 1, 2, \dots, g, \quad i = 1, 2, \dots, n_r; \text{ e}$$

d) diferença quadrática dos logaritmos das probabilidades estimadas

$$DQLP_{ri} = \sum_{j=1}^g \sum_{k=1}^{n_j} \frac{[\log \hat{p}_j(\mathbf{x}_{jk}) - \log \hat{p}_{j(r)}(\mathbf{x}_{jk})]^2}{n}, \quad r = 1, 2, \dots, g, \quad i = 1, 2, \dots, n_r.$$

Na definição destas medidas, foi usado o estimador não viciado para Σ . No entanto, Fung (1995-b) verificou que, se usadas outras estimativas como a de máxima verossimilhança ou estimativas Bayesianas, os resultados seriam similares.

Em um estudo de simulação, Fung (1995-b) observou que as medidas quadráticas *DQP* e *DQLP* fornecem resultados similares entre si e melhores do que as medidas absolutas *DAP* e *DALP*, e as medidas *DALP* e *DQLP*, que utilizam logaritmos, são superiores. Sendo assim, a medida *DQLP* seria preferível dentre as quatro medidas propostas.

6. Aplicação das medidas em um conjunto de dados

Nesta seção, será feita uma aplicação das medidas apresentadas a um conjunto de dados reais. Para esse fim, foi utilizado um subconjunto do conjunto de dados obtido no Centro de Estatística Aplicada - CEA do Instituto de Matemática e Estatística da Universidade de São Paulo - IME-USP. Tratam-se dos dados referentes ao Relatório de Análise Estatística (RAE-CEA-9415) sobre o projeto: Biologia de uma comunidade de marsupiais e roedores, em floresta atlântica de montanha, no Parque Estadual da Serra do Tabuleiro, Santa Catarina, Brasil (Leite, Singer e Lourenço, 1994).

Em cada animal capturado, foram medidas 26 características. Na análise discriminante feita no projeto, foram testados oito conjuntos de funções discriminantes com suas respectivas variáveis. O modelo adotado foi o que apresentou melhor relação custo/benefício, pois utilizava apenas quatro medidas, com 1% de animais classificados incorretamente, e se baseia na informação de 88 animais, 13 do grupo 1, de marsupiais, e 75 do grupo 2, de roedores. As quatro variáveis são comprimento total (mm), comprimento da cauda (mm), comprimento do pé (mm) e comprimento do pelo do dorso (mm).

Assumindo que os dados têm distribuição normal multivariada, a análise de diagnóstico presente em Reigada (2005) iniciou-se com o teste de homogeneidade de matrizes de covariâncias (Morrison, 1976, pg. 252) para determinar a medida mais adequada ao conjunto de dados. Toda a programação foi feita no *software S-plus*.

Uma vez que o teste apontou para matrizes de covariância diferentes, utilizou-se primeiramente a medida $LOQRG_i$, para análise discriminante quadrática:

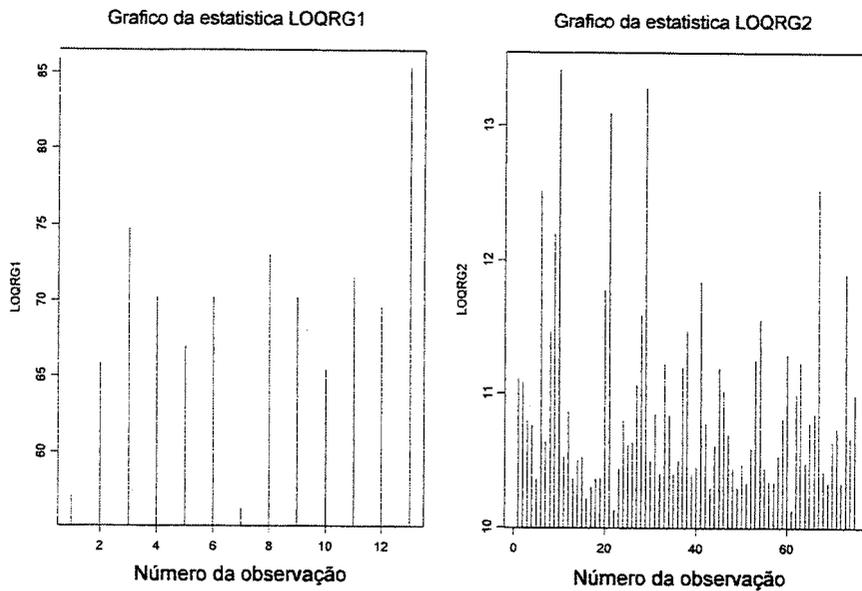
$$LOQRG_i = \frac{1}{n_1} \sum_{k=1}^{n_1} \left[\log \hat{f}_1(\mathbf{x}_{1k}) - \log \hat{f}_{1(i)}(\mathbf{x}_{1k}) \right]^2 = \frac{1}{n_1} \sum_k R_1^2(\mathbf{x}_{1k})$$

em que

$$2R_1(\mathbf{x}_{1k}) = -\log \left| \hat{\Sigma}_1 \right| - (\mathbf{x}_{1k} - \bar{\mathbf{x}}_1)' \hat{\Sigma}_1^{-1} (\mathbf{x}_{1k} - \bar{\mathbf{x}}_1) + \log \left| \hat{\Sigma}_{1(i)} \right| + (\mathbf{x}_{1k} - \bar{\mathbf{x}}_{1(i)})' \hat{\Sigma}_{1(i)}^{-1} (\mathbf{x}_{1k} - \bar{\mathbf{x}}_{1(i)}).$$

Tal medida apontou as observações de números 3, 8 e 13 do grupo 1 e 6, 9, 10, 20, 21, 29, 41, 67 e 73 do grupo 2 como as mais influentes no logaritmo da razão das probabilidades estimadas, conforme Figura 6.1.

Figura 6.1 - Gráficos da estatística LOQRG

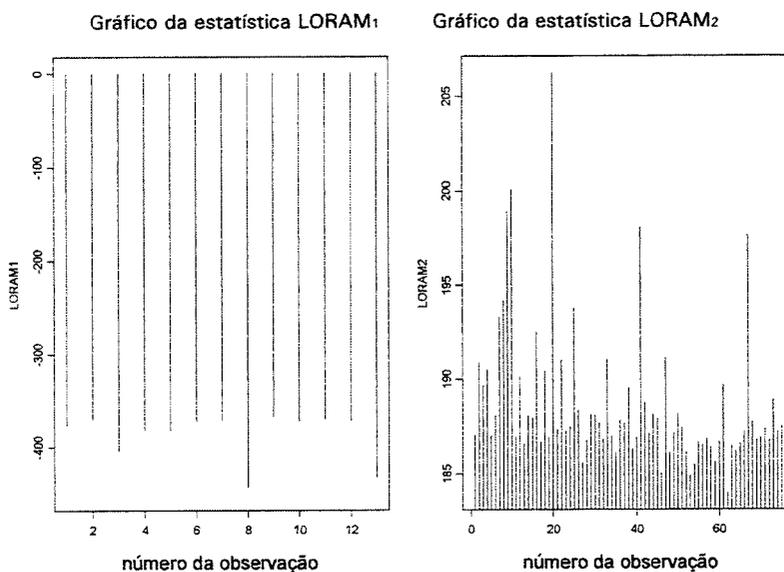


Apesar do teste de homogeneidade de matrizes de covariâncias indicar matrizes de covariâncias diferentes, para efeito de ilustração, foi verificado também o desempenho das demais medidas. As Figuras 6.2 a 6.6 são os gráficos dos valores destas medidas de diagnóstico para ambos os grupos.

- $$LORAM_i = n^{-1} \sum_{j=1}^2 \sum_{k=1}^{n_j} \ln \left\{ \frac{o(\mathbf{x}_{jk})}{o_{(i)}(\mathbf{x}_{jk})} \right\} \text{ medida Bayesiana}$$

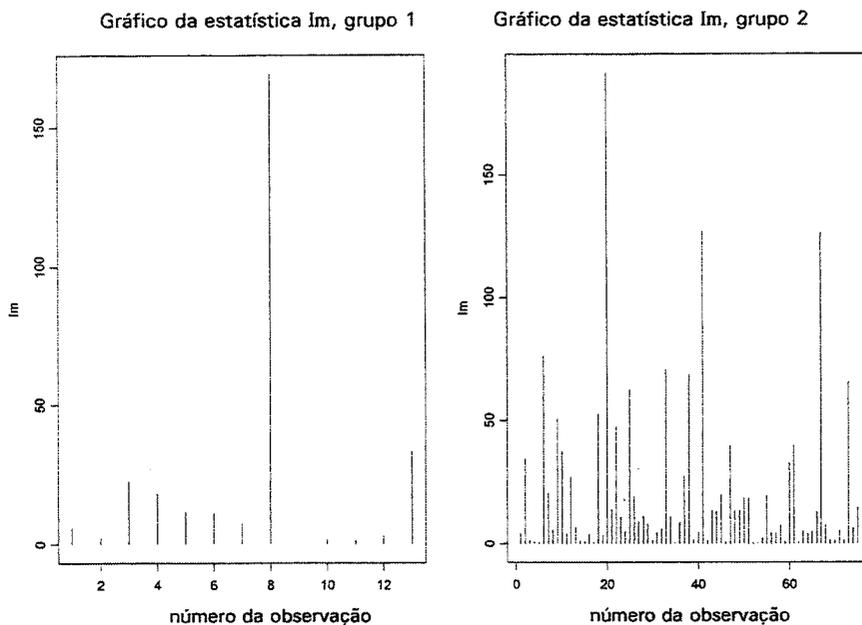
em que $\ln(o(\mathbf{y})) = \ln \frac{q}{(1-q)} + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} \mathbf{y} - \frac{1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2),$

Figura 6.2 - Gráficos da estatística LORAM para o grupo 1 e o grupo 2



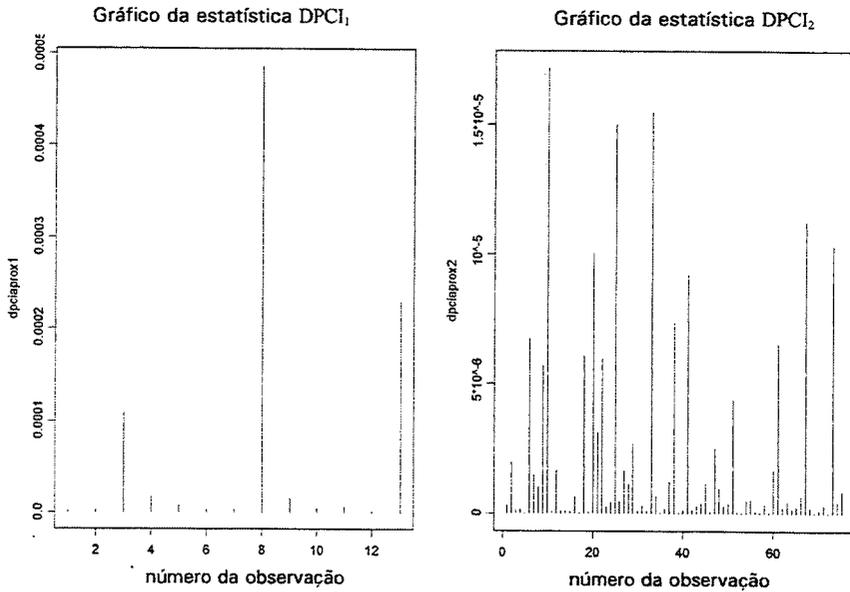
- $\hat{I}_m(\mathbf{x}; \Delta^2) = w_1 (\hat{\phi} - w_1^{-1})^2$

Figura 6.3 - Gráfico da estatística $\hat{I}_m(\mathbf{x}; \Delta^2)$



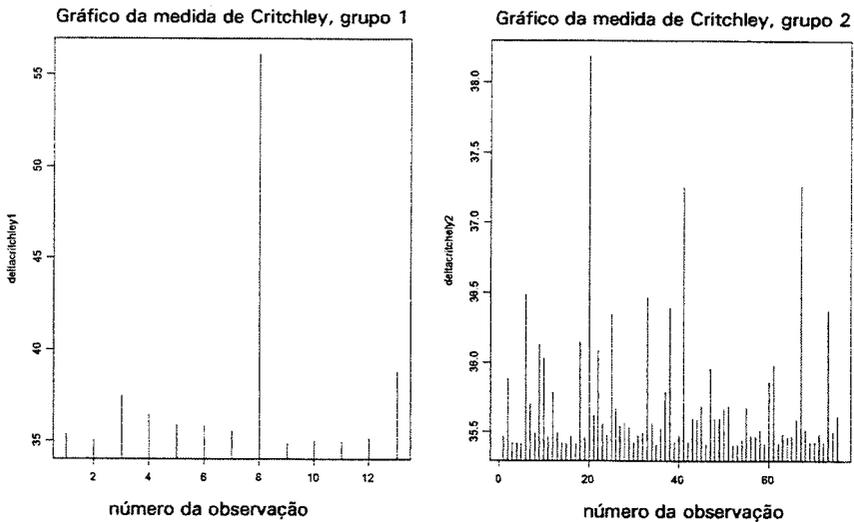
- $$DPCI_i \cong \frac{\Phi\left(-\frac{1}{2}D\right)}{4D(n_1-1)^2} \left[(1-w_1\hat{\phi}_i)^2 (a_i^2 - \hat{\phi}_i^2 / D^2) + \frac{1}{4}\hat{\phi}_i^2 \right],$$

Figura 6.4 – Gráfico da estatística $DPCI$ para o grupo 1 e o grupo 2



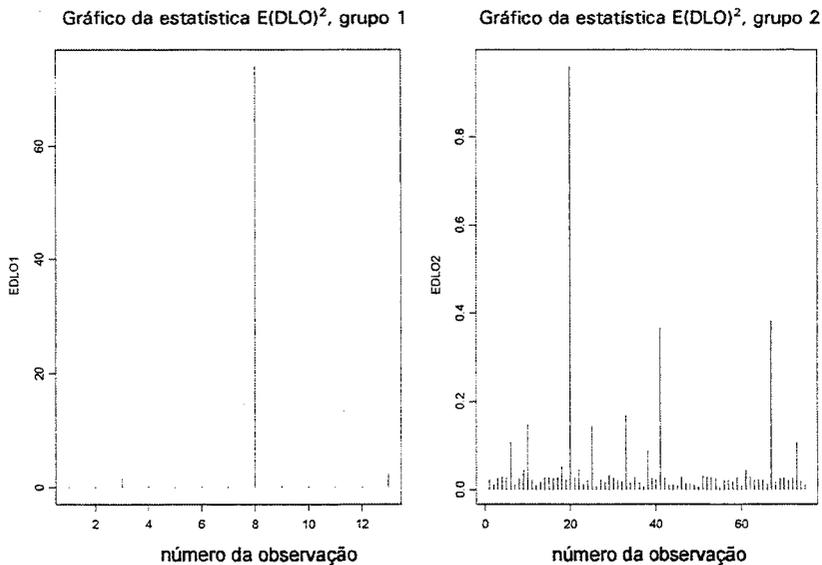
- $$\hat{\Delta}_{1(i)}^2 = \left(\frac{n_1 + n_2 - 1}{n_1 + n_2} \right) \left[\hat{\Delta}^2 - \frac{1}{w_1(n_1 - 1)} + \frac{\{e(\mathbf{x}_{1i}) - w_1^{-1}\}^2}{w_1^{-1}(n_1 - 1) - \hat{\alpha}_1^2(\mathbf{x}_{1i})} \right],$$

Figura 6.5 – Gráfico da medida $\Delta_{1(i)}^2$



- $E(DLO_i)^2 = (w_1B_1 + w_2B_2)^2 + V$,

Figura 6.6 – Gráfico da estatística $E(DLO)^2$



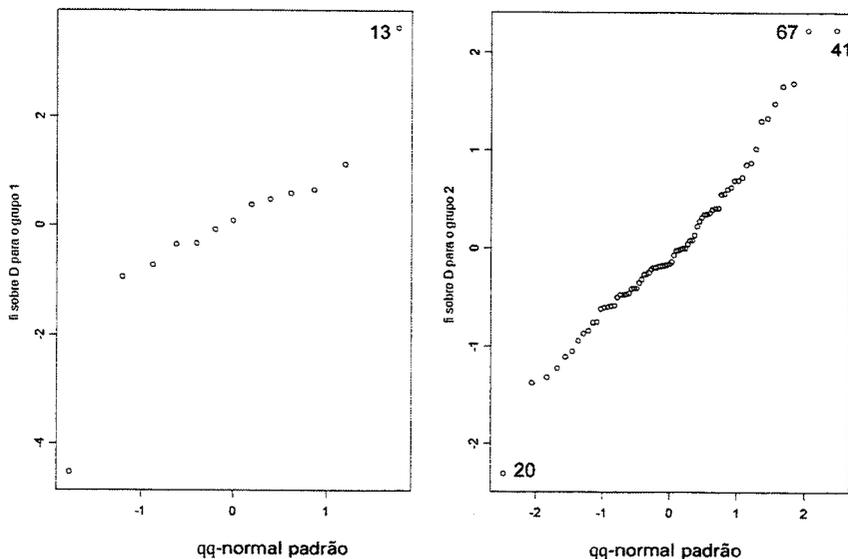
Analisando os gráficos, selecionou-se as observações que apresentaram os maiores valores para as medidas em relação às demais, dentro de cada grupo.

Nota-se que, para o grupo 1, as observações de número 8 e 13 foram apontadas por todas as medidas. Já para o grupo 2, as observações 20 e 67 foram indicadas em todas as medidas. No grupo 1, destacam-se as observações 3, 8 e 13, e no grupo 2, as observações 6, 20, 25, 33, 38, 41, 67 e 73.

Complementando a análise gráfica, são apresentados gráficos do tipo *qq plot*, que também apontam as observações discrepantes de ambos os grupos. Apresenta-se na

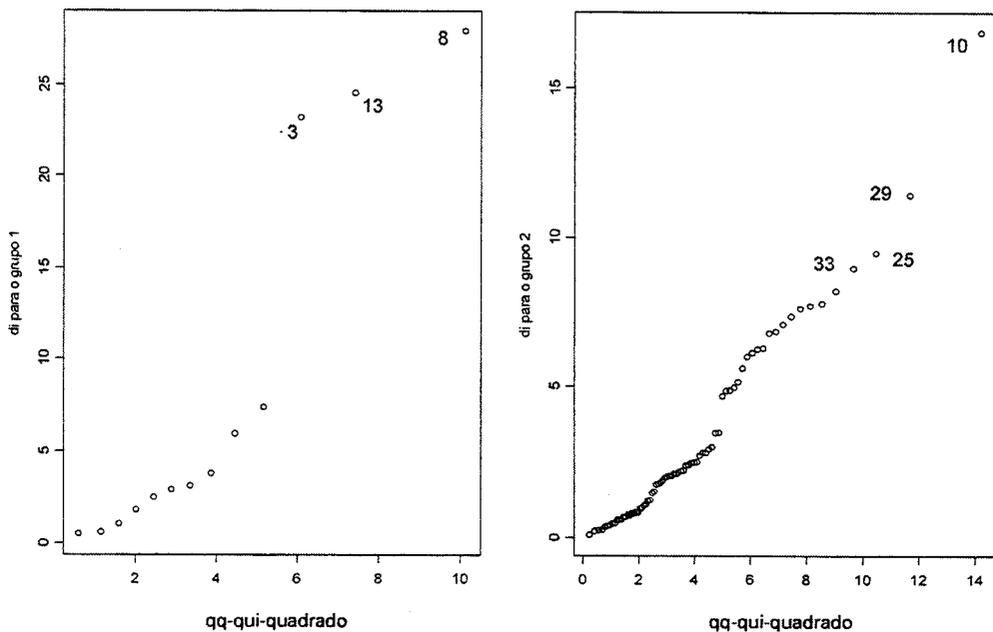
Figura 6.7 os valores da estatística $\frac{\hat{\phi}_i}{D}$ contra os quantis da distribuição normal padrão.

Figura 6.7 – Gráfico da estatística $\frac{\hat{\phi}_i}{D}$ contra os quantis da distribuição normal padrão



No gráfico da Figura 6.8, observa-se os valores da estatística d_i^2 contra os quantis da distribuição chi-quadrado.

Figura 6.8– Gráfico da estatística d_i^2 contra quantis da distribuição chi-quadrado



Finalizando este estudo, foram feitas análises discriminantes linear e quadrática para o conjunto de dados, através do *software* Minitab.

A Tabela A-1 indica as observações que se mostraram mais influentes e os correspondentes valores de cada medida de influência.

Na análise discriminante quadrática para o conjunto completo dos dados, a proporção de classificação correta foi de 0,989, sendo apontada apenas a observação 8 do grupo 1, classificada incorretamente, com probabilidade 0,33 de ser do grupo 1 e 0,67 de pertencer ao grupo 2. Na análise discriminante linear, a proporção correta de classificação foi de 0,989, apontando também a observação do grupo 1 como classificada incorretamente.

Procedeu-se à análise discriminante na ausência dos pontos 8 do grupo 1 e 20, 41 e 67 do grupo 2, por serem os mais discrepantes.

Eliminando-se somente a observação 8 do grupo 1 tanto na análise discriminante quadrática, quanto na análise discriminante linear, obteve-se uma proporção de 100% de classificação correta. A distância de Manalanobis, que anteriormente era 35,8264, passou para 56,6779, fato que, na análise discriminante linear, implica numa menor probabilidade de classificação incorreta. Com a retirada da observação 20 do grupo 2, a proporção de classificação correta é 0,989 tanto na análise linear, quanto na quadrática, (observação 8 do grupo 1 classificada incorretamente). A distância de Mahalanobis passou para 37,42, valor este que difere pouco da distância D^2 do conjunto de dados completo, que é de 35,8264. Excluindo-se a observação 41 do grupo 2, verificam-se as mesmas proporções de classificação correta obtidas quando da retirada da observação 20 do grupo 2, o mesmo ocorrendo na ausência da observação 67 do grupo 2.

Excluída a observação 41 do grupo 2, a distância de Mahalanobis é de 38,0738 e excluída a observação 67 do grupo 2, essa distância é de 38,0813.

Uma vez que a observação 10 do grupo 2 foi apontada pela medida *LOQRG* (quadrática) e apontada também por algumas medidas (Bayesiana, *DPCI*), foi feita a análise discriminante quadrática e linear sem esse ponto. Observa-se que a distância de Mahalanobis nessa situação é de 36,5536 que é bem próxima da distância de Mahalanobis para o conjunto de dados completo, e a única observação classificada incorretamente é a de número 8 do grupo 1.

Observou-se também os elementos 21 e 29 do grupo 2, apontados pela medida *LOQRG* como mais influentes. Na análise discriminante linear realizada na ausência desses pontos, obteve-se distância de Mahalanobis de 35,9158 e 35,4021, respectivamente, valores estes bem próximos de 35,8264 que é a distância de Mahalanobis para o conjunto completo dos dados. As funções lineares discriminantes também foram próximas àquelas para o conjunto completo, e tanto na análise discriminante quadrática, quanto na análise discriminante linear é apontada apenas a observação 8 como classificada incorretamente.

Nota-se que os três pontos indicados pela estatística *LOQRG* como os mais influentes, casos 10, 21 e 29 do grupo 2, são aqueles que apresentam altos valores para d_i^2 , no entanto procedendo-se à análise discriminante na ausência desses pontos, as alterações não são muito significativas.

Examinando a observação 8 do grupo 1, verifica-se que apresenta comprimento total 155 mm, comprimento da cauda 57mm, comprimento do pé 17mm e comprimento do pelo 4 mm. Se excluído esse caso do grupo 1, os valores mínimos para essas variáveis seriam 250mm, 148mm, 20mm e 8 mm respectivamente. Portanto, todos os valores das variáveis da observação 8 do grupo 1 estão abaixo desses valores mínimos, identificando-a claramente como um *outlier*.

No grupo 2, os valores mínimos são de 130mm, 46mm, 17mm e 5 mm, sendo que a observação 8, a princípio, estaria melhor classificada nesse grupo. Sugere-se então que se verifique as outras variáveis medidas nesse elemento a fim de explicar o porquê de sua inclusão no grupo 1, e questionar se essa classificação é realmente correta, uma vez que sua permanência no grupo 1 causa grande impacto não só na distância de Mahalanobis como também na probabilidade de classificação incorreta.

Deste exemplo, pode-se concluir que as medidas de diagnóstico apresentadas têm o potencial de apontar observações atípicas. Por outro lado, a escolha das medidas de discrepância depende do objetivo inferencial e também das características do problema apresentado.

Considerando o estudo de todas as medidas apresentadas no presente trabalho, sugere-se o uso da medida $\hat{\Delta}_{(i)}^2$ de Critchley e Vitiello (1991):

$$\hat{\Delta}_{1(i)}^2 = \left(\frac{n_1 + n_2 - 1}{n_1 + n_2} \right) \left[\hat{\Delta}^2 - \frac{1}{w_1(n_1 - 1)} + \frac{\{e(\mathbf{x}_{1i}) - w_1^{-1}\}^2}{w_1^{-1}(n_1 - 1) - \hat{\alpha}_1^2(\mathbf{x}_{1i})} \right],$$

uma vez que ela fornece diretamente a distância de Mahalanobis quando da eliminação da observação, e, conseqüentemente, pode-se avaliar também a alteração da probabilidade de classificação incorreta, visto que $PCI = \Phi\left(-\frac{\Delta}{2}\right)$.

7. Estudo de simulação

Finalizando, foi realizado um estudo de simulação com dados gerados sob a distribuição normal multivariada, testando-se as medidas de função de influência I e I_m (seção 2.1), DPCI e Critchley (seção 2.3), EDLO (seção 2.4) e LOQRG (seção 4). Foram efetuadas as análises descritas, a seguir:

Análise 1: Para o grupo 1, foram geradas 24 observações (X_1, X_2) da distribuição normal bivariada com vetor de médias (3,10) e para o grupo 2, 34, observações com vetor de médias (4,7) e matrizes de covariância iguais para ambos os grupos. Foi então acrescentado o ponto (10,3) no grupo 1 e (7,4) no grupo 2;

Análise 2: O ponto (7,4), *outlier* do grupo 2, foi substituído pelo ponto (2,11). Esse ponto acrescentado no grupo 2, a rigor não é um *outlier*, visto que suas coordenadas não ultrapassam os limites dos valores de cada variável. Além disso, feito um diagrama de dispersão de X_1 e X_2 , verificou-se que o ponto não se distanciava muito do conjunto de dados. Tal ponto não foi detectado pelas medidas;

Análise 3: Neste caso, partindo-se do conjunto de dados da análise 1, excluíram-se os outliers dos dois grupos e acrescentou-se o ponto (10,2) para o grupo 1 e (12,1) para o grupo 2;

Análise 4: Adotando-se uma nova matriz de covariância para o grupo 2, de modo a se estudar a situação em que as matrizes de covariância são desiguais, gerou-se novos conjuntos de observações para os dois grupos, mantendo-se os vetores de médias

fixados anteriormente. No grupo 1, foi acrescentado o ponto (7,4) e no grupo 2, o ponto (3,10). No entanto, da mesma forma que na Análise 2, o ponto incluído no grupo 2 não pode ser considerado um *outlier*: suas coordenadas não ultrapassam os limites dos valores de cada variável e no diagrama de dispersão, o ponto não se distancia muito do conjunto de dados. Tal ponto também não foi detectado pelas medidas;

Análise 5: Foi realizado o teste excluindo-se o *outlier* do passo anterior do grupo 2 e acrescentou-se o ponto (20,21);

Análise 6: Foram gerados conjuntos de dados com 3 variáveis, vetor de médias (10,15,5) para o grupo 1 e (20,25,15) para o grupo 2 e matrizes de covariância iguais. Acrescentou-se, então, o ponto (20, 25,15) no grupo 1 e (10,15,5) no grupo 2;

Análise 7: Repetiu-se a análise 6, com matrizes de covariância desiguais;

Análise 8: Trocou-se o *outlier* do grupo 1 pelo ponto (20,25,35); e

Análise 9: O último teste foi feito com quatro variáveis, matrizes de covariância identidade para os dois grupos e médias (1,2,0,3) para o grupo 1 e (3,-2,1,2) para o grupo 2, sendo acrescentados como *outliers* os pontos (-5,-3,-4,7) e (9,-7,4,4) para os grupos 1 e 2 respectivamente.

Em todos os testes, a medida EDLO mostrou-se mais eficiente, seguida pelas medidas de Critchley, função de influência e I_m

A Tabela 6.1 mostra quais medidas detectaram os *outliers* nos diversos testes realizados.

Tabela 6.1 – Medidas que detectaram os pontos acrescidos aos conjuntos de dados gerados na simulação¹

Análise	Critichley		Dpci		Edlo		I		Im		Loqrg quadr	
	G1	G2	G1	G2	G1	G2	G1	G2	G1	G2	G1	G2
1	x	x	x	x	x	x	x	x	x	x		x
2	x		x		x		x		x			
3	x	x	x		x	x	x	x	x	x		x
4	x		x		x		x		x		x	
5	x	x			x	x	x	x	x	x	x	x
6	x	x	x	x	x	x	x	x	x	x		x
7	x	x		x	x	x	x	x	x	x		x
8	x	x	x	x	x	x	x	x	x	x	x	x
9		x	x	x	x	x		x		x	x	x

x : foi detectado corretamente o *outlier* introduzido.

8. Considerações finais

Técnicas de diagnóstico são exaustivamente estudadas em regressão, mas pouco exploradas em análise discriminante. No presente artigo, procurou-se descrever as técnicas de diagnóstico em análise discriminante, devido à sua pouca utilização, apesar da extensa literatura existente.

Concluindo, sugere-se, como primeiro passo da análise, examinar aquelas observações identificadas como mais influentes, pois algumas delas podem se tratar de erros, devendo ser corrigidas ou descartadas. Para as outras observações influentes que permanecerem no conjunto de dados, é importante verificar como as inferências de interesse seriam alteradas se esses casos não tivessem sido observados. Sendo assim, o pesquisador poderá decidir sobre a permanência ou não desses *outliers* em seu estudo, e talvez até partir para uma análise mais aprofundada dessas observações.

A continuidade desse trabalho está sendo desenvolvida através da construção da medida baseada na “Diferença de Verossimilhança”, ainda não introduzida na literatura de diagnóstico em análise discriminante.

Todos os cálculos das medidas apresentadas estão descritos com detalhes em Reigada (2005).

Apêndice

Tabela A.1 - Observações mais influentes e respectivos valores das medidas de diagnóstico

Casos		Estatísticas							
Grupo 1	$LOQR_G$	$LORAM$	$\hat{I}(x, \Delta^2)$	$\hat{I}_m(x, \Delta^2)$	$DPCI$	$\hat{\Delta}_{1(i)}^2$	$EDLO$	$\hat{\phi}_i$	d_i^2
3	74,74	-403,40	-10,67	2,27E+01	1,09E-04	37,49	1,57	-5,64	23,17
8	73,06	-442,90	-157,23	1,69E+02	4,86E-04	56,13	73,70	-	27,93
9	70,22	-367,70	12,06	2,83E-04	1,48E-04	34,87	0,18	6,73	7,33
13	85,32	-432,30	-21,29	3,34E+01	2,29E-04	38,81	2,30	21,8	24,53
Grupo 2									
6	12,50	188,09	-44,35	76,06	6,75E-06	36,49	0,11	-8,27	4,66
8	11,45	194,18	26,20	5,52	1,02E-06	35,49	0,03	3,72	5,97
9	12,19	198,89	-18,75	50,46	5,70E-06	36,14	0,05	8,87	5,58
10	13,41	200,08	-5,65	37,37	1,72E-06	36,03	0,15	7,8	16,85
20	11,77	206,22	-159,73	191,45	1,00E-05	38,19	0,96	-	6,82
21	13,09	187,30	17,68	14,03	3,12E-06	35,62	0,00	5,23	7,76
25	10,61	193,78	-30,54	62,25	1,50E-05	36,34	0,14	-7,37	9,46
29	13,28	188,11	23,72	7,99	2,68E-06	35,53	0,03	-1,89	11,39
33	11,22	191,02	-38,91	70,62	1,55E-05	36,46	0,17	-7,93	8,95
38	11,46	189,52	-36,51	68,22	7,35E-06	36,39	0,09	10,12	6,08
41	11,84	198,06	-95,32	127,03	9,21E-06	37,26	0,37	13,38	7,05
67	12,52	197,65	-94,66	126,37	1,13E-05	37,26	0,38	13,35	7,59
73	11,89	188,87	-33,64	65,35	1,03E-05	36,37	0,11	9,93	7,68

Referências bibliográficas

- CAMPBELL, N. A. (1978). The Influence Function as an Aid in Outlier Detection in Discriminant Analysis. *Applied Statistics*, 27 nr 3: 251-258.
- CRITCHLEY, F. ; VITIELLO, C. (1991). The Influence of Observations Missclassification Probability Estimates in Linear Discriminant Analysis. *Biometrika*, 78: 677-690.
- FUNG, W. K. (1992). Some Diagnostic Measures in Discriminant Analysis. *Statistics and Probability Letters*, 13: 279-285.
- FUNG, W. K. (1995-a). Diagnostics in Linear Discriminant Analysis. *Journal of the American Statistical Association*, 90 nr 43: 952-956.
- FUNG, W. K. (1995-b). Detecting Influential Observations for Estimated Probabilities in Multiple Discriminant Analysis. *Computational Statistics & Data analysis*, 20: 557-568.
- FUNG, W. K. (1996-a). *Diagnosing Influential Analysis*. *Biometrics*, 52: 1235-1241.
- FUNG, W. K. (1996-b). *The Influence of Observations for Local Log-odds in Linear Discriminant Analysis*. *Communications in Statistics – Theory and Methods*, 25: 257-268.

- FUNG, W. K. (1998). *On the Equivalence of Two Diagnostic Measures in Discriminant Analysis*. Communications in Statistics – Theory and Methods, 27(8): 1915-1922.
- GEISSER, S. (1964). *Posterior Odds for Multivariate Normal Classifications*. Journal of the Royal Statistical Society, Ser B, 26: 69-76.
- GEISSER, S. (1966). " *Predictive Discrimination*" in *Proceedings of the International Symposium on Multivariate Analysis*. New York: Academic Press, pg 149-163.
- HAMPEL, F. R. (1974). *The Influence Curve and its Role in Robust Estimation*. Journal of the American Statistical Association, 69: 383-393.
- HEALY, M. J. R. (1968). *Multivariate Normal Plotting*. Applied Statistics 17 : 157-161.
- JOHNSON, W. (1987). *The Detection of Influential Observations for Allocation, Separation, and the Determination of Probabilities in a Bayesian Framework*. Journal of Business & Economic Statistics 5: 369-381.
- KULLBACK, S.; LEIBLER, R. A. (1951). *On Information and Sufficiency*. The Annals of Statistics, 22: 79-86.
- LEITE, J.G.; SINGER, J. M. E LOURENÇO, F.C. (1994). Relatório de análise Estatística sobre o projeto: Biologia de uma comunidade de marsupiais e roedores, em Floresta Atlântica de montanha, no Parque Estadual da Serra do Tabuleiro, Santa Catarina, Brasil. São Paulo, IME - USP, 33 p (RAE-CEA-94P11)
- MORRISON, D. F. (1976). *Multivariate Statistical Methods* (2nd ed.) New York McGraw Hill.
- REIGADA, S. M. B. (2005). Diagnóstico em Análise Discriminante. Dissertação de Mestrado. IME -USP, SP.

Abstract

This work presents the measures of diagnostics in discriminant analysis in two main situations: for two populations with matrix of covariance constant and two populations with different matrices of covariance (quadratic discriminant analysis). The Bayesian approach is still studied for the detection of influential observations. In the end of the work it is examined the measures of diagnostics for the multiple discriminant analysis when there is more than two populations. The study is complemented with the presentation of the main measures applied to a real data set.

Intervalos de confiança *Bootstrap* baseados na amostragem por conjuntos ordenados para a variância populacional de uma distribuição normal

Clécia Aparecida Garcia*
Maria Cecília Mendes Barreto**

Resumo

A amostragem por conjuntos ordenados é um delineamento amostral que produz estimadores de diversos parâmetros de interesse mais precisos do que os obtidos a partir da amostragem aleatória simples (Barnett e Moore, 1997; Yu et al., 1999; Chen, 2000). Entretanto, a estimação pontual nem sempre é suficiente sob o ponto de vista prático. Em muitas situações, a estimação intervalar é mais interessante, pois apresenta mais informação sobre o parâmetro de interesse, incorporando à estimativa pontual informações a respeito de sua variabilidade. Por outro lado, poucos trabalhos foram realizados sobre a estimação intervalar em amostragem por conjuntos ordenados (Chen, 2000; Cesário, 2001; Raminelli, 2003). Neste trabalho apresenta-se um estudo da adequabilidade de intervalos de confiança *bootstrap* padrão, *t-bootstrap*, *t-bootstrap* transformado, percentil e BCa para a variância populacional de uma distribuição normal baseados na amostragem por conjuntos ordenados, utilizando dois estimadores: Stokes (1980) e Yu et al. (1999). Resultados dos estudos por simulação indicam que os intervalos *t-bootstrap* transformado, considerando os coeficientes de confiança 90, 95 e 99%, apresentam probabilidade de cobertura estimada mais próxima da nominal. Já os intervalos BCa e percentil mostram menores amplitudes do que os demais mas com probabilidade de cobertura um pouco distante da nominal.

Palavras-chave: Amostragem por conjuntos ordenados, estimadores da variância, distribuição normal, intervalos de confiança *bootstrap*.

*Mestre em Estatística pelo Dept^o de Estatística da Universidade Federal de São Carlos – UFSCar – e-mail: cleciagarcia@yahoo.com.br

**Departamento de Estatística da Universidade Federal de São Carlos. e-mail: cbarreto@power.ufscar.br

1. Introdução

Um dos objetivos da Estatística é utilizar dados amostrais para estimar uma característica numérica desconhecida da população. Para a obtenção dos dados amostrais, existem técnicas de amostragem com propriedades importantes que permitem selecionar os indivíduos que comporão a amostra, como, por exemplo, a aleatoriedade e a independência dos dados. Essas propriedades definem técnicas eficientes para se obter resultados coerentes e garantir estimativas precisas.

Em muitos levantamentos, os indivíduos selecionados para compor a amostra são difíceis de serem mensurados ou o processo de mensuração apresenta alto custo. Quando antes da mensuração é possível se fazer uma ordenação de custo reduzido nos indivíduos selecionados, a amostragem por conjuntos ordenados (McIntyre, 1952) apresenta uma eficiência maior quando comparada a amostragem aleatória simples na estimação de diversos parâmetros, entre eles a média e variância populacionais, os quantis e a função densidade.

A amostragem por conjuntos ordenados (ranked set sampling - RSS) pode ser descrita em três etapas:

Primeiro, selecionam-se n amostras aleatórias simples, cada uma contendo n elementos; a seguir considera-se em cada amostra sua ordenação crescente através de inspeção visual ou por alguma variável concomitante; após a ordenação, mensura-se um elemento de cada amostra: na primeira amostra, o menor elemento, na segunda, o segundo menor e assim por diante, até que o maior seja mensurado na n -ésima amostra.

Este procedimento pode ser repetido m vezes, até que um total de mn^2 elementos tenham sido retirados da população, mas somente $N = mn$ tenham sido mensurados. Uma amostra obtida a partir da amostragem por conjuntos ordenados pode ser indicada por $X_{(11)1}, X_{(22)1}, \dots, X_{(nn)1}, \dots, X_{(11)m}, X_{(22)m}, \dots, X_{(nn)m}$, onde $X_{(ii)j}$ corresponde a j -ésima réplica da i -ésima estatística de ordem na n -ésima amostra, $i = 1, \dots, n$ e $j = 1, \dots, m$. A amostra por conjuntos ordenados (ACO) corresponde a um conjunto de estatísticas de ordem cujos elementos são independentes já que provêm de amostras escolhidas independentemente, porém não são identicamente distribuídos.

Diversos trabalhos trataram da estimação pontual sob o delineamento por conjuntos ordenados. Por outro lado, a estimação pontual nem sempre é suficiente sob o ponto de vista prático. Em muitas situações, a estimação intervalar é mais interessante, pois apresenta mais informação sobre o parâmetro de interesse, incorporando à estimativa pontual informações a respeito de sua variabilidade. Entretanto, poucos trabalhos foram realizados sobre a estimação intervalar em amostragem por conjuntos ordenados. Por exemplo, Chen (2000) propôs a estimação de quantis populacionais através de intervalos de confiança não-paramétricos e testes aproximados utilizando normalidade assintótica. Cesário (2001) abordou a construção e análise de intervalos de confiança via método *bootstrap* em amostragem por conjuntos ordenados para o parâmetro de locação da distribuição normal e para os parâmetros de um modelo de regressão linear simples. Raminelli (2003) também realizou estudos sobre procedimentos de estimação intervalar de um quantil populacional em amostragem por conjuntos ordenados.

Neste trabalho, apresenta-se um estudo da adequabilidade de intervalos de confiança *bootstrap* padrão, *t-bootstrap*, *t-bootstrap* transformado, percentil e BCa para a variância populacional de uma distribuição normal baseados na amostragem por conjuntos ordenados, utilizando dois estimadores: Stokes (1980) e Yu et al. (1999).

2. Desenvolvimento

Uma forma de se obter medidas de incerteza que expressem o erro cometido na estimação do parâmetro de interesse é estimar o erro padrão da estimativa pontual, $\hat{\theta}$. Entretanto, essas medidas nem sempre são obtidas analiticamente, pois podem apresentar expressões altamente complexas. Por outro lado, a construção de intervalos de confiança baseada na teoria assintótica, exige que as amostras sejam suficientemente grandes, a fim de garantir a validade das aproximações em questão, caso contrário, a eficácia do método é seriamente comprometida.

O método *bootstrap* é uma alternativa eficiente, através da qual obtém-se a estimativa do erro padrão de $\hat{\theta}$, livre de complexidades algébricas e, também, possibilita

a obtenção de intervalos de confiança sem a necessidade de pressupostos sobre a distribuição do estimador. Classificado no grupo de métodos computacionais intensivos, tem sido um instrumento alternativo e eficiente em várias situações, principalmente quando as amostras são de tamanho pequeno ou moderado. O *bootstrap* é essencialmente um método de reamostragem não-paramétrica que posteriormente foi desenvolvido na forma paramétrica. O *bootstrap* paramétrico reamostra a partir de um modelo probabilístico cujos parâmetros são estimados a partir da amostra original. O *bootstrap* não-paramétrico faz a reamostragem com reposição, a partir da amostra aleatória simples original. A estimação intervalar via método *bootstrap* é indicada quando há dificuldades analíticas de se encontrar intervalos de confiança exatos e, quando aproximados, existe a necessidade de ter grandes amostras a fim de garantir a validade das aproximações em questão.

Neste trabalho, apresenta-se um estudo da adequabilidade de intervalos de confiança *bootstrap* para a variância populacional de uma distribuição normal baseados na amostragem por conjuntos ordenados. Sendo assim, é considerado apenas o *bootstrap* paramétrico.

A reamostragem paramétrica consiste, primeiramente, em selecionar n conjuntos com n elementos cada, considerando m réplicas, gerados a partir de uma distribuição normal, cujos parâmetros são obtidos a partir da amostra por conjuntos ordenados original. Em seguida, os elementos são ordenados em cada conjunto e toma-se apenas o i -ésimo elemento ordenado do i -ésimo conjunto na j -ésima replicação, $i=1, \dots, n$ e $j=1, \dots, m$. Esse processo é repetido B vezes. Depois de simuladas as B reamostras, os intervalos de confiança *bootstrap* padrão, *t-bootstrap*, *t-bootstrap* transformado, percentil e BCa, são calculados de modo análogo à amostragem aleatória simples (Efron e Tibshirani, 1993).

Para a construção dos intervalos de confiança *bootstrap* para a variância populacional foram utilizados dois estimadores:

i) Estimador não-paramétrico da Stokes (1980)

$$\hat{\sigma}^2 = \sum_{j=1}^m \sum_{i=1}^n \frac{(X_{(ii)j} - \mu'_{rep})^2}{mn-1},$$

onde $\mu'_{rep} = \sum_{j=1}^m \sum_{i=1}^n X_{(ii)j} / (mn)$. Stokes (1980) mostrou que este estimador é assintoticamente não-viciado para a variância populacional σ_x^2 .

Supondo que a distribuição pertence à família de distribuições locação e escala (por exemplo, a distribuição normal), sabe-se que $E(X_{(ii)} - \mu_{(i)})^k = \sigma^k E(U_{(i)} - \alpha_{(i)})^k$ e que $\alpha_i = (\mu_{(i)} - \mu) / \sigma \Rightarrow \mu_{(i)} - \mu = \sigma \alpha_i$ (Garcia, 2004), onde $X_{(ii)}$ são as estatísticas de ordem i para uma amostra de tamanho n , $\mu_{(i)}$ são os valores esperados de $X_{(ii)}$, $U_{(i)} = (X_{(ii)} - \mu) / \sigma$ são as estatísticas de ordem padronizadas, α_i são as esperanças das estatísticas de ordem padronizadas em uma população com distribuição normal padronizada e, μ e σ são, respectivamente, os parâmetros de locação e escala. O resultado mostrado em (Garcia, 2004), $E(X_{(ii)} - \mu_{(i)})^k / \sigma^k = E(U_{(i)} - \alpha_{(i)})^k$, em palavras diz que quando a variável de interesse pertence à família de distribuições locação e escala, o k -ésimo momento central padronizado da r -ésima estatística de ordem, $E[X_r - \mu_{(r)}]^k / \sigma^k$, independe de μ e σ . É importante ressaltar que a média populacional de X , μ_x , não é necessariamente o parâmetro de locação, μ , e o desvio padrão populacional de X , σ_x , não é necessariamente o parâmetro de escala, σ .

De acordo com o resultado acima, a variância de $\hat{\sigma}^2$ apresentada em (Stokes, 1980), pode ser reescrita da seguinte forma

$$\begin{aligned} Var(\hat{\sigma}^2) = & \frac{m\sigma^4}{(mn-1)^2} \left\{ \left(\frac{mn-1}{mn} \right)^2 \sum_{i=1}^n E(U_{(i)} - \alpha_i)^4 + 4 \sum_{i=1}^n \alpha_i^2 E(U_{(i)} - \alpha_i)^2 \right. \\ & + 4 \left(\frac{mn-1}{mn} \right) \sum_{i=1}^n \alpha_i E(U_{(i)} - \alpha_i)^3 + \frac{4m}{(mn)^2} \sum_{i=1}^n \sum_{i < s} E(U_{(i)} - \alpha_i)^2 E(U_{(s)} - \alpha_s)^2 \\ & \left. + \frac{2(m-1) - (mn-1)^2}{(mn)^2} \sum_{i=1}^n [E(U_{(i)} - \alpha_i)^2]^2 \right\}, \end{aligned} \quad (1)$$

onde $E(U_{(i)} - \alpha_i)^k = E(X_{(ii)} - \mu_{(i)})^k / \sigma^k$. Note que esta variância é função de $(\sigma^2)^2$, ou seja, existe uma dependência da $Var(\hat{\sigma}^2)$ com relação a σ^2 ;

ii) Estimador da variância modificado de Yu et al. (1999)

Considerando que o estimador $\hat{\sigma}^2$, introduzido por Stokes (1980), é assintoticamente não-viciado, Yu et al. (1999) propuseram uma versão não-viciada que é dada por

$$\hat{\sigma}_{MS}^2 = \left(mn - 1 + \sum_{i=1}^n \frac{\alpha_i^2}{n} \right)^{-1} \sum_{j=1}^m \sum_{i=1}^n (X_{(ii)j} - \mu'_{rep})^2,$$

onde $\mu'_{rep} = \sum_{j=1}^m \sum_{i=1}^n X_{(ii)j} / (mn)$ e α_i é a esperança da i -ésima estatística de ordem padronizada em uma população com distribuição normal padronizada. Pode-se notar que a diferença entre $\hat{\sigma}^2$ e $\hat{\sigma}_{MS}^2$ está no denominador, com o acréscimo de $\sum_{i=1}^n \alpha_i^2 / n$ ao denominador $\hat{\sigma}^2$. Vale ressaltar que este estimador foi desenvolvido especialmente para a distribuição normal.

A variância do estimador, $\hat{\sigma}_{MS}^2$, obtida por Yu et al. (1999), é dada por

$$\begin{aligned} Var(\hat{\sigma}_{MS}^2) = & \frac{m\sigma^4}{\left(mn - 1 + \sum_{i=1}^n \frac{\alpha_i^2}{n} \right)^2} \left\{ \left(\frac{mn-1}{mn} \right)^2 \sum_{i=1}^n v_{4(i)} + 4 \sum_{i=1}^n \alpha_i^2 v_{2(i)} \right. \\ & + 4 \left(\frac{mn-1}{mn} \right) \sum_{i=1}^n \alpha_i v_{3(i)} + \frac{4m}{(mn)^2} \sum_{i=1}^n \sum_{i < s} v_{2(i)} v_{2(s)} \\ & \left. + \frac{2(m-1) - (mn-1)^2}{(mn)^2} \sum_{i=1}^n v_{2(i)}^2 \right\}, \end{aligned}$$

onde $v_{k(i)} = \mu_{k(i)} / \sigma^k = E(X_{(ii)} - \mu_{(i)})^k / \sigma^k$. Observe que esta variância é semelhante a obtida em (1). Novamente nota-se que a variância de $\hat{\sigma}_{MS}^2$ é uma função de $(\sigma^2)^2$, ou seja, existe uma dependência da $Var(\hat{\sigma}_{MS}^2)$ com relação σ^2 .

Segundo Davison e Hinkley (1997), quando a variância do estimador depende do parâmetro é interessante estudar uma transformação do parâmetro, $h(\hat{\theta})$, de tal modo que a variância dessa transformação, $Var[h(\hat{\theta})]$, seja aproximadamente independente de

θ . Segundo Canty (2003), o intervalo de confiança *t-bootstrap*, por exemplo, apresenta melhores resultados quando a $Var(\hat{\theta})$ não depende de θ . Sendo assim, como ambos estimadores dependem do parâmetro foi estudada uma transformação com a finalidade de melhorar o intervalo *t-bootstrap*, resultando no intervalo de confiança *t-bootstrap* transformado.

Seja $\hat{\theta}$ o estimador com variância $v(\theta)$, então a função $h(\cdot)$ que estabiliza a variância é dada por

$$h(t) = \int_b^t \frac{du}{\sqrt{v(u)}}. \quad (2)$$

A equação (2) é conhecida como a transformação estabilizadora da variância.

Uma aproximação para a variância de $h(\hat{\theta})$, segundo Davison e Hinkley (1997), é dada por

$$Var[h(\hat{\theta})] \approx \left| \frac{dh(t)}{dt} \right|_{\theta}^2 v(\theta). \quad (3)$$

Dois casos específicos são apresentados por Canty (2003): quando $v(\hat{\theta}) \propto \theta^2$ então $h(t) \propto \ln(t)$ e quando $v(\hat{\theta}) \propto \theta$ então $h(t) \propto \sqrt{t}$. Tendo encontrado $h(\cdot)$ então transforma-se $\hat{\theta}$ e aplicam-se as estimativas intervalares a essa escala.

Considere o intervalo de confiança *t-bootstrap*

$$\left[\hat{\theta} - \hat{t}_{\left(\frac{1+\gamma}{2}\right)} se(\hat{\theta}), \hat{\theta} - \hat{t}_{\left(\frac{1-\gamma}{2}\right)} se(\hat{\theta}) \right],$$

onde $se(\hat{\theta}) = \sqrt{Var(\hat{\theta})}$ é o erro padrão de $\hat{\theta}$, $t_{\left(\frac{1+\gamma}{2}\right)}$ é tal que $\# \left\{ Z^*(b) \leq t_{\left(\frac{1+\gamma}{2}\right)} \right\} / B = \left(\frac{1+\gamma}{2}\right)$,

sendo

$$Z^*(b) = (\hat{\theta}^*(b) - \hat{\theta}) / \hat{se}(\hat{\theta}^*(b)), \hat{se}(\hat{\theta}^*(b)) = \left\{ \sum_{b=1}^B \left[(\hat{\theta}^*(b) - \hat{\theta}^*(\cdot))^2 \right] / (B-1) \right\}^{1/2}$$

o erro padrão estimado para cada replicação *bootstrap*, $\hat{\theta}^*(b)$, $b = 1, \dots, B$.

Como já dito anteriormente, se a $Var(\hat{\theta})$ depender de θ , deve-se então fazer uma transformação adequada para melhorar o intervalo *t-bootstrap*. Logo, segundo Davison & Hinkley (1997), pode-se obter um intervalo de confiança *t-bootstrap* transformado, onde $\hat{\theta}$ é agora $h(\hat{\theta})$, $Var(\hat{\theta})$ é dada em (3). Assim, pode-se calcular a quantidade $Z^*(b)$, $b = 1, \dots, B$, da seguinte forma

$$Z^T(b) = \frac{h(\hat{\theta}^*(b)) - h(\hat{\theta})}{\left| \frac{dh(t)}{dt} \right|_{\hat{\theta}^*(b)} se(\hat{\theta}^*(b))}, b = 1, \dots, B, \quad (4)$$

onde $h(\hat{\theta}^*(b)) = \ln(\hat{\theta}^*(b))$, $h(\hat{\theta}) = \ln(\hat{\theta})$ e $se(\hat{\theta}^*(b)) = \sqrt{Var(\hat{\theta}^*(b))}$. O $\left(\frac{1+\gamma}{2}\right)$ -ésimo percentil de $Z^T(b)$ é estimado pelo valor $t_{\left(\frac{1+\gamma}{2}\right)}^T$ tal que

$$\# \left\{ Z^T(b) \leq t_{\left(\frac{1+\gamma}{2}\right)}^T \right\} / B = \left(\frac{1+\gamma}{2}\right).$$

Vale ressaltar, que na equação (4), foi aplicada uma transformação logarítmica, já que a variância dos estimadores $\hat{\sigma}^2$ e $\hat{\sigma}_{MS}^2$ depende de $(\sigma^2)^2$.

Finalmente, o intervalo de confiança *t-bootstrap* com transformação do parâmetro (ou ainda, *t-bootstrap* transformado) é dado por

$$\left[h^{-1} \left(h(\hat{\theta}) - t_{\left(\frac{1+\gamma}{2}\right)}^T \left| \frac{dh(t)}{dt} \right|_{\theta} \sqrt{\text{Var}(\hat{\theta})} \right); h^{-1} \left(h(\hat{\theta}) + t_{\left(\frac{1-\gamma}{2}\right)}^T \left| \frac{dh(t)}{dt} \right|_{\theta} \sqrt{\text{Var}(\hat{\theta})} \right) \right]. \quad (5)$$

Além da obtenção dos intervalos *bootstrap*, também foi calculada a probabilidade de cobertura estimada ou coeficiente de confiança estimado, $\hat{\gamma}$, definido como o número de vezes em que o verdadeiro valor do parâmetro está contido no intervalo em uma série de repetições, D , do procedimento de construção do intervalo de confiança. Em outras palavras, repetisse o procedimento de construção do intervalo de confiança D vezes e verifica-se em cada repetição se o verdadeiro valor do parâmetro está contido dentro de cada intervalo. Assim, a probabilidade de cobertura estimada, $\hat{\gamma}$, é definida como

$$\frac{\text{\#número de vezes em que o verdadeiro valor do parâmetro está contido no intervalo.}}{D}$$

Segundo Raminelli (2003), por exemplo, $\hat{\gamma}$ é definido por

$$\hat{\gamma} = \frac{\sum_{i=1}^D I(vvp \in IC)}{D},$$

onde $I(\cdot)$ é a função indicadora, vvp denota o verdadeiro valor do parâmetro e D é o número de repetições realizadas. Considere γ a probabilidade de cobertura nominal. Se $\hat{\gamma} > \gamma$, então o procedimento do intervalo de confiança é dito conservativo, caso contrário, é dito não-conservativo.

Segundo Carrasco (2002), por exemplo, a amplitude de um intervalo de confiança é outro critério para comparação de intervalos de confiança. Considerando a mesma probabilidade de cobertura, procedimentos de intervalos de confiança que possuem menores amplitudes são considerados melhores (Jeng e Meeker, 1999). Procedimentos

de intervalos de confiança conservativos tendem a ter maiores amplitudes do que os procedimentos não-conservativos.

A verificação da adequabilidade da probabilidade de cobertura estimada, também pode ser baseada no cálculo de constantes, a e b , de tal modo que a probabilidade de uma variável com distribuição binomial, com parâmetros g e γ , estar entre a e b é aproximadamente igual a 95%, ou seja,

$$P(a \leq X \leq b) = 0,95, P(X \leq b) = 0,975, P(X < a) = 0,025, \quad (6)$$

onde g é o tamanho da simulação e γ o coeficiente de confiança do intervalo.

3. Especificações e resultados

O estudo por simulação teve por base a geração de amostras por conjuntos ordenados a partir da distribuição normal. Assim, foram consideradas n amostras aleatórias simples de tamanho n , com m réplicas, geradas a partir de uma distribuição normal com parâmetros inicialmente conhecidos e fixados, $\mu = 0$ e $\sigma^2 = 1$. Seguindo o delineamento por conjuntos ordenados, os elementos gerados são ordenados em cada amostra e toma-se como elemento da amostra o i -ésimo elemento ordenado do i -ésimo conjunto na j -ésima replicação, $i = 1, \dots, n$ e $j = 1, \dots, m$, obtendo um total de $N = nm$ elementos que compõem a amostra por conjuntos ordenados, dita ACO original.

Em seguida, foram obtidas a média μ_{rep} e a variância $\hat{\sigma}^2$ (ou $\hat{\sigma}_{MS}^2$), para a ACO original.

A partir desta média e desta variância foram geradas $B = 1000$ amostras por conjuntos ordenados, também com distribuição normal, mas agora com média μ_{rep} e a variância $\hat{\sigma}^2$ (ou $\hat{\sigma}_{MS}^2$). A partir dessas amostras geradas foram construídos os intervalos de confiança *bootstrap*.

No cálculo das probabilidades de cobertura estimadas o procedimento descrito acima foi repetido $D = 1000$ vezes. Fixando $\gamma = 90, 95$, e 99% , para os 1000

intervalos obtidos foi verificado se o verdadeiro valor do parâmetro, $\hat{\sigma}^2 = 1$, estava contido em cada intervalo correspondente ao coeficiente de confiança especificado.

Além disso, a partir de (6), para $g = 1000$ e, $\gamma = 90, 95, \text{ e } 99\%$, os intervalos obtidos são, respectivamente:

0, 90 \rightarrow [881; 918];

0, 95 \rightarrow [936; 963]; e

0, 99 \rightarrow [983; 996].

Logo, se o número de intervalos que contêm o verdadeiro valor do parâmetro estiver dentro desses intervalos, considerando seus respectivos coeficientes de confiança, a probabilidade de cobertura estimada, $\hat{\gamma}$, é adequada.

A linguagem de programação utilizada foi o S-Plus, sendo que nas gerações sempre se usou a mesma semente, *set.seed(5)*.

Os tamanhos amostrais considerados foram: $N = 6$ ($n = 2$ e $m = 3$, $n = 3$ e $m = 2$), $N = 10$ ($n = 2$ e $m = 5$, $n = 5$ e $m = 2$) e $N = 15$ ($n = 3$ e $m = 5$, $n = 5$ e $m = 3$).

Nas Tabelas 1-6 é observado o número de intervalos que contêm o verdadeiro valor do parâmetro para diferentes configurações amostrais, e nas Tabelas 7-12, são observadas as médias dos limites inferior e superior. Todos os resultados foram obtidos a partir da simulação da probabilidade de cobertura para os intervalos de confiança *bootstrap* utilizando os estimadores (E) da variância populacional, $\hat{\sigma}^2$ e $\hat{\sigma}_{MS}^2$, considerando uma população $N(0, 1)$.

De acordo com a Tabela 1 tem-se que o intervalo de confiança que apresenta o melhor desempenho, em termos de probabilidade de cobertura estimada, é o *t-bootstrap* transformado. Considerando este intervalo, note que o número de intervalos que contêm o verdadeiro valor do parâmetro são iguais para os dois estimadores, considerando $\gamma = 90, 95, \text{ e } 99\%$ (Tabela 6). Isto se deu devido à transformação estabilizadora da variância, que fez com que os intervalos apresentassem os mesmos limites inferior e superior, com diferença na casa decimal de número 8, o que implicou em resultados iguais.

O segundo melhor intervalo de confiança foi o BCa seguido do intervalo percentil, apresentando probabilidades de cobertura estimadas próximas da nominal, considerando os dois estimadores e $\gamma = 90, 95, \text{ e } 99\%$.

Os intervalos *bootstrap* padrão e *t-bootstrap* apresentaram probabilidades de cobertura estimadas bem distantes da nominal. Isso se deve ao fato de que o intervalo padrão está relacionado à aproximação assintótica, e considerando que o tamanho amostral ($N = 6$) é pequeno, os intervalos foram imprecisos violando a propriedade da preservação de amplitude (Raminelli, 2003). O intervalo *t-bootstrap* pode apresentar amplitudes muito maior do que os demais, já que este método depende da estimativa do erro padrão em cada amostra *bootstrap* e além disso os dois estimadores, $\hat{\sigma}^2$ e $\hat{\sigma}_{MS}^2$, apresentam variância que depende de $(\hat{\sigma}^2)^2$ e isso causou uma alta variabilidade no intervalo.

Todos os procedimentos de intervalos de confiança *bootstrap* foram não-conservativos, apresentando probabilidade de cobertura estimada, $\hat{\gamma}$, menor que a probabilidade de cobertura nominal para os estimadores, $\hat{\sigma}^2$ e $\hat{\sigma}_{MS}^2$, e considerando $\gamma = 90, 95, \text{ e } 99\%$.

Tabela 1 - Número de intervalos que contêm o verdadeiro valor do parâmetro obtidos na simulação da probabilidade de cobertura para os intervalos de confiança *bootstrap* utilizando os estimadores (E) da variância, $\hat{\sigma}^2$ e $\hat{\sigma}_{MS}^2$, considerando uma população $N(0,1)$ para uma amostra de tamanho $N = 6$ ($n = 2$ e $m = 3$)

E	Intervalos de Confiança	Número de intervalos		
		90%	95%	99%
$\hat{\sigma}^2$	<i>bootstrap</i> padrão	821	846	879
	<i>t-bootstrap</i>	752	734	737
	<i>t-bootstrap</i> transformado	900	947	984
	<i>bootstrap</i> percentil	847	878	923
	<i>bootstrap</i> BCa	869	903	930
$\hat{\sigma}_{MS}^2$	<i>bootstrap</i> padrão	788	820	860
	<i>t-bootstrap</i>	748	758	760
	<i>t-bootstrap</i> transformado	900	947	984
	<i>bootstrap</i> percentil	816	854	903
	<i>bootstrap</i> BCa	862	892	922

Tabela 2 - Número de intervalos que contêm o verdadeiro valor do parâmetro obtidos na simulação da probabilidade de cobertura para os intervalos de confiança *bootstrap* utilizando os estimadores (E) da variância, $\hat{\sigma}^2$ e $\hat{\sigma}_{MS}^2$, considerando uma população $N(0,1)$ para uma amostra de tamanho $N=6$ ($n=3$ e $m=2$)

E	Intervalos de Confiança	Número de intervalos		
		90%	95%	99%
$\hat{\sigma}^2$	<i>bootstrap</i> padrão	852	871	906
	<i>t-bootstrap</i>	704	708	708
	<i>t-bootstrap</i> transformado	888	938	987
	<i>bootstrap</i> percentil	875	906	945
	<i>bootstrap</i> BCa	880	909	948
$\hat{\sigma}_{MS}^2$	<i>bootstrap</i> padrão	814	843	875
	<i>t-bootstrap</i>	750	753	753
	<i>t-bootstrap</i> transformado	888	938	987
	<i>bootstrap</i> percentil	835	869	917
	<i>bootstrap</i> BCa	862	902	940

Tabela 3 - Número de intervalos que contêm o verdadeiro valor do parâmetro obtidos na simulação da probabilidade de cobertura para os intervalos de confiança *bootstrap* utilizando os estimadores (E) da variância, $\hat{\sigma}^2$ e $\hat{\sigma}_{MS}^2$, considerando uma população $N(0,1)$ para uma amostra de tamanho $N=10$ ($n=2$ e $m=5$)

E	Intervalos de Confiança	Número de intervalos		
		90%	95%	99%
$\hat{\sigma}^2$	<i>bootstrap</i> padrão	853	885	927
	<i>t-bootstrap</i>	762	765	765
	<i>t-bootstrap</i> transformado	902	952	988
	<i>bootstrap</i> percentil	877	917	857
	<i>bootstrap</i> BCa	888	936	965
$\hat{\sigma}_{MS}^2$	<i>bootstrap</i> padrão	838	867	913
	<i>t-bootstrap</i>	788	791	793
	<i>t-bootstrap</i> transformado	902	952	988
	<i>bootstrap</i> percentil	853	893	946
	<i>bootstrap</i> BCa	887	931	962

Tabela 4 - Número de intervalos que contêm o verdadeiro valor do parâmetro obtidos na simulação da probabilidade de cobertura para os intervalos de confiança *bootstrap* utilizando os estimadores (E) da variância, $\hat{\sigma}^2$ e $\hat{\sigma}_{MS}^2$, considerando uma população $N(0,1)$ para uma amostra de tamanho $N = 10$ ($n = 5$ e $m = 2$)

E	Intervalos de Confiança	Número de intervalos		
		90%	95%	99%
$\hat{\sigma}^2$	<i>bootstrap</i> padrão	886	919	952
	<i>t-bootstrap</i>	741	752	755
	<i>t-bootstrap</i> transformado	900	944	986
	<i>bootstrap</i> percentil	892	938	978
	<i>bootstrap</i> BCa	890	933	977
$\hat{\sigma}_{MS}^2$	<i>bootstrap</i> padrão	844	878	930
	<i>t-bootstrap</i>	795	801	802
	<i>t-bootstrap</i> transformado	900	944	986
	<i>bootstrap</i> percentil	860	910	955
	<i>bootstrap</i> BCa	873	929	968

Tabela 5 - Número de intervalos que contêm o verdadeiro valor do parâmetro obtidos na simulação da probabilidade de cobertura para os intervalos de confiança *bootstrap* utilizando os estimadores (E) da variância, $\hat{\sigma}^2$ e $\hat{\sigma}_{MS}^2$, considerando uma população $N(0,1)$ para uma amostra de tamanho $N = 15$ ($n = 3$ e $m = 5$)

E	Intervalos de Confiança	Número de intervalos		
		90%	95%	99%
$\hat{\sigma}^2$	<i>bootstrap</i> padrão	890	913	956
	<i>t-bootstrap</i>	798	810	813
	<i>t-bootstrap</i> transformado	912	954	990
	<i>bootstrap</i> percentil	886	936	974
	<i>bootstrap</i> BCa	902	940	980
$\hat{\sigma}_{MS}^2$	<i>bootstrap</i> padrão	868	898	941
	<i>t-bootstrap</i>	823	834	839
	<i>t-bootstrap</i> transformado	912	954	990
	<i>bootstrap</i> percentil	878	917	968
	<i>bootstrap</i> BCa	898	937	976

Tabela 6 - Número de intervalos que contêm o verdadeiro valor do parâmetro obtidos na simulação da probabilidade de cobertura para os intervalos de confiança *bootstrap* utilizando os estimadores (E) da variância, $\hat{\sigma}^2$ e $\hat{\sigma}_{MS}^2$, considerando uma população $N(0,1)$ para uma amostra de tamanho $N = 15$ ($n = 5$ e $m = 3$)

<i>E</i>	Intervalos de Confiança	Número de intervalos		
		90%	95%	99%
$\hat{\sigma}^2$	<i>bootstrap</i> padrão	899	930	965
	<i>t-bootstrap</i>	791	804	808
	<i>t-bootstrap</i> transformado	900	950	987
	<i>bootstrap</i> percentil	889	941	980
	<i>bootstrap</i> BCa	900	941	980
$\hat{\sigma}_{MS}^2$	<i>bootstrap</i> padrão	874	902	944
	<i>t-bootstrap</i>	827	842	848
	<i>t-bootstrap</i> transformado	900	950	987
	<i>bootstrap</i> percentil	876	923	970
	<i>bootstrap</i> BCa	895	936	975

Tabela 7 - Médias dos limites inferior e superior, obtidos na simulação da probabilidade de cobertura para os intervalos de confiança *bootstrap* utilizando os estimadores (E) da variância, $\hat{\sigma}^2$ e $\hat{\sigma}_{MS}^2$, considerando uma população $N(0,1)$ para uma amostra de tamanho $N = 6$ ($n = 2$ e $m = 3$)

<i>E</i>	Intervalos de Confiança	Média dos limites		
		90%	95%	99%
$\hat{\sigma}^2$	<i>bootstrap</i> padrão	(-0,0824;2,1245)	(-0,2937;2,3359)	(-0,7069;2,7491)
	<i>t-bootstrap</i>	(0,7674;13,1852)	(0,7662;26,0967)	(0,7658;113,5966)
	<i>t-bootstrap</i> transf.	(0,4413;4,0625)	(0,3825;5,5734)	(0,2995;11,0366)
	<i>bootstrap</i> percentil	(0,2576;2,3619)	(0,1884;2,7294)	(0,0975;3,4926)
	<i>bootstrap</i> BCa	(0,3223;6,6772)	(0,2423;3,0928)	(0,1361;3,9150)
$\hat{\sigma}_{MS}^2$	<i>bootstrap</i> padrão	(-0,0154;1,9353)	(-0,2022;2,1221)	(-0,5674;2,4873)
	<i>t-bootstrap</i>	(0,7220;14,1571)	(0,7205;27,9868)	(0,7200;121,4049)
	<i>t-bootstrap</i> transf.	(0,4413;4,0625)	(0,3825;5,5734)	(0,2995;11,0366)
	<i>bootstrap</i> percentil	(0,2277;2,0903)	(0,1666;2,4125)	(0,0862;3,0870)
	<i>bootstrap</i> BCa	(0,3382;2,5820)	(0,2568;2,9604)	(0,1451;3,6201)

Tabela 8 - Médias dos limites inferior e superior, obtidos na simulação da probabilidade de cobertura para os intervalos de confiança *bootstrap* utilizando os estimadores (E) da variância, $\hat{\sigma}^2$ e $\hat{\sigma}_{MS}^2$, considerando uma população $N(0,1)$ para uma amostra de tamanho $N = 6$ ($n = 3$ e $m = 2$)

E	Intervalos de Confiança	Média dos limites		
		90%	95%	99%
$\hat{\sigma}^2$	<i>bootstrap</i> padrão	(-0,0625;2,3021)	(-0,2889;2,5285)	(-0,7316;2,9712)
	<i>t-bootstrap</i>	(0,8414;11,1297)	(0,8402;21,2703)	(0,8399;85,6842)
	<i>t-bootstrap</i> transf.	(0,4851;3,9457)	(0,4234;5,3227)	(0,3324;10,1333)
	<i>bootstrap</i> percentil	(0,3190;2,5872)	(0,2371;2,9654)	(0,1275;3,7852)
$\hat{\sigma}_{MS}^2$	<i>bootstrap</i> BCa	(0,3426;2,6783)	(0,2561;3,0645)	(0,1386;3,8871)
	<i>bootstrap</i> padrão	(0,0370;2,0073)	(-0,1516;2,1960)	(-0,5205;2,5649)
	<i>t-bootstrap</i>	(0,7689;12,3648)	(0,7672;23,6052)	(0,7667;94,6296)
	<i>t-bootstrap</i> transf.	(0,4851;3,9457)	(0,4234;5,3227)	(0,3324;10,1333)
	<i>bootstrap</i> percentil	(0,2658;2,1558)	(0,1976;2,4708)	(0,1062;3,1541)
	<i>bootstrap</i> BCa	(0,3708;2,5423)	(0,2827;2,8894)	(0,1571;3,5719)

Tabela 9 - Médias dos limites inferior e superior, obtidos na simulação da probabilidade de cobertura para os intervalos de confiança *bootstrap* utilizando os estimadores (E) da variância, $\hat{\sigma}^2$ e $\hat{\sigma}_{MS}^2$, considerando uma população $N(0,1)$ para uma amostra de tamanho $N = 10$ ($n = 2$ e $m = 5$)

E	Intervalos de Confiança	Média dos limites		
		90%	95%	99%
$\hat{\sigma}^2$	<i>bootstrap</i> padrão	(0,2140;1,8555)	(0,0568;2,1027)	(-0,2505;2,3201)
	<i>t-bootstrap</i>	(0,7791;5,2467)	(0,7769;8,0765)	(0,7761;20,5392)
	<i>t-bootstrap</i> transf.	(0,5361;2,6652)	(0,4779;3,2600)	(0,3900;5,0052)
	<i>bootstrap</i> percentil	(0,4025;1,9985)	(0,3294;2,2426)	(0,2165;2,7522)
$\hat{\sigma}_{MS}^2$	<i>bootstrap</i> BCa	(0,4684;2,2306)	(0,3945;2,5437)	(0,2814;3,1032)
	<i>bootstrap</i> padrão	(0,2338;1,7651)	(0,0871;1,9117)	(-0,1995;2,1984)
	<i>t-bootstrap</i>	(0,7535;1,7651)	(0,7506;8,4055)	(0,7496;21,3707)
	<i>t-bootstrap</i> transf.	(0,5361;2,6652)	(0,4779;3,2560)	(0,3900;5,0052)
	<i>bootstrap</i> percentil	(0,3755;1,8643)	(0,3073;2,0920)	(0,2020;2,5674)
	<i>bootstrap</i> BCa	(0,4749;2,2037)	(0,4012;2,4976)	(0,2872;2,9722)

Tabela 10 - Médias dos limites inferior e superior, obtidos na simulação da probabilidade de cobertura para os intervalos de confiança *bootstrap* utilizando os estimadores (E) da variância, $\hat{\sigma}^2$ e $\hat{\sigma}_{MS}^2$, considerando uma população $N(0,1)$ para uma amostra de tamanho $N = 10$ ($n = 5$ e $m = 2$)

E	Intervalos de Confiança	Média dos limites		
		90%	95%	99%
$\hat{\sigma}^2$	<i>bootstrap</i> padrão	(0,3112;1,8411)	(0,1648;1,9875)	(-0,1217;2,2740)
	<i>t-bootstrap</i>	(0,8108;3,6115)	(0,8081;5,1233)	(0,8072;11,0102)
	<i>t-bootstrap</i> transf.	(0,5777;2,2735)	(0,5210;2,6894)	(0,4325;3,8310)
	<i>bootstrap</i> percentil	(0,5100;2,0058)	(0,4316;2,2245)	(0,3050;2,6830)
	<i>bootstrap</i> BCa	(0,4934;1,9626)	(0,4165;2,1800)	(0,2941;2,6354)
$\hat{\sigma}_{MS}^2$	<i>bootstrap</i> padrão	(0,3379;1,6717)	(0,2102;1,7993)	(-0,0395;2,0491)
	<i>t-bootstrap</i>	(0,7606;3,8816)	(0,7555;5,5303)	(0,7537;11,9162)
	<i>t-bootstrap</i> transf.	(0,5776;2,2735)	(0,5210;2,6894)	(0,4325;3,8310)
	<i>bootstrap</i> percentil	(0,4446;1,7487)	(0,4356;2,1134)	(0,2659;2,3390)
	<i>bootstrap</i> BCa	(0,5126;1,9153)	(0,4356;2,1137)	(0,3108;2,5141)

Tabela 11 - Médias dos limites inferior e superior, obtidos na simulação da probabilidade de cobertura para os intervalos de confiança *bootstrap* utilizando os estimadores (E) da variância, $\hat{\sigma}^2$ e $\hat{\sigma}_{MS}^2$, considerando uma população $N(0,1)$ para uma amostra de tamanho $N = 15$ ($n = 5$ e $m = 3$)

E	Intervalos de Confiança	Média dos limites		
		90%	95%	99%
$\hat{\sigma}^2$	<i>bootstrap</i> padrão	(0,4575;1,6367)	(0,3445;1,7497)	(0,1237;1,9704)
	<i>t-bootstrap</i>	(0,7964;2,5574)	(0,7890;3,2888)	(0,7856;5,6102)
	<i>t-bootstrap</i> transf.	(0,6296;1,8847)	(0,5769;2,1408)	(0,4931;2,7635)
	<i>bootstrap</i> percentil	(0,5821;1,7420)	(0,5128;1,9015)	(0,3985;2,2268)
	<i>bootstrap</i> BCa	(0,5761;1,7390)	(0,5088;1,9092)	(0,4005;2,2583)
$\hat{\sigma}_{MS}^2$	<i>bootstrap</i> padrão	(0,4621;1,5407)	(0,3588;1,6439)	(0,1569;1,8459)
	<i>t-bootstrap</i>	(0,7678;2,6663)	(0,7574;3,4431)	(0,7515;5,8989)
	<i>t-bootstrap</i> transf.	(0,6296;1,8847)	(0,5769;2,1408)	(0,4931;2,7635)
	<i>bootstrap</i> percentil	(0,5324;1,5932)	(0,4690;1,7391)	(0,3645;2,0367)
	<i>bootstrap</i> BCa	(0,5849;1,7183)	(0,5185;1,8808)	(0,4072;2,1880)

Tabela 12 - Médias dos limites inferior e superior, obtidos na simulação da probabilidade de cobertura para os intervalos de confiança *bootstrap* utilizando os estimadores (E) da variância, $\hat{\sigma}^2$ e $\hat{\sigma}_{MS}^2$, considerando uma população $N(0,1)$ para uma amostra de tamanho $N = 15$ ($n = 3$ e $m = 5$)

E	Intervalos de Confiança	Média dos limites		
		90%	95%	99%
$\hat{\sigma}^2$	<i>bootstrap</i> padrão	(0,4058;1,6673)	(0,2850;1,7881)	(0,0488;2,0243)
	<i>t-bootstrap</i>	(0,7860;3,0060)	(0,7800;4,0408)	(0,7776;7,5727)
	<i>t-bootstrap</i> transf.	(0,6085;2,0369)	(0,5550;2,3550)	(0,4703;3,1597)
	<i>bootstrap</i> percentil	(0,5280;1,7664)	(0,4570;1,9373)	(0,3423;2,2886)
	<i>bootstrap</i> BCa	(0,5582;1,8494)	(0,4908;2,0502)	(0,3806;2,4482)
$\hat{\sigma}_{MS}^2$	<i>bootstrap</i> padrão	(0,4125;1,5922)	(0,2996;1,7051)	(0,0787;1,9260)
	<i>t-bootstrap</i>	(0,7637;3,1084)	(0,7557;4,1894)	(0,7520;7,8692)
	<i>t-bootstrap</i> transf.	(0,6085;2,0369)	(0,5550;2,3550)	(0,4703;3,1597)
	<i>bootstrap</i> percentil	(0,4937;1,6518)	(0,4273;1,8117)	(0,3200;2,1401)
	<i>bootstrap</i> BCa	(0,5642;1,8327)	(0,4973;2,0234)	(0,3863;2,3713)

Observando as Tabelas 2 e 8, note que diminuindo o número de réplicas ($m = 2$) e aumentando o tamanho da amostra ($n = 3$) não há muita diferença com relação aos resultados apresentados nas Tabelas 1 e 7.

Agora aumentando o tamanho amostral $N = 10$ ($n = 2$ e $m = 5$) (Tabelas 3 e 9), noto que há uma melhora em todos os intervalos, considerando os dois estimadores, para os diferentes coeficientes de confiança, mas os resultados são análogos aos das Tabela 1 e 7.

A Tabela 5, mostra que o intervalo de confiança *t-bootstrap* transformado apresenta probabilidade de cobertura estimada, $\hat{\gamma}$, maior que a probabilidade de cobertura nominal para $\gamma = 90$ e 95% , considerando os dois estimadores. O intervalo de confiança BCa apresenta boa probabilidade de cobertura, porém com $\hat{\gamma} < \gamma$ para $\gamma = 95$ e 99% , ou seja, é não-conservativo e o intervalo percentil é ligeiramente inferior que o intervalo BCa considerando $\hat{\sigma}^2$. Considerando o estimador $\hat{\sigma}_{MS}^2$, o método BCa continua apresentando melhor desempenho. Os intervalos padrão e *t-bootstrap* ainda continuam inferiores comparado aos demais para os dois estimadores e considerando $\gamma = 90, 95$ e 99% , a Tabela 6, mostra que diminuindo o número de réplicas ($m = 3$) e aumentando o tamanho da amostra ($n = 5$), os resultados são, em geral, inferiores aos obtidos na

Tabela 5. Logo, quando o número de réplicas é maior, os resultados são melhores. As análises são análogas às tabelas anteriores.

É importante ressaltar que o resultado dos intervalos depende muito do tamanho amostral N . Como para RSS é recomendável que a amostra, n , não seja muito grande, aumenta-se o número de réplicas, m , ao invés de n . As Tabelas 5 e 6 deixam claro que realmente é mais aconselhável aumentar o número de réplicas.

Através das Tabelas 1-12, nota-se que apenas alguns intervalos de confiança *bootstrap* são aconselháveis para a estimação da variância populacional sob RSS, tendo em vista que a ACO deve ser de tamanho pequeno e aproximações assintóticas não são viáveis.

Se comparar o número de intervalos que contêm o verdadeiro valor do parâmetro das Tabelas 6-12, com os intervalos obtidos através da distribuição binomial para $\gamma = 90, 95$ e 99% , nota-se que apenas o intervalo *t-bootstrap* transformado está adequado à probabilidade de cobertura estimada. Conforme aumenta o tamanho amostral, N , os intervalos BCa e percentil também vão ficando adequados.

Quanto às médias dos limites inferior e superior, em geral, os intervalos *t-bootstrap* e *t-bootstrap* transformado apresentam maiores amplitudes se comparados aos demais fixando $\gamma = 90, 95$ e 99% , considerando os dois estimadores e variando o tamanho amostral N .

Porém, se a avaliação for em termos da probabilidade de cobertura estimada, os melhores intervalos são, em ordem, o *t-bootstrap* transformado, BCa e percentil. Seria interessante verificar se essas propriedades (ou características) se mantêm quando existe uma má especificação do modelo, ou seja, os dados têm uma distribuição diferente da normal, por exemplo. Outro aspecto a ser futuramente considerado se refere ao número de reamostragens *bootstrap* para o intervalo de confiança BCa (Davison e Hinkley, 1997).

Em Garcia (2004) além das diferentes configurações amostrais, também foi realizado o mesmo estudo apresentado neste trabalho, porém com distribuição normal com média 5 e variância 10. Os resultados foram análogos ao da distribuição normal com média 0 e variância 1.

4. Conclusão

Os resultados dos estudos por simulação indicam que os intervalos *t-bootstrap* transformado, considerando $\gamma = 90, 95$ e 99% , apresentam probabilidade de cobertura estimada mais próxima da nominal e os intervalos BCa e percentil, mostram menores amplitudes do que os demais, mas com probabilidade de cobertura um pouco distante da nominal.

Os intervalos *bootstrap* aqui estudados são as únicas alternativas viáveis, até o momento, para a estimação intervalar da variância de uma população normal.

Referências bibliográficas

- BARNETT, V. ; MOORE, K. (1997). Best linear unbiased estimates in ranked set sampling with particular reference to imperfect ordering. *Journal of Applied Statistics*, v.24, n.6, p.697-710.
- CANTY, A. (2003). *The bootstrap and confidence intervals*. Disponível em: <www.mathstat.concordia.ca/anty/teaching/mast679t.html>. Acesso em: 10 ago.
- CARRASCO, C. G. (2002). *Comparação de procedimentos para construção de intervalos de confiança para os parâmetros do modelo poli-log-logístico na presença de censuras aleatórias*. 66f. Dissertação (Mestrado em Estatística) - Centro de Ciências Exatas e Tecnologia - Departamento de Estatística, Universidade Federal de São Carlos, São Carlos.
- CESÁRIO, L. C.(2001). *Intervalos de confiança bootstrap em amostragem por conjuntos ordenados*. 103f. Dissertação (Mestrado em Estatística)-Centro de Ciências Exatas e Tecnologia-Departamento de Estatística, Universidade Federal de São Carlos, São Carlos.
- CHEN, Z. (2000). On ranked set sample quantiles and their applications. *Journal of Statistical Planning and Inference*, v.83, p.125-135.
- DAVISON, A. C.; HINKLEY, D. V. (1997). *Bootstrap methods and their application*. New York: Cambridge University Press, 582p.
- DELL, T. R.;CLUTTER, J. L. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, v.28, p.545-555.
- EFRON, B.; TIBSHIRANI, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman e Hall, 436p.
- GARCIA, C. A. (2004). *Precisão relativa e intervalos de confiança bootstrap para a variância populacional de uma distribuição normal baseados na amostragem por conjuntos ordenados*. 127f. Dissertação (mestrado em estatística) - Centro de Ciências Exatas e Tecnologia - Departamento de Estatística, Universidade Federal de São Carlos, São Carlos.
- JENG, S. L.;MEEKER,W. Q. (1999). Comparisons of Approximate Condence Interval Procedures for Type I Censored Data. *Technometrics*, v.42, p.135-148.

- MCINTYRE, G. A. A. (1952). Method for unbiased selective sampling using ranked sets. *Australian Journal of Agricultural Research*, v.3, p.385-390.
- RAMINELLI, J. A. (2003). *Procedimentos de estimação intervalar de um quantil populacional em amostragem por conjuntos ordenados*. 81f. Dissertação (mestrado em estatística) - Centro de Ciências Exatas e Tecnologia - Departamento de Estatística, Universidade Federal de São Carlos, São Carlos.
- STOKES, S. L. (1980). Estimation of variance using judgement ordered ranked set samples. *Biometrics*, v.36, p.35-42.
- YU, PHILIP L. H.; LAM, K.; SINHA, BIMAL K. (1999). Estimation of normal variance basead on balanced and unbalanced ranked set samples. *Environmental and ecological statistics*, v.6, p.23-46.

Abstract:

The estimators based on the design of ranked set sampling are more precise than those based on the single random sampling (Barnett & Moore, 1997; Yu at al., 1999; Chen, 2000). We present in this paper a study of the performance of several bootstrap confidence intervals for two estimators (Stokes, 1980; Yu et al., 1999) of the variance of normal population under the design of ranked sets. The results of the simulation study has shown that the confidences estimated are very close to the nominal ones of t-bootstrap intervals under 90%, 95% and 99% confidence coefficients. And the BCa and percentil intervals have smallest range but confidences a little different form the nominal ones.

Keywords: ranked set sampling, variance estimators, normal distribution, bootstrap confidence intervals.

Estimação de parâmetros no ajuste de modelos teóricos a semivariogramas experimentais

*Miguel Angel Uribe-Opazo**
*Célia Rupp Kavanagh***

Resumo

Este trabalho teve como objetivo estudar o ajuste de modelos teóricos a semivariogramas experimentais, importante etapa na análise geostatística de dados espacialmente referenciados. Os estudos foram desenvolvidos a partir de processos gaussianos multivariados, simulados pelo método de decomposição de Cholesky. As simulações realizadas correspondem a "coletas de dados" em grades quadradas 10x10 (100 pontos amostrais) e grades duplas, compostas de 20 conglomerados 10x10 (2000 pontos amostrais). No ajuste dos modelos teóricos, foram utilizados e comparado o método de mínimos quadrados ordinários (OLS), o método de mínimos quadrados ponderados, com pesos diretamente proporcionais ao número de pares de pontos amostrais correspondentes a cada semivariância estimada (WLS1) e o método de mínimos quadrados ponderados, com pesos diretamente proporcionais ao número de pares de pontos amostrais e inversamente proporcionais ao quadrado das semivariâncias estimadas (WLS2). Dos resultados obtidos concluiu-se que, mesmo que a hipótese de estacionaridade de segunda ordem seja satisfeita, 100 pontos amostrais podem ser insuficientes para que os semivariogramas experimentais detectem a estrutura de dependência espacial correspondente, caso em que qualquer dos métodos de ajuste

* Endereços para correspondências: Prof. Associado do Centro de Ciências Exatas e Tecnológicas da Univ. Estadual do Oeste do Paraná – UNIOESTE – Campus de Cascavel-PR, Rua Universitária, 2069, sala 65, CEP 85 819 – 110 – Brasil – e-mail mopazo@unioeste.br.

** Mestra em Engenharia Agrícola, com Área de Concentração em Sistemas Agroindustriais da Universidade Estadual do Oeste do Paraná – UNIOESTE – Campus de Cascavel-PR, Brasil – e-mail celiakavanagh@hotmail.com.

considerados pode estimar parâmetros que diferem em muito dos verdadeiros. Para amostragens intensas (2000 pontos amostrais) e com distâncias entre os pontos amostrais adequadas à descrição da estrutura de dependência espacial correspondente, qualquer um dos métodos produziu bons resultados. Dos resultados se concluiu que a maior dificuldade no ajuste é devida a estimativas incorretas das semivariâncias, relacionadas com amostragens insuficientes.

Palavras-chave: Geoestatística, simulações, mínimos quadrados ponderados.

1. Introdução

A Geoestatística vem sendo intensamente utilizada no estudo da variabilidade espacial de propriedades físico, química e hídrica do solo e da produtividade. A Geoestatística se fundamenta em um modelo probabilístico, no qual se considera que os dados espacialmente localizados podem ser modelados por um processo estocástico, ou função de variáveis aleatórias, $\{Z(s):s \in D\}$, onde D é um subconjunto do espaço d -dimensional \mathbf{R}^d , e s varia continuamente sobre D (CRESSIE, 1988). Segundo o modelo, cada valor do atributo em estudo, amostrado em uma posição s_i , representa uma realização de uma variável aleatória e cada conjunto de valores amostrados, em posições s_1, s_2, \dots, s_{n_r} , representa uma realização incompleta de uma função de variáveis aleatórias. Se alguns parâmetros do processo estocástico podem ser estimados a partir da amostra, predições de valores do atributo em estudo, em pontos não amostrados, podem ser feitas através da krigagem para fins de construção de mapas temáticos de contorno (isolinhas) e de superfícies que podem ser utilizados em técnicas de Agricultura de Precisão. A estimação dos parâmetros é obtida, de modo geral, ajustando-se um modelo teórico a um semivariograma experimental (ISAAKS & SRIVASTAVA, 1989). Um ponto sobre o qual ainda não há consenso é qual o melhor critério a ser adotado no ajuste. Os métodos de ajuste mais freqüentemente disponibilizados por *softwares* geoestatísticos comerciais são métodos de mínimos quadrados ordinário e ponderado, embora alertas sobre suas limitações venham sendo feitos (DIGGLE & RIBEIRO JR., 2000). A decisão sobre a validade ou não de um critério de ajuste, ou da superioridade de um critério sobre outro, não pode ser tomada a partir de dados reais, uma vez que, nesse caso, não se conhece a verdadeira estrutura de dependência espacial

correspondente. Assim sendo, buscou-se, neste trabalho, simular dados espacialmente referenciados, que satisfazem as hipóteses básicas da teoria geoestatística e têm estrutura de dependência espacial conhecida e, a partir destes, estudar a problemática do ajuste e comparar diferentes critérios de atribuição de pesos nos ajustes por métodos de mínimos quadrados.

2. Materiais e métodos

O trabalho foi desenvolvido a partir de simulações de conjuntos de dados espacialmente correlacionados $\{Z(s_1), Z(s_2), \dots, Z(s_n)\}$, tais que cada conjunto representa uma realização de um processo estocástico gaussiano multivariado, isotrópico e com estrutura de dependência espacial conhecida. As simulações foram feitas pelo método de decomposição de Cholesky (CRESSIE, 1993). Segundo esse método, escolhido o vetor $\mu = (E(Z(s_1)), E(Z(s_2)), \dots, E(Z(s_n)))^T$ de ordem $n \times 1$, que representa a média do processo, e a matriz de covariância Σ , de ordem $n \times n$, $\Sigma = [C(Z(s_i), Z(s_j))]$, onde $C(Z(s_i), Z(s_j))$ representa a covariância entre as variáveis $Z(s_i)$ e $Z(s_j)$, pode-se obter o vetor Z , tal que $Z = (Z(s_1), Z(s_2), \dots, Z(s_n))^T$ através da relação

$$Z = \mu + L\varepsilon \quad (1)$$

Na Equação (1), L é uma matriz triangular inferior tal que $LL^T = \Sigma$ e $\varepsilon = (\varepsilon(s_1), \varepsilon(s_2), \dots, \varepsilon(s_n))^T$ é um vetor de variáveis aleatórias não-correlacionadas. Pode ser demonstrado (JOHNSON & WICHERN, 1982), que se $\varepsilon(s_1), \varepsilon(s_2), \dots, \varepsilon(s_n)$ são variáveis aleatórias identicamente distribuídas, com distribuição normal de média igual a zero e variância igual a um, e $E(Z(s_1)) = E(Z(s_2)) = \dots = E(Z(s_n)) = \mu$, o vetor Z , obtido através da Equação (1), tem distribuição normal multivariada com vetor de médias $\mu = \mu \mathbf{1}$ (onde $\mathbf{1}$ é um vetor de 1's de ordem $n \times 1$) e matriz de covariância Σ . A média do processo μ não influi na análise do semivariograma e, sem perda de generalidade, pode ser escolhida igual a zero. A matriz de covariância foi escolhida de modo a corresponder a um processo isotrópico, com modelo de semivariância teórico e parâmetros *nugget* (C_0), *sill* (C_1) e alcance (a) apresentados no Quadro 1.

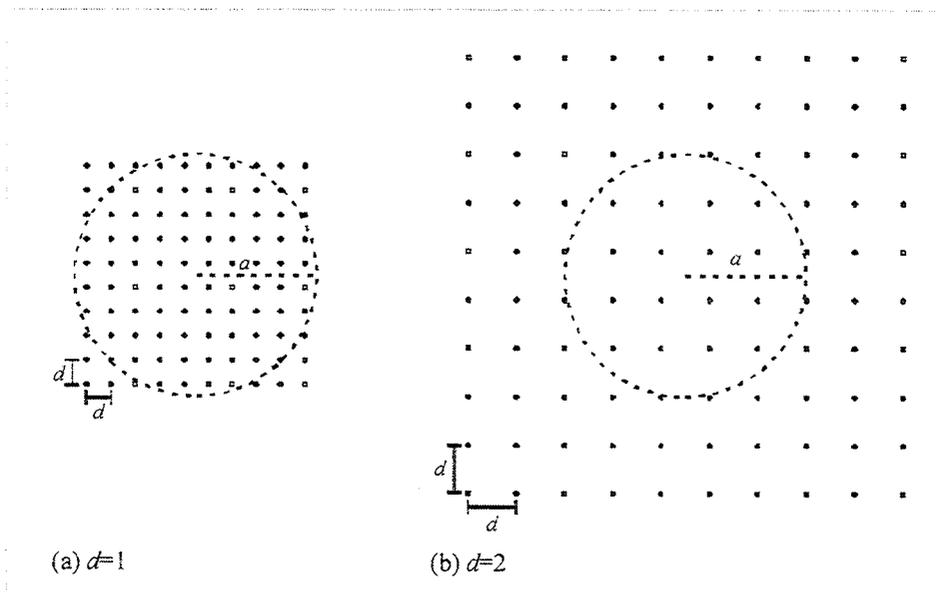
Quadro 1 - Modelos de semivariogramas dos processos simulados

Modelos Teóricos	(C_0, C_1, a)
Esférico	(0, 1, 5)
Esférico	(0.2, 0.8, 5)
Exponencial	(0, 1, 5)
Exponencial	(0.2, 0.8, 5)

Os valores $Z(s)$ foram simulados através de experimentos Monte Carlo, correspondendo a coletas de dados em nós de grades quadradas 10×10 (cem pontos amostrais). Para fins de análise geoestatística, tão importante quanto à configuração geométrica e ao tamanho da amostragem é a relação entre o alcance da dependência espacial e as distâncias entre os pontos amostrais. Por certo não se terá êxito na caracterização da estrutura de dependência espacial se as distâncias entre os pontos amostrais forem todas muito menores ou todas muito maiores que o alcance. Para favorecer a "captura" da dependência espacial dos processos simulados, duas diferentes relações entre a distância entre pontos amostrais consecutivos e o alcance a foram consideradas e são apresentadas na Figura 1. Os círculos, de raio igual ao alcance, indicam a área de influência de um ponto situado no centro da "área amostrada". O que se busca verificar, com as diferentes relações, não é qual a melhor maneira de amostrar uma determinada área (observa-se que as áreas abrangidas em (a) e (b) não são iguais). O que se pretende é comparar as qualidades dos diferentes métodos de estimação em estudo, em duas diferentes situações: quando o semivariograma experimental possui mais informações correspondentes a valores de h menores que o alcance (correspondente à grade amostral da Figura 1a), e quando possui mais informações correspondentes a valores de h maiores que o alcance (correspondente à grade amostral da Figura 1b). Na Figura 1a, a menor distância entre os pontos amostrais, d , é igual a 20% do alcance, e a maior distância entre os pontos amostrais, correspondente a pontos situados nos extremos das diagonais das áreas de amostragem, é aproximadamente igual a 2,5 vezes o alcance. Na Figura 1b, a menor distância entre

pontos amostrais é igual a 40% do alcance, e a maior distância entre pontos amostrais é aproximadamente igual a cinco vezes o alcance.

Figura 1 - Grades de amostragem correspondentes às simulações realizadas



Para cada modelo de semivariograma e vetor de parâmetros especificados no Quadro 1, foram obtidos vinte conjuntos de dados simulados em malhas 10x10 com distância entre nós consecutivos igual a uma unidade (Figura 1a) e vinte conjuntos de dados simulados em malhas 10x10 com distância entre nós consecutivos igual a duas unidades (Figura 1b). Com o objetivo de simplificar a notação, modelos esféricos com parâmetros C_0 , C_1 e a são por vezes denotados por $\text{esf}(C_0, C_1, a)$ e modelos exponenciais com parâmetros C_0 , C_1 e a por $\text{exp}(C_0, C_1, a)$. Para as componentes dos vetores utiliza-se, neste trabalho, o sistema inglês de escrita numérica. A partir de cada conjunto de dados simulados, que representa uma coleta, foi construído um semivariograma experimental $\hat{\gamma}(h)$, com a utilização do estimador de Matheron (MATHERON, 1963), segundo o qual

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(s_i + \mathbf{h}) - Z(s_i)]^2 \quad (2)$$

onde, $N(h)$ é o número de pares de pontos amostrais separados por vetores \mathbf{h} tais que $|\mathbf{h}| = h$.

Não há consenso, na literatura, a respeito do número mínimo de pares de pontos amostrais suficientes para a estimativa da semivariância experimental a cada *lag*. Segundo WOLLENHAUPT et al. (1997), JOURNEL & HUIJBREGTS (1978) recomendam um mínimo de 30 pares, enquanto WEBSTER & OLIVER (1992) recomendam um mínimo de 100 pares. Neste trabalho, buscando favorecer a acurácia das estimativas, somente foram consideradas as semivariâncias estimadas por um mínimo de 150 pares de pontos amostrais, com nenhuma tolerância no *lag*. A cada semivariograma experimental, foram ajustados os modelos esférico e exponencial, das formas:

$$\gamma(h) = \begin{cases} C_0 + C_1 \left[\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right], & \text{se } 0 \leq h \leq a \\ C_0 + C_1, & \text{se } h > a \end{cases} \quad (3)$$

e

$$\gamma(h) = C_0 + C_1(1 - e^{-3h/a}), \quad \text{se } h \geq 0. \quad (4)$$

onde, C_0 representa o efeito *nugget*, $C_0 + C_1$ o patamar e a o alcance. Os ajustes foram feitos por três métodos diferentes, aqui designados OLS, WLS1 e WLS2.

O método OLS é o método de mínimos quadrados ordinários. Segundo esse método, o vetor de parâmetros a ser escolhido é o que minimiza a expressão (5):

$$\sum_{j=1}^k [\hat{\gamma}(h_j) - \gamma(h_j, \theta)]^2 \quad (5)$$

em que,

k é o número de *lags* que constituem o semivariograma experimental;

$\hat{\gamma}(h_j)$ é a semivariância estimada correspondente ao j -ésimo *lag*;

$\gamma(h_j, \theta)$ é o modelo de semivariograma selecionado para o ajuste, dependente do vetor de parâmetros $\theta = (C_0, C_1, a)$.

O método de estimação WLS1 (McBRATNEY & WEBSTER, 1986) é um método de estimação de parâmetros de mínimos quadrados ponderados, com pesos diretamente proporcionais ao número de pares de pontos amostrais que contribuem para a semivariância estimada a cada *lag*, $N(h_j)$. Segundo esse método, o vetor de parâmetros a ser escolhido é o que minimiza a expressão (6):

$$\sum_{j=1}^k N(h_j) [\hat{\gamma}(h_j) - \gamma(h_j, \theta)]^2 \quad (6)$$

O método WLS2 (EASTMAN, 2001; MARCHANT & LARK, 2004) é um método de mínimos quadrados ponderados, com pesos diretamente proporcionais ao número de pares de pontos amostrais que contribuem para a semivariância estimada a cada *lag* e inversamente proporcionais ao quadrado da semivariância estimada. Este método de estimação assemelha-se ao proposto por CRESSIE (1985). O vetor de parâmetros a ser escolhido é o que minimiza a expressão (7):

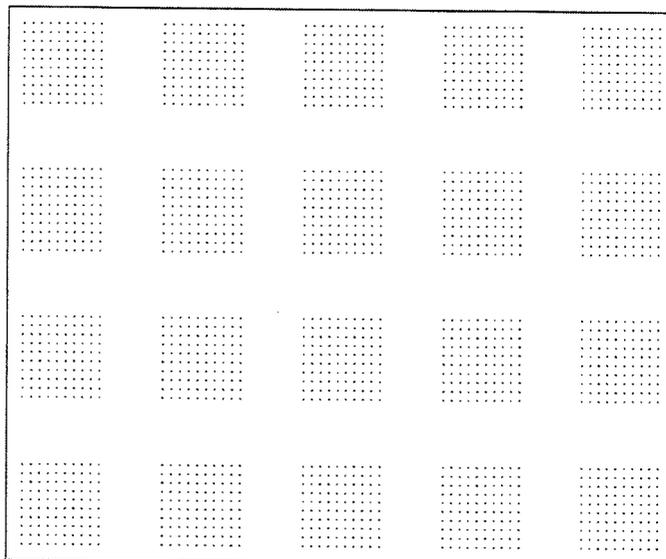
$$\sum_{j=1}^k \frac{N(h_j)}{[\hat{\gamma}(h_j)]^2} [\hat{\gamma}(h_j) - \gamma(h_j, \theta)]^2 \quad (7)$$

A estimação dos parâmetros, por qualquer um dos métodos considerados neste trabalho, consiste em resolver um problema de programação não-linear. A única restrição imposta no ajuste foi que os parâmetros estimados fossem não negativos. Uma vez que o verdadeiro vetor de parâmetros a partir do qual os dados foram simulados são conhecidos, os resultados obtidos pelos diferentes métodos puderam ser comparados e as principais dificuldades do ajuste identificadas.

Os resultados de vinte simulações em malhas 10x10 foram reunidos para simular uma intensa coleta de dados (2000 pontos amostrais), em uma grade dupla, na qual os resultados de cada uma das vinte simulações passaram a representar os valores $Z(s)$ em um pequeno conglomerado, em uma coleta em que os conglomerados estão separados uns dos outros por distâncias maiores que o alcance (Figura 2). O que possibilita atribuir significados à reunião dos conjuntos de dados simulados referentes aos pequenos conglomerados para simular a intensa coleta de dados, é a configuração de grade dupla proposta, com os conglomerados separados por distâncias maiores que o alcance, pois para distâncias maiores que o alcance a dependência espacial é igual a zero no caso

modelos de semivariogramas esféricos e desprezível no caso de modelos semivariogramas exponenciais. Os semivariogramas experimentais foram construídos considerando-se apenas semivariâncias estimadas para valores de h (*lags*) menores que a metade da distância máxima entre pontos amostrais nos conglomerados. Os ajustes dos modelos teóricos foram feitos pelos métodos OLS, WLS1 e WLS2. A matriz de covariância Σ e a matriz L , necessárias para a realização das simulações, foram obtidas com a utilização de um programa computacional.

Figura 2 - Amostragem em grade dupla



3. Resultados e discussão

Nas Figuras 3 e 4 são apresentados alguns dos semivariogramas experimentais obtidos a partir das simulações correspondentes aos modelos $esf(0, 1, 5)$ e $esf(0.2, 0.8, 5)$, em malhas 10x10. Em cada figura, a linha contínua corresponde ao modelo teórico verdadeiro a partir do qual os dados foram gerados e pontos de formas diferentes correspondem a semivariogramas experimentais obtidos de diferentes realizações. Observou-se, dos gráficos apresentados, que os valores estimados para as semivariâncias podem diferir em muito do valor $\gamma(h)$ correspondente ao modelo verdadeiro a partir do qual os dados foram gerados. Resultados similares foram obtidos

nos semivariogramas experimentais correspondentes às simulações de dados com estrutura de dependência espacial $exp(0, 1, 5)$ e $exp(0.2, 0.8, 5)$. Os resultados obtidos trazem informações a respeito do que pode acontecer no estudo de dados reais. Imagine-se que um determinado atributo do solo seja amostrado em áreas distantes entre si, mas que sejam homogêneas no que diz respeito à estrutura de dependência espacial do atributo em questão. Os semivariogramas experimentais construídos a partir de dados coletados nas diferentes áreas, podem apresentar aspectos tão diferentes entre si quanto aos dos semivariogramas de cada uma das Figuras 3 e 4.

Figura 3 - Semivariogramas experimentais obtidos de três diferentes realizações, correspondentes ao modelo $esf(0, 1, 5)$, em grades quadradas 10×10 , com distância entre nós consecutivos igual a uma unidade de comprimento (a) e igual a duas unidades de comprimento (b)

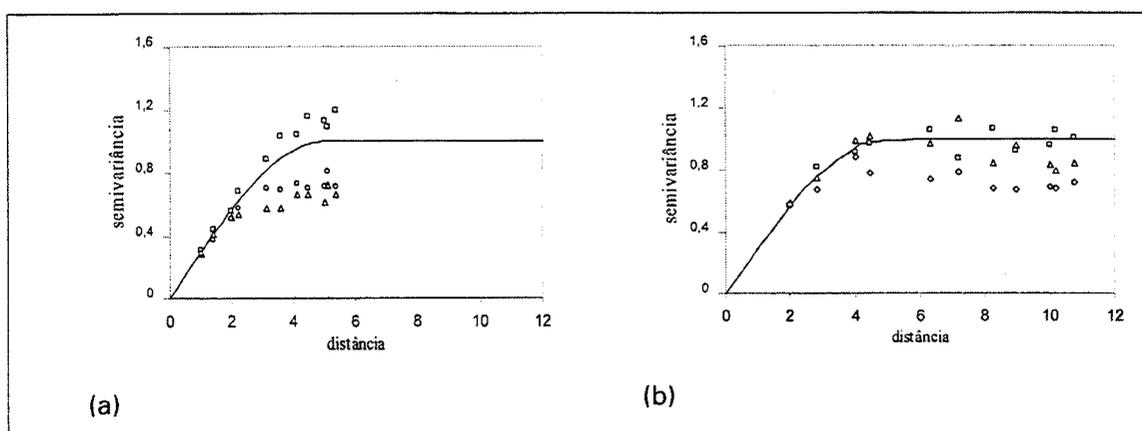
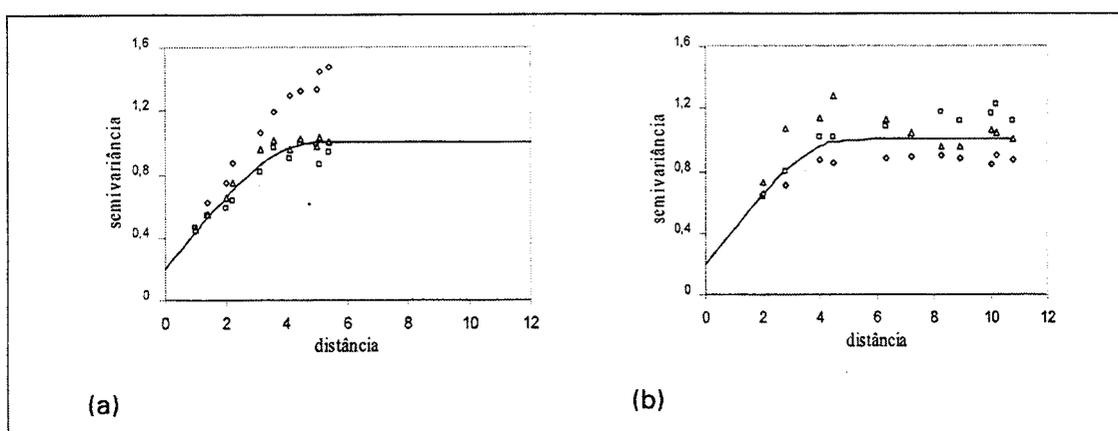


Figura 4 - Semivariogramas experimentais obtidos de três diferentes realizações, correspondendo ao modelo $esf(0.2, 0.8, 5)$, em grades quadradas 10×10 , com distância entre nós consecutivos igual a uma unidade de comprimento (a) e igual a duas unidades de comprimento (b)



Os resultados de cada conjunto de 20 realizações de processos correspondentes a modelos esférico e exponencial, com diferentes parâmetros, são resumidos nas Figuras 5 a 8. Em cada figura encontram-se representados o modelo verdadeiro, os valores médios das semivariâncias estimadas em cada conjunto e barras de erro, cujos extremos indicam a média das semivariâncias estimadas mais um desvio padrão e a média menos um desvio padrão. Da análise das Figuras 5 a 8, seguem-se três considerações. Inicialmente, observa-se que as médias das semivariâncias estimadas são muito próximas dos valores verdadeiros de $\gamma(h)$, o que confirma a qualidade de não-viesado do estimador de Matheron, para os processos simulados. A segunda consideração é a de que existe tendência a um aumento na dispersão das semivariâncias estimadas à medida que h aumenta. Tal fato, demonstrado por diversos autores, tais como CRESSIE (1985), PARDO-IGÚZQUIZA & DOWD (2001) e ORTIZ & DEUTSCH (2002) e verificado, experimentalmente, através de simulações de processos estocásticos espaciais, por outros, como McBRATNEY & WEBSTER (1986) e MARCHANT & LARK (2004), é o motivo pelo qual se procura, no ajuste de modelos por métodos de mínimos quadrados, atribuir maior peso aos pontos do semivariograma experimental correspondentes aos menores valores de h . Finalmente, constatou-se que as dispersões das semivariâncias estimadas, apresentadas nas Figuras 5b, 6b, 7b e 8b, são menores que as apresentadas nas Figuras 5a, 6a, 7a e 8a, o que sugere que as estimativas das semivariâncias melhoram à medida que a área de amostragem abrange vários alcances.

Figura 5 - Valores médios das semivariâncias estimadas e barras de erro indicando mais e menos um desvio padrão, para modelos $esf(0, 1, 5)$, grades quadradas 10×10 e distância entre nós consecutivos igual a uma unidade de comprimento (a) e igual a duas unidades de comprimento (b)

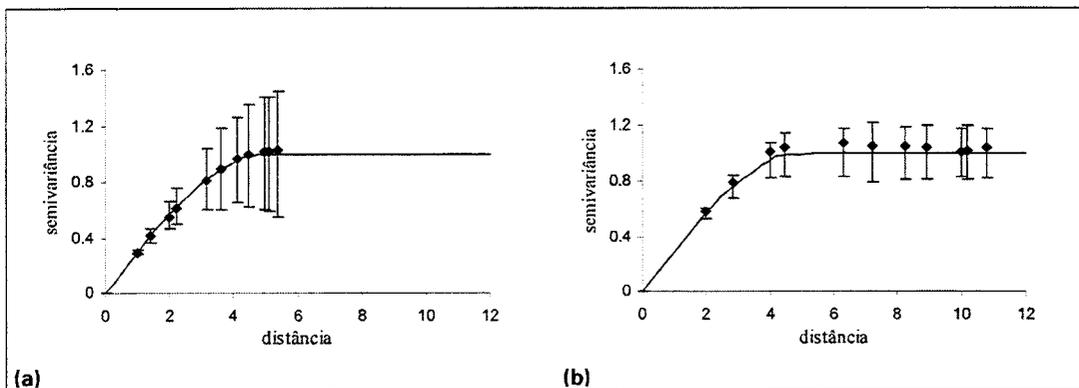


Figura 6 - Valores médios das semivariâncias estimadas e barras de erro indicando mais e menos um desvio padrão, para modelos $esf(0.2, 0.8, 5)$, grades quadradas 10×10 e distância entre nós consecutivos igual a uma unidade de comprimento (a) e igual a duas unidades de comprimento (b)

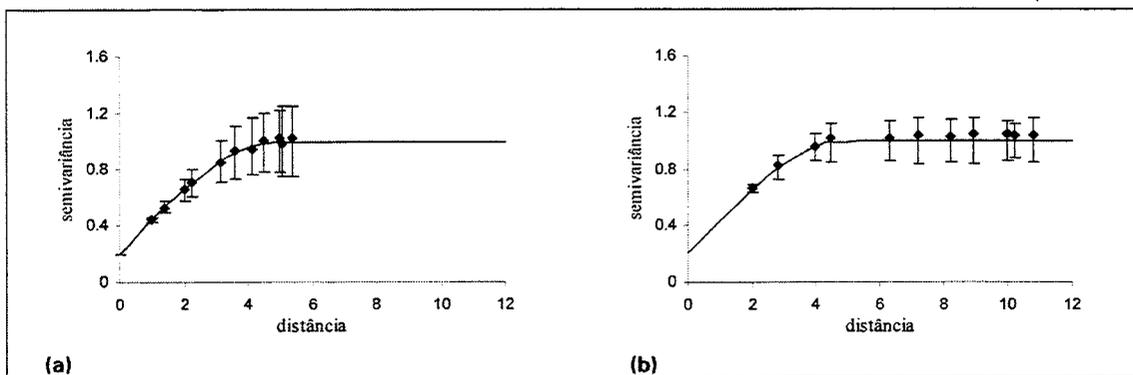


Figura 7 - Valores médios das semivariâncias estimadas e barras de erro indicando mais e menos um desvio padrão, para modelos $exp(0, 1, 5)$, grades quadradas 10×10 e distância entre nós consecutivos igual a uma unidade de comprimento (a) e igual a duas unidades de comprimento (b)

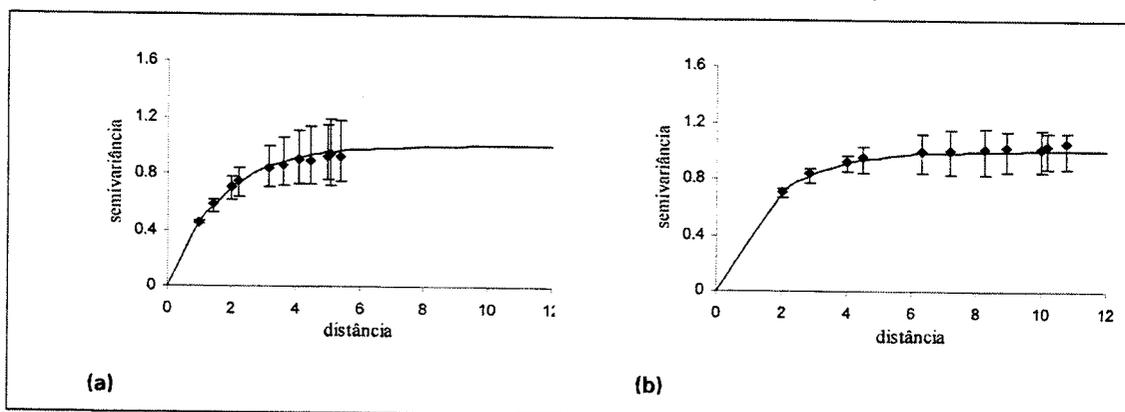
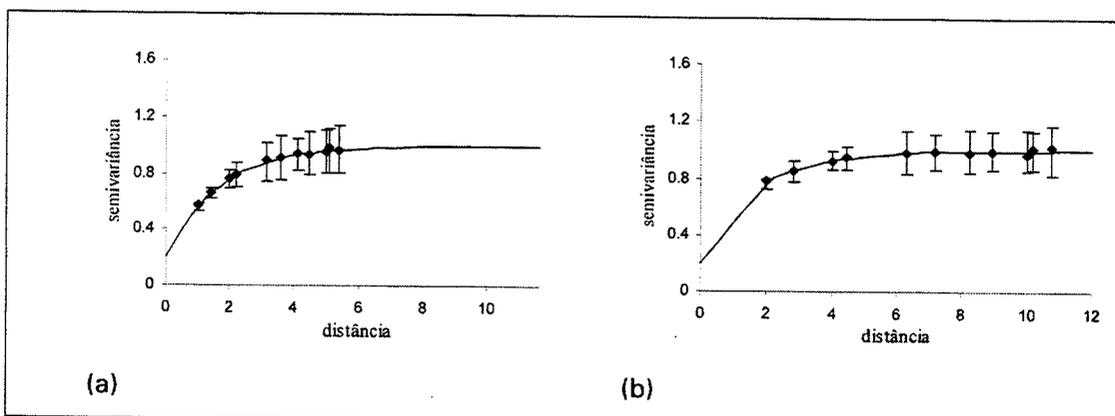


Figura 8 - Valores médios das semivariâncias estimadas e barras de erro indicando mais e menos um desvio padrão, para modelos $exp(0.2, 0.8, 5)$, grades quadradas 10×10 e distância entre nós consecutivos igual a uma unidade de comprimento (a) e igual a duas unidades de comprimento (b)



No Quadro 2 é apresentado um resumo estatístico dos resultados obtidos na estimação dos parâmetros, segundo cada método de ajuste, impondo-se como única restrição que os valores estimados para C_0 , C_1 e α fossem não-negativos. Comparando as estatísticas referentes aos parâmetros estimados pelos diferentes métodos, não foi possível concluir a respeito da superioridade de nenhum dos métodos de estimação considerados, nem quanto à tendenciosidade (avaliada pela média dos parâmetros estimados) nem quanto à eficiência (avaliada pelos intervalos semiquartílicos).

Pode ser observado que, em alguns casos, as médias dos parâmetros C_1 e a estimados são muito maiores que os valores verdadeiros correspondentes. Isso indica que, nas condições em que os experimentos foram realizados e os ajustes foram feitos, nenhum dos métodos de estimação garante a qualidade das estimativas. A estimação dos parâmetros apresentou-se mais problemática para o modelo $exp(0.2, 0.8, 5)$. Acredita-se que isto se deve ao modelo ser, dentre os estudados, aquele ao qual corresponde à estrutura de menor continuidade espacial, tendo ficado mais prejudicado que os demais pela falta de informações a pequenas distâncias.

Quadro 2 - Estatísticas referentes aos parâmetros estimados pelos métodos OLS, WLS1 e WLS2 em cada conjunto de 20 simulações em grades quadradas 10x10, com distância (d) entre nós consecutivos igual uma ou duas unidades de comprimento, impondo-se como única restrição que os parâmetros fossem não-negativos

Modelo	Método	C_0					C_1					a				
		M	Q1	Q2	Q3	I_q	M	Q1	Q2	Q3	I_q	M	Q1	Q2	Q3	I_q
$esf(0, 1, 5)$ $d=1$	OLS	0,03	0,00	0,00	0,04	0,02	7,07	0,64	1,08	1,52	0,44	33,2	3,83	4,74	6,91	1,54
	WLS1	0,03	0,00	0,00	0,04	0,02	5,90	0,64	1,08	1,52	0,44	28,2	3,78	4,76	6,82	1,52
	WLS2	0,02	0,00	0,00	0,03	0,01	10,6	0,65	1,07	1,55	0,45	52,7	3,37	5,00	7,12	1,88
$esf(0, 1, 5)$ $d=2$	OLS	0,03	0,00	0,00	0,03	0,02	1,01	0,92	1,01	1,14	0,11	5,07	4,60	4,95	5,42	0,41
	WLS1	0,03	0,00	0,00	0,05	0,02	1,01	0,93	1,00	1,15	0,11	5,08	4,60	4,92	5,34	0,37
	WLS2	0,03	0,00	0,00	0,05	0,02	1,01	0,92	1,00	1,13	0,11	5,08	4,62	4,96	5,60	0,49
$esf(0.2, 0.8, 5)$ $d=1$	OLS	0,17	0,12	0,14	0,21	0,04	0,87	0,76	0,88	1,05	0,14	5,05	4,16	4,84	6,13	0,99
	WLS1	0,17	0,11	0,15	0,21	0,05	0,88	0,77	0,88	1,07	0,15	5,07	4,10	4,92	6,08	0,99
	WLS2	0,18	0,13	0,14	0,21	0,04	0,86	0,73	0,86	1,06	0,17	4,99	4,06	5,13	5,81	0,88
$esf(0.2, 0.8, 5)$ $d=2$	OLS	0,19	0,05	0,18	0,27	0,11	2,88	0,81	0,94	0,98	0,09	55,1	4,59	4,90	6,10	0,76
	WLS1	0,20	0,07	0,19	0,32	0,12	1,70	0,77	0,92	1,00	0,12	26,1	4,56	4,86	6,23	0,83
	WLS2	0,18	0,03	0,18	0,25	0,11	1,13	0,78	0,91	1,01	0,11	12,3	4,50	4,72	6,07	0,79
$exp(0, 1, 5)$ $d=1$	OLS	0,08	0,00	0,00	0,10	0,05	1,09	0,78	1,00	1,06	0,14	10,4	3,76	4,76	9,28	2,76
	WLS1	0,07	0,00	0,00	0,07	0,03	1,22	0,79	1,01	1,08	0,15	12,4	3,74	4,70	9,36	2,81
	WLS2	0,08	0,00	0,00	0,15	0,07	1,14	0,79	1,02	1,18	0,20	10,5	4,06	4,93	7,23	1,58
$exp(0, 1, 5)$ $d=2$	OLS	0,22	0,00	0,08	0,48	0,24	0,99	0,80	0,99	1,09	0,15	20,1	4,11	6,06	12,3	4,08
	WLS1	0,26	0,00	0,06	0,55	0,28	1,17	0,89	1,00	1,14	0,12	36,3	3,98	6,10	12,5	4,25
	WLS2	0,24	0,00	0,03	0,50	0,25	2,27	0,76	0,95	1,07	0,16	118	4,12	6,20	12,8	4,34
$exp(0.2, 0.8, 5)$ $d=1$	OLS	0,19	0,00	0,14	0,35	0,18	187	0,80	0,85	1,25	0,23	18650	3,31	5,12	8,80	2,75
	WLS1	0,19	0,00	0,16	0,33	0,17	63,0	0,80	0,96	1,17	0,19	5311	3,44	5,06	8,44	2,50
	WLS2	0,19	0,00	0,19	0,33	0,16	27,9	0,78	0,94	1,15	0,18	1826	3,44	5,24	7,45	2,01
$exp(0.2, 0.8, 5)$ $d=2$	OLS	0,37	0,00	0,22	0,71	0,35	4,49	0,51	0,95	1,00	0,25	1289	3,33	7,52	14,9	5,81
	WLS1	0,36	0,00	0,18	0,73	0,36	15,8	0,55	0,96	1,02	0,24	5424	3,12	6,82	11,6	4,25
	WLS2	0,35	0,00	0,16	0,71	0,36	40,6	0,66	0,95	1,08	0,21	15245	3,25	6,93	15,4	6,08

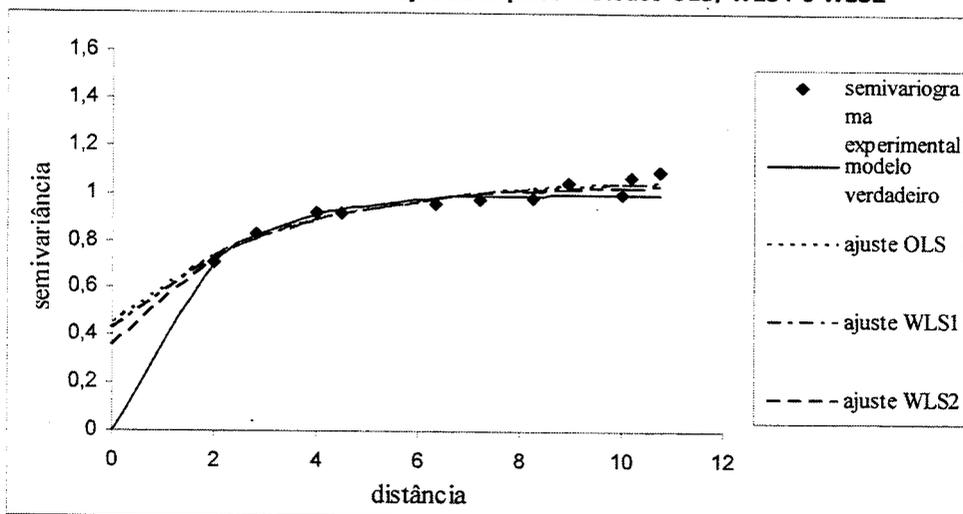
M representa a média, Q1 o primeiro quartil, Q2 o segundo quartil (ou mediana), Q3 o terceiro quartil e I_q o intervalo semi-quartilico $((Q3-Q1)/2)$.

São discutidas, a seguir, algumas das situações em que os parâmetros estimados mais diferiram dos valores verdadeiros.

- **Efeito Pepita Superestimado**

Na Figura 9 é apresentado um dos semivariogramas experimentais obtidos a partir de dados simulados, o gráfico do modelo verdadeiro, $exp(0, 1, 5)$, a partir do qual os dados foram gerados e os gráficos dos modelos exponenciais ajustados pelos métodos OLS, WLS1 e WLS2. Os vetores de parâmetros estimados pelos três métodos são, respectivamente, (0.45, 0.62, 10.03), (0.43, 0.63, 9.48) e (0.35, 0.69, 8.00).

Figura 9 - Semivariograma experimental correspondente a uma realização do modelo $exp(0, 1, 5)$, modelo verdadeiro e modelos ajustados pelos métodos OLS, WLS1 e WLS2



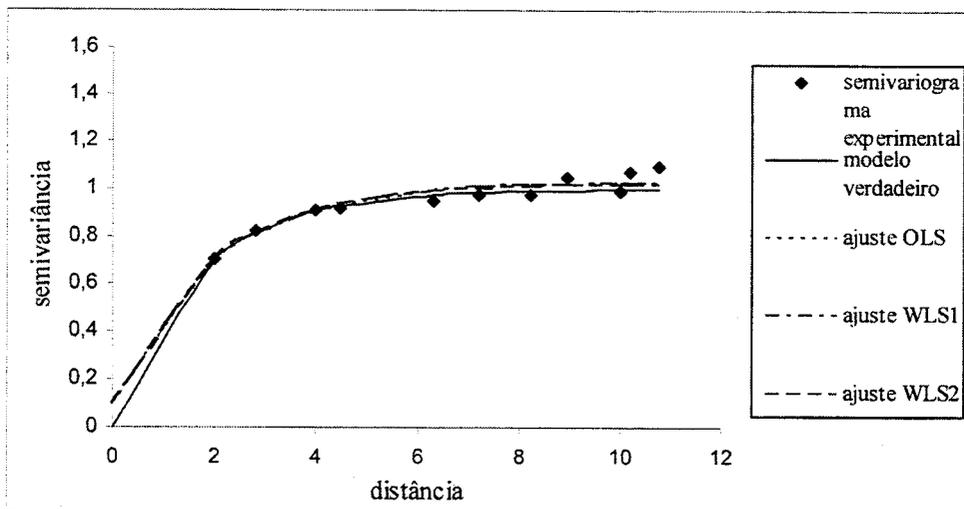
Pode-se observar que as curvas correspondentes aos modelos ajustados pelos métodos OLS, WLS1 e WLS2 são bem próximas dos pontos correspondentes ao semivariograma experimental. São também bem próximas à curva correspondente ao modelo verdadeiro, para distâncias maiores do que duas unidades. O mesmo não ocorre para as pequenas distâncias, pois o efeito pepita estimado, qualquer que seja o método de ajuste considerado, é muito grande. O erro na estimativa do efeito pepita tem

implicações, tanto na krigagem quanto na caracterização da estrutura de dependência espacial em si.

De resultados como os apresentados na Figura 9, o que se pode aprender com simulações, para a modelagem da dependência espacial de dados reais, é que não se deve concluir a respeito de altos valores do efeito pepita tendo por base apenas métodos automáticos de ajuste. É preciso que o pesquisador, com base em seu conhecimento a respeito do atributo em estudo, decida se a grande descontinuidade do semivariograma na origem é razoável. Se há dúvida, pode-se, em uma nova coleta de dados, buscar mais informações para pequenas distâncias. Se, por outro lado, o pesquisador está certo de que um valor de efeito pepita tão grande é incompatível com o atributo em estudo, ele pode impor um limite superior ao valor de C_0 , na resolução do problema de otimização. Na Figura 10, são mostradas as curvas ajustadas ao mesmo semivariograma experimental da Figura 9, acrescentando-se a restrição de que o parâmetro C_0 fosse menor ou igual a 0,1.

Os vetores de parâmetros estimados, correspondentes aos ajustes pelos métodos OLS, WLS1 e WLS2, apresentados na Figura 10, foram (0.1, 0.93, 5.97), (0.1, 0.93, 5.62) e (0.1, 0.93, 5.62), respectivamente. Observe-se que, com essa restrição, melhora sensivelmente a qualidade do ajuste para os menores valores de h , sem prejuízo para o ajuste a maiores distâncias. Além disso, a restrição imposta a C_0 contribuiu para melhorar também as estimativas dos parâmetros C_j e a .

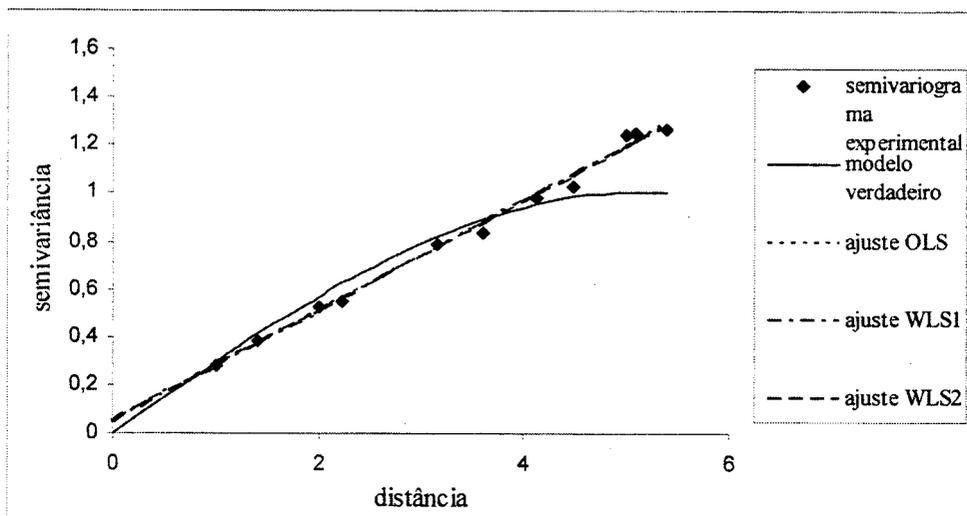
Figura 10 - Semivariograma experimental correspondente a uma realização do modelo $exp(0, 1, 5)$, modelo verdadeiro e modelos ajustados pelos métodos OLS, WLS1 e WLS2, com restrição imposta ao valor máximo do efeito pepita



• Patamar e Alcance Superestimados

Na Figura 11 apresenta-se um semivariograma experimental, o gráfico do modelo $esf(0, 1, 5)$, a partir do qual os dados foram gerados, e os modelos esféricos ajustados. Os vetores de parâmetros estimados pelos métodos OLS, WLS1 e WLS2 foram (0.05, 27.17, 177.53), (0.04, 25.34, 164.31) e (0.05, 44.93, 295.88), respectivamente.

Figura 11 - Semivariograma experimental correspondente a uma realização do modelo $esf(0, 1, 5)$, modelo verdadeiro e modelos ajustados pelos métodos OLS, WLS1 e WLS2



O primeiro fato que pode causar estranheza, nos resultados obtidos, é que os valores estimados para os parâmetros C_I e a são muito grandes, com relação aos valores verdadeiros. Este resultado, no entanto, pode ser compreendido, observando-se que a curva a ser ajustada é do tipo esférico definido na Equação (3), e que os pontos do semivariograma experimental estão alinhados muito próximos de uma reta. Assim sendo, é de se esperar que o valor estimado para o alcance seja bastante grande, de forma a tornar desprezível a contribuição do termo de terceiro grau, para as distâncias consideradas. Um grande valor estimado para o alcance implica um grande valor estimado para C_I , uma vez que $3C_I/2a$ é o coeficiente angular da “reta” ajustada. O segundo fato que pode causar estranheza é que, embora com parâmetros C_I e a tão diferentes, estimados pelos métodos OLS, WLS1 e WLS2, não é possível distinguir, no intervalo de distâncias para as quais os gráficos foram plotados, as curvas correspondentes aos modelos ajustados pelos diferentes métodos. A razão por que isto acontece é que, embora os valores de C_I e a , estimados pelos diferentes métodos de ajuste, variem, a razão C_I/a é mantida praticamente constante.

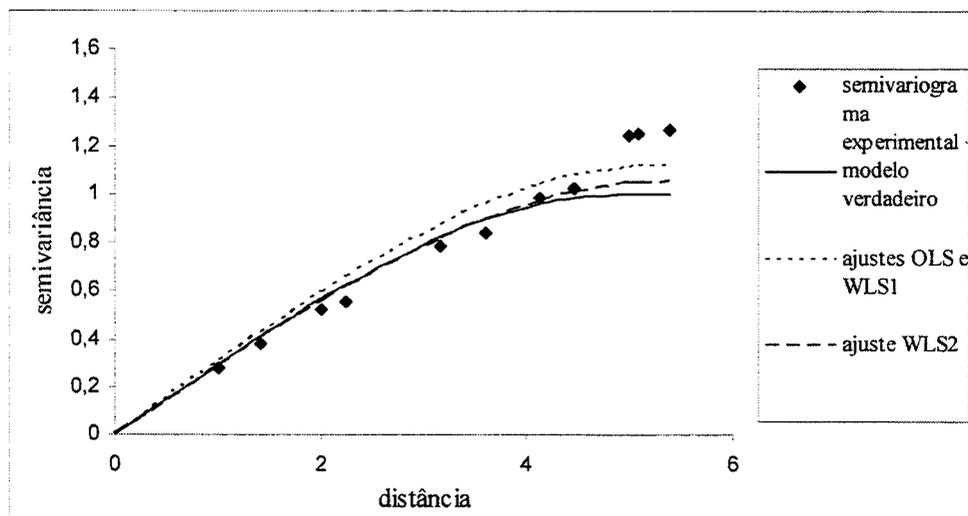
Embora os altos valores estimados para os parâmetros não tragam prejuízo para a krigagem, uma vez que os valores das semivariâncias correspondentes aos modelos ajustados são, para pequenos valores de h , muito próximos, tanto das semivariâncias experimentais quanto do modelo que gerou os dados, não parece razoável supor que, a partir de semivariâncias experimentais estimadas apenas para valores de h que não chegam a seis unidades de comprimento, seja possível concluir a respeito de alcances da ordem de 300 unidades. Uma maneira de contornar o problema é acrescentar alguma restrição ao ajuste. Uma restrição que parece plausível, quando se trata de ajustar modelos com patamar, é supor que o alcance não seja maior do que o maior valor da distância para a qual a semivariância experimental foi estimada. Acrescentando-se essa restrição (oferecida por *default* em alguns *softwares* geoestatísticos comerciais e à qual o usuário deve estar atento...), e fazendo novamente o ajuste de modelos esféricos pelos métodos OLS, WLS1 e WLS2, obtiveram-se, respectivamente, os vetores de parâmetros (0, 1.12, 5.39), (0, 1.12, 5.39) e (0, 1.05, 5.39).

Na Figura 12, é apresentado o semivariograma experimental, o modelo a partir do

qual os dados foram gerados e os modelos ajustados pelos métodos OLS, WLS1 e WLS2, com restrição no alcance. Ressalta-se que, embora nesse caso específico a restrição com respeito ao alcance tenha resolvido com sucesso o problema do ajuste e a estimação dos parâmetros, em muitos casos deve levar a valores do alcance subestimados.

Dos casos particulares discutidos nesta seção, pode-se perceber que, com os tamanhos de amostragem aqui considerados, os métodos de mínimos quadrados mostraram-se insuficientes para garantir a acurácia das estimativas dos parâmetros que caracterizam a estrutura de dependência espacial. No entanto, pode-se também observar que as estimativas podem ser melhoradas com o uso de restrições.

Figura 12 - Semivariograma experimental correspondente a uma realização do modelo $sf(0, 1, 5)$, modelo verdadeiro e modelos ajustados pelos métodos OLS, WLS1 e WLS2, com restrição imposta ao valor máximo do alcance



No Quadro 3, são apresentados os resultados obtidos na estimação dos parâmetros, segundo cada método de ajuste, para cada conjunto de 20 simulações realizadas em malhas 10x10, com distância entre nós consecutivos igual a uma unidade ($d=1$) e igual a duas unidades de comprimento ($d=2$), impondo-se aos ajustes, além das restrições de que os parâmetros sejam não-negativos, a restrição de que o alcance estimado seja menor ou igual ao maior valor do *lag* considerado no semivariograma experimental. Assim sendo, a maior estimativa admitida para o alcance, quando $d=1$ é

igual a 5,39 unidades e, quando $d=2$ é igual a 10,8 unidades. Comparando-se as estatísticas dos parâmetros estimados pelos métodos OLS, WLS1 e WLS2, apresentadas no Quadro 2, pode-se verificar que são todas muito próximas umas das outras. Nas condições em que os experimentos foram realizados, nenhum dos métodos de estimação considerados mostrou-se menos tendencioso ou mais eficiente que os demais.

Comparando-se as estatísticas apresentadas no Quadro 3 às estatísticas correspondentes, apresentadas no Quadro 2, pode-se verificar que a restrição imposta ao alcance contribuiu muito para a melhoria dos resultados obtidos, tanto no que diz respeito às médias quanto às dispersões dos parâmetros estimados. Comparando os resultados referentes a cada quartil, pode-se constatar que o primeiro quartil e o segundo quartil (ou mediana) permaneceram inalterados em todos os conjuntos de parâmetros estimados. Já o terceiro quartil foi afetado pela restrição acrescentada, mostrando-se alterado em quase todas as situações. Pode ser observado, no Quadro 3, que em todos os conjuntos de dados correspondentes a $d=1$, o terceiro quartil relativo ao alcance é igual ao valor máximo admitido (5,39 unidades de comprimento), o que significa que, nessas situações, a restrição no alcance foi determinante no ajuste.

Reunindo as 20 realizações correspondentes a um mesmo processo estocástico e mesma distância entre os nós da malha, em um único experimento, segundo a amostragem apresentada na Figura 2, foram obtidas quatro novas amostras correspondentes a modelos de semivariogramas esféricos e quatro novas amostras correspondentes a modelos de semivariogramas exponenciais, cada uma com 2000 pontos amostrais.

Quadro 3 - Estatísticas referentes aos parâmetros estimados pelos métodos OLS, WLS1 e WLS2 em cada conjunto de 20 simulações em grades duplas 10x10, com distância (d) entre nós consecutivos igual uma ou duas unidades de comprimento, impondo-se restrição nos limites inferiores dos parâmetros e no valor máximo do alcance

Modelo	Método	C ₀					C ₁					a				
		M	Q1	Q2	Q3	I _q	M	Q1	Q2	Q3	I _q	M	Q1	Q2	Q3	I _q
esf(0, 1, 5) d=1	OLS	0,03	0,00	0,00	0,02	0,01	0,98	0,64	1,06	1,28	0,32	4,57	3,83	4,74	5,39	0,78
	WLS1	0,03	0,00	0,00	0,02	0,01	0,97	0,64	1,06	1,28	0,32	4,57	3,78	4,76	5,39	0,81
	WLS2	0,02	0,00	0,00	0,02	0,01	0,90	0,65	1,01	1,22	0,29	4,54	3,37	5,00	5,39	1,01
esf(0, 1, 5) d=2	OLS	0,03	0,00	0,00	0,03	0,02	1,01	0,92	1,01	1,14	0,11	5,07	4,60	4,95	5,42	0,41
	WLS1	0,03	0,00	0,00	0,05	0,02	1,01	0,93	1,00	1,15	0,11	5,08	4,60	4,92	5,34	0,37
	WLS2	0,03	0,00	0,00	0,05	0,02	1,01	0,92	1,00	1,13	0,11	5,08	4,62	4,96	5,60	0,49
esf(0,2, 0,8, 5) d=1	OLS	0,16	0,11	0,13	0,19	0,04	0,85	0,73	0,88	1,04	0,15	4,75	4,18	5,05	5,39	0,61
	WLS1	0,16	0,10	0,13	0,19	0,05	0,85	0,74	0,88	1,05	0,15	4,77	4,12	5,11	5,39	0,64
	WLS2	0,17	0,12	0,14	0,22	0,05	0,83	0,71	0,86	0,99	0,14	4,80	4,23	5,19	5,39	0,58
esf(0,2, 0,8, 5) d=2	OLS	0,19	0,05	0,18	0,27	0,11	0,86	0,76	0,92	0,97	0,11	5,44	4,59	4,90	6,10	0,76
	WLS1	0,20	0,07	0,19	0,32	0,12	0,85	0,74	0,88	0,98	0,12	5,53	4,56	4,86	6,23	0,83
	WLS2	0,18	0,03	0,18	0,25	0,11	0,85	0,78	0,90	0,99	0,11	5,35	4,50	4,72	6,07	0,79
exp(0, 1, 5) d=1	OLS	0,04	0,00	0,00	0,00	0,00	0,91	0,73	0,99	1,05	0,16	4,37	3,76	4,76	5,39	0,82
	WLS1	0,04	0,00	0,00	0,00	0,00	0,91	0,73	0,98	1,05	0,16	4,33	3,74	4,70	5,39	0,83
	WLS2	0,02	0,00	0,00	0,00	0,00	0,90	0,72	1,00	1,05	0,16	4,45	4,06	4,93	5,39	0,66
exp(0, 1, 5) d=2	OLS	0,18	0,00	0,08	0,40	0,20	0,88	0,70	0,92	1,03	0,17	7,01	4,11	6,06	10,8	3,33
	WLS1	0,21	0,00	0,06	0,40	0,20	0,84	0,64	0,92	1,03	0,19	7,23	4,25	6,84	10,8	3,26
	WLS2	0,20	0,00	0,04	0,41	0,21	0,80	0,59	0,88	1,02	0,22	7,12	4,30	7,13	10,7	3,22
exp(0,2, 0,8, 5) d=1	OLS	0,11	0,00	0,02	0,20	0,10	0,88	0,73	0,93	1,09	0,18	4,35	3,31	4,86	5,39	1,04
	WLS1	0,11	0,00	0,01	0,23	0,12	0,89	0,72	0,94	1,09	0,19	4,34	3,44	4,83	5,39	0,98
	WLS2	0,12	0,00	0,05	0,23	0,12	0,86	0,70	0,89	1,08	0,19	4,38	3,44	4,92	5,39	0,97
exp(0,2, 0,8, 5) d=2	OLS	0,30	0,00	0,14	0,56	0,28	0,72	0,42	0,89	0,97	0,28	6,44	3,12	7,24	10,8	3,82
	WLS1	0,29	0,00	0,14	0,58	0,29	0,73	0,40	0,90	0,99	0,30	6,43	3,12	6,81	10,8	3,83
	WLS2	0,29	0,00	0,15	0,56	0,28	0,72	0,41	0,88	0,97	0,28	6,39	3,17	6,46	10,8	3,80

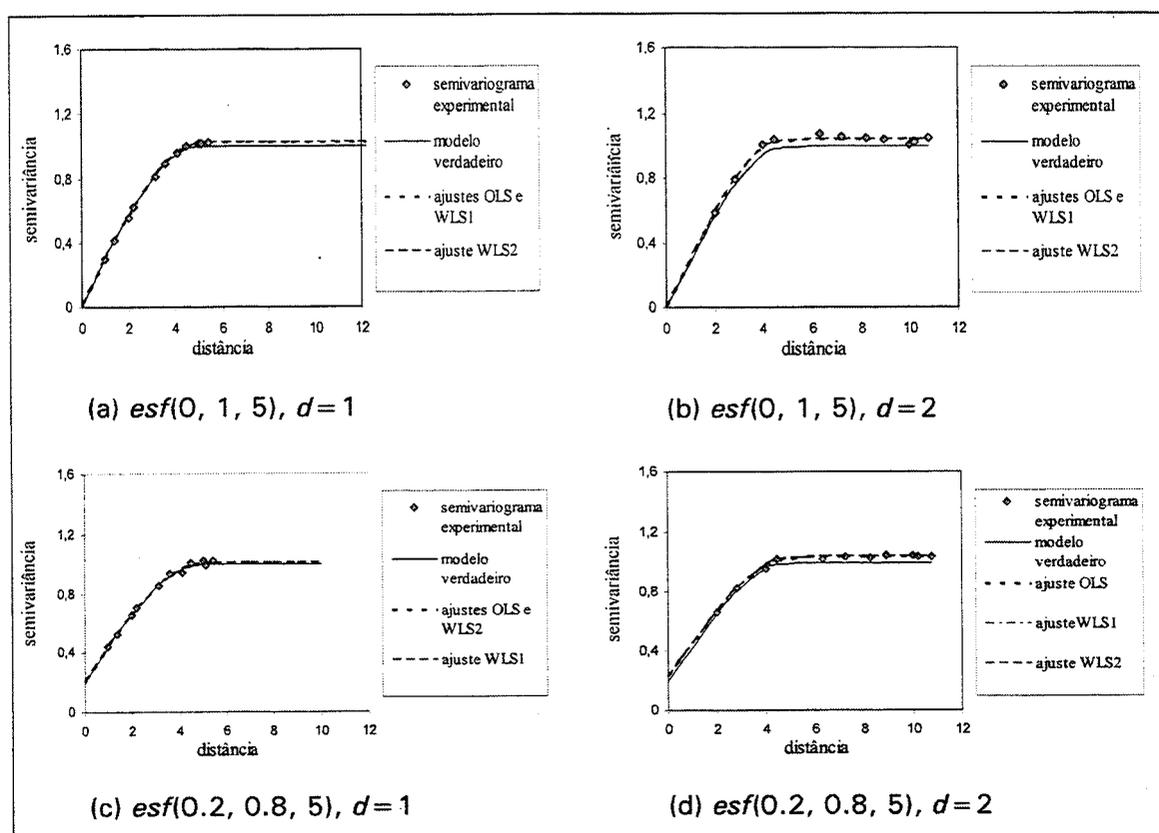
M representa a média, Q1 o primeiro quartil, Q2 o segundo quartil (ou mediana), Q3 o terceiro quartil e I_q o intervalo semi-quartilício ((Q3-Q1)/2).

No Quadro 4, são apresentados, para os processos estocásticos simulados, correspondentes a modelos esféricos, os vetores de parâmetros estimados pelos métodos OLS, WLS1 e WLS2, a partir de cada uma das amostragens em grade dupla. A única restrição imposta aos ajustes foi que os parâmetros fossem não-negativos. Dos resultados apresentados no Quadro 4, pode ser observado que os parâmetros estimados, por qualquer um dos métodos considerados, são muito próximos dos valores verdadeiros. Na Figura 13, são apresentados os gráficos correspondentes.

Quadro 4 - Vetores de parâmetros estimados, para modelos esféricos, a partir das simulações em grades duplas, pelos métodos OLS, WLS1 e WLS2

Modelo de semivariograma	Amostragem com distância d	OLS (C_0, C_1, a)	WLS1 (C_0, C_1, a)	WLS2 (C_0, C_1, a)
$esf(0, 1, 5)$	$d = 1$	(0.01, 1.02, 5.26)	(0.01, 1.02, 5.26)	(0.01, 1.02, 5.27)
	$d = 2$	(0.00, 1.04, 4.95)	(0.00, 1.04, 4.95)	(0.00, 1.04, 5.00)
$esf(0.2, 0.8, 5)$	$d = 1$	(0.20, 0.81, 5.10)	(0.20, 0.81, 5.11)	(0.20, 0.81, 5.10)
	$d = 2$	(0.23, 0.81, 5.27)	(0.22, 0.82, 5.22)	(0.22, 0.82, 5.21)

Figura 13 - Semivariogramas experimentais construídos a partir de simulações em grades duplas (2000 pontos amostrais), modelos esféricos a partir dos quais os dados foram gerados e modelos ajustados pelos métodos OLS, WLS1 e WLS2



Observa-se, na Figura 13, que há uma sobreposição quase total das curvas correspondentes aos modelos ajustados, sendo, as mesmas, muito próximas das curvas correspondentes aos modelos verdadeiros. Mesmo nos casos das Figuras 13a e 13c, em que os semivariogramas experimentais são constituídos de poucas semivariâncias estimadas a distâncias maiores que o alcance, os patamares dos modelos ajustados são bem próximos do valor verdadeiro.

No Quadro 5, são apresentados, para os processos estocásticos simulados, correspondentes a modelos exponenciais, os vetores de parâmetros estimados pelos métodos OLS, WLS1 e WLS2. A única restrição imposta aos ajustes foi a de que os parâmetros C_0 , C_1 e a fossem não-negativos. Pode-se observar que os parâmetros estimados diferem mais de uma amostragem para outra do que de um método de estimação para outro. Observe-se que, na amostragem em que houve maior falta de informação a pequenas distâncias ($d = 2$), o valor do efeito pepita, estimado por qualquer um dos métodos, aparece superestimado.

Quadro 5 - Vetores de parâmetros estimados, para modelos exponenciais, a partir de simulações em grades duplas, pelos métodos OLS, WLS1 E WLS2

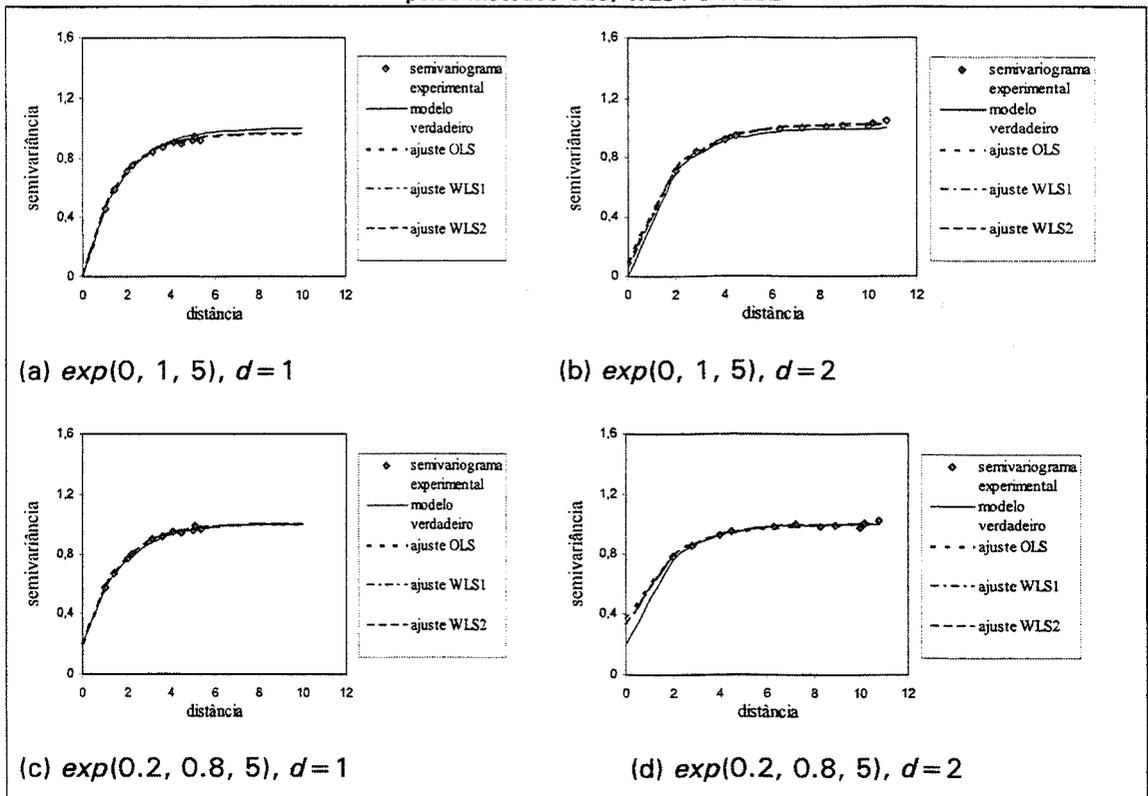
Modelo de semivariograma	Amostragem com distância d	OLS (C_0, C_1, a)	WLS1 (C_0, C_1, a)	WLS2 (C_0, C_1, a)
<i>exp</i> (0, 1, 5)	$d = 1$	(0.00, 0.96, 4.51)	(0.00, 0.96, 4.50)	(0.00, 0.96, 4.55)
	$d = 2$	(0.09, 0.94, 5.49)	(0.08, 0.95, 5.41)	(0.05, 0.97, 5.28)
<i>exp</i> (0.2, 0.8, 5)	$d = 1$	(0.19, 0.81, 4.73)	(0.18, 0.82, 4.70)	(0.19, 0.81, 4.77)
	$d = 2$	(0.37, 0.63, 5.64)	(0.33, 0.67, 5.31)	(0.33, 0.67, 5.32)

Na Figura 14, são apresentados os semivariogramas experimentais e os gráficos correspondentes aos modelos exponenciais ajustados com os parâmetros apresentados no Quadro 5. Como já observado para os semivariogramas correspondentes aos modelos esféricos, as curvas correspondentes aos ajustes pelos métodos OLS, WLS1 e WLS2, para os modelos exponenciais, são quase coincidentes e, na maioria dos casos, bastante próximas da curva correspondente ao modelo verdadeiro. As maiores diferenças entre as semivariâncias estimadas pelos modelos ajustados e as semivariâncias verdadeiras são observadas na Figura 14d. O que diferencia essa situação das demais é a combinação de rápido decréscimo da correlação espacial

(devido ao efeito pepita e ao modelo exponencial) com a falta de informação a respeito de semivariâncias experimentais para pequenos *lags*.

Busca-se aprender com dados simulados para poder utilizar as conclusões obtidas no estudo de variáveis reais. Nesse sentido, cabe ressaltar que, embora tenha sido obtido sucesso na "captura" da estrutura de dependência espacial simulando a amostragem apresentada na Figura 2, tal amostragem dificilmente poderia ser utilizada com dados reais, pois a coleta em 2000 pontos amostrais pode ter um custo proibitivo. Além disso, nessa configuração de amostragem os mapeamentos seriam prejudicados pela falta de dados nos espaços entre os conglomerados. O que esses experimentos Monte Carlo ensinam é que os esquemas de amostragem considerados possibilitam que se conclua a respeito do padrão de dependência espacial. Uma vez conhecido o padrão de dependência espacial de um atributo, em um certo domínio, esse padrão pode ser utilizado para fins de mapeamento a partir de menor número de pontos amostrais, regularmente espaçados.

Figura 14 - Semivariogramas experimentais construídos a partir de simulações em grades duplas (2000 pontos amostrais), modelos exponenciais a partir dos quais os dados foram gerados e modelos ajustados pelos métodos OLS, WLS1 e WLS2



5. Conclusões

Do trabalho desenvolvido, concluiu-se que o sucesso da estimação de parâmetros que caracterizam a estrutura de dependência espacial de um atributo depende de semivariâncias experimentais estimadas com acurácia (o que requer muitos pontos amostrais), mais do que do método de ajuste de um modelo teórico ao semivariograma experimental.

A análise das semivariâncias estimadas a partir de conjuntos de 100 pontos amostrais, mostrou que, mesmo que a população satisfaça a hipótese de estacionaridade de segunda ordem e que a amostragem possibilite que sejam estimadas semivariâncias tanto a distâncias menores como a distâncias maiores que o alcance, 100 pontos amostrais podem ser insuficientes para que os semivariogramas experimentais detectem a estrutura de dependência espacial correspondente.

Foi verificada a maior eficiência do estimador de Matheron para menores valores da distância, pelo que se reitera a importância de priorizar a qualidade do ajuste para os pequenos *lags*. No entanto, os maiores pesos atribuídos aos menores *lags* nos métodos de mínimos quadrados ponderados considerados neste trabalho mostraram-se insuficientes para superar as limitações dos métodos dos mínimos quadrados ordinários. Constatou-se que a qualidade do ajuste pode ser melhorada com o uso de métodos de mínimos quadrados aliados a restrições aos valores dos parâmetros, restrições essas que podem ser impostas a partir do conhecimento do pesquisador a respeito do atributo em estudo, e que fazem do ajuste de um modelo teórico ao semivariograma experimental um verdadeiro exercício de modelagem, e não um procedimento automático.

Na análise das semivariâncias estimadas a partir de 2000 pontos amostrais, com a amostragem correspondente a grades duplas proposta, constatou-se que as mesmas são bastante próximas das semivariâncias correspondentes aos modelos teóricos de semivariogramas a partir dos quais os dados foram gerados. Neste caso, tanto os métodos de mínimos quadrados ponderados, como o método de mínimos quadrados ordinários, produziram bons resultados na estimação dos parâmetros e no ajuste de modelos teóricos aos semivariogramas experimentais.

Verificou-se, ainda, que ajustar um modelo teórico ao semivariograma experimental, obtendo uma função que, para pequenos valores de h , conduza a valores das semivariâncias próximos dos verdadeiros (que é só o que é requerido para fins de krigagem), é uma tarefa mais simples do que estimar com acurácia os parâmetros efeito pepita, patamar e alcance, ou o percentual de dependência espacial.

No experimento simulado em que a área amostrada abrangeu vários alcances, sem deixar de contemplar um grande número de pontos amostrais separados por pequenas distâncias, qualquer um dos métodos de ajuste considerados produziu bons resultados.

Para avaliações mais abrangentes, espera-se em trabalhos futuros realizar simulações com modelos da família Matérn e família exponencial potência, assim como trabalhar com grades não-regulares e com os estimadores de máxima verossimilhança e máxima verossimilhança restrita.

Referências bibliográficas

- CRESSIE, N. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology*, New York, v. 17, n. 4, p. 563-586.
- CRESSIE, N. (1988). Spatial prediction and ordinary kriging. *Mathematical Geology*, New York, v. 20, n. 4, p. 405-421.
- CRESSIE, N. A. (1993). *Statistics for Spacial Data*. New York: John Wiley, 900p.
- DIGGLE, P. J.; RIBEIRO JR, P. J. (2000) *Model Based Geoestistics*. 14^o SINAPE-CAXAMBU- 70p.
- EASTMAN, J. R. IDRISI 32 Release 2 - Guide To GIS and Image Processing, 2001 - Manual Version 32.20, Clark Labs, Clark University, Worcester, EUA, vol. I, 2001. 161 p.
- ISAAKS, E. H.; SRIVASTAVA, R. M. (1989). *Applied Geoestistics*. New York: Oxford University Press, 561p.
- JOHNSON, R. A.; WICHERN, D. W. (1982). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice-Hall, 593p.
- JOURNAL, A. G.; HUIJBREGTS, CH. J. (1978). *Mining Geoestistics*. London, Academic Press, 600p.
- MARCHANT, B. P.; LARK, R. M. (2004). Estimating Variogram Uncertainty. *Mathematical Geology*, New York, v. 36, n. 8, p. 867-898.
- MATHERON, G. (1963). Principles of geoestistics. *Economic Geology*, Lacaster, v. 58, p.1246-1266.

- MCBRATNEY, A. B.; WEBSTER, R. (1986). Choosing functions for semivariograms of soil properties and fitting them to samples estimates. *Journal of Soil Science*. Oxford, p.617-639.
- ORTIZ, C. J.; DEUTSCH, C. V. (2002). Calculation of uncertainty in the variogram. *Mathematical Geology*, New York, v. 34, n. 2, p. 169-183.
- PARDO-IGÚZQUIZA, E.; DOWN, P. A. (2001). Variance-covariance matrix of the experimental variogram: Assessing Variogram Uncertainty. *Mathematical Geology*, New York, v. 33, n. 4, p. 397-419.
- WEBSTER, R; OLIVER, M.A. (2001). Sample adequately to estimate variograms of soil properties. *Journal of Soil Science*, Oxford, 43: 177-192.
- WOLLENHAUPT, N. C.; MULLA, D.J.; CRAWFORD, G. (1997). Soil Sampling and Interpolation Techniques for Mapping Spatial Variability of Soil Properties. In: PIERCE, F. J.; SADLER, E.J. *The State of Site-Specific Management for Agriculture*. Madison, p.19-53.

Abstract

This work had as objective to study the fitting of models to experimental semivariograms, an important stage in geostatistical analysis of spatially referenced data. It was developed starting from gaussian spatial processes, simulated by the Cholesky-decomposition method. The sampling design in which the simulations were accomplished are 10x10 square grids (100 sample points), and double grids composed of twenty 10x10 conglomerates (2000 sample points). In the fitting of models, were utilized and compared the method of ordinary least squares (OLS), the method of weighted least squares, with weights which are directly proportional to the number of pairs of sample points correspondent to each estimated semivariance (WLS1) and the method of weighted least squares with weights which are directly proportional to the number of pairs of sample points and inversely proportional to the square of the estimated semivariances (WLS2). From the results, it follows that the difficulty of the fitting is associated to insufficient samplings. For intense samplings (2000 sample points) and with distances between the sample points adequate for description of the correspondent spatial dependence, any one of the methods produces good results. From the results, it follows that the biggest difficulty in the fitting is due to incorrect estimates of the semivariances, related with insufficient samplings.

Key words: Geoestistics, simulations, weighted least squares.

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq e à Fundação Araucária do Paraná, pelo apoio financeiro ao desenvolvimento desta pesquisa.

POLÍTICA EDITORIAL

A Revista Brasileira de Estatística - RBEs publica trabalhos relevantes em Estatística Aplicada, não havendo limitação no assunto ou matéria em questão. Como exemplos de áreas de aplicação citamos as áreas de advocacia, ciências física e biomédica, criminologia, demografia, economia, educação, estatísticas governamentais, finanças, indústria, medicina, meio ambiente, negócios, políticas públicas, psicologia e sociologia, entre outras. A RBEs publicará, também, artigos abordando os diversos aspectos de metodologias relevantes para usuários e produtores de estatísticas públicas, incluindo planejamento, avaliação e mensuração de erros em censos e pesquisas, novos desenvolvimentos em metodologia de pesquisa, amostragem e estimação, imputação de dados, disseminação e confiabilidade de dados, uso e combinação de fontes alternativas de informação e integração de dados, métodos e modelos demográfico e econométrico. Os artigos submetidos devem ser inéditos e não devem ter sido submetidos simultaneamente a qualquer outro periódico.

O periódico tem como objetivo a apresentação de artigos que permitam fácil assimilação por membros da comunidade em geral. Os artigos devem incluir aplicações práticas como assunto central, com análises estatísticas exaustivas e apresentadas de forma didática. Entretanto, o emprego de métodos inovadores, apesar de ser incentivado, não é essencial para a publicação.

Artigos contendo exposição metodológica são também incentivados, desde que sejam relevantes para a área de aplicação pela qual os mesmos foram motivados, auxiliem na compreensão do problema e contenham interpretação clara das expressões algébricas apresentadas.

A RBEs tem periodicidade semestral e também publica artigos convidados e resenhas de livros, bem como incentiva a submissão de artigos voltados para a educação estatística.

Artigos em espanhol ou inglês só serão publicados caso nenhum dos autores seja brasileiro e nem resida no País.

Todos os artigos submetidos são avaliados quanto à qualidade e à relevância por dois especialistas indicados pelo Comitê Editorial da RBEs.

O processo de avaliação dos artigos submetidos é do tipo 'duplo cego', isto é, os artigos são avaliados sem identificação de autoria e os comentários dos avaliadores também são repassados aos autores sem identificação.

INSTRUÇÃO PARA SUBMISSÃO DE ARTIGOS À RBEs

O processo editorial da RBEs é eletrônico. Os artigos devem ser submetidos via *e-mail* para: hortega@ibge.gov.br

Após a submissão, o autor correspondente receberá um código para acompanhar o processo de avaliação do artigo. Caso não receba um aviso com este número no prazo de uma semana, fazer contato com a secretaria da revista no endereço:

Revista Brasileira de Estatística

IBGE – Diretoria de Pesquisas - Coordenação de Métodos e Qualidade

Av. República do Chile, nº 500, 10º andar

Centro, Rio de Janeiro – RJ

CEP: 20031-170

Tel.: 55 21 2142-0472

55 21 2142-4549

Fax: 55 21 2142-0039

INSTRUÇÕES PARA PREPARO DOS ORIGINAIS

Os originais entregues para publicação devem obedecer às normas seguintes:

1. Originais processados pelo editor de textos Word for Windows são preferidos. Entretanto, serão aceitos também, originais processados em Látex desde que sejam encaminhados e acompanhados de versões em pdf, conforme descrito no item 3 a seguir;
2. A primeira página do original (folha de rosto) deve conter o título do artigo, seguido do(s) nome(s) completo(s) do(s) autor(es), indicando-se, para cada um, a afiliação e endereço para correspondência. Agradecimentos a colaboradores e instituições, e auxílios recebidos, também, devem figurar nesta página;
3. No caso da submissão não ser em Word for Windows, três arquivos do original devem ser enviados. O primeiro deve conter os originais no processador de texto utilizado (por exemplo, Látex). O segundo e terceiro devem ser no formato pdf, sendo um com a primeira página, como descrito no item 2, e outro contendo apenas o título, sem identificação do(s) autor(es) ou outros elementos que possam permitir a identificação da autoria;
4. A segunda página do original deve conter resumos em português e inglês (*abstract*), destacando os pontos relevantes do artigo. Cada resumo deve ser digitado seguindo o mesmo padrão do restante do texto, em um único parágrafo, sem fórmulas, com, no máximo, 150 palavras;
5. O artigo deve ser dividido em seções, numeradas progressivamente, com títulos concisos e apropriados. Todas as seções e subseções devem ser numeradas e receber título apropriado;

6. Tratamentos algébricos exaustivos devem ser evitados ou alocados em apêndices;
7. A citação de referências no texto e a listagem final de referências devem ser feitas de acordo com as normas da ABNT;
8. As tabelas e gráficos devem ser precedidos de títulos que permitam perfeita identificação do conteúdo. Devem ser numeradas seqüencialmente (Tabela 1, Figura 3, etc.) e referidas nos locais de inserção pelos respectivos números. Quando houver tabelas e demonstrações extensas ou outros elementos de suporte, podem ser empregados apêndices. Os apêndices devem ter título e numeração, tais como as demais seções de trabalho; e
9. Gráficos e diagramas para publicação devem ser incluídos nos arquivos com os originais do artigo. Caso tenham que ser enviados em separado, devem ter nomes que facilitem a sua identificação e posicionamento correto no artigo (ex.: Gráfico 1; Figura 3; etc.). É fundamental que não existam erros, quer no desenho, quer nas legendas ou títulos.