**A INCORPORAÇÃO DA INCERTEZA DEVIDO À
ESCOLHA DE MODELOS NA INFERÊNCIA
ESTATÍSTICA COM APLICAÇÃO EM MODELOS DE
REGRESSÃO LINEAR**

CECILIA CANDOLO

Bacharel em Estatística

Orientadora: Prof.^a Dr.^a CLARICE G. B. DEMÉTRIO

Tese apresentada à Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, para obtenção do título de Doutor em Agronomia, Área de Concentração: Estatística e Experimentação Agronômica.

PIRACICABA

Estado de São Paulo - Brasil

Maior - 2001

ERRATA

- pag. 2, linha 2: substituir á por à;
- pag. 2, linha 12: substituir ser'a por será;
- pag. 4, linha 13: substituir BMA por ponderação Bayesiana de modelos;
- pag. 5, substituir a formulação de $\text{var}(\theta \mid D)$ por

$$\begin{aligned}\text{var}(\theta \mid D) &= \sum_{k=1}^K \text{var}(\theta \mid M_k, D) \Pr(M_k \mid D) \\ &\quad + \sum_{k=1}^K (\text{E}(\theta \mid M_k, D) - \text{E}(\theta \mid D))^2 \Pr(M_k \mid D);\end{aligned}$$

- pag. 6, linha 5: substituir *Markov chain Monte Carlo* por *Monte Carlo Markov chain*;
- pag. 13, linha 11: substituir $1 = 1, \dots, n$ por $i = 1, \dots, n$;
- pag. 16, linha 9: $I(\cdot)$ representa uma função indicadora que assume valor 1 quando o evento é verdadeiro e valor 0 em caso contrário;
- pag. 17 e 18: nas equações (13) e (16) substituir MSE por EQM;
- pag. 23, linha 9: substituir usando por usado;
- pag. 34, linha 17, e pag. 52, linha 11: substituir varância por variância;
- pag. 53, linha 19: substituir $\beta_1 = 0.7$, $\beta_2 = -0.1$ e $\sigma^2 = 0.61$ por $\beta_1 = 0,7$, $\beta_2 = -0,1$ e $\sigma^2 = 0,61$;
- pag. 57, linha 9: substituir Cimatológicas por Climatológicas.

Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - Campus "Luiz de Queiroz"/USP

Candolo, Cecilia

A incorporação da incerteza devido à escolha de modelos na inferência estatística com aplicação em modelos de regressão linear / Cecilia Candolo. - - Piracicaba, 2001.
80 p.

Tese (doutorado) - Escola Superior de Agricultura Luiz de Queiroz, 2001.
Bibliografia.

1. Análise estatística 2. Inferência estatística 3. Método de Bayes 4. Método estatístico
I. Título

CDD 519.54

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

Aos meus pais,
Gaspar e Lúcia,
ofereço.

Aos meus filhos,
Márcio e João,
dedico.

AGRADECIMENTOS

Agradeço a todos aqueles que, direta ou indiretamente, contribuíram para a realização deste trabalho.

À Prof.^a Dr.^a Clarice G. B. Demétrio a orientação e a confiança no desenvolvimento deste trabalho;

Ao Professor Anthony Davison, da Escola Politécnica Federal de Lausane, Suíça, a co-orientação e apoio;

À CAPES/PICDT a bolsa de estudos;

Ao CNPq a bolsa de programa sanduíche no exterior;

Aos professores e funcionários do Departamento de Ciências Exatas da Escola Superior de Agricultura “Luiz de Queiroz” os ensinamentos, a atenção e a amizade;

Aos professores e funcionários do Departamento de Estatística da Universidade Federal de São Carlos o incentivo, a confiança e a amizade;

Aos colegas de curso a ajuda, a amizade e o companheirismo;

Aos meus pais, irmãs, cunhados, sobrinhas e sobrinhos a companhia e os impagáveis almoços de domingo na chácara;

Aos meus filhos, fontes de inspiração, todo imenso carinho, cuidado, atenção, e compreensão.

SUMÁRIO

	Página
LISTA DE FIGURAS	v
LISTA DE TABELAS	viii
RESUMO	ix
SUMMARY	xi
1 INTRODUÇÃO	1
2 REVISÃO DA LITERATURA	3
2.1 Regressão linear	3
2.2 Ponderação Bayesiana de modelos	4
2.2.1 Formulação geral	4
2.2.2 BMA para regressão linear	6
2.3 Ponderação de modelos	9
2.4 O método <i>bootstrap</i>	11
2.4.1 O método <i>bootstrap</i> em modelos de regressão	12
3 METODOLOGIA	14
3.1 Ponderação de modelos em regressão linear simples	14
3.1.1 Variância conhecida	14
3.1.2 Variância desconhecida	16
3.2 Ponderação de modelos em regressão linear múltipla	19
3.2.1 Modelos com duas covariáveis	19
3.2.2 Modelos com três covariáveis	25
3.2.3 Caso Geral	29
3.3 O método <i>bootstrap</i> para calcular a variância do estimador ponderado	31

	iv
3.3.1 Planos de reamostragem	31
3.4 Comparação entre os métodos	32
4 RESULTADOS E DISCUSSÃO	34
4.1 Comparação entre o estimador ponderado e a metodologia usual baseada na escolha de um modelo	34
4.1.1 Regressão linear simples	34
4.1.2 Regressão linear múltipla	39
4.2 Comparação entre o estimador ponderado, o estimador BMA e resul- tados usando o método <i>bootstrap</i>	51
4.2.1 Exemplo 1: Efeito da duração da área foliar sadia na produção de feijão	51
4.2.2 Exemplo 2: Concentração de fósforo	52
4.2.3 Exemplo 3: Volume de cerejeiras	54
4.2.4 Exemplo 4: Temperatura do ar	57
5 CONCLUSÕES	59
REFERÊNCIAS BIBLIOGRÁFICAS	61
APÊNDICE 1	64
APÊNDICE 2	75

LISTA DE FIGURAS

Página

- 1 Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear simples com variância conhecida. Esquerda: $E(Q_+) - \delta$ (pontos), $\text{var}(Q_+)$ (sólido) e $\text{var}(Q_+) + \{E(Q_+) - \delta\}^2$ (traços) para $\alpha = 0,025$. Direita: funções correspondentes para o estimador ponderado usando Q 35
- 2 Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear simples para diferentes valores de α . Linhas fortes: funções $E(Q_+) - \delta$ (pontos), $\text{var}(Q_+)$ (sólido) e $\text{var}(Q_+) + (E(Q_+) - \delta)^2$ (traços) para $\alpha = 0,01, 0,025, 0,05$ e $0,1$. Linhas finas: mesmas funções para Q 36
- 3 Propriedades dos estimadores de θ em regressão linear simples com variância conhecida. As figuras mostram $E(Q_{\text{BIC}}) - \delta$ (pontos), $\text{var}(Q_{\text{BIC}})$ (sólido) e $\text{var}(Q_{\text{BIC}}) + \{E(Q_{\text{BIC}}) - \delta\}^2$ (traços) para $n = 11, 41$ e 101 em comparação com o gráfico correspondente ao uso do AIC, no alto à esquerda. 37
- 4 Propriedades dos estimadores $\hat{\theta}$ e $\hat{\theta}_+$ no modelos de regressão linear simples com variância desconhecida. Esquerda: $E(Q'_+) - \delta$ (pontos), $\text{var}(Q'_+)$ (sólido) e $\text{var}(Q'_+) + (E(Q'_+) - \delta)^2$ (traços) para $\alpha = 0,025$ e $n = 11, 41$ e 101 . Direita: o mesmo para Q' 38
- 5 Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com duas covariáveis e variância conhecida. Esquerda: funções $[E(Q_{+1}) - \delta_1] + [E(Q_{+2}) - \delta_2]$, $\text{var}(Q_{+1}) + \text{var}(Q_{+2})$ e o correspondente EQM, com $\alpha = 0,025$. Direita: mesmas funções para Q_1 e Q_2 40

- 6 Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com duas covariáveis e variância conhecida, fixando-se valores de δ_2 . Esquerda: perfis das funções $[E(Q_{+1}) - \delta_1] + [E(Q_{+2}) - \delta_2]$ (pontos), $\text{var}(Q_{+1}) + \text{var}(Q_{+2})$ (sólido) e o correspondente EQM (traços), com $\alpha = 0,025$. Direita: o mesmo para as funções Q_1 e Q_2 41
- 7 Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com duas covariáveis e variância conhecida quando uma covariável é ignorada. Esquerda: superfícies de $[E(Q_{+1}) - \delta_1] + [E(Q_{+2}) - \delta_2]$, $\text{var}(Q_{+1}) + \text{var}(Q_{+2})$ e o correspondente EQM, com $\alpha = 0,025$. Direita: mesmas superfícies para Q_1 e Q_2 42
- 8 Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com duas covariáveis e variância conhecida, fixando-se valores de δ_2 . Esquerda: perfis das funções $[E(Q_{+1}) - \delta_1] + [E(Q_{+2}) - \delta_2]$ (pontos), $\text{var}(Q_{+1}) + \text{var}(Q_{+2})$ (sólido) e o correspondente EQM (traços), com $\alpha = 0,025$. Direita: o mesmo para as funções Q_1 e Q_2 43
- 9 Propriedades dos estimadores de θ em regressão linear com duas covariáveis e variância conhecida. Topo à esquerda: $\text{var}(Q_1) + \text{var}(Q_2)$ usando AIC. Outras superfícies: o mesmo quando usando BIC no cálculo do peso, para vários valores de n 44
- 10 Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com duas covariáveis e variância desconhecida. Esquerda: superfícies correspondentes ao EQM de $\hat{\theta}_+$, com $\alpha = 0,025$. Direita: superfícies correspondentes ao EQM de $\hat{\theta}$. De alto a baixo: $n = 11, 41$ e 101 45
- 11 Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com duas covariáveis e variância desconhecida, fixando-se valores de δ_2 e $n = 11$. Esquerda: perfis para vício (pontos), variância (sólido) e EQM (traços), com $\alpha = 0,025$, para $\hat{\theta}_+$. Direita: o mesmo para $\hat{\theta}$ 46

- 12 Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com duas covariáveis e variância desconhecida, fixando-se valores de δ_2 e $n = 101$. Esquerda: perfis para vício (pontos), variância (sólido) e EQM (traços), com $\alpha = 0,025$, para $\hat{\theta}_+$. Direita: o mesmo para $\hat{\theta}$ 47
- 13 Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com três covariáveis e variância conhecida, fixando-se $\delta_1 = 2$. Esquerda: superfícies correspondentes ao vício, à variância e ao EQM de $\hat{\theta}_+$, com $\alpha = 0,025$. Direita: mesmas superfícies para $\hat{\theta}$ 48
- 14 Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com três covariáveis e variância conhecida, fixando-se $\delta_2 = 2$. Esquerda: superfícies correspondentes ao vício, à variância e ao EQM de $\hat{\theta}_+$, com $\alpha = 0,025$. Direita: mesmas superfícies para $\hat{\theta}$ 49
- 15 Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com três covariáveis e variância conhecida, fixando-se $\delta_3 = 2$. Esquerda: superfícies correspondentes ao vício, à variância e ao EQM de $\hat{\theta}_+$, com $\alpha = 0,025$. Direita: mesmas superfícies para $\hat{\theta}$ 50
- 16 Histogramas de $\hat{\theta}$ simulado. Esquerda: ponderado; Direita: BMA. A linha pontilhada está sobre o valor verdadeiro de θ 54
- 17 Histogramas de $\hat{\theta}^*$ para três amostras simuladas. No alto: amostra 1, 2 e 3 para o *bootstrap* 1; Embaixo: o mesmo para *bootstrap* 2. A linha pontilhada está sobre o valor verdadeiro de θ e a linha tracejada corresponde às médias *bootstrap* de $\hat{\theta}^*$ 56

LISTA DE TABELAS

Página

1	Variâncias de $\hat{\theta}$ para o exemplo de produção de feijão. V1: variância em (16); V2: variância em (3).	52
2	Modelos, pesos e probabilidades a posteriori dos modelos para o exemplo do fósforo.	53
3	Resultados da simulação para o exemplo do fósforo.	54
4	Resultados para três amostras do estudo de simulação do exemplo do fósforo.	55
5	Modelos, pesos e probabilidades a posteriori dos modelos para o exemplo das cerejeiras.	55
6	Variâncias de $\hat{\theta}$ para o exemplo das cerejeiras. V1: variância em (33); V2: variância em (3).	57
7	Modelos, pesos e probabilidades a posteriori dos modelos para o exemplo da temperatura.	58
8	Variâncias de $\hat{\theta}$ para o exemplo da temperatura. V1: variância em (39); V2: variância em (3).	58

A INCORPORAÇÃO DA INCERTEZA DEVIDO À ESCOLHA DE MODELOS NA INFERÊNCIA ESTATÍSTICA COM APLICAÇÃO EM MODELOS DE REGRESSÃO LINEAR

Autora: CECILIA CANDOLO

Orientadora: Prof.^a Dr.^a CLARICE G. B. DEMÉTRIO

RESUMO

Uma abordagem típica de análise estatística consiste em vários estágios: exploração descritiva do conjunto de dados, definição da classe de modelos a ser considerada, seleção do melhor modelo dentro desta classe de acordo com algum critério pré-estabelecido e obtenção de inferências baseadas no modelo selecionado. Este ciclo é geralmente iterativo e envolve, além da aplicação dos conceitos e técnicas estatísticas, também considerações subjetivas. A conclusão obtida ao final deste processo depende do(s) modelo(s) escolhido(s), sem levar em consideração a incerteza devida à escolha do(s) modelo(s), o que pode resultar na subestimação da variabilidade de quantidades de interesse e em inferências super-otimistas ou viciadas.

Este trabalho considera o problema de como incorporar a incerteza devido à escolha do modelo na inferência estatística aplicando a abordagem de Buckland et al. (1997), que propuseram um estimador ponderado para um parâmetro

comum a todos modelos em estudo, sendo que os pesos desta ponderação são obtidos a partir do uso de critérios de informação ou do método *bootstrap*. Esta abordagem é aplicada a modelos de regressão linear e comparada com o procedimento usual de escolha de um modelo, com ponderação Bayesiana de modelos e com o uso do método *bootstrap*.

O estimador ponderado apresentou erro quadrático médio substancialmente reduzido nas situações em que os coeficientes da regressão são pouco significativos, e comportamento similar ao estimador obtido via ponderação Bayesiana de modelos e com o uso do método *bootstrap*.

INCORPORATING MODEL SELECTION UNCERTAINTY INTO STATISTICAL INFERENCE IN LINEAR REGRESSION

Author: CECILIA CANDOLO

Adviser: Prof.^a Dr.^a CLARICE G. B. DEMÉTRIO

SUMMARY

Statistical data analysis typically has several stages: exploration of the data set; deciding on a class or classes of models to be considered; selecting the best of them according to some criterion and making inferences based on the selected model. The cycle is usually iterative and will involve subject-matter considerations as well as statistical insights. The conclusion reached after such a process depends on the model(s) selected, but the consequent uncertainty is not usually incorporated into the inference. This may lead to underestimation of the uncertainty about quantities of interest and overoptimistic and biased inferences.

This work consider the problem of how to incorporate model selection uncertainty into statistical inference applying the approach of Buckland et al. (1997), who proposed a weighted estimator of a parameter common to all models under study, where the weights are obtained using information criteria or the bootstrap.

This approach is applied to linear regression models and compared with the usual model-choice framework, with Bayesian model averaging and the *bootstrap*.

The weighted estimator behaves similarly to model averaging and the *bootstrap*, with smaller mean squared error than the usual model-selection based estimator when the regression coefficients are not significant.

1 INTRODUÇÃO

Uma abordagem típica de análise estatística consiste em vários estágios: exploração descritiva do conjunto de dados, definição da classe de modelos a ser considerada, seleção do melhor modelo dentro desta classe de acordo com algum critério pré-estabelecido e obtenção de inferências baseadas no modelo selecionado. Este ciclo é geralmente iterativo e envolve, além da aplicação dos conceitos e técnicas estatísticas, também considerações subjetivas. A conclusão obtida ao final deste processo depende do(s) modelo(s) escolhido(s), sem levar em consideração a incerteza devida à escolha do(s) modelo(s), o que pode resultar na subestimação da variabilidade de quantidades de interesse e em inferências super-otimistas ou viciadas. Este fato não é desconhecido, mas tem se mantido assim por muitos anos. A questão da incorporação da incerteza devida à escolha do modelo na análise final começou a ser tratada de forma mais sistemática apenas recentemente.

O maior volume de sistematização e discussão deste assunto tem sido sob o enfoque Bayesiano. Chatfield (1995) e Draper (1995) discutem o custo de ignorar a incerteza devido à escolha do modelo, com interpretação Bayesiana enquanto que a revisão sobre ponderação Bayesiana de modelos apresentada por Hoeting et al. (1999) fornece uma perspectiva histórica sobre combinação de modelos além de outras referências. Nessa interpretação, a seleção do modelo é substituída por uma média ponderada obtida para todos os modelos possíveis, para a estimação da quantidade de interesse. O método pode ser sensível à escolha de distribuições a priori e se houver muitos modelos possíveis o procedimento de ponderação poderá não ser

uma proposição prática.

Mais recentemente, um procedimento frequentista proposto por Buckland et al. (1997), incorpora a incerteza devido á seleção de modelos na inferência estatística usando métodos mais simples de ponderação, onde os pesos são obtidos a partir de critérios de informação ou com o uso do método *bootstrap*. Esta abordagem fornece ao usuário técnicas de aplicação mais fácil e através de exemplos parece ter bons resultados práticos, mas teoricamente não está bem justificada.

A proposta deste trabalho é estudar este procedimento de ponderação de modelos em situações de regressão linear simples e múltipla obtendo vício, variância e erro quadrático médio. Estas quantidades serão comparadas com as obtidas no procedimento padrão de seleção de um modelo. Também, através de exemplos da literatura, esta ponderação ser'á comparada com o procedimento Bayesiano e com resultados obtidos com o uso do método *bootstrap*. O processo de seleção de modelos a ser considerado resume-se à seleção de covariáveis. O método *bootstrap* será utilizado no contexto da obtenção de propriedades do estimador ponderado, principalmente a variância. O *bootstrap* tem a vantagem de utilizar reamostragens para a construção de instrumentos inferenciais, evitando o desenvolvimento analítico.

É importante ressaltar que a ponderação de modelos apenas faz sentido se as quantidades que estão sendo ponderadas têm a mesma interpretação para todos os modelos em consideração. Assim, ponderar valores de parâmetros ou estimativas relativos a um modelo particular não será pertinente ao contexto de ponderação de modelos. Por este motivo, neste trabalho, o enfoque é na estimação da predição, tratando os modelos como caixas pretas que produzem as estimativas, sem ter um interesse intrínseco no conteúdo das caixas.

Este estudo está concentrado em situações simples de regressão linear, mas este tipo de situação não só é um paradigma para outros problemas de predição como também serve como uma aproximação para situações mais complexas.

2 REVISÃO DA LITERATURA

2.1 Regressão linear

A Análise de Regressão é usada para estudar relações entre variáveis mensuráveis. Regressão linear é usada para uma classe especial de relações: aquelas que podem ser descritas por linhas retas, ou por generalizações de linhas retas para muitas dimensões, especificamente, regressão linear simples e múltipla. Estas técnicas são aplicadas a dados de quase todas as áreas da ciência, e as razões para se ajustarem modelos de regressão linear são tantas quanto suas aplicações, mas as mais comuns são a descrição de uma relação e a predição de um valor futuro. Em muitas aplicações de regressão linear múltipla, uma das propostas da análise é decidir quais covariáveis devem ser incluídas no modelo final.

A literatura na área de análise de regressão linear e seleção de covariáveis é vasta. Citam-se aqui, como referências básicas Draper & Smith (1981) e Weisberg (1985). A metodologia básica para selecionar a melhor equação de regressão é baseada, dentre outros, em três critérios: R^2 , obtido a partir do ajuste de mínimos quadrados, s^2 , o quadrado médio dos resíduos e a estatística C_p de Mallows, Mallows (1973), que leva em consideração a soma de quadrados dos resíduos, o número de parâmetros do modelo e o quadrado médio dos resíduos relativo ao ajuste do modelo contendo todas as covariáveis em estudo.

Em um contexto mais amplo que o de escolha de modelos a partir da seleção de variáveis, estão o Critério de Informação de Akaike, AIC, proposto por Akaike (1973) e uma modificação deste, o Critério Bayesiano de Informação, BIC,

proposto por Schwarz (1978). Ambos são baseados na função de verossimilhança maximizada do modelo em estudo. O AIC assume que o modelo verdadeiro para os dados requer uma grande quantidade de parâmetros, mas busca apenas o melhor modelo para o ajuste, não sendo um critério para a estimação da ordem do modelo. Em contraste, o BIC fornece uma estimação consistente da ordem do modelo, assumindo que o verdadeiro modelo existe e tem dimensão pequena.

2.2 Ponderação Bayesiana de modelos

A idéia de combinar os modelos foi primeiro mencionada por Barnard (1963), em um contexto de controle de qualidade, que ao propor o uso dos desvios acumulados ao invés de desvios puros, indicou uma maior sensibilidade em detectar mudanças persistentes. Já num contexto Bayesiano, Roberts (1965) sugeriu uma distribuição que combina dois modelos, produzindo uma média ponderada de duas distribuições a posteriori de modelos, que é similar ao BMA. De uma forma paralela, durante a década de 70, houve muitas publicações na área de economia combinando predições a partir de diferentes modelos de previsão, como relatado por Clemen (1989) em uma revisão detalhada.

Foi a partir da década de 90 que a metodologia de ponderação Bayesiana de modelos foi sistematizada, apesar de ter sido Leamer (1978) quem apresentou a idéia básica, apontando para uma solução Bayesiana envolvendo a incerteza contida na escolha de modelos. Referências básicas, além do já citado artigo de Hoeting et al. (1999) são Hoeting (1994), Madigan & York (1993), Madigan & Raftery (1994), Kass & Raftery (1995). À parte, outros trabalhos descreveram os riscos de ignorar a incerteza devido à escolha do modelo em uma análise estatística, como os já citados Draper (1995) e Chatfield (1995).

2.2.1 Formulação geral

Ponderação Bayesiana de modelos, ou BMA (do inglês *Bayesian Model Averaging*), é uma solução Bayesiana padrão ao problema de como incorporar a

incerteza devido à seleção de modelos na inferência estatística. Se θ é a quantidade de interesse, tal como um valor de predição num modelo linear, e $\mathcal{M} = \{M_1, \dots, M_K\}$ denota o conjunto de todos os modelos possíveis, então a distribuição a posteriori de θ dados os dados D é

$$\Pr(\theta \mid D) = \sum_{k=1}^K \Pr(\theta \mid M_k, D) \Pr(M_k \mid D),$$

que é uma média das distribuições a posteriori sob cada modelo considerado, $M_1, \dots, M_K$, ponderados pelas probabilidades a posteriori dos modelos respectivos. A probabilidade a posteriori do modelo é

$$\Pr(M_k \mid D) = \frac{\Pr(D \mid M_k) \Pr(M_k)}{\sum_{l=1}^K \Pr(D \mid M_l) \Pr(M_l)},$$

onde

$$\Pr(D \mid M_k) = \int \Pr(D \mid \eta_k, M_k) \Pr(\eta_k \mid M_k) d\eta_k$$

é a integral da verossimilhança do modelo M_k (ou sua verossimilhança marginal), η_k é o vetor de parâmetros do modelo M_k (para regressão, $\eta_k = (\beta_k, \sigma^2)$), $\Pr(\eta_k \mid M_k)$ é a densidade a priori de η_k sob o modelo M_k , $\Pr(D \mid \eta_k, M_k)$ é a verossimilhança e $\Pr(M_k)$ é a probabilidade a priori de que M_k é o modelo verdadeiro. Todas as probabilidades são implicitamente condicionais a $\mathcal{M}$.

A média e a variância a posteriori de θ são dadas por

$$\begin{aligned} \mathbb{E}(\theta \mid D) &= \sum_{k=1}^K \mathbb{E}(\theta \mid M_k, D) \Pr(M_k \mid D) \quad \text{e} \\ \text{var}(\theta \mid D) &= \sum_{k=1}^K \text{var}(\theta \mid M_k, D) \Pr(M_k \mid D) \\ &\quad + \sum_{k=1}^K (\mathbb{E}(\theta \mid M_k, D) - \mathbb{E}(\theta \mid D))^2 \Pr(M_k \mid D). \end{aligned}$$

A implementação do BMA pode ser dificultada por duas razões: dificuldade no cálculo das integrais em $\Pr(D \mid M_k)$ e o número de termos na soma em $\Pr(\theta \mid D)$ pode ser enorme. Uma alternativa para resolver o segundo problema é obter a média para um subconjunto dos modelos que sejam mais indicados pelos dados. Madigan & Raftery (1994) propuseram um método chamado *Occam's window*,

com o qual se obtém a média de um conjunto de modelos parsimoniosos e indicados pelos dados, selecionados a partir da aplicação de técnicas padrões da pesquisa científica. Outra alternativa é aproximar a soma em $\Pr(\theta \mid D)$ usando a abordagem de Monte Carlo via cadeias de Markov, também conhecida por MCMC (do inglês *Markov chain Monte Carlo*). Madigan & York (1995) apresentaram uma metodologia chamada composição de modelos via MCMC, que denominaram de MC³, a qual gera um processo estocástico que se move através do espaço de modelos. Para isso, constrói-se uma cadeia de Markov $\{M(t), t = 1, 2, \dots\}$ com espaço de estados $\mathcal{M}$ e distribuição de equilíbrio $\Pr(M_i \mid D)$, e esta cadeia de Markov é simulada para obter observações $M(1), \dots, M(N)$. Então, sob certas condições de regularidade, para qualquer função $g(M_i)$ definida em $\mathcal{M}$, a média

$$\hat{G} = \frac{1}{N} \sum_{t=1}^N g(M(t))$$

é um estimador de $E(g(M))$. Para calcular $\Pr(\theta \mid D)$ desta maneira faz-se $g(M) = \Pr(\theta \mid D)$.

Quanto à dificuldade no cálculo das integrais em $\Pr(D \mid M_k)$, Hoeting et al. (1999) fornecem detalhes de implementação do BMA para algumas classes de modelos, incluindo, além de regressão linear, modelos lineares generalizados, análise de sobrevivência e modelos gráficos. O uso do método de Laplace pode fornecer boas aproximações para $\Pr(D \mid M_k)$, como pode ser visto em Tierney & Kadane (1986).

2.2.2 BMA para regressão linear

O desenvolvimento de BMA para regressão linear é apresentado por Hoeting (1994) e por Raftery et al. (1997), que fornecem o cálculo apropriado de $\Pr(D \mid M_k)$, usando como classe de distribuições a priori conjugadas, normais padrões, e a distribuição a posteriori preditiva a partir da qual são obtidas $E(\theta \mid M_k, D)$ e $\text{var}(\theta \mid M_k, D)$.

Seja

$$\mathbf{Y} = \beta_0 + \beta_1 \mathbf{x}_{1j} + \dots + \beta_p \mathbf{x}_{pj} + \varepsilon_j = \mathbf{X}\beta + \varepsilon$$

a forma de cada modelo considerado, onde $\mathbf{Y}$ é um vetor $n \times 1$ de variáveis dependentes, $\mathbf{X}$ é uma matrix $n \times (p+1)$ de covariáveis observadas e β é o vetor $(p+1) \times 1$ de parâmetros. Assume-se que os erros são independentes com distribuição $N(0, \sigma^2)$ e β e σ^2 são desconhecidos. Para o cálculo das integrais em $\Pr(D \mid M_k)$, Hoeting (1994) adotou a classe das distribuições a priori conjugadas gama-normal padrão, com a seguinte parametrização

$$\beta \sim \mathbf{N}(\mu, \sigma^2 \mathbf{V}) \quad \text{e}$$

$$\frac{\nu\lambda}{\sigma^2} \sim \chi_\nu^2,$$

onde ν , λ , a matrix $(p+1) \times (p+1)$ $\mathbf{V}$ e o vetor $(p+1)$ μ são hiperparâmetros a serem escolhidos. Para covariáveis não categorizadas, Hoeting (1994) usa como priori

$$\mu = (\hat{\beta}_0, 0, 0, \dots, 0),$$

onde $\hat{\beta}_0$ é o estimador de mínimos quadrados ordinário de β_0 , e a matrix de covariância $\mathbf{V}$ que tem diagonal dada por

$$(s_Y^2, \phi^2 s_1^{-2}, \dots, \phi^2 s_p^{-2}),$$

onde s_Y^2 é a variância amostral de $\mathbf{Y}$, s_i^2 é a variância amostral de $\mathbf{X}_i$, para $i = 1, \dots, p$, e ϕ é um hiperparâmetro a ser escolhido. Os valores dos hiperparâmetros ν , λ e ϕ são obtidos, também, a partir do trabalho de Hoeting (1994), que fornece completa argumentação para a escolha de $\nu = 2,58$, $\lambda = 0,28$ and $\phi = 2,85$.

A função de verossimilhança marginal de $\mathbf{Y}$ sob um modelo M_k , com base nas distribuições a priori descritas, é dada por

$$\begin{aligned} \Pr(\mathbf{Y} \mid \mu_k, \mathbf{V}_k, \mathbf{X}_k, M_k) &= \frac{\Gamma\left(\frac{\nu+n}{2}(\mathbf{Y}-\mathbf{X}_k\mu_k)\right)(\nu\lambda)^{\nu/2}}{\pi^{n/2}\Gamma\left(\frac{\nu}{2}\right)|I+\mathbf{X}_k\mathbf{V}_k\mathbf{X}_k^T|^{1/2}} \times \left[\lambda\nu + (\mathbf{Y}-\mathbf{X}_k\mu_k)^T \right. \\ &\quad \left. \times (I + \mathbf{X}_k\mathbf{V}_k\mathbf{X}_k^T)^{-1}(\mathbf{Y}-\mathbf{X}_k\mu_k)\right]^{-(\nu+n)/2}, \end{aligned}$$

onde $\mathbf{X}_k$ é a matrix de covariáveis, μ_k é o vetor de médias de β e $\mathbf{V}_k$ é a matrix de covariâncias de β correspondendo ao modelo k . Esta distribuição é uma t de

Student não-central de dimensão n com ν graus de liberdade, média $\mathbf{X}\mu$ e variância $[\nu/(\nu-2)]\lambda(I + \mathbf{X}\mathbf{V}\mathbf{X}^T)$. Assumindo que todos os modelos possíveis, ou seja, todas as possíveis combinações de covariáveis, são igualmente prováveis a priori, então

$$\Pr(M_k | D) = \frac{\Pr(\mathbf{Y}|\mu_k, \mathbf{V}_k, \mathbf{X}_k, M_k)}{\sum_{l=1}^K \Pr(\mathbf{Y}|\mu_k \lambda, \mathbf{V}_l, \mathbf{X}_l, M_l)}.$$

Considerando que $\theta = \mathbf{Z}\beta + \varepsilon$, com $\varepsilon \sim N(0, \sigma^2 I)$, é um valor de predição, onde $\mathbf{Z}$ é uma matriz $1 \times (p+1)$ de preditores conhecidos, β é o vetor de parâmetros $(p+1)$, e que as distribuições a priori dos parâmetros são

$$\beta \sim \mathbf{N}_{p+1}(\mu', \sigma^2 \mathbf{V}') \quad \text{e}$$

$$\frac{\nu' \lambda'}{\sigma^2} \sim \chi_{\nu'}^2,$$

onde μ' , $\mathbf{V}'$, ν' e λ' são os parâmetros a posteriori dados por

$$\begin{aligned} \mu' &= (\mathbf{X}^T \mathbf{X} + \mathbf{V}^{-1})^{-1} (\mathbf{X}^T \mathbf{Y} + \mathbf{V}^{-1} \mu) \\ \mathbf{V}' &= (\mathbf{X}^T \mathbf{X} + \mathbf{V}^{-1})^{-1} \\ \nu' &= n + \nu \\ \lambda' &= \frac{1}{n+\nu} \left[\nu \lambda + \mu^T \mathbf{V}^{-1} \mu + \mathbf{Y}^T \mathbf{Y} - \right. \\ &\quad \left. (\mathbf{X}^T \mathbf{Y} + \mathbf{V}^{-1} \mu)^T (\mathbf{X}^T \mathbf{X} + \mathbf{V}^{-1})^{-1} (\mathbf{X}^T \mathbf{Y} + \mathbf{V}^{-1} \mu) \right], \end{aligned}$$

então, a distribuição a posteriori preditiva é

$$\begin{aligned} f(\theta | \mathbf{Y}) &= \frac{\Gamma\left(\frac{n+\nu+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \Gamma((n+\nu)/2)} (n+\nu)^{(n+\nu)/2} |\lambda'(\mathbf{Z}\mathbf{V}'\mathbf{Z}^T + \mathbf{1})|^{-1/2} \\ &\quad \times \left\{ (n+\nu) + \frac{(\theta - \mathbf{Z}\mu')^2}{\lambda'(\mathbf{Z}\mathbf{V}'\mathbf{Z}^T + \mathbf{1})} \right\}^{-(n+\nu+1)/2}, \end{aligned}$$

que é uma distribuição t de Student com média $\mathbf{Z}\mu'$, variância $\frac{n+\nu}{n+\nu-2} \lambda'(\mathbf{Z}\mathbf{V}'\mathbf{Z}^T)$ e $(n+\nu)$ graus de liberdade, onde n é o número de dados observados. Desta forma, obtém-se

$$E(\theta | M_k, D) = \mathbf{Z}\mu'_k$$

e, seguindo a formulação aqui utilizada para a variância da predição,

$$\text{var}(\theta | M_k, D) = \frac{n+\nu}{n+\nu-2} \lambda'_k(\mathbf{Z}\mathbf{V}'_k \mathbf{Z}^T).$$

Finalmente, a média e a variância a posteriori de θ são

$$\begin{aligned} E(\theta | D) &= \sum_{l=1}^K E(\theta | M_l, D) \Pr(M_l | D) \quad \text{e} \\ \text{var}(\theta | D) &= \sum_{l=1}^K \text{var}(\theta | M_l, D) \Pr(M_l | D) \\ &\quad + \sum_{l=1}^K (E(\theta | M_l, D) - E(\theta | D))^2 \Pr(M_l | D). \end{aligned}$$

Hoeting et al. (1999) fornecem indicação de como obter programas em S-PLUS para o cálculo das probabilidades a posteriori dos modelos e também para o uso dos métodos *Occam's window* e MC³.

2.3 Ponderação de modelos

No contexto frequentista, além das aplicações em economia citadas em 2.2, o trabalho de Buckland et al. (1997) buscou uma metodologia de fácil aplicação, indicando o uso de critérios de informação e do método *bootstrap* como forma de ponderar os modelos.

Esta abordagem de ponderação de modelos assume que em uma situação de K modelos em estudo, $M_1, \dots, M_K$, onde o objetivo é estimar um parâmetro de interesse θ , cada modelo ajustado fornece um estimador deste parâmetro θ , $\hat{\theta}_k$, e um peso W_k , de forma que $\sum W_k = 1$. Assim, o estimador geral de θ é

$$\hat{\theta} = \sum_k W_k \hat{\theta}_k. \quad (1)$$

Os pesos W_k podem ser obtidos usando o Critério de Informação de Akaike, AIC (Akaike, 1973), o Critério de Informação Bayesiano, BIC (Schwarz, 1978), ou através do uso do *bootstrap*, no qual W_k é estimado pela proporção de amostras *bootstrap* nas quais M_k é identificado como o melhor modelo. Usando um dos critérios de informação, que são dados pela expressão $I_k = -2 \log L_k + q_k$, onde L é a função de verossimilhança maximizada para o modelo k e q é uma penalidade igual a $2p$ para o AIC e $p \log(n)$ para o BIC, com p sendo o número de parâmetros do modelo k , os

pesos são definidos como

$$W_k = \frac{\exp(-I_k/2)}{\sum_{i=1}^K \exp(-I_i/2)}, \quad k = 1, \dots, K. \quad (2)$$

Para estimar a variância de (1) quando não é usado o método *bootstrap*, Buckland et al. (1997) supõem que os pesos W_k são constantes conhecidas, que o vício na estimação de θ , B_k , é zero e que os estimadores $\hat{\theta}_k$ e $\hat{\theta}_l$, $k \neq l$, são perfeitamente correlacionados. Sob estas condições obtiveram a seguinte expressão

$$\text{var}(\hat{\theta}) = \left\{ \sum_k W_k \sqrt{\text{var}(\hat{\theta}_k | B_k) + B_k^2} \right\}^2, \quad (3)$$

que pode ser estimada substituindo o vício $B_k = E(\hat{\theta}_k) - \theta$ por $\hat{B}_k = \hat{\theta}_k - \hat{\theta}$, e usando $\widehat{\text{var}}(\hat{\theta}_k | B_k)$ calculada pelos métodos usuais de inferência, assumindo que o modelo k é o verdadeiro.

Uma outra abordagem para obter $\text{var}(\hat{\theta})$ é através do uso do método *bootstrap*, onde o estimador $\hat{\theta}$ pode ser calculado em cada reamostra para estimar a variância e gerar intervalos de confiança. No caso de situações de regressão, usualmente a reamostragem é feita a partir dos resíduos da regressão, como indicado por Efron (1979) e demonstrado por Freedman (1981), pois desta forma a análise continua condicional aos valores das covariáveis. No presente caso, o uso do *bootstrap* provoca uma influência muito grande do modelo a partir do qual os resíduos foram calculados. Pode parecer mais adequado, neste caso, amostrar os vetores de observações, mas algumas alternativas têm sido consideradas. Buckland et al. (1997) sugerem três alternativas:

- (i) gerar todas as amostras *bootstrap* a partir do modelo selecionado na análise preliminar dos dados originais,
- (ii) gerar as amostras a partir do modelo mais completo e
- (iii) selecionar os modelos a partir dos pesos como calculados na equação (2) e, então, gerar a próxima amostra a partir do modelo selecionado.

2.4 O método *bootstrap*

O método *bootstrap*, introduzido por Efron (1979), tem a vantagem de evitar desenvolvimentos analíticos e tem sido uma das técnicas mais utilizadas e expandidas, com aplicações nas mais diversas áreas. Referências básicas são os livros de Efron & Tibshirani (1993) e Davison & Hinkley (1997).

Seja $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ o resultado das variáveis aleatórias independente e identicamente distribuídas $Y_1, \dots, Y_n$, com distribuição de probabilidade desconhecida F , que depende de um parâmetro desconhecido θ , dado por $\theta = s(F)$, sendo $s(\cdot)$ a função que define θ . O parâmetro θ é estimado por $\hat{\theta} = s(\hat{F})$, onde $\hat{F}$ é um estimador de F obtido a partir da amostra $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$. Seja $\hat{F}$ a distribuição empírica de $\mathbf{y}$, que coloca probabilidades iguais a $1/n$ para cada valor da amostra, y_i , $i = 1, \dots, n$, isto é,

$$\hat{F}(y) = \frac{\#\{y_i \leq y\}}{n}, \quad (4)$$

onde $\#\{A\}$ indica o número de vezes em que o evento A ocorre. O método *bootstrap* tem como princípio substituir a distribuição desconhecida F por $\hat{F}$ para estimar θ , aproximando a distribuição de $\theta = s(F)$ pela de $\hat{\theta}^* = s(\mathbf{y}^*, \hat{F})$, onde $\mathbf{y}^*$ é uma amostra aleatória de tamanho n , retirada de $\hat{F}$. A amostra $\mathbf{y}^* = (\mathbf{y}_1^*, \dots, \mathbf{y}_n^*)$ é chamada de amostra *bootstrap* e sua definição é análoga a obter uma amostra aleatória de tamanho n retirada com reposição da população de tamanho n ($y_1, \dots, y_n$). A partir da amostra *bootstrap* é calculada uma repetição *bootstrap* de $\hat{\theta}$, $\hat{\theta}^* = s(\mathbf{y}^*)$. Repetindo este procedimento um número suficientemente grande de vezes, calcula-se uma distribuição empírica de $\hat{\theta}^*$ e, a partir desta, obtêm-se média, erro padrão, intervalo de confiança etc. Observa-se que a única suposição feita é a de que as observações ($y_1, \dots, y_n$) são independentes e identicamente distribuídas.

Quando $\hat{F}$ é definida como em (4) o método *bootstrap* é denominado de *bootstrap* não-paramétrico, e, quando $\hat{F}$ é definida como uma distribuição de probabilidade com os parâmetros substituídos pelas estimativas obtidas na amostra ($y_1, \dots, y_n$), o método *bootstrap* é denominado de *bootstrap* paramétrico.

As propriedades assintóticas do método *bootstrap* começaram a ser

desenvolvidas a partir do artigo de Bickel & Freedman (1981) que discutiram algumas condições de regularidade sob as quais o *bootstrap* é consistente, considerando estatísticas *estudentizadas*. Este trabalho foi seguido por outros e ainda hoje é crescente a literatura em desenvolvimentos teóricos concernentes ao método *bootstrap*, propondo modificações, como suavização dos dados e sub-amostragem, entre outras, para garantir a consistência do *bootstrap* para classes de estatísticas mais complexas.

Outro avanço na teoria do *bootstrap* teve início com o trabalho de Singh (1981) que introduziu o uso de expansões de Edgeworth para explicar o desempenho do método *bootstrap*, também para estatísticas *estudentizadas*. Este assunto foi abordado por vários autores e o trabalho de Hall (1992) é uma obra de referência na teoria de expansões de Edgeworth aplicadas ao *bootstrap*.

2.4.1 O método *bootstrap* em modelos de regressão

Um aspecto do uso do método *bootstrap* de interesse a este trabalho é a aplicação em modelos de regressão. Este assunto foi primeiramente abordado por Freedman (1981), que mostrou que a aproximação *bootstrap* para a distribuição dos estimadores de mínimos quadrados é válida para modelos de regressão e correlação. O principal ponto a ser abordado aqui diz respeito à forma de reamostragem em problemas de regressão, pois se a simulação é feita de maneira consistente com o modelo adotado, o resultado assintótico obtido pelo método *bootstrap* será o mesmo que aquele obtido pelos métodos clássicos. Seja o modelo de regressão $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, onde $\mathbf{y}$ é o vetor de respostas, $\mathbf{X}$ é a matriz de covariáveis não aleatória, β é o vetor de parâmetros desconhecidos e ε é o vetor de erros independentes e identicamente distribuídos, com distribuição F , média zero e variância σ^2 , onde F e σ^2 são geralmente desconhecidos. O plano de reamostragem para que se obtenha o mesmo delineamento que têm os dados, isto é $x_j^* \equiv x_j$, especifica que y_j^* tenha distribuição condicional dado x_j^* , e é obtido por

$$y_j^* = X\hat{\beta} + e_j^*, \quad j = 1, \dots, n,$$

onde $\hat{\beta}$ é a estimativa de β . Os erros e_j^* são aleatoriamente amostrados de $\hat{F}_e$, a

distribuição empírica dos resíduos $e_i = r_i - \bar{r}$, onde $r_i = y_i - \hat{y}_i$ e $\bar{r}$ é sua média, para $i = 1, \dots, n$. Davison & Hinkley (1997), Cap. 6, indicam que para melhores resultados práticos é melhor usar os resíduos *estudentizados*

$$r_i = \frac{y_i - \hat{y}_i}{(1 - h_i)^{1/2}},$$

onde h_i é o i -ésimo elemento da diagonal da matriz $X(X^T X)^{-1} X^T$, pois a variância deste resíduo concorda com a de ε . A partir dos valores $y_1^*, \dots, y_n^*$ obtém-se $\hat{\beta}^* = (X^T X)^{-1} X^T Y^*$ e a distribuição de $\sqrt{n}(\hat{\beta}^* - \hat{\beta})$ aproxima a distribuição de $\sqrt{n}(\hat{\beta} - \beta)$.

Uma abordagem diferente de reamostragem ocorre quando os dados são considerados como amostras de uma distribuição bivariada de (X, Y) . Neste caso $\hat{F}$ é a distribuição empírica dos vetores de observações $(\mathbf{x}_i, y_i)$, colocando probabilidade $1/n$ a cada um deles, para $i = 1, \dots, n$, obtendo-se, então, uma amostra *bootstrap* de n pares. Há diferenças importantes entre estes dois métodos de reamostragem: o segundo método não faz suposições quanto à homogeneidade de variância, tendo por um lado mais robustez à heterocedasticidade, mas por outro lado pode ser ineficiente se o modelo de variância constante é o correto. Outra diferença é que as amostras são obtidas com diferentes delineamentos, uma vez que os valores de X são obtidos aleatoriamente.

3 METODOLOGIA

3.1 Ponderação de modelos em regressão linear simples

3.1.1 Variância conhecida

Considere o modelo de regressão linear simples

$$y_j = \beta_0 + \beta_1 x_j + \varepsilon_j, \quad j = 1, \dots, n,$$

sendo β_0 e β_1 os parâmetros desconhecidos, $x_1, \dots, x_n$ os valores da covariável com a restrição $\sum x_j = 0$, ou seja $x_j = X_j - \bar{X}$ é uma variável centrada, e $\varepsilon_j \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$, onde σ^2 é conhecido. O objetivo é prever a média $\theta = \beta_0 + \beta_1 x_+$ de uma observação futura Y_+ para o valor de $x = x_+$. Os dois estimadores possíveis são

$$\begin{aligned} M_1 &: \hat{\theta}_1 = \hat{\beta}_0 + \hat{\beta}_1 x_+ \quad \text{e} \\ M_2 &: \hat{\theta}_2 = \hat{\beta}_0. \end{aligned}$$

Como $\sum x_j = 0$, $\hat{\beta}_0 = \bar{y}$ sob os dois modelos. Se $\beta_1 \neq 0$, $\hat{\theta}_1$ é não viciado com variância $\sigma^2/n + x_+^2 v$, onde $v = \sigma^2 / \sum x_j^2$ é a variância de $\hat{\beta}_1$, e $\hat{\theta}_2$ tem vício $\hat{\beta}_1 x_+$ e variância σ^2/n , o que resulta nos erros quadráticos médios

$$\sigma^2/n + x_+^2 v \quad \text{e} \quad \sigma^2/n + \hat{\beta}_1^2 x_+^2,$$

respectivamente. Nesta situação, considerando $\delta = \beta_1/v^{1/2}$, o coeficiente angular padronizado, tem-se

$$\sigma^2/n + x_+^2 v \quad \text{e} \quad \sigma^2/n + x_+^2 v \delta^2,$$

e então M_2 é melhor que M_1 quando $|\delta| < 1$.

Usando-se (1), obtém-se

$$\hat{\theta} = W_1 \hat{\theta}_1 + W_2 \hat{\theta}_2 = \hat{\beta}_0 + x_+ W_1 \hat{\beta}_1,$$

onde,

$$W_1 = \frac{\exp(-AIC_1/2)}{\exp(-AIC_1/2) + \exp(-AIC_2/2)} = \frac{\exp((AIC_2 - AIC_1)/2)}{1 + \exp((AIC_2 - AIC_1)/2)}.$$

Como a variância é conhecida, o logaritmo da função de verossimilhança maximizada para o modelo k é

$$\log L_k = const - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \text{SQR}_k,$$

sendo SQR_k a soma de quadrados dos resíduos da regressão para o modelo k . Assim,

$$AIC_2 - AIC_1 = \frac{1}{\sigma^2} (\text{SQR}_2 - \text{SQR}_1) + 2(p_2 - p_1),$$

onde p é o número de parâmetros do modelo. Como $\sum x_j = 0$, $\text{SQR}_2 - \text{SQR}_1 = \hat{\beta}_1^2 \sum x_j^2$ e escrevendo $\hat{\beta}_1 = v^{1/2}(Z + \delta)$, com $Z = (\hat{\beta}_1 - \beta_1)/v^{1/2}$ sendo uma variável normal padrão, chega-se a

$$AIC_2 - AIC_1 = \hat{\beta}_1^2/v - 2 = (Z + \delta)^2 - 2.$$

Então

$$W_1 \hat{\beta}_1 = v^{1/2}(\delta + Z)G\{(Z + \delta)^2/2 - 1\},$$

com $G(u) = \exp(u)/(1 + \exp(u))$, e

$$\hat{\theta} = \hat{\beta}_0 + x_+ v^{1/2} Q,$$

onde $Q = (Z + \delta)G\{(Z + \delta)^2/2 - 1\}$. O vício, a variância e o erro quadrático médio de $\hat{\theta}$ ficam dados por

$$E(\hat{\theta} - \theta) = x_+ v^{1/2} [E(Q) - \delta], \quad (5)$$

$$\text{var}(\hat{\theta}) = \frac{\sigma^2}{n} + x_+^2 v \text{var}(Q) \quad \text{e} \quad (6)$$

$$\text{EQM}(\hat{\theta}, \theta) = \frac{\sigma^2}{n} + x_+^2 v [\text{var}(Q) + (E(Q) - \delta)^2]. \quad (7)$$

Desta forma, obtêm-se as propriedades do estimador sem a necessidade de suposições como para a obtenção de (3).

Um dos objetivos deste trabalho é comparar o estimador obtido a partir da ponderação dos modelos com a situação usual onde se obtém a estimativa a partir do modelo selecionado. Para isso, considere que a escolha entre os modelos M_1 e M_2 é feita com base em que M_1 é indicado quando $|\hat{\beta}_1|/v^{1/2} > z_{1-\alpha}$, que corresponde a um teste bilateral da hipótese $\beta_1 = 0$ ao nível $(1 - 2\alpha) \times 100\%$. Um estimador geral de θ é, então,

$$\hat{\theta}_+ = \hat{\beta}_0 + x_+ \hat{\beta}_1 \left\{ I(\hat{\beta}_1/v^{1/2} < z_\alpha) + I(\hat{\beta}_1/v^{1/2} > z_{1-\alpha}) \right\}.$$

Fazendo os mesmos cálculos como na situação anterior, $\hat{\theta}_+$ pode ser escrito como

$$\hat{\theta}_+ = \hat{\beta}_0 + x_+ v^{1/2} Q_+,$$

onde $Q_+ = (Z + \delta)\{I(Z < z_\alpha - \delta) + I(Z > z_{1-\alpha} - \delta)\}$, e então obtêm-se

$$E(\hat{\theta}_+ - \theta) = x_+ v^{1/2} [E(Q_+) - \delta], \quad (8)$$

$$\text{var}(\hat{\theta}_+) = \frac{\sigma^2}{n} + x_+^2 v \text{var}(Q_+) \quad \text{e} \quad (9)$$

$$\text{EQM}(\hat{\theta}_+, \theta) = \frac{\sigma^2}{n} + x_+^2 v [\text{var}(Q_+) + (E(Q_+) - \delta)^2]. \quad (10)$$

Usando-se BIC em W_1 no lugar de AIC obtêm-se as mesmas expressões para vício, variância e EQM como em (5), (6) e (7), com Q substituído por

$$Q_{\text{BIC}} = (\delta + Z)G[(Z + \delta)^2/2 - (\log n)/2],$$

que depende n .

3.1.2 Variância desconhecida

Em situações de aplicações em problemas reais, σ^2 é desconhecido. Se σ^2 é estimado a partir do modelo completo M_1 , a variância estimada de $\hat{\beta}_1$ é

$\hat{v} = \hat{\sigma}^2 / \sum x_j^2$, e assim M_1 é indicado quando $|\hat{\beta}_1| / \hat{v}^{1/2} > t_{n-2, 1-\alpha}$ e então

$$\hat{\theta}_+ = \hat{\beta}_0 + x_+ \hat{\beta}_1 \{I(\hat{\beta}_1 / \hat{v}^{1/2} < -t_{n-2, 1-\alpha}) + I(\hat{\beta}_1 / \hat{v}^{1/2} > t_{n-2, 1-\alpha})\}.$$

Como $(n-2)\hat{\sigma}^2/\sigma^2$ tem distribuição qui-quadrado com $n-2$ graus de liberdade,

$$\hat{v} = \frac{\sigma^2}{\sum x_j^2} \frac{C}{(n-2)} = v \frac{C}{(n-2)},$$

onde $C \sim \chi_{n-2}^2$ e $v = \sigma^2 / \sum x_j^2$. Também, sabe-se que

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{v}^{1/2}} \sim t_{n-2},$$

e então, $\hat{\beta}_1 = v^{1/2}([C/(n-2)]^{1/2}T + \delta)$, onde $T \sim t_{n-2}$. Além disso, $\hat{v} = vC/(n-2)$ e $\beta_1 = \delta v^{1/2}$. Como $T = Z/\sqrt{C/(n-2)}$ obtém-se $\hat{\beta}_1 = v^{1/2}(Z + \delta)$. Usando-se estes resultados, tem-se

$$\frac{\hat{\beta}_1}{\hat{v}^{1/2}} = \frac{(Z+\delta)}{\sqrt{C/(n-2)}}.$$

Pode-se então escrever

$$\hat{\theta}_+ = \hat{\beta}_0 + x_+ v^{1/2} Q'_+,$$

onde

$$Q'_+ = (Z + \delta) \left\{ I \left(\frac{(Z+\delta)}{\sqrt{C/(n-2)}} < t_{n-2, \alpha} \right) + I \left(\frac{(Z+\delta)}{\sqrt{C/(n-2)}} > t_{n-2, 1-\alpha} \right) \right\}.$$

O vício, a variância e o EQM de $\hat{\theta}_+$ com variância σ^2 desconhecida são dados por

$$E(\hat{\theta}_+ - \theta) = x_+ v^{1/2} [E(Q'_+) - \delta], \quad (11)$$

$$\text{var}(\hat{\theta}_+) = \frac{\hat{\sigma}^2}{n} + x_+^2 v \text{var}(Q'_+) \quad \text{e} \quad (12)$$

$$\text{MSE}(\hat{\theta}_+, \theta) = \frac{\hat{\sigma}^2}{n} + x_+^2 v [\text{var}(Q'_+) + (E(Q'_+) - \delta)^2]. \quad (13)$$

Para o cálculo de $\hat{\theta}$ quando a variância é desconhecida, o logaritmo da função de verossimilhança maximizada para o modelo k é

$$\log L_k = \text{const} - \frac{n}{2} \log(\text{SQR}_k/n).$$

Assim,

$$\begin{aligned} \text{AIC}_2 - \text{AIC}_1 &= n \log(\text{SQR}_2/n) - n \log(\text{SQR}_1/n) + 2(p_2 - p_1) \\ &= n \log(\text{SQR}_2/\text{SQR}_1) + 2(1 - 2) = n \log(\text{SQR}_2/\text{SQR}_1) - 2, \end{aligned}$$

onde se tem

$$\frac{\text{SQR}_2}{\text{SQR}_1} = 1 + \frac{\text{SQR}_2 - \text{SQR}_1}{\text{SQR}_1}.$$

Como anteriormente, $\text{SQR}_2 - \text{SQR}_1 = \hat{\beta}_1^2 \sum x_j^2$ e, agora tem-se que $\text{SQR}_1 \sim \sigma^2 \chi_{n-2}^2$, independente de $\hat{\beta}_1$, e então

$$\frac{\text{SQR}_2}{\text{SQR}_1} = 1 + \frac{\hat{\beta}_1^2 \sum x_j^2}{\sigma^2 C},$$

onde $C \sim \chi_{n-2}^2$. Seja $a = 1/\sum x_j^2$ e, então, $\hat{\beta}_1 \sim N(\beta_1, \sigma^2 a)$. Logo, $\hat{\beta}_1/\sqrt{\sigma^2 a} - \delta \sim N(0, 1)$ e pode-se escrever $\hat{\beta}_1/\sqrt{\sigma^2 a} = (Z + \delta)$. Assim, tem-se

$$\frac{\text{SQR}_2}{\text{SQR}_1} = 1 + \frac{\hat{\beta}_1^2}{\sigma^2 a C} = 1 + \frac{(Z + \delta)^2}{C},$$

com Z independente de C , e

$$\frac{1}{2}(\text{AIC}_2 - \text{AIC}_1) = \frac{n}{2} \log \left(1 + \frac{(Z + \delta)^2}{C} \right) - 1.$$

Definindo-se $U' = (Z + \delta)^2/C$ e

$$G'(u) = \frac{e^{-1}(1+u)^{n/2}}{1+e^{-1}(1+u)^{n/2}},$$

pode-se escrever

$$\hat{\theta} = \hat{\beta}_0 + x_+ v^{1/2} Q',$$

sendo $Q' = (Z + \delta)G'(U')$. O vício, a variância e o EQM de $\hat{\theta}$ ficam

$$\text{E}(\hat{\theta} - \theta) = x_+ v^{1/2} [\text{E}(Q') - \delta] \quad (14)$$

$$\text{var}(\hat{\theta}) = \frac{\hat{\sigma}^2}{n} + x_+^2 v \text{var}(Q') \quad \text{e} \quad (15)$$

$$\text{MSE}(\hat{\theta}, \theta) = \frac{\hat{\sigma}^2}{n} + x_+^2 v [\text{var}(Q') + (\text{E}(Q') - \delta)^2]. \quad (16)$$

3.2 Ponderação de modelos em regressão linear múltipla

A extensão do modelo de regressão linear simples descrito na Seção 3.1 é

$$y_j = \beta_0 + \beta_1 x_{1j} + \cdots + \beta_p x_{pj} + \varepsilon_j,$$

$j = 1, \dots, n$, sendo $\beta_0, \beta_1, \dots, \beta_p$ os parâmetros desconhecidos, $x_{1j}, \dots, x_{pj}$ os valores das covariáveis com a restrição de que as colunas correspondentes a elas são ortogonais entre si, e $\varepsilon_j \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$. A variância σ^2 pode ser conhecida ou não.

3.2.1 Modelos com duas covariáveis

Seja o modelo de regressão

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \varepsilon_j,$$

onde o objetivo é prever a média $\theta = \beta_0 + \beta_1 x_{1+} + \beta_2 x_{2+}$ de uma variável futura Y_+ para valores x_{1+} e x_{2+} .

Considerando, inicialmente, σ^2 conhecido o estimador geral é

$$\begin{aligned} \hat{\theta}_+ &= \hat{\beta}_0 + x_{1+} \hat{\beta}_1 \{I(\hat{\beta}_1/v_1^{1/2} < -z_{1-\alpha}) + I(\hat{\beta}_1/v_1^{1/2} > z_{1-\alpha})\} \\ &\quad + x_{2+} \hat{\beta}_2 \{I(\hat{\beta}_2/v_2^{1/2} < -z_{1-\alpha}) + I(\hat{\beta}_2/v_2^{1/2} > z_{1-\alpha})\}, \end{aligned}$$

sendo $v_1 = \sigma^2 / \sum x_{1j}^2$ e $v_2 = \sigma^2 / \sum x_{2j}^2$ as variâncias de $\hat{\beta}_1$ e $\hat{\beta}_2$, respectivamente. Substituindo $\hat{\beta}_1 = v_1^{1/2}(Z_1 + \delta_1)$ e $\hat{\beta}_2 = v_2^{1/2}(Z_2 + \delta_2)$, com $Z_1 \sim N(0, 1)$, $Z_2 \sim N(0, 1)$ e os coeficientes angulares padronizados $\delta_1 = \beta_1/v_1^{1/2}$ e $\delta_2 = \beta_2/v_2^{1/2}$, tem-se

$$\begin{aligned} \hat{\theta}_+ &= \hat{\beta}_0 + x_{1+} v_1^{1/2} (Z_1 + \delta_1) \{I(Z_1 < z_\alpha - \delta_1) + I(Z_1 > z_{1-\alpha} - \delta_1)\} + \\ &\quad + x_{2+} v_2^{1/2} (Z_2 + \delta_2) \{I(Z_2 < z_\alpha - \delta_2) + I(Z_2 > z_{1-\alpha} - \delta_2)\}, \end{aligned}$$

ou

$$\hat{\theta}_+ = \hat{\beta}_0 + x_{1+} v_1^{1/2} Q_{+1} + x_{2+} v_2^{1/2} Q_{+2},$$

sendo

$$\begin{aligned} Q_{+1} &= (Z_1 + \delta_1)\{I(Z_1 < z_\alpha - \delta_1) + I(Z_1 > z_{1-\alpha} - \delta_1)\} \quad \text{e} \\ Q_{+2} &= (Z_2 + \delta_2)\{I(Z_2 < z_\alpha - \delta_2) + I(Z_2 > z_{1-\alpha} - \delta_2)\}. \end{aligned}$$

O vício, a variância e o EQM de $\hat{\theta}_+$ são dados por

$$E(\hat{\theta}_+ - \theta) = x_{1+}v_1^{1/2}[E(Q_{+1}) - \delta_1] + x_{2+}v_2^{1/2}[E(Q_{+2}) - \delta_2], \quad (17)$$

$$\text{var}(\hat{\theta}_+) = \frac{\sigma^2}{n} + x_{1+}^2v_1\text{var}(Q_{+1}) + x_{2+}^2v_2\text{var}(Q_{+2}) \quad \text{e} \quad (18)$$

$$\begin{aligned} \text{EQM}(\hat{\theta}_+, \theta) &= \frac{\sigma^2}{n} + x_{1+}^2v_1[\text{var}(Q_{+1}) + (E(Q_{+1}) - \delta_1)^2] \\ &\quad + x_{2+}^2v_2[\text{var}(Q_{+2}) + (E(Q_{+2}) - \delta_2)^2] \\ &\quad + 2x_{1+}v_1^{1/2}x_{2+}v_2^{1/2}(E(Q_{+1}) - \delta_1)(E(Q_{+2}) - \delta_2). \end{aligned} \quad (19)$$

Como Z_1 e Z_2 são independentes, a covariância entre Q_1 e Q_2 é nula.

Para construir o estimador ponderado (1) nesta situação, são considerados todos os 2^p modelos possíveis, que são 4 modelos neste exemplo, como se segue:

$$\begin{aligned} M_1 : \hat{y}_{1+} &= \hat{\beta}_0; & M_3 : \hat{y}_{3+} &= \hat{\beta}_0 + \hat{\beta}_2x_{2+}; \\ M_2 : \hat{y}_{2+} &= \hat{\beta}_0 + \hat{\beta}_1x_{1+}; & M_4 : \hat{y}_{4+} &= \hat{\beta}_0 + \hat{\beta}_1x_{1+} + \hat{\beta}_2x_{2+}; \end{aligned}$$

onde vale observar que, como as colunas correspondentes às covariáveis são ortogonais entre si, então, $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\beta}_2$ são os mesmos em todos os modelos onde aparecem. O estimador $\hat{\theta}$ é

$$\begin{aligned} \hat{\theta} &= W_1\hat{\beta}_0 + W_2(\hat{\beta}_0 + \hat{\beta}_1x_{1+}) + W_3(\hat{\beta}_0 + \hat{\beta}_2x_{2+}) + W_4(\hat{\beta}_0 + \hat{\beta}_1x_{1+} + \hat{\beta}_2x_{2+}) \\ &= \hat{\beta}_0 + (W_2 + W_4)\hat{\beta}_1x_{1+} + (W_3 + W_4)\hat{\beta}_2x_{2+}, \end{aligned}$$

sendo os pesos dados por

$$W_k = \frac{\exp(-AIC_k/2)}{\sum_{l=1}^4 \exp(-AIC_l/2)}, \quad k = 2, 3, 4,$$

e então

$$W_2 + W_4 = \frac{\exp(-AIC_2/2) + \exp(-AIC_4/2)}{\sum_{k=1}^4 \exp(-AIC_k/2)} = \frac{1 + \exp((AIC_4 - AIC_2)/2)}{\sum_{k=1}^4 \exp((AIC_4 - AIC_k)/2)},$$

e

$$W_3 + W_4 = \frac{1 + \exp((AIC_4 - AIC_3)/2)}{1 + \sum_{k=1}^3 \exp((AIC_4 - AIC_k)/2)}.$$

Para σ^2 conhecido,

$$AIC_4 - AIC_k = \frac{1}{\sigma^2}(\text{SQR}_4 - \text{SQR}_k) + 2(p_4 - p_k),$$

para $k = 1, 2, 3$. Usando o fato de que as colunas correspondentes às covariáveis são ortogonais entre si, obtêm-se as expressões

$$\begin{aligned} (\text{SQR}_4 - \text{SQR}_1) &= -\hat{\beta}_1^2 \sum x_{1j}^2 - \hat{\beta}_2^2 \sum x_{2j}^2 = -\frac{\hat{\beta}_1^2 \sigma^2}{v_1} - \frac{\hat{\beta}_2^2 \sigma^2}{v_2}, \\ (\text{SQR}_4 - \text{SQR}_2) &= -\hat{\beta}_2^2 \sum x_{2j}^2 = -\frac{\hat{\beta}_2^2 \sigma^2}{v_2} \quad \text{e} \\ (\text{SQR}_4 - \text{SQR}_3) &= -\hat{\beta}_1^2 \sum x_{1j}^2 = -\frac{\hat{\beta}_1^2 \sigma^2}{v_1}, \end{aligned}$$

e, substituindo $\hat{\beta}_1$ e $\hat{\beta}_2$ por $v_1^{1/2}(Z_1 + \delta_1)$ e $v_2^{1/2}(Z_2 + \delta_2)$, tem-se que

$$\begin{aligned} (AIC_4 - AIC_1)/2 &= -(Z_1 + \delta_1)^2/2 - (Z_2 + \delta_2)^2/2 + 2 = U_1 + U_2, \\ (AIC_4 - AIC_2)/2 &= -(Z_2 + \delta_2)^2/2 + 1 = U_2 \quad \text{e} \\ (AIC_4 - AIC_3)/2 &= -(Z_1 + \delta_1)^2/2 + 1 = U_1, \end{aligned}$$

sendo $U_1 = -(Z_1 + \delta_1)^2/2 + 1$ e $U_2 = -(Z_2 + \delta_2)^2/2 + 1$. Desta forma, tem-se

$$\begin{aligned} W_2 + W_4 &= (1 + \exp(U_2))/(1 + \exp(U_1) + \exp(U_2) + \exp(U_1 + U_2)) \quad \text{e} \\ W_3 + W_4 &= (1 + \exp(U_1))/(1 + \exp(U_1) + \exp(U_2) + \exp(U_1 + U_2)). \end{aligned}$$

Seja

$$G_2(u) = (1 + \exp(u))/(1 + \exp(u_1) + \exp(u_2) + \exp(u_1 + u_2)),$$

então, $\hat{\theta}$ pode ser escrito como

$$\hat{\theta} = \hat{\beta}_0 + x_{1+}v_1^{1/2}Q_1 + x_{2+}v_2^{1/2}Q_2,$$

sendo

$$Q_1 = (Z_1 + \delta_1)G_2(U_2) \quad \text{e} \quad Q_2 = (Z_2 + \delta_2)G_2(U_1).$$

O vício, a variância e o EQM de $\hat{\theta}$ são dados por

$$E(\hat{\theta} - \theta) = x_{1+}v_1^{1/2}[E(Q_1) - \delta_1] + x_{2+}v_2^{1/2}[E(Q_2) - \delta_2], \quad (20)$$

$$\text{var}(\hat{\theta}) = \frac{\sigma^2}{n} + x_{1+}^2v_1\text{var}(Q_1) + x_{2+}^2v_2\text{var}(Q_2) \quad \text{e} \quad (21)$$

$$\begin{aligned} \text{EQM}(\hat{\theta}, \theta) &= \frac{\sigma^2}{n} + x_{1+}^2v_1[\text{var}(Q_1) + (E(Q_1) - \delta_1)^2] \\ &\quad + x_{2+}^2v_2[\text{var}(Q_2) + (E(Q_2) - \delta_2)^2] + \\ &\quad 2x_{1+}v_1^{1/2}x_{2+}v_2^{1/2}(E(Q_1) - \delta_1)(E(Q_2) - \delta_2). \end{aligned} \quad (22)$$

Novamente, não há covariância, pois $\hat{\beta}_1$ e $\hat{\beta}_2$ são independentes.

É interessante analisar o comportamento destes estimadores quando se considera o modelo verdadeiro com duas covariáveis mas, na prática, ignora-se a existência de uma delas. Ignorando-se a existência de x_2 os estimadores $\hat{\theta}_+$ e $\hat{\theta}$ tornam-se

$$\begin{aligned} \hat{\theta}_+ &= \hat{\beta}_0 + x_+v^{1/2}Q_+ \quad \text{e} \\ \hat{\theta} &= \hat{\beta}_0 + x_+v^{1/2}Q, \end{aligned}$$

e o vício, a variância e o EQM são, para $\hat{\theta}_+$, dados por

$$E(\hat{\theta}_+ - \theta) = x_{1+}v_1^{1/2}[E(Q_+) - \delta_1] - x_{2+}v_2^{1/2}\delta_2, \quad (23)$$

$$\text{var}(\hat{\theta}_+) = \frac{\sigma^2}{n} + x_{1+}^2 v_1 \text{var}(Q_+) \quad \text{e} \quad (24)$$

$$\begin{aligned} \text{EQM}(\hat{\theta}_+, \theta) &= \frac{\sigma^2}{n} + x_{1+}^2 v_1 [\text{var}(Q_+) + (E(Q_+) - \delta_1)^2] \\ &\quad + x_{2+}^2 v_2 \delta_2^2 + 2x_{1+} v_1^{1/2} x_{2+} v_2^{1/2} (E(Q_+) - \delta_1) \delta_2, \end{aligned} \quad (25)$$

e para $\hat{\theta}$, são

$$E(\hat{\theta} - \theta) = x_{1+} v_1^{1/2} [E(Q) - \delta_1] - x_{2+} v_2^{1/2} \delta_2, \quad (26)$$

$$\text{var}(\hat{\theta}) = \frac{\sigma^2}{n} + x_{1+}^2 v_1 \text{var}(Q) \quad \text{e} \quad (27)$$

$$\begin{aligned} \text{EQM}(\hat{\theta}, \theta) &= \frac{\sigma^2}{n} + x_{1+}^2 v_1 [\text{var}(Q) + (E(Q) - \delta_1)^2] \\ &\quad + x_{2+}^2 v_2 \delta_2^2 + 2x_{1+} v_1^{1/2} x_{2+} v_2^{1/2} (E(Q) - \delta_1) \delta_2. \end{aligned} \quad (28)$$

Se for usando o BIC na construção dos pesos, as funções envolvidas nos vícios e variâncias continuam as mesmas como em (20), (21) e (22), mudando-se apenas os valores de U_1 e U_2 para

$$U_1 = -(Z_1 + \delta_1)^2/2 - \log(n)/2 \quad \text{e}$$

$$U_2 = -(Z_2 + \delta_2)^2/2 - \log(n)/2.$$

Quando a variância σ^2 é desconhecida,

$$\hat{\theta}_+ = \hat{\beta}_0 + x_{1+} \hat{\beta}_1 \{I(\hat{\beta}_1/\hat{v}_1^{1/2} < t_{n-3,\alpha}) + I(\hat{\beta}_1/\hat{v}_1^{1/2} > t_{n-3,1-\alpha})\}$$

$$+ x_{2+} \hat{\beta}_2 \{I(\hat{\beta}_2/\hat{v}_2^{1/2} < t_{n-3,\alpha}) + I(\hat{\beta}_2/\hat{v}_2^{1/2} > t_{n-3,1-\alpha})\},$$

onde $\hat{v}_1 = \hat{\sigma}^2 / \sum x_{1j}^2$ e $\hat{v}_2 = \hat{\sigma}^2 / \sum x_{2j}^2$. Como no caso de regressão linear simples, seja

$C \sim \chi_{n-3}^2$, então

$$\begin{aligned}\hat{\beta}_1/\hat{v}_1^{1/2} &= (Z_1 + \delta_1)/\sqrt{C/(n-3)} \quad \text{e} \\ \hat{\beta}_2/\hat{v}_2^{1/2} &= (Z_2 + \delta_2)/\sqrt{C/(n-3)},\end{aligned}$$

e assim,

$$\hat{\theta}_+ = \hat{\beta}_0 + x_{1+}v_1^{1/2}Q'_{+1} + x_{2+}v_2^{1/2}Q'_{+2},$$

onde

$$\begin{aligned}Q'_{+1} &= (Z_1 + \delta_1)\{I[(Z_1 + \delta_1)/\sqrt{C/(n-3)} < t_{n-3,\alpha}] \\ &\quad + I[(Z_1 + \delta_1)/\sqrt{C/(n-3)} > t_{n-3,1-\alpha}]\} \quad \text{e} \\ Q'_{+2} &= (Z_2 + \delta_2)\{I[(Z_2 + \delta_2)/\sqrt{C/(n-3)} < t_{n-3,\alpha}] \\ &\quad + I[(Z_2 + \delta_2)/\sqrt{C/(n-3)} > t_{n-3,1-\alpha}]\}.\end{aligned}$$

O vício, a variância e o EQM de $\hat{\theta}_+$ para variância σ^2 desconhecida são

$$E(\hat{\theta}_+ - \theta) = x_{1+}v_1^{1/2}[E(Q'_{+1}) - \delta_1] + x_{2+}v_2^{1/2}[E(Q'_{+2}) - \delta_2], \quad (29)$$

$$\text{var}(\hat{\theta}_+) = \frac{\hat{\sigma}^2}{n} + x_{1+}^2v_1\text{var}(Q'_{+1}) + x_{2+}^2v_2\text{var}(Q'_{+2}) \quad \text{e} \quad (30)$$

$$\begin{aligned}\text{EQM}(\hat{\theta}_+, \theta) &= \frac{\hat{\sigma}^2}{n} + x_{1+}^2v_1[\text{var}(Q'_{+1}) + (E(Q'_{+1}) - \delta_1)^2] \\ &\quad + x_{2+}^2v_2[\text{var}(Q'_{+2}) + (E(Q'_{+2}) - \delta_2)^2] \\ &\quad + 2x_{1+}v_1^{1/2}x_{2+}v_2^{1/2}(E(Q'_{+1}) - \delta_1)(E(Q'_{+2}) - \delta_2).\end{aligned} \quad (31)$$

No cálculo do estimador ponderado $\hat{\theta}$ tem-se que

$$\begin{aligned}(\text{AIC}_4 - \text{AIC}_1)/2 &= -\frac{n}{2} \log \left(1 + \frac{\text{SQR}_1 - \text{SQR}_4}{\text{SQR}_4} \right) + 2 \\ &= -\frac{n}{2} \log \left(1 + \frac{(Z_1 + \delta_1)^2}{C} + \frac{(Z_2 + \delta_2)^2}{C} \right) + 2, \\ (\text{AIC}_4 - \text{AIC}_2)/2 &= -\frac{n}{2} \log \left(1 + \frac{\text{SQR}_2 - \text{SQR}_4}{\text{SQR}_4} \right) + 1 = -\frac{n}{2} \log \left(1 + \frac{(Z_2 + \delta_2)^2}{C} \right) + 1 \quad \text{e} \\ (\text{AIC}_4 - \text{AIC}_3)/2 &= -\frac{n}{2} \log \left(1 + \frac{\text{SQR}_3 - \text{SQR}_4}{\text{SQR}_4} \right) + 1 = -\frac{n}{2} \log \left(1 + \frac{(Z_1 + \delta_1)^2}{C} \right) + 1,\end{aligned}$$

com $Z_1 \sim N(0, 1)$ e $Z_2 \sim N(0, 1)$, independentes de $C \sim \chi_{n-3}^2$. Fazendo-se

$$U'_1 = (Z_1 + \delta_1)^2/C \quad \text{e} \quad U'_2 = (Z_2 + \delta_2)^2/C$$

tem-se que

$$\begin{aligned} W_2 + W_4 &= \frac{1 + e^1(1+U'_2)^{-n/2}}{1 + e^1(1+U'_1)^{-n/2} + e^1(1+U'_2)^{-n/2} + e^2(1+U'_1+U'_2)^{-n/2}} \quad \text{e} \\ W_3 + W_4 &= \frac{1 + e^1(1+U'_1)^{-n/2}}{1 + e^1(1+U'_1)^{-n/2} + e^1(1+U'_2)^{-n/2} + e^2(1+U'_1+U'_2)^{-n/2}}, \end{aligned}$$

e, assim

$$\hat{\theta} = \hat{\beta}_0 + x_{1+}v_1^{1/2}Q'_1 + x_{2+}v_2^{1/2}Q'_2,$$

sendo

$$Q_1 = (Z_1 + \delta_1)G'_2(U'_2) \quad \text{e} \quad Q_2 = (Z_2 + \delta_2)G'_2(U'_1),$$

e

$$G'_2(u') = \frac{1 + e^1(1+u')^{-n/2}}{1 + e^1(1+u'_1)^{-n/2} + e^1(1+u'_2)^{-n/2} + e^2(1+u'_1+u'_2)^{-n/2}}.$$

O vício, a variância e o EQM de $\hat{\theta}$ são dados por

$$E(\hat{\theta} - \theta) = x_{1+}v_1^{1/2}[E(Q'_1) - \delta_1] + x_{2+}v_2^{1/2}[E(Q'_2) - \delta_2], \quad (32)$$

$$\text{var}(\hat{\theta}) = \frac{\hat{\sigma}^2}{n} + x_{1+}^2v_1\text{var}(Q'_1) + x_{2+}^2v_2\text{var}(Q'_2) \quad \text{e} \quad (33)$$

$$\begin{aligned} \text{EQM}(\hat{\theta}, \theta) &= \frac{\hat{\sigma}^2}{n} + x_{1+}^2v_1[\text{var}(Q'_1) + (E(Q'_1) - \delta_1)^2] \\ &\quad + x_{2+}^2v_2[\text{var}(Q'_2) + (E(Q'_2) - \delta_2)^2] \\ &\quad + 2x_{1+}v_1^{1/2}x_{2+}v_2^{1/2}(E(Q_1) - \delta_1)(E(Q_2) - \delta_2). \end{aligned} \quad (34)$$

3.2.2 Modelos com três covariáveis

Considerando um modelo com três covariáveis e a variância conhecida, o estimador geral é

$$\hat{\theta}_+ = \hat{\beta}_0 + x_{1+}\hat{\beta}_1\{I(\hat{\beta}_1/v_1^{1/2} < -z_{1-\alpha}) + I(\hat{\beta}_1/v_1^{1/2} > z_{1-\alpha})\}$$

$$+x_{2+}\hat{\beta}_2\{I(\hat{\beta}_2/v_2^{1/2} < -z_{1-\alpha}) + I(\hat{\beta}_2/v_2^{1/2} > z_{1-\alpha})\}$$

$$+x_{3+}\hat{\beta}_3\{I(\hat{\beta}_3/v_3^{1/2} < -z_{1-\alpha}) + I(\hat{\beta}_3/v_3^{1/2} > z_{1-\alpha})\},$$

e obtêm-se

$$E(\hat{\theta}_+ - \theta) = \sum_{i=1}^3 x_{i+} v_i^{1/2} [E(Q_{+i}) - \delta_i], \quad (35)$$

$$\text{var}(\hat{\theta}_+) = \frac{\sigma^2}{n} + \sum_{i=1}^3 x_{i+}^2 v_i \text{var}(Q_{+i}) \quad \text{e} \quad (36)$$

$$\text{EQM}(\hat{\theta}_+, \theta) = \text{var}(\hat{\theta}_+) + [E(\hat{\theta}_+ - \theta)]^2, \quad (37)$$

sendo

$$Q_{+i} = (Z_i + \delta_i) \{I(Z_i < z_\alpha - \delta_i) + I(Z_i > z_{1-\alpha} - \delta_i)\},$$

$i = 1, \dots, 3$.

No caso de variância desconhecida, vício, variância e EQM para $\hat{\theta}_+$ são calculados da mesma maneira, substituindo-se Q_{+i} em (35), (36) e (37) por

$$\begin{aligned} Q'_{+i} = & (Z_i + \delta_i) \{I([Z_i + \delta_i]/\sqrt{C/(n-4)} < t_{n-4,\alpha}) \\ & + I([Z_i + \delta_i]/\sqrt{C/(n-4)} > t_{n-4,1-\alpha})\}, \end{aligned}$$

$i = 1, \dots, 3$.

Para calcular $\hat{\theta}$, são possíveis $2^3 = 8$ modelos, que estão descritos a seguir

$$\begin{aligned} M_1 : \hat{y}_{1+} &= \hat{\beta}_0; & M_5 : \hat{y}_{5+} &= \hat{\beta}_0 + \hat{\beta}_1 x_{1+} + \hat{\beta}_2 x_{2+}; \\ M_2 : \hat{y}_{2+} &= \hat{\beta}_0 + \hat{\beta}_1 x_{1+}; & M_6 : \hat{y}_{6+} &= \hat{\beta}_0 + \hat{\beta}_1 x_{1+} + \hat{\beta}_3 x_{3+}; \\ M_3 : \hat{y}_{3+} &= \hat{\beta}_0 + \hat{\beta}_2 x_{2+}; & M_7 : \hat{y}_{7+} &= \hat{\beta}_0 + \hat{\beta}_2 x_{2+} + \hat{\beta}_3 x_{3+}; \\ M_4 : \hat{y}_{4+} &= \hat{\beta}_0 + \hat{\beta}_3 x_{3+}; & M_8 : \hat{y}_{8+} &= \hat{\beta}_0 + \hat{\beta}_1 x_{1+} + \hat{\beta}_2 x_{2+} + \hat{\beta}_3 x_{3+}; \end{aligned}$$

e, então,

$$\begin{aligned} \hat{\theta} = & \hat{\beta}_0 + (W_2 + W_5 + W_6 + W_8) \hat{\beta}_1 x_{1+} + (W_3 + W_5 + W_7 + W_8) \hat{\beta}_2 x_{2+} \\ & + (W_4 + W_6 + W_7 + W_8) \hat{\beta}_3 x_{3+}, \end{aligned}$$

com os pesos dados por

$$W_k = \frac{\exp([AIC_8 - AIC_k]/2)}{1 + \sum_{l=1}^7 \exp([AIC_8 - AIC_l]/2)},$$

para $k = 1, \dots, 8$.

Quando a variância é conhecida, seja

$$U_1 = -(Z_1 + \delta_1)^2/2 + 1, \quad U_2 = -(Z_2 + \delta_2)^2/2 + 1 \quad \text{e}$$

$$U_3 = -(Z_3 + \delta_3)^2/2 + 1,$$

e, então,

$$(AIC_8 - AIC_1)/2 = U_1 + U_2 + U_3,$$

$$(AIC_8 - AIC_2)/2 = U_2 + U_3,$$

$$(AIC_8 - AIC_3)/2 = U_1 + U_3,$$

$$(AIC_8 - AIC_4)/2 = U_1 + U_2,$$

$$(AIC_8 - AIC_5)/2 = U_3,$$

$$(AIC_8 - AIC_6)/2 = U_2 \quad \text{e}$$

$$(AIC_8 - AIC_7)/2 = U_1.$$

Fazendo-se

$$G_3(u) = (1 + \exp(u)) / \prod_{i=1}^3 (1 + e_i^u),$$

obtém-se

$$\hat{\theta} = \hat{\beta}_0 + x_{1+} v_1^{1/2} Q_1 + x_{2+} v_2^{1/2} Q_2 + x_{3+} v_3^{1/2} Q_3,$$

sendo

$$Q_1 = (Z_1 + \delta_1)[G_3(0) + G_3(u_2) + G_3(u_3) + G_3(u_2 + u_3)],$$

$$Q_2 = (Z_2 + \delta_2)[G_3(0) + G_3(u_1) + G_3(u_3) + G_3(u_1 + u_3)],$$

$$Q_3 = (Z_3 + \delta_3)[G_3(0) + G_3(u_1) + G_3(u_2) + G_3(u_1 + u_2)],$$

e

$$E(\hat{\theta} - \theta) = \sum_{i=1}^3 x_{i+} v_i^{1/2} [E(Q_i) - \delta_i], \quad (38)$$

$$\text{var}(\hat{\theta}) = \frac{\sigma^2}{n} + \sum_{i=1}^3 x_{i+}^2 v_i \text{var}(Q_i) \quad \text{e} \quad (39)$$

$$\text{EQM}(\hat{\theta}, \theta) = \text{var}(\hat{\theta}) + [E(\hat{\theta} - \theta)]^2. \quad (40)$$

Quando a variância σ^2 é desconhecida as diferenças entre os AIC's são dadas por

$$\begin{aligned} (\text{AIC}_8 - \text{AIC}_1)/2 &= -\frac{1}{2} \log(1 + U'_1 + U'_2 + U'_3) + 3, \\ (\text{AIC}_8 - \text{AIC}_2)/2 &= -\frac{1}{2} \log(1 + U'_2 + U'_3) + 2, \\ (\text{AIC}_8 - \text{AIC}_3)/2 &= -\frac{1}{2} \log(1 + U'_1 + U'_3) + 2, \\ (\text{AIC}_8 - \text{AIC}_4)/2 &= -\frac{1}{2} \log(1 + U'_1 + U'_2) + 2, \\ (\text{AIC}_8 - \text{AIC}_5)/2 &= -\frac{1}{2} \log(1 + U'_3) + 1, \\ (\text{AIC}_8 - \text{AIC}_6)/2 &= -\frac{1}{2} \log(1 + U'_2) + 1 \quad \text{e} \\ (\text{AIC}_8 - \text{AIC}_7)/2 &= -\frac{1}{2} \log(1 + U'_1) + 1, \end{aligned}$$

onde

$$U'_1 = (Z_1 + \delta_1)^2/C, \quad U'_2 = (Z_2 + \delta_2)^2/C \quad \text{e} \quad U'_3 = (Z_3 + \delta_3)^2/C,$$

com $Z_1 \sim N(0, 1)$, $Z_2 \sim N(0, 1)$ e $Z_3 \sim N(0, 1)$ independentes de $C \sim \chi_{n-4}^2$.

Fazendo-se

$$\begin{aligned} G'_3(u', a) &= e^a (1 + u')^{n/2} / \left[1 + \sum_{i=1}^3 e^1 (1 + u'_i)^{n/2} \right. \\ &\quad + e^2 (1 + u'_1 + u'_2)^{n/2} + e^2 (1 + u'_1 + u'_3)^{n/2} \\ &\quad \left. + e^2 (1 + u'_2 + u'_3)^{n/2} + e^3 (1 + u'_1 + u'_2 + u'_3)^{n/2} \right], \end{aligned}$$

o vício, a variância e o EQM para $\hat{\theta}$, quando a variância σ^2 é desconhecida, são obtidos substituindo-se σ^2 , Q_1 , Q_2 e Q_3 em (38), (39) e (40) por $\hat{\sigma}^2$ e

$$Q'_1 = (Z_1 + \delta_1)[G'_3(0, 0) + G'_3(U'_2, 1) + G'_3(U'_3, 1) + G'_3(U'_2 + U'_3, 2)],$$

$$\begin{aligned}
Q'_2 &= (Z_2 + \delta_2)[G'_3(0, 0) + G'_3(U'_1, 1) + G_3(U'_3, 1) + G'_3(U'_1 + U'_3, 2)] \quad \text{e} \\
Q'_3 &= (Z_3 + \delta_3)[G'_3(0, 0) + G'_3(U'_1, 1) + G_3(U'_2, 2) + G'_3(U'_1 + U'_2, 2)].
\end{aligned}$$

3.2.3 Caso Geral

Para um valor genérico de p , o estimador geral, com variância σ^2 conhecida, é

$$\begin{aligned}
\hat{\theta}_+ &= \hat{\beta}_0 + \sum_{i=1}^p x_{i+} \hat{\beta}_i \{I(\hat{\beta}_i/v_i^{1/2} < -z_{1-\alpha}) + I(\hat{\beta}_i/v_i^{1/2} > z_{1-\alpha})\} \\
&= \hat{\beta}_0 + \sum_{i=1}^p x_{i+} v_i^{1/2} Q_{i+},
\end{aligned}$$

sendo

$$Q_{i+} = (\delta_i + Z_i) \{I(Z_i < z_\alpha - \delta_i) + I(Z_i > z_{1-\alpha} - \delta_i)\}.$$

Então, o vício e a variância são dados por

$$E(\hat{\theta}_+ - \theta) = \sum_{i=1}^p x_{i+} v_i^{1/2} [E(Q_{i+}) - \delta_i] \quad \text{e} \quad (41)$$

$$\text{var}(\hat{\theta}_+) = \frac{\sigma^2}{n} + \sum_{i=1}^p x_{i+}^2 v_i \text{var}(Q_{i+}). \quad (42)$$

Quando a variância é desconhecida, σ^2 e Q_{i+} são substituídos em (41) e (42) por $\hat{\sigma}^2$ e

$$\begin{aligned}
Q'_{i+} &= (Z_i + \delta_i) \{I([Z_i + \delta_i]/\sqrt{C/(n-p+1)} < t_{n-p+1, \alpha}) \\
&\quad + I([Z_i + \delta_i]/\sqrt{C/(n-p+1)} > t_{n-p+1, 1-\alpha})\}.
\end{aligned}$$

Para obter $\hat{\theta}$, é definido $\mathcal{M} = \{M_1, \dots, M_K\}$ como o conjunto de todos $2^p = K$ modelos possíveis, o qual contém $\binom{1}{1}$ modelo sem covariáveis, $\binom{p}{1}$ modelos com uma covariável, $\binom{p}{2}$ modelos com duas covariáveis, ..., $\binom{p}{p}$ modelos com p covariáveis. O estimador ponderado para uma regressão linear múltipla com p covariáveis é

$$\hat{\theta} = \hat{\beta}_0 + \sum_{i=1}^p v_i^{1/2} x_{i+} Q_i,$$

sendo, para variância conhecida,

$$Q_i = (Z_i + \delta_i) \left[G_p(0) + \sum_{l=1, l1 \neq i}^p G_p(U_l) \right. \\ \left. + \sum_{\substack{l_1, l_2=1 \\ l_1 \neq l_2 \neq i}}^p G_p(U_{l_1} + U_{l_2}) + \cdots + \sum_{\substack{l_1, l_2=\dots=l_{p-1}=1 \\ l_1 \neq l_2 \neq \dots \neq l_{p-1} \neq i}}^p G_p(U_{l_1} + U_{l_2} + \cdots + U_{l_{p-1}}) \right],$$

com $i = 1, \dots, p$, $U_i = -(Z_i + \delta_i)^2/2 + 1$ e

$$G_p(u) = (1 + \exp(u)) / \prod_{i=1}^p (1 + e_i^u).$$

Com estas definições, obtêm-se

$$E(\hat{\theta} - \theta) = \sum_{i=1}^p x_{i+} v_i^{1/2} [E(Q_i) - \delta_i] \quad (43)$$

e

$$\text{var}(\hat{\theta}) = \frac{\sigma^2}{n} + \sum_{i=1}^p x_{i+}^2 v_i \text{var}(Q_i). \quad (44)$$

Quando a variância σ^2 é desconhecida, substitui-se σ^2 e Q_i em (43) e (44) por $\hat{\sigma}^2$ e

$$Q'_i = (Z_i + \delta_i) \left[G'_p(0, 0) + \sum_{l=1, l1 \neq i}^p G'_p(U'_l, 1) \right. \\ \left. + \sum_{\substack{l_1, l_2=1 \\ l_1 \neq l_2 \neq i}}^p G'_p(U'_{l_1} + U'_{l_2}, 2) + \cdots + \sum_{\substack{l_1, l_2=\dots=l_{p-1}=1 \\ l_1 \neq l_2 \neq \dots \neq l_{p-1} \neq i}}^p G'_p(U'_{l_1} + U'_{l_2} + \cdots + U'_{l_{p-1}}, p-1) \right],$$

para $U'_i = (Z_i + \delta_i)^2/C$, $i = 1, \dots, p$, onde $Z_i \sim N(0, 1)$ são independentes de $C \sim \chi_{n-p-1}^2$, e

$$G'_p(u', a) = e^a (1 + u')^{n/2} / \left[1 + \sum_{i=1}^p e^1 (1 + u'_i)^{-n/2} + \right. \\ \left. + \sum_{c \in \mathcal{C}_2} e^2 (1 + c)^{-n/2} + \sum_{c \in \mathcal{C}_3} e^3 (1 + c)^{-n/2} + \right. \\ \left. + \cdots + \sum_{c \in \mathcal{C}_p} e^p (1 + c)^{-n/2} \right].$$

sendo que $\mathcal{C}_m$ representa todas combinações das m covariáveis e c é a soma dos u_{im} 's que correspondem a estas combinações, para $i, m = 1, \dots, p$.

3.3 O método *bootstrap* para calcular a variância do estimador ponderado

Outra maneira de se estimar a variância de $\hat{\theta}$, sem a necessidade de suposições como em (3) e nem de desenvolvimentos analíticos, é através do uso do método *bootstrap*. Também, no contexto das situações descritas em 3.1 e 3.2, o uso do *bootstrap* não requer ortogonalidade entre os vetores correspondentes às covariáveis.

Como visto em 2.3, o trabalho de Buckland et al (1997) considera três alternativas para obtenção das amostras *bootstrap*, além da amostragem dos vetores de observações. Aqui, será utilizada uma alternativa diferente, além da amostragem dos vetores de observações. Esta alternativa consiste em gerar K amostras a partir dos resíduos de cada modelo e calcular w_k e $\hat{\theta}_k$ nas respectivas amostras para, então, calcular $\hat{\theta}$.

3.3.1 Planos de reamostragem

O Algoritmo 1 descreve os passos para a reamostragem dos vetores de observações, aplicado aos modelos de regressão apresentados em 3.1 e 3.2. Neste caso, a ortogonalidade entre as colunas referentes às covariáveis não será mantida, mas esta ortogonalidade não é necessária para o cálculo da variância *bootstrap* de $\hat{\theta}$.

Algoritmo 1

- (i) amostrar $i_1^*, \dots, i_n^*$ aleatoriamente, com reposição, de $\{1, 2, \dots, n\}$;
- (ii) fazer $(y_j^*, \mathbf{x}_j^*) = (y_{i_j^*}, \mathbf{x}_{i_j^*})$, para $j = 1, \dots, n$, obtendo a amostra *bootstrap*
 $(y_1^*, \mathbf{x}_1^*), \dots, (y_n^*, \mathbf{x}_n^*)$;
- (iii) repetir os passos 1 e 2, B vezes.

A segunda abordagem é a reamostragem baseada no modelo de regressão linear. Como não se sabe qual o modelo correto a ser utilizado para a

obtenção dos resíduos, uma opção é gerar amostras a partir dos resíduos de cada um dos modelos $M_1, \dots, M_K$, em cada um deles calcular os respectivos pesos W_k , como definidos em 2.3, e estimativas $\hat{\theta}_k$ e então calcular $\hat{\theta}^*$. Desta forma, os K modelos estarão sendo ajustados em amostras diferentes o que caracteriza um estimador diferente daquele definido em (1). Mas, parece interessante verificar o comportamento desta alternativa. O Algoritmo 2 descreve os passos para esta alternativa.

Algoritmo 2

- (i) seja $e_j = r_j - \bar{r}$, para $j = 1, \dots, n$, onde r_j é o resíduo de Pearson padronizado e $\bar{r}$ é a média de r_j para o modelo k ;
- (ii) amostrar $e_1^*, \dots, e_n^*$ aleatoriamente, com reposição, dos $e_1, \dots, e_n$ de (i);
- (iii) fazer $y_j^* = \hat{\mu}_{kj} + e_j^*$, para $j = 1, \dots, n$, obtendo a amostra *bootstrap* $y_1^*, \dots, y_n^*$, onde $\hat{\mu}_{kj}$ é o modelo M_k ajustado à amostra original;
- (iv) ajustar o modelo M_k à amostra *bootstrap* e obter AIC_k^* e $\hat{\theta}_k^*$;
- (v) repetir os passos 1 a 4 para $k = 1, \dots, K$, calcular $w_1^*, \dots, w_K^*$ e $\hat{\theta}^*$;
- (vi) repetir os passos 1 a 5 B vezes.

3.4 Comparação entre os métodos

A comparação entre o estimador ponderado, $\hat{\theta}$, e o estimador obtido ao ser escolhido um modelo, $\hat{\theta}_+$, será feita comparando-se diretamente as expressões de vício, variância e erro quadrático médio, obtidas para estes estimadores em 3.1 e 3.2, através de gráficos. Em todas as expressões de (5) a (44) é possível separar uma parte constante e igual para os dois estimadores, correspondente a σ^2 , n e os valores das covariáveis, e obter a variação para vício, variância e EQM, ao se utilizarem diferentes valores do coeficiente angular padronizado δ .

Para comparar o estimador ponderado com o estimador ponderado Bayesiano serão utilizados exemplos da literatura. Nestes exemplos serão, também,

comparados os resultados da variância obtida em (3) e as variâncias obtidas pelas duas versões *bootstrap*. Dentro dos exemplos serão, também, realizados alguns estudos de simulação para que as conclusões possam ser aprofundadas. Na aplicação da ponderação Bayesiana de modelos, serão considerados o conjunto completo de modelos possíveis e, conseqüentemente, não serão aplicados os métodos *Occam's window* e MC^3 .

4 RESULTADOS E DISCUSSÃO

4.1 Comparação entre o estimador ponderado e a metodologia usual baseada na escolha de um modelo

A comparação entre os estimadores $\hat{\theta}$ e $\hat{\theta}_+$ é feita comparando-se os vícios, as variâncias e os EQM's correspondentes, para as situações de regressão linear com uma, duas e três covariáveis. Em todas as equações correspondentes a estas propriedades, os termos referentes aos valores da variância σ^2 (ou sua estimativa), aos valores da variância dos parâmetros β e aos valores de x_+ são comuns aos dois métodos. Sendo assim, nesta comparação, estes valores serão ignorados ou mantidos iguais a um, e, será analisada apenas a parte que define a diferença entre as quantidades comparadas, correspondente às funções envolvendo os valores denominados, de uma forma geral, de Q .

Para o cálculo da média e variância das funções Q foram utilizados 10.000 valores para as variáveis Z e C . Os cálculos e gráficos foram feitos no S-PLUS e os programas estão no Apêndice 1.

4.1.1 Regressão linear simples

A Figura 1 mostra a comparação entre $\hat{\theta}_+$, lado esquerdo, para $\alpha = 0,025$, e $\hat{\theta}$, lado direito, para regressão linear simples com variância conhecida, utilizando-se as expressões de (5) a (10). Como já esperado, os dois estimadores são viciados para valores de δ próximos de zero, devido à possibilidade de o esti-

mador mais simples $\hat{\theta}_2$ ser escolhido mesmo se $\beta_1 \neq 0$ e, à medida que $|\delta| \rightarrow \infty$, o vício diminui até o valor que corresponde a condicionar a estimação ao modelo M_1 . Mas, o vício é menor para $\hat{\theta}$, assim como a variância, e, conseqüentemente seu erro quadrático médio foi consideravelmente reduzido. Uma dificuldade com $\hat{\theta}_+$ é que o processo de decisão entre os modelos, representado pelas variáveis indicadoras, inflaciona o vício e a variância, enquanto que a ponderação em $\hat{\theta}$ evita isso.

Se o valor de α for aumentado obtêm-se valores aproximados das funções envolvendo Q e Q_+ para $\alpha = 0, 1$, correspondendo a escolher entre os modelos M_1 e M_2 a um nível bilateral de 80%, como está mostrado na Figura 2.

Figura 1: Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear simples com variância conhecida. Esquerda: $E(Q_+) - \delta$ (pontos), $\text{var}(Q_+)$ (sólido) e $\text{var}(Q_+) + \{E(Q_+) - \delta\}^2$ (traços) para $\alpha = 0,025$. Direita: funções correspondentes para o estimador ponderado usando Q .

A Figura 3 mostra os gráficos de $E(Q_{\text{BIC}}) - \delta$, $\text{var}(Q_{\text{BIC}})$ e $[\text{var}(Q_{\text{BIC}}) + (E(Q_{\text{BIC}}) - \delta)^2]$ para vários valores de δ e $n = 11, 41$ e 101 em comparação com os gráficos correspondentes ao uso do AIC. Observa-se que à medida que n aumenta, o vício e a variância também aumentam, sendo que a variância somente diminui para

Figura 2: Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear simples para diferentes valores de α . Linhas fortes: funções $E(Q_+) - \delta$ (pontos), $\text{var}(Q_+)$ (sólido) e $\text{var}(Q_+) + (E(Q_+) - \delta)^2$ (traços) para $\alpha = 0,01, 0,025, 0,05$ e $0,1$. Linhas finas: mesmas funções para Q .

valores de δ bem próximos de zero. Este resultado está provavelmente relacionado ao fato de $\bullet$ BIC, em contraste com o AIC, proporcionar uma seleção de modelos mais consistente, no sentido de que a probabilidade de que o modelo seja escolhido tende a um, à medida que n aumenta.

A comparação entre $\hat{\theta}$ e $\hat{\theta}_+$ quando a variância é desconhecida está

Figura 3: Propriedades dos estimadores de θ em regressão linear simples com variância conhecida. As figuras mostram $E(Q_{\text{BIC}}) - \delta$ (pontos), $\text{var}(Q_{\text{BIC}})$ (sólido) e $\text{var}(Q_{\text{BIC}}) + \{E(Q_{\text{BIC}}) - \delta\}^2$ (traços) para $n = 11, 41$ e 101 em comparação com o gráfico correspondente ao uso do AIC, no alto à esquerda.

na Figura 4, usando-se $\alpha = 0,025$ e as expressões dadas em (11) a (16). Como neste caso as funções Q' e Q'_+ dependem de n , foram usados $n = 11, 41$ e 101 . O comportamento é similar ao caso com variância conhecida e como se podia esperar, observa-se que mudanças no tamanho de n não provocam mudanças nas conclusões anteriores.

Figura 4: Propriedades dos estimadores $\hat{\theta}$ e $\hat{\theta}_+$ no modelos de regressão linear simples com variância desconhecida. Esquerda: $E(Q'_+) - \delta$ (pontos), $\text{var}(Q'_+)$ (sólido) e $\text{var}(Q'_+) + (E(Q'_+) - \delta)^2$ (traços) para $\alpha = 0,025$ e $n = 11, 41$ e 101 . Direita: o mesmo para Q' .

4.1.2 Regressão linear múltipla

Na Figura 5 estão as superfícies das funções de Q_{+1} e Q_{+2} , usando-se $\alpha = 0,025$, lado esquerdo, e Q_1 e Q_2 , lado direito, para valores de δ_1 e δ_2 , obtidas nas expressões de (17) a (22), e, para melhor visualização, na Figura 6 estão os perfis referentes às estas superfícies fixando δ_2 em -4 , -2 e 0 . Como no caso de regressão linear simples, $\hat{\theta}$ tem propriedades melhores do que $\hat{\theta}_+$.

As superfícies e perfis correspondentes às expressões (23) a (28), para o caso onde uma covariável é ignorada, estão nas Figuras 7 e 8. Como já podia ser esperado o vício é considerável nos dois estimadores, mas $\hat{\theta}$ ainda tem variância menor.

Para o caso onde é usado o BIC no cálculo dos pesos, em comparação com o uso do AIC, as superfícies referentes a $\text{var}(Q_1) + \text{var}(Q_2)$ e $\text{var}(Q_{BIC_1}) + \text{var}(Q_{BIC_2})$, para tamanhos de amostra $n = 11, 41, 101, 201$ e 1001 estão na Figura 9. Novamente observa-se que a variância aumenta quando n cresce, como foi visto no caso de regressão linear simples.

A Figura 10 traz as superfícies referentes aos EQM's de $\hat{\theta}_+$, esquerda, e $\hat{\theta}$, direita, considerando a variância desconhecida, utilizando as expressões (31) e (34), para valores de n iguais a $11, 41$ e 101 . As Figuras 11 e 12 apresentam os perfis referentes às superfícies para vício, variância e EQM, fixando δ_2 em $-4, -2$ e 0 , para $n = 11$ e 101 , respectivamente. Como no caso de regressão linear simples $\hat{\theta}$ tem propriedades melhores que $\hat{\theta}_+$ e mudanças em n não alteram as conclusões.

Para três covariáveis, as Figuras 13, 14 e 15 trazem as superfícies de vício, variância e EQM para $\hat{\theta}_+$ e $\hat{\theta}$ fixando-se, respectivamente, δ_1, δ_2 e δ_3 em 2 para o caso em que a variância é conhecida. Novamente confirma-se que as propriedades de $\hat{\theta}$ são melhores do que as de $\hat{\theta}_+$. Para variância desconhecida obtêm-se resultados semelhantes, como já foi verificado para os casos de uma e duas covariáveis.

Figura 5: Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com duas co-variáveis e variância conhecida. Esquerda: funções $[E(Q_{+1}) - \delta_1] + [E(Q_{+2}) - \delta_2]$, $\text{var}(Q_{+1}) + \text{var}(Q_{+2})$ e o correspondente EQM, com $\alpha = 0,025$. Direita: mesmas funções para Q_1 e Q_2 .

Figura 6: Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com duas co-variáveis e variância conhecida, fixando-se valores de δ_2 . Esquerda: perfis das funções $[E(Q_{+1}) - \delta_1] + [E(Q_{+2}) - \delta_2]$ (pontos), $\text{var}(Q_{+1}) + \text{var}(Q_{+2})$ (sólido) e o correspondente EQM (traços), com $\alpha = 0,025$. Direita: o mesmo para as funções Q_1 e Q_2 .

Figura 7: Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com duas co-variáveis e variância conhecida quando uma covariável é ignorada. Esquerda: superfícies de $[E(Q_{+1}) - \delta_1] + [E(Q_{+2}) - \delta_2]$, $\text{var}(Q_{+1}) + \text{var}(Q_{+2})$ e o correspondente EQM, com $\alpha = 0,025$. Direita: mesmas superfícies para Q_1 e Q_2 .

Figura 8: Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com duas co-variáveis e variância conhecida, fixando-se valores de δ_2 . Esquerda: perfis das funções $[E(Q_{+1}) - \delta_1] + [E(Q_{+2}) - \delta_2]$ (pontos), $\text{var}(Q_{+1}) + \text{var}(Q_{+2})$ (sólido) e o correspondente EQM (traços), com $\alpha = 0,025$. Direita: o mesmo para as funções Q_1 e Q_2 .

Figura 9: Propriedades dos estimadores de θ em regressão linear com duas covariáveis e variância conhecida. Topo à esquerda: $\text{var}(Q_1) + \text{var}(Q_2)$ usando AIC. Outras superfícies: o mesmo quando usando BIC no cálculo do peso, para vários valores de n .

Figura 10: Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com duas co-variáveis e variância desconhecida. Esquerda: superfícies correspondentes ao EQM de $\hat{\theta}_+$, com $\alpha = 0,025$. Direita: superfícies correspondentes ao EQM de $\hat{\theta}$. De alto a baixo: $n = 11, 41$ e 101 .

Figura 11: Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com duas co-variáveis e variância desconhecida, fixando-se valores de δ_2 e $n = 11$. Esquerda: perfis para vício (pontos), variância (sólido) e EQM (traços), com $\alpha = 0,025$, para $\hat{\theta}_+$. Direita: o mesmo para $\hat{\theta}$.

Figura 12: Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com duas co-variáveis e variância desconhecida, fixando-se valores de δ_2 e $n = 101$. Esquerda: perfis para vício (pontos), variância (sólido) e EQM (traços), com $\alpha = 0,025$, para $\hat{\theta}_+$. Direita: o mesmo para $\hat{\theta}$.

Figura 13: Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com três co-variáveis e variância conhecida, fixando-se $\delta_1 = 2$. Esquerda: superfícies correspondentes ao vício, à variância e ao EQM de $\hat{\theta}_+$, com $\alpha = 0,025$. Direita: mesmas superfícies para $\hat{\theta}$.

Figura 14: Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com três co-variáveis e variância conhecida, fixando-se $\delta_2 = 2$. Esquerda: superfícies correspondentes ao vício, à variância e ao EQM de $\hat{\theta}_+$, com $\alpha = 0,025$. Direita: mesmas superfícies para $\hat{\theta}$.

Figura 15: Propriedades dos estimadores $\hat{\theta}_+$ e $\hat{\theta}$ em regressão linear com três co-variáveis e variância conhecida, fixando-se $\delta_3 = 2$. Esquerda: superfícies correspondentes ao vício, à variância e ao EQM de $\hat{\theta}_+$, com $\alpha = 0,025$. Direita: mesmas superfícies para $\hat{\theta}$.

4.2 Comparação entre o estimador ponderado, o estimador BMA e resultados usando o método *bootstrap*

As comparações entre o estimador ponderado, o estimador BMA e resultados usando-se o método *bootstrap* são feitas a partir de quatro exemplos da literatura, considerando um caso de regressão linear simples, dois casos de regressão com duas covariáveis e um exemplo de regressão com três covariáveis. Nestas aplicações práticas, os valores de δ e σ^2 são substituídos pelas estimativas obtidas a partir do ajuste do modelo completo.

Os cálculos e gráficos foram feitos usando-se o S-PLUS e os programas estão no Apêndice 1. Para os cálculos envolvendo o BMA foi utilizada a função `MC3.REG.logpost`, desenvolvida por Jennifer Hoeting usando-se o S-PLUS e obtida como indicado em Hoeting et al. (1999)

4.2.1 Exemplo 1: Efeito da duração da área foliar sadia na produção de feijão

Seis experimentos foram conduzidos em Londrina, PR, com os cultivares Carioca, Iapar-57 e Rosinha com o objetivo de quantificar os danos causados pela infecção conjunta de *Phaeoisariopsis griseola* e *Collectotrichum lindemuthianum* e avaliar o efeito das doenças sobre a área foliar, as variáveis de área foliar sadia e os componentes de produção. A partir destes experimentos foram selecionados quatro conjunto de dados que correspondem às medidas de HDA (duração da área foliar sadia), x , e produção de feijão (g/m^2), y , dos cultivares Carioca e Iapar-57 em 1996 e 1998, denominados Car96, Car98, Iap96 e Iap98. O experimento completo é detalhado e analisado em Carneiro (1999). Os dados usados neste exemplo estão no Apêndice 2. Os tamanhos de amostra para cada conjunto são, respectivamente, $n = 47, 48, 48$ e 47 . Os valores de x foram centrados e a última observação foi utilizada como x_+ . A Tabela 1 mostra as variâncias de $\hat{\theta}$ para os estimadores ponderados usando-se (16) e (3), o estimador BMA e os dois estimadores *bootstrap* detalhados em 3.4.

Tabela 1: Variâncias de $\hat{\theta}$ para o exemplo de produção de feijão. V1: variância em (16); V2: variância em (3).

Método	Cultivar e ano			
	Car96	Car98	Iap96	Iap98
V1	282	1257	90	513
V2	281	1252	90	512
BMA	320	1185	85	485
Bootstrap 1	342	2078	83	1762
Bootstrap 2	302	1785	88	596

As variâncias ponderadas, V1 e V2, têm valores muito próximos e comparadas com a variância BMA são bastante similares também. Os valores *bootstrap* ficaram relativamente próximos dos anteriores para as produções do ano 96, mas foram substancialmente maiores no ano 98. Observa-se que neste ano a variabilidade foi maior e então o *bootstrap* parece estar sendo muito influenciado por isso. Ao comparar as variâncias da Tabela 1 com as variâncias correspondentes aos valores preditos pelo modelo completo, que são, respectivamente, 281, 1252, 90 e 512, observa-se pouquíssima diferença, principalmente para V2. Isto se dá porque o nível de significância para a inclusão da covariável no modelo é muito pequeno (quase zero) e então a ponderação tem influência pequena.

4.2.2 Exemplo 2: Concentração de fósforo

Este exemplo é uma pesquisa sobre a fonte da qual plantas de milho de vários solos do estado de Iowa, EUA, obtêm fósforo. As concentrações de fósforo inorgânico (x_1) e orgânico (x_2) no solo foram determinadas quimicamente e foi medido o conteúdo (y) de fósforo no milho crescido nos vários solos. A fonte deste exemplo é Snedecor e Cochran (1967), pag. 384, e os dados estão no Apêndice 2. Há $n = 18$ observações, das quais foi excluído um *outlier*. Os dados foram padronizados

a fim de simplificar o uso dos procedimentos envolvidos nos cálculos BMA, e, e a última observação foi reservada para predição. Os dados foram padronizados novamente e as colunas correspondentes às covariáveis foram ortogonalizadas, seguindo o procedimento de ortogonalização de Gram-Schmidt (Draper & Smith, 1966, p. 275). A ortogonalização foi devida à suposição colocada para o estimador ponderado. Os quatro modelos possíveis foram ajustados e os pesos e as probabilidades a posteriori dos modelos estão na Tabela 2. As variâncias de $\hat{\theta}$ para (33), (3), BMA, e para os dois *bootstrap*, foram respectivamente 0,27, 0,21, 0,26, 0,24, 0,24, e são bastante similares. Verifica-se ainda que os pesos atribuídos aos diferentes modelos são muito semelhantes para os dois métodos.

Tabela 2: Modelos, pesos e probabilidades a posteriori dos modelos para o exemplo do fósforo.

Preditor	peso	Probabilidade a Posteriori
$\hat{y}_{1+} = \hat{\beta}_0$	0,0158	0,0545
$\hat{y}_{2+} = \hat{\beta}_0 + \hat{\beta}_1 x_{1+}$	0,6926	0,8477
$\hat{y}_{3+} = \hat{\beta}_0 + \hat{\beta}_2 x_{2+}$	0,0062	0,0056
$\hat{y}_{4+} = \hat{\beta}_0 + \hat{\beta}_1 x_{1+} + \hat{\beta}_2 x_{2+}$	0,2854	0,0922

A partir dos parâmetros estimados neste exemplo foram geradas 10.000 amostras ($\beta_0 = 0$, $\beta_1 = 0.7$, $\beta_2 = -0.1$ e $\sigma^2 = 0.61$) e $\hat{\theta}$ foi calculado para o método ponderado e Bayesiano. As médias e variâncias estão na Tabela 3 e os histogramas respectivos estão na Figura 16. O valor verdadeiro de θ é 1.28 e o vício observado pode ser explicado pelo pequeno tamanho da amostra, $n = 16$. De uma forma geral, analisando-se os histogramas da Figura 16, a distribuição de $\hat{\theta}$ é semelhante para os dois métodos.

Como ilustração, foram sorteadas três amostras, dentre as 10.000 amostras simuladas, e os cálculos das variâncias relativos a elas estão na Tabela 4. Novamente, observa-se que os valores são bastante similares. A Figura 17 mostra

Tabela 3: Resultados da simulação para o exemplo do fósforo.

	ponderado	BMA
média de $\hat{\theta}$	1,19	1,12
variância $\hat{\theta}$	0,12	0,18

Figura 16: Histogramas de $\hat{\theta}$ simulado. Esquerda: ponderado; Direita: BMA. A linha pontilhada está sobre o valor verdadeiro de θ .

os histogramas de $\hat{\theta}^*$ para cada uma das três amostras, considerando os dois tipos de reamostragem *bootstrap*. O *bootstrap* 2, correspondente a compor $\hat{\theta}^*$ a partir de amostras *bootstrap* obtidas pela reamostragem dos resíduos de cada modelo em estudo, parece estar reproduzindo melhor a distribuição do lado esquerdo da Figura 16, referente ao estimador ponderado.

4.2.3 Exemplo 3: Volume de cerejeiras

Este exemplo foi obtido de Ryan et al. (1976), pag. 278, e é também uma regressão linear com duas covariáveis. O objetivo é determinar o volume de uma

Tabela 4: Resultados para três amostras do estudo de simulação do exemplo do fósforo.

Amostra	$\hat{\theta}_w$	$\hat{\theta}_{BMA}$	Variância				
			V1	V2	BMA	Bootstrap 1	Bootstrap 2
1	1,27	1,29	0,22	0,19	0,21	0,17	0,40
2	1,62	1,61	0,12	0,11	0,10	0,18	0,11
3	1,39	1,45	0,14	0,14	0,13	0,16	0,15

Tabela 5: Modelos, pesos e probabilidades a posteriori dos modelos para o exemplo das cerejeiras.

Preditor	peso	Probabilidade a Posteriori
$\hat{y}_{1+} = \hat{\beta}_0$	0,0000	0,0000
$\hat{y}_{2+} = \hat{\beta}_0 + \hat{\beta}_1 x_{1+}$	0,0001	0,0194
$\hat{y}_{3+} = \hat{\beta}_0 + \hat{\beta}_2 x_{2+}$	0,0000	0,0000
$\hat{y}_{4+} = \hat{\beta}_0 + \hat{\beta}_1 x_{1+} + \hat{\beta}_2 x_{2+}$	0,9999	0,9806

árvore (y) em função de seu diâmetro (x_1) e de sua altura (x_2). Os dados são referentes a árvores de cerejas pretas provenientes de Allegheny National Forest, Pennsylvania, EUA, com 31 observações. Trabalhou-se com o logaritmo das observações. Uma observação foi excluída por se tratar de um *outlier* e as outras observações receberam o mesmo tratamento descrito no Exemplo 2 referente à padronização e ortogonalização, também reservando a última observação para previsão. As Tabelas 5 e 6 mostram os pesos e probabilidades a posteriori dos quatro modelos possíveis e as variâncias correspondentes a V1, V2, BMA e os dois casos *bootstrap*.

Os dois métodos produziram valores similares para os pesos e observa-se que o modelo contendo as duas covariáveis é uma escolha unânime. É de se esperar que as variâncias tenham praticamente o mesmo valor que a variância correspondente

Figura 17: Histogramas de $\hat{\theta}^*$ para três amostras simuladas. No alto: amostra 1, 2 e 3 para o *bootstrap* 1; Embaixo: o mesmo para *bootstrap* 2. A linha pontilhada está sobre o valor verdadeiro de θ e a linha tracejada corresponde às médias *bootstrap* de $\hat{\theta}^*$.

ao valor predito pelo modelo completo, que é 0,0043, o que ocorre, a não ser para a variância BMA que foi praticamente o dobro deste valor.

Tabela 6: Variâncias de $\hat{\theta}$ para o exemplo das cerejeiras. V1: variância em (33); V2: variância em (3).

V1	V2	BMA	<i>bootstrap 1</i>	<i>bootstrap 2</i>
0,0043	0,0043	0,0083	0,0040	0,0044

4.2.4 Exemplo 4: Temperatura do ar

Este exemplo refere-se a uma regressão linear múltipla com 3 co-variáveis. O objetivo é determinar a temperatura média do ar em Janeiro (y), para cidades do Estado de São Paulo, como função da latitude (x_1), longitude (x_2) e altitude (x_3). Os dados, obtidos das Normas Climatológicas 1961–1990 do DNMET e do Instituto Agrônomo de Campinas, estão no Apêndice 2 contendo informações para 21 cidades. Uma observação foi excluída por se tratar de uma cidade acima de 1.000 metros de altitude. As outras observações receberam o mesmo tratamento descrito no Exemplo 2 referente a padronização e ortogonalização, também reservando-se a última observação para previsão. O número de modelos possíveis é $2^3 = 8$ e os modelos estão descritos na Tabela 7, onde também estão os pesos e as probabilidades a posteriori dos modelos.

Os valores dos pesos e das probabilidades a posteriori dos modelos são similares e como esperado dão peso maior ao modelo contendo a latitude e a altitude. A Tabela 8 mostra as variâncias e, novamente, observam-se valores similares, como nos exemplos anteriores.

Tabela 7: Modelos, pesos e probabilidades a posteriori dos modelos para o exemplo da temperatura.

Preditor	peso	Probabilidade a Posteriori
$\hat{y}_{1+} = \hat{\beta}_0$	0,0003	0,0050
$\hat{y}_{2+} = \hat{\beta}_0 + \hat{\beta}_1 x_{1+}$	0,0041	0,0175
$\hat{y}_{3+} = \hat{\beta}_0 + \hat{\beta}_2 x_{2+}$	0,0001	0,0005
$\hat{y}_{4+} = \hat{\beta}_0 + \hat{\beta}_3 x_{3+}$	0,0057	0,0282
$\hat{y}_{5+} = \hat{\beta}_0 + \hat{\beta}_1 x_{1+} + \hat{\beta}_2 x_{2+}$	0,0019	0,0018
$\hat{y}_{6+} = \hat{\beta}_0 + \hat{\beta}_1 x_{1+} + \hat{\beta}_3 x_{3+}$	0,6310	0,8395
$\hat{y}_{7+} = \hat{\beta}_0 + \hat{\beta}_2 x_{2+} + \hat{\beta}_3 x_{3+}$	0,0026	0,0029
$\hat{y}_{8+} = \hat{\beta}_0 + \hat{\beta}_1 x_{1+} + \hat{\beta}_2 x_{2+} + \hat{\beta}_3 x_{3+}$	0,3541	0,1046

Tabela 8: Variâncias de $\hat{\theta}$ para o exemplo da temperatura. V1: variância em (39); V2: variância em (3).

V1	V2	BMA	<i>bootstrap 1</i>	<i>bootstrap 2</i>
0,0757	0,0663	0,0703	0,0862	0,0857

5 CONCLUSÕES

Neste trabalho foram estudadas as propriedades de um estimador ponderado para a predição, em modelos de regressão linear, que incorpora de forma simples a incerteza devido à escolha de um modelo. As propriedades deste estimador foram comparadas com a situação usual de escolha de um modelo, com a ponderação Bayesiana de modelos e com duas abordagens usando o método *bootstrap*.

A comparação entre o estimador ponderado e o obtido a partir da escolha de um modelo, dentro do contexto de regressão linear, mostrou que:

- quando os testes para os coeficientes da regressão, β , são pouco significativos, o erro quadrático médio da predição pode ser substancialmente reduzido pelo uso do estimador ponderado;
- quando os testes para os coeficientes da regressão são altamente significativos, o uso do estimador ponderado não é justificado, apresentando resultados equivalentes aos obtidos com o uso do modelo selecionado.

Com relação à comparação entre o estimador ponderado e o BMA conclui-se que, para esta situação simples de regressão linear, os dois são similares. Vale observar que o estimador ponderado é de mais fácil implementação e é independente da escolha de distribuições a priori.

De um modo geral, as formulações diferentes para o cálculo da variância forneceram resultados parecidos, embora seja necessário maior desenvolvimento teórico e empírico para a generalização deste resultado.

As duas alternativas para se utilizar o método *bootstrap* também forneceram resultados similares, havendo indicação de que o uso do *bootstrap* é uma boa alternativa. Este é um resultado importante para a continuação deste trabalho, no sentido de estender o estudo de ponderação de modelos para situações mais complexas, onde certamente os desenvolvimentos analíticos apresentam maior dificuldade.

REFERÊNCIAS BIBLIOGRÁFICAS

- AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In **Breakthroughs in Statistics**, v.1, eds. S. Kotz & N. L. Johnson, p.610-624. New York: Springer. 1973.
- BARNARD, G. A. New methods of quality control. **Journal of the Royal Statistical Society, Série A**, v.126, p.255-258, 1963.
- BICKEL, J. & FREEDMAN, D. A. Some asymptotic theory for the bootstrap. **The Annals of Statistics**, v.9, p.1196-1217, 1981.
- BUCKLAND, S. T., BURNHAN, K. P. and AUGUSTIN, N. H. Model selection: An integral part of inference. **Biometrics**, V.53, P.603-618, 1997.
- CARNEIRO, S. M. T. P. G. Efeito da infecção conjunta de *Phaeoisariopsis griseola* e *Colletotrichum lindemuthianum* nos componentes de produção e nas variáveis de área foliar do feijoeiro. Piracicaba, 1999. 89p. Tese (Doutorado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.
- CHATFIELD, C. Model uncertainty, data mining and statistical inference (with Discussion). **Journal of the Royal Statistical Society, Série A**, v.158, p.419-466, 1995.
- CLEMEN, R. T. Combining forecasts: A review and annotated bibliography. **International Journal of Forecasting**, v.5, p.559-583, 1989.
- DAVISON, A. C. & HINKLEY, D. V. **Bootstrap Methods and their Application**. Cambridge: Cambridge University Press, 1997. 582p.

- DRAPER, N. R. & SMITH, H. (1981) **Applied Regression Analysis** 2.ed. New York: Wiley, 1981. 709p.
- DRAPER, D. Assessment and propagation of model uncertainty (with Discussion). **Journal of the Royal Statistical Society, Série B**, v.57, p.45–97. 1995.
- EFRON, B. Bootstrap methods: another look at the jackknife. **Annals of Statistics**, v.7, p.1-26, 1979.
- EFRON, B. & TIBSHIRANI, R. J. **An Introduction to the Bootstrap**. New York: Chapman & Hall, 1993. 436p.
- FREEDMAN, D. A. Bootstrapping regression models. **Annals of Statistics**, v.9, p.1218-1228, 1981.
- HALL, P. **The Bootstrap and Edgeworth Expansion**. New York: Springer, 1992. 352p.
- HOETING, J. A. (1994) Accounting for model uncertainty in linear regression. Seattle, 1994. 167p. Thesis (Ph.D.) – University of Washington.
- HOETING, J. A., MADIGAN, D., RAFTERY, A. E. & VOLINSKY, C. T. Bayesian model averaging: a tutorial (with Discussion). **Statistical Science**, v.14, p.382–417, 1999.
- KASS, R. E. & RAFTERY, A. E. Bayes Factors. **Journal of the American Statistical Association**, v.90, p.773–795, 1995.
- LEAMER, E. E. **Specification Searches**. New York: Wiley, 1978. 370p.
- MADIGAN, D. & RAFTERY, A. E. Model selection and accounting for model uncertainty in graphical models using Occam's window. **Journal of the American Statistical Association**, v.89, p.1535–1546, 1994.
- MADIGAN, D. & YORK, J. Bayesian graphical models for discrete data. **International Statistical Review**, v.63, p.215–232, 1995.

- MALLOWS, C. L. Some comments on C_p . **Technometrics**, v.15, p.661-675, 1973.
- RAFTERY, A. E., MADIGAN, D. & HOETING, J. A. Bayesian model averaging for linear regression models. **Journal of the American Statistical Association**, v.92, p.179-191, 1997.
- ROBERTS, H. V. Probabilistic prediction. **Journal of the American Statistical Association**, v.60, p.50-62, 1965.
- RYAN, B.F, JOINER, B. L. & RYAN Jr., T. A. **Minitab Student Handbook**. North Scituate: Duxbury Press, 1976. 341p.
- SCHWARZ, G. Estimating the dimensions of a model. **Annals of Statistics**, v.6, p.461-463, 1978.
- SNEDECOR, G. W. & COCHRAN, W. G. **Statistical Methods**. 6.ed. Ames: Iowa State University Press, 1967. 593p.
- SINGH, K. On the asymptotic accuracy of Efrons's bootstrap. **The Annals of Statistics**, v.9, p. 1187-1195, 1981.
- TIERNEY, L. & KADANE, J. B. Accurate approximations for posterior moments and marginal densities. **Journal of the American Statistical Association**, v.81, p.82-86, 1986.
- WEISBERG, S. **Applied Linear Regression**. 2.ed. New York: Wiley, 1985. 324p.

APÊNDICE 1

Programas

(a) PROGRAMAS EM S-PLUS UTILIZADOS PARA GERAR OS GRÁFICOS DE 4.1

```

=====
#regressao linear simples: variancia conhecida - figura 1
options(object.size=8e6)
delta<-seq(-4,4,length=100)
zalfa<-c(-2.58,-1.96,-1.64,-1.28,-0.84,0)
deltaz<-rep(delta,rep(10000,100))
zz<-rep(z,100)

Q<-(deltaz+zz)*(I(zz<(zalfa[2]-deltaz))+I(zz>(-zalfa[2]-deltaz)))
mu.bias2<-tapply(Q,list(deltaz),mean) - delta
mu.var2<-tapply(Q,list(deltaz),var)
mu.mse2<-mu.var2 + (mu.bias2)^2
Qw<-(deltaz+zz)*(exp((deltaz+zz)^2/2 -1)/(1+exp((deltaz+zz)^2/2 -1)))
muw.bias<-tapply(Qw,list(deltaz),mean) - delta
muw.var<-tapply(Qw,list(deltaz),var)
muw.mse<-muw.var + (muw.bias)^2

par(mfrow=c(1,2))
plot(delta,mu24w.m[,1],type="n",xlab="delta",ylab="Vicio,variancia,EQM",
ylim=c(-1,3))
lines(delta,mu24w.m[,1],type="l",lty=1)
lines(delta,mu24w.m[,2],lty=7); lines(delta,mu24w.m[,3],lty=8)

plot(delta,mu24w.m[,4],type="n",xlab="delta",ylab="Vicio,variancia,EQM",
ylim=c(-1,3))
lines(delta,mu24w.m[,4],type="l",lty=1)
lines(delta,mu24w.m[,5],lty=7); lines(delta,mu24w.m[,6],lty=8)

=====
#regressao linear simples: variancia desconhecida - figura 4

options(object.size=8e6)
delta<-seq(-4,4,length=100)
deltaz<-rep(delta,rep(10000,100))
z<-rnorm(10000,0,1); zz<-rep(z,100)
n<-101
c101<-sqrt(rchisq(10000,n-2)/(n-2)); cc101<-rep(c101,100)
c<-cc101
aux3<-qt(0.025,(n-2))
Q2<-(zz+deltaz)*(I((1/c)*(zz+deltaz)< aux3)+I((1/c)*(zz+deltaz) > -aux3))
TT<-(zz+deltaz)/c
Qw2<-(deltaz+zz)*(exp((n/2)*log(1+TT^2/(n-2)) -1)/(1+exp((n/2)*log(1+TT^2/
(n-2)) -1)))

muu101.bias2<-tapply(Q2,list(deltaz),mean) - delta
muu101.var2<-tapply(Q2,list(deltaz),var)
muu101.mse2<-muu101.var2 + (muu101.bias2)^2
muuw101.bias<-tapply(Qw2,list(deltaz),mean) - delta
muuw101.var<-tapply(Qw2,list(deltaz),var)
muuw101.mse<-muuw101.var + (muuw101.bias)^2

par(mfrow=c(3,2))
plot(delta,muu11.var2,type="l",lty=1,xlab="delta",ylab="Vicio,variancia,EQM",
ylim=c(-1,3))
lines(delta,muu11.bias2,lty=2); lines(delta,muu11.mse2,lty=8)
title(main=" n=11")

```



```

plot(delta,muuw11.var,type="l",xlab="delta",ylab="Vicio,variancia,EQM",
ylim=c(-1,3))
lines(delta,muuw11.bias,lty=2); lines(delta,muuw11.mse,lty=8)
title(main=" n=11")

plot(delta,muu41.var2,type="l",lty=1,xlab="delta",ylab="Vicio,variancia,EQM",
ylim=c(-1,3))
lines(delta,muu41.bias2,lty=2); lines(delta,muu41.mse2,lty=8)
title(main=" n=41")

plot(delta,muuw41.var,type="l",xlab="delta",ylab="Vicio,variancia,EQM",
ylim=c(-1,3))
lines(delta,muuw41.bias,lty=2); lines(delta,muuw41.mse,lty=8)
title(main=" n=41")

plot(delta,muu101.var2,type="l",lty=1,xlab="delta",ylab="Vicio,variancia,EQM",
ylim=c(-1,3))
lines(delta,muu101.bias2,lty=2); lines(delta,muu101.mse2,lty=8)
title(main=" n=101")

plot(delta,muuw101.var,type="l",xlab="delta",ylab="Vicio,variancia,EQM",
ylim=c(-1,3))
lines(delta,muuw101.bias,lty=2); lines(delta,muuw101.mse,lty=8)
title(main=" n=101")

=====
#regressao linear multipla: variancia conhecida - superficies e
#perfis - figuras 5 e 6
options(object.size=15e6)
delta1<-seq(-4,4,length=30)
delta2<-seq(-4,4,length=30)
z1<-rnorm(10000,0,1)
z2<-rnorm(10000,0,1)
GG<-function(a,b){
  (1+exp(a))/(1+exp(b[,1])+exp(b[,2])+exp(b[,1]+b[,2]))}
mq1<-rep(0,900); mq2<-rep(0,900)
vq1<-rep(0,900); vq2<-rep(0,900)
mqw1<-rep(0,900); mqw2<-rep(0,900)
vqw1<-rep(0,900); vqw2<-rep(0,900)
k<-1
for(i in 1:30){
  for(j in 1:30){
    Q1<-(delta1[i]+z1)*(I(z1<(zalfa[2]-delta1[i]))+I(z1>(-zalfa[2]-delta1[i])))
    Q2<-(delta2[j]+z2)*(I(z2<(zalfa[2]-delta2[j]))+I(z2>(-zalfa[2]-delta2[j])))
    mq1[k]<-mean(Q1) - delta1[i]; mq2[k]<-mean(Q2) - delta2[j]
    vq2[k]<-var(Q1); vq2[k]<-var(Q2)
    u1<--(delta1[i]+z1)^2/2 + 1; u2<--(delta2[j]+z2)^2/2 + 1
    uu<-cbind(u1,u2)
    Q1w<-(delta1[i]+z1)*GG(u2,uu)
    Q2w<-(delta2[j]+z2)*GG(u1,uu)
    mqw1[k]<-mean(Q1w) - delta1[i]
    mqw2[k]<-mean(Q2w) - delta2[j]
    vqw2[k]<-var(Q1w) ; vqw2[k]<-var(Q2w)
    k<-k+1
  }}

#superficies
par(mfrow=c(3,2))

```

```

inter<-interp(delta1,delta2,mu2d.bias2)
persp(inter,xlab="delta1",ylab="delta2",zlab="vicio",zlim=c(-1.5,1.5))
inter<-interp(delta1,delta2,muw2d.bias)
persp(inter,xlab="delta1",ylab="delta2",zlab="vicio",zlim=c(-1.5,1.5))

inter<-interp(delta1,delta2,mu2d.var2)
persp(inter,xlab="delta1",ylab="delta2",zlab="variancia",zlim=c(0,5))
inter<-interp(delta1,delta2,muw2d.var)
persp(inter,xlab="delta1",ylab="delta2",zlab="variancia",zlim=c(0,5))

inter<-interp(delta1,delta2,mu2d.mse2)
persp(inter,xlab="delta1",ylab="delta2",zlab="EQM",zlim=c(0,6))
inter<-interp(delta1,delta2,muw2d.mse)
persp(inter,xlab="delta1",ylab="delta2",zlab="EQM",zlim=c(0,6))

#perfis
d1<-seq(-4,4,length=30); par(mfrow=c(3,2))
#delta= -4
bias<-mq1[1:30] + mq2[1:30] ; var<-vq1[1:30] + vq2[1:30]
mse<-var + bias^2
plot(d1,var,type="l",xlab="delta1",ylab="Vicio,Variancia,EQM",ylim=c(-1,3.5))
lines(d1,bias,lty=7); lines(d1,mse,lty=8)
title(main="delta2 = -4")

biasw<-mqw1[1:30] + mqw2[1:30] ; varw<-vqw1[1:30] + vqw2[1:30]
msew<-varw + biasw^2
plot(d1,varw,type="l",xlab="delta1",ylab="Vicio,Variancia,EQM",ylim=c(-1,3.5))
lines(d1,biasw,lty=7); lines(d1,msew,lty=8)
title(main="delta2 = -4")

#delta= -2
bias<-mq1[211:240] + mq2[211:240]
var<-vq1[211:240] + vq2[211:240]; mse<-var + bias^2
plot(d1,var,type="l",xlab="delta1",ylab="Vicio,Variancia,EQM",ylim=c(-1,3.5))
lines(d1,bias,lty=7); lines(d1,mse,lty=8)
title(main="delta2 = -2")

biasw<-mqw1[211:240] + mqw2[211:240];
varw<-vqw1[211:240] + vqw2[211:240]; msew<-varw + biasw^2
plot(d1,varw,type="l",xlab="delta1",ylab="Vicio,Variancia,EQM",ylim=c(-1,3.5))
lines(d1,biasw,lty=7); lines(d1,msew,lty=8)
title(main="delta2 = -2")

#delta= 0
bias<-mq1[451:480] + mq2[451:480];
var<-vq1u11[451:480] + vq2u11[451:480] ; mse<-var + bias^2
plot(d1,var,type="l",xlab="delta1",ylab="Vicio,Variancia,EQM",ylim=c(-1,3.5))
lines(d1,bias,lty=7); lines(d1,mse,lty=8)
title(main="delta2 = 0")

biasw<-mqw1[451:480] + mqw2[451:480];
varw<-vq1u11[451:480] + vq2u11[451:480] ; msew<-varw + biasw^2
plot(d1,varw,type="l",xlab="delta1",ylab="Vicio,Variancia,EQM",ylim=c(-1,3.5))
lines(d1,biasw,lty=7); lines(d1,msew,lty=8)
title(main="delta2 = 0")

=====
#regressao linear multipla: variancia desconhecida -superficies e

```

```

#perfis - figuras 10 e 11

options(object.size=15e6)
delta1<-seq(-4,4,length=30); delta2<-seq(-4,4,length=30)
z1<-rnorm(10000,0,1); z2<-rnorm(10000,0,1)
G2u<-function(u,n,U){
  (1+exp(1)*(1+u)^(-n/2))/(1 + exp(1)*(1+U[,1])^(-n/2) +
exp(1)*(1+U[,2])^(-n/2) + exp(2)*(1+U[,1]+U[,2])^(-n/2))}

=====
n<-11
c<-rchisq(10000,n-3); t1<-qt(0.975,n-3)
mq1u112<-rep(0,900); mq2u112<-rep(0,900)
vq1u112<-rep(0,900); vq2u112<-rep(0,900)
mqw1u112<-rep(0,900); mqw2u112<-rep(0,900)
vqw1u112<-rep(0,900); vqw2u112<-rep(0,900)
k<-1
for(i in 1:30){
  for(j in 1:30){
    a1<-sqrt(n-3)*(z1+delta1[i])/sqrt(c)
    Q1<-(delta1[i]+z1)*(I(a1 < -t1)+I(a1 > t1))
    a2<-sqrt(n-3)*(z2+delta2[j])/sqrt(c)
    Q2<-(delta2[j]+z2)*(I(a2 < -t1)+I(a2 > t1))
    mq1u112[k]<-mean(Q1) - delta1[i]; mq2u112[k]<-mean(Q2) - delta2[j]
    vq1u112[k]<-var(Q1); vq2u112[k]<-var(Q2)
    u1<-(delta1[i]+z1)^2/c; u2<-(delta2[j]+z2)^2/c
    uu<-cbind(u1,u2)
    Q1w<-(delta1[i]+z1)*(G2u(u2,n,uu))
    Q2w<-(delta2[j]+z2)*(G2u(u1,n,uu))
    mqw1u112[k]<-mean(Q1w) - delta1[i]
    mqw2u112[k]<-mean(Q2w) - delta2[j]
    vqw1u112[k]<-var(Q1w) ; vqw2u112[k]<-var(Q2w)
    k<-k+1
  }}

#superficies
par(mfrow=c(3,2))
bias<-mq1u11+mq2u11; biasw<-mqw1u11+mqw2u11
var<-vq1u11 + vq2u11 ; varw<-vqw1u11 + vqw2u11
mse<-var + bias^2; msew<-varw + biasw^2
inter<-interp(delta1,delta2,mse)
persp(inter,xlab="delta1",ylab="delta2",zlab="EQM",zlim=c(0,6))
title(main="n=11")
inter<-interp(delta1,delta2,msew)
persp(inter,xlab="delta1",ylab="delta2",zlab="EQM",zlim=c(0,6))
title(main="n=11")

bias<-mq1u41+mq2u41; biasw<-mqw1u41+mqw2u41
var<-vq1u41 + vq2u41 ; varw<-vqw1u41 + vqw2u41
mse<-var + bias^2; msew<-varw + biasw^2
inter<-interp(delta1,delta2,mse)
persp(inter,xlab="delta1",ylab="delta2",zlab="EQM",zlim=c(0,6))
title(main="n=41")
inter<-interp(delta1,delta2,msew)
persp(inter,xlab="delta1",ylab="delta2",zlab="EQM",zlim=c(0,6))
title(main="n=41")

bias<-mq1u101+mq2u101; biasw<-mqw1u101+mqw2u101

```

```

var<-vq1u101 + vq2u101 ; varw<-vqw1u101 + vqw2u101
mse<-var + bias^2, msew<-varw + biasw^2
inter<-interp(delta1,delta2,mse)
persp(inter,xlab="delta1",ylab="delta2",zlab="EQM",zlim=c(0,6))
title(main="n=101")
inter<-interp(delta1,delta2,msew)
persp(inter,xlab="delta1",ylab="delta2",zlab="EQM",zlim=c(0,6))
title(main="n=101")

#perfis
d1<-seq(-4,4,length=30)
#n=11
bias<-mq1u11+mq2u11; biasw<-mqw1u11+mqw2u11
var<-vq1u11 + vq2u11 ; varw<-vqw1u11 + vqw2u11
mse<-var + bias^2; msew<-varw + biasw^2
par(mfrow=c(3,2))
#delta= -4
b<-bias[1:30] ; v<-var[1:30] ; m<-mse[1:30]
plot(d1,v,type="l",xlab="delta1",ylab="Vicio,Variancia,EQM",ylim=c(-3,6))
lines(d1,b,lty=7); lines(d1,m,lty=8)
title(main="delta2 = -4, n=11")
b<-biasw[1:30]; v<-varw[1:30]; m<-msew[1:30]
plot(d1,v,type="l",xlab="delta1",ylab="Vicio,Variancia,EQM",ylim=c(-3,6))
lines(d1,b,lty=7); lines(d1,m,lty=8)
title(main="delta2 = -4, n=11")

#delta= -2
b<-bias[211:240]; v<-var[211:240]; m<-mse[211:240]
plot(d1,v,type="l",xlab="delta1",ylab="Vicio,Variancia,EQM",ylim=c(-3,6))
lines(d1,b,lty=7); lines(d1,m,lty=8)
title(main="delta2 = -2, n=11")
b<-biasw[211:240]; v<-varw[211:240]; m<-msew[211:240]
plot(d1,v,type="l",xlab="delta1",ylab="Vicio,Variancia,EQM",ylim=c(-3,6))
lines(d1,b,lty=7); lines(d1,m,lty=8)
title(main="delta2 = -2, n=11")

#delta= 0
b<-bias[451:480]; v<-var[451:480]; m<-mse[451:480]
plot(d1,v,type="l",xlab="delta1",ylab="Vicio,Variancia,EQM",ylim=c(-3,6))
lines(d1,b,lty=7); lines(d1,m,lty=8)
title(main="delta2 = 0, n=11")
b<-biasw[451:480]; v<-varw[451:480]; m<-msew[451:480]
plot(d1,v,type="l",xlab="delta1",ylab="Vicio,Variancia,EQM",ylim=c(-3,6))
lines(d1,b,lty=7); lines(d1,m,lty=8)
title(main="delta2 = 0, n=11")

```

(b) FUNÇÕES S-PLUS USADAS NOS CÁLCULOS DO EXEMPLOS DE 4.2

```

=====
#funcoes para modelo com uma covariavel e sem padronizar os dados
=====
weight.fun1x<-function(n,y,x,Xsp,zc){
#funcao que calcula os pesos, a var. buck, o vicio e a variancia com
# sigma desconhecido, para modelos com uma covariavel
#n: tamanho da amostra;
#ydata:vetor das respostas; x: vetor da covariavel

```

```

#Xsp: valor futuro de x; zc: matriz contendo as var. z1 e c1 (10000 de cada)
aux<-matrix(0,ncol=5,nrow=2); aux2<-matrix(0,ncol=5,nrow=1)
fit1<-lm(y~1)
aux[1,1]<-deviance(fit1)
#numero de parametros no modelo
aux[1,2]<-1
aux[1,3]<-fit1$df.residual
aux[1,4]<-fit1$coef
#variancia de b0
aux[1,5]<-(deviance(fit1)/fit1$df.residual)/n
yx<-data.frame(y,x); xplus<-data.frame(x=Xsp)
fit1<-lm(y~.,data=yx)
pred1<-predict.lm(fit1,newdata=xplus,se.fit=T)
aux[2,1]<-deviance(fit1)
#numero de parametros do modelo (contando com b0)
aux[2,2]<-2
aux[2,3]<-fit1$df.residual
#valor predito e seu desvio padrao
aux[2,4]<-pred1$fit
aux[2,5]<-pred1$se.fit^2
coef<-fit1$coef
a<-n*log(aux[,1]/n) +2*aux[,2] ; w<-exp(-a/2)/sum(exp(-a/2))
yh<-sum(w*aux[,4])
#varw buck
aux2[1,1]<-(aux[2,1]/aux[2,3])
aux2[1,2]<-yh
aux2[1,3]<-((w%*(sqrt(aux[,5] + (aux[,4]-yh)^2)))^2
v<-(aux[2,1]/aux[2,3])/sum(x^2); d<-coef[2]/sqrt(v)
u<-(n/2)*log(1 + ((zc[,1]+d)/sqrt(zc[,2]/(n-2)))^2/(n-2)) - 1
Q<-(d +zc[,1])*(G1(u))
# bw1 e vw1: vicio e variancia com sigma desconhecido
aux2[1,4]<-Xsp*sqrt(v)*(mean(Q)-d)
aux2[1,5]<-aux2[1,1]/n + Xsp^2*v*var(Q)
aux2
}

=====
bma.fun1x<-function(n,y,x,Xsp){
#funcao que calcula a posterior model probability, o estimador, a
#esperanca e a variancia BMA, para modelos com 1 cov. sem escalonar
aux<-matrix(0,ncol=2,nrow=2); aux2<-matrix(0,ncol=2,nrow=1)
fit1<-lm(y~1)
aux[1,1]<-fit1$coef
aux[1,2]<-MC3.REG.logpost1x1(as.matrix(y),nu=2.58,lambda=0.28,phi=2.85)
yx<-data.frame(y,x); xplus<-data.frame(x=Xsp)
fit1<-lm(y~.,data=yx)
pred1<-predict.lm(fit1,newdata=xplus,se.fit=T)
aux[2,1]<-pred1$fit
aux[2,2]<-MC3.REG.logpost1x(as.matrix(y), x, nu=2.58, lambda=0.28, phi=2.85)
pp<-exp(aux[,2])/sum(exp(aux[,2]))
aux4<-matrix(0,ncol=2,nrow=2)
aux4[1,1:2]<-yhatbayes1.fun1x(as.matrix(y),nu=2.58, lambda=0.28, phi=2.85)
aux4[2,1:2]<-yhatbayes.fun1x(as.matrix(y),as.matrix(x),Xsp,
nu=2.58, lambda=0.28, phi=2.85)
#calculo da esperanca e variancia BMA de cada yhat+
aux2[1,1]<-sum(pp*aux4[,1])
aux5<-c(aux2[1,1],aux2[1,1])
aux2[1,2]<-sum(pp*aux4[,2]) + sum(pp*(aux4[,1] - aux5)^2)

```

```
aux2
}
```

```
=====
yhatbayes.fun1x<-function(yy,xx,xxplus,nu,lambda,phi){
#funcao que calcula media e variancia da preditice posterior distribution
#para modelos com uma covariavel sem escalonar os dados
#yy: vetor da variavel resposta (entrado como matriz)
#xx: matriz de preditores do modelo em estudo (sem a coluna de uns)
#xxplus: valor futuro de x
#nu, lambda, phi: hiperparametros da regressao
n <- dim(yy)[1]; ones <- rep(1, n); A <- cbind(ones, xx)
#MU ee o vetor das medias dos Betas (priori) (considerado zero a priori)
MU<-as.matrix(c(mean(yy),0))
# V ee a matriz de covariancia dos Betas (priori)
V <- diag(c(var(yy), phi^2*var(xx)))
V1<-solve(V); V2<-t(A)%*%A + V1; V3<-solve(V2)
V4<-t(A)%*%yy + V1%*%MU
lambda2<-(1/(n+nu))*(nu*lambda + t(MU)%*%V1%*%MU + t(yy)%*%yy -
t(V4)%*%V3%*%V4)
xxplus<-matrix(c(1,xxplus),nrow=1,ncol=2)
hat1<-t(xxplus)%*%(V3%*%V4))
hat2<-t(as.matrix(diag(((n+nu)/(n+nu-2))*as.vector(lambda2)*
(xxplus)%*%V3%*%t(xxplus))))
cbind(hat1,hat2 )
}
```

```
=====
yhatbayes1.fun1x<-function(yy,nu,lambda,phi){
#funcao que calcula media e variancia da preditice posterior
#distribution para o modelo nulo (ou seja apenas com b0)
#yy: vetor da variavel resposta (entrado como matriz)
#nu, lambda, phi: hiperparametros da regressao
n <- dim(yy)[1]; A<- as.matrix(rep(1, n))
#MU ee o vetor das medias dos Betas (priori) (considerado zero a priori)
MU<-as.matrix(mean(yy))
# V ee a matriz de covariancia dos Betas (priori)
V<- matrix(var(yy),ncol=1,nrow=1)
V1<- 1/V; V2<-t(A)%*%A + V1; V3<-1/V2
V4<-t(A)%*%yy + V1%*%MU
lambda2<-(1/(n+nu))*(nu*lambda + t(MU)%*%V1%*%MU + t(yy)%*%yy -
t(V4)%*%V3%*%V4)
xxplus<-matrix(1,ncol=1,nrow=1)
hat1<-t(xxplus)%*%(V3%*%V4))
hat2<-t(as.matrix(diag(((n+nu)/(n+nu-2))*as.vector(lambda2)*
(xxplus)%*%V3%*%t(xxplus))))
cbind(hat1,hat2 )
}
```

```
=====
MC3.REG.logpost1x<-function(Y, X, nu, lambda, phi){
#Splus program to compute the log of the PMP for models with 1 covariate
#Refer to the program "MC3.REG.logpost" for further description.
#Inputs:
# Y - the column vector of responses
# X - covariate
# nu, lambda, phi - regression hyperparameters
#Outputs:
```

```

# The log of the posterior for the current model
n <- dim(Y)[1] ; ones <- rep(1, n); A <- cbind(ones, X)
mu<-c(mean(Y),0)
V <- diag(c(var(Y), phi^2*var(X)))
det <- diag(ones) + A %*% V %*% t(A)
divs<- prod(eigen(det, T, T)$values)^0.5
denom <- (t(Y - A%*%mu) %*% solve(det)%*%(Y - A%*%mu)) + (nu * lambda)
lgamma((n + nu)/2) + log(nu * lambda)*(nu/2) - log(pi)*(n/2) -
lgamma(nu/2) - log(divs) - ((nu + n)/2) * log(denom)
}

=====
MC3.REG.logpost1x1<-function(Y, nu, lambda, phi){
#Splus program to compute the log of the PMP for models with 0 covariate
#Refer to the program "MC3.REG.logpost" for further description.
#Inputs:
# Y - the column vector of responses
# nu, lambda, phi - regression hyperparameters
#Outputs:
# The log of the posterior for the current model
n <- dim(Y)[1]; ones <- rep(1, n); A <- as.matrix(ones)
mu<-as.matrix(mean(Y))
V <- as.matrix(var(Y))
det <- diag(ones) + A %*% V %*% t(A)
divs<- prod(eigen(det, T, T)$values)^0.5
denom <- (t(Y - A%*%mu) %*% solve(det)%*%(Y - A%*%mu)) + (nu * lambda)
lgamma((n + nu)/2) + log(nu * lambda)*(nu/2) - log(pi)*(n/2) -
lgamma(nu/2) - log(divs) - ((nu + n)/2) * log(denom)
}

=====
boot1x.fun<-function(y,x,Xsp){
#funcao que calcula a variancia bootstrap para o boot1 e boot2 para
#modelos com uma covariavel
aux2<-matrix(0,ncol=2,nrow=1); library(boot,first=T)
y.df1<-data.frame(y,x)
fit1<-glm(y~1,data=y.df1); y.diag<-glm.diag(fit1)
res1<-(y.diag$res - mean(y.diag$res))*y.diag$sd
fit2<-glm(y~1+x,data=y.df1); y.diag<-glm.diag(fit2)
res2<-(y.diag$res - mean(y.diag$res))*y.diag$sd
y.df2<-data.frame(y,x,y1=y,y2=y,res1=res1,res2=res2,
ft1=fitted(fit1),ft2=fitted(fit2))
xx<-Xsp
aux<-boot(y.df2,boot1x.fun1,R=999)
aux2[1,1]<-var(aux$t[,1]); aux2[1,2]<-var(aux$t[,2])
aux2
}

=====
boot1x.fun2<-function(data2,data1,xx){
n<-length(data1$y)
#ajuste de cada modelo a dados gerados dos pares de obs.
fit1<-lm(y~1,data=data1); fit2<-lm(y~1+x,data=data1)
a1<-n*log(deviance(fit1)/n) +2*1
a2<-n*log(deviance(fit2)/n) +2*2
w1<-exp(-a1/2)/(exp(-a1/2)+exp(-a2/2))
w2<-exp(-a2/2)/(exp(-a1/2)+exp(-a2/2))
theta1<-w1*fit1$coef + w2*(fit2$coef[1]+fit2$coef[2]*xx)

```

```

#ajuste de cada modelo a dados gerados de residuos de cada modelo
fit1<-lm(y1~1,data=data2); fit2<-lm(y2~1+x,data=data2)
a1<-n*log(deviance(fit1)/n) +2*1
a2<-n*log(deviance(fit2)/n) +2*2
w1<-exp(-a1/2)/(exp(-a1/2)+exp(-a2/2))
w2<-exp(-a2/2)/(exp(-a1/2)+exp(-a2/2))
theta2<-w1*fit1$coef + w2*(fit2$coef[1]+fit2$coef[2]*xx)
c(theta1,theta2) }

boot1x.fun1<-function(data,i) {
d<-data
d$y1<-d$ft1 + d$res1[i]; d$y2<-d$ft2 + d$res2[i]
boot1x.fun2(d,data[i,],xx)}

=====

#funcoes para modelo com duas covariaveis e padronizando os dados
options(object.size=8e6)
=====
weight.fun2x<-function(n,p,y,x1,x2,Xsp,zc,XM2){
#funcao que calcula os pesos, a var. buck, o vicio e a variancia com
# sigma desconhecido, para modelos com duas covariaveis
#n: tamanho da amostra;
#y:vetor das respostas(escalonado); x1 e x2: covariaveis (escalonadas)
#Xsp: valor futuro de x;
#zc: matriz contendo as var. z1, z2 e c (10000 de cada)
K<-2^p
aux<-matrix(0,ncol=5,nrow=K); aux2<-matrix(0,ncol=4,nrow=1)
Xs<-cbind(x1,x2); dimnames(Xs)<-list(NULL,c("X1","X2"))
fit1<-lm(y~1)
aux[1,1]<-deviance(fit1)
aux[1,2]<-1
aux[1,3]<-fit1$df.residual
aux[1,4]<-fit1$coef
#variancia de b0
aux[1,5]<-(deviance(fit1)/fit1$df.residual)/n
for (i in 2:K){
x<-Xs[,XM2[i,]]
yx<-data.frame(y,x)
if (i < (p+2)) xplus<-data.frame(x=Xsp[XM2[i,]])
else xplus<-data.frame(X1=Xsp[1],X2=Xsp[2])
fit1<-lm(y~.,data=yx)
pred1<-predict.lm(fit1,newdata=xplus,se.fit=T)
aux[i,1]<-deviance(fit1)
#numero de parametros do modelo (contando com b0)
aux[i,2]<-length(fit1$coef)
aux[i,3]<-fit1$df.residual
#valor predito e seu desvio padrao
aux[i,4]<-pred1$fit
aux[i,5]<-pred1$se.fit^2
}
coef<-fit1$coef
a<-n*log(aux[,1]/n) +2*aux[,2] ; w<-exp(-a/2)/sum(exp(-a/2))
yh<-sum(w*aux[,4])
#varw buck
aux2[1,1]<-((w%*(sqrt(aux[,5] + (aux[,4]-yh)^2)))^2
v<-c((aux[K,1]/aux[K,3])/sum(x1^2),(aux[K,1]/aux[K,3])/sum(x2^2))
d<-c(coef[2]/sqrt(v[1]),coef[3]/sqrt(v[2]))

```



```

u1<-((zc[,1]+d[1])^2)/zc[,3]; u2<-((zc[,2]+d[2])^2)/zc[,3]
U<-cbind(u1,u2)
Q1<-(d[1] + zc[,1])*G2x(u2,U,n)
Q2<-(d[2] + zc[,2])*G2x(u1,U,n)
# bw1 e vw1: vicio e variancia com sigma desconhecido
aux2[1,2]<-Xsp[1]*sqrt(v[1])*(mean(Q1)-d[1]) + Xsp[2]*sqrt(v[2])*(
  mean(Q2)-d[2])
aux2[1,3]<-(aux[K,1]/aux[K,3])/n + Xsp[1]^2*v[1]*var(Q1) +
  Xsp[2]^2*v[2]*var(Q2) + 2*Xsp[1]*Xsp[2]*sqrt(v[1]*v[2])*var(Q1,Q2)
aux2[1,4]<-aux2[3] - 2*Xsp[1]*Xsp[2]*sqrt(v[1]*v[2])*var(Q1,Q2)
aux2
}

G2x<-function(u,U,n){
  (1+exp(1)*(1+u)^(-n/2))/(1 + exp(1)*(1+U[,1])^(-n/2) +
  exp(1)*(1+U[,2])^(-n/2) + exp(2)*(1+U[,1]+U[,2])^(-n/2))}

=====
bma.fun2x<-function(n,y,x1,x2,Xsp,XM2){
  #funcao que calcula a posterior model probability, o estimador, a
  #esperanca e a variancia BMA, para modelos com 2 cov. escalonando
  aux<-matrix(0,ncol=2,nrow=4); aux2<-matrix(0,ncol=2,nrow=1)
  Xs<-cbind(x1,x2); dimnames(Xs)<-list(NULL,c("X1","X2"))
  fit1<-lm(y~1)
  aux[1,1]<-fit1$coef
  aux[1,2]<-MC3.REG.logpost(as.matrix(y),Xs,0,0,NULL,0,nu=2.58,lambda=0.28,
  phi=2.85)
  for(i in 2:4){
    x<-Xs[,XM2[i,]]
    yx<-data.frame(y,x)
    if (i < (p+2)) xplus<-data.frame(x=Xsp[XM2[i,]])
      else xplus<-data.frame(X1=Xsp[1],X2=Xsp[2])
    fit1<-lm(y~.,data=yx)
    pred1<-predict.lm(fit1,newdata=xplus,se.fit=T)
    aux[i,1]<-pred1$fit
    aux[i,2]<-MC3.REG.logpost(as.matrix(y), Xs, XM2[i,], sum(XM2[i,]), NULL, 0,
    nu=2.58, lambda=0.28, phi=2.85)}
  pp<-exp(aux[,2])/sum(exp(aux[,2]))
  aux4<-matrix(0,ncol=2,nrow=4)
  aux4[1,1:2]<-yhatbayes1.fun(as.matrix(y),nu=2.58, lambda=0.28, phi=2.85)
  for(i in 2:4){
    x<-Xs[,XM2[i,]]; xplus<-Xsp[XM2[i,]]
    aux4[i,1:2]<-yhatbayes.fun(as.matrix(y),as.matrix(x),t(as.matrix(xplus)),
    sum(XM2[i,]),
    nu=2.58, lambda=0.28, phi=2.85)
  }
  #calculo da esperanca e variancia BMA de cada yhat+
  aux2[1,1]<-sum(pp*aux4[,1])
  aux5<-rep(aux2[1,1],4)
  aux2[1,2]<-sum(pp*aux4[,2]) + sum(pp*(aux4[,1] - aux5)^2)
  aux2
}

=====

```

APÊNDICE 2

Conjuntos de dados

DADOS DO EXEMPLO 1

planta	Cultivar Carioca			
	1996		1998	
	HDA (x)	produção (y)	HDA (x)	produção (y)
1	62.28344	182.8	165.97570	432.8
2	71.29486	222.4	117.77177	325.2
3	75.01710	206.4	88.61334	205.6
4	50.57304	167.0	103.96401	298.6
5	44.89198	129.0	164.28780	574.4
6	72.50233	235.6	189.05415	580.4
7	75.32795	165.6	132.46552	186.6
8	44.65743	127.2	85.49199	202.6
9	96.55272	362.8	90.64075	240.0
10	57.50205	248.4	205.89086	628.0
11	73.59484	233.0	134.41990	359.6
12	72.79220	222.2	171.92430	527.6
13	60.81360	104.4	25.17353	0.0
14	44.35783	144.2	90.42368	31.4
15	31.47039	13.2	39.52577	0.0
16	54.68722	158.4	64.84983	8.6
17	64.26305	272.8	49.67627	31.6
18	85.79504	235.2	27.13072	12.0
19	60.95003	142.0	42.31018	0.0
20	57.55370	127.8	40.44843	11.4
21	28.86643	19.8	96.31012	53.8
22	48.12824	143.6	67.84115	19.8
23	65.85987	237.8	85.39750	8.2
24	50.46424	172.8	84.96226	16.6
25	52.34580	220.0	90.84430	17.4
26	70.97219	303.8	93.05982	46.6
27	46.23859	158.0	77.88562	54.8
28	50.00018	132.2	45.73644	9.6
29	52.64035	175.0	38.20497	3.0
30	57.26229	211.4	62.23485	28.6
31	33.76885	71.8	24.31876	0.0
32	79.08407	212.8	36.41527	21.8
33	46.49055	162.2	56.73585	32.4
34	30.07167	106.0	73.20585	71.4
35	49.47947	188.4	32.01439	0.4
36	35.01779	33.6	66.86218	14.0
37	55.07975	183.8	59.97460	73.8
38	68.71914	171.0	72.82077	51.8
39	87.53766	214.8	34.91190	11.8
40	42.63445	24.0	54.67723	28.4
41	77.24690	229.4	55.21344	12.2
42	85.96758	323.2	89.13840	7.4
43	68.72971	227.6	75.88872	16.0
44	51.76310	62.0	86.67220	45.0
45	43.74570	106.4	63.99704	22.6
46	67.33340	216.4	56.85665	8.8
47	60.61112	180.6	72.80906	2.8
48			85.61449	16.2

Fonte: CARNEIRO, S. M. T. P. G. Efeito da infecção conjunta de *Phaeoisariopsis griseolae* *Colletotrichum lindemuthianum* nos componentes de produção e nas variáveis de área foliar do feijoeiro. Piracicaba, 1999. 89p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.

DADOS DO EXEMPLO 1 (continuação)

planta	Cultivar Iapar-57			
	1996		1998	
	HDA (x)	produção (y)	HDA (x)	produção (y)
1	77.41144	221.8	118.776237	402.4
2	64.71351	162.0	200.100285	663.6
3	56.11975	205.6	200.823811	367.2
4	54.25853	168.6	170.083989	434.2
5	70.98862	231.4	158.494440	304.2
6	53.31411	107.0	128.553280	233.0
7	52.60751	154.4	225.367984	393.8
8	45.10528	125.6	199.021156	553.0
9	23.39282	41.6	178.678577	503.6
10	68.42072	207.0	180.443421	334.2
11	44.97979	132.8	218.948645	631.6
12	30.14567	59.8	19.854011	19.6
13	57.37962	110.0	18.387978	0.0
14	40.26148	113.6	39.033333	20.8
15	60.94629	104.2	44.058094	51.0
16	48.45775	111.6	10.761542	0.0
17	74.48482	225.6	31.442035	26.4
18	44.65173	86.0	20.708696	0.0
19	27.08103	35.8	29.698139	3.0
20	73.93426	170.4	56.924184	100.0
21	51.48275	79.8	15.524443	0.0
22	48.98530	138.8	9.698739	0.0
23	39.43555	130.8	16.087410	3.6
24	32.07299	88.0	37.590081	34.0
25	45.16823	193.2	91.049506	141.0
26	78.32821	234.4	56.402728	0.0
27	32.09437	77.2	28.805119	0.4
28	71.18227	198.0	31.985071	31.2
29	58.77839	190.4	35.870283	1.0
30	51.67246	209.0	18.882361	0.0
31	79.92522	228.0	39.293434	0.8
32	37.96867	92.0	121.775772	330.2
33	30.93311	112.8	40.890089	34.4
34	44.12345	150.0	36.867811	59.8
35	42.33336	133.0	45.034731	39.0
36	25.00188	31.0	88.823785	184.0
37	53.99080	143.0	72.398250	118.6
38	57.85782	195.2	102.079272	323.6
39	20.81355	46.4	48.855192	72.4
40	81.90295	227.4	21.407977	15.4
41	74.03815	304.0	45.499557	66.6
42	65.94940	200.0	41.383758	36.6
43	62.10889	190.8	16.768704	0.0
44	54.84199	135.2	32.057892	52.8
45	44.93056	160.0	24.455303	0.8
46	80.13001	225.0	30.279019	0.0
47	50.80276	140.8	34.805903	5.4
48	54.36784	152.2		

Fonte: CARNEIRO, S. M. T. P. G. Efeito da infecção conjunta de *Phaeoisariopsis griseolae* *Colletotrichum lindemuthianum* nos componentes de produção e nas variáveis de área foliar do feijoeiro. Piracicaba, 1999. 89p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.

DADOS DO EXEMPLO 2

Amostra	fósforo inorgânico (x_1)	fósforo orgânico (x_2)	fósforo na planta (y)
1	0,4	53	64
2	0,4	23	60
3	3,1	19	71
4	0,6	34	61
5	4,7	24	54
6	1,7	65	77
7	9,4	44	81
8	10,1	31	93
9	11,6	29	93
10	12,6	58	51
11	10,9	37	76
12	23,1	46	96
13	23,1	50	77
14	21,6	44	93
15	23,1	56	95
16	1,9	36	54
17	26,8	58	168
18	29,9	51	99

Fonte: SNEDECOR, G.W & COCHRAN, W.G. (1967). *Statistical Methods*. The Iowa State Press University. pág. 384.

DADOS DO EXEMPLO 3

Amostra	diâmetro (x_1)	altura (x_2)	volume (y)
1	8,3	70	10,3
2	8,6	65	10,3
3	8,8	63	10,2
4	10,5	72	16,4
5	10,7	81	18,8
6	10,8	83	19,7
7	11,0	66	15,6
8	11,0	75	18,2
9	11,1	80	22,6
10	11,2	75	19,9
11	11,3	79	24,2
12	11,4	76	21,0
13	11,4	76	21,4
14	11,7	69	21,3
15	12,0	75	19,1
16	12,9	74	22,2
17	12,9	85	33,8
18	13,3	86	27,4
19	13,7	71	25,7
20	13,8	64	24,9
21	14,0	78	34,5
22	14,2	80	31,7
23	14,5	74	36,3
24	16,0	72	38,3
25	16,3	77	42,6
26	17,3	81	55,4
27	17,5	82	55,7
28	17,9	80	58,3
29	18,0	80	51,5
30	18,0	80	51,0
31	20,6	87	77,0

Fonte: RYAN, B.F.; JOINER, B.L.; RYAN Jr., T.A. (1976). *Minitab Student Handbook*. Duxbury Press. pág. 329.

DADOS DO EXEMPLO 4

cidade	latitude (x_1)	longitude (x_2)	altitude (x_3)	temperatura do ar (y)
1	23.0	22.90	47.08	674
2	23.2	24.03	48.37	702
3	24.8	21.13	48.97	536
4	23.9	22.30	49.55	522
5	21.5	23.95	48.88	647.18
6	19.2	24.95	49.25	1150
7	23.9	22.28	48.57	580
8	23.8	23.20	46.88	715
9	22.6	22.53	47.45	639
10	23.0	23.17	49.33	589
11	23.5	21.47	47.02	665
12	22.2	22.68	46.72	777
13	24.0	22.12	48.00	528
14	23.5	22.92	45.00	560
15	24.1	21.22	48.93	562
16	23.6	21.18	47.80	621
17	21.6	22.02	47.88	856
18	22.1	23.50	46.62	792.06
19	22.5	23.53	47.13	850
20	23.8	21.48	47.55	617.39
21	23.6	23.33	47.87	600
22	23.5	23.12	47.72	538

Fonte: NORMAS CLIMATOLÓGICAS 1961-1990 DO DNMET e do INSTITUTO AGRONÔMICO.