

PONDERAÇÃO BAYESIANA DE MODELOS EM REGRESSÃO LINEAR CLÁSSICA

HÉLIO RUBENS DE CARVALHO NUNES

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

P I R A C I C A B A
Estado de São Paulo - Brasil
Agosto - 2005

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

PONDERAÇÃO BAYESIANA DE MODELOS EM REGRESSÃO LINEAR CLÁSSICA

HÉLIO RUBENS DE CARVALHO NUNES

Bacharel em Estatística

Orientadora: Prof^ª Dr^ª ROSELI APARECIDA LEANDRO

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

P I R A C I C A B A
Estado de São Paulo - Brasil
Agosto - 2005

Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP

Nunes, Hélio Rubens de Carvalho
Ponderação bayesiana de modelos em regressão linear clássica / Hélio Rubens de
Carvalho Nunes. - - Piracicaba, 2005.
54 p.

Dissertação (Mestrado) - - Escola Superior de Agricultura Luiz de Queiroz, 2005.
Bibliografia.

1. Análise de regressão 2. Inferência bayesiana I. Título

CDD 519.536

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

DEDICATÓRIA

Aos meus pais,

Waldir Nunes e Maria Magdalena de Carvalho Nunes.

Todos esses anos dedicados à minha educação e formação certamente mereciam muito mais que essas humildes palavras. Nunca me esquecerei das orientações e conselhos nos momentos difíceis, dos risos compartilhados em ocasiões festivas e principalmente do amor que cada um a seu modo sente por mim. Que este pequeno trecho consiga transmitir o amor que sinto por vocês!

À minha esposa,

Paula Roberta Mendes.

A amizade, o respeito e a compreensão sempre foram marcantes em nosso relacionamento.

AGRADECIMENTOS

À Prof^ª Dr^ª **Roseli Aparecida Leandro** pela orientação. Sua conduta sempre firme e sincera me fez enxergar os requisitos que a vida acadêmica exige de um profissional. Foram dois anos e meio de muito aprendizado e observação. Vou sentir saudades!

À Prof^ª Dr^ª **Clarice Garcia Borges Demétrio** pela capacidade de liderança, vocação para o ensino e por saber aconselhar e cobrar nos momentos mais adequados. Buscarei sempre a objetividade!

Ao Prof^º Dr^º **Edwin Moisés Marcos Ortega** pela amizade e apoio.

À Prof^ª Dr^ª **Cecília Candolo** da Universidade Federal de São Carlos pela troca de informação acerca do tema deste trabalho.

Ao Prof^º Dr^º **Sílvio Sandoval Zocchi**. Por ter me apoiado em períodos complicados.

Ao pesquisador **Bill Venables**, do CSIRO, Cleveland, Austrália.

Ao pesquisador **Antonio de Oliveira Lobão**, do CESAHO, Piracicaba.

À **Solange de Assis Paes Sabadin** pela disposição sincera e pela eficiência com que cuidou dos assuntos extra-classe.

À **Luciane Brajão** pelo atendimento sempre simpático.

A **Jorge Alexandre Wiendl** pelo apoio técnico.

Aos amigos do mestrado: Alfredo, Édila, Elisabeth, Francine, Juliana, Ana, Júlio, Lúcio, Melissa, Moita, Sandra e William.

Aos amigos do doutorado: Afrânio, Ana, Andréia, David, Denise, Elisabeth, Fogo, Giovana, Genevile, Idemauro, João Maurício, Luciana, Luciano, Osmar e Pedro.

Aos professores da DEs/UFSCar.

Aos membros dos grupos “GDP-1” e “Agronomia” do CESAHO.

Ao CNPQ e à CAPES pelo apoio financeiro.

SUMÁRIO

	Página
LISTA DE FIGURAS	vi
LISTA DE TABELAS	vii
RESUMO	viii
SUMMARY	x
1 INTRODUÇÃO	1
2 REVISÃO DE LITERATURA	3
2.1 A seleção de modelos em regressão linear clássica	3
2.2 A incerteza envolvida no processo de seleção de modelos	7
2.3 A ponderação de modelos	8
2.3.1 A ponderação Bayesiana de modelos	10
2.3.2 Desempenho preditivo da ponderação Bayesiana de modelos	12
3 MATERIAL E MÉTODOS	14
3.1 Material	14
3.1.1 Para a aplicação da metodologia BMA	15
3.1.2 Para a avaliação do desempenho preditivo do BMA	19
3.2 Métodos	20
3.2.1 Ponderação Bayesiana de modelos de regressão linear clássica	20
4 RESULTADOS E DISCUSSÃO	29
4.1 Estudo de Simulação	36
5 CONCLUSÕES	41
REFERÊNCIAS BIBLIOGRÁFICAS	42
ANEXOS	46

LISTA DE FIGURAS

	Página
1	Processo de seleção de modelos 7
2	Distribuições a <i>priori</i> para σ^2 segundo valores de ν e λ 24
3	Boxplot para o número de condição das amostras simuladas para cada combinação entre cultivar e tamanho amostral 37
4	Comportamento da média do Logaritmo do Escore Preditivo da metodologia BMA (vermelho) e dos critérios: R^2 (preto), \bar{R}^2 (preto tracejado), $PRESS$ (azul) e C_p (azul Tracejado) 40

LISTA DE TABELAS

	Página
1	Dados referentes ao cultivar C-111, Paraná, 1976 16
2	Dados referentes ao cultivar Piranão, Paraná, 1976 17
3	Dados referentes ao cultivar Flint Composto, Paraná, 1976 18
4	Ilustração de um específico conjunto de dados simulado 20
5	Opções para os valores dos hiperparâmetros 22
6	Estimativas de mínimos quadrados dos coeficientes de regressão dos modelos selecionados através do critério \bar{R}^2 - Cultivar C-111 30
7	Estimativas de mínimos quadrados dos coeficientes de regressão dos modelos selecionados através do critério \bar{R}^2 - Cultivar Piranão 31
8	Estimativas de mínimos quadrados dos coeficientes de regressão dos modelos selecionados através do critério \bar{R}^2 - Cultivar Flint Composto 31
9	Análises clássica e Bayesiana dos dados do cultivar Flint Composto 33
10	Modelos selecionados pelo <i>Occam's Window simétrico</i> , executado com diferentes valores de a 34
11	Esperança e variância <i>a posteriori</i> de cada coeficiente obtida pela ponderação dos modelos selecionados pelo OW para cada valor de a 35
12	Resumo descritivo para o número de condição (k^*) das amostras geradas para cada combinação tamanho amostral e cultivar 37
13	Resumo dos valores do LEP nas amostras simuladas (cultivar C-111) 38
14	Resumo dos valores do LEP nas amostras simuladas (cultivar Piranão) 39
15	Resumo dos valores do LEP nas amostras simuladas (cultivar Flint Composto) 39

PONDERAÇÃO BAYESIANA DE MODELOS EM REGRESSÃO LINEAR CLÁSSICA

Autor: HÉLIO RUBENS DE CARVALHO NUNES

Orientadora: Prof^ª Dr^ª ROSELI APARECIDA LEANDRO

RESUMO

Este trabalho tem o objetivo de divulgar a metodologia de ponderação de modelos ou *Bayesian Model Averaging* (BMA) entre os pesquisadores da área agrônômica e discutir suas vantagens e limitações. Com o BMA é possível combinar resultados de diferentes modelos acerca de determinada quantidade de interesse, com isso, o BMA apresenta-se como sendo uma metodologia alternativa de análise de dados frente os usuais métodos de seleção de modelos tais como o Coeficiente de Determinação Múltipla (R^2), Coeficiente de Determinação Múltipla Ajustado (\bar{R}^2), Estatística de Mallows (C_p) e Soma de Quadrados de Predição (*PRESS*). Vários trabalhos foram, recentemente, realizados com o objetivo de comparar o desempenho do BMA em relação aos métodos de seleção de modelos, porém, há ainda muitas situações para serem exploradas até que se possa chegar a uma conclusão geral acerca desta metodologia. Neste trabalho, o BMA foi aplicado a um conjunto de dados proveniente de um experimento agrônômico. A seguir, o desempenho preditivo do BMA foi comparado com o desempenho dos métodos de seleção acima citados por meio de um estudo de simulação variando o grau de multicolinearidade e o tamanho amostral. Em cada uma dessas situações, foram utilizadas 1000 amostras geradas a partir de medidas descritivas de conjuntos de dados

reais da área agronômica. O desempenho preditivo das metodologias em comparação foi medido pelo Logaritmo do Escore Preditivo (LEP). Os resultados empíricos obtidos indicaram que o BMA apresenta desempenho semelhante aos métodos usuais de seleção de modelos nas situações de multicolinearidade exploradas neste trabalho.

BAYESIAN MODEL AVERAGING IN CLASSIC LINEAR REGRESSION MODELS

Author: HÉLIO RUBENS DE CARVALHO NUNES

Adviser: Prof^ª Dr^ª ROSELI APARECIDA LEANDRO

SUMMARY

The objective of this work was divulge to Bayesian Model Averaging (BMA) between the researchers of the agronomy area and discuss its advantages and limitations. With the BMA is possible combine results of different models about determined quantity of interest, with that, the BMA presents as being a methodology alternative of data analysis front the usual models selection approaches, for example the Coefficient of Multiple Determination (R^2), Coefficient of Multiple Determination Adjusted (\bar{R}^2), Mallows (C_p Statistics) and Prediction Error Sum Squares ($PRESS$). Several works recently were carried out with the objective of compare the performance of the BMA regarding the approaches of models selection, however, there is still many situations for will be exploited to that can arrive to a general conclusion about this methodology. In this work, the BMA was applied to data originating from an agronomy experiment. It follow, the predictive performance of the BMA was compared with the performance of the approaches of selection above cited by means of a study of simulation varying the degree of multicollinearity, measured by the number of condition of the matrix standardized $\mathbf{X}'\mathbf{X}$ and the number of observations in the sample. In each one of those situations, were utilized 1000 samples generated from the descriptive information of agronomy data . The predictive performance of the methodologies in comparison was

measured by the Logarithm of the Score Predictive (LEP). The empirical results obtained indicated that the BMA presents similar performance to the usual approaches of selection of models in the situations of multicollinearity exploited.

1 INTRODUÇÃO

Métodos de seleção de modelos (ou variáveis) são freqüentemente utilizados em trabalhos que envolvem modelagem estatística porque geralmente, nesses trabalhos, há uma disposição natural por parte do pesquisador em observar uma grande quantidade de variáveis na expectativa de que ao serem incluídas em um modelo estatístico, este ajuste-se aos dados amostrais da melhor forma possível.

Porém, sem um bom conhecimento teórico acerca do fenômeno em estudo, a observação de muitas variáveis pode trazer o problema da multicolinearidade, ou seja, algumas das variáveis observadas podem ser altamente correlacionadas. A multicolinearidade faz com que algumas das variáveis sejam irrelevantes para a explicação do fenômeno em estudo, além disso, a inclusão dessas variáveis em um modelo aumenta a variância dos estimadores de mínimos quadrados de seus coeficientes de regressão (Rao, 1971). Além dos problemas de estimação, modelos superparametrizados podem ser inconvenientes do ponto de vista prático para seu usuário que pode ser: um pesquisador; um instituto de pesquisa ou uma empresa. Por outro lado, a omissão de variáveis importantes em um modelo introduz vício nos mesmos estimadores de mínimos quadrados, além disso, modelos subparametrizados podem não representar adequadamente o fenômeno em estudo.

Por essas e outras razões é que os métodos de seleção de modelos são tão utilizados. Espera-se que o modelo selecionado contorne os problemas trazidos pela sub ou superparametrização da melhor maneira possível.

Entretanto, a utilização de métodos de seleção normalmente condiciona as conclusões e tomadas de decisão do pesquisador a um único modelo, descartando com isso a possibilidade de outros modelos que se ajustam razoavelmente bem aos dados serem levados em consideração. Em algumas situações talvez existam razões científicas que justifiquem a adoção de um único modelo, mas em muitas outras, como por exemplo em regressão linear múltipla com muitas variáveis, muitos modelos podem se ajustar igualmente bem aos dados,

explicando-os de forma plausível e fornecendo estimativas substancialmente distintas de um mesmo coeficiente de regressão.

Situações como esta pode provocar um certo desconforto e uma sensação de incerteza no(s) responsável(eis) pela escolha do modelo, pois pequenas diferenças nas estimativas de um coeficiente podem representar, na realidade, grandes diferenças orçamentárias ou mudanças significativas das estratégias de um instituto de pesquisa/empresa privada.

A incerteza diante da escolha de um modelo pode ser contornada por meio de uma combinação ponderada entre todos os modelos propostos ao ajuste dos dados ou apenas entre os melhores modelos, indicados com o auxílio de algum critério. Esta combinação pode ser feita seguindo os princípios da estatística clássica ou Bayesiana. Sob os princípios da estatística Bayesiana esta combinação é conhecida por ponderção Bayesiana de modelos ou *Bayesian Model Averaging* (BMA) (Leamer, 1978). O BMA é essencialmente a sistematização do conceito de se concluir e tomar decisões com base em uma combinação de resultados obtidos sob diferentes modelos.

A partir da década de 1990, vários trabalhos acerca do BMA foram realizados com o objetivo de compará-lo com os métodos de seleção de modelos, no entanto, há ainda muitas situações para serem exploradas até que se possa chegar a uma conclusão geral acerca desta metodologia.

O BMA é aplicável em diversas classes de modelos. Atualmente, apenas em regressão linear clássica os seus termos podem ser obtidos de forma analítica exata sem a necessidade de recorrer a métodos de aproximação. Além da combinação de informações acerca de coeficientes de regressão e de predições da variável resposta, o BMA também pode ser utilizado para contornar a incerteza diante da escolha da transformação mais adequada a um conjunto de dados ou para identificação de *outliers*.

Muitas sugestões de pesquisa acerca do BMA foram propostas em trabalhos recentes, como por exemplo o estudo das vantagens em implementá-lo para ponderar modelos pertencentes a diferentes classes e o desenvolvimento de formas analíticas exatas dos seus termos.

O presente trabalho apresenta a metodologia BMA e tem por objetivo divulgá-la entre os pesquisadores da área agrônômica. Através da sua aplicação, pretende-se discutir suas vantagens e desvantagens em relação aos procedimentos de seleção de modelos normalmente utilizados e avaliar seu desempenho preditivo na classe dos modelos de regressão linear clássica.

2 REVISÃO DE LITERATURA

Nesta seção, apresenta-se um breve histórico acerca da seleção de modelos em regressão linear clássica e uma das possíveis formas de caracterização da incerteza diante da escolha de um modelo.

2.1 A seleção de modelos em regressão linear clássica

A análise de regressão linear é uma técnica estatística muito utilizada em várias áreas de pesquisa. Com ela o pesquisador pode investigar a possível existência de uma relação entre as variáveis observadas durante seu estudo. Esta relação é representada, matematicamente, por uma equação (ou modelo) entre a variável resposta Y (a de maior interesse no estudo) e p variáveis X_1, \dots, X_p que supostamente explicam a variabilidade de Y . A forma desta equação para cada uma das n observações da amostra é:

$$Y_i = \beta_0 + \sum_{j=1}^p \beta_j X_{ij} + \varepsilon_i, \quad i = 1, \dots, n, \quad (1)$$

podendo $\beta_0, \beta_1, \dots, \beta_p$ serem estimados pelo método dos mínimos quadrados. O termo ε_i representa a influência do acaso sobre o valor de Y_i .

O conhecimento teórico do pesquisador sobre o fenômeno em estudo e as investigações exploratórias sobre os dados amostrais podem sugerir que sejam acrescentadas na expressão (1) algumas interações ou funções das variáveis explicativas.

Sempre que possível o pesquisador tende a observar muitas variáveis explicativas na expectativa de conseguir controlar grande parte da variabilidade de Y , entretanto, a inclusão de todas essas variáveis no modelo geralmente contraria o princípio da parcimônia¹, além de superestimar (*overfitting*) o modelo. Por essa razão, recomenda-se que sejam incluídas no modelo apenas as variáveis mais relevantes.

¹Criada por William de Ockham por volta de 1315, o princípio da parcimônia diz que, dentre várias hipóteses que explicam um acontecimento, a mais plausível é quase sempre a mais simples.

Segundo Neter et al. (1983), a seleção de variáveis (ou de modelos) é uma das questões mais importantes de uma análise de regressão linear. Neste trabalho esta questão é discutida levando-se em conta apenas modelos cuja forma é dada por (1), restringindo a discussão a um conjunto de exatamente $\sum_{r=0}^p \binom{p}{r} = 2^p$ modelos, dos quais, $\binom{p}{r}$ possuem r das p variáveis observadas.

O processo de seleção de modelos é uma das partes que compõe toda a questão da seleção de modelos. Este processo pode ser realizado de duas formas: ajustando todos os possíveis modelos e comparando-os por meio de um critério pré-estabelecido ou automaticamente através de métodos do tipo *Stepwise*, por exemplo. No primeiro caso, os ajustes de cada modelo são avaliados, e o modelo que apresentar o melhor desempenho conforme o critério estabelecido é então escolhido para representar os dados amostrais. No segundo caso, o processo de seleção é executado por passos, sendo que, em cada passo, uma variável explicativa pode ser incluída ou excluída do conjunto das mais relevantes de acordo com o resultado de uma estatística pré-determinada. No fim da execução do método, as variáveis consideradas mais relevantes deverão compor um modelo que, conseqüentemente, será considerado o melhor. Tais métodos baseiam-se nos princípios conhecidos por Seleção *Forward* e Eliminação *Backward*, como por exemplo, a Regressão *Stepwise* (Efroymson, 1960). Portanto, o processo de seleção é o meio operacional pelo qual o modelo que melhor representa as observações amostrais é identificado. O desempenho e a escolha deste processo pode ser influenciado de forma direta ou indireta por vários fatores, entre eles os interesses do pesquisador pelo modelo, o enfoque adotado na análise (Bayesiano ou freqüentista), as possíveis falhas de especificação do modelo (*misspecification*) etc.

Vários critérios são freqüentemente utilizados nos processos de seleção de modelos de regressão linear, sendo a estatística da Razão de Verossimilhança o principal deles, no entanto, quando modelos aninhados são comparados e quando o comportamento assintótico é válido, a Razão de Verossimilhança torna-se inconsistente, tendendo a favorecer o modelo completo (Gelfand & Ghosh, 1998). Para evitar tal favorecimento, foram criados critérios que penalizam os modelos que possuem muitas variáveis. Alguns desses critérios são: a *Deviance* (Nelder & Wedderburn, 1972); o critério de informação de Akaike (Akaike, 1973) e o critério de informação de Bayes (Schwarz, 1978). Há, também, critérios diretamente baseados na soma de quadrados residual (SQRes), sendo que os mais utilizados na estatística aplicada, conforme Hocking (1976) são:

O quadrado médio residual,

$$QMRes_p = \frac{SQRes_p}{n - p}.$$

O coeficiente de determinação,

$$R_p^2 = 1 - \frac{SQRes_p}{SQT_{total}}.$$

O coeficiente de determinação ajustado,

$$\bar{R}_p^2 = 1 - \frac{(n - 1)}{(n - p)}(1 - R_p^2).$$

A variância de predição média,

$$J_p = \frac{(n + p)}{n} QMRes_p.$$

O erro quadrático total,

$$C_p = \frac{SQRes_p}{\hat{\sigma}^2} + 2p - n.$$

A soma de quadrados de predição,

$$PRESS_p = \mathbf{e}_p' \mathbf{D}_p^{-1} \mathbf{e}_p,$$

sendo $\mathbf{e}_p = \mathbf{y} - \hat{\mathbf{y}}_p$ e $\mathbf{D}_p = \text{diag}(\mathbf{I} - \mathbf{X}_p(\mathbf{X}_p' \mathbf{X}_p)^{-1} \mathbf{X}_p')$.

Alguns desses critérios são mais adequados para determinados objetivos e não há uma regra definida sobre qual deles utilizar em cada situação. Uma ampla revisão sobre o assunto pode ser encontrada em Hocking (1976). Segundo o autor, tanto o R_p^2 como o \bar{R}_p^2 são muito utilizados quando o pesquisador quer descrever os dados amostrais resumindo as informações de um banco de dados por meio de um modelo. Quando o objetivo é a predição de um valor futuro de Y ou do seu valor médio, Wetherill et al. (1986) sugerem os critérios C_p e $PRESS_p$. No entanto, se o objetivo for o controle de um determinado processo (situações nas quais busca-se determinar o intervalo de variação de uma variável explicativa que leve a desejáveis mudanças em Y), Chatterjee & Price (1977) sugerem modelos que apresentam pequenas estimativas dos erros padrão associadas aos estimadores dos coeficientes inclusos na equação do modelo.

Historicamente, o processo de seleção de modelos foi pouco discutido até o fim da década de 1940, quando surgiram as primeiras calculadoras manuais. Na década de

1950, trabalhos como os de Snedecor (1956) e Lyle (1957) trataram do assunto considerando situações onde apenas três ou quatro variáveis explicativas são observadas.

Somente a partir da utilização dos computadores, no final dessa mesma década, é que o processo de seleção começou a se desenvolver de forma efetiva. Foram criados algoritmos capazes de ajustar automaticamente grandes quantidades de modelos, como por exemplo, os algoritmos de Garside (1965), Newton & Spurrel (1967) e de Furnival (1971). Todos esses algoritmos geram todas as 2^p combinações de modelos baseados no “Operador Sweep” (Beaton, 1964), exceto o algoritmo de Newton & Spurrel (1967), que utiliza “elementos básicos” presentes na soma de quadrados de regressão de todos os modelos.

Assim como esses algoritmos, os processos de seleção do tipo *Stepwise* foram criados com o objetivo de permitir que o pesquisador identifique o melhor modelo sem que para isso tenha de ajustar todos os 2^p modelos.

Uma das vantagens desses algoritmos em relação aos critérios de seleção é a redução dos cálculos necessários para a obtenção de um modelo considerado o melhor, no entanto, como apenas uma parcela da totalidade dos 2^p modelos são avaliados, não se pode afirmar que o modelo obtido seja realmente o melhor dentre todos, pois no decorrer da execução, a procura pode seguir por uma direção na qual a melhor combinação de variáveis não foi considerada. Para contornar este problema, foram criados os algoritmos de procura otimizada, como por exemplo, o “Leaps and Bounds” (Furnival & Wilson, 1974). Atualmente vários desses algoritmos encontram-se implementados em *softwares* tais como o *SAS* (*Statistical Analysis System*), *S – Plus*, *Statistica*, *R* entre outros.

Do ponto de vista Bayesiano, o processo de seleção de modelos é conceitualmente simples (Laud & Ibrahim, 1996 e Lee III, 1999). Seja M_k , $k = 1, \dots, 2^p$ um modelo com probabilidade *a priori* $P(M_k)$ e vetor de parâmetros $\boldsymbol{\theta}_k$ com distribuição *a priori* $\pi(\boldsymbol{\theta}_k|M_k)$ e função de verossimilhança $L(\boldsymbol{\theta}_k; \mathbf{y})$. A função de verossimilhança marginal correspondente ao modelo M_k , é dada por:

$$L_k(\mathbf{y}) = \int \dots \int L(\boldsymbol{\theta}_k; \mathbf{y}) \pi(\boldsymbol{\theta}_k|M_k) d\boldsymbol{\theta}_k. \quad (2)$$

Aplicando o teorema de Bayes, pode-se obter a probabilidade de M_k , condicionada a ocorrência da amostra \mathbf{y} :

$$P(M_k|\mathbf{y}) = \frac{L_k(\mathbf{y})P(M_k)}{\sum_{l=1}^{2^p} L_l(\mathbf{y})P(M_l)}, \quad k = 1, \dots, 2^p. \quad (3)$$

Com isso, basta selecionar o modelo que tiver a maior probabilidade condi-

cionada dada por (3).

2.2 A incerteza envolvida no processo de seleção de modelos

Segundo Tversky (1974), a incerteza é um elemento essencial da condição humana. Com base nesta afirmação, pode-se dizer que a incerteza está presente em quase todas as situações nas quais decisões precisam ser tomadas, como por exemplo nos processos de seleção de modelos. Nesse caso, a incerteza se caracteriza e mostra-se relevante quando em cada modelo são observadas diferentes estimativas e erros padrão de um mesmo coeficiente de regressão. Mesmo que resultados empíricos não indiquem diferenças estatisticamente significativas, as mesmas podem representar na prática diferenças significativas de orçamento, de políticas a serem adotadas etc.

Um processo de seleção de modelos pode ser visto como uma metodologia que tem por objetivo identificar modelos que combinem de forma satisfatória dois fatores: **qualidade de ajuste aos dados** e **conseqüências gerenciais causadas pela sua escolha**. O primeiro destes fatores é medido, numericamente, conforme o critério de seleção pré-estabelecido e por isso contribui menos para a geração da incerteza diante da escolha de um modelo do que o segundo fator, que tem caráter mais subjetivo, já que as regras para se decidir por um modelo muitas vezes não são tão claras. A incerteza envolvida nos processos de seleção se deve mais às razões subjetivas de uma tomada de decisão.

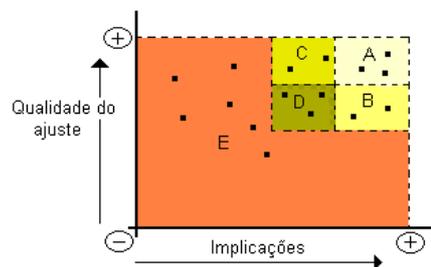


Figura 1 - Processo de seleção de modelos

Como ilustração, suponha que, em um experimento, quatro variáveis explicativas sejam observadas, possibilitando o ajuste de 2^4 modelos e que o resultado da combinação dos fatores acima citados em cada modelo esteja representado pelos pontos da Figura 1.

Verifica-se que os modelos da região A ajustam-se muito bem aos dados e, além disso, as conseqüências práticas, econômicas e políticas devido a escolha de um modelo pertencente a essa região serão as melhores possíveis. Já os modelos da região B apresentam um desempenho inferior em termos do critério de seleção utilizado, entretanto, as conseqüências da escolha de um modelo dessa região são as mesmas observadas na região A. Nas regiões C e D estão os modelos caros, de difícil manutenção e causadores de mudanças mais significativas na política e nas estratégias de uma empresa/instituto de pesquisa. Finalmente, na região E, encontram-se os modelos que não atingem níveis desejáveis em pelo menos um dos fatores, sendo portanto, desaconselhável a adoção de qualquer um desses para a representação dos dados amostrais.

As áreas A, B, C e D são regiões de incerteza nas quais o pesquisador não consegue identificar, dentro de cada uma delas, o modelo que produz a melhor combinação entre os fatores já citados, ou seja, há diferenças significativas entre as regiões, o que não ocorre entre os modelos dentro das mesmas. Essas regiões podem, dependendo da situação, reagruparem-se, formando novas regiões ou até mesmo uma única região de incerteza diante da escolha convicta de um modelo que possa ser considerado o gerador dos dados amostrais, pois os demais modelos, mesmo com suas irrelevantes diferenças, podem fornecer informações substancialmente distintas sobre predição e sobre o efeito de uma determinada variável explicativa.

2.3 A ponderação de modelos

A idéia de se concluir e tomar decisões com base numa combinação de opiniões ou modelos foi inicialmente proposta na década de 1960. Essa idéia foi sistematizada por Leamer (1978) e só a partir da década de 1990 a ponderação de modelos passou a ser mais estudada e aplicada em diversas situações. Assim como muitas outras metodologias estatísticas, a ponderação de modelos pode ser feita sob os enfoques freqüentista e Bayesiano. Em ambos os contextos há o consenso de que a incerteza diante da escolha de um modelo deve ser totalmente incorporada no processo de análise dos dados sempre que a estimação dos parâmetros for sensível a escolha do modelo (Leamer, 1978, p. 91-92 e Buckland et al., 1997). Na concepção de Leamer (1978), a sensibilidade na estimação dos parâmetros pode ser detectada quando diferentes modelos fornecem diferentes estimativas para um mesmo coeficiente

de regressão.

No contexto freqüentista, Buckland et al. (1997) apresentam uma forma de ponderar modelos. Os pesos de cada modelo podem ser obtidos de três maneiras: pelo critério de informação AIC (Akaike, 1973); pelo critério de informação BIC (Schwarz, 1978) e pelo método *bootstrap*. Porém, os autores alertam para o fato de que em regressão linear, a utilização do *bootstrap* pode causar problemas se as amostras *bootstrap* forem obtidas a partir dos resíduos resultantes do ajuste de um modelo fixado. Os autores defendem sua metodologia argumentando que os resultados obtidos através da aplicação da ponderação Bayesiana podem ser sensíveis a escolha da distribuição *a priori* para os parâmetros dos modelos. Entretanto, Hoeting (1994) defende a ponderação Bayesiana através de um estudo de sensibilidade, no qual adota três conjuntos de valores para os parâmetros da distribuição *a priori* conjugada normal-gama atribuída aos parâmetros do modelo, e mostra que os resultados empíricos, condicionados a cada *priori*, não se diferenciam de forma significativa. Conclusões semelhantes foram obtidas por Raftery et al. (1997) utilizando os mesmos conjuntos de valores.

Candolo (2001) estuda a ponderação de modelos proposta por Buckland em situações de regressão linear simples e múltipla e a compara com o procedimento padrão de seleção de um modelo e com a metodologia de ponderação Bayesiana em termos das propriedades dos estimadores obtidos por cada uma dessas metodologias. Seus resultados mostram que o erro médio quadrático da predição da média da variável resposta Y pode ser substancialmente reduzido pelo uso do estimador ponderado de Buckland nas situações em que os testes para os coeficientes de regressão são pouco significativos. Já nas situações em que os testes para os coeficientes de regressão são altamente significativos, a aplicação da metodologia de Buckland não é justificada, apresentando resultados equivalentes aos obtidos com o uso do modelo selecionado. Os resultados de Candolo (2001) revelam ainda que o cálculo da variância do estimador para a predição sob as diferentes metodologias comparadas forneceram resultados semelhantes, embora seja necessário maior desenvolvimento teórico e empírico para a generalização deste resultado.

Outro trabalho que compara o desempenho das metodologias de ponderação freqüentista e Bayesiana de modelos é o de Candolo (2004), que através de dados simulados, apresenta resultados empíricos que indicam desempenho semelhante à variância dos estimadores obtidos por cada uma das metodologias.

2.3.1 A ponderação Bayesiana de modelos

A ponderação Bayesiana de modelos ou *Bayesian Model Averaging* (BMA) possibilita ao pesquisador obter uma distribuição a *posteriori* da quantidade de interesse Δ , que pode ser o efeito de uma variável explicativa, representado por β ou a predição da variável resposta Y condicionada a determinados valores das variáveis explicativas, representada por $Y_0 = Y|\mathbf{x}_0$. A estrutura geral do BMA é dada por:

$$\pi(\Delta|\mathbf{y}) = \sum_{M_k \in \mathcal{C}} \pi(\Delta|M_k, \mathbf{y}) p_{M_k} \quad (4)$$

A expressão (4) é uma mistura ponderada das distribuições a *posteriori* de Δ , $\pi(\Delta|M_k, \mathbf{y})$, sob cada modelo pertencente ao conjunto \mathcal{C} . A definição do conjunto \mathcal{C} depende da quantidade populacional a ser representada por Δ . Assim, se Δ for $Y|\mathbf{x}_0$, \mathcal{C} será formado por todos os 2^p modelos sem interação que podem ser propostos para representar os dados amostrais quando p variáveis explicativas são observadas, pois neste caso, todos os modelos dão informação acerca de Δ . Se, por outro lado, a quantidade de interesse for β_j , \mathcal{C} será formado apenas pelos modelos que contém X_j em seu componente determinístico.

No processo de ponderação, o peso do k -ésimo modelo com vetor de parâmetros $\boldsymbol{\theta}_k$ é dado por:

$$p_{M_k} = \frac{P(M_k|\mathbf{y})}{\sum_{M_l \in \mathcal{C}} P(M_l|\mathbf{y})}, \quad l = 1, \dots, n(\mathcal{C}),$$

sendo $n(\mathcal{C})$ o número de modelos contidos em \mathcal{C} , e $P(M_k|\mathbf{y})$ a probabilidade condicionada aos dados \mathbf{y} do modelo M_k dada por:

$$P(M_k|\mathbf{y}) = \frac{L_k(\mathbf{y}) P(M_k)}{\sum_{M_l \in \mathcal{C}} L_l(\mathbf{y}) P(M_l)}, \quad (5)$$

sendo $P(M_k)$ a probabilidade a *priori* de M_k , $L_k(\mathbf{y})$ a função de verossimilhança marginal correspondente ao modelo M_k , dada por:

$$L_k(\mathbf{y}) = \int \dots \int L(\boldsymbol{\theta}_k; \mathbf{y}) \pi(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k, \quad (6)$$

sendo $L(\boldsymbol{\theta}_k; \mathbf{y})$ a função de verossimilhança de $\boldsymbol{\theta}_k$ e $\pi(\boldsymbol{\theta}_k)$ a distribuição a *priori* de $\boldsymbol{\theta}_k$ sob o modelo M_k .

Geralmente, o pesquisador não tem preferência por um ou outro modelo. Portanto, é comum definir a distribuição uniforme sobre o conjunto \mathcal{C} , ou seja:

$$P(M_k) = \frac{1}{n(\mathcal{C})}.$$

Utilizando a medida de divergência da informação de Kullback-Leibler, Madigan & Raftery (1994) mostram que:

$$E \left[-\log \left(\sum_{M_k \in \mathcal{C}} \pi(\Delta | M_k, \mathbf{y}) p_{M_k} \right) \right] \leq E[-\log \pi(\Delta | M_k, \mathbf{y})],$$

em que as esperanças são referentes a distribuição a *posteriori* conjunta de Δ e M_k . Ou seja, a ponderação Bayesiana de todos os modelos do conjunto \mathcal{C} apresenta, teoricamente, melhor desempenho preditivo em relação ao desempenho de qualquer um dos modelos de \mathcal{C} quando avaliada através do logaritmo do escore preditivo (LEP) proposto por Good (1952).

O principal obstáculo para a aplicação do *BMA* é a quantidade de modelos contidos em \mathcal{C} , que pode ser muito grande dependendo do número de variáveis explicativas observadas. Dentre as principais propostas para contornar este problema estão o *Occam's Window* (Madigan & Raftery, 1994), a Composição MCMC (Madigan & York, 1995) e o *stochastic search variable selection* (George & McCulloch, 1993).

O *Occam's Window* (Madigan & Raftery, 1994) é um método criado com o propósito de reduzir a quantidade de modelos em \mathcal{C} . Segundo os autores, a concepção de pesquisa científica como um processo iterativo onde as hipóteses (ou modelos) menos plausíveis são descartadas(os) em cada iteração até se chegar no conjunto das hipóteses (ou modelos) mais aceitáveis não é representada quanto (4) é diretamente aplicada, pois até mesmo os modelos menos verossímeis de \mathcal{C} diante da amostra obtida são levados em conta.

Portanto, a primeira etapa do *Occam's Window* tem o objetivo de descartar os modelos menos plausíveis conforme a seguinte regra:

1º) Eliminar de \mathcal{C} os modelos pertencentes ao conjunto A , definido por:

$$A = \left\{ M_k \in \mathcal{C} : \frac{\max_l P[M_l | \mathbf{y}]}{P[M_k | \mathbf{y}]} > a \right\}, \quad (7)$$

sendo a definido conforme o contexto da pesquisa. Em seu trabalho, Madigan & Raftery (1994) utilizam $a = 20$ sem uma argumentação mais detalhada. Jeffreys² (1961), citado por Madigan & Raftery (1994), recomenda o valor de a entre 10 e 100 independentemente do contexto.

²JEFFREYS, H. Theory of probability. 3. ed. Oxford: Oxford University Press, 1961.

Além disso, o Princípio da Parcimônia³ ou *Occam's Razor* criado por William de Ockham por volta de 1315, preconiza que dentre todas as hipóteses aceitáveis que explicam um acontecimento seja escolhida a mais simples. Esta recomendação é a base para a segunda regra do *Occam's Window*:

2º) Descartar de $A^c = \mathcal{C} - A$, ou seja, descartar do conjunto dos modelos que restaram após a primeira eliminação, os modelos pertencentes ao conjunto B , definido por:

$$B = \left\{ M_k \in A^c : \exists M_l \in A^c, M_l \subset M_k, \frac{P[M_l|\mathbf{y}]}{P[M_k|\mathbf{y}]} > O_R \right\}, \quad (8)$$

sendo O_R definido pelo usuário. Geralmente, define-se $O_R = 1$ (Madigan & Raftery, 1994).

Em suma, a primeira regra promove o descarte dos modelos menos plausíveis enquanto que a segunda regra promove o descarte dos modelos que mostram-se menos aceitáveis em relação aos seus sub-modelos encaixados. Quando apenas a primeira regra é aplicada, diz-se que o *Occam's Window* é simétrico e quando as duas regras são aplicadas, diz-se que o *Occam's Window* é exato.

Em Madigan & Raftery (1994) podem ser encontrados algoritmos para a implementação computacional do *Occam's Window* (OW). Através de resultados empíricos, os autores mostram que a ponderação Bayesiana feita apenas com os modelos selecionados pelo OW apresenta desempenho preditivo muito próximo do desempenho da ponderação Bayesiana feita com todos os modelos.

2.3.2 Desempenho preditivo da ponderação Bayesiana de modelos

Acredita-se que a multicolinearidade presente em um conjunto de dados possa informar o pesquisador acerca da incerteza diante da escolha de um modelo e, por isso, diferentes situações de incerteza podem ser simuladas através de diferentes graus de multicolinearidade.

O grau de multicolinearidade de um conjunto de dados pode ser medido pelo número de condição (k^*), dado por:

$$k^* = \frac{\lambda_{max}}{\lambda_{min}},$$

sendo λ_{max} e λ_{min} , respectivamente, o maior e o menor autovalor da matriz $\mathbf{X}'\mathbf{X}$, sendo \mathbf{X} a matriz de valores observados das variáveis explicativas na forma padronizada, sendo que a

³<http://astro.if.ufrgs.br/metodo.htm>

j -ésima variável padronizada, x_{pij} , é dada por:

$$x_{pij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{\sum_{i=1}^n x_{ij}^2 - \frac{(\sum_{i=1}^n x_{ij})^2}{n}}}, i = 1, \dots, n, \text{ (Montgomery \& Peck, 1992).}$$

Segundo os autores, valores de k^* menores que 100 indicam que não há um problema sério de multicolinearidade, já valores de k^* entre 100 e 1000 indicam moderada a forte multicolinearidade, e valores acima de 1000 indicam severa multicolinearidade.

Um dos objetivos de Hoeting (1994) é investigar o desempenho preditivo do BMA através de diferentes valores de k^* . Para isso, a autora utiliza dois conjuntos de dados reais. Seus resultados empíricos mostraram que o desempenho preditivo do BMA foi superior ao desempenho dos métodos de seleção de modelos em condições de extrema multicolinearidade ($k^* \geq 1000000$).

Outro trabalho que avalia o desempenho preditivo do BMA explorando a questão da quantidade de incerteza envolvida no processo de seleção medida pela correlação existente entre as variáveis é o trabalho de Raftery et al. (1997). Neste trabalho, foi realizado um estudo de simulação e o desempenho preditivo foi medido através da *cobertura preditiva*, medida pela proporção das observações de um conjunto teste que se encontram dentro dos limites de um intervalo de predição com nível de significância estabelecido pelos autores de 0,10. Seus resultados empíricos mostraram que a cobertura preditiva da ponderação Bayesiana feita utilizando OW foi superior a capacidade preditiva dos métodos usuais de seleção de modelos.

Apesar dos estudos de comparação e avaliação da metodologia BMA estarem atualmente mais em destaque devido, ao ainda, pouco tempo de exploração desta metodologia, podem ser encontrados, na literatura, trabalhos com enfoque mais aplicado, como por exemplo, os trabalhos de Lee III (1999) voltado para a aplicação do BMA em redes neurais e Viallefont et al. (2001) que aplica a ponderação Bayesiana de modelos na epidemiologia.

3 MATERIAL E MÉTODOS

3.1 Material

Conforme dados da FAO - Food Agriculture and Organization of the United States, estima-se que 10% da produção mundial de grãos é perdida durante o período de armazenamento devido ao ataque de pragas, tais como a *Sitophilus oryzae* (Linné, 1763) (Gallo et. al, 1988). No Brasil, a perda ocorrida durante as fases de colheita e armazenamento de milho, arroz, trigo e feijão na safra de 1991/1992 foi estimada em 20% de acordo com o IEA - Instituto de Economia Aplicada de São Paulo (Fontes et al., 2003).

Segundo o mesmo IEA⁴, a estimativa da produção total de milho em 2004 no Estado de São Paulo foi de 74.070.779 sc.60Kg (sacas de 60kg). Supondo que o índice de perda de grãos tenha se mantido em 20% e que essa perda distribui-se uniformemente sobre as culturas citadas, pode-se concluir que a perda estimada do milho, somente no ano de 2004, foi de aproximadamente 3.703.539 sc.60kg. Considerando-se que, no ano de 2004, os produtores tenham recebido por saca do milho o valor de R\$ 16,50, pode-se concluir que a perda devido ao armazenamento e ao ataque de pragas foi de R\$ 61.108.394,00.

Portanto, o aperfeiçoamento dos processos de colheita do milho e o melhor controle de pragas podem acarretar em aumento da receita dos produtores. Por essa razão, estudos com o propósito de entender o comportamento das pragas e os danos causados por elas nas diferentes culturas são freqüentemente realizados.

Embora o presente trabalho não esteja voltado para uma discussão sobre o impacto destas questões na economia, é importante que o leitor compreenda os motivos pelos quais os dados apresentados a seguir foram utilizados neste trabalho.

⁴Informações obtidas em www.iea.sp.gov.br em 29/07/2005.

3.1.1 Para a aplicação da metodologia BMA

Para a aplicação do *BMA* foram utilizados os dados⁵ amostrais contidos em Matsuo (1986), provenientes de um experimento agrônomico conduzido dentro de uma câmara de condições controladas da Universidade Federal do Paraná. Seu objetivo foi verificar os efeitos da intensidade da população de *Sitophilus oryzae* (Linné, 1763) (Coleoptera - Curculionidae) e do período de armazenamento sobre a perda de peso (em %) do milho de três diferentes variedades: C-111, Flint-Composto e Piranão.

Os dados de Matsuo, além de contribuírem para o entendimento de uma questão de impacto no cenário econômico, onde novidades metodológicas em análise de dados trazem benefícios importantes para a economia, possuem uma característica técnica importante e diretamente relacionada com os propósitos deste trabalho - que é o fato dos três conjuntos de dados possuírem graus de multicolinearidade bem diferentes um do outro, possibilitando com isso a ampliação das conclusões acerca do BMA.

A respeito de como o experimento foi realizado, Matsuo (1986) explica que cada parcela experimental foi constituída por um frasco contendo 500 (quinhentos) grãos de milho. O delineamento experimental utilizado foi o inteiramente casualizado em esquema fatorial com os seguintes fatores: número de casais de *Sitophilus oryzae* em 4 níveis (0, 5, 10 e 20 casais), e tempo de armazenamento em 3 níveis (60, 105 e 150 dias), caracterizando um fatorial 4x3. O experimento foi repetido 3 vezes, totalizando 36 observações (3 observações para cada um dos 12 tratamentos). Este mesmo esquema foi seguido em cada um dos cultivares estudados. Além das duas variáveis controladas, denotadas aqui por X_1 e X_2 , foram observadas outras seis, que são apresentadas a seguir:

Y : Perda de peso (%)

X_1 : Número de casais de *Sitophilus oryzae*.

X_2 : Produto entre o tempo de armazenamento e 0 (zero) se a parcela não foi infestada, e 1 (um) se a parcela foi infestada.

X_3 : Teor de carboidratos (%)

X_4 : Teor de umidade (%)

⁵As variáveis Y e X_1 já se encontram transformadas em Matsuo para que as suposições do modelo de regressão fossem satisfeitas. No entanto, Matsuo não diz qual função foi utilizada para realizar a transformação.

X_5 : Teor de matéria seca (%)

X_6 : Teor de proteínas (%)

X_7 : Teor de ácidos graxos (%)

X_8 : Índice de acidez dos ácidos graxos (%)

Os dados referentes aos três cultivares estudados encontram-se nas Tabelas 1,

2 e 3.

Tabela 1. Dados referentes ao cultivar C-111, Paraná, 1976

i	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	4,360	0,707	0	72,086	12,391	0,987	10,005	4,530	24,336
2	5,437	0,707	0	72,865	12,607	0,951	8,936	4,641	26,212
3	5,425	0,707	0	72,568	12,529	0,951	9,629	4,322	22,630
4	9,418	6,819	60	70,977	12,949	1,031	10,469	4,572	22,113
5	8,148	6,519	60	72,167	12,581	0,999	9,777	4,474	28,147
6	7,649	4,743	60	72,195	12,615	1,105	9,495	4,588	27,263
7	15,370	12,786	60	71,073	13,142	1,041	9,909	4,735	44,133
8	12,485	9,823	60	72,216	12,979	1,001	9,035	4,767	36,757
9	14,940	13,473	60	72,634	13,039	1,107	8,576	4,644	40,565
10	17,276	13,209	60	71,390	13,005	1,040	9,673	4,891	42,461
11	14,586	11,379	60	71,801	13,280	1,030	9,027	4,861	40,492
12	16,118	12,980	60	71,956	13,048	1,035	9,341	4,620	41,236
13	6,812	0,707	0	72,680	12,951	1,155	8,792	4,420	31,440
14	6,632	0,707	0	72,867	12,835	1,224	8,520	4,552	33,835
15	7,927	0,707	0	72,400	13,024	1,235	8,897	4,444	29,245
16	13,644	8,031	105	70,700	13,371	1,287	9,897	4,744	42,660
17	17,484	11,022	105	71,325	13,498	1,281	9,323	4,571	43,550
18	9,842	6,123	105	71,334	13,206	1,209	9,660	4,590	36,945
19	22,991	14,983	105	69,840	13,556	1,393	10,273	4,938	67,534
20	25,814	17,762	105	69,274	13,659	1,340	10,856	4,870	97,399
21	25,718	18,207	105	69,728	13,457	1,411	10,221	5,182	93,434
22	22,640	14,747	105	70,049	13,118	1,326	10,346	5,161	56,625
23	31,502	22,726	105	70,028	13,412	1,412	10,473	4,674	148,556
24	30,361	21,737	105	68,504	13,951	1,445	11,883	4,215	172,521
25	4,009	0,707	0	72,932	12,470	1,210	9,147	4,240	15,842
26	4,887	0,707	0	72,225	12,777	1,146	9,456	4,395	15,647
27	5,239	0,707	0	72,872	12,842	1,207	8,680	4,397	15,934
28	23,085	13,693	150	70,312	12,898	1,226	9,734	4,628	37,867
29	24,806	15,475	150	69,680	13,428	1,356	10,740	4,795	36,988
30	22,747	14,300	150	71,289	13,503	1,228	9,199	4,780	34,643
31	22,317	13,583	150	69,964	13,953	2,418	9,892	3,773	41,520
32	24,088	14,439	150	69,531	13,587	2,495	9,588	4,798	35,405
33	30,147	17,930	150	70,322	13,241	1,939	9,849	4,648	43,233
34	31,096	19,634	150	69,936	13,345	2,307	10,195	4,217	71,769
35	24,646	16,688	150	71,369	13,000	1,961	9,023	4,646	38,676
36	30,432	19,912	150	70,026	12,934	2,472	10,377	4,189	66,272

Tabela 2. Dados referentes ao cultivar Piranão, Paraná, 1976

i	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	4,761	0,707	0	69,498	12,096	1,036	12,849	4,520	20,684
2	4,131	0,707	0	69,014	12,124	1,000	13,068	4,791	18,622
3	4,360	0,707	0	68,621	12,112	0,850	13,584	4,830	13,974
4	10,167	7,176	60	67,591	12,509	1,028	14,067	4,801	13,759
5	12,762	11,423	60	66,703	13,057	1,076	14,049	5,112	17,155
6	10,120	8,573	60	67,670	12,649	1,157	13,895	4,628	15,751
7	11,394	10,793	60	67,106	12,831	1,271	13,910	4,879	17,238
8	11,762	10,024	60	67,667	12,835	1,201	13,553	4,743	16,273
9	8,730	7,314	60	67,891	12,126	1,194	14,133	4,653	15,453
10	11,155	9,974	60	69,535	12,981	1,182	12,807	4,493	13,039
11	11,269	10,464	60	67,635	12,489	1,254	14,087	4,533	15,607
12	11,955	9,624	60	66,860	12,969	1,301	14,195	4,672	20,872
13	6,971	0,707	0	68,479	12,716	1,306	13,175	4,321	10,775
14	6,802	0,707	0	68,663	12,721	1,408	13,008	4,296	9,189
15	7,158	0,707	0	68,937	12,648	1,362	12,939	4,112	8,624
16	14,485	9,513	105	68,593	13,031	1,311	12,531	4,541	20,206
17	15,435	13,583	105	68,719	13,003	1,379	12,420	4,477	20,411
18	17,299	11,379	105	68,932	12,725	1,337	12,489	4,514	20,352
19	20,418	16,598	105	67,298	13,071	1,344	13,597	4,687	31,702
20	17,248	14,781	105	67,670	12,875	1,351	13,553	4,548	21,293
21	12,644	10,977	105	68,197	12,951	1,369	13,237	4,244	15,885
22	19,985	16,263	105	66,351	13,090	1,427	14,610	4,552	33,712
23	17,427	12,780	105	67,155	13,820	1,386	13,674	3,962	47,854
24	21,564	17,902	105	67,300	12,915	1,309	14,082	4,396	21,446
25	5,802	0,707	0	68,850	12,780	1,558	12,680	4,129	10,712
26	6,083	0,707	0	66,584	12,713	1,949	14,394	4,359	10,222
27	5,724	0,707	0	68,251	12,478	1,812	12,952	4,505	9,307
28	18,006	11,247	150	67,515	13,258	1,450	13,282	4,491	20,511
29	23,489	14,815	150	63,129	14,215	1,716	16,712	4,227	49,043
30	17,133	11,510	150	64,194	13,373	1,582	16,271	4,578	32,980
31	21,939	13,209	150	66,350	13,594	1,467	14,478	4,109	34,350
32	23,001	16,263	150	65,977	13,807	2,092	13,439	4,682	21,882
33	17,022	10,885	150	65,310	13,398	3,064	13,863	4,363	26,330
34	32,163	24,259	150	63,145	14,825	3,453	15,762	2,812	125,530
35	29,271	17,449	150	61,989	14,708	2,039	17,031	4,230	66,520
36	26,660	21,644	150	65,631	14,611	3,000	14,329	2,427	86,761

Tabela 3. Dados referentes ao cultivar Flint Composto, Paraná, 1976

i	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	4,314	0,707	0	68,457	11,953	0,958	13,810	4,819	52,665
2	5,367	0,707	0	68,382	12,183	1,007	13,689	4,736	46,486
3	4,592	0,707	0	68,400	12,120	0,910	13,740	4,819	43,469
4	6,398	4,301	60	67,695	12,294	1,263	14,010	4,735	49,628
5	10,861	7,516	60	67,152	12,645	1,309	14,119	4,773	45,535
6	9,390	7,382	60	66,999	12,810	1,283	14,138	4,767	49,102
7	11,683	11,811	60	66,246	12,924	1,353	14,393	5,082	73,487
8	10,216	8,689	60	67,148	12,495	1,333	14,012	5,010	67,999
9	10,504	9,772	60	67,616	12,974	1,218	13,660	4,529	72,560
10	14,998	13,509	60	66,833	13,258	1,026	13,645	5,235	76,588
11	12,828	11,853	60	67,052	12,703	1,104	14,082	5,057	51,152
12	13,156	10,653	60	67,379	12,692	1,327	13,651	4,949	59,561
13	8,177	0,707	0	67,103	13,029	1,808	13,435	4,622	31,989
14	7,321	0,707	0	67,544	12,759	1,592	13,470	4,633	32,437
15	7,636	0,707	0	67,285	12,690	1,602	13,633	4,787	31,176
16	16,674	12,469	105	66,420	13,138	1,768	13,734	4,936	49,933
17	15,150	9,354	105	65,472	13,516	1,637	14,442	4,931	39,769
18	10,488	5,244	105	66,463	13,245	1,326	14,059	4,905	47,242
19	16,882	13,729	105	66,817	13,061	1,668	13,675	4,777	78,162
20	11,922	8,514	105	65,875	13,042	1,914	14,350	4,817	81,616
21	16,342	11,510	105	66,094	13,540	2,254	13,418	4,732	86,797
22	17,532	13,838	105	65,098	13,521	2,048	14,530	4,801	57,121
23	15,808	12,629	105	66,961	13,142	1,349	13,767	4,779	53,729
24	16,647	14,017	105	66,134	13,219	1,569	14,451	4,624	55,329
25	6,447	0,707	0	68,095	12,979	1,497	13,071	4,356	26,710
26	7,156	0,707	0	66,437	13,066	1,460	14,253	4,782	22,299
27	7,606	0,707	0	67,919	13,203	1,462	12,968	4,445	21,305
28	22,904	14,815	150	63,292	14,523	2,387	14,951	4,844	42,634
29	16,041	9,513	150	64,680	13,978	1,592	14,762	4,729	31,535
30	21,393	14,089	150	64,384	14,040	1,734	14,894	4,946	35,211
31	23,779	13,656	150	65,650	14,069	1,435	14,085	4,759	47,302
32	19,382	13,133	150	66,222	13,689	1,186	14,039	4,861	40,428
33	26,115	16,896	150	64,783	14,049	1,421	15,028	4,716	49,292
34	24,675	16,263	150	63,391	14,422	1,139	16,703	4,343	64,506
35	22,855	15,732	150	64,398	13,752	1,217	15,662	4,969	43,935
36	32,691	21,459	150	62,353	15,646	1,434	18,235	2,330	115,699

3.1.2 Para a avaliação do desempenho preditivo do BMA

A comparação entre o desempenho preditivo da ponderação Bayesiana e o desempenho preditivo dos critérios⁶ de seleção de modelos foi feita através de dados simulados obtidos a partir da utilização de medidas descritivas referentes aos dados dos três cultivares. As medidas descritivas utilizadas foram: $\bar{\mathbf{x}}$ (vetor de médias amostral), \mathbf{S} (matriz de variâncias e covariâncias amostral), $\hat{\boldsymbol{\beta}}$ (estimativa do vetor dos coeficientes de regressão do modelo com todas as variáveis explicativas observadas) e $QMRes$ (a estimativa de σ^2 obtida pelo ajuste do mesmo modelo).

Os graus de multicolinearidade medido pelo número de condição da matriz $\mathbf{X}'\mathbf{X}$ padronizada encontrados foram: 391,39 (cultivar C-111), 1232,68 (cultivar Piranão) e 12826,39 (Flint Composto), os quais indicam, respectivamente, moderada, forte e severa multicolinearidade.

Segundo Hoeting (1994), grau de multicolinearidade elevado pode contribuir para aumentar a incerteza do pesquisador diante da escolha de um modelo. Em situações deste tipo pode ser interessante a utilização de métodos que permitem combinar informações de diferentes modelos acerca de determinadas quantidades populacionais.

Foram geradas amostras de tamanho amostral 50, 75 e 100. Para cada tamanho amostral considerado foram geradas 1000 amostras utilizando as medidas descritivas de cada um dos cultivares, totalizando 9000 amostras, sendo 1000 para cada combinação entre tamanho amostral e cultivar.

As variáveis explicativas e a variável resposta foram respectivamente geradas da seguinte forma:

$$\mathbf{X}'_i = (X_{i1}, X_{i2}, \dots, X_{i8}) \sim N(\bar{\mathbf{x}}, \mathbf{S}); \quad i = 1, 2, \dots, n.$$

$$Y_i = \mathbf{X}'_i \hat{\boldsymbol{\beta}} + \varepsilon_i, \quad i = 1, 2, \dots, n$$

sendo $\varepsilon_i \sim N(0, QMRes)$. Portanto, todos os conjuntos simulados possuem oito variáveis explicativas X_1, X_2, \dots, X_8 e uma variável resposta Y . logo, a cada simulação foi obtido um conjunto de dados com a seguinte estrutura:

⁶Os critérios avaliados foram: Maior R^2 ; Maior \bar{R}^2 ; Menor C_p e PRESS. Todos esses critérios encontram-se definidos em 2.1.

Tabela 4. Ilustração de um específico conjunto de dados simulado

i	y_i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	x_{i6}	x_{i7}	x_{i8}
1	y_1	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}
2	y_2	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}	x_{28}
\vdots									
n	y_n	x_{n1}	x_{n2}	x_{n3}	x_{n4}	x_{n5}	x_{n6}	x_{n7}	x_{n8}

3.2 Métodos

A ponderação Bayesiana pode ser aplicada em várias classes de modelos com o propósito de combinar informações acerca de uma determinada quantidade de interesse, representada neste trabalho por Δ . Nesse trabalho serão considerados, somente, modelos de regressão linear clássica.

3.2.1 Ponderação Bayesiana de modelos de regressão linear clássica

Considere o modelo de regressão linear clássica representado por:

$$Y_i = \beta_0 + \sum_{j=1}^p \beta_j X_{ij} + \varepsilon_i, \quad (9)$$

com $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, \dots, n$, sendo que:

$$\text{cov}(\varepsilon_i, \varepsilon_{i'}) = \begin{cases} 1 & \text{se } i = i'; \\ 0 & \text{se } i \neq i'. \end{cases}$$

Desde que sejam observadas p variáveis explicativas em um experimento, o número de modelos cuja forma é dada por (9) é igual a 2^p . Representando por \mathcal{C} o conjunto que contém todos esses modelos, tem-se então que $n(\mathcal{C}) = 2^p$.

O k -ésimo modelo de \mathcal{C} pode ser reescrito na forma matricial da seguinte forma:

$$M_k : \mathbf{Y} \sim N(\mathbf{X}_k \boldsymbol{\beta}_k, \sigma^2 \mathbf{I}_k),$$

em que

- (i) \mathbf{Y} é o vetor aleatório n dimensional que representa a variável resposta;

- (ii) \mathbf{X}_k é a matriz de dimensão $n \times (p_k + 1)$ dos valores observados das p_k variáveis explicativas contidas no modelo M_k ;
- (iii) $(\boldsymbol{\beta}_k, \sigma^2)' = \boldsymbol{\theta}_k$, sendo $\boldsymbol{\beta}_k = (\beta_0, \beta_1, \dots, \beta_{p_k})'$ o vetor dos coeficientes de regressão de dimensão $(p_k + 1) \times 1$.

A quantidade de interesse do pesquisador, Δ , pode ser o efeito de uma variável explicativa sobre Y , suponha que seja β_j , neste caso, para a aplicação do BMA deve-se considerar apenas os modelos pertencentes a \mathcal{C} que contêm o coeficiente de regressão β_j . Logo, a ponderação Bayesiana será feita sobre os modelos pertencentes ao conjunto \mathcal{C}_Δ , definido por:

$$\mathcal{C}_\Delta = \{M_k \in \mathcal{C} | \beta_j \in \boldsymbol{\beta}_k\}.$$

Por outro lado, a quantidade de interesse do pesquisador, Δ , pode ser a predição da variável resposta condicionada a determinados valores das variáveis explicativas, $Y | \mathbf{x}_0$, neste caso, a ponderação deve ser feita levando-se em conta todos os 2^p modelos, pois todos transmitem informação acerca do valor futuro de Y , logo, $\mathcal{C}_\Delta = \mathcal{C}$.

De acordo com a estrutura geral do BMA dada pelas expressões (4), (5) e (6), deve-se a princípio ser definida:

- (i) $\pi(\boldsymbol{\theta}_k)$, a distribuição a *priori* para $\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k', \sigma^2)$ e;
- (ii) $P(M_k)$, as probabilidades a *priori* para os modelos de \mathcal{C}_Δ .

Com relação a (i), foi adotada neste trabalho a distribuição *priori* conjugada normal-gama com a seguinte parametrização:

$$\begin{aligned} \boldsymbol{\beta}_k | \sigma^2 &\sim N(\boldsymbol{\mu}_k, \sigma^2 \mathbf{V}_k) \\ \sigma^2 &\sim \text{Gamma} - \text{invertida} \left(\frac{\nu}{2}, \frac{\nu\lambda}{2} \right), \end{aligned}$$

e dessa forma tem-se que

$$\pi(\boldsymbol{\theta}_k) = (2\pi)^{-(p_k+1)/2} |\sigma^2 \mathbf{V}_k|^{-1/2} e^{\left\{ -\frac{1}{2} (\boldsymbol{\beta}_k - \boldsymbol{\mu}_k)' (\sigma^2 \mathbf{V}_k)^{-1} (\boldsymbol{\beta}_k - \boldsymbol{\mu}_k) \right\}} \frac{(\nu\lambda/2)^{\nu/2}}{\Gamma(\nu/2)} \sigma^{2-(\nu/2-1)} e^{-\frac{\nu\lambda}{2\sigma^2}},$$

sendo a opinião subjetiva do pesquisador representada pelos valores de $\boldsymbol{\mu}_k$, \mathbf{V}_k , ν , λ e ϕ , definidos por:

$$\boldsymbol{\mu}_k = \begin{pmatrix} \widehat{\beta}_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

sendo $\widehat{\beta}_0$ a estimativa de mínimos quadrados de β_0 sob o modelo M_k e

$$\mathbf{V}_k = \begin{pmatrix} s_Y^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \phi^2/s_{X_1}^2 & 0 & & \vdots & & \vdots & \vdots \\ \vdots & 0 & \phi^2/s_{X_1}^2 & & \vdots & & \vdots & \vdots \\ \vdots & \vdots & 0 & \ddots & 0 & & \vdots & \vdots \\ \vdots & \vdots & \vdots & & \phi^2/s_{X_j}^2 & & \vdots & \vdots \\ \vdots & \vdots & \vdots & & 0 & \ddots & 0 & \vdots \\ \vdots & \vdots & \vdots & & \vdots & & \phi^2/s_{X_{p_k-1}}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi^2/s_{X_{p_k}}^2 \end{pmatrix},$$

com todos os elementos fora da diagonal principal iguais a zero, sendo s_Y^2 a variância amostral de Y e $s_{X_j}^2$ a variância amostral da j -ésima variável explicativa contida em M_k e ν , λ e ϕ podendo assumir um dos três conjuntos de valores (Tabela 6).

Tabela 5. Opções para os valores dos hiperparâmetros

Conjunto	ν	λ	ϕ
1	0,17	0,59	5,29
2	0,28	1,29	2,86
3	2,58	0,28	2,85

Essas possibilidades de valores para ν , λ e ϕ são os resultados obtidos por Hoeting (1994) após a resolução de um sistema com as seguintes restrições:

- (i) A distribuição *a priori* para $(\beta_0, \beta_1, \dots, \beta_{p_k})$ deve ser razoavelmente “flat” sobre o hiper-cubo $[-1, 1]^{p_k}$;
- (ii) A distribuição *a priori* para σ^2 deve ser razoavelmente “flat” sobre o intervalo $(a, 1)$ para um a pequeno;

(iii) Seja o evento $\{\sigma^2 < 1\}$. A probabilidade de ocorrência deste evento deve ser maximizada.

Dizemos que há uma conjugação entre as distribuições a *priori* e *posteriori* quando a distribuição a *posteriori* possui a mesma forma paramétrica da distribuição a *priori*. A adoção de famílias conjugadas é algo freqüente em análise de dados sob enfoque Bayesiano pela conveniência matemática e computacional trazida por ela, além de facilitar a interpretação dos resultados pelo fato dos parâmetros da *posteriori* serem basicamente os parâmetros da *priori* atualizados com a informação da amostra, transmitida através da função de verossimilhança. Porém, tais conveniências não devem se sobrepor ao objetivo principal do processo Bayesiano de análise, que é a representação adequada da opinião subjetiva do pesquisador. Sendo assim, famílias conjugadas só devem ser adotadas se o objetivo principal puder ser atendido, caso contrário, deve-se priorizar a representatividade da opinião do pesquisador com uma *priori* mais adequada, mesmo que a *posteriori* resultante tenha uma forma desconhecida, ou seja, não tenha a forma de alguma das funções densidade conhecidas da literatura.

Embora a *priori* aqui adotada não comprometa o interesse principal deste trabalho, que é a avaliação do desempenho preditivo do BMA, é bom ressaltar que, para trabalhos nos quais o objetivo é a análise dos dados e não estudos comparativos, as informações do pesquisador acerca do fenômeno em estudo devem estar devidamente representadas, logo, os hiperparâmetros devem assumir valores que representem as opiniões do pesquisador.

Os hiperparâmetros ν e λ definem a forma da distribuição probabilística de σ^2 (a variância do termo aleatório dos modelos). Já o conhecimento prévio acerca dos coeficientes de regressão linear é expresso através de ϕ , de modo que valores maiores de ϕ representam maior dispersão em torno do valor esperado dos coeficientes. Por exemplo, valores entre 2,50 a 3,50 representam informações mais precisas a respeito do efeito das variáveis explicativas que valores entre 4,00 a 5,00.

A Figura 2 ilustra as diferentes formas da distribuição a *priori* de σ^2 para cada conjunto de valores da Tabela 6. Todos os conjuntos representam a opinião subjetiva do pesquisador acerca do erro experimental, que de acordo com o que se pode observar, a opinião subjetiva do pesquisador a respeito do controle do erro experimental realizado é a de que os fatores que mais influenciam a variabilidade da resposta foram controlados.

Com relação a (ii) (as probabilidades a *priori* para os modelos de \mathcal{C}_Δ), geral-

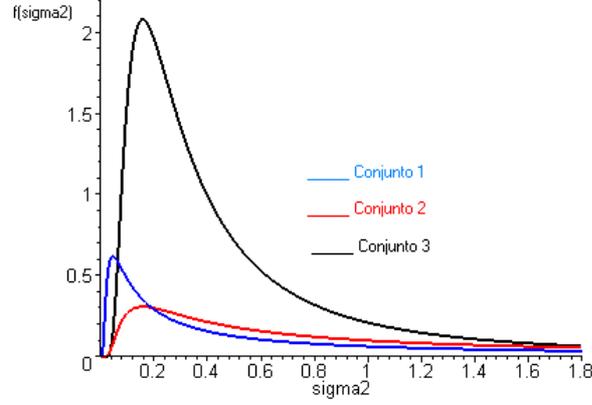


Figura 2 - Distribuições a *priori* para σ^2 segundo valores de ν e λ

mente o pesquisador não tem preferência por um modelo específico de \mathcal{C}_Δ antes da realização do experimento e por essa razão, costuma-se atribuir probabilidades iguais para todos os modelos, deste modo:

$$P(M_k) = \frac{1}{n(\mathcal{C}_\Delta)}, \forall M_k \in \mathcal{C}_\Delta.$$

Após as definições de (i) e (ii), que constituem as etapas do BMA nas quais as opiniões subjetivas são tratadas probabilisticamente, e etapa seguinte envolve a obtenção da função de verossimilhança marginal $L_k(\mathbf{y})$ conforme (6). Para isso, é necessário combinar a distribuição a *priori* e a função de verossimilhança de $\boldsymbol{\theta}_k$. A função de verossimilhança de $\boldsymbol{\theta}_k$, definida por:

$$L(\boldsymbol{\theta}_k; \mathbf{y}_1, \dots, \mathbf{y}_n) = \prod_{i=1}^n f_{\mathbf{Y}_i}(\mathbf{y}_i | \boldsymbol{\theta}_k) = \prod_{i=1}^n f_{\mathbf{Y}_i}(\mathbf{y}_i | \boldsymbol{\beta}_k, \sigma^2),$$

sendo $f_{\mathbf{Y}_i}(\mathbf{y}_i | \boldsymbol{\beta}_k, \sigma^2)$ a função densidade de probabilidade de \mathbf{Y}_i . Como no modelo de regressão linear clássica as n observações da variável Y podem ser vistas como elementos de um vetor \mathbf{y} de dimensão n que segue distribuição normal com média $\mathbf{X}_k \boldsymbol{\beta}_k$ e variância $\sigma^2 \mathbf{I}$, a função de verossimilhança de $\boldsymbol{\theta}_k$ é dada por:

$$L(\boldsymbol{\theta}_k; \mathbf{y}) = |2\pi\sigma^2\mathbf{I}|^{-n/2} e^{\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{X}_k\boldsymbol{\beta}_k)'(\sigma^2\mathbf{I})^{-1}(\mathbf{y}-\mathbf{X}_k\boldsymbol{\beta}_k)\right\}}.$$

Assim, a função de verossimilhança marginal $L_k(\mathbf{y})$ correspondente ao k -ésimo

modelo de regressão linear clássica é dada por:

$$\begin{aligned}
L_k(\mathbf{y}) &= \int_{-\infty}^{\infty} \int_0^{\infty} \dots \int_0^{\infty} L(\boldsymbol{\theta}_k; \mathbf{y}) \pi(\boldsymbol{\beta}_k | \sigma^2) \pi(\sigma^2) d\sigma^2 d\boldsymbol{\beta}_k \\
&= \frac{\Gamma\left(\frac{n+\nu}{2}\right) |\lambda(\mathbf{I} + \mathbf{X}_k \mathbf{V}_k \mathbf{X}_k')|^{-1/2}}{\Gamma(\nu/2) \nu^{n/2} \pi^{n/2}} \\
&\quad \times \left[1 + \frac{1}{\nu} (\mathbf{y} - \mathbf{X}_k \boldsymbol{\mu}_k)' (\lambda(\mathbf{I} + \mathbf{X}_k \mathbf{V}_k \mathbf{X}_k'))^{-1} (\mathbf{y} - \mathbf{X}_k \boldsymbol{\mu}_k) \right]^{-\frac{(n+\nu)}{2}}
\end{aligned}$$

que corresponde a função densidade de probabilidade de uma distribuição t de *Student* n dimensional com ν graus de liberdade, média $\mathbf{X}_k \boldsymbol{\mu}_k$ e variância $\frac{\nu}{(\nu-2)} \lambda(\mathbf{I} + \mathbf{X}_k \mathbf{V}_k \mathbf{X}_k')$.

Diferentemente do que ocorre nas demais classes de modelos, a função de verossimilhança marginal $L_k(\mathbf{y})$ pode ser obtida de forma analítica na classe dos modelos de regressão linear clássica sem que seja preciso recorrer a métodos de integração numérica. Isso permite a obtenção exata da probabilidade condicionada aos dados \mathbf{y} $P(M_k | \mathbf{y})$. Para outras classes de modelos, Raftery (1995) sugere a seguinte aproximação para $L_k(\mathbf{y})$:

$$\log L_k(\mathbf{y}) \approx \log f(\mathbf{y} | \widehat{\boldsymbol{\theta}}_k, M_k) - (d_k/2) \log(n),$$

sendo $\widehat{\boldsymbol{\theta}}_k$ o estimador de máxima verossimilhança de $\boldsymbol{\theta}_k$, cuja dimensão é d_k e n o tamanho amostral. Para os modelos de regressão linear em geral, o estimador de máxima verossimilhança de $\widehat{\boldsymbol{\theta}}_k$ pode ser obtido sem dificuldade mesmo que a sua dimensão seja muito grande, entretanto, para modelos não-lineares, o estimador de máxima verossimilhança de $\boldsymbol{\theta}_k$ deve ser obtido por meio de métodos iterativos, sendo que a dificuldade de obtenção aumenta a medida que aumenta a dimensão de $\boldsymbol{\theta}_k$.

Após a obtenção de $L_k(\mathbf{y})$ e $P(M_k | \mathbf{y})$, a próxima etapa envolve a obtenção de $\pi(\Delta | M_k, \mathbf{y})$, a distribuição a *posteriori* de Δ para cada modelo de \mathcal{C}_Δ . Para os casos em que $\Delta = \beta_j$, deve-se obter a distribuição a *posteriori* de β_{jk} , ou seja, a distribuição de β_j sob cada modelo de \mathcal{C}_Δ . Recomenda-se que primeiramente seja obtida a distribuição de $\boldsymbol{\beta}_k$, ou seja, do vetor de coeficientes de regressão do modelo M_k . Há duas possibilidades:

- (i) Obter $\pi(\boldsymbol{\beta}_k | M_k, \mathbf{y})$ via distribuição marginal a *posteriori* de $(\boldsymbol{\beta}_k, \sigma^2)$. Isto é:

$$\pi(\boldsymbol{\beta}_k | \mathbf{y}) = \int_0^{\infty} \pi(\boldsymbol{\beta}_k, \sigma^2 | \mathbf{y}) d\sigma^2,$$

cujos resultados são $\boldsymbol{\beta}_k | \mathbf{y} \sim t_{\nu_*}(\boldsymbol{\mu}_{k*}, \lambda_* \mathbf{V}_{k*})$, ou seja, o vetor dos coeficientes de regressão de M_k segue distribuição t de *Student* $(p_k + 1)$ dimensional com vetor de locação:

$$\boldsymbol{\mu}_{k*} = (\mathbf{X}'_k \mathbf{X}_k + \mathbf{V}_k^{-1})^{-1} (\mathbf{X}'_k \mathbf{y} + \mathbf{V}_k^{-1} \boldsymbol{\mu}_k),$$

matriz de escala:

$$\lambda_* \mathbf{V}_{k*} = \lambda_* (\mathbf{X}'_k \mathbf{X}_k + \mathbf{V}_k^{-1})^{-1},$$

e ν_* graus de liberdade, sendo que:

$$\nu_* = n + \nu$$

e

$$\lambda_* = \frac{1}{\nu_*} [\nu \lambda + \boldsymbol{\mu}'_k \mathbf{V}_k^{-1} \boldsymbol{\mu}_k + \mathbf{y}' \mathbf{y} - (\mathbf{X}'_k \mathbf{y} + \mathbf{V}_k^{-1} \boldsymbol{\mu}_k)' (\mathbf{X}'_k \mathbf{X}_k + \mathbf{V}_k^{-1})^{-1} (\mathbf{X}'_k \mathbf{y} + \mathbf{V}_k^{-1} \boldsymbol{\mu}_k)].$$

(ii) Obter a distribuição a *posteriori* de $\boldsymbol{\beta}_k$ condicionada a σ^2 , $\pi(\boldsymbol{\beta}_k | \sigma^2, \mathbf{y})$. Isto é:

$$\pi(\boldsymbol{\beta}_k | \sigma^2, \mathbf{y}) \propto L(\boldsymbol{\beta}_k, \sigma^2; \mathbf{y}) \pi(\boldsymbol{\beta}_k | \sigma^2),$$

cujos resultados são $\boldsymbol{\beta}_k | \sigma^2, \mathbf{y} \sim N(\boldsymbol{\mu}_{k*}; \sigma^2 \mathbf{V}_{k*})$, ou seja, o vetor de coeficientes de regressão de M_k segue distribuição normal $(p_k + 1)$ dimensional condicionada a σ^2 com média $\boldsymbol{\mu}_{k*}$ e matriz de variâncias e covariâncias $\sigma^2 \mathbf{V}_{k*}$ apresentadas em (i).

Escolhida uma das opções, a análise prossegue e o próximo passo é a obtenção de $\pi(\beta_j | \mathbf{y})$, ou equivalentemente, $\pi(\beta_j | M_k, \mathbf{y})$, que se dá pela obtenção da distribuição marginal de β_j em relação aos outros parâmetros de $\boldsymbol{\beta}_k$. Para a opção (ii), tem-se:

$$\beta_j | \sigma^2, M_k, \mathbf{y} \sim N(\mu_{k*j}; \sigma^2 V_{k*jj}), \quad (10)$$

em que μ_{k*j} e V_{k*jj} correspondem, respectivamente, ao j -ésimo elemento de $\boldsymbol{\mu}_{k*}$ e ao j -ésimo elemento da diagonal principal de \mathbf{V}_{k*} .

Viallefont et al. (2001) apresenta a probabilidade

$$P(\beta_j \neq 0 | \mathbf{y}) = \sum_{C_\Delta} P(M_k | \mathbf{y})$$

como alternativa para testar hipótese acerca da significância de β_j . Uma regra convencional para se concluir a respeito do efeito de X_1 , segundo Viallefont et al. (2001) é dizer que não há evidências que comprovem sua significância se $P(\beta_j \neq 0 | \mathbf{y}) < 50\%$. Já valores entre 50% e 75% indicam fraca evidência, enquanto que valores acima de 75% até 95% indicam evidência positiva e valores acima de 95% indicam uma evidência muito forte.

Se por outro lado o interesse do pesquisador for a predição de $Y|\mathbf{x}_0$, então $\Delta = Y_0$ e portanto o termo $\pi(\Delta|M_k, \mathbf{y})$ representa a função densidade preditiva a *posteriori* de $Y|\mathbf{x}_0$ sob M_k , dada por:

$$f(y_0|M_k, \mathbf{y}) = \frac{\Gamma(\frac{\nu_*+1}{2})}{\Gamma(1/2)\Gamma(\nu_*/2)} \nu_*^{\nu_*/2} |\lambda_*(\mathbf{x}'_0 \mathbf{V}_{k*} \mathbf{x}_0 + 1)|^{-1/2} \left[\nu_* + \frac{(y_0 - \mathbf{x}'_0 \boldsymbol{\mu}_{k*})^2}{[\lambda_*(\mathbf{x}'_0 \mathbf{V}_{k*} \mathbf{x}_0 + 1)]} \right]^{-(\nu_*+1)/2}, \quad (11)$$

que corresponde a função densidade de uma distribuição *t* de *Student* univariada com ν_* graus de liberdade, média $\mathbf{x}'_0 \boldsymbol{\mu}_{k*}$ e variância $\frac{\nu_*}{(\nu_*-2)} \lambda_*(\mathbf{x}'_0 \mathbf{V}_{k*} \mathbf{x}_0 + 1)$.

A última etapa do BMA compreende a ponderação de todas as distribuições de Δ sob os modelos de \mathcal{C}_Δ conforme (4), sendo a média e variância ponderada dadas respectivamente por:

$$E(\Delta|\mathbf{y}) = \sum_{M_k \in \mathcal{C}_\Delta} E(\Delta|M_k, \mathbf{y}) p_{M_k},$$

$$\begin{aligned} V(\Delta|\mathbf{y}) &= \sum_{M_k \in \mathcal{C}_\Delta} V(\Delta|M_k, \sigma^2, \mathbf{y}) p_{M_k} \\ &+ \sum_{M_k \in \mathcal{C}_\Delta} [E(\Delta|M_k, \sigma^2, \mathbf{y}) - E(\Delta|\sigma^2, \mathbf{y})]^2 p_{M_k}. \end{aligned}$$

Para avaliar o desempenho preditivo do BMA, cada uma das 9000 amostras simuladas foi dividida aleatoriamente em duas partes: a primeira parte contendo $(n-25)$ observações, destinada à aplicação dos critérios de seleção R^2 , \overline{R}^2 , C_p e *PRESS* além da aplicação da ponderação de modelos e a segunda parte contendo 25 observações, utilizada para o cálculo do Logaritmo do Escore Preditivo (LEP). O LEP foi proposto por Good (1952) em seu trabalho de caráter filosófico que teve como objetivo estabelecer as relações entre a Teoria das Probabilidades e a Teoria do Comportamento Racional. Para um modelo qualquer M , selecionado por um determinado critério de seleção, o LEP é calculado da seguinte maneira:

$$LEP = - \sum_{i=1}^{25} \log f(y_{0i}|M, \mathbf{y}),$$

em que $f(y_{0i}|M, \mathbf{y})$ é o valor da expressão (11) calculado para cada observação simulada $(\mathbf{x}_{0i}, y_{0i})$, $i = 1, \dots, 25$, enquanto que o LEP da ponderação Bayesiana é dado por:

$$LEP = - \sum_{i=1}^{25} \log \left\{ \sum_{M_k \in \mathcal{C}} f(y_{0i}|M_k, \mathbf{y}) P(M_k|\mathbf{y}) \right\}.$$

Quanto menor é o valor do LEP melhor é considerada a capacidade preditiva do modelo utilizado para seu cálculo. Dessa maneira, o LEP possibilita medir o desempenho preditivo do critério que o selecionou. Como foram simuladas 1000 amostras para cada combinação entre tamanho amostral e cultivar, foram obtidos 1000 valores de LEP para cada um dos critérios de seleção e para a ponderação Bayesiana em cada combinação. Os programas utilizados para a obtenção dos valores do LEP podem ser vistos nos Anexos A e B.

4 RESULTADOS E DISCUSSÃO

As Tabelas 6, 7 e 8 são apresentadas com o propósito de mostrar quão difícil pode ser o processo de escolha de um modelo por meio de critérios de seleção, como por exemplo o Coeficiente de Determinação Múltipla Ajustado, \bar{R}^2 . O problema da incerteza diante da escolha de um modelo foi explorado utilizando os resultados do \bar{R}^2 pelo fato deste critério ser um dos mais utilizados na estatística aplicada. As respectivas Tabelas 6, 7 e 8 contêm os dois melhores ajustes dentre os modelos com 1,2,...,7 variáveis explicativas e o ajuste do modelo completo com base nos dados dos cultivares C-111, Piranão e Flint Composto. Nos três conjuntos de dados analisados pode-se observar que os valores de \bar{R}^2 dos modelos são muito semelhantes, além disso, pode-se observar que os modelos fornecem diferentes estimativas para um mesmo coeficiente de regressão, sendo que as maiores diferenças podem ser observadas entre as estimativas fornecidas pelos modelos ajustados para os dados do cultivar Flint Composto, o qual possui o maior grau de multicolinearidade dentre os três cultivares, com número de condição da matrix $\mathbf{X}'\mathbf{X}$ igual a 12826,69.

O fato das maiores variações entre as estimativas de um mesmo coeficiente terem sido observadas justamente no conjunto de dados com maior grau de multicolinearidade é coerente com a observação feita por Hoeting (1994, p.65) de que a indecisão do pesquisador diante da escolha de um modelo pode aumentar a medida que o grau de multicolinearidade aumenta. Leamer (1978, p.92) também faz uma observação nesse sentido, dizendo que a incerteza sobre o modelo a ser escolhido é irrelevante para o processo de inferência dos coeficientes se e somente se suas estimativas forem as mesmas para todos os modelos.

Devido a maior multicolinearidade ter sido observada nos dados do cultivar Flint Composto foi acrescentada, na Tabela 8, uma coluna referente a probabilidade condicionada $P(M_k|\mathbf{y})$ ⁷ de cada modelo. Esta informação será utilizada em análise subsequente.

A análise descritiva dos valores de \bar{R}^2 dos modelos ajustados mostra que:

- (i) Para o cultivar C-111, 75% dos 256 modelos de regressão linear clássica sem interação

⁷Calculada através da expressão (10), seção 3.2.

Tabela 6. Estimativas de mínimos quadrados dos coeficientes de regressão dos modelos selecionados através do critério \bar{R}^2 - Cultivar C-111

M_k	Coeficientes									\bar{R}^2
	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	
1	3,50	1,25								0,9320
2	5,51		0,14							0,7548
3	-0,48	1,11				3,96				0,9579
4	117,67	1,02		-1,56						0,9447
5	77,38	0,97		-1,06		3,43				0,9630
6	-0,97	1,04				4,25			0,01	0,9588
7	151,13	0,94		-1,92		2,81	-1,18			0,9638
8	-0,55	0,82	0,02			3,41			0,03	0,9629
9	159,01	0,89		-1,98		2,98	-1,55		0,01	0,9642
10	-9,68	0,76	0,02			4,25		1,82	0,04	0,9638
11	278,01	0,87		-3,09	-2,23	2,66	-2,63		0,02	0,9656
12	475,09	0,89		-5,06	-3,77		-4,66	-2,96	0,02	0,9651
13	247,23	0,78	0,01	-2,67	-2,33	2,47	-2,45		0,03	0,9656
14	360,09	0,88		-3,90	-2,89	1,62	-3,47	-1,44	0,02	0,9650
15	305,08	0,80	0,01	-3,26	-2,75	1,82	-3,03	-0,93	0,03	0,9647

que podem ser propostos ao ajuste dos dados possuem \bar{R}^2 acima de 0,91;

- (ii) Para o cultivar Piranão, 75% dos 256 modelos possuem \bar{R}^2 acima de 0,90, enquanto que;
- (iii) Para o cultivar Flint Composto, 75% possuem \bar{R}^2 acima de 0,87.

Com base nesses resultados, verifica-se uma distribuição bem concentrada dos valores de \bar{R}^2 . Esta concentração pode gerar uma região de incerteza na qual o pesquisador pode ter dificuldades em selecionar um único modelo baseando-se apenas no caráter objetivo dos critérios de seleção.

As diferenças entre as estimativas, mesmo quando não comprovadas estatisticamente, podem representar, na prática, diferentes e importantes decisões, provocando, no responsável pela tomada de decisão, uma indecisão diante do modelo a ser escolhido. Uma alternativa para contornar este problema é obter estimativas para cada coeficiente de regressão por meio de uma ponderação de Bayesiana de modelos, tendo como objetivo a “acomodação” dessas variações entre as estimativas de um mesmo coeficiente.

Por exemplo, para os dados do cultivar Flint Composto, cada coeficiente pode ser estimado pela sua esperança a *posteriori*, obtida pela ponderação da sua esperança a *posteriori* sob:

- (i) Cada um dos 256 modelos;

Tabela 7. Estimativas de mínimos quadrados dos coeficientes de regressão dos modelos selecionados através do critério \bar{R}^2 - Cultivar Piranão

M_k	Coeficientes									\bar{R}^2
	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	
1	3,82	1,06								0,8746
2	5,08		0,11							0,8132
3	-50,10	0,69			4,40					0,9474
4	81,39	0,86		-1,12						0,9201
5	-44,42	0,54	0,02		3,93					0,9544
6	-17,62	0,69		-0,34	3,68					0,9484
7	-34,86	0,48	0,03		3,15				0,02	0,9551
8	-21,17	0,54	0,02	-0,24	3,44					0,9544
9	-17,24	0,50	0,02	-0,20	2,84				0,02	0,9546
10	-39,16	0,47	0,03		3,28			0,56	0,03	0,9542
11	-17,65	0,49	0,02	-0,20	2,92	-0,23			0,02	0,9532
12	-0,16	0,50	0,02	-0,37	2,69		-0,23		0,02	0,9531
13	43,63	0,48	0,02	-0,86	2,55	-0,79	-0,84		0,03	0,9524
14	276,45	0,51	0,02	-3,12		-3,20	-3,04	-2,95	0,02	0,9523
15	157,23	0,50	0,02	-1,97	1,46	-2,08	-1,91	-1,64	0,02	0,9517

(ii) Os modelos contidos na Tabela 8, ou;

(iii) Os modelos selecionados por outros métodos. Nesse trabalho será utilizado o método *Occam's Window*.

Tabela 8. Estimativas de mínimos quadrados dos coeficientes de regressão dos modelos selecionados através do critério \bar{R}^2 - Cultivar Flint Composto

M_k	Coeficientes									\bar{R}^2	$P(M_k \mathbf{y})(\%)$
	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8		
1	4,27	1,07								0,8367	<0,00001
2	-95,78				8,30					0,8099	<0,00001
3	-52,71	0,64			4,60					0,9566	39,36
4	146,04	0,66		-2,08						0,9207	0,0004
5	-49,05	0,71			4,37				-0,02	0,9592	14,19
6	-49,78	0,57	0,01		4,36					0,9570	12,35
7	-49,97	0,71			3,96		0,47		-0,03	0,9598	3,39
8	-35,62	0,78			3,74			-1,02	0,04	0,9598	2,15
9	-29,34	0,72	0,01		3,35			-1,35	-0,03	0,9599	2,15
10	-95,00	0,74		0,51	4,48		0,76		-0,03	0,9596	0,21
11	-113,01	0,67	0,01	0,74	4,53		0,89		-0,03	0,9595	0,06
12	-33,70	0,70	0,01		3,31		0,26	-1,07	-0,03	0,9591	0,09
13	-1003,00	0,74		9,57	13,53	9,25	9,93	9,07	-0,04	0,9605	<0,00001
14	-165,08	0,70	0,01	1,35	4,63	1,08	1,50		-0,04	0,9589	0,0051
15	-1301,00	0,63	0,01	12,63	16,17	12,16	12,89	11,60	-0,04	0,9628	<0,00001

Os resultados desta ponderação utilizando os modelos da Tabela 8 podem ser vistos na Tabela 9, onde são apresentadas as estimativas de mínimos quadrados de cada um dos coeficientes de regressão, seus respectivos erros padrão (SE) e seus respectivos p -valores, associados ao teste de hipótese acerca da significância das variáveis explicativas.

Além desses resultados, obtidos pelo procedimento clássico de análise de dados, a Tabela 9 apresenta também as estimativas da esperança e variância a *posteriori* ponderada de cada coeficiente de regressão e a probabilidade $P(\beta_j \neq 0|\mathbf{y})$ calculada de duas formas: utilizando os modelos da Tabela 8, e utilizando todos os 256 modelos, identificado por $P'(\beta_j \neq 0|\mathbf{y})$. Para β_5 por exemplo, sua esperança, variância a *posteriori* ponderada e sua probabilidade a *posteriori*⁸ de ser diferente de zero são dadas respectivamente por:

$$E(\beta_5|\sigma^2, \mathbf{y}) = \sum_{k=13}^{15} E(\beta_5|M_k, \sigma^2, \mathbf{y})p_{M_k},$$

sendo $E(\beta_5|M_k, \sigma^2, \mathbf{y})$ a média a *posteriori* de β_5 sob o modelo M_k cuja distribuição é dada pela expressão (10),

$$\begin{aligned} V(\beta_5|\mathbf{y}) &= \sum_{k=13}^{15} V(\beta_5|M_k, \sigma^2, \mathbf{y})p_{M_k} \\ &+ \sum_{k=13}^{15} [E(\beta_5|M_k, \sigma^2, \mathbf{y}) - E(\beta_5|\sigma^2, \mathbf{y})]^2 p_{M_k}, \end{aligned}$$

em que

$$p_{M_k} = \frac{P(M_k|\mathbf{y})}{\sum_{l=13}^{15} P(M_l|\mathbf{y})},$$

$$P(\beta_j \neq 0|\mathbf{y}) = \sum_{i=13}^{15} P(M_i|\mathbf{y}),$$

e

$$P'(\beta_j \neq 0|\mathbf{y}) = \sum_{M_k: \beta_j \in M_k} P(M_k|\mathbf{y}),$$

sendo que o cálculo de $P'(\beta_j \neq 0|\mathbf{y})$ é feito através do programa apresentado no Anexo C.

Pode-se observar que os resultados apresentados pelas duas formas de análise apresentam algumas concordâncias. Os coeficientes de maior significância segundo a análise clássica são β_0 (intercepto), β_1 e β_4 , exatamente os coeficientes que apresentaram as maiores

⁸Comentários acerca desta probabilidade podem ser encontrados na seção 2.3.1, p.27.

Tabela 9. Análises clássica e Bayesiana dos dados do cultivar Flint Composto

	Análise clássica			BMA			
	$\hat{\beta}$	SE	p	$E(\beta \sigma^2, \mathbf{y})$	$V(\beta \sigma^2, \mathbf{y})(\%)$	$P(\beta \neq 0 \mathbf{y})(\%)$	$P'(\beta \neq 0 \mathbf{y})(\%)$
Intercepto	-1301,00	580,70	0,03	-50,31	50,58	73,95	100
β_1	0,63	0,11	$1,07e^{-5}$	0,64	0,006	73,94	100
β_2	0,01	0,01	0,10	0,01	0,0001	14,64	22,66
β_3	12,63	5,80	0,03	0,12	1,39	0,26	4,97
β_4	16,17	5,89	0,01	4,41	0,23	73,94	99,97
β_5	12,16	5,76	0,04	1,07	0,69	0,0051	9,17
β_6	12,89	5,83	0,03	0,49	0,13	3,75	13,16
β_7	11,60	5,86	0,05	-1,17	0,34	4,39	8,37
β_8	-0,04	0,01	0,04	-0,02	0,0002	22,24	28,06

probabilidades a *posteriori*. Todos os demais coeficientes apresentaram $P(\beta_j \neq 0|\mathbf{y}) < 0,5$, resultado que, conforme Viallefont et al. (2001) pode indicar a não significância das variáveis associadas a estes coeficientes, enquanto que pela análise clássica, esses coeficientes podem ser considerados significativos pelo fato de seus p -valores associados estarem abaixo do valor usualmente utilizado de 0,05. O coeficiente β_2 foi o único coeficiente a apresentar p -valor acima de 0,05, além disso, sua probabilidade a *posteriori* também indica que não há evidências para considerá-lo significativo.

Levando-se em conta a alta multicolinearidade presente nos dados, algumas variáveis explicativas podem ser irrelevantes à explicação de Y , os valores obtidos pelo cálculo de $P(\beta_j \neq 0|\mathbf{y})$ parecem ser mais coerentes que os resultados obtidos pelos p -valores no sentido de identificar com maior eficiência as variáveis mais importantes. Com base na análise clássica, vários modelos poderiam ser propostos ao pesquisador, dentre eles M_3 e M_4 (Tabela 8), enquanto que com base na análise Bayesiana pode ser identificado com maior clareza um único modelo:

$$Y_i = -50,31 + 0,64X_1 + 4,41X_4. \quad (12)$$

Como já comentado anteriormente, a ponderação poderia ter sido realizada levando-se em conta todos os 256 modelos ou outro grupo de modelos, escolhidos com base em outro método, como por exemplo o *Occam's Window* (OW). De acordo com Madigan & Raftery (1994) o *Occam's Window* é um método eficiente para identificar os modelos que melhor se ajustam aos dados. Os modelos identificados pelo OW podem ser utilizados na ponderação Bayesiana dos coeficientes de regressão evitando com isso que modelos super-

parametrizados sejam levados em consideração. No OW, o rigor no processo de descarte dos modelos pode ser regulado pelos valores definidos para a e O_R (respectivamente primeira e segunda regra do OW) de modo que quanto maior forem a e O_R maior será o rigor no processo de descarte. A Tabela 10 apresenta os modelos selecionados pelo OW⁹ simétrico considerando três valores diferentes para a : 5, 20 (sugestão de Madigan & Raftery (1994)) e 35.

Tabela 10. Modelos selecionados pelo *Occam's Window simétrico*, executado com diferentes valores de a

M_k	Coeficientes								$P(M_k \mathbf{y})(\%)$	
	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7		β_8
1 ⁽¹⁾⁽²⁾⁽³⁾	-52,71	0,64			4,60					39,36
2 ⁽¹⁾⁽²⁾⁽³⁾	-49,78	0,57	0,01		4,36					12,35
3 ⁽¹⁾⁽²⁾	-40,96	0,63		-0,13	4,39					2,00
4 ⁽¹⁾⁽²⁾	-53,77	0,63			4,76	-0,68				3,80
5 ⁽¹⁾	-50,67	0,55	0,01		4,52	-0,82				1,51
6 ⁽¹⁾⁽²⁾	-53,58	0,63			4,40		0,24			4,75
7 ⁽¹⁾	-50,62	0,55	0,01		4,11		0,29			1,71
8 ⁽¹⁾⁽²⁾	-54,10	0,64			4,66			0,12		2,48
9 ⁽¹⁾⁽²⁾⁽³⁾	-49,05	0,71			4,37				-0,02	14,19
10 ⁽¹⁾⁽²⁾	-47,99	0,67	0,005		4,28				-0,02	2,70
11 ⁽¹⁾	-50,11	0,71			4,54	-0,69			-0,02	1,45
12 ⁽¹⁾⁽²⁾	-49,97	0,71			3,96		0,47		-0,03	3,39
13 ⁽¹⁾⁽²⁾	-35,62	0,78			3,74			-1,02	-0,04	2,15

(1) Modelos selecionados pelo OW com $a=35$.

(2) Modelos selecionados pelo OW com $a=20$.

(3) Modelos selecionados pelo OW com $a=5$.

Observa-se que dos 256 modelos, 13 foram selecionados pelo OW simétrico com $a=35$, já pelo OW com $a=20$, 10 modelos foram selecionados enquanto que pelo OW com $a=5$ apenas 3 modelos foram selecionados. A Tabela 11 apresenta as esperanças, as variâncias e os valores de $P(\beta_j \neq 0|\mathbf{y})$ para cada coeficiente de regressão obtidos pela ponderação dos modelos selecionados pelos três diferentes valores de a .

⁹O programa que executa o OW encontra-se no Anexo C.

Tabela 11. Esperança e variância a *posteriori* de cada coeficiente obtida pela ponderação dos modelos selecionados pelo OW para cada valor de a

	$OW_{a=5}$			$OW_{a=20}$			$OW_{a=35}$		
	$E(\beta \sigma^2, \mathbf{y})$	$V(\beta \sigma^2, \mathbf{y})$	$P(\beta \neq 0 \mathbf{y})$	$E(\beta \sigma^2, \mathbf{y})$	$V(\beta \sigma^2, \mathbf{y})$	$P(\beta \neq 0 \mathbf{y})$	$E(\beta \sigma^2, \mathbf{y})$	$V(\beta \sigma^2, \mathbf{y})$	$P(\beta \neq 0 \mathbf{y})$
Intercepto	-51,29	27,13	65,9	-50,81	38,47	87,17	-50,79	37,84	91,84
β_1	0,63	0,05	65,9	0,64	0,005	87,17	0,64	0,006	91,84
β_2	0,01	$8,26e^{-5}$	12,35	0,009	$8,55e^{-5}$	15,05	0,009	$8,49e^{-5}$	18,27
β_3	0	-0,13	0,01	2,0	-0,13	0,01	2,0
β_4	4,49	0,16	65,9	4,45	0,20	87,17	4,45	0,20	91,84
β_5	0	-0,66	0,56	3,80	-0,69	0,54	6,76
β_6	0	0,35	0,14	8,14	0,34	0,13	9,85
β_7	0	-0,41	0,62	4,63	-0,41	0,62	4,63
β_8	-0,02	0,0002	14,19	-0,02	0,0002	22,43	-0,02	0,0002	23,88

Alguns aspectos acerca dos resultados da Tabela 11 merecem ser explorados, como por exemplo o fato da taxa de aumento de $P(\beta_j \neq 0|\mathbf{y})$ em função do aumento de a para os coeficientes β_0 , β_1 e β_4 ser 2,67 vezes maior que a taxa dos demais coeficientes, o que significa que as variáveis mais importantes são logo identificadas para pequenos valores de a e que parece não haver riscos de se concluir acerca da relevância de uma determinada variável quando na realidade as evidências indicariam o contrário se valores maiores para a fossem adotados. Outro resultado observado foi a obtenção de variâncias menores para β_0 e β_4 quando $a=5$, indicando que a variância dos coeficientes pode ser reduzida se a ponderação for realizada com uma quantidade menor de modelos mais plausíveis diante da amostra obtida. Os coeficientes β_0 , β_1 e β_4 foram os que apresentaram os maiores valores de $P(\beta \neq 0|\mathbf{y})$, logo, os modelos propostos ao pesquisador após a ponderação Bayesiana dos coeficientes de regressão com base no OW com $a=5$, 20 e 35 são, respectivamente:

$$Y_i = -51,29 + 0,63X_1 + 4,49X_4,$$

$$Y_i = -50,81 + 0,64X_1 + 4,45X_4$$

e

$$Y_i = -50,79 + 0,64X_1 + 4,45X_4.$$

Observando os valores dos coeficientes desses modelos, os valores dos coeficientes do modelo dado pela expressão (12) e as estimativas de β_0 , β_1 e β_4 apresentadas na Tabela 8 pode-se dizer que a aplicação da ponderação Bayesiana permitiu incorporar a incerteza envolvida no processo de escolha de um modelo.

4.1 Estudo de Simulação

A comparação entre o desempenho preditivo da ponderação Bayesiana e o desempenho preditivo dos critérios de seleção R^2 , \overline{R}^2 , C_p e $PRESS$ foi realizada com amostras simuladas, obtidas a partir de medidas descritivas dos dados de cada cultivar. A Tabela 12 e a Figura 3 apresentam, respectivamente, as medidas que resumem o comportamento dos valores do número de condição k^* e o boxplot dos valores de k^* encontrados nas 1000 amostras simuladas para cada uma das 9 combinações entre tamanho amostral e cultivar. Estas combinações são dadas por:

- (i) Combinação 1: $\bar{\mathbf{x}}$, \mathbf{S} , $\hat{\boldsymbol{\beta}}$ e $QMRes$ obtidos a partir dos dados do cultivar C-111, e tamanho amostral $n=50$;
- (ii) Combinação 2: $\bar{\mathbf{x}}$, \mathbf{S} , $\hat{\boldsymbol{\beta}}$ e $QMRes$ obtidos a partir dos dados do cultivar C-111, e tamanho amostral $n=75$;
- (iii) Combinação 3: $\bar{\mathbf{x}}$, \mathbf{S} , $\hat{\boldsymbol{\beta}}$ e $QMRes$ obtidos a partir dos dados do cultivar C-111, e tamanho amostral $n=100$;
- (iv) Combinação 4: $\bar{\mathbf{x}}$, \mathbf{S} , $\hat{\boldsymbol{\beta}}$ e $QMRes$ obtidos a partir dos dados do cultivar Piranão, e tamanho amostral $n=50$;
- (v) Combinação 5: $\bar{\mathbf{x}}$, \mathbf{S} , $\hat{\boldsymbol{\beta}}$ e $QMRes$ obtidos a partir dos dados do cultivar Piranão, e tamanho amostral $n=75$;
- (vi) Combinação 6: $\bar{\mathbf{x}}$, \mathbf{S} , $\hat{\boldsymbol{\beta}}$ e $QMRes$ obtidos a partir dos dados do cultivar Piranão, e tamanho amostral $n=100$;
- (vii) Combinação 7: $\bar{\mathbf{x}}$, \mathbf{S} , $\hat{\boldsymbol{\beta}}$ e $QMRes$ obtidos a partir dos dados do cultivar Flint Composto, e tamanho amostral $n=50$;
- (viii) Combinação 8: $\bar{\mathbf{x}}$, \mathbf{S} , $\hat{\boldsymbol{\beta}}$ e $QMRes$ obtidos a partir dos dados do cultivar Flint Composto, e tamanho amostral $n=75$;
- (ix) Combinação 9: $\bar{\mathbf{x}}$, \mathbf{S} , $\hat{\boldsymbol{\beta}}$ e $QMRes$ obtidos a partir dos dados do cultivar Flint Composto, e tamanho amostral $n=100$;

Tabela 12. Resumo descritivo para o número de condição (k^*) das amostras geradas para cada combinação tamanho amostral e cultivar

Estatística	C-111			Piranão			F. Composto		
	50	75	100	50	75	100	50	75	100
Mínimo	171,9	164,1	191,5	598,8	545,9	611,4	4262,0	5657,0	6808,0
1ºQuartil	378,0	374,1	367,2	1186,0	1164,0	1152,0	12050,0	12120,0	11960,0
Mediana	463,3	440,8	427,7	1448,0	1366,0	1334,0	14970,0	14130,0	13860,0
Média	491,1	453,2	435,7	1526,0	1411,0	1364,0	15810,0	14660,0	14150,0
3ºQuartil	572,5	518,5	490,3	1795,0	1621,0	1540,0	18450,0	16810,0	16010,0
Máximo	1367,0	929,8	917,7	4120,0	2952,0	2802,0	42100,0	29590,0	29120,0
Desvio-Padrão	163,1	113,5	94,6	494,3	352,8	299,7	5325,5	3676,1	3105,5
C.V. (em %)	33,2	25,0	21,7	32,3	25,0	21,9	33,6	25,0	21,9
Combinação	1	2	3	4	5	6	7	8	9

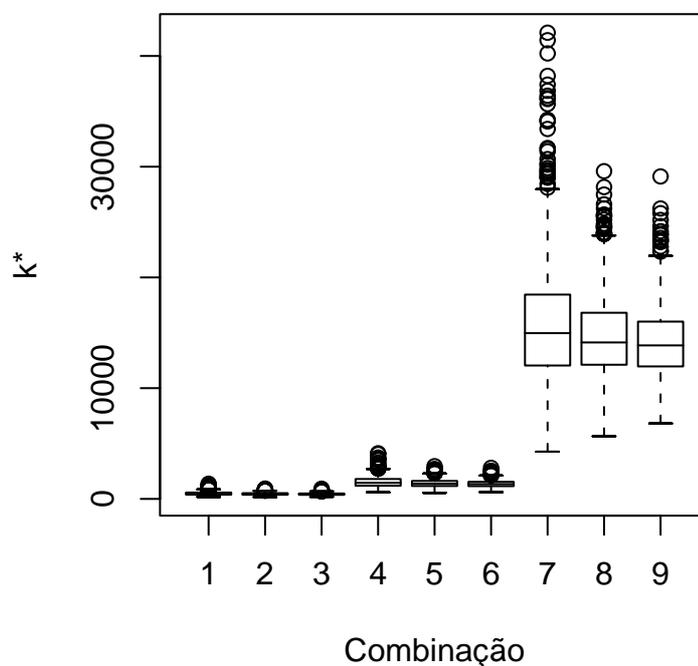


Figura 3 - Boxplot para o número de condição das amostras simuladas para cada combinação entre cultivar e tamanho amostral

As 9000 amostras produziram uma amplitude considerável dos valores de k , propiciando uma boa cobertura para as conclusões acerca do desempenho preditivo do BMA.

As Tabelas 13, 14 e 15 apresentam os resumos descritivos dos valores do LEP para cada uma das combinações. Esses resultados são representados graficamente através da Figura 4, que mostra o comportamento do Logaritmo do Escore Preditivo (LEP) de cada metodologia em função do grau de multicolinearidade e do tamanho amostral. Os resultados empíricos mostram que, para as amostras de tamanho 50, o desempenho preditivo do BMA parece melhorar a medida que o grau de multicolinearidade dos dados aumenta (1a), enquanto que para as amostras de tamanho 75 e 100, o desempenho preditivo dos critérios de seleção se equipararam com o desempenho do BMA (1b e 1c). O mesmo comportamento pode ser observado fixando-se o grau de multicolinearidade e variando o tamanho amostral (2a, 2b e 2c).

Tabela 13. Resumo dos valores do LEP nas amostras simuladas (cultivar C-111)

n	Metodologia	Mínimo	1 ^o Quartil	Mediana	Média	3 ^o Quartil	Máximo	Variância
50	<i>BMA</i>	42,46	51,75	54,57	55,07	57,60	89,05	26,83
	R^2	41,29	51,41	55,09	55,97	59,11	90,31	40,87
	\bar{R}^2	40,99	51,98	55,91	57,03	61,03	94,27	51,40
	<i>PRESS</i>	41,05	51,89	55,83	56,97	60,98	94,27	50,82
	C_p	40,99	52,40	56,24	57,33	61,12	90,06	49,73
	<i>BMA</i>	40,29	49,75	52,33	52,67	55,10	70,81	17,39
75	R^2	39,22	49,13	51,64	52,13	54,50	74,44	19,62
	\bar{R}^2	39,73	49,34	52,00	52,58	55,20	74,17	21,85
	<i>PRESS</i>	39,77	49,30	52,04	52,59	55,24	74,17	22,12
	C_p	39,99	49,62	52,30	52,98	55,56	75,15	23,99
	<i>BMA</i>	41,82	48,79	51,30	51,50	53,87	77,32	15,71
	R^2	39,94	47,81	50,25	50,62	52,99	72,70	16,62
100	\bar{R}^2	39,87	47,97	50,60	50,87	53,48	74,69	17,80
	<i>PRESS</i>	39,87	47,94	50,58	50,87	53,45	74,69	17,78
	C_p	39,87	48,15	50,86	51,14	53,61	81,82	18,64

Tabela 14. Resumo dos valores do LEP nas amostras simuladas (cultivar Piranão)

n	Metodologia	Mínimo	1 ^o Quartil	Mediana	Média	3 ^o Quartil	Máximo	Variância
50	<i>BMA</i>	39,55	49,26	52,05	52,92	56,02	88,83	31,31
	R^2	39,94	49,80	53,49	54,21	57,57	98,62	39,32
	\bar{R}^2	39,54	50,25	54,19	55,30	59,05	102,30	54,59
	<i>PRESS</i>	39,35	49,97	54,07	55,20	58,86	102,30	54,66
	C_p	39,35	50,17	54,40	55,54	59,22	98,80	57,67
75	<i>BMA</i>	38,71	47,32	49,98	50,25	52,53	66,99	16,32
	R^2	37,38	47,44	50,01	50,49	52,91	70,37	19,70
	\bar{R}^2	37,76	47,61	50,29	50,80	53,33	70,64	22,21
	<i>PRESS</i>	37,76	47,72	50,34	50,83	53,25	70,64	21,95
	C_p	37,85	47,72	50,45	50,91	53,42	70,64	22,23
100	<i>BMA</i>	39,59	46,40	48,95	49,11	51,32	64,63	14,68
	R^2	38,33	46,33	48,89	49,17	51,68	70,65	17,03
	\bar{R}^2	38,33	46,42	49,11	49,40	51,95	72,72	18,16
	<i>PRESS</i>	38,33	46,44	49,18	49,38	51,89	72,72	17,88
	C_p	38,85	46,52	49,42	49,59	52,26	72,72	18,30

Tabela 15. Resumo dos valores do LEP nas amostras simuladas (cultivar Flint Composto)

n	Metodologia	Mínimo	1 ^o Quartil	Mediana	Média	3 ^o Quartil	Máximo	Variância
50	<i>BMA</i>	36,84	45,36	48,03	48,82	51,37	82,50	27,69
	R^2	38,32	48,85	52,20	53,30	57,21	81,44	38,88
	\bar{R}^2	35,30	48,77	53,34	53,81	58,01	89,94	50,88
	<i>PRESS</i>	35,30	48,49	53,30	53,75	58,06	89,94	52,75
	C_p	36,93	47,04	51,98	53,00	58,01	93,95	60,03
75	<i>BMA</i>	37,52	44,02	46,28	46,76	49,12	69,66	15,91
	R^2	36,24	45,29	47,20	47,35	49,28	70,45	10,70
	\bar{R}^2	35,07	45,50	47,40	47,65	49,74	74,73	12,45
	<i>PRESS</i>	35,55	45,49	47,10	47,63	49,67	72,21	12,21
	C_p	34,59	45,76	47,83	48,18	50,39	74,73	16,67
100	<i>BMA</i>	35,06	43,00	45,48	45,74	48,07	67,49	14,91
	R^2	36,01	43,26	45,20	45,25	47,13	60,61	8,56
	\bar{R}^2	35,96	43,39	45,34	45,38	47,21	63,14	9,22
	<i>PRESS</i>	35,96	43,39	45,34	45,39	47,21	63,14	9,36
	C_p	36,01	43,60	45,66	45,74	47,56	69,94	11,37

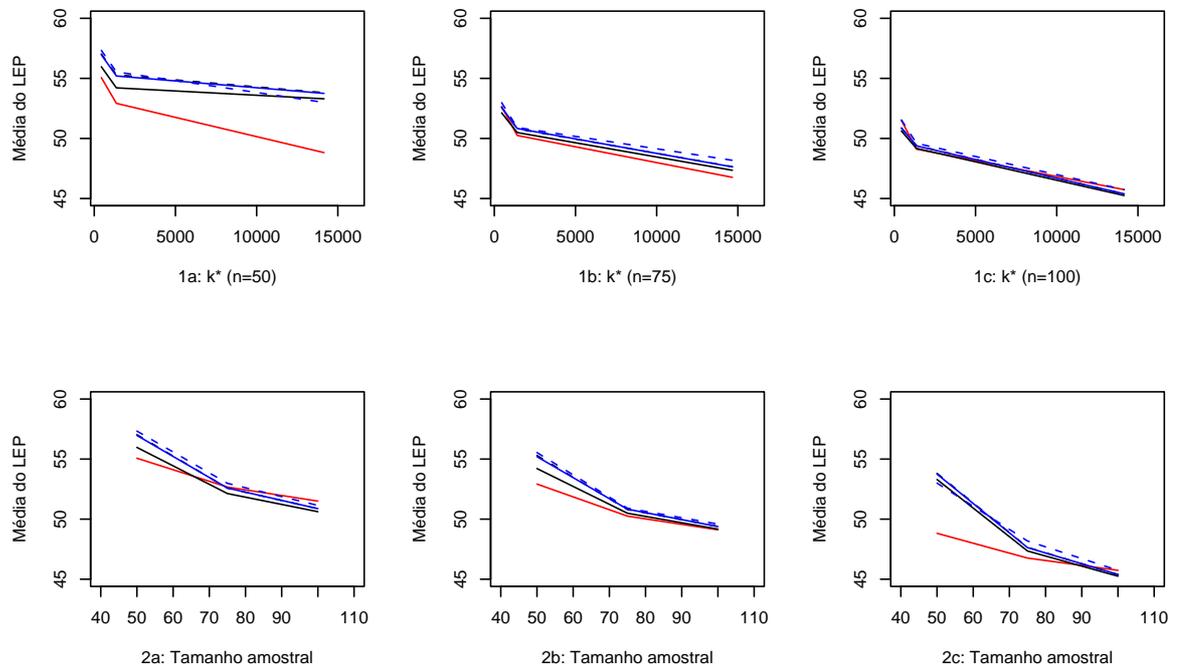


Figura 4 - Comportamento da média do Logaritmo do Escore Preditivo da metodologia BMA (vermelho) e dos critérios: R^2 (preto), \bar{R}^2 (preto tracejado), $PRESS$ (azul) e C_p (azul Tracejado)

5 CONCLUSÕES

A aplicação da metodologia de ponderação de modelos sob enfoque Bayesiano para a análise do conjunto de dados do cultivar Flint Composto permitiu que incerteza diante da escolha de um modelo fosse de fato incorporada. A aplicação também mostrou que o processo de seleção dos melhores modelos a serem levados em conta na ponderação também pode ser realizado utilizando critérios usuais de seleção, como por exemplo o \bar{R}^2 , pois ao contrário do *Occam's Window* que precisa ser programado, os critérios usuais de seleção estão disponíveis para uso na maioria dos softwares estatísticos.

Com relação ao desempenho preditivo da ponderação Bayesiana, os resultados empíricos indicaram que esta metodologia apresenta desempenho semelhante aos métodos usuais de seleção de modelos nas situações de multicolinearidade exploradas neste trabalho.

Pretende-se, futuramente, pesquisar alternativas que permitam aprimorar o processo de descarte de modelos promovido pelo *Occam's Window* (OW). Seria interessante incorporar, na execução do OW, as conseqüências políticas e gerências causadas pela inclusão de uma específica variável. Embora o OW utilize, durante sua execução, a probabilidade a *posteriori* de cada modelo condicionada aos dados, $P(M_k|\mathbf{y})$, cuja informação subjetiva do pesquisador a respeito de cada modelo é levada em conta no seu cálculo, seria interessante incrementar o OW com informações acerca do custo operacional de cada variável e as implicações causadas pela inclusão dessas no modelo final.

REFERÊNCIAS BIBLIOGRÁFICAS

- AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In: INTERNATIONAL SYMPOSIUM OF INFORMATION THEORY, 2., Budapest, 1973. **Proceedings**. Budapest: Akademiai Kiado, 1973. p.267-281.
- BEATON, A.E. **The use of special matrix operators in statistical calculus**, Princeton, New Jersey. Education Testing Service, 1964. 1v. (Res. Bull RB, 64-51).
- BUCKLAND, S.T.; BURNHAM, K.P.; AUGUSTIN, N.H. Model selection: An integral part of inference. **Biometrics**, v.53, n.2, p.603-618, Jun. 1997.
- CANDOLO, C. A incorporação da incerteza devido à escolha de modelos na inferência estatística com aplicação em modelos de regressão linear. Piracicaba, 2001. 80p. Tese(Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- CANDOLO, C. Um estudo de simulação considerando a incorporação da incerteza na seleção de modelos em regressão linear. **Revista de Matemática e Estatística**, v.20, n.2, p.111-126, 2004.
- CHATTERJEE, S.; PRICE, B. **Regression analysis by example**. New York: John Wiley, 1977. 228p.
- EFROYMSON, M.A. **Multiple regression analysis. Mathematical methods for digital computers**. Edit. A. Ralston; A.S. Wiley. (Ed). New York; Wiley, 1960. 1v
- FONTES, L.S.; ALMEIDA FILHO, A.J.de; ARTHUR, V. Danos causados por *Sitophilus Oryzae* (Linné, 1763) e *Sitophilus Zeanaeis* Motschulsky. 1855(Coleoptera: Curculionidal) em cultivares de arroz (*Oryza Sativa* L.). **Arquivos do Instituto Biológico**, v.70, n.3, p.303-307, 2003.

- FURNIVAL, G.M. All possible regressions with less computation. **Technometrics**, v.13, p.403-408, 1971.
- FURNIVAL, G.M.; WILSON, R.W. Regression by leaps and bounds. **Technometrics**, v.16, p.499-512, 1974.
- GALLO, D. ;NAKANO, O.; SILVEIRA NETO, S. et al.**Manual de entomologia agrícola**. São Paulo: Agronômica Ceres, 1988. 649p.
- GARSIDE, M.J. The best subset in multiple regression analysis. **Applied Statistics**, v.14, n.2, p.196-200, 1965.
- GELFAND, A.E.; GHOSH, S.K. Model choice: a minimum poeriori predictive loss approach. **Biometrika**, v.85, n.1, p.1-11, 1998.
- GEORGE, E.I.; McCULLOCH, R.E. Variable selection via gibbs sampling. **Journal of the American Statistical Association**, v.88, n.423, p.881-889, Sep. 1993.
- GOOD, I.J. Rational decisions. **Journal of the Royal Statistical Society**, v.14, n.1, p.107-114, 1952.
- HOCKING, R.R. The analysis and selection of variables in linear regression. **Biometrics**, v.32, n.1, p.1-49, Mar. 1976.
- HOETING, J.A. Accounting for model uncertainty in linear regression. Washington, 1994. 170p. Thesis(Ph.D.) - University of Washington.
- LAUD, P.W.; IBRAHIM, J.G. Predictive specification of priori model probabilities in variable selection. **Biometrika**, v.83, n.2, p.267-274, Jun. 1996.
- LEAMER, E.E. **Specification searches**. New York: John Wiley, 1978. 370p.
- LEE III, H.K.H. Model selection and model averaging for neural networks. Pittsburg, 1999. 113p. Thesis(Ph.D.) - Carnegie Mellon University.
- LYLE, P. **Regression analysis of production costs and factory operations**. New York: Publishing, 1957. 204p.

- 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, n.428, p.1535-1546, Dec. 1994.
- MADIGAN, D.; YORK, J. Bayesian graphical models for discrete data. **International Statistical Review**, v.63, p.215-232, 1995.
- MATSUO, T. O uso da regressão de cumeieira em experimentos agrônômicos. Piracicaba, 1986. 89p. Dissertação(Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- MONTGOMERY, D.C.; PECK, E.A. **Introduction to linear regression analysis**. New York: John Wiley, 1992. 527p.
- NELDER, J.A.; WEDDERBURN, R.W.M. Generalized linear models. **Journal of the Royal Statistical Society**, v.135, n.3, p.370-384, 1972.
- NETER, J.; WASSERMAN, W.; KUTNER, M.H. **Applied linear regression models**. Homewood: Irwin, 1983. 547p.
- NEWTON, R.G.; SPURRELL, D.J. A development of multiple regression for the analysis of routine data. **Applied Statistics**, v.16, n.1, p.51-65, 1967.
- RAFTERY, A.E. Bayesian model selection in social research. **Sociological Methodology**, v.25, p.111-163, 1995.
- RAFTERY, A.E.; MADIGAN, D.; HOETING, J.A. Bayesian model averaging for linear regression models. **Journal of the American Statistical Association**, v.92, n.437, p.179-191, Mar. 1997.
- RAO, P. Some notes on misspecification in multiple regressions. **The American Statistician**, v.25, n.5, p.37-39, Dec. 1971.
- R FOUNDATION FOR STATISTICAL COMPUTING. **R 2.0.1**, 2004. Disponível em: <<http://www.cran.r-project.org>>.
- SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, v.6, n.2, p.461-464, Mar. 1978.

- SNEDECOR, W.G. **Statistical methods**. Ames: Iowa State College Press, 1956. 534p.
- TVERSKY, A. Assessing uncertainty. **Journal of the Royal Statistical Society**, v.36, n.2, p.148-159, 1974.
- VIALLEFONT, V.; RAFTERY, A.E.; RICHARDSON, S. Variable selection and bayesian model averaging in case-control studies. **Statistical in Medicine**, v.20, p.3215-3230, Mar. 2001.
- WETHERILL, G.B.; DUNCOMBE, P.; KENWARD, M.; et al. **Regression analysis with applications**. London: Chapman & Hall, 1986. 311p.

ANEXOS

ANEXO A - Programa em *R* v. 2.0.1 utilizado para obter o Logaritmo do Escore Preditivo (LEP) dos critérios de seleção de modelos.

```

rm(list=ls(all=TRUE)) require(mvtnorm)
#dataname="C-111.txt","Piranão.txt" ou "Flint Composto.txt" #
n=50,75 ou 100 dpred<-function(dataname,n,sem){
d<-read.table(dataname,header=T) attach(d)
matx<-cbind(x1,x2,x3,x4,x5,x6,x7,x8) mod<-summary(lm(y~matx))
coef<-mod$coefficients[,1]
med<-c(mean(x1),mean(x2),mean(x3),mean(x4),mean(x5),mean(x6),mean(x7),mean(x8))
vc<-cov(matx) set.seed(sem) datx<-rmvnorm(n,med,vc)
dat<-cbind(1,datx) e<-rnorm(n,0,mod$sigma) y<-rep(0,n) for(i in1:n){
y[i]<-sum(coef*dat[i,])+e[i]} Xg<-cbind(y,datx)
s<-sample(1:n,(n-25),FALSE) yt<-Xg[s,] yp<-Xg[-s,] phi<-2.85
s2<-c(var(yt[,1]),(phi^2)*(1/var(yt[,2])),(phi^2)*(1/var(yt[,3])),(phi^2)*
(1/var(yt[,4])),(phi^2)*(1/var(yt[,5])),(phi^2)*(1/var(yt[,6])),(phi^2)*
(1/var(yt[,7])),(phi^2)*(1/var(yt[,8])),(phi^2)*(1/var(yt[,9])))
p<-ncol(datx) w<-NULL for(j in 1:p)
w<-rbind(cbind(w,FALSE),cbind(w,TRUE)) w<-cbind(TRUE,w) k<-1:(p+1)
v<-2.58 n<-nrow(yt) vp<-n+v lamb<-0.28
K<-(gamma((vp+1)/2)/(gamma(1/2)*gamma(vp/2)))*vp^(vp/2)
b<-matrix(0,nrow(w),ncol(w)) X<-cbind(1,yt[,-1])
xx0<-cbind(1,yp[,-1]) Mf<-matrix(0,nrow(yp),nrow(w))
r2<-rep(0,nrow(w)) r2aj<-rep(0,nrow(w)) press<-rep(0,nrow(w))
cp<-rep(0,nrow(w)) c<-sum(yt[,1])^2/nrow(X) ve<-(mod$sigma)^2

for(j in 1:nrow(w)){
ind<-k[w[j,]]
X0<-X[,ind,drop=F]
r2[j]<-((t(solve(t(X0)%*%X0)%*%t(X0)%*%yt[,1]))%*%t(X0)%*%
yt[,1])-c)/((t(yt[,1])%*%yt[,1])-c)
r2aj[j]<-1-(((nrow(X0)-1)*(1-r2[j]))/(nrow(X0)-nrow(solve(t(X0)%*%X0)%*%
t(X0)%*%yt[,1])))

```

```

cp[j]<-(((t(yt[,1]))**yt[,1])-(t(solve(t(X0)**X0)**t(X0)**yt[,1])**t(X0)**
yt[,1]))/ve)+2*nrow(solve(t(X0)**X0)**t(X0)**yt[,1])-nrow(X0)
I<-diag(1,nrow(X0),nrow(X0))
D<-I-X0**solve(t(X0)**X0)**t(X0)
D<-diag(diag(D),nrow(X0),nrow(X0))
e<-yt[,1]-(X0**(solve(t(X0)**X0)**t(X0)**yt[,1]))
press[j]<-t(e)**solve(D)**e
x0<-xx0[,ind,drop=F]
tmp<-lsfit(X0,yt[,1],int=F)
b[j,ind]<-tmp$coefficients
mu<-c(b[j,1],rep(0,length(ind)-1))
V0<-diag(s2[ind,drop=F],length(ind))
mup<-(solve(t(X0)**X0+solve(V0))**(t(X0)**yt[,1]+solve(V0)**mu)
Vp<-solve(t(X0)**X0+solve(V0))
mu<-as.matrix(mu)
y<-as.matrix(yt[,1])
lambp<-(1/(n+v))*((v*lamb)+(t(mu)**solve(V0)**mu)+
(t(y)**y)-(t(t(X0)**y+solve(V0)**mu))**
(solve(t(X0)**X0+solve(V0))**(t(X0)**y+solve(V0)**mu))
a1<-NULL
a2<-NULL
for(i in 1:25){
a1[i]<-(lambp*(t(as.matrix(x0[i,]))**Vp**x0[i,]+1))^(1/2)
y0<-yp[,1]
a2[i]<-(vp+(((y0[i]-(t(as.matrix(x0[i,]))**mup))^2)/
(lambp*(t(as.matrix(x0[i,]))**
Vp**x0[i,]+1))))^(1/2)
Mf[i,j]<-K*a1[i]*a2[i]} lep<-rep(0,256) for(i in 1:256){
lep[i]<--sum(log(Mf[,i]))} lepr2<-lep[which.max(r2)]
lepr2aj<-lep[which.max(r2aj)] leppress<-lep[which.min(press)]
lepcp<-lep[which.min(cp)]
list(LEPR2=lepr2,LEPR2aj=lepr2aj,LEPPress=leppress,LEPCp=lepcp)}

```

```
#criando as sementes
set.seed(*)
s<-round(runif(1000,0,99999999),digits=0)
LEP<-matrix(0,1000,4)
for(i in 1:1000){
LEP[i,]<-cbind(dpred(dataname,n,s[i])$LEPR2,dpred(dataname,n,s[i])$LEPR2aj,
dpred(dataname,n,s[i])$LEPPress,dpred(dataname,n,s[i])$LEPCp)
if((i%%1)*1 == i) save(LEP,file="<caminho>/LEP.Rdata")}
# (*) Foram utilizados os seguintes valores iniciais para a produção
#das sementes: #32815368 para n=50, 69429889 para n=75 e 76094802
#para n=100.
```

ANEXO B - Programa em *R* v. 2.0.1 utilizado para obter o Logaritmo do Escore Preditivo (LEP) da ponderação Bayesiana de modelos.

```
# dataname = "C-111.txt", "Piranão.txt" ou "Flint Composto.txt" #
#n=50, 75 ou 100 rm(list=ls(all=TRUE)) require(mvtnorm)
dpred<-function(dataname,n,sem){ d<-read.table(dataname,header=T)
attach(d) matx<-cbind(x1,x2,x3,x4,x5,x6,x7,x8)
mod<-summary(lm(y~matx)) coef<-mod$coefficients[,1]
med<-c(mean(x1),mean(x2),mean(x3),mean(x4),mean(x5),mean(x6),mean(x7),mean(x8))
vc<-cov(matx) set.seed(sem) datx<-rmvnorm(n,med,vc)
dat<-cbind(1,datx) e<-rnorm(n,0,mod$sigma) y<-rep(0,n) for(i in1:n){
y[i]<-sum(coef*dat[i,])+e[i]} Xg<-cbind(y,datx)
s<-sample(1:n,(n-25),FALSE) yt<-Xg[s,] yp<-Xg[-s,] phi<-2.85
s2<-c(var(yt[,1]),(phi^2)*(1/var(yt[,2])),(phi^2)*(1/var(yt[,3])),(phi^2)*
(1/var(yt[,4])),(phi^2)*(1/var(yt[,5])),(phi^2)*(1/var(yt[,6])),(phi^2)*
(1/var(yt[,7])),(phi^2)*(1/var(yt[,8])),(phi^2)*(1/var(yt[,9])))
p<-ncol(datx) w<-NULL

for(j in 1:p)w<-rbind(cbind(w,FALSE),cbind(w,TRUE))
w<-cbind(TRUE,w)
k<-1:(p+1)
v<-2.58
n<-nrow(yt)
vp<-n+v
lamb<-0.28
K<-(gamma((vp+1)/2)/(gamma(1/2)*gamma(vp/2)))*vp^(vp/2)
b<-matrix(0,nrow(w),ncol(w))
X<-cbind(1,yt[,-1])
xx0<-cbind(1,yp[,-1])
Mf<-matrix(0,nrow(yp),nrow(w))

for(j in 1:nrow(w)){
ind<-k[w[j,]]
```

```

X0<-X[,ind,drop=F]
x0<-xx0[,ind,drop=F]
tmp<-lsfit(X0,yt[,1],int=F)
b[j,ind]<-tmp$coefficients
mu<-c(b[j,1],rep(0,length(ind)-1))
V0<-diag(s2[ind,drop=F],length(ind))
mup<-(solve(t(X0)%*%X0+solve(V0))%*%(t(X0)%*%yt[,1]+solve(V0)%*%mu)
Vp<-solve(t(X0)%*%X0+solve(V0))
mu<-as.matrix(mu)
y<-as.matrix(yt[,1])
lambp<-(1/(n+v))*((v*lamb)+(t(mu)%*%solve(V0)%*%mu)+(t(y)%*%y)-
(t(t(X0)%*%y+solve(V0)%*%mu))%*%(solve(t(X0)%*%X0+solve(V0))%*%
(t(X0)%*%y+solve(V0)%*%mu))
a1<-NULL a2<-NULL for(i in 1:25){
a1[i]<-(lambp*(t(as.matrix(x0[i,]))%*%Vp%*%x0[i,]+1))^(1/2)
y0<-yp[,1]
a2[i]<-(vp+(((y0[i]-t(as.matrix(x0[i,]))%*%mup))^2)/
(lambp*(t(as.matrix(x0[i,]))%*%Vp%*%x0[i,]+1))))^(1/2)
Mf[i,j]<-K*a1[i]*a2[i]}
lep<-rep(0,256)
for(i in 1:256){
lep[i]<--sum(log(Mf[,i]))}
k1<-gamma((n+v)/2)
k2<-gamma(v/2)*(v^(n/2))*(pi^(n/2))
py<-NULL I<-diag(1,n)

for(i in 1:nrow(w)){
ind<-k[w[i,]]
X0<-X[,ind,drop=F]
tmp<-lsfit(X0,y,int=F)
b[i,ind]<-tmp$coefficients
m<-c(b[i,1],rep(0,length(ind)-1))

```

```

V0<-diag(s2[ind,drop=F],length(ind))
py[i]<-k1*((det(lamb*(I+X0%*%V0%*%t(X0))))^(-1/2))*((1+((1/v)*t(y-X0%*%m)%*%(solve(
  pm<-NULL for(i in 1:nrow(w)){
pm[i]<-py[i]*(1/nrow(w))/sum(py*(1/nrow(w)))}
pm<-matrix(pm,256,1)
leppond<--sum(log(Mf%*%pm))
list(leppond=leppond)}
#criando as sementes
set.seed(*)
s<-round(runif(1000,0,99999999),digits=0)
LEPBMA<-NULL

for(i in 1:1000){
LEPBMA[i]<-dpred(dataname,n,s[i])$leppond
if((i/%1)*1 == i)save(LEPBMA,file="c:/hm/LEPBMA.Rdata")}

# (*) Foram utilizados os seguintes valores iniciais para a produção
#das sementes:
#32815368 para n=50, 69429889 para n=75 e 76094802
#para n=100.

```

ANEXO C - Programa em R v. 2.0.1 utilizado para executar o *Occam's Window* e calcular $P(\beta_j \neq 0 | \mathbf{y})$

```

rm(list=ls(all=TRUE))
pond<-function(dataname,X,y,nu,lambda,phi,a,OR){
attach(read.table(dataname,header=T))
X<-cbind(x1,x2,x3,x4,x5,x6,x7,x8)
p<-ncol(X)
w<-NULL for(j in 1:p)
w<-rbind(cbind(w,FALSE),cbind(w,TRUE))
w<-cbind(TRUE,w)
s2<-NULL
for(i in 1:ncol(X)){
s2[i]<-var(X[,i])}
s2<-c(var(y),s2)
s2<-c(s2[1],(phi^2)*(1/s2[2:length(s2)])) X<-cbind(1,X)
#res<-list()
k<-1:(p+1)
n<-nrow(X)
b<-matrix(0,nrow(w),ncol(w))
k1<-gamma((n+nu)/2)
k2<-gamma(nu/2)*(nu^(n/2))*(pi^(n/2))
py<-NULL
I<-diag(1,n)
for(i in 1:nrow(w)){
ind<-k[w[i,]]
X0<-X[,ind,drop=F]
tmp<-lsfit(X0,y,int=F)
b[i,ind]<-tmp$coefficients
m<-c(b[i,1],rep(0,length(ind)-1))
V0<-diag(s2[ind,drop=F],length(ind))
py[i]<-k1*((det(lambda*(I+X0%*%V0%*%t(X0))))^(-1/2))*
((1+((1/nu)*t(y-X0%*%m)%*%(solve(lambda*(I+X0%*%V0%*%t(X0))))))%*%

```

```

(y-X0%*%m))^(-(n+nu)/2))/k2}
# Obtenção de P(Mk|y) pm<-NULL for(i in 1:nrow(w)){
pm[i]<-py[i]*(1/nrow(w))/sum(py*(1/nrow(w)))}
pm<-round(100*pm,digits=4) # Occam Window - 1ºEtapa#
r<-matrix(FALSE,1,nrow(w)) for(i in 1:nrow(w)){ if(max(pm)/pm[i]<a)
r[i]<-TRUE} kk<-1:nrow(w) aux<-kk[r[1,]]
W<-matrix(0,length(aux),ncol(X)) for(i in 1:length(aux)){
W[i,]<-w[aux[i],]} pmm<-pm[aux] R<-cbind(aux,W,pmm)
R<-as.data.frame(R)
names(R)<-c("Modelo","b0","b1","b2","b3","b4","b5","b6","b7","b8","P(Mk|y)")
# Occam Window - 2º Etapa
R1<-as.matrix(R)
s<-NULL

for(i in 1:nrow(R1)){ s[i]<-sum(R1[i,3:(ncol(R)-1)])}
ind<-which(s==min(s)) R1[ind,] ind2<-which(R1[ind,] == 1)
o<-matrix(FALSE,1,nrow(R1)) for(i in 1:nrow(R1)){
if((R1[i,ind2[1]]==1) & (R1[i,ind2[2]]==1) &
(R1[i,ind2[3]]==1))o[i]<-TRUE} ind3<-which(o[,TRUE])
p<-matrix(FALSE,1,length(o)) for(i in 1:length(o)){
if((R1[ind,11]/R1[i,11])<OR) p[i]<-TRUE} ind4<-which(p[1,TRUE])
R2<-rbind(R1[ind,],R1[ind4,]) # P(bj!=0|y) #
k<-1:ncol(w)
pb<-rep(0,ncol(w))
for(i in 1:ncol(w)){
ind5<-which(w[,i], arr.ind =FALSE)
pb[i]<-sum(pm[ind5])}
pb
list(py=py,pm=pm,R=R,R2=R2,pb=pb)}

```

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)