# THE FOCUSED INFORMATION CRITERION IN LOGISTIC REGRESSION TO PREDICT REPAIR OF DENTAL RESTORATIONS

Cecilia CANDOLO[1]

▪ *ABSTRACT: Statistical data analysis typically has several stages: exploration of the data set; deciding on a class or classes of models to be considered; selecting the best of them according to some criterion and making inferences based on the selected model. The cycle is usually iterative and will involve subject-matter considerations as well as statistical insights. The conclusion reached after such a process depends on the model(s) selected, but the consequent uncertainty is not usually incorporated into the inference. This may lead to underestimation of the uncertainty about quantities of interest and overoptimistic and biased inferences. This framework has been the aim of research under the terminology of model uncertainty and model averanging in both, frequentist and Bayesian approaches. The former usually uses the Akaike´s information criterion (AIC), the Bayesian information criterion (BIC) and the bootstrap method. The last weigths the models using the posterior model probabilities. This work consider model selection uncertainty in logistic regression under frequentist and Bayesian approaches, incorporating the use of the focused information criterion (FIC) (CLAESKENS and HJORT, 2003) to predict repair of dental restorations. The FIC takes the view that a best model should depend on the parameter under focus, such as the mean, or the variance, or the particular covariate values. In this study, the repair or not of dental restorations in a period of eighteen months depends on several covariates measured in teenagers. The data were kindly provided by Juliana Feltrin de Souza, a doctorate student at the Faculty of Dentistry of Araraquara - UNESP. The results showed that the BIC, FIC and Bayesian averaging matches and the weights calculated enhanced the discussion concerning the choice of a best model.*

▪ *KEYWORDS: AIC; Bayesian model averaging; FIC; logistic regression; model averaging.*

[1]Universidade Federal de São Carlos – UFSCar, Centro de Ciências Exatas e de Tecnologia, Departamento de Estatística, Postal Box 676, Postal code 13.565-905, São Carlos, São Paulo, Brasil. E-mail: *cecilia@ufscar.br*

## 1 Introduction

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

Chatfield (1995) and Draper (1995) discuss the costs of ignoring model uncertainty and focus on Bayesian interpretation. The tutorial on Bayesian model averaging given by Hoeting *et al.* (1999) provides a historical perspective on the combination of models and gives further references. In the Bayesian framework model selection is replaced by weighted averaging over all possible models. This can be sensitive to the choice of prior and if there are many possible models the averaging will not be a practical proposition.

A related frequentist procedure proposed by Buckland *et al.* (1997) integrates model selection with statistical inference using simple weighting methods, where weights are obtained from information criteria or the bootstrap. That paper gave a number of examples that suggested that the method works fairly well in practice, but contained little theoretical justification. Candolo *et al.* (2003) studied this weighting procedure in linear model cases and compared it with some other options.

More recently, Claeskens e Hjort (2008) discussed several approaches in model selection and model averaging, including the use of the focused information criterion that they proposed early (Claeskens e Hjort, 2003). Claeskens e Hjort (2003) proposed the focused information criterion, FIC, that takes the view that a *best model* should depend on the parameter under focus, such as the mean, or the variance, or the particular covariate values, etc. In this way, the FIC allows different models to be selected for different parameters of interest. The FIC is defined in terms of a focus that can be any continuous function of the parameters, with non-zero continuous first derivative. For example, the focus could be the prediction for a given new observation. Again, because the FIC depends on the focus, it will take different values, and possibly select different models depending on which focus is used. The standard version of the FIC measures the mean squared error of the estimator of the quantity of interest in the selected model.

The purpose of this work is to use the several approaches of model averaging in views to predict the odds of an event of interest in a logistic regression framework. This will be done by means of an application to a study on repair of dental restorations. In this study, the repair or not of dental restorations in a period of eighteen months depends on several covariates mesured in teenagers. The data were kindly provided by Juliana Feltrin de Souza, from the Faculty of Dentistry of Araraquara, UNESP, Brazil.

In Section 2 we describe the methodology for frequentist and Bayesian model averaging, as well as the FIC for logistic regression. Section 3 shows the results of the application and Section 4 contains brief discussion.

# 2 Methodology

## 2.1 General framework

Consider a multiple linear regression model, in which independent responses $Y_1, \ldots, Y_n$ with corresponding covariates $X_{1,j}, \ldots, X_{n,j}$, $j = 1, \ldots, r$, have means $\beta_0 + \beta_1 X_{i,1} + \cdots + \beta_1 X_{i,r}$, $i = 1, \ldots, n$, and common variance $\sigma^2$ are supposed to come from a density of the form

$$f_Y(y) = f(y, \beta, \sigma^2).$$

The ordinary least square estimators of the parameters $\beta$ are $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, in matricial notation, and if $f_Y(y)$ is a normal distribuition the maximum likelihood estimators coincide. When many covariates are used we could atempt to use them all to model their influence on a response, or only a subset of them, which would make it easier to interpret the results. In this case there are $2^r$ possible models. For selecting a model among a list of candidates, Akaikes's information criterion (AIC) (AKAIKEe, 1973) is among the most popular and versatile strategies. Its essence is a penalized version of the attained maximum log–likelihood, for each model $k$,

$$AIC(k) = 2 \log likelihood_{max}(k) - 2dim(k),$$

where *dim(k)* is the length of the parameter vector of model *k*. Thus AIC acts as a penalised log–likelihood criterion, affording a balance between good fit (high value of log-likelihood) and complexity (complex models are penalised more than simple ones). The model with the highest AIC score is then selected. There are some natural generalisations and modifications of AIC that in various situations aim at performing more accurately (CLAESKENS and HJORT, 2008). The idea of to pick the candidate model with the higest probability given the data, has been formalised inside a Bayesian framework, involving priors probabilities on candidate models along with priors densities on all parameter vectors in the models. The Bayesian information criterion (BIC) (SCHWARZ, 1978) takes the form of a penalised log–likelihood function where the penalty is equal to the logarithm of the sample size times the number of estimated parameters in the model,

$$BIC(k) = 2 \log likelihood_{max}(k) - \log n dim(k),$$

again with *dim(k)* is the length of the parameter vector of model *k* and *n* the sample size of the data. The model with the highest BIC value is chosen as the best model. The BIC is constructed in a manner quite similar to the AIC, with a stronger penalty for complexity (as long as $n \geq 8$). Claeskens and Hjort (2008) presents a comparison of some information criteria with respect to consistency and efficiency. In a framework with increasing sample size, the AIC is not strongly consistent, though it is efficient, while the opposite is true for the BIC. Emiliano *et al.* (2014) discuss the statistical concepts of information and entropy concerning the AIC and BIC.

The model selection methods, such as AIC and BIC and others, have one thing in common: they select one single *best model*, which should then be used to explain all aspects of the mechanisms underlying the data and predict all future data points.

### 2.2 Frequentist model averaging

Buckland *et al.* (1997) discussed how model selection uncertainty may be incorporated into statistical inference using weighting methods. If the models $M_0, \ldots, M_K$ provide estimates $\hat{\theta}_k$ of $\theta$ and weights $W_k$ with $\sum W_k = 1$, the overall estimate of $\theta$ is taken to be

$$\hat{\theta} = \sum_{k=0}^{K} W_k \hat{\theta}_k. \quad (1)$$

Suppose that model $M_k$ has $p_k$ parameters. The $W_k$ may be obtained using the Akaike information criterion (AIC) (AKAIKE, 1973) or the Bayes information criterion (BIC) (SCHWARZ, 1978). Using information criteria of the form $I_k = -2\log \hat{L}_k + q_k$, where $\log \hat{L}_k$ is the maximized log likelihood for model $k$ and $q_k$ is a penalty equal to $2p_k$ for AIC and $p_k \log n$ for BIC, the weights are defined as

$$W_k = \frac{\exp(-I_k/2)}{\sum_{i=0}^{K} \exp(-I_i/2)}, \quad k = 1, \ldots, M,$$

and the averaged estimator is

$$\hat{\theta} = \sum_{k=1}^{M} W_k \hat{\theta}_k.$$

For logistic regression models the formulation is the same.

### 2.3 Bayesian model averaging

Bayesian model averaging is a standard way to incorporate model uncertainty into statistical inference (HOETING *et al.*, 1999). If $\theta$ is the quantity of interest, such as a future observable, and $\mathcal{M} = \{M_0, \ldots, M_K\}$ denotes the set of all models being considered, then the posterior distribution of $\theta$ given data $D$ is

$$P(\theta \mid D) = \sum_{k=0}^{K} P(\theta \mid M_k, D) P(M_k \mid D),$$

an average of the posterior distributions weighted by the posterior model probabilities. The posterior probability for model $M_k$ is

$$P(M_k \mid D) = \frac{P(D \mid M_k) P(M_k)}{\sum_{l=0}^{K} P(D \mid M_l) P(M_l)},$$

where

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

is the integrated likelihood for model $M_k$, $\eta_k$ is the parameter vector for $M_k$, $P(\eta_k \mid M_k)$ is the prior density of $\eta_k$ under $M_k$ and $P(M_k)$ is the prior probability that $M_k$ is the true model. All probabilities are implicitly conditioned on $\mathcal{M}$.

The posterior mean and variance of $\theta$ are

$$\begin{aligned} E(\theta \mid D) &= \sum_{k=0}^{K} E(\theta \mid M_k, D) P(M_k \mid D) \\ Var(\theta \mid D) &= \sum_{k=0}^{K} Var(\theta \mid M_k, D) P(M_k \mid D) \\ &\quad + \sum_{k=1}^{K} \{E(\theta \mid M_k, D) - E(\theta \mid D)\}^2 P(M_k \mid D). \end{aligned}$$

For logistic regression models, Raftery (1995) considered that when all models are equal a priori, the posterior probability for model $M_k$ is approximated by

$$P(M_k/D) \approx \exp(-\frac{1}{2} BIC_k) / \sum_{l=1}^{k} \exp(-\frac{1}{2} BIC_l).$$

### 2.4 The focused information criterion

Consider a *focus parameter* $\mu = \mu(\theta, \gamma)$, i.e. a parameter of direct interest and that is wished to estimate with good precision. For the estimation all components of $\theta$are to be included in the model, but it is not sure which components of of $\gamma$ to include when forming a final estimate. Perhaps all $\gamma_j$ shall be included, perhaps none. This leads to considering estimators of the form $\widehat{\mu} = \mu(\widehat{\theta_S}, \widehat{\gamma_S}, \gamma_{0,S^c})$. The *best* model for estimation of the focus parameter $\mu$ is the model for which the mean squared error of $\sqrt{n}(\widehat{\mu_S} - \mu_{true})$ is the smallest. The focused information criterion (FIC) is based on an estimator of these mean squared errors. The model with the lowest value of the FIC is selected.

To simplify the notation $(\hat{\theta}, \hat{\gamma})$ signals maximun likelihood estimation in the full $p+q$ parameter model. Let $D_n = \sqrt{n}(\hat{\gamma} - \gamma_0)$ and $\omega = \mathbf{J}_{10}\mathbf{J}_{00}^{-1} \partial\mu/\partial\theta - \partial\mu/\partial\gamma$. The FIC score is defined for each of the submodels indexed by $S$. Claeskens and Hjort (2008) present its proper derivation and several equivalent formulae that may be used, including

$$\begin{aligned} FIC(S) &= \hat{\omega}^t (I_q - \hat{G}_S) D_n D_n^t (I_q - \hat{G}_S)^t \hat{\omega} + 2\hat{\omega}^t \hat{Q}_S^0 \hat{\omega} && (2) \\ &= n\hat{\omega}^t (I_q - \hat{G}_S)(\hat{\gamma} - \gamma_0)(\hat{\gamma} - \gamma_0)^t (I_q - \hat{G}_S)^t \hat{\omega} + 2\hat{\omega}^t \hat{Q}_S^0 \hat{\omega} && (3) \\ &= (\hat{\psi}_{wide} - \hat{\psi}_S)^2 + 2\hat{\omega}_S^t \hat{Q}_S \hat{\omega}_S, && (4) \end{aligned}$$

where $\hat{\psi} = \hat{\omega}^t D_n$ and $\hat{\psi}_S = \hat{\omega}^t G_S D_n$ may be seen as the wide model based and the S–based estimates of $\psi = \omega^t \delta$. The FIC is a criterion used for selecting the best model, with smaller values of FIC($S$) favoured over larger ones. A bigger $S$ makes the first term small and the second big; correspondingly, a smaller $S$ makes the first term big and the second small. The two extremes are $2\hat{\omega}^t \hat{Q} \hat{\omega}$ for the wide model and $(\hat{\omega}^t D_n)^2$ for the narrow model. In practice the FIC value is computed for each of the models that are deemed plausible a priori, i.e. not always for the full list of the $2^q$ submodels. When $\hat{Q}$ is diagonal matrix $diag(\hat{\kappa}_1^2, \ldots, \hat{\kappa}_q^2)$, the FIC expression simplifies to

$$FIC(S) = (\sum_{j \notin S} \hat{\omega}_j D_{n,j})^2 + 2 \sum_{j \in S} \hat{\omega}_j^2 \hat{\kappa}_j^2. \tag{5}$$

The first term is a squared bias component for those parameters not in the set $S$, while the second term is twice the variance for those parameter estimators of which the index belongs to the index set $S$.

### 2.5 The FIC in logistic regression models

Claeskens, Croux and Van Kerckhoven (2006) and Claeskens and Hjort (2008) proposed versions of the FIC for variable selection in logistic regression. For a logistic regression model, there is a binary outcome variable $Y_i$ which is either one or zero. The most widely used model for relating probabilities to the covariates takes

$$P(Y_i = 1 | x_i, z_i) = p_i = \frac{\exp(x_i^t \beta + z_i^t \gamma)}{1 + \exp(x_i^t \beta + z_i^t \gamma)},$$

or $logit\{P(Y_i = 1|x_i, z_i)\} = x_i^t \beta + z_i^t \gamma$, where $logit(v) = \log\{v/(1-v)\}$. The vector of covariates is split in two parts; $x_i = (x_{i,1}, \ldots, x_{i,p})^t$ has the protected covariates, meant to be present in all of the models considered, while $z_i = (z_{i,1}, \ldots, z_{i,q})^t$ are the open or non protected variables from which we select the most adequate or important ones. Likewise, the coefficients are vectors $\beta = \theta = (\beta_1, \ldots, \beta_p)^t$ and $\gamma = (\gamma_1, \ldots, \gamma_q)^t$. In order to obtain the matrix $\mathbf{J}_{wide}$, or its empirical version $\mathbf{J}_{n,wide}$, the second–order partial derivatives of the log-likelihood function $\ell_n$ are calculated. The inverse logit function $H(\mu) = \exp(u)/\{1 + \exp(u)\}$ has $H'(u) = H(u)\{1 - H(U)\}$, so $p_i = H(x_i^t \beta + z_i^t \gamma)$ has $\partial p_i / \partial \beta = p_i(1 - p_i) x_i$ and $\partial p_i / \partial \gamma = p_i(1 - p_i) z_i$, which leads to

$$\begin{bmatrix} \partial \ell_n / \partial \beta \\ \partial \ell_n / \partial \gamma \end{bmatrix} = \sum_{i=1}^{n} \{y_i - p(x_i, z_i)\} \begin{bmatrix} x_i \\ z_i \end{bmatrix}.$$

It follows that for a logistic regression model

$$\mathbf{J}_{n,wide} = n^{-1} \sum_{i=1}^{n} p_i(1 - p_i) \begin{bmatrix} x_i x_i^t & x_i z_i^t \\ z_i x_i^t & z_i z_i^t \end{bmatrix} = \begin{bmatrix} \mathbf{J}_{n,00} & \mathbf{J}_{n,01} \\ \mathbf{J}_{n,10} & \mathbf{J}_{n,11} \end{bmatrix}.$$

This matrix and the matrices $Q$,$Q_S$ and $G_S$ are estimated using the estimates of the $\beta$ and $\gamma$ parameters in the full model. The vector $\delta/\sqrt{n}$ measures the departure distance between the smallest and the largest model and is estimated by $\hat{\gamma} - \gamma_0$. The narrow model in this application corresponds to $\gamma_0 = 0_{q\times 1}$, and then $\hat{\delta} = \sqrt{n}\hat{\gamma}$. Let the odds of an event of interest be the focus parameter to be estimated. The odds of an event of interest taken place, at a given position in the covariate space is

$$\mu(\theta, \gamma) = \frac{p(x, z))}{1 - p(x, z)} = \exp(x^t\beta + z^t\gamma).$$

The vector $\omega$ is given by

$$\omega(x, z) = \frac{p(x, z))}{1 - p(x, z)}(\mathbf{J}_{n,10}\mathbf{J}_{n,00}^{-1}x - z).$$

With this information, the FIC may be computed for each of the models of interest and the model with the lowest value of FIC is selected.

Other focus parameter that may be of interest is the probability that an event occurs. For a specified set of covariates $x, z$, select variables in a logistic regression model to estimate $\mu(\beta, \gamma; x, z) = P(Y = 1|x, z)$ will serve this purpose. The required FIC component $\omega$ is for this focus parameter equal to $p(x, z)\{1 - p(x, z)\}(\mathbf{J}_{n,10}\mathbf{J}_{n,00}^{-1}x - z)$. When $\omega$ has been estimated for this situation, the values of the FIC are calculated for all models in the search path.

The model averaging using the FIC follows the same framework as the frequentist version. Claeskens and Hjort (2008) state that the weights are defined as

$$W_k = \exp(-\frac{1}{2}\kappa\frac{FIC_k}{\hat{\omega}^t\hat{Q}\hat{\omega}})/\sum_{k=1}^{M}\exp(-\frac{1}{2}\kappa\frac{FIC_k}{\hat{\omega}^t\hat{Q}\hat{\omega}}).$$

## 3 Application to a study on repair of dental restorations

In this study, $Y$ is an indicator of the repair or not of occlusal dental restorations in a period of eighteen months and a list of potencially influential covariates were mesured in teenagers. The covariates are: plaque( yes or not ); caries (yes or not); material (chemically activated or resin modified); cusp (yes or not); isolation (absolute or relative) and severity (spots, loss of structure or atypical restorations). For the present illustration the $p$ covariates included in every candidate model were the intercept, plaque and caries. So, we have $x_1 = 1$, $x_2$ =plaque and $x_3$ =caries. The other covariates, from which we wish to select a relevant subset, are $z_1$ =material, $z_2$ =cusp, $z_3$ =isolation and $z_4$ =severity. The covariate severity is a factor with 3 levels and then it produces two covariates vectors in the design matrix, $z_4$= level 2 of severity and $z_5$= level 3. The others covariates are all factors with 2 levels. In this way, we have 5 columns in the matrix $Z$ of covariates to enter in the model, which means $2^5 = 32$ possible models. Table 1 presents the notation for all models.

The focus parameter we consider for the FIC is the probability that the event of interest takes place which in this application is need of a repair ($Y = 1$). The

Table 1 - Notation used to define the 32 models

| Model | Codes | Variables | Model | Codes | Variables |
|---|---|---|---|---|---|
| 1 | 1 1 1 1 1 1 1 | $x_2$, $x_3$, $z_1$, $z_2$, $z_3$, $z_4$, $z_5$ | 17 | 1 1 0 1 1 1 1 | $x_2$, $x_3$, $z_2$, $z_3$, $z_4$, $z_5$ |
| 2 | 1 1 1 1 1 1 0 | $x_2$, $x_3$, $z_1$, $z_2$, $z_3$, $z_4$ | 18 | 1 1 0 1 1 1 0 | $x_2$, $x_3$, $z_2$, $z_3$, $z_4$ |
| 3 | 1 1 1 1 1 0 1 | $x_2$, $x_3$, $z_1$, $z_2$, $z_3$, $z_5$ | 19 | 1 1 0 1 1 0 1 | $x_2$, $x_3$, $z_2$, $z_3$, $z_5$ |
| 4 | 1 1 1 1 1 0 0 | $x_2$, $x_3$, $z_1$, $z_2$, $z_3$ | 20 | 1 1 0 1 1 0 0 | $x_2$, $x_3$, $z_2$, $z_3$ |
| 5 | 1 1 1 1 0 1 1 | $x_2$, $x_3$, $z_1$, $z_2$, $z_4$, $z_5$ | 21 | 1 1 0 1 0 1 1 | $x_2$, $x_3$, $z_2$, $z_4$, $z_5$ |
| 6 | 1 1 1 1 0 1 0 | $x_2$, $x_3$, $z_1$, $z_2$, $z_4$ | 22 | 1 1 0 1 0 1 0 | $x_2$, $x_3$, $z_2$, $z_4$ |
| 7 | 1 1 1 1 0 0 1 | $x_2$, $x_3$, $z_1$, $z_2$, $z_5$ | 23 | 1 1 0 1 0 0 1 | $x_2$, $x_3$, $z_2$, $z_5$ |
| 8 | 1 1 1 1 0 0 0 | $x_2$, $x_3$, $z_1$, $z_2$ | 24 | 1 1 0 1 0 0 0 | $x_2$, $x_3$, $z_2$ |
| 9 | 1 1 1 0 1 1 1 | $x_2$, $x_3$, $z_1$, $z_3$, $z_4$, $z_5$ | 25 | 1 1 0 0 1 1 1 | $x_2$, $x_3$, $z_3$, $z_4$, $z_5$ |
| 10 | 1 1 1 0 1 1 0 | $x_2$, $x_3$, $z_1$, $z_3$, $z_4$ | 26 | 1 1 0 0 1 1 0 | $x_2$, $x_3$, $z_3$, $z_4$ |
| 11 | 1 1 1 0 1 0 1 | $x_2$, $x_3$, $z_1$, $z_3$, $z_4$, $z_5$ | 27 | 1 1 0 0 1 0 1 | $x_2$, $x_3$, $z_3$, $z_5$ |
| 12 | 1 1 1 0 1 0 0 | $x_2$, $x_3$, $z_1$, $z_3$ | 28 | 1 1 0 0 1 0 0 | $x_2$, $x_3$, $z_3$ |
| 13 | 1 1 1 0 0 1 1 | $x_2$, $x_3$, $z_1$, $z_4$, $z_5$ | 29 | 1 1 0 0 0 1 1 | $x_2$, $x_3$, $z_4$, $z_5$ |
| 14 | 1 1 1 0 0 1 0 | $x_2$, $x_3$, $z_1$, $z_4$ | 30 | 1 1 0 0 0 1 0 | $x_2$, $x_3$, $z_4$ |
| 15 | 1 1 1 0 0 0 1 | $x_2$, $x_3$, $z_1$, $z_5$ | 31 | 1 1 0 0 0 0 1 | $x_2$, $x_3$, $z_5$ |
| 16 | 1 1 1 0 0 0 0 | $x_2$, $x_3$, $z_1$ | 32 | 1 1 0 0 0 0 0 | $x_2$, $x_3$ |

data has 38 observations and there are 32 models to be fitted and then we have 32 values of the weights calculated by the AIC (WAIC), BIC (WBIC), posterior probabilities (WBAY) and the FIC (WFIC). Also were calculated the 38 values of averaged probability that the event of interest takes place. Figure 1 shows the values of the weights for each model in two versions: separately for each method (at the top) and all together (at the bottom). We observe that the weights using the BIC and the posterior probabilities are virtually the same.

Table 2 shows the weight values for the two best models in each approach, and clearly models 16 and 32 are the winners. Model 32 has only the X covariate, plaque and caries, and model 16 includes the material. This last information is very important for the dentist research and, the model 31, that was enlighten by the FIC methodology and includes the level 3 of severity is also important.

Table 2 - Weight values for the two best models (in bold) in each approach

| Codes | Model | WAIC | WBIC | WFIC | WBAY |
|---|---|---|---|---|---|
| 1 1 1 0 0 0 0 | 16 | **0.0969** | **0.1481** | 0.0629 | **0.1481** |
| 1 1 0 0 0 0 1 | 31 | 0.0532 | 0.0813 | **0.06292** | 0.08127 |
| 1 1 0 0 0 0 0 | 32 | **0.0769** | **0.2665** | **0.0629** | **0.2665** |

Figure 2 shows the averaged estimates using the four approaches, and again a plot for each method at the top and a plot with all graphs at the bottom. The BIC, FIC and Bayesian average matches and the AIC produces averaged estimates a little different for some values. All the calculations for the FIC were performed using the functions available in the web site of Claeskens and Hjort (2008) for the package R and for the Bayesian averaging were used functions from Hoeting *et al.* (1999).

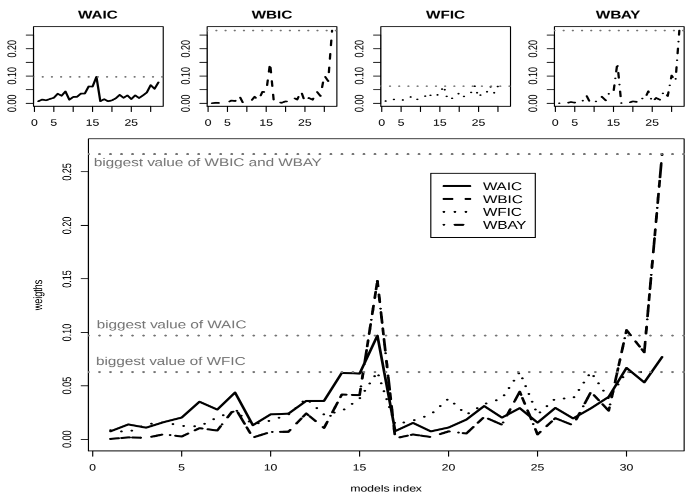


Figure 1 - Values of the weights calculated by the AIC (WAIC), BIC (WBIC), posterior probabilities (WBAY) for each model: separate plots at the top and at the bottom a plot with all graphs. The horizontal lines correspond to the biggest value of each weight.

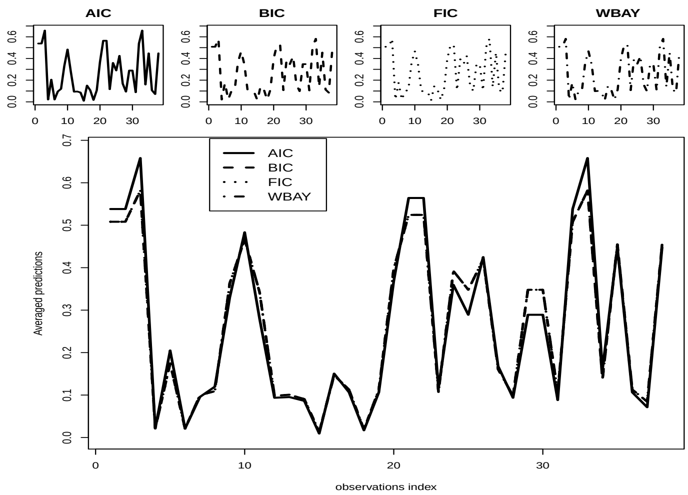


Figure 2 - Averaged estimates using the four approaches (AIC, BIC, FIC and Bayesian averaging: separate plots at the top and at the bottom a plot with all graphs.

## Conclusions

This paper has studied the use of a model averaging to allow for model uncertainty in logistic regression by frequentist and Bayesian approaches. The focused information criterion was one of the frequentist possibilities. We used one application for logistic regression models applied to a problem of repair of dental restorations. The aim was to explore the possibility to find more models to study and consider as a final result. The previous analysis of these data, under others model selection approaches showed that only the covariates plaque and caries were important. The use of the FIC improved this result by indicating the possibility of adding severity and isolation.

CANDOLO, C. Uso do *focused information criterion* em regressão logística para predizer o reparo de restaurações dentárias. *Rev. Bras. Biom.,* São Paulo, v.31, n.4, p.547-557, 2013.

- RESUMO: Neste trabalho considera-se a incerteza devido à seleção de modelos em regressão logística segundo as abordagens de ponderação de modelos frequentista, Bayesiana e também usando o FIC (focused informaion criterion) para predizer o reparo de restaurações dentárias. Os resultados mostram que as abordagens de ponderação usando o BIC, o FIC e as probabilidades posteriores são praticamente iguais, enquanto que o usos dos pesos calculados para a escolha de modelos ajudaram a proceder à seleção de modelos.
- PALAVRAS-CHAVE: AIC; FIC; ponderação Bayesiana de modelos; ponderação de modelos; regressão logística.

## References

AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In KOTZ, S. and JOHNSON, N. L. (Eds.) *Breakthroughs in Statistics*, **1**, New York: Springer, 1973, p.610-624.

BUCKLAND, S. T.; BURNHAN, K. P.; AUGUSTIN, N. H. Model selection: an integral part of inference. *Biometrics*, v.53, p.603-618, 1997.

CANDOLO, C., DAVISON, A. C.; DEMÉTRIO, C. G. B. A Note on model uncertainty in linear regression. *Journal of the Royal Statistical Society, Series D: The Statistician*, v.52, p.165-177, 2003.

CHATFIELD, C. Model uncertainty, data mining and statistical inference (with Discussion). *Journal of the Royal Statistical Society, Series A*, v.158, p.419-466, 1995.

CLAESKENS, G.; CROUX, C.; van KERCKHOVEN, J. Variable selection for logistic regression using a prediction-focused information criterion. *Biometrics*, v.62, p.972-979, 2006.

CLAESKENS, G.; HJORT, N. L. The focused information criterion (with discussion). *Journal of the American Statistical Association*, v.98, p.900-945, 2003.

CLAESKENS, G.; HJORT, N. L. *Model selection and Mmodel averaging.* Cambridge: Cambridge University Press, 2008, 320p.

DRAPER, D. Assessment and propagation of model uncertainty (with Discussion). *Journal of the Royal Statistical Society, Series B* v.57, p.45-97, 1995.

EMILIANO, P. C.; VIVANCO, M. J. F.; MENEZES, F. S. Information criteria: how do they behave in different models? *Computational Statistics and Data Analysis*, v.69, p.141-153, 2014.

HOETING, J. A.; MADIGAN, D.; RAFTERY, A. E.; VOLINSKY, C. T. Bayesian model averaging: a tutorial (with Discussion). *Statistical Science*, v.14, p.382-417, 1999.

RAFTERY, A.E. Bayesian model selection in social research (with Discussion). *Sociological Methodology*, p.111-196, 1995.

SSHWARZ, G. Estimating the dimension of a model. *The Annals of Statistics*, v.6, p.461-464, 1978.

Received in 16.10.2013.

Approved after revised in 29.01.2014.