3190

RELATÓRIO DE PESQUISA 1994

BEST SUBSETS FOR EXTENTIONS OF DICHOTOMOUS LOGISTIC REGRESSION

Tirza Aydar

and

Eliana H. de F. Marques

Outubro RP 58/94

INSTITUTO DE MATEMÁTICA ESTATÍSTICA E CIÊNCIA DA COMPUTAÇÃO

UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS

Best Subsets for Extentions of Dichotomous Logistic Regression

Tirza Aydar[1] and Eliana H. de F. Marques[2]

[1]Lara-Mara, Km 3.5 Rodovia Heitor Penteado
Jardim Santa Marcelina CEP 13094-430
[2]Universidade Estadual de Campinas
13081 – 970, Campinas, São Paulo, Brasil

SUMMARY

The problem of selecting independent variables to be employed in a logistic model may be approached via linear regression technique. In this article, we extend based on Hastie and Tibishirani (1987, *Applied Statistics* 36, 260 – 276), the paper of Hosmer, Jovanovic and Lemeshow (1989, *Biometrics* 45, 1265 – 1270) to yield a best subset selection of covariates for the analysis of polytomous and proportional odds logistic regression models.

1. Introduction

In some instances researchers are forced to prospect for independent variables that could conceivably be related to the dependent variable. Under these circumstances studies will have a large set of independent variables. An important step in model building strategy for these types of datasets is the selection of a best set of independent variables such that with a shortened list of such variables and a small number of parameters one is able to describe and coherently interpret the relationship between the response variable Y and the vector of independent variables $\underset{\sim}{X}$.

When dealing with regression analysis for a binary response, the dichotomous logistic regression model has become a standard method in many fields of research. Hosmer, Jovanovic and Lemeshow (1989) present a method for selection of meaningful covariates for the case of binary logistic regression based on Lawless and Singhal (1978) method of efficiently screening nonnormal regression models. The Lawless-Singhal analysis may be performed with any best subsets linear regression program that allows

Key words: Proportional odds logistic model; Weighted least squares; Variable selection; Mallow's C_p.

for weights. The approach proposed by Hosmer et al. (1989) has the appeal of utilizing less complex algorithms, and most important, requires programs for linear regression analysis which are readily available. Basically, it identifies from the set of independent variables, key subsets of q ($q < p$) variables by fitting models with a linear regression package using a pseudo-variable Z and a diagonal matrix W of weights. The Wald statistic is used for comparisons of hierarchical models and Mallow's coefficient is used for comparing competing models of same size q.

In this note we show that the method of best subsets by Hosmer et al. (1989) can be extended to other members of the family of logistic regression models, more specifically the polytomous and the proportional odds model.

2. Methodology

Even though logistic regression is frequently employed to model the relationship between a dichotomous outcome variable and a set of covariates, generalization to an outcome variable with k categories ($k>2$) is sometimes appropriate for certain types of studies.

Consider the polytomous logistic model with k categories based on multiple logits of probability of $(k-1)$ categories in relation to a selected one chosen as a standard. This model does not take into consideration an ordered structure of the response variable so it is appropriate to nominal outcome.

Let T be an indicator variable such that,

$$t_{ij} = \begin{cases} 1 & \text{if } Y_i = j \\ 0 & \text{otherwise} \end{cases} \qquad i = 1, 2, \ldots, n \quad j = 1, 2, \ldots, k$$

and

$$\pi_j = P[\, t_{ij} = 1 / \underset{\sim}{X}_i \,],$$

where Y_i is the outcome variable and $\underset{\sim}{X}_i' = (x_{i0}, x_{i1}, \ldots, x_{ip})$ is a vector of $p+1$ independent variables for the i^{th} observation.

If the k^{th} category is used as reference, the polytomous logistic model is defined as:

$$\eta_{ij} = \ln\left(\frac{\pi_{ij}}{\pi_{ik}}\right) = \underset{\sim}{X}_i' \underset{\sim}{\beta}_j$$

and

$$\pi_{ij} = \frac{\exp(\underset{\sim}{X}_i' \underset{\sim}{\beta}_j)}{\sum_{j=1}^{k} \exp(\underset{\sim}{X}_i' \underset{\sim}{\beta}_j)} \qquad (1)$$

with $\underset{\sim}{\beta}_j' = \left(\underset{\sim}{\beta}_{j0}, \underset{\sim}{\beta}_{j1}, \ldots, \underset{\sim}{\beta}_{jp}\right)$, the vector of unknown parameters, $\underset{\sim}{\beta}_k = \underset{\sim}{0}$ and $j = 1, 2, \ldots, k-1$.

Begg and Craig (1984) suggest that model (1) may be fit via $(k-1)$ individual dichotomous logistic regressions. They demonstrate that the estimates obtained are consistent with efficiency close to that of the polytomous fit. They also mention that this efficiency will decrease with increase of number of parameters and low frequencies in the reference category.

Extension of Hosmer et al. (1989) results to the polytomous logistic model is straightforward. The best subsets linear regression approach can be performed either on individual dichotomous logistic regressions following Begg and Craig (1984) or on the polytomous model with:

$$\underset{\sim}{Z}_i = \underset{\sim}{X}_i^{*\prime} \hat{\underset{\sim}{\xi}} + \underset{\sim}{W}_i^{-1}(\underset{\sim}{y} - \hat{\underset{\sim}{\pi}});$$

where

$$\hat{\underset{\sim}{\xi}}' = (\hat{\underset{\sim}{\beta}}_1', \hat{\underset{\sim}{\beta}}_2', \ldots, \hat{\underset{\sim}{\beta}}_{k-1}'),$$

$$\underset{\sim}{W}_i = \operatorname{diag} \frac{\partial \pi_j}{\partial \eta_j} = \operatorname{diag}(\pi_j(1-\pi_j))$$

and

$$\underset{\sim}{X}_i^{*\prime} = \left[\begin{array}{cccc|ccc|ccc|c|ccc} 1 & 0 & \cdots & 0 & x_{i1} & \cdots & x_{ip} & 0 & \cdots & 0 & & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & & 0 & x_{i1} & \cdots & x_{ip} & & & & \\ \vdots & & & \vdots & \vdots & & \vdots & \vdots & & \vdots & \cdots & & & \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & & x_{i1} & \cdots & x_{ip} \end{array}\right];$$

where $\underset{\sim}{X}_i^{*\prime}$ is $(k-1) \times (k-1)(p+1)$ and the $\underset{\sim}{W}_i$'s ($i = 1, 2, \ldots, n$) are the blocks of a square block diagonal matrix $\underset{\sim}{W}$ of dimension $n(k-1)$.

Since the dimension of $\underset{\sim}{X}^*$ increases with p and k care need to be taken so that the procedure does not lose its ability to compare models. For example, if $p = 10$ and $k = 4$, $\underset{\sim}{X}^{*\prime}$ will have 3n lines and ($3 \times 11 = 33$) columns, and will generate 2^{11} possible models to be compared. In this situation analysis with individual dichotomous logistic regressions would be preferred.

Differently from the polytomous logistic model, the proportional odds model takes into consideration an ordinal scale outcome variable. If Y is the response variable with k ordered categories then

$$\pi_j = P(Y = j/\underset{\sim}{X})$$

and

$$\gamma_j = P(Y \leq j/\underset{\sim}{X}) = \sum_{l=1}^{j} \pi_l \qquad j = 1, 2, \ldots, k$$

with $\gamma_k = 1$, the cumulative response probability of all categories of Y.

The model proposed by McCullag (1980) is defined as

$$\ln\left(\frac{P(Y \leq j/\underset{\sim}{X})}{1 - P(Y \leq j/\underset{\sim}{X})}\right) = \ln\left(\frac{\gamma_j}{1-\gamma_j}\right) = \theta_j + \underset{\sim}{X}'\underset{\sim}{\beta}$$

where $\underset{\sim}{\theta}$ and $\underset{\sim}{\beta}$ are vectors of unknown parameters and $\underset{\sim}{\theta}$ must satisfy $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_{k-1}$. This model has a smaller number of parameters when compared to (1) due to the fact that it takes into account the ordered structure presented by the data. The parameter estimates of $\underset{\sim}{\xi} = (\underset{\sim}{\theta}, \underset{\sim}{\beta})$ are obtained by iterative methods.

Thompson and Baker (1981) show that for generalized linear models with more than one predictive value for each observation, as in the proportional odds model, the ML estimator of $\underset{\sim}{\xi}$ may be obtained by the iterative reweighted least squares (IRLS) method through a link function. Green (1984) demonstrates that the method, in this case, is equivalent to the solution of the equations generated by $\partial L/\partial \underset{\sim}{\xi} = 0$ using the Newton Raphson algorithm with Fisher's Scores.

Hastie and Tibshirani (1987) using the results from Thompson, Baker and Green show the equivalence of Fisher's scores and IRSL for the proportional odds logistic model. The extension of Hosmer et al. best subsets dichotomous logistic regression technique to the proportional odds model presented below draws on results of Hastie and Tibshirani (1987).

Consider the proportional odds model with k ordered responses. Define the pseudo-variable $\underset{\sim}{Z}$ and its weight $\underset{\sim}{W}$ by:

$$\underset{\sim}{Z} = \underset{\sim}{C}^{-1}(\underset{\sim}{\gamma} - \underset{\sim}{\hat{\gamma}}) + \underset{\sim}{X}^{*\prime}\underset{\sim}{\xi}$$

or

$$\underset{\sim}{Z} = \begin{bmatrix} \frac{1}{\gamma_1(1-\gamma_1)} & 0 & \cdots & 0 \\ 0 & \frac{1}{\gamma_2(1-\gamma_2)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\gamma_{k-1}(1-\gamma_{k-1})} \end{bmatrix} \begin{bmatrix} \gamma_1-\hat{\gamma}_1 \\ \gamma_2-\hat{\gamma}_2 \\ \cdots \\ \gamma_{k-1}-\hat{\gamma}_{k-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 & \cdots & 0 & | & x_1 & x_2 & \cdots & x_p \\ 0 & 1 & \cdots & 0 & | & x_1 & x_2 & \cdots & x_p \\ \vdots & \vdots & \vdots & \vdots & | & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & | & x_1 & x_2 & \cdots & x_p \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{k-1} \\ \cdots \\ \underset{\sim}{\beta} \end{bmatrix} \qquad (2)$$

and $\underset{\sim}{W}$, a tridiagonal $(k-1)\times(k-1)$ matrix ($w_{ij} = 0 \ \forall$ i, j, s.t. $|i-j| > 1$) with diagonal elements $w_{jj}(j = 1, 2, \ldots, k-1)$, superdiagonal elements $w_{j,j+1}$ and subdiagonal elements $w_{j,j-1}$ described as :

$$w_{jj} = \left(\gamma_1^2(1-\gamma_1)^2\left(\frac{1}{\gamma_1} + \frac{1}{\gamma_2-\gamma_1}\right), \gamma_1^2(1-\gamma_2)^2\left(\frac{1}{\gamma_2-\gamma_1}\right)\left(\frac{1}{\gamma_3+\gamma_2}\right) \ldots \gamma_{k-1}^2(1-\gamma_{k-1})^2\left(\frac{1}{\gamma_{k-1}-\gamma_{k-2}} + \frac{1}{1-\gamma_{k-1}}\right)\right) \qquad (3)$$

$$w_{j,j+1} = w_{j,j-1} =$$

$$=\left(-\gamma_1(1-\gamma_1)-\gamma_2(1-\gamma_2)\left(\frac{1}{\gamma_2-\gamma_1}\right), -\gamma_2(1-\gamma_2)-\gamma_3(1-\gamma_3)\left(\frac{1}{\gamma_3-\gamma_2}\right), \ldots -\gamma_{k-2}(1-\gamma_{k-2})-\gamma_{k-1}(1-\gamma_{k-1})\left(\frac{1}{\gamma_{k-1}-\gamma_{k-2}}\right)\right) \qquad (4)$$

with

$$\underset{\sim}{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_{k-1}) = (\pi_1, \pi_1 + \pi_2, \ldots, \pi_1 + \pi_2 + \ldots + \pi_{k-1}).$$

$$\underset{\sim}{\pi} = (\pi_1, \pi_2, \ldots, \pi_k) \text{ and } \sum_{i=1}^{k} \pi_i = 1.$$

If it is of interest to select from p independent variables a subset q of these variables. This can be done using a linear regression program that starts with the WLS fit of the full model and the MLE $\hat{\underset{\sim}{\xi}}$ of $\underset{\sim}{\xi}$.

Let SSE(p) be the sum of squares of residuals from the fitted model

$$\underset{\sim}{Z} = [\underset{\sim}{1} \mid 1\underset{\sim}{X}]'\xi$$

where

$$\underset{\sim}{Z} = \underset{\sim}{C}^{-1}(\gamma - \hat{\gamma}) + \underset{\sim}{X}^{*\prime}\hat{\underset{\sim}{\xi}}$$

and

$$\hat{\gamma} = 1 \,/\, 1 + \exp\ -(\underset{\sim}{X}^{*\prime}\hat{\underset{\sim}{\xi}}).$$

Then,

$$SSE(p) = [\underset{\sim}{Z} - \hat{\underset{\sim}{Z}}(p)]'\ \underset{\sim}{W}[\underset{\sim}{Z} - \hat{\underset{\sim}{Z}}(p)].$$

$$= \sum_{i=1}^{n} (\underset{\sim}{\gamma}_i - \hat{\gamma}_i)\ C_i^{-1}\underset{\sim}{W}_i\underset{\sim}{C}_i^{-1}(\underset{\sim}{\gamma}_i - \hat{\gamma}_i).$$

By writing Pearson's χ^2 in terms of $\underset{\sim}{\gamma}$, one easily finds that SSE(p) is equivalent to the Pearson Chi-square goodness of fit statistic . This allows for the utilization of algorithms for selection of models with q variables based on the difference between the residual sums of squares of such models and that of the full model. The fitting of the model with q independent variables is performed by partitioning matrices X^*, ξ^* and I such that:

$$\underset{\sim}{Z}(q) = \underset{\sim}{X}^{*\prime}\hat{\xi}$$

with

$$\begin{aligned}\hat{\xi} &= (\hat{\underset{\sim}{\theta}}, \hat{\underset{\sim}{\beta}}_1 \mid \hat{\underset{\sim}{\beta}}_2) \\ &= (\hat{\underset{\sim}{\theta}}, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_q \mid \hat{\beta}_{q+1}, \ldots, \hat{\beta}_p) \\ &= (\hat{\underset{\sim}{\xi}}_1, \hat{\underset{\sim}{\xi}}_2),\end{aligned}$$

$$\underset{\sim}{X}^{*\prime} = (X_1^* X_2^*)' = \begin{bmatrix} 1 & 0 & \cdots & 0 & x_{11} & \cdots & x_{1q} & \mid x_{1(q+1)} \cdots & x_{1p} \\ 0 & 1 & \cdots & 0 & x_{21} & \cdots & x_{2q} & \mid \vdots & \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \mid \vdots & \\ 0 & 0 & \cdots & 1 & x_{n1} & \cdots & x_{nq} & \mid x_{n(q+1)} \cdots & x_{np} \end{bmatrix}$$

and

$$I = \begin{bmatrix} \underset{\sim}{X}_1^* \underset{\sim}{W} \underset{\sim}{X}_1^{*\prime} & \underset{\sim}{X}_1^* \underset{\sim}{W} \underset{\sim}{X}_2^{*\prime} \\ \underset{\sim}{X}_2^* \underset{\sim}{W} \underset{\sim}{X}_1^{*\prime} & \underset{\sim}{X}_2^* \underset{\sim}{W} \underset{\sim}{X}_2^{*\prime} \end{bmatrix} = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}.$$

The estimator of $\underset{\sim}{\xi}_1$ is given by

$$\tilde{\underset{\sim}{\xi}}_1 = (I_{11})^{-1} \underset{\sim}{X}^* \underset{\sim}{W} \underset{\sim}{Z}.$$

Substituting (2) in this last expression,

$$\begin{aligned}\tilde{\underset{\sim}{\xi}}_1 &= (I_{11})^{-1} \underset{\sim}{X}_1^* \underset{\sim}{W}[\underset{\sim}{C}^{-1}(\gamma - \hat{\gamma}) + \underset{\sim}{X}^{*\prime} \hat{\underset{\sim}{\xi}}] \\ &= (I_{11})^{-1} \underset{\sim}{X}_1^* \underset{\sim}{W}\underset{\sim}{C}^{-1}(\gamma - \hat{\gamma}) + (I_{11})^{-1} \underset{\sim}{X}_1^* \underset{\sim}{W}(\underset{\sim}{X}_1^* \underset{\sim}{W}(\underset{\sim}{X}_1^{*\prime} \hat{\underset{\sim}{\xi}}_1 + \underset{\sim}{X}_2^{*\prime} \hat{\underset{\sim}{\xi}}_2) \\ &= \hat{\underset{\sim}{\xi}}_1 + I_{11}^{-1} I_{12} \hat{\underset{\sim}{\xi}}_2.\end{aligned}$$

The residual sums of squares for the fitted model with q variables is given by

$$\begin{aligned}SSE(q) &= [\underset{\sim}{Z} - \tilde{\underset{\sim}{Z}}(q)]' \underset{\sim}{W}[\underset{\sim}{Z} - \tilde{\underset{\sim}{Z}}(q)] \\ &= \underset{\sim}{Z}' \underset{\sim}{W} \underset{\sim}{Z} - 2[\tilde{\underset{\sim}{\xi}}_1' \underset{\sim}{X}_1^*] W[\underset{\sim}{C}^{-1}(\underset{\sim}{\gamma} - \hat{\underset{\sim}{\gamma}}) + \underset{\sim}{X}^{*\prime} \hat{\underset{\sim}{\xi}}) + \tilde{\underset{\sim}{\xi}}_1' \underset{\sim}{X}_1^* \underset{\sim}{W} \underset{\sim}{X}_1^{*\prime} \tilde{\xi}_1 \\ &= X^2 - \hat{\underset{\sim}{\xi}}' I \hat{\underset{\sim}{\xi}} + \hat{\underset{\sim}{\xi}}_1' I_{11} \hat{\xi}_1 \\ &= X^2 + \hat{\underset{\sim}{\xi}}_2' [I_{22} - I_{21} I_{11}^{-1} I_{12}] \hat{\underset{\sim}{\xi}}_2 \\ &= X^2 + \lambda^*. \qquad (5)\end{aligned}$$

From (5) it follows that λ^* is the Wald Statistic for the test of H_0: $\underset{\sim}{\xi}_2 = 0$. The WLS regression of $\underset{\sim}{Z}$ on $\underset{\sim}{X}^{*\prime}$ with weight $\underset{\sim}{W}$ (where $\underset{\sim}{Z}$ and $\underset{\sim}{W}$ are obtained with the help of $\hat{\xi}$) generates the sums of squares necessary for comparisons of models with q variables using Mallow's coefficient.

The results presented in this section permit the utilization of WLS regression programs for the selection of covariates in polytomous and proportional odds logistic models. Comparisons among selected models can be done using Mallows C_p, or the Wald Statistic λ^* as an approximation of the likelihood ratio statistic (Lawless and Singhal, 1978) in the same way as presented by Hosmer et al (1989).

3. Example

The method of Hosmer et al. (1989) extended for the proportional odds logistic model is used to examine data from a cocoa study done in Brazil. This study was concerned with qualitative evaluation of cocoa areas that were developed under a national expansion program of incentives to cocoa farmers. Even though quantitatively the government program had attained its goals in terms of planted areas, questions related to the use of improper areas for such crops remained to be examined, since this might result in decrease of productivity and necessity of extra expanditure for treatment of the soil.

The response variable was "technical opinion" of the soil specialist visiting the (planted) area and interviewing the owner or manager of the land, with outcomes "excellent", "good", "bad" and "very bad". There were 14 independent variables appearing in the 1173 questionnaires answered by the farmers. Because of incomplete data, only 1100 questionnaires were used in this application.

Initially the full model with 14 variables was fitted using the procedure logistic from SAS. Table 1 shows the maximum likelihood estimates $\hat{\beta}_j$, for the full model along with the standard errors, the weighted least squares estimates, $\tilde{\underset{\sim}{\beta}}$, and their standard errors for the model $\underset{\sim}{Z} = \underset{\sim}{X}'\underset{\sim}{\beta}$, along with the Wald statistic and significance for H_o: $\beta_j = 0$. Then through weighted least squares regression fit of all models with q variables ($q < 14$) and use of $\underset{\sim}{Z}$ (as in (2)) and $\underset{\sim}{W}$ (as in (4)), the models with the smallest Mallow's C(q) were selected. These results are presented in Table 2.

(Table 1 here)

Table 3 presents the maximum likelihood parameter estimates for the selected models. Use of ratio likelihood test for comparison of hierarchical models resulted in that all parameters of model (6) were different from zero.

Table 2

Models with q variables selected by WLS, residual sum of squares, Wald statistic, maximum likelihood ratio, level of significance and Mallow's coefficient.

Variables Included the Selected Models	SSE(q)	λ^*	$-2\ln L(\hat{\beta})$	λ	LS(%)	(df)	C(q)
FER, PROF, Prv1-Prv3, CPE, DEN, LIM	1885.06	53.07	2131.88	61.74	0.00	10	95.99
FER, PROF, Prv1-Prv3, CPE, DEN, LIM, FRVIS	1874.24	42.25	2118.74	47.60	0.00	9	78.89
FER, PROF, Prv1-Prv3, CPE, DEN, LIM, FRVIS, AGE	1862.60	31.61	2105.39	34.25	0.00	8	60.05
FER, PROF, Prv1-Prv3, CPE, DEN, LIM, FRVIS, AGE, ADUB	1853.28	21.29	2095.24	24.10	1.00	7	45.37
FER, PROF, Prv1-Prv3, CPE, DEN, LIM, FRVIS, AGE, ADUB, Def1, Def2	1841.71	9.72	2082.84	11.70	5.00	5	28.75
FER, PROF, Prv1-Prv3, CPE, DEN, LIM, FRVIS, AGE, ADUB, Def1, Def2, FRASS2	1834.12	2.13	2074.71	3.54	25.00	4	17.07

Table 3

Maximum likelihood estimates for the selected models

Variable	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)
FER	0.643	0.710	0.729	0.768	0.712	0.707
PROF	0.831	0.833	0.830	0.826	0.889	0.837
Prv1	- 1.203	- 1.093	- 1.120	- 1.129	- 1.171	- 1.134
Prv2	- 1.247	- 1.237	- 1.269	- 1.275	- 1.534	- 1.169
Prv3	0.244	0.220	0.115	0.105	0.178	0.193
Def1					- 0.527	- 0.546
Def2					- 0.441	- 0.461
ADUB				0.541	0.522	0.506
CPE	1.126	1.159	1.094	0.918	0.885	0.788
DEN	1.658	1.637	1.619	1.603	1.543	1.558
AGE			0.107	0.103	0.103	0.100
LIM	0.630	0.610	0.622	0.589	0.592	0.560
FRVIS		0.207	0.202	0.200	0.203	0.187
FRASS2						0.398

Analysis for the dichotomous situation of first two and last two categories being put together and also of the proportional odds model with k = 3 categories where first and second categories were again put together were also performed and showed very similar results to the analysis with four outcomes (k = 4) presented above. The estimates retained under each of the three analyses showed to be stable in relation to inclusion or exclusion of new variables to the models. Results obtained by applying this technique were coincident with those obtained by using the Stepwise procedure from SAS.

While statistical methods discussed in this paper were motivated by a cocoa study, the methodology presented is suitable for data from other research areas as well. The contribution of this work is to offer an alternative way of performing polytomous and proportional odds logistic analysis to the Stepwise procedure only available in expensive statistical packages. The problem of selecting independent variables to be employed in a logistic model may be major if there is no access to specific statistical programs and so, these extensions give the researcher an additional tool to perform his/her analysis using weighted least squares regression which is available in most basic statistical packages.

ACKNOWLEDGEMENTS

The authors thank Professor Gary Koch for his comments and Cara Davis for manuscript preparation.

REFERENCES

Begg, C.B. and Gray, R. (1984). Calculation of polychotomous logistic regression parameters using individualized regression. *Biometrika* 71, 11 – 18.

Draper, N.R. and Smith , H. (1981). "Applied Regression Analysis". 2nd edition. J. Willey & Sons, New York.

Green, J.P. (1984). Iteratively reweighted least square for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of Royal Statistics Society, Series B* 46, 149 – 192.

Hastie, T. & Tibshirani, R. (1987). Nonparametric logistic and proportional odds regression. *Applied Statistics* 36, 260 – 276.

Hosmer, W.D., Jovanovic, B. and Lemeshow, S. (1989). Best subset logistic regression. *Biometrics* 45, 1265 – 1270.

Lawless, J.F. and Singhal, K. (1978). Efficient screening of non normal regression models. *Biometrics* 34, 318 – 327.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of Royal Statistical Society, Series B* 42, 109 – 142.

McCullagh, P. and Nelder, J.A. (1989). "Generalized Linear Models". 2nd edition London, Academic, Chapman and Hall.

Thompson, R. and Baker, R.J. (1981). Composite linear models. *Applied Statistics* 30, 125 – 131.

Table 1

Parameter estimates of the full proportional odds model, Wald Statistic and level significancefor H_o: $\beta_j = 0$.

Variable	Abbreviation	M.L. Estimate	M.L. S.E.	W.L.S. Estimate	W.L.S. S.E.	Wald Statistic	Level of Significance(%)
Constant	Const1	− 11.363	0.734	− 11.228	0.610	239.449	0.01
	Const2	− 8.273	0.692	− 8.218	0.547	142.850	0.01
	Const3	− 5.615	0.664	− 5.564	0.507	71.569	0.01
Soil Type	Sol1	0.226	0.176	0.225	0.132	1.644	19.20
	Sol2	− 0.007	0.274	− 0.007	0.205	0.007	97.88
Soil Fertility	FER	0.615	0.136	0.609	0.102	20.609	0.01
Soil Depth	PROF	0.881	0.156	0.874	0.118	31.910	0.01
Temporary Shading	Prv1	− 1.190	0.207	− 1.180	0.157	33.055	0.01
	Prv2	− 1.176	0.182	− 1.170	0.139	41.523	0.01
	Prv3	0.190	0.164	0.189	0.122	1.352	24.49
Definite Shading	Def1	− 0.522	0.155	− 0.515	0.116	11.343	0.08
	Def2	− 0.446	0.212	− 0.440	0.158	4.437	3.52
Use of Fertilizer	ADUB	0.512	0.172	0.510	0.129	8.842	0.29
Control of Insects	CPE	0.803	0.177	0.801	0.133	20.643	0.01
Distance Between Plants	DEN	1.570	0.146	1.554	0.115	115.393	0.01
Problems with Finding Workers	PMO	− 0.061	0.132	− 0.061	0.099	0.213	64.46
Age of Plants	AGE	0.094	0.030	0.093	0.023	9.860	0.17
Number of Clean-Ups of Area	LIM	0.562	0.090	0.556	0.068	38.897	0.01
Number of Owner's Visits to the Land	FRVIS	0.197	0.060	0.195	0.045	10.907	0.10
Frequency of Technical Assistance	FRASS2	0.389	0.143	0.386	0.107	7.454	0.63
Size of Planted Area	STRATUM	0.016	0.028	0.016	0.021	0.310	57.78

RELATÓRIOS DE PESQUISA — 1994

01/94 Stability of the Sucker Rod's Periodic Solution — *Aloisio Freiria Neves.*

02/94 A New Strategy for Solving Variational Inequalities in Bounded Polytopes — *Ana Friedlander, José Mario Martínez* and *Sandra Augusta Santos.*

03/94 Prime and Maximal Ideals in Polynomial Rings — *Miguel Ferrero.*

04/94 A New Globalization Strategy for the Resolution of Nonlinear Systems of Equations — *Ana Friedlander, Márcia A. Gomes-Ruggiero, José Mario Martínez* and *Sandra Augusta Santos.*

05/94 An Ambrosetti-Prodi Type Result for a System of Elliptic Equations via Leray-Schauder Degree — *Daniel Cordeiro de Morais Filho.*

06/94 On Attractivity of Discontinuous Systems — *Marco Antonio Teixeira.*

07/94 Weakly Elliptic Systems of Variational Inequalities: a 2×2 Model Problem with Obstacles in both Components — *D. R. Adams* and *H. J. Nussenzveig Lopes.*

08/94 A New Method for Large-Scale Box Constrained Convex Quadratic Minimization Problems — *Ana Friedlander, José Mario Martínez* and *Marcos Raydan.*

09/94 Weak - Strong Continuity of Multilinear Mappings Pelczyǹski - Pitt Theorem — *Raymundo Alencar* and *Klaus Floret.*

10/94 Removable Singularities Theorems for the Yang-Mills Functional — *Antonella Marini.*

11/94 Magneto-Micropolar Fluid Motion: Existence and Uniqueness of Strong Solution — *Marko A. Rojas-Medar.*

12/94 Absolutely Summing Analytic Operators and the Generalized Khintchine Inequality — *Mário C. Matos.*

13/94 Minimal Immersions of Surfaces Into n-Dimensional Space Forms — *Irwen Valle Guadalupe.*

14/94 Inexact - Newton Methods and the Computation of Singular Points — *Daniel N. Kozakevich, José Mario Martínez* and *Sandra Augusta Santos.*

15/94 Existence and non-existence of radial solutions for elliptic equations with critical growth in R^2 — *Djairo G. de Figueiredo* and *B. Ruf.*

16/94 Numerical Solution of the Leptonic Model — *M. F. Tome, J. M. Martínez* and *Waldyr A. Rodrigues Jr.*

17/94 On Acyclic Knots — *Ricardo N. Cruz.*

18/94 **Optimal Stopping Time for a Poisson Point Process** — *Nancy Lopes Garcia.*

19/94 **Field Theory of the Spinning Electron: I – Internal Motions** — *Giovanni Salesi* and *Erasmo Recami.*

20/94 **Field Theory of the Spinning Electron: II – The New, Non-Linear Field Equations** — *Erasmo Recami* and *Giovanni Salesi.*

21/94 **The Weak Solutions and Reproductive Property for a System of Evolution Equations of Magnetohydrodynamic Type** — *Marko A. Rojas-Medar* and *José Luiz Boldrini.*

22/94 **A Fixed Point Theorem of Banach in the Fuzzy Context** — *Marko A. Rojas-Medar, Heriberto Román-Flores* and *Rodney C. Bassanezi.*

23/94 **Spectral Galerkin Approximations for the Equations of Magneto Hydrodynamic Type: Local in Time Error Estimates** — *Marko Rojas-Medar* and *José Luiz Boldrini.*

24/94 **Solving Nonsmooth Equations by Means of Quasi-Newton Methods with Globalization** — *Marcia A. Gomes-Ruggiero, José Mario Martínez* and *Sandra Augusta Santos.*

25/94 **Os Trabalhos de Leopoldo Nachbin (1922-1993)** — *Jorge Mujica.*

26/94 **On a New Class of Polynomials** — *D. Gomes* and *E. Capelas de Oliveira.*

27/94 **Level-Convergence and Fuzzy Integral** — *H. Román-Flores, A. Flores-Franulic* and *Rodney C. Bassanezi, Marko A. Rojas-Medar.*

28/94 **Estimating Leaf Area Index for Canopies using Coverage Processes** — *Nancy Lopes Garcia.*

29/94 **The Isoperimetric Problem in Cylindrical Spaces I - Symmetrization and General Results** — *Renato H.L. Pedrosa.*

30/94 **The Isoperimetric Problem in Cylindrical Spaces II - Classification Results** — *Renato H.L. Pedrosa.*

31/94 **Chain Control Sets for Semigroup Actions on Homogeneous Spaces** — *Carlos José Braga Barros* and *Luiz A.B. San Martin.*

32/94 **Linear Semigroups Acting on Stiefel Manifolds** — *Luiz A.B. San Martin.*

33/94 **Wave Series Expansions for Stratified Acoustic Media** — *Lúcio Tunes dos Santos, Bjørn Ursin* and *Martin Tygel.*

34/94 **Estudo de Fractais gerados por Iterações de Ponto Fixo** — *Ricardo Caetano Azevedo Biloti* and *Lúcio Tunes dos Santos.*

35/94 **On Global Controllability of Discrete-Time Control Systems** — *Luiz A. B. San Martin.*

36/94 **Closed Formulas for the Exponentiation of Lie Algebra Elements and Applications to Kaluza Klein Theories** — *Alex I. Shimabukuro* and *Márcio A. F. Rosa.*

37/94 **On Thirring's Approach to Mach's Principle: Criticisms and Speculations on Extensions of his Original Work** — *Fabio M. Peixoto* and *Márcio A. F. Rosa.*

38/94 **The Clifford Bundle and the Dynamics of the Superparticle** — *W.A. Rodrigues, Jr., J. Vaz, Jr.* and *M. Pavšič.*

39/94 **On the Level-continuity of Fuzzy Integrals** — *H. Román-Flores, A. Flores-Franulic, R.C. Bassanezi* and *M.Rojas-Medar.*

40/94 **On the Resolution of the External Penalization Problem** — *José Mario Martínez* and *Lúcio Tunes dos Santos.*

41/94 **On the Convergence of Quasi-Newton Methods for Nonsmooth Problems** — *Vera L. R. Lopes* and *José Mario Martínez.*

42/94 **On the Phenomenology of Tachyon Radiation** — *Ron Folman* and *Erasmo Recami.*

43/94 **Nonsingular Charged Particles in a Classical Field Theory** — *Werner M. Vieira.*

44/94 **Gradient-like Flows on High Dimensional Manifolds** — *R. N. Cruz* and *K. A. de Rezende.*

45/94 **Critical Point Theorems and Applications to a Semilinear Elliptic Problem** — *E. Alves de B. e Silva.*

46/94 **On the Convergence Rate of Spectral Galerkin Approximations for a Generalized Boussinesq Model** — *Sebastián A. Lorca* and *José Luiz Boldrini.*

47/94 **Um Modelo Para a Dengue Hemorrágica com Força de Infecção Periódica** — *R.C. Bassanezi, R. Zotin* e *M.B.F. Leite.*

48/94 **Approximations by Periodic Biorthogonal Wavelets: Numerical Algorithms** — *Sônia M. Gomes, Elsa Cortina* and *Irene Moroz.*

49/94 **Incompressible Tori Transverse to Anosov Flows in 3-Manifolds** — *Sérgio R. Fenley.*

50/94 **Non-Equilibrium Waterflood in Heterogeneous Porous Media** — *Cristina Cunha, Deise Ferreira* and *Antonio C. Corrêa.*

51/94 **Fixed Points Theorem and Stability Results for Fuzzy-Multivalued Mapping** — *Marko A. Rojas-Medar* and *R. C. Bassanezi .*

52/94 **Density Estimation via Hybrid Splines** — *Ronaldo Dias.*

53/94 **The Equations of a Viscous Incompressible Chemical Active Fluid II: Regularity of Solutions** — *M.A. Rojas-Medar* and *S.A. Lorca.*

54/94 **Existence of Solutions for Quasilinear Elliptic Equations** — *João Marcos Bezerra do Ó.*

55/94 **Multiple Critical Points for Asymptotically Quadratic Functionals** — *E. Alves de B. e Silva.*

56/94 **On the Differentiability of Functions in Banach Spaces** — *Marko A. Rojas-Medar* and *Heriberto Román-Flores.*

57/94 **A Model for Distributions and Correlation Structure of Multivariate Polytomous Data From a Two-Stage Cluster Sample** — *Eliana H. de F. Marques* and *Gary G. Koch.*