Mýnmad Logýstýc Regressýon in Multý-Objectýve Programmýng

Assoc.Prof.Dr. Ayº en APAYDIN
Ankara University, Faculty of Science
Department of Statistics, Ankara, TÜRKÝYE
e-mail: apaydin@science.ankara.edu.tr

Res.Assist. Nimet YAPICI
Selcuk University, Art&Science Faculty
Department of Statistics, Konya, TÜRKÝYE
e-mail: nimet@selcuk.edu.tr

Instr.Özlem Müge AYDIN
Baskent University, Faculty of Health Sciences,
Department of Health Care Administration,
Ankara,TÜRKÝYE
e-mail: ozlem@baskent.edu.tr

1.Introduction

Classifying multivariate statistical data is a necessary type of analysis for various statistical methods in applications. There are three methods for grouping the observations. These are; clustring analysis, discriminant analysis and logistic regression analysis. In clustring analysis, for the observations it is unknown how many group numbers they have been assigned. But in discriminant and logistic regression analysis the group numbers have been known[2,4].

Logistic model has been used since 1845. It is generally used in searches including social-economic topics. In earlier times, it was used for identifying an analytic mathematical model of population increasement for socities[2].

Under the lack of the assumptions of normality, common variance, logistic regression analysis is an alternative one for discriminant analysis and crosswise tables. If the dependent (response) variable is a discrete variable having double or more levels like 0 and 1; because of the lack of normality assumptions; logistic regression model is an alternative method to linear regression anlaysis. Because of having no constraints of the assumptions, the solved model is very flexible and easy interpereted. So logistic regression model is preferred [2,4].

Least square method is used for estimating the parameters in linear regression models. But this method has some disadvantages, so there may be difficulties in operations. Because of this reason in the mathematical programming models; MINMAD method is used alternatively for this approach. In the first study done in 1955; for the linear regression models MINMAD method is used and it is established a relation between MINMAD and linear programming method [1,3].

In this study, firstly, logistic regression model is modelled as MINMAD problem. Then, to solve for this constituted problem is used multi-objective programming approach.

2. Logistic Regression Model

Multiple regression model is defined as follows,

$$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i \qquad i=1,2,...,n$$

(1)Where, $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$. Equation (1) can be rewritten by the means of expected value as,

$$E(Y_i / x_{i1}, x_{i2}, ..., x_{ik}) = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i \qquad (2)$$

In this model, while there is a continiousness condition on dependent variable Y. So that dependent variable Y can be valued with all values between $-\infty$ and $+\infty$. In Equation (1), Y having only values 0 and 1 this rule is omited and as $P(Y_i = 1)$, (the probability of i th observation valued 1) then the expected value is;

$$E(Y_i) = 1 \times P(Y_i = 1) + 0 \times P(Y_i = 0) = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i \qquad (3)$$

Equation (3) can have the probabilities in the interval [0,1]. This model is called “linear probability model”. In this model, when dependent variable is 0 or 1, the error term ε_i is ,

$$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i = 0 \Rightarrow \varepsilon_i = -\beta_0 - \sum_{j=1}^{k} \beta_j X_{ij} \quad (4)$$

$$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i = 1 \Rightarrow \varepsilon_i = 1 - \beta_0 - \sum_{j=1}^{k} \beta_j X_{ij} \quad (5)$$

Regarding to the error term the assumptions

$$E(\varepsilon_i) = 0 \quad (6)$$

and

$$Var(\varepsilon_i) = P(Y_i = 1) \times [1 - P(Y_i = 1)] = [-\beta_0 - \sum_{j=1}^{k} \beta_j X_{ij}] \times [1 - \beta_0 - \sum_{j=1}^{k} \beta_j X_{ij}] \quad (7)$$

are proved. Error terms has the binom distibution with such parameters [2,4].

3. Multi-Objective Programming For The MINMAD Logistic Regression Model

The first study about the application of the optimization methods in statistics is made in 1955. In this study, minimum absolute mean deviation (MINMAD) also known as the L_1-norm is used in regression model and it is formulated as linear programming and solved by simplex algorithm [1,3].

Later, for the MINMAD regression problems solution modelled as linear programming problem solved by the simplex algorithm. In this algorithm, some modifications has been applied for variables which one enter and remove from basis. Then in the selection of basic and nonbasic variables, an algorithm was given with changes[3].

There are a lot of studies having used both L_1-norm and optimization methods. In solving the multiple linear regression problems by linear programming methods, making use of simplex method, for the purpose of special methods and unbiased estimators, special algorithms put forward. Using the relationship between primal and dual problems, there can be done an analysis of duality and sensitivitv[3]

For the logistic regression, MINMAD regression approach is considered an,

$$\begin{aligned} &Minimize \sum |d_1| \\ &Minimize \sum |d_2| \\ &X\beta + d_1^+ - d_2^+ = 0 \\ &X\beta + d_1^+ - d_2^+ = 1 \\ &\beta \text{ isareti belirtilmemis} \\ &d_1^-, d_1^+, d_2^-, d_2^+ \geq 0 \end{aligned} \quad (8)$$

is obtained. Where $|d_1| = [-\beta_0 - \sum_{j=1}^{k} \beta_j X_{ij}]$ and $|d_2| = [1 - \beta_0 - \sum_{j=1}^{k} \beta_j X_{ij}]$.

REFERENCES

1. **Arthanari,T.S., Dodge, Y., "**Mathematical Programming in Statistics**",** John Wilwy&Sons Inc., 1981

2. Gürcan, M.,"Logistic Regression Analysis and An Application", MSc Thesis (unpublished), 19 Mayýs University, Graduate School of Natural and Applied Science, Samsun, 1998

3. Sanlý, K., "Parameter Estimation of Randomized Block Design Model Equation by MINMAD Method", MSc Thesis (unpublished), Ankara University, Graduate School of Natural and Applied Science , Ankara, 1999.

4. **Tatlýdil, H.,** "Multivariate Statistical Analysis with Application", Academy Press, Ankara, 1996

Resumé

Dans ce travaille, on proposera l' approche de "plusieurs buts programme" pour le modele logistique regression qui a modelisé par "MÝNMAD" probléme.