

Marginal Regression for a Bi-variate Response with Diabetes Mellitus Study

¹G. V. Arunamayi, ²K.Nagendra Kumar, , ³Bisrat Misganaw
^{1,2,3}Department of Statistics, Andhra University, Visakhapatnam-530 003, Andhra Pradesh.

ABSTRACT

In this paper, we have developed an “A Bivariate response” model to determine for ‘A Diabetic Mellitus’ Study affects large number of people of all social conditions throughout the world. Continuous to grow despite of existing advances in the past few years in virtually every field of diabetes research and in-protect care for improved treatment. This is sometimes accompanied by symptoms of severe thirst, profuse urination, weight loss and stopper. We tested by SPSS software by taking 200 samples and using Logistic Regression to estimate the relationship between the response probability whether a diabetic patient had B.P. or not.

Keywords- Binary Logistic Regression, logit model, Odds Ratio, Model validation, Hosmer and Lemeshow Test.

Date of Submission: 11 May 2017



Date of Accepted: 30 May 2017

I. INTRODUCTION

Classical linear models and least square began with the work of Gauss and Legendre (Stigler, 1981) who applied the method to astronomical data. Their data were usually measurements of continuous quantities such as the position and magnitude of the heavenly bodies, and at least in the astronomical investigations, the variability in the observations was largely the effect of measurement error. The normal or Gaussian distribution was a mathematical construction developed to describe the properties of such errors; later in the nineteenth century the distribution was used to describe the variations of individual in biological population in respect of a character such as height, an application quite different in kind from its use for describing measurement error, and leading to the numerous biological applications of linear models.

Gauss introduced the normal distribution of errors as a device for describing variability but he showed that important properties of least squares estimates depend not on normality but on one assumption of constant variance. A closely related property applies to all generalized linear models. In other words, although we make reference at various points to standard distributions such as Normal, Binomial, Poisson, Exponential or Gamma, the second order properties of the parameter estimates do not depend on the assumed distributional form but only on the variance to mean relationship and uncorrelated.

Another standard in the history of statistics is the development of methods for dealing with discrete events rather than with continuously varying quantities. The enumeration of probabilities of configuration in cards and dice was a matter of keen interest in the eighteenth century, and from this grew methods of dealing with data in the form of counts of events. The basic distribution here is that named after the Poisson and it has been widely applied to the analysis of such data, which may refer to many diverse kinds of events: a famous example concerns unfortunate soldiers killed by Prussian horses (Bortkewitsch, 1898); others might concern the infection of an agar plate in a pathology experiment, the presence or absence of disease in medical patients, arrival of neutrons from source and so on. Generalized linear models allow us to develop models for the analysis of counts analogous to classical linear models for continuous quantities. Closely related are the models for the analysis of data in the form of proportions, or ratios of counts, such as occur when for example disease incidence is related to factors such as age, social classes, exposure to pollutants and so on.

The primary objective of this work is the study of the dependence of the probability of ‘success’ of a binary response variable, Y on a vector X of explanatory variables. Therefore a key issue is how such dependencies are to be specified mathematically.

The introduction of specific models for dependence is much more recent, and in a sense stems from the analysis of dose-response curves in bioassay (Finney, 1952, 1964) and more specifically from the use of maximum likelihood to estimate the 50% point of such curves (Fisher, 1935).

The importance of logistic distribution in bioassay and the advantages of estimates obtained by weighted least squares applied to transformed values (minimum logit chi-squared) were stressed in a series of papers by Berkson (1944, 1951, 1953, 1957, 1960). Another use of the empirical logit transform was treated by Woolf (1955). Much of the literature on the logistic binary regression focused largely on the logistic curve as a

alternative to the probit (integrated, normal), the discussion of relative merits being somewhat confounded with a dispute about maximum likelihood versus empirically weighted least square (minimum, logit chi-squared) as a point estimation procedure for slope and 50% point modifications to the empirical logit transform to remove bias originate with Haldane(1955) **and have been very thoroughly studied by Gant, Pettigrew and Thomas(1985).**

The broader use of logistic regression probably dates from the work connected with the framing harm study(a major US hypertension investigation), especially the contribution of J.cornfield .

II. REVIEW OF LITERATURE

Guangyong Zou and Allan Donner(2004) studied confidence Internal Estimation of the Intra-class Correlation Coefficient for Binary Outcome Data. They obtained closed-form asymptotic variance formulae for three point estimator of the intra-class correlation coefficient that may applied to binary outcome data arising in clusters of variable size. Their result include a special case those that have previously appeared in the literature(Fleiss and Cuzick, 1979, Applied Physiological Measurement 3, 537-542; Bloch and kraemer,1989,Biometrics 45,269-287; Altaye, Donner, and Klar,2001, Biometrics 57, 584-588). Simulation results indicate that confidence intervals based on the estimator proposed by Fleiss and Cuzick provide coverage levels close to nominal over a wide range of parameter combinations.

Brain J. Reich, James S. Hodges, and Bradley P. Carlin(2007) analyzed attachment loss data that can be used as conditionally autoregressive (CAR) prior distribution that smooth fitted values toward neighboring values. However, it may be desirable to have more than one class of neighbor relation in the spatial structure, so that the different classes of neighbor relations can induce different degrees of smoothing,. Adequate modeling of the spatial structure may improve the monitoring of periodontal disease progression. They showed that the prior distribution on these parameters has little effect on the posterior of the fixed effects but has a marked influence on the posterior of both the random effects and the smoothing parameters. Their analysis of attachment loss data also suggests that the spatial structure itself varies between individuals.

Nicole H. Augustin, Stefan Lang, Monica Musio and Klaus von Wilpert (2007) jointly monitored a survey which was carried out in 1994 in the forests of Baden-Wurttemberg, a federal state in the south-western region of Germany. The survey is a part of large monitoring scheme that has been carried out since the 1980s at different spatial and temporal resolutions to observe the increase in forest damage. One indicator for tree vitality is tree defoliation to observe the increase in forest damage. Once indicator for tree vitality is tree defoliation, which is mainly caused by intrinsic factors, age and stand conditions, but also by biotic (e.g. insects) and a biotic stresses (e.g. industrial emissions). In the survey, needle loss of pine-trees and many potential covariates are recorded at about 580 grid points of a 4 km x 4 km grid. The aim is to identify a set of predictors for needle loss and to investigate the relationships between the needle loss and the predictors. The response variable needle loss is recorded as a percentage in 5% steps estimated by eye using binoculars and categorized into healthy trees (10% or less), intermediate trees (10-25%) and damaged trees 25% or more). They used a Bayesian cumulative threshold model with non-linear functions of continuous variables and a random effect for spatial heterogeneity, for both the non-linear functions and the spatial random effect we use Bayesian versions of P-splines as priors. Our methods are novel in that it deals with several non-standard data requirements; the ordinal response variable (the categorized version of needle loss), non-linear effects of covariates, spatial heterogeneity and prediction with missing covariates. The model is a special case of models with a geoadditive or more generally structured additive predictor. Inference can be based on Markov chain Monte Carlo techniques or mixed model technology.

Jerome A. Dupuis (2006) considered the problem of estimating the number of species of an animal community. It is assumed that it is possible to draw up a list of species liable to be present in this community. Data are collected from quadrat sampling. The parameterization enables us to incorporate prior information on the presence detect ability, and spatial density of species. Moreover the elaborated procedures to build the prior distributions on these parameters from information furnished by external data. A simulation study is carried out to examine the influence of different priors on the performances of our estimator.

Bo Cai and David B. Suson (2006) generalized linear mixed model (GLMM), which extends the generalized linear model (GLM) to incorporate random effects characterizing heterogeneity among subjects, is widely used in analyzing correlated and longitudinal data. Although there is often interest in identifying the subset of predictors that have random effects, random effects selection can be challenging, particularly when outcome distributions are non-normal. They proposed a fully Bayesian approach to the problem of simultaneous selection of fixed and random effects in GLMMs. Integrating out of random effects induces a covariance structure on the multivariate outcome data, and an important problem that we also consider is that of covariance selection. Their approach relies on variable selection-type mixture priors for the components in a special Cholesky decomposition of the random effects covariance. A stochastic search MCMC algorithm is developed, which relies on Gibbs sampling with Taylor series expansions used to approximate intractable integrals.

Simulated data examples are presented for different exponential family distributions, and the approach is applied to discrete survival data from a time-to-pregnancy study.

Alan E. Gelfand, Athanasios Kottas, and Steven N. Maceachern (2005) considered that customary modeling for continuous point-referenced data assuming a Gaussian process that is often taken to be stationary. When such models are fitted within a Bayesian framework, the unknown parameters of the process are assumed to be random, so a random Gaussian process results. Here they proposed a novel spatial Dirichlet process mixture model to produce a random spatial process that is neither Gaussian nor stationary. They first developed a spatial Dirichlet process model for spatial data and discuss its properties. Because of familiar limitations associated with direct use of Dirichlet process models. They introduced mixing by convolving this process with a pure error process. They examined the properties of models created through such Dirichlet process mixing. In the Bayesian framework, they implemented posterior inference using Gibbs sampling, spatial prediction raises interesting questions, but these can be handled. Finally they illustrated the approach they illustrated then approach using simulated data, as well as dataset involving precipitation measurements over the Languedoc – Roussillon region in the south France.

Howard D. Bondell (2005) considered the model from the case control view point, it is clear that current techniques can exhibit poor behavior in many common situations. A new robust classes of estimation procedures is introduced. The estimators are constructed via minimum distance approach after identifying the model with a semi parametric based sampling model. The approach is developed under the case control sampling scheme, yet is shown to be applicable under prospective sampling as well.

Dietmar Zellnar, Frieder Keller, and Gunter E. Zellner (2004) overcome the problems with selection procedures to obtain the authentic variables, they compared the performance of stepwise selection procedures with a bagging method proposed Sauerbraten. The use of re-sampling methods to regression models in the statistics. The bootstrap method with a variable selection from the full logistic regression model was applied. Logistic regression models were conducted to compare the performance of the selection procedures.

K.S Chan and H. Tong (2004) studied testing for multimodality with dependent data. They proposed a test for multimodality with dependent data by resampling from a suitable constructed transition probability Kernel, which includes Silverman's test with independent data as a special case. They extended some theoretical properties of Silverman's test with independent and identical distributed data to weekly dependent data, and also discussed the robustness of Silverman's test against departure from independence.

III. DIABETES MELLITUS

Diabetes afflicts larger number of people of all social conditions throughout the world. The personal and public health problem of diabetes. Already of vast proportions. Continues to grow despite of existing advances in the past few years in virtually every field of diabetes research and in-patient care for e.g. improved treatment. Protection against complications increased self-care and even prevention of some forms of diabetes. The underlying cause of diabetes is the defective production or action of the hormone insulin.

Definition:

In the untreated state, diabetes mellitus is recognized by chronic elevation of the concentration of glucose in the blood (Hyperglycemia). This is sometimes accompanied by symptoms of severe thirst, profuse urination. Weight loss and stupor. Culminating in coma and death in the absence of effective treatment. The high concentration of blood glucose and other biochemical abnormalities result from deficient production or action of insulin, a hormone that controls glucose, fat and amino acid metabolism. The severity of its symptoms is largely determined by the degree to which the insulin action is deficient.

IV. DIFFERENT TYPES OF DIABETES

The basis for the distinction between the major sub classes of diabetes mellitus is the patient dependence for survival on insulin. Such dependence is judged to be present when the "classical" symptoms of diabetes (increased thirst, polyuria, wasting and ultimately and coma) are associated with greatly raised concentrations of glucose and ketone bodies in the blood and urine. All other patients meeting the glycemic criteria for diabetes fall into the non-insulin dependent (NIDDM) class unless they have malnutrition-related or gestational diabetes or are among the small number who qualify for one of the other special categories.

Insulin dependent diabetes (IDDM) can occur at any age though it most commonly occurs during youth. This type of diabetes used to be known as juvenile-onset diabetes. NIDDM; additional insulin is not usually required to sustain life. NIDDM is much more common than IDDM. This type is less severe than IDDM and starts more slowly. Some patients with type of diabetes produce some or at times even excessive, insulin in their pancreas. It either is not enough for proper function or is not being produced quickly enough to maintain insulin glucose levels in the blood effectively. This type of disease occurs in middle or older age and frequently in

overweight people it can most often be controlled by diet alone poor by a combination of diet, exercise and oral meditation.

4.1. Causes of Diabetes

Although the cause of diabetes is still unknown, medical science does know that certain factors contribute to its development. One factor is heredity a child of non-diabetic can become diabetic, however since the disease may skip generations because of genetic coding that prevents it from appearing in every generation. Stresses that affect the cells of the body seem to set the stage for diabetes in these people. One such stress is extra weight. Obesity, affecting insulin utilization. Contributes to disease.

Stresses can be emotional or physical such as surgery or a serious infection, an accident or emotional shock. Many medications affect the body in a stressful way. Pregnancy also places extra stresses on the body, and diabetes is often diagnosed in pregnant women or women who have repeated miscarriages.

4.2. Diagnosing Diabetes:

When adequately distributed throughout body, glucose gives the energy. If the glucose level in blood or urine high, it means that the glucose is not being channeled properly. This happens when too little of the harmonic insulin is produced or when the insulin that is produced is not used appropriately by the body. Insulin is responsible for clearing sugar out of the blood stream and getting it into cells, where it is stored or used. In a diabetic because enough insulin is not there or is not working properly glucose just accumulates in the blood. After a certain point the blood becomes so high that the kidney, which filters waste products from the blood, being to overflow some of the glucose into the urine. Glucose level can be measured through several tests using urine and blood specimens.

Urine testing is part of most physicians' routines in giving complete physical examinations. Urine tests are basic diagnostic tools that can give physicians valuable information about how the body is working in a form of urine testing known as semi quantitative, the doctor may use either a tablet or one several brands of paper or plastic strips to determine the approximate sugar in the urine. Only a small sample of urine is needed, and the test takes only a few seconds.

An advantage of semi quantitative urine testing is that it is fast and can be easily by the doctor. The disadvantage of urine tests, in general, however is that they are less accurate and harder to interpret than blood tests. Urine test readings may not always accurately reflect the level of glucose in the blood.

Blood glucose tests are used to indicate whether or not too much or not enough sugar in the blood present. The easiest and fastest procedure for measuring glucose in the blood is to use plastic or paper strips like those used in urine tests. These strips give a range, not an exact value. Other methods called rapid analytical system are used by many physicians. These tests use chemicals to test blood samples in a rapid single-step process

Blood glucose concentration is one of the best measurements of the presence of diabetes as well as diabetic control. But because this test measures the blood glucose level at only one instant, it may not represent patient's usual condition. Evaluating the fasting or before glucose intake level indicates the extent to which carbohydrates from the patient's last meal have been removed from the blood stream. The fasting glucose concentration is the measure of blood glucose after a lack of intake for more than four hours or overnight.

Diabetes can be controlled through effective management of the balance of glucose and insulin, to enable the body function well. Similarly, by Using simple regression $FG=297.67-1.909 AGE$

4.3. Stepwise Regression Method:

The explanatory variable $x_i(i=1,2,3,\dots,8)$ viz AGE, income, gender, weight, type of disease (TD), family history(FH), blood pressure (BP), (presence or absence) disease history (DH)(i.e., the no. of completed years after first diagnosis) be related to the dependent variable fasting blood glucose level(FG) the data are included in the APPENDIX for significant variable only. The minimum income of the patients was Rs. 4500 p.m. and the maximum was Rs.8000 p.m.

A stepwise Regression method was carried out using the statistical package SPSS to determine the relationship between the dependent variable FG and the explanatory variables.

4.4. Diabetes and Hypertension:

Diabetes is not a disease is not a disease of the blood sugar alone. Like no other disease, diabetes involves our entire body from out mind to all organ systems. Therefore, controlling not only blood sugar but the factors controlling blood sugar as well, is important High blood pressure is a worldwide condition of almost epidemic proportions. Yet, there are no table symptoms. It is often overlooked or treated casually. Although hypertension is more common among diabetics than among non-diabetics does not mean he or she will have hypertension. In some cases diabetics, high blood pressure, kidney, diseases are linked closely. Diabetics who

have kidney disease often develop high blood pressure as a result of the impaired function of their kidneys. Because the problems of diabetics, hypertension and kidney diseases are so closely interrelated preventing or controlling one disease helps to reduce the complications of others.

Suresh (2005) had studied the dependence of the presence or absence of the response variable BP on the covariates AGE, DH, and TD was considered and discussed. The diabetic patients who were diagnosed for BP at a later date were coded as 1 and others as 0 for the purpose of analysis and treated as a binary response. The other covariates that were excluded in the linear regression as discussed in the previous chapter were ignored in this study also.

4.5. Binary Logistic regression

Suppose that for each individual or experimental unit the response Y_i can take only one of two possible values, denoted for convenience each by 0 and 1 respectively. Observations of this nature arise for four example in medical trials where at the end of the trial period the patient has either recovered $Y=1$ or has not $Y=0$ similarly, in this study a patient has blood pressure.

$$P(Y_i = 0) = 1 - \theta_i$$

$$P(Y_i = 1) = \theta_i$$

In most investigations, we have associated with each individual or experimental unit a vector of covariates or explanatory variables (x_1, x_2, \dots, x_k) . the covariates may be qualitative, quantitative or a combination of both. The principal objective is to investigate the relationship between the response probability θ and the explanatory variables x_1, x_2, \dots, x_k . Often otherwise, a subset of the x is of primary importance but due allowance must also be made for the effect of the remaining covariates.

The simplest empirical regression is to suppose that θ_i , the value of θ for the i^{th} individual, is linearly related to the explanatory variables, i.e., for the i^{th} individual.

$$\theta_i = \beta_0 + \sum_{j=1}^k x_{ij} \beta_j \quad \text{---} \quad (4.5.1)$$

Where β_0 is an unknown intercept and β_j 's are the unknown regression coefficients; x_{ij} is the value of the j^{th} explanatory variable on the i^{th} individual with the condition. $0 \leq \theta_i \leq 1$

A model in which is automatically satisfied is the logistic distribution (Cox. D,R et al) with location $\beta_0 + X \beta$ and unit scale.

$$\frac{\exp(\mu - \beta_0 - X \beta)}{\{1 + \exp(\mu - \beta_0 - X \beta)\}}, \text{so that}$$

$$F(0; x) = \frac{1}{\{1 + \exp(\beta_0 + x \beta)\}},$$

From which it follows that

$$\theta = P(Y = 1; x) = \frac{\exp(\beta_0 + x \beta)}{\{1 + \exp(\beta_0 + x \beta)\}}$$

$$1 - \theta = P(Y = 0; x) = \frac{1}{\{1 + \exp(\beta_0 + x \beta)\}}$$

Where X is the vector of explanatory variables and β is the vector of unknown regression coefficients. This relation is linearized by the transformation.

$$\log \left(\frac{\theta}{1 - \theta} \right) = \beta_0 + x \beta \quad \text{---} \quad (4.5.2)$$

Which is called the logit function.

V. LIKELIHOOD FUNCTION FOR BINARY DATA

Let the response y_1, y_2, \dots, y_n correspond to independent random variables Y_1, Y_2, \dots, Y_N . where Y_i is assumed to be binomially distributed with index n_i and parameter θ_i . Ungrouped data are considered with $n_i=1(i=1,2,\dots,N)$. The likelihood considered as a function of the vector

$$\theta = (\theta_1, \theta_2, \dots, \theta_N) \text{ is}$$

$$l(\theta; Y) = \sum_{i=0}^N \left[y_i \ln \left(\frac{\theta_i}{1 - \theta_i} \right) + n_i n (1 - \theta_i) \right] \quad \dots \quad (4.5.3)$$

In particular, if

$$\ln \left(\frac{\theta_i}{1 - \theta_i} \right) = \beta_0 + x\beta$$

We have

$$l(\theta; Y) = \sum_i \sum_j y_i x_{ij} \beta_j - \left[\sum_i n_i \ln \left[1 + \exp \sum_j x_{ij} \beta_j \right] \right] \quad \dots \quad (4.5.4)$$

The β vector of regression coefficients can be estimated by maximizing. Many statistical packages have logistic regression as a module and β can be estimated directly.

5.1. Fitting of a Logistic Regression:

The data on the diabetic patients was to estimate the relationship between the response probability θ (where the dependent variable Y is whether a diabetic patient had B.P. or not) and the relevant covariates are Age, DH,TD. The calculations were performed using the statistical package STATISTIKA by regressing B.P on the covariates AGE, DH and TD obtained the following results.

5.2. Model: Logistic regression

$-2 * \log(\text{likelihood}) = 235.4902$

Chi-square=57303, d.f=3, p=135517

M1 estimates the regression coefficients;

$\hat{\beta}_0 = -1.7466, \hat{\beta}_1 = 0.0121, \hat{\beta}_2 = 0.00614, \hat{\beta}_3 = -0.2314$

Thus the estimated logit function of (from above) is given by

$$\log \left(\frac{\theta_i}{1 - \theta_i} \right) = -1.7466 + 0.012 \text{ AGE} + 0.00614 \text{ DH} - 0.2314 \text{ TD} \quad \dots \quad (5.1.1)$$

Using (5.1.1) the chance that a diabetic patient will be diagnosed for B>P at a later date can be obtained. For instance the patient (s.no=1) whose AGE=30, DH=1 and TD=0 has the chance of 0.21 being diagnosed for BP at a later date. It can be noticed from (5) that DH and AGE are the contributors for the response probability. Of these two DH is the larger contributor with coefficient value 0.0614. this indicates that the effect of DH is more on the fasting blood glucose(FG) levels as seen in the previous chapter and on the chance that a diabetic patient is diagnosed for BP at a later date. So, DH must also be considered as a factor being diagnosed for BP. In such a situation, the problem of interest was to assess a threshold, if any, for the covariate DH. Since AGE and DH are the contributors for BP a threshold model was developed for these covariates.

Table :VII Classification table

Observed	Predicted		
	FG	1.00	Percentage Corrected
Step 1 FG	.00	2	48
	1.00	2	148
Overall Percentage			75.0
Step 2 FG	.00	2	48
	1.00	2	148
Overall Percentage			75.0

a. The cut value is .500

Table VIII Variables in the equations

	B	S.E	Wald	df	Sig.	Exp(B)
Step 1 ^a						
AGE	-.047	.017	8.136	1	.004	.954
DH	.112	.053	4.541	1	.033	1.119
TD	-.210	.417	.254	1	.614	.811
Constant	3.108	.847	13.460	1	.000	22.374
Step 2 ^a						
AGE	-.045	.016	7.991	1	.005	.956
DH	.107	.051	4.349	1	.037	1.113
Constant	2.971	.799	13.833	1	.000	19.521

a. Variable (Entered on step one:AGE,DH,TD).

Table IX Model if Term Removed

Variable	Model Log likelihood	Change in likelihood	-2log	df	Significance of the change
Step 1					
AGE	-111.632	8.566		1	.003
DH	-109.999	5.300		1	.021
TD	-107.475	.251		1	.616
Step 2					
AGE	-111.665	8.380		1	.004
DH	-110.000	5.050		1	.025

a. Variable (Entered on step one: AGE, DH, and TD).

VI. CONCLUSION

In this paper we examined the suitability Logistic regression analysis of a bivariate binary response values. They are two factors are significant out of three factors were tested and identified as having influence significantly performance on AGE and DH. From the SPSS output tables VIII and IX. When the age of the respondents were increase there is a chance of getting diabetes and due to family history also. A random sample of diabetes patients collected from King George Hospital, Visakhapatnam, were interviewed and the information on characteristics such AGE, Disease history(DH), whether the patient has Blood Pressure(BP), family history of the diseases, types of diseases, type of medicine were recorded for each patient. The linear regression, binary regression model and marginal regression were discussed their properties, estimation of parameters and testing aspects for the response variables such as AGE and DH.

REFERENCES

- [1] Allcroft,D.J and Glasbey.C.A.(2003). A latent Gaussian markov random field model for spatiotemporal rainfall disaggregation. *Appl.Statist.* 52, 487-498.
- [2] Baggerly,K.A.(1998). Empirical likelihood as goodness-of-fit measure. *Biomaterial* 85, 535-47.
- [3] Bianco,A.M. & Yohai, V.J.(1996). Robust estimation in the logistic regression model. In *Robust Statistics, Data Analysis and Computer Intensive Methodes, Lecture Notes in Statistics*, 109,Ed.H. Rieder, pp.17-34,Newyork, Springer-Verlag.
- [4] Blotzheim, U.N. and Bauer, K.M.(1997) Pyrrhula, Gimpel,Dompfaff. In *Handbuch der Vogel Mitteleuropas*, Band 14/2, Teil 5, 1130-1181. Wiesbaden: Aula-verlag
- [5] Boot,J.G. & Buttler, R.W.(1990). Randomization distributions with saddle-point approximations in generalized linear models. *Biometrika* 77, 787-96.
- [6] Bunge, J. and Fitzpatrick, M. (1993). Estimating the number of species: A review *Journal of the American Statistical Association* 8,364-373.
- [7] Carroll R.J. & Pederson.S.(1993). On robust estimation in the logistic regression model. *J.R. statist.Soc. B* 55,693-706.
- [8] Caroll R.J, Wang,S. & Wang,C.Y.(1995) Prospective analysis of logistic case-control studies. *J.Am.Statist.Assoc.* 90,157-69.
- [9] Chen.M.H and Shao,Q.M(1999) Existence of Bayesian estimates for the polychotomous quantal response model. *Annals of the institute of statistics and mathematics* 51,637-656.
- [10] Chib,S and Greenberg,E.(1998) Analysis of multivariate probit models. *Biometrika* 85, 347-361.
- [11] Ging Qin and Biao Zhang (2003) Using Logistic Regression Procedures for Estimating receiver Operating Characteristic Curves; *Biometrika*,90,585-596
- [12] Garrett M. Fitzmaurice, Nan. M.Laird, & Gwendolyn SE.P.Zahner(1996), *Multivariate Logistic Models for incomplete Binary Response*; *JASA*,Vol.91,99-108.
- [13] Kleinbaum DG,Klein M.(2002). *Logistic Regerssion: A Self-Learning Text*. 2nd Ed.New York,Springer-Verlag.
- [14] Kutner M.H; Nachtsheim C.J and Neter J. (2004),”*Applied linear regression models (fourth edition)*”
- [15] Lin, D.Y.,Zeger,S.L. & Qaqish,B.(1992). Multivariate regression analysis for categorical data (with discussion). *J.R.statist.Soc.B* 54,340.
- [16] Mitchell C. And Dayton (1992),” *Logistic Regression Analysis*”,University of Maryland.