1. INTRODUCTION

The word 'statistics ' means the 'science which deals with collection, analysis and interpretation of the data'. In today's world, the data analysis has got top priority in every field. There are number of methods of analysing the data depending upon what type of analysis is required. Some of these are 'operations research (OR)', 'statistical quality control (SQC)', 'time series analysis', 'method of correlation and regression' and so on. One such method is fitting an appropriate model to the data and then using fitted model for predicting future values. This method of analysis is useful in the situations where a variable under study (Y, say) is thought to be related to one or more measurements X_j $(j=1,2,\ldots,k)$ made usually made on the same objects. The purpose of model fitting is to use data to estimate the form of this relationship approximately. Generally, the dependent variable Y is called as 'response' variable and the 'independent' variables X_j $(j=1,2,\ldots,k)$ are known as 'stimulus' variables or 'covariates'. Here by independent variables X_j, we dose not mean that variables X_j $(j=1,2,\ldots,k)$ are not having any relationship between themselves, but just we mean that each of them may have seperate effect on the response variate Y. e.g. if there are two stimulus variates X_1 and X_2 in some model, they may be X and X^2 also. Now we define the general class of models.

Definition-1 : General class of models :

Suppose $Y_i (i=1,2,\ldots,n)$ are n independent observations on the response variate Y with mean of Y_i as μ_i and variance $Var(Y_i)$. Let $\underline{x}_j$ $(j=1,2,\ldots,k)$ be the vector of known values of the stimulus variates X_j $(j=1,2,\ldots,k)$, and β_j $(j=1,2,\ldots,k)$ are the coefficients of the functions of $\underline{x}_j$. Frequently, the quantities β_j $(j=1,\ldots,k)$ are called as the model parameters.

Then the general class of models is of the form

$$E(Y_i) = \tau(X, \underline{\beta}), \quad \text{for } i=1,2,\ldots,n, \qquad (1)$$

where

(1) $X = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k)'$ is called as incidence matrix

and

(2) τ is some functions of X and $\underline{\beta}$.

We can have particular subsets of this general class of models in two different ways.

1. Restricting the function τ to a specific type;
2. demanding the particular type of distribution for the responses.

Now to have better understanding of the above definition consider the 'classical linear model (CLM)', which is the simplest and the oldest model. It can be verified that under the assumptions,

(i) $\tau(X, \underline{\beta}) = X\underline{\beta}$,

and

(ii) the responses Y_i $(i=1,2,\ldots,n)$ are independently distributed $N(\mu_i, \sigma^2)$ variates,

equation (1) is equivalent to the classical linear model. We will justify this fact in chapter 2 immediately after the definition of classical linear model.

Though the theory of classical linear models is well known to the students of statistics, for the sake of completeness and to develope the further models and their need, the chapter 2 is devoted to the theory of classical linear models.

In chapter 2, classical linear models (CLMs) are studied explicitly. Various methods of fitting the model, like least

square method, method of maximum likelihood are studied in detail. Many other methods are also mentioned. The chapter also covers different hypothesis testing problems about the model parameters β_j (j=1,2,...,k). Residual analysis and its use to check the adequacy of the model is given in comprehensive form. Many times in the real life situations, the residual analysis clearly indicates that classical linear model is not suitable to analyse the data. In such situations the two ways of analysing the data, namely, 'response variable transformations' and fitting 'generalised linear model (GLM)' are useful. This chapter covers in brief the response variable transformations along with disadvantages of using it instead of fitting generalised linear model. Thus chapter 2 concludes with focusing on the importance of generalised linear model.

The chapter 3 includes the theory of generalised linear models (GLMs), which are introduced by Nelder & Wedderburn (1972) for the responses having the distribution from 'one parameter natural exponential family'. It is shown that this class of models is a particular subset of the general class of models. In section (3.4) procedure of fitting the model to the data is explained in detail. Sections (3.5) to (3.7) are useful to check the adequacy of the fitted model. Chapter also includes 'generalised linear models with varying dispersion' and the procedure for fitting these models. One recent alternative method of fitting generalised linear model is also discussed in brief. In socio economic fields, generally we come across the discrete responses. Hence the fourth chapter is devoted to the methods of analysing discrete data. This chapter covers methods of analysing various discrete responses like binomial, Poisson. Analysis of binomial responses is carried out in three different

Analysis of binomial responses is carried out in three different ways by taking three link functions, namely, 'logistic', 'log-log' and 'complementory log-log' functions. Chapter also includes analysis of the data in the form of table of counts. This analysis is carried out by fitting 'log linear model' in two different ways. One approach is as discussed by Bishop (1969) and another by using the theory of generalised linear model discussed by Nelder & Wedderburn (1972). It is natural to observe that both approaches leads to the same results. This fact is demonstrated with the help of numerical example.

Most of the times distributional form of the response variate is unknown. In such situations if there is a known relationship between mean and variance of the response variate, then the models, namely 'quasi likelihood models' introduced by Wedderburn (1974) are useful. The chapter 5 provides the theory related to these models and the 'extended quasi likelihood models' suggested by Nelder & Pregibon (1987) are explained. The model fitting procedures are also given. Further the cases of model fitting for 'over dispersed' discrete data are described in the chapter.

Last chapter gives guideline to the reader to analyse the data. For this purpose few artificial examples by obtaining data using simulation technique and some real life situations are taken. Then for every example, it is illustrated how one may proceed to analyse the data.

Appendix-1 contains some important proofs required at some stage of dissertation. Appendix-2 contains two PC-based software packages in FORTRAN-77 to analyse discrete data. We have developed these packages independently. One package is useful to analyse binomial data and another is useful to analyse data in

the form of table of counts.