# CHAPTER 0

# PREFACE

# CHAPTER - 0

# PREFACE

## 0.1 INTRODUCTION

In recent years, the use of statistically designed experiments in industries is greatly expanded, because they provide an effective method for improving quality and increasing productivity. A basic tool for the statistical analysis for such experiments is the classical linear model, whose analysis is based on certain assumptions like normally distributed errors, constant error variances, and the independence of observations. Moreover, from simplicity point of view and easy interpretation of results usually the models are preferred to be additive. The data in quality characteristics frequently are in the form of a set of counts, or proportions or lifetime data. For such type of data, some of the above mentioned assumptions are not satisfied. That is, the errors will not be normally distributed, the variances change with the values of mean, and the systematic effects are likely to combine additively on some scale different from the original one. For example, counts of defects are generally modeled by the Poisson distribution, where the variance is not constant but equal to the mean, and the systematic effects are belived to be multiplicative rather than additive, or they are additive on the logarithmic scale.

In this dissertation, our main aim is to handle situations where the response variates are nonnormally distributed. The traditional approach to the analysis of such responses is to transform the response into a new quantity that behaves more like a normal random variable. A modern approach is to use an analysis procedure based on generalized linear model (GLM). The Generalized linear models were first introduced by Nelder and Wedderburn(1972). The fitting of GLM is complicated, tedious and time consuming, hence use of GLM is restricted. But in recent years with the help of statistical packages like Matlab, SAS, etc. we can fit GLM very easily. In GLM approach, we can examine the significance of effects using either a $F$-test based on model deviances corresponding to the full and the restricted models or by using Wald test.

Designed experiments are used to build a model that describes the relationship of a response variable to one or more independent variables. Confidence intervals constructed around the estimate of the mean for each experimental run, and it's associated length or precision provides information about performance of model at points of interest in the experimental region. Therefore we have attempted in this dissertation, to compare the Transformation approach and the GLM approach with respect to confidence intervals around the predicted responses. Results of these comparisons are reported in Chapter 3 and 4. Chapter 3 presents a comparison based on some real data sets and Chapter 4 presents a comparison based on simulation studies.

The comparison based on the Type I error or confidence level actually attained by these tests and their powers is also important. We have carried out such comparisons and the results of these comparisons are presented in Chapter 4. Next we present chapterwise summary.

## 0.2 CHAPTERWISE SUMMARY

This dissertation includes four Chapters,excluding this preface Chapter, Chapters 1 and 2 cover a review of the theoretical framework for the transformation approach and the GLM approach, respectively. Results of these chapters are used in the development of the chapters 3 and 4.

### Chapter 1 :

The analysis of classical linear model is based on some assumptions namely constant variance, normality and uncorrelatedness of the observations. The optimality properties of ordinary least squares estimators and validity of the underlying $F$-tests and t-tests depend on these assumptions. Hence, violation of any assumptions prohibits the use of classical linear model for modelling the data. If the response variable does not satisfy any of these assumptions, then traditionally an appropriate transformation is employed. So that the transformed data satisfies the assumptions to the best possible extent. This transformed response is then analyzed using a linear model. We refer this technique as the "Transformation approach". Chapter 1 deals with the transformation approach. In section 1.1, we discuss the need of transformation.The variance stabilizing transformations are discussed in section 1.2, to handle the problem of nonconstant variance. Section 1.3 deals with the problem of nonnormality. In section 1.4, we discuss the detection of nonadditivity and transformations for removal of nonadditivity with illustration of one example. In section 1.5, we summarize some results of classical linear model which are used in analyzing transformed data, and point out some limitations of classical linear model.

**Chapter 2 :**

In Chapter 2, we discuss the modern GLM approach to analyze nonnormal data. The definition of GLM with it's components and main features are discussed in Section 2.2. Section 2.3 deals with fitting of GLM. Fitting of GLM is done with maximum likelihood estimation method. The 'Iterative reweighted least square method' for obtaining maximum likelihood estimate is described and properties of estimators are discussed in this Section. In Section 2.4, we describe the procedure of testing significance of regression parameters. A method for obtaining confidence intervals on the mean responses in GLM are described in Section 2.5.

**Chapter 3 :**

Chapter 3 deals with the analysis of real data sets, involving nonnormal response variables, based on the theory discussed in Chapters 1 and 2. A general procedure for analyzing these data sets and comparing the performance of the transformation approach and the GLM approach for these data sets is outlined in Section 3.2. In Section 3.3, we analyze four data sets involving Poisson, Gamma, and Binomial distributed response variables, and compare the two approaches based on confidence intervals,around the estimated responses.

**Chapter 4 :**

Chapter 4 deals with the comparison between transformation and GLM approach based on a simulation study at the parameter values belonging to a sub set of the parameter space. The selection of this subset was made based on a pilot study. Section 4.1 illustrates with examples the need for such study. Section 4.2 outline details of the simulation procedure and a general method that is adopted for carrying out these comparisons.

The comparison is carried out based on the following three aspects.

(i) Length of the confidence intervals around the estimated responses. Results of this comparison are reported in Section 4.3.

(ii) Type-I error actually attained by the tests.

(iii) Powers of the tests.

For comparisons with respect to the last two aspects we consider three tests namely, the transformation approach- $F$-tests, GLM approach- Wald test and GLM approach- $F$-test( )The results of these comparisons are reported in Section 4.4 overall conclusions are reported in Section 4.5. The entire material presented in this Chapter is our original work.