CHAPTER 3

GENERALIZED LINEAR MODELS

3.1 INTRODUCTION

Generalized linear models extend normal theory linear models to response variables whose distributions belong to the exponential family or can be characterized by a quasi-likelihood (Section 2.5). This class of models includes fixed effects analysis of variance as well as regression and analysis of covariance models that do not contain random effects. A generalized linear model consists of three components:

- a stochastic component that defines the probability distribution or quasi-likelihood of the response variable,
- a linear predictor that is a systematic component describing the linear model defined by the explanatory variables,
- a link function that relates the mean of the response variable to a linear combination of the explanatory variables.

More specifically, let $Y$ be the response variable with a probability distribution from the exponential family with mean $E[Y] = \mu$ and let $x_1, ..., x_p$ be a set of $p$ explanatory variables. For the $j$th observation from a random sample of size $n$, the systematic component can be expressed as

$$
\eta_j = g(\mu_j) = \beta_0 + \sum_{i=1}^{p} \beta_i x_{ij}, \ j = 1, ..., n
$$

where $g(\cdot)$ is the link function, $\mu_j$ is the mean of the $j$th observation, and $x_{ij}$ is the observed value of the $i$th explanatory variable for the $j$th observation.

Writing the above system of equations in matrix form,

$$
\eta = g(\mu) = X\beta
$$

where $g(\mu)' = [g(\mu_1), ..., g(\mu_n)]'$ is the $n \times 1$ vector of link functions, $X$ is the $n \times (p + 1)$ design matrix, and $\beta$ is the $(p + 1) \times 1$ vector of unknown parameters. It is important to emphasize that a generalized linear model involves relating the transformed mean of the response variable $Y$ to the explanatory variables but does not involve...