CHAPTER 1

Generalized Linear Models

In statistics, generalized linear models (GLMs) are a flexible generalization of ordinary linear regression models for response variables that are not normally distributed. If you're already familiar with multiple linear regression, you're well on your way to understanding GLMs.

1.1 Components of a Generalized Linear Model

A GLM consists of three components:

1. A random component: This is a specification of the probability distribution of the response variable (e.g., normal, binomial, Poisson distributions, etc.). This differs from ordinary linear regression, which assumes that the response variable follows a normal distribution.
2. A systematic component: This is the linear predictor, a linear combination of the explanatory variables, just as in ordinary linear regression.
3. A link function: This is a function that connects the mean of the response variable

to the linear predictor. The choice of link function depends on the nature of the response variable and the range of its possible values.

1.2 Formulation of a Generalized Linear Model

The GLM can be formulated as follows:

$$g(E(Y)) = \eta = X\beta \tag{1.1}$$

Here, Y is the response variable, X represents the matrix of explanatory variables, β is the vector of parameters to be estimated, η is the linear predictor, $E(Y)$ represents the expected value of Y, and $g(.)$ is the link function.

1.3 Fitting a Generalized Linear Model

The parameters β in a GLM are typically estimated using maximum likelihood estimation (MLE). The specifics of this process depend on the probability distribution of the response variable and the link function.

1.4 Examples of Generalized Linear Models

Examples of GLMs include:

- Logistic regression: This is a GLM with a binomial response variable and a logit link function.
- Poisson regression: This is a GLM with a Poisson response variable and a log link function.

Appendix A

Answers to Exercises

Index