general linear model

Canonical name	GeneralLinearModel
Date of creation	2013-03-22 14:31:23
Last modified on	2013-03-22 14:31:23
Owner	CWoo (3771)
Last modified by	CWoo (3771)
Numerical id	6
Author	CWoo (3771)
Entry type	Definition
Classification	msc 62J10
Classification	msc 62J05
Synonym	normal linear model
Defines	analysis of variance
Defines	ANOVA
Defines	analysis of covariance
Defines	ANCOVA

In statistical modeling of N data observations ($N < \infty$), two types of variables are usually defined. One is the response variable or variate, usually denoted by Y, and the other is the explanatory variable or covariate X. While there is only one response variable, there may be one or more than one explanatory variables. The response variable is considered random, where as the explanatory variable(s) may or may not be random.

Based on the above setup, a *general linear model*, or *normal linear model*, is a statistical model with the following assumptions:

1. the response variable Y is a continuous random variable

2. the response variable Y can be expressed as a linear combination of functions $z_i(\mathbf{X})$, of the explanatory variables, plus a random error term ε:
$$Y = \beta_0 z_0(\mathbf{X}) + \cdots + \beta_k z_k(\mathbf{X}) + \varepsilon.$$
The portion of Y without the error term is known as the *systematic component* of Y.

3. the error component and the systematic component are independent

4. random error variables ε_i for the N observations are iid normal with mean 0 and variance σ^2

Remarks

- Conditioning on the explanatory variables, the random variables Y_i corresponding to the individual responses are independent, normally distributed, with mean
$$\mu = \mathrm{E}[Y \mid \mathbf{X} = \boldsymbol{x}] = \beta_0 z_0(\mathbf{X}) + \cdots + \beta_k z_k(\mathbf{X})$$
and variance σ^2.

- A linear regression model is a special case of the general linear model where all explanatory variables are assumed to be continuous.

- *Analysis of variance* model, or *ANOVA*, is another special case of the general linear model, where all of the explantory variables are categorical in nature (for example, gender, marital status, etc..).

- *Analysis of covariance*, or *ANCOVA*, sits between a linear regression model and the ANOVA, where some of the explanatory variables are continuous and some are categorical.

- The general linear model is a special case of the *generalized linear model*, where the assumption that the response variable Y has a normal distribution is dropped.