In Neil Salkind (Ed.), ***Encyclopedia of Research Design.***
Thousand Oaks, CA: Sage. 2010

The General Linear Model

Hervé Abdi

1 Overview

The general linear model (GLM) provides a general framework for a large set of models whose common goal is to explain or predict a quantitative dependent variable by a set of independent variables which can be categorical of quantitative. The GLM encompasses techniques such as Student's t test, simple and multiple linear regression, analysis of variance, and covariance analysis. The GLM is adequate only for *fixed* effect models. In order to take into account *random* effect model, the GLM needs to be extended and becomes the *mixed* effect model.

2 Notations

Vectors are denoted with bold lower case letters (*e.g.,* $\mathbf{y}$), matrices are denoted with bold upper case letters (*e.g.,* $\mathbf{X}$). The transpose of a matrix is denoted by the superscript $^{\mathsf{T}}$, the inverse of a matrix is de-

Hervé Abdi
The University of Texas at Dallas
Address correspondence to:
Hervé Abdi
Program in Cognition and Neurosciences, MS: Gr.4.1,
The University of Texas at Dallas,
Richardson, TX 75083–0688, USA
E-mail: herve@utdallas.edu http://www.utd.edu/∼herve

noted by the superscript $^{-1}$. There are I observations. The values of a quantitative dependent variable describing the I observations are stored in an I by 1 vector denoted $\mathbf{y}$. The values of the independent variables describing the I observations are stored in an I by K matrix denoted $\mathbf{X}$, K is smaller than I and $\mathbf{X}$ is assumed to have rank K (*i.e.,* $\mathbf{X}$ is full rank on its columns). A quantitative independent variable can be directly stored in $\mathbf{X}$, but a qualitative independent variable needs to be recoded with as many columns as they are degrees of freedom for this variable. Common coding schemes include dummy coding, effect coding, and contrast coding.

2.1 The general linear model core equation

For the GLM, the values of the dependent variable are obtained as a *linear* combination of the values of the independent variables. The vector for the coefficients of the linear combination are stored in a K by 1 vector denoted $\mathbf{b}$. In general, the values of $\mathbf{y}$ cannot be perfectly obtained by a linear combination of the columns of $\mathbf{X}$ and the difference between the actual and the predicted values is called the *prediction error*. The values of the error are stored in an I by 1 vector denoted $\mathbf{e}$. Formally the GLM is stated as:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}\ . \tag{1}$$

The predicted values are stored in an I by 1 vector denoted $\widehat{\mathbf{y}}$ and, therefore, Equation 1 can be rewritten as

$$\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{e} \quad \text{with} \quad \widehat{\mathbf{y}} = \mathbf{X}\mathbf{b}\ . \tag{2}$$

Putting together Equations 1 and 2 shows that

$$\mathbf{e} = \mathbf{y} - \widehat{\mathbf{y}}\ . \tag{3}$$

2.2 Additional assumptions of the general linear model

The independent variables are assumed to be fixed variables (*i.e.,* their values will not change for a replication of the experiment analyzed by the GLM, and they are measured without error). The error is interpreted as a *random* variable and in addition the I components of the error are assumed to be independently and identically distributed ("i.i.d.") and their distribution is assumed to be a normal distribution with a zero mean and a variance denoted $\sigma^2_{\mathbf{e}}$. The values of the dependent variable are assumed to be a random sample of a population of interest. Within this framework, the vector $\mathbf{b}$ is seen as an estimation of the population parameter vector $\boldsymbol{\beta}$.

3 Least square estimate for the general linear model

Under the assumptions of the GLM, the population parameter vector $\boldsymbol{\beta}$ is estimated by $\mathbf{b}$ which is computed as

$$\mathbf{b} = \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{y}. \tag{4}$$

This value of $\mathbf{b}$ minimizes the residual sum of squares (*i.e.,* $\mathbf{b}$ is such that $\mathbf{e}^{\mathsf{T}}\mathbf{e}$ is minimum).

3.1 Sums of squares

The total sum of squares of $\mathbf{y}$ is denoted SS_{Total}, it is computed as

$$SS_{\text{Total}} = \mathbf{y}^{\mathsf{T}}\mathbf{y} \; . \tag{5}$$

Using Equation 2, the total sum of squares can be rewritten as

$$SS_{\text{Total}} = \mathbf{y}^{\mathsf{T}}\mathbf{y} = (\widehat{\mathbf{y}} + \mathbf{e})^{\mathsf{T}}(\widehat{\mathbf{y}} + \mathbf{e}) = \widehat{\mathbf{y}}^{\mathsf{T}}\widehat{\mathbf{y}} + \mathbf{e}^{\mathsf{T}}\mathbf{e} + 2\widehat{\mathbf{y}}^{\mathsf{T}}\mathbf{e}, \tag{6}$$

but it can be shown that $2\widehat{\mathbf{y}}^{\mathsf{T}}\mathbf{e} = 0$, and therefore Equation 6 becomes

$$SS_{\text{Total}} = \mathbf{y}^{\mathsf{T}}\mathbf{y} = \widehat{\mathbf{y}}^{\mathsf{T}}\widehat{\mathbf{y}} + \mathbf{e}^{\mathsf{T}}\mathbf{e} \; . \tag{7}$$

The first term of Equation 7 is called the *model* sum of squares, it is denoted SS_{Model} and it is equal to

$$SS_{\text{Model}} = \widehat{\mathbf{y}}^{\mathsf{T}}\widehat{\mathbf{y}} = \mathbf{b}^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}\mathbf{X}\mathbf{b}. \tag{8}$$

The second term of Equation 7 is called the *residual* or the *error* sum of squares, it is denoted SS_{Residual} and it is equal to

$$SS_{\text{Residual}} = \mathbf{e}^{\mathsf{T}}\mathbf{e} = (\mathbf{y} - \mathbf{X}\mathbf{b})^{\mathsf{T}}(\mathbf{y} - \mathbf{X}\mathbf{b}). \tag{9}$$

3.2 Sampling distributions of the sums of squares

Under the assumptions of normality and i.i.d for the error, we find that the ratio of the residual sum of squares to the error variance

$\frac{SS_{\text{Residual}}}{\sigma^2_{\mathbf{e}}}$ is distributed as a χ^2 with a number of degrees of freedom of $\nu = I - K - 1$. This is abbreviated as

$$\frac{SS_{\text{Residual}}}{\sigma^2_{\mathbf{e}}} \;\sim\; \chi^2(\nu) \;. \tag{10}$$

By contrast, the ratio of the model sum of squares to the error variance $\frac{SS_{\text{Model}}}{\sigma^2_{\mathbf{e}}}$ is distributed as a *non-central* χ^2 with $\nu = K$ degrees of freedom and non centrality parameter

$$\lambda = \frac{2}{\sigma^2_{\mathbf{e}}} \boldsymbol{\beta}^\mathsf{T} \mathbf{X}^\mathsf{T} \mathbf{X} \boldsymbol{\beta}.$$

This is abbreviated as

$$\frac{SS_{\text{Model}}}{\sigma^2_{\mathbf{e}}} \;\sim\; \chi^2(\nu,\ \lambda) \;. \tag{11}$$

From Equations 10 and 11, it follows that the ratio

$$F = \frac{SS_{\text{Model}}/\sigma^2_{\mathbf{e}}}{SS_{\text{Residual}}/\sigma^2_{\mathbf{e}}} \times \frac{I-K-1}{K} = \frac{SS_{\text{Model}}}{SS_{\text{Residual}}} \times \frac{I-K-1}{K} \tag{12}$$

is distributed as a non-central Fisher's F with $\nu_1 = K$ and $\nu_2 = I - K - 1$ degrees of freedom and non-centrality parameter equal to

$$\lambda = \frac{2}{\sigma^2_{\mathbf{e}}} \boldsymbol{\beta}^\mathsf{T} \mathbf{X}^\mathsf{T} \mathbf{X} \boldsymbol{\beta}.$$

In the specific case when the null hypothesis of interest states that $H_0 :\ \boldsymbol{\beta} = \mathbf{0}$, the non-centrality parameter vanishes and then the F ratio from Equation 12 follows a standard (*i.e.*, central) Fisher's distribution with $\nu_1 = K$ and $\nu_2 = I - K - 1$ degrees of freedom.

4 Test on subsets of the parameters

Often we are interested in testing only a subset of the parameters. When this is the case, the I by K matrix $\mathbf{X}$ can be interpreted as composed of two blocks: an I by K_1 matrix $\mathbf{X}_1$ and an I by K_2 matrix $\mathbf{X}_2$ with $K = K_1 + K_2$. This is expressed as

$$\mathbf{X} = \left[\mathbf{X}_1 \vdots \mathbf{X}_2\right] \;. \tag{13}$$

Vector $\mathbf{b}$ is partitioned in a similar manner as

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \cdots \\ \mathbf{b}_2 \end{bmatrix} . \tag{14}$$

In this case the model corresponding to Equation 1 is expressed as

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e} = \begin{bmatrix} \mathbf{X}_1 \vdots \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \cdots \\ \mathbf{b}_2 \end{bmatrix} + \mathbf{e} = \mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2\mathbf{b}_2 + \mathbf{e} . \tag{15}$$

For convenience, we will assume that the test of interest concerns the parameters $\boldsymbol{\beta}_2$ estimated by vector $\mathbf{b}_2$ and that the null hypothesis to be tested corresponds to a *semi partial* hypothesis namely that adding $\mathbf{X}_2$ *after* $\mathbf{X}_1$ does not improve the prediction of $\mathbf{y}$. The first step is to evaluate the quality of the prediction obtained when using $\mathbf{X}_1$ alone. The estimated value of the parameters is denoted $\widetilde{\mathbf{b}}_1$—a new notation is needed because in general $\mathbf{b}_1$ is different from $\widetilde{\mathbf{b}}_1$ ($\mathbf{b}_1$ and $\widetilde{\mathbf{b}}_1$ are equal only if $\mathbf{X}_1$ and $\mathbf{X}_2$ are two orthogonal blocks of columns). The model relating $\mathbf{y}$ to $\mathbf{X}_1$ is called a *reduced* model. Formally, this reduced model is obtained as:

$$\mathbf{y} = \mathbf{X}_1\widetilde{\mathbf{b}}_1 + \widetilde{\mathbf{e}}_1 \tag{16}$$

(where $\widetilde{\mathbf{e}}_1$ is the error of prediction for the reduced model). The model sum of squares for the reduced model is denoted $SS_{\widetilde{\mathbf{b}}_1}$ (see Equation 9 for its computation). The semi partial sum of squares for $\mathbf{X}_2$ is the sum of squares *over and above* the sum of squares already explained by $\mathbf{X}_1$. It is denoted $SS_{\mathbf{b}_2|\mathbf{b}_1}$ and it is computed as

$$SS_{\mathbf{b}_2|\mathbf{b}_1} = SS_{\text{Model}} - SS_{\widetilde{\mathbf{b}}_1} . \tag{17}$$

The null hypothesis test indicating that $\mathbf{X}_2$ does not improve the prediction of $\mathbf{y}$ over and above $\mathbf{X}_1$ is equivalent to testing the null hypothesis that $\boldsymbol{\beta}_2$ is equal to $\mathbf{0}$. It can be tested by computing the following F ratio:

$$F_{\mathbf{b}_2|\mathbf{b}_1} = \frac{SS_{\mathbf{b}_2|\mathbf{b}_1}}{SS_{\text{Residual}}} \times \frac{I - K - 1}{K_2} . \tag{18}$$

When the null hypothesis is true, $F_{\mathbf{b}_2|\mathbf{b}_1}$ follows a Fisher's F distribution with $\nu_1 = K_2$ and $\nu_2 = I - K - 1$ degrees of freedom and therefore $F_{\mathbf{b}_2|\mathbf{b}_1}$ can be used to test the null hypothesis that $\boldsymbol{\beta}_2 = \mathbf{0}$.

5 Specific cases of the general linear model

The GLM comprises several standard statistical techniques. Specifically, linear regression is obtained by augmenting the matrix of independent variables by a column of ones (this additional column codes for the intercept). Analysis of variance is obtained by coding the experimental effect in an appropriate way. Various schemes can be used such as effect coding, dummy coding, or contrast coding (with as many columns as there are degrees of freedom for the source of variation considered). Analysis of covariance is obtained by combining the quantitative independent variables expressed as such and the categorical variables expressed in the same way as for an analysis of variance.

6 Limitations and extensions of the general linear model

The general model, despite its name, is not completely general and has several limits which have spurred the development of "generalizations" of the general linear model. Some of the most notable limits and some palliatives are listed below.

The general linear model requires $\mathbf{X}$ to be full rank, but this condition can be relaxed by using, (*cf.* Equation 4) the Moore-Penrose generalized inverse (often denoted $\mathbf{X}^{+}$ and sometime called a "pseudo-inverse") in lieu of $\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}$. Doing so, however makes the problem of estimating the model parameters more delicate and requires the use of the notion of *estimable* functions.

The general linear model is a *fixed* effect model and therefore, it does not naturally works with random effect models (including multifactorial repeated or partially repeated measurement designs). In this case (at least for balanced designs), the sums of squares are correctly computed but the F tests are likely to be incorrect. A palliative to this problem is to compute expected values for the different sums of squares and to compute F-tests accordingly. Another, more general, approach is to model separately the fixed effects and the random effects. This is done with *mixed* effect models.

Another obvious limit of the general *linear* model is to model only linear relationship. In order to include some non linear models (such as, *e.g.,* logistic regression) the GLM needs to be expended to the class of the *generalized* linear models.

Related entries

Analysis of variance, analysis of covariance, contrast, degrees of freedom, dummy coding, effect coding, experimental design, fixed effect models, Gauss-Markov theorem, Homoscedasticity, least squares (method of), linear regression, matrix algebra, mixed model design, multiple linear regression, normality assumption, random error, sampling distributions, Student's t statistics, t test.

Further readings

1. Abdi, H., Edelman, B., Valentin, D., & Dowling, W.J. (2009). *Experimental design and analysis for psychology.* Oxford: Oxford University Press.
2. Brown, H., & Prescott, R. (2006). *Applied mixed models in medicine.* London: Wiley.
3. Fox J. (2008). *Applied regression analysis and generalized linear models.* Thousand Oaks: Sage.
4. Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, **70**, 426–443.
5. Hocking, R.R. (2003). *Methods and applications of linear models.* New York: Wiley.
6. Graybill, F.A. (1976). *Theory and application of the linear model.* North Scituate, MA: Duxbury.
7. Rencher, A.C., Schaafe, G.B. (2008). *Linear models in statistics.* New York: Wiley.
8. Searle, S.R. (1971). *Linear models.* New York: Wiley.