

Introduction to General Linear Models

Moo K. Chung
University of Wisconsin-Madison
mkchung@wisc.edu

September 27, 2014

In this chapter, we introduce *general linear models* (GLM) that have been widely used in brain imaging applications. The GLM is a very flexible and general statistical framework encompassing a wide variety of fixed effect models such as the multiple regressions, the analysis of variance (ANOVA), the multivariate analysis of variance (MANOVA), the analysis of covariance (ANCOVA) and the multivariate analysis of covariance (MANCOVA) [5]. Note that the term *linear* is misleading in a sense that the model can also include mathematically nonlinear model terms such as the higher degree polynomials. This chapter is mainly based on [2].

The GLM provides a framework for testing various associations and hypotheses while accounting for nuisance covariates in the model in a straightforward fashion. The effect of age, sex, brain size and possibly IQ can have severe confounding effects on the final outcome of many anatomical and functional imaging studies. Older population's reduced functional activation could be the consequence of age-related atrophy of neural systems [3]. Brain volumes are significantly larger for children with autism 12 years old and younger compared with normally developing children [1]. Therefore, it is desirable to account for various confounding factors such as age and sex in the model. This can be done using GLM. The parameters of the GLM are mainly estimated by the least squares estimation and has been implemented in many statistical packages such as R (www.r-project.org) or Splus [4] and brain imaging packages such as SPM (www.fil.ion.ucl.ac.uk/spm) and fMRISTAT (www.math.mcgill.ca/keith/fmristat).

1 General Linear Models

Let y_i be the response variable, which is mainly coming from images and $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ to be the variables of interest and $\mathbf{z}_i = (z_{i1}, \dots, z_{ik})$ to be nuisance variables corresponding to the i -th subject. We assume there are n subjects, i.e. $i = 1, \dots, n$. We are interested in testing the significance of variables \mathbf{x}_i while accounting for nuisance covariates \mathbf{z}_i . Then we set up the following GLM

$$y_i = \mathbf{z}_i \boldsymbol{\lambda} + \mathbf{x}_i \boldsymbol{\beta} + \epsilon_i \quad (1)$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)'$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ are unknown parameter vectors to be estimated. We assume ϵ to be the usual zero mean Gaussian noise although the distributional assumption is not necessary for the least squares estimation.

The significance of the variable of interests \mathbf{x}_i is determined by testing the null hypothesis

$$H_0 : \boldsymbol{\beta} = 0 \text{ vs. } H_1 : \boldsymbol{\beta} \neq 0.$$

The fit of the reduced model corresponding to $\boldsymbol{\beta} = 0$, i.e.

$$y_i = \mathbf{z}_i \boldsymbol{\lambda}, \quad (2)$$

is measured by the sum of the squared errors (SSE):

$$\text{SSE}_0 = \sum_{i=1}^n (y_i - \mathbf{z}_i \widehat{\boldsymbol{\lambda}}_0)^2,$$

where $\widehat{\boldsymbol{\lambda}}_0$ is the least squares estimation obtained from the reduced model. The reduced model (2) can be written in a matrix form

$$\underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} z_{11} & \cdots & z_{1k} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nk} \end{pmatrix}}_{\mathbf{Z}} \underbrace{\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}}_{\boldsymbol{\lambda}}. \quad (3)$$

By multiplying \mathbf{Z}' on the both sides, we obtain

$$\mathbf{Z}'\mathbf{y} = \mathbf{Z}'\mathbf{Z}\boldsymbol{\lambda}.$$

Now the matrix $\mathbf{Z}'\mathbf{Z}$ is a full rank and can be invertible if $n \geq k$, which is the usual case in brain imaging. Therefore, the matrix equation can be solved by performing a matrix inversion

$$\widehat{\boldsymbol{\lambda}}_0 = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}.$$

Similarly the fit of the full model corresponding to $\boldsymbol{\beta} \neq 0$, i.e.

$$y_i = \mathbf{z}_i \boldsymbol{\lambda} + \mathbf{x}_i \boldsymbol{\beta} \quad (4)$$

is measured by

$$\text{SSE}_1 = \sum_{i=1}^n (y_i - \mathbf{z}_i \widehat{\boldsymbol{\lambda}}_1 - \mathbf{x}_i \widehat{\boldsymbol{\beta}}_1)^2,$$

where $\widehat{\boldsymbol{\lambda}}_1$ and $\widehat{\boldsymbol{\beta}}_1$ are the least squares estimation from the full model. The full model can be written in a matrix form by concatenating the row vectors \mathbf{z}_i and \mathbf{x}_i into a larger row vector $(\mathbf{z}_i, \mathbf{x}_i)$, and the column vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ into a larger column vector, i.e.

$$y_i = (\mathbf{z}_i, \mathbf{x}_i) \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\beta} \end{bmatrix}.$$

The full model can be also written in a matrix form

$$\mathbf{y} = [\mathbf{Z} \ \mathbf{X}] \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\beta} \end{bmatrix}.$$

Then the parameters of the full model are estimated similarly in the least squares fashion. Note that

$$\begin{aligned} \text{SSE}_1 &= \min_{\boldsymbol{\lambda}_1, \boldsymbol{\beta}_1} \sum_{i=1}^n (y_i - \mathbf{z}_i \boldsymbol{\lambda}_1 - \mathbf{x}_i \boldsymbol{\beta}_1)^2 \\ &\leq \min_{\boldsymbol{\lambda}_0} \sum_{i=1}^n (y_i - \mathbf{z}_i \boldsymbol{\lambda}_0)^2 = \text{SSE}_0. \end{aligned}$$

So the larger the value of $\text{SSE}_0 - \text{SSE}_1$, the more significant the contribution of the coefficients $\boldsymbol{\beta}$ is. Under the assumption of the null hypothesis H_0 , the test statistic is the ratio

$$F = \frac{(\text{SSE}_0 - \text{SSE}_1)/p}{\text{SSE}_1/(n - p - k)} \sim F_{p, n-p-k}. \quad (5)$$

The larger the F value, it is more unlikely to accept H_0 .

2 T-statistic

Consider a special case of GLM (1) with $p = 1$, i.e. When $p = 1$, the test statistic F is distributed as $F_{1, n-1-k}$, which is the square of the student t -distribution with $n - 1 - k$ degrees of freedom, i.e. t_{n-1-k}^2 . Instead of using F -statistic, we can use t -statistic as follows. Let $c = (\underbrace{0, \dots, 0}_k, \underbrace{1, 0, \dots, 0}_{p-1})'$ be

the contrast vector of size $k + p$. Although we will not show in detail, the incorporation of the contrast vector makes algebraic derivation straightforward. Then the least squares estimation of β_1 in the full model is given by

$$\widehat{\beta}_1 = c' \begin{bmatrix} \widehat{\boldsymbol{\lambda}}_1 \\ \widehat{\boldsymbol{\beta}}_1 \end{bmatrix}$$

Under the distributional assumption of $\epsilon_i \sim N(0, \sigma^2)$, we can show that $\widehat{\beta}_1$ is unbiased, i.e. $\mathbb{E}\widehat{\beta}_1 = \beta_1$. Then we can show the variance of $\widehat{\beta}_1$ is given by

$$\mathbb{V}\widehat{\beta}_1 = c' \mathbb{V} \begin{bmatrix} \widehat{\boldsymbol{\lambda}}_1 \\ \widehat{\boldsymbol{\beta}}_1 \end{bmatrix} c = \sigma^2 c' ([\mathbf{Z} \ \mathbf{X}]' [\mathbf{Z} \ \mathbf{X}])^{-1} c.$$

The unbiased estimator of σ^2 is given by $\text{SSE}_1/(n - 1 - k)$. We will plug this estimator into σ^2 . Then the test statistic testing for $\beta_1 = 0$ is given by t -statistic

$$T = \frac{\widehat{\beta}_1}{\sqrt{\mathbb{V}\widehat{\beta}_1}} \sim t_{n-1-k}.$$

In the special case when $p = 1$, it is better to use the t -statistic. The advantage of using the t -statistic is that unlike the F -statistic, it has two sides so we can actually use it to test for one sided alternative hypothesis $H_1 : \beta_1 \geq 0$ or $H_1 : \beta_1 \leq 0$. Therefore, the t -statistic map can provide the direction of the group difference that the F -statistic map cannot provide.

3 R-square

The R-square of a model explains the proportion of variability in measurement that is accounted for by the model. Sometimes R-square is called the coefficient of determination and it is given as the square of a correlation coefficient for a very simple model. For a linear model involving the response variable y_i , the total sum of squares (SST) measures total variation in response y_i and is defined as

$$\text{SST} = \sum_{i=1}^n (y_i - \bar{y})^2,$$

where \bar{y} is the sample mean of y_i .

On the other hand, SSE measures the amount of variability in y_i that is not explained by the model. Note that SSE is the minimum of the sum of squared residual of any linear model, SSE is always smaller than SST. Therefore, the amount of variability explained by the model is SST-SSE. The proportion of variability explained by the model is then

$$R^2 = \frac{\text{SST} - \text{SSE}}{\text{SST}},$$

which is the coefficient of determination. The R-square ranges between 0 and 1 and the value larger than 0.5 is usually considered as significant.

References

- [1] E.H. Aylward, N.J. Minshew, G. Goldstein, N.A. Honeycutt, A.M. Augustine, K.O. Yates, P.E. Bartra, and G.D. Pearlson. Mri volumes of amygdala and hippocampus in nonmentally retarded autistic adolescents and adults. *Neurology*, 53:2145–2150, 1999.
- [2] M.K. Chung, K.M. Dalton, A.L. Alexander, and R.J. Davidson. Less white matter concentration in autism: 2D voxel-based morphometry. *NeuroImage*, 23:242–251, 2004.
- [3] M. Mather, T. Canli, T. English, S. Whitfield, P. Wais, K. Ochsner, J.D.E. Gabrieli, and L.L. Carstensen. Amygdala responses to emotionally valenced stimuli in older and younger adults. *Psychological Science*, 15:259–263, 2004.

- [4] J.C. Pinehiro and D.M. Bates. *Mixed Effects Models in S and S-Plus*. Springer, 3rd edition, 2002.
- [5] N.H. Timm and T.A. Mieczkowski. *Univariate and Multivariate General Linear Models: Theory and Applications using SAS Software*. SAS Publishing, 1997.