The General Linear Statistical Model

Study Notes December 26, 2013 Kenneth Chu

Let $Y : \Omega \longrightarrow \mathbb{R}^n$ be an $\mathbb{R}^n$-valued random variable defined on the probability space Ω. We assume that the expected value $E[\,Y\,]$ of Y exists. Then, trivially, we have $E[\,Y\,] \in \mathbb{R}^n$.

1 Assumption on the expected value of the response variable Y

The most fundamental assumption of the General Linear Model is that the expected value of the response variable Y lies in a model-specific subspace of $\mathbb{R}^n$ (this subspace will be called the *estimation space* of the model), in the following sense: One of the "components" of a general linear model is its *model matrix* $X \in \mathbb{R}^{n\times p}$, and the expected value of the response variable Y is assumed to lie in the column space $\mathcal{C}(X) \subset \mathbb{R}^n$.

In other words:

The Estimation Space Assumption

$$E[\,Y\,] \in \mathcal{C}(X); \;\; \text{equivalently,} \;\; E[\,Y\,] = X\beta, \; \text{for some (unknown) } \beta \in \mathbb{R}^p, \tag{1.1}$$

where $\mathcal{C}(X) \subset \mathbb{R}^n$ is the column space of the model matrix $X \in \mathbb{R}^{n\times p}$.

We will call $\mathbb{R}^n$ the *observation space*, and $\mathcal{C}(X)$ the *estimation space* of the model.

2 Assumption of the distribution of the response variable Y

In order to make estimation and hypothesis testing computationally feasible, we need to make certain assumptions on the distribution of the response variable Y.

Assumptions on the distribution of Y:

1. The response variable Y has a multivariate normal distribution.
2. The components of Y are independent $\mathbb{R}$-valued random variables.
3. The variances of the components of Y are all equal.

The assumptions on the expected value and distribution on Y together are equivalent to the following:

$$Y \sim N\big(X\beta, \sigma^2 I_n\big)\,, \text{ for some (unknown but fixed) } \beta \in \mathbb{R}^p, \text{ and some (unknown but fixed) } \sigma > 0. \tag{2.1}$$

Define $\varepsilon := Y - X\beta$. Then, $\varepsilon : \Omega \longrightarrow \mathbb{R}^n$ is also an $\mathbb{R}^n$-valued random variable, with

$$\varepsilon \sim N\big(0, \sigma^2 I_n\big)\,, \;\; \text{for some } \; \sigma > 0. \tag{2.2}$$

Proposition 2.1 (Distribution of the full-model error sum-of-squares)
Let $P_{\mathcal{C}(X)^\perp} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ denote the orthogonal projection operator onto the subspace $\mathcal{C}(X)^\perp$. Then,

$$\frac{\|\,P_{\mathcal{C}(X)^\perp}(Y)\,\|^2}{\sigma^2} \;\sim\; \chi^2\big(\mathrm{rank}\big(\mathcal{C}(X)^\perp\big)\big)$$

3 Testing the hypothesis that $H_0 : E[\,Y\,] \in \mathcal{C}(X_0) \subset \mathcal{C}(X)$

Proposition 3.1
Let $P_{\mathcal{C}(X_0)^\perp \cap \mathcal{C}(X)} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ denote the orthogonal projection operator onto the subspace $\mathcal{C}(X_0)^\perp \cap \mathcal{C}(X)$. Then,

$$\frac{\|\,P_{\mathcal{C}(X_0)^\perp \cap \mathcal{C}(X)}(Y)\,\|^2}{\sigma^2} \;\sim\; \chi^2\left(\mathrm{rank}\big(\mathcal{C}(X_0)^\perp \cap \mathcal{C}(X)\big)\;,\; \frac{\|\,P_{\mathcal{C}(X_0)^\perp \cap \mathcal{C}(X)} X\beta\,\|^2}{2\,\sigma^2}\right)$$

Corollary 3.2 (Distribution of F-statistics under validity of full model)

$$\frac{\|\, P_{\mathcal{C}(X_0)^\perp \cap \mathcal{C}(X)}(Y)\,\|^2 \,/\, \operatorname{rank}\big(\mathcal{C}(X_0)^\perp \,\cap\, \mathcal{C}(X)\big)}{\|\, P_{\mathcal{C}(X)^\perp}(Y)\,\|^2 \,/\, \operatorname{rank}(\mathcal{C}(X)^\perp)} \;\sim\; F\left(\operatorname{rank}\big(\mathcal{C}(X_0)^\perp \,\cap\, \mathcal{C}(X)\big)\;,\;\operatorname{rank}\big(\mathcal{C}(X)^\perp\big)\;;\;\frac{\|\, P_{\mathcal{C}(X_0)^\perp \cap \mathcal{C}(X)} X\beta\,\|^2}{2\,\sigma^2}\right)$$

Corollary 3.3 (Distribution of F-statistics under validity of reduced model)

$$\frac{\|\, P_{\mathcal{C}(X_0)^\perp \cap \mathcal{C}(X)}(Y)\,\|^2 \,/\, \operatorname{rank}\big(\mathcal{C}(X_0)^\perp \,\cap\, \mathcal{C}(X)\big)}{\|\, P_{\mathcal{C}(X)^\perp}(Y)\,\|^2 \,/\, \operatorname{rank}(\mathcal{C}(X)^\perp)} \;\sim\; F\big(\operatorname{rank}\big(\mathcal{C}(X_0)^\perp \,\cap\, \mathcal{C}(X)\big)\;,\;\operatorname{rank}\big(\mathcal{C}(X)^\perp\big)\;;\;0\big)$$

4 Model adequacy checking

Model adequacy checking is large done via examination of the residuals of the model fit. Recall that the least-squares estimator $\widehat{Y} : \Omega \longrightarrow \mathcal{C}(X)$ of the response variable $Y : \Omega \longrightarrow \mathbb{R}^n$ is given by:

$$\widehat{Y} \;=\; X \cdot \big(X^t \cdot X\big)^{-1} \cdot X^t \cdot Y \;=\; H \cdot Y\,,$$

where $H := X \cdot (X^t \cdot X)^{-1} \cdot X^t$ is called the **hat matrix** of the model. Recall also that, geometrically speaking, the hat matrix H is simply the orthogonal projection operator, defined on $\mathbb{R}^n$ (the observation space), onto the column space $\mathcal{C}(X)$ of X (the estimation space, or the model space). The **residual** $\mathbf{e} : \Omega \longrightarrow \mathcal{C}(X)^\perp$ is defined to be:

$$\mathbf{e} \;:=\; Y - \widehat{Y} \;=\; (I_n - H) \cdot Y\,,$$

where $I - H$ is the orthogonal projection operator defined on $\mathbb{R}^n$ (the observation space) onto the orthogonal complement $\mathcal{C}(X)^\perp$ of $\mathcal{C}(X)$. Note that $\mathcal{C}(X)^\perp$ can be regarded as the **error space** of the model. Recall that our model assumption is:

$$Y \;=\; X \cdot \beta \,+\, \varepsilon\,,$$

with $\varepsilon \sim N\big(0, \sigma^2 I_n\big)$; see (2.2). Note that in general, the codomain of the error term $\varepsilon : \Omega \longrightarrow \mathbb{R}^n$ is NOT $\mathcal{C}(X)^\perp$ but all of the observation space $\mathbb{R}^n$. On the other hand, observe that

$$\mathbf{e} \;=\; (I_n - H) \cdot Y \;=\; (I_n - H) \cdot (X \cdot \beta \,+\, \varepsilon) \;=\; (I_n - H) \cdot \varepsilon\,,$$

since $I_n - H$ is the orthogonal projection operator onto $\mathcal{C}(X)^\perp$, which maps $X \cdot \beta \in \mathcal{C}(X)$ to zero. We thus see that the residual $\mathbf{e} : \Omega \longrightarrow \mathcal{C}(X)^\perp$ is the orthogonal projection of the error term $\varepsilon : \Omega \longrightarrow \mathbb{R}^n$ onto the error space $\mathcal{C}(X)^\perp$. Or, more strictly speaking, the residual $\mathbf{e} : \Omega \longrightarrow \mathcal{C}(X)^\perp$ is the composition

$$\mathbf{e} :\; \Omega \;\overset{\varepsilon}{\longrightarrow}\; \mathbb{R}^n \;\overset{I_n - H}{\longrightarrow}\; \mathcal{C}(X)^\perp$$

Furthermore, note that

$$\begin{aligned}
\operatorname{Var}(\mathbf{e}) \;&=\; \operatorname{Var}[\,(I_n - H) \cdot \varepsilon\,] \;=\; (I_n - H) \cdot \operatorname{Var}[\,\varepsilon\,] \cdot (I_n - H)^t \;=\; (I_n - H) \cdot \operatorname{Var}[\,\varepsilon\,] \cdot (I_n - H) \\
&=\; (I_n - H) \cdot \sigma^2 I_n \cdot (I_n - H) \;=\; \sigma^2 \cdot (I_n - H) \cdot (I_n - H) \\
&=\; \sigma^2 \cdot (I_n - H)\;,
\end{aligned}$$

where the symmetry and idempotence of the orthogonal projection operator $I_n - H$ is used in the above derivation. The above observations lead to the following "model adequacy checks":

- Generate the scatter plot of the observed residuals $\mathbf{e}$ against the fitted values $\widehat{y}$. Examine this scatter plot for trends between the observed residuals and the fitted values; any trend between the observed residuals and the fitted values may indicate violations of model assumptions.

  This adequacy check is based on the following fact:

$$\operatorname{Cov}\Big(\,\widehat{Y}\,,\,\mathbf{e}\,\Big) \;=\; \operatorname{Cov}(\,H \cdot Y\,,\,(I_n - H) \cdot Y\,) \;=\; H \cdot \operatorname{Cov}(\,Y\,,\,Y\,) \cdot (I_n - H) \;=\; H \cdot \sigma^2 I_n \cdot (I_n - H) \;=\; 0_{n \times n}$$

- Generate the scatter plot of the observed residuals $\mathbf{e}$ against the observed values of each of the predictor variables (columns of the model matrix X). Any trends in any of these scatter plots may indicate violations of model assumptions.
- Generate the QQ-plot of the **Studentized residuals** against the theoretical quantiles of the standard Gaussian distribution, where the Studentized residuals are defined as follows:

$$r_i \;:=\; \frac{e_i}{\sqrt{\mathrm{MS}_{\mathrm{error}}\,(1-h_{ii})}}$$

  where e_i is the i^{th} component of the observed residual $\mathbf{e}$, h_{ii} is the i^{th} diagonal element of the hat matrix $H := X\cdot\left(X^t\cdot X\right)^{-1}\cdot X^t$, and $\mathrm{MS}_{\mathrm{error}}$ is the mean squared error of the model fit, which is defined as follows:

$$\mathrm{MS}_{\mathrm{error}} \;:=\; \frac{1}{n-p}\sum_{i=1}^{n}\left(y_i-\widehat{y}_i\right)^2$$

  Large deviations of the data points on this QQ-plot from the $y=x$ line may indicate violations of model assumptions. This model adequacy check is based on the observations that (1) $\mathrm{MS}_{\mathrm{error}}$ is an unbiased estimator of σ^2, and (2):

$$\mathbf{e} \;\sim\; N\big(\,0\,,\,\sigma^2\,(I_n-H)\,\big)\;,$$

  which in turn implies that, for each $i=1,\ldots,n$,

$$\frac{e_i}{\sqrt{\sigma^2\,(1-h_{ii})}} \;\sim\; N(0,1)$$