Stat 455 Lecture 1: Introduction and Normal theory (08-28-2007)

1 Motivations

Previously, we applied linear models (LM) (regression, ANO(CO)VA ...) and generalized linear models (GLM) to estimate fixed effects (regression parameters) and report the results. Previous model diagnostics and model checking methods may give you some rough hints on the accountability of these models. However, these tools may fail to fail the fixed effects models when they are insufficient. Sometimes we need to improve the fixed effects (G)LMs by incorporating random effects. This work is more difficult since random effects are unobservable, and is important since they describe the realistic error structures, i.e., true story in the real-life.

2 Linear Model

2.1 Least Squares (LS) Estimation

We assume

1) The nature follows some certain rules

2) Uncertainty gives us the characteristics of statistical models.

A classic linear model with data points: y_i,$x_{i,1}$,$x_{i,2}$,. . .,$x_{i,p-1}$, $i = 1, 2, \ldots, n$. We may assume

$$y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_{p-1} x_{i,p-1} + \epsilon_i, (i = 1, 2, \ldots, n). \tag{1}$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & \ldots & x_{1,p-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,1} & \ldots & x_{n,p-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \tag{2}$$

or, $Y = X\beta + \epsilon$. Where error ϵs are independent with mean 0 and variance σ^2.

Remark

1) Although ϵs are errors, we prefer not to name them "random effects" here and reserve this term to later chapters. We may do regression analysis by least squares estimation (projection onto X space) without distributional assumption on ϵs. i.e.,

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{3}$$

2)) We do not incorporate variance of ϵs for estimating β since the correlation matrix is identity (trivial).

3) More complicated error structure will demand more realistic solutions (later). i.e., if the correlation matrix for *errors* is not I, we may try iterative algorithms (loops between estimating β given V and estimating V given β)

2.2 Estimation with Constrained Model

For example, we know three inner angles of a triangle has constraint: $\sum_{i=1}^{3}\beta_i = 180°$. We have linear model (get independent measurements at three triangle vertex):

$$y_{1i} = \beta_1 + e_{1i},\ (i = 1, \ldots, n_1);$$

$$y_{2i} = \beta_2 + e_{2i},\ (i = 1, \ldots, n_2);$$

$$y_{3i} = \beta_3 + e_{3i},\ (i = 1, \ldots, n_2);$$

for which we **must** follow model constraint for parameter estimation. For linear model: $y = X\beta + e$, where $e \sim N(0, \sigma^2 I)$, we estimate parameter β under constraint: $G\beta = g$. We have

- ⋄ $\tilde{\beta} = \hat{\beta} - (X^TX)^{-1}G^T[G(X^TX)^{-1}G^T]^{-1}(G\hat{\beta} - g)$
- ⋄ $\tilde{\sigma}^2 = \frac{1}{N-r}(y - X\tilde{\beta})^T(y - X\tilde{\beta})$, where r=rank of $I - (X^TX)^{-1}G^T[G(X^TX)^{-1}G^T]^{-1}G$

Example
The constraint is:

1) β_1=β_2=$\cdots$=β_p=0

2) β_1+β_2=0

3) ....

is corresponding to

1) constraint: $I\beta = 0$, then $\tilde{\beta} = \hat{\beta} - (X^TX)^{-1}[(X^TX)^{-1}]^{-1}\hat{\beta}$=0

2) constraint: $(1, -1, 0, \ldots, 0)\beta$=0, then

$$\tilde{\beta} = \hat{\beta}\text{-}\ (X^TX)^{-1}(1, -1, 0, \ldots, 0)^T[(1, -1, 0, \ldots, 0)(X^TX)^{-1}(1, -1, 0, \ldots, 0)^T]^{-1}((1, -1, 0, \ldots, 0)\hat{\beta})=0$$

♠ You may need plug in specific G and g values for problem solving.

2.3 General Linear Hypothesis Test with Unconstrained Model

Based on likelihood ratio test,

- ⋄ We test hypothesis H_0: $H\beta = h$ v.s. H_a: $H\beta \neq h$.
- ⋄ The H is a matrix of $v \times p$, and v is the full rank of H.
- ⋄ We calculate: $N_H = (H\hat{\beta} - h)^T[H(X^TX)^{-1}H^T]^{-1}(H\hat{\beta} - h)$.
- ⋄ The test statistic: $F = \frac{N_H/v}{Q(\hat{\beta})/(N-p)} \sim F(v, N-p)$ under H_0, where p is rank of full model, N is the number of observations.
- ⋄ $\hat{\beta}$ comes from full model estimation: $\hat{\beta} = (X^TX)^{-1}X^Ty$
- ⋄ $Q(\hat{\beta}) = (y - X\hat{\beta})^T(y - X\hat{\beta})$

♠ You may need to write down the specific hypothesis in terms of H.

Example

Consider the cell means model

$$y_{ij} = \mu_i + e_{ij},\ i = 1, \ldots, p,\ j = 1, \ldots, n$$

Test: H_0: $\mu_1 = \mu_2 = \cdots = \mu_p$.

Solution

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n} \\ \vdots \\ y_{p1} \\ y_{p2} \\ \vdots \\ y_{pn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \tag{4}$$

The null hypothesis is H_0: $H\beta = 0$.

Where, H could be

$$\begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & & & & \\ 1 & 0 & 0 & \cdots & -1 \end{bmatrix} \tag{5}$$

with size (p-1) by (p). $h = 0$.

3 Results from Normal Theory

3.1 Distribution-free Results

If y has mean μ and covariance Σ, then

- Var(By)=$B\Sigma B^T$, Cov($B_1 y$, $B_2 y$)=$B_1 \Sigma B_2^T$
- E(tr($y^T Ay$))=E(tr(Ayy^T))=tr(AV)+$\mu^T A\mu$

3.2 Univariate Normal Distribution

y_i, $i = 1, \ldots, N$ be a random sample from a normal distribution with mean μ and variance σ^2. The estimates of these parameters are

- sample mean: $\bar{y} = \frac{1}{N}\sum_{i=1}^N y_i = \frac{1}{N}J^T\tilde{y}$, this is a linear form of y.
- sample variance: $s_y^2 = \frac{1}{N-1}\sum_{i=1}^N (y_i - \bar{y})^2 = \frac{1}{N-1}\tilde{y}^T(I - \frac{1}{N}U)\tilde{y}$, where U is N by N matrix of ones. This is a quadratic form of y.
- $\bar{y} \sim N(\mu, \frac{\sigma^2}{N})$
- $s_y^2 \sim \frac{\sigma^2}{N-1}\chi^2(N-1)$

3.3 Multivariate Normal Distribution

For y_i $(i = 1, \ldots, n)$, a random sample from a normal distribution with mean μ and variance σ^2, we may take it as a N-dimensional multivariate normal random variable with mean (μ,μ,$\ldots$, μ) and variance matrix $\sigma^2 I_n$.

3.3.1 Conditional Distribution

◇ Suppose

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \tag{6}$$

follows multivariate normal distribution with

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \tag{7}$$

then

$$(y_2|y_1) \sim N(\mu_2 + V_{21}V_{11}^{-1}(y_1 - \mu_1), V_{22} - V_{21}V_{11}^{-1}V_{12})$$

3.3.2 Linear Form

◇ If $y \sim N(\mu, V)$, and $x = By + b$, then $X \sim N(B\mu + b, BVB^T)$.

◇ What is the distribution of $\bar{y}$?

3.3.3 Quadratic Form

◇ $y^T A_1 y$ and $y^T A_2 y$ are independent $\Longleftrightarrow A_1 V A_2 = 0$

◇ $y^T Ay$ and By are independent $\Longleftrightarrow BVA = 0$

♠ How to show that, sample average and sample variance are independent?

◇ If $y \sim N(\mu, V)$ and $q = y^T Ay$, then $q \sim \chi^2(r, \lambda)$, with $r = r(A)$ and $\lambda = \frac{1}{2}\mu^T A\mu$, if and only if AV is idempotent. i.e., $AVAV = AV$.

♠ Given special A, say $A = V^{-1}$, how to get the χ^2 distribution?

4 Skew Normal Distribution

Normal distribution is symmetric and widely used. However, sometimes there is a presence of high skewness (say salary distribution with some very rich people). Dey and Liu (2005) decomposed skewed elliptical distribution (normal is a special case) into an original symmetric component and accumulated linearly constrained (skewed) portion. A simple skew normal random variable u (Sahu, Dey and Branco, 2003) can be represented by

$$u = \delta Z + \nu,\ Z \sim \mathrm{N}^+(0, \sigma_Z^2) \quad \text{and} \quad \nu \sim \mathrm{N}(0, \sigma_\nu^2)$$

with N^+ indicates the folded (positive part) normal distribution. The likelihood function for u is

$$\pi(u|\delta) = \frac{2}{\sqrt{\sigma_\nu^2 + \delta^2\sigma_Z^2}}\phi(\frac{u}{\sqrt{\sigma_\nu^2 + \delta^2\sigma_Z^2}})\Phi(\frac{\delta\sigma_Z u}{\sigma_\nu\sqrt{\sigma_\nu^2 + \delta^2\sigma_Z^2}}). \tag{8}$$

Remark

1) It is unimodal (Azzalini, 1985).

2) What if $\delta = 0$?

3) Closed-form density carries over to general multivariate elliptical distributions including Students'-t and others, and correla

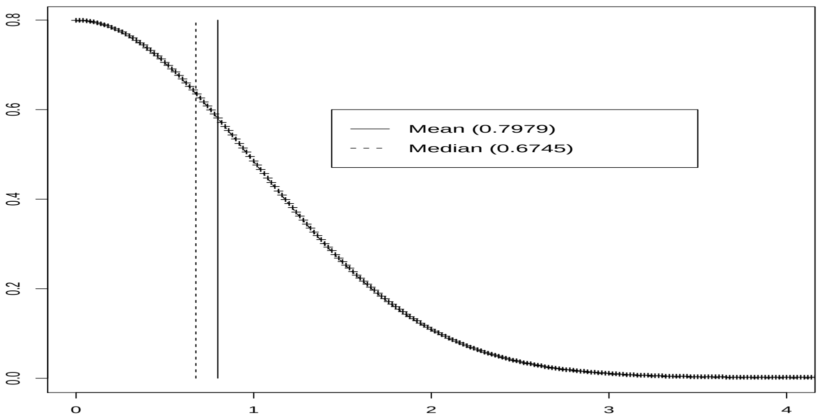

Figure 1: Folded normal density

5 References

◇ Azzalini, A. (1985), A class of distributions which includes the normal ones, Scand. J. Stat. 12, 171-178.

◇ Dey, D. K. and Liu, J. (2005). A new construction for skew multivariate distribution. Journal of Multivariate Analysis, Vol 95 No 2, 323-344.

◇ Jiang, J. (2006), Linear and Generalized Linear Mixed Models and Their Applications. Springer. ◇

◇ Hocking R. R. (1996), Methods and Applications of Linear Models: Regression and the Analysis of Variance. John Wiley and Sons, Inc.

◇ Sahu, S., Dey, D.K. and Branco, M. (2003). A new class of multivariate skew distributions with application to Bayesian regression models. The Canadian Journal of Statistics 31, 129-150.