1. Partitioning SS   2. nonlinear models   3. glm

Note Title   11/10/2009

1. $\underset{n\times 1}{y} = \underset{n\times p}{X}\underset{p\times 1}{\beta} + \underset{n\times 1}{\varepsilon}$

$\hat\beta = (X^TX)^{-1}X^Ty$

$X_1 = \underset{\sim}{1}_n^T$

$y^Ty = \hat\beta^TX^TX\hat\beta + (y-\hat y)^T(y-\hat y)$

$= (y - \hat y + \hat y)^T(y - \hat y + \hat y)$

$y^Ty - n\bar y^2 = \hat\beta^TX^TX\hat\beta - n\bar y^2 + (y-\hat y)^T(y-\hat y)$

Anova for reg.	total SS (corr'd)	= SS_{reg}	SS_{resid}
df	$n-1$	$p-1$	$n-p$

- we discussed partitioning SS_{reg} into $p-1$ individual pieces, one for each column of X (watch the order)

2. one-way anova $y_{tj} = \alpha + \gamma_t + \varepsilon_{tj}$ (2) $t=1,\dots,T$, $j=1,\dots,J$

$y = X\beta + \varepsilon$ $\underset{\sim}{\beta} = (\alpha, \gamma_1, \dots, \gamma_T)$ $\boxed{E(\bar y_{\cdot\cdot}) = \alpha \text{ iff } \sum_{t=1}^T \gamma = 0}$

$y_i = \beta_0 + \beta_1 d_{1i} + \beta_2 d_{2i} + \dots + \beta_{T-1}d_{T-1,i} + \varepsilon_i$ (1)

$i = 1,\dots,TJ$ $d_{1i} = \begin{cases} 1 & \text{if } y_i \text{ is in gp } 1 \\ 0 & \text{o.w.} \end{cases}$

$d_{2i} = \begin{cases} 1 & \text{if } y_i \text{ is in gp } 2 \\ 0 & \text{o.w.} \end{cases}$

$\vdots$

$d_{T-1,i} = \begin{cases} 1 & \text{if } y_i \text{ is in gp } T-1 \\ 0 & \text{o.w.} \end{cases}$

~~these~~ (1) & (2) are the same model with different param's

$X = \begin{pmatrix} 1 & d_{11} & \cdots & d_{T-1,1} \\ \vdots & \vdots & & \vdots \\ 1 & d_{1n} & & d_{T-1,n} \end{pmatrix}$ $E(y_{3j}) \underset{(2)}{=} \alpha + \gamma_3 \underset{(1)}{=} \beta_0 + \beta_3$

$E(y_{5j}) \underset{(2)}{=} \alpha + \gamma_5 \underset{(1)}{=} \beta_0$

$E(y_{1j}) = \alpha + \gamma_1 = \beta_0 + \beta_1$

$E(y_{2j}) = \alpha + \gamma_2 = \beta_0 + \beta_2$

$E(y_{4j}) = \alpha + \gamma_4 = \beta_0 + \beta_4$

$\beta_0 \quad \beta_1 \cdots \beta_4 \qquad \alpha \quad \gamma_1 \cdots \gamma_5$ need to put some constraint

a) $E(y_{tj}) = \alpha + \gamma_t \qquad \gamma_1 = 0$ constraint

$E(y_{1j}) = \alpha \qquad E(y_{2j}) = \alpha + \gamma_2 \quad \cdots \quad E(y_{5j}) = \alpha + \gamma_5$

$\hat{\gamma}_2 \quad \hat{\gamma}_3 \quad \hat{\gamma}_4 \quad \hat{\gamma}_5$

$E(y_{2j}) - E(y_{1j}) = \gamma_2$

$E(y_{5j}) - E(y_{1j}) = \gamma_5 \qquad E\hat{\gamma}_5 = E(y_{5j}) - E(y_{1j})$

b) $E(y_{tj}) = \alpha + \gamma_t \qquad \gamma_5 = 0$ constraint

$E\hat{\gamma}_1 = E(y_{1j}) - E(y_{5j})$

> options (contrasts)

[1] "contr.treatment" "contr.poly"

↑ factors ↑ ordered factors

translation: $\gamma_1 = 0$

...

[1] "contr.sum" "contr.poly"

↑ $\Sigma \gamma_t = 0$

$E(y_{2j}) = \alpha + \gamma_2$

$E(y_{5j})$

$= \alpha - (\gamma_1 + \cdots + \gamma_4)$

$E(y_{5j}) - E(y_{2j}) = \alpha + \gamma_5 - \alpha - \gamma_2 =$

$= \gamma_5 - \gamma_2$

$$\cancel{E} \quad y_{tj} = \alpha + \gamma_t + \varepsilon_{tj}$$

R default $\quad \gamma_1 = 0 \qquad \hat\gamma_1 = 0 \quad \hat\gamma_2 = \bar y_{2\cdot} - \bar y_{1\cdot}, \ldots, \hat\gamma_5 = \bar y_{5\cdot} - \bar y_{1\cdot}$

$$\hat\alpha = \bar y_{1\cdot}$$

SAS default $\rightarrow \gamma_5 = 0 \qquad \hat\gamma_1 = \bar y_{1\cdot} - \bar y_{5\cdot}, \ \hat\gamma_2 = \bar y_{2\cdot} - \bar y_{5\cdot}, \ldots, \hat\gamma_5 = 0$

R option $\quad \sum \gamma_t = 0 \qquad \hat\alpha = \bar y_{\cdot\cdot} \quad \hat\gamma_t = \bar y_{t\cdot} - \bar y_{\cdot\cdot}$

SS in this linear model

$$\sum_{tj} (y_{tj} - \bar y_{\cdot\cdot})^2 = \sum_{tj} (\bar y_{t\cdot} - \bar y_{\cdot\cdot})^2 + \sum_{tj} (y_{tj} - \bar y_{t\cdot})^2$$

SS_{total} $\qquad SS_{\text{between groups}}$ $\qquad SS_{\text{within groups}}$

df $\qquad TJ - 1 \qquad T - 1 \qquad T(J-1)$

$$J \sum_{t=1}^{T} (\bar y_{t\cdot} - \bar y_{\cdot\cdot})^2 = J \left[\frac{\left\{ \sum_{t=1}^{T} c_{1t} (\bar y_{t\cdot} - \bar y_{\cdot\cdot}) \right\}^2}{\underline{\sum_{t=1}^{T} c_{1t}^2}} + \frac{\left\{ \sum_{t=1}^{T} c_{2t} (\bar y_{t\cdot} - \bar y_{\cdot\cdot}) \right\}^2}{\underline{\sum c_{2t}^2}} + \ldots + \frac{\left\{ \sum_{t=1}^{T} c_{T-1,t} (\bar y_{t\cdot} - \bar y_{\cdot\cdot}) \right\}^2}{\underline{\sum c_{T-1,t}^2}} \right]$$

vector $\bar{y} = (\bar y_{1\cdot}, \ldots, \bar y_{T\cdot})$ $\qquad \bar y_{\cdot\cdot} = \frac{1}{T} \sum_{t=1}^{T} \bar y_{t\cdot}$ scalar

$$(\bar{y} - \underset{\sim}{1} \bar y)^T (\bar{y} - \underset{\sim}{1} \bar y)$$

$$= (\bar{y} - \underset{\sim}{1} \bar y)^T \underline{\tilde C^T \tilde C} (\bar{y} - \underset{\sim}{1} \bar y)$$

$1 \times T$

$$\underset{T-1 \times T}{C} = \begin{pmatrix} c_{11} & \cdots & c_{1T} \\ c_{21} & \cdots & c_{2T} \\ \vdots & & \\ c_{T-1,1} & \cdots & c_{T-1,T} \end{pmatrix}$$

$*$ $C^T C = \text{Diagonal}_T$
$= \text{Diag}\left(\sum_{t=1} c_{kt}^2 \right)$

$$\tilde{C}^T \tilde{C} = I = \begin{pmatrix} \sum c_{1,t}^2 & & O \\ & \ddots & \\ O & & \sum c_{T-1,t}^2 \end{pmatrix}$$

$$\left\{ \begin{array}{cccc} \underset{\sim}{c}_1 & -1 & 0 & 1 \\ \underset{\sim}{c}_2 & +1 & -2 & +1 \end{array} \right\}$$

$$R \quad \left. \begin{array}{cccc} \underset{\sim}{\tilde{c}}_1 & \frac{-1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \underset{\sim}{\tilde{c}}_2 & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{array} \right]$$

behind the scenes, if factor is ordered

§ 9.3.2 has a slightly different version.

1) $\text{logit } p_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$

$\alpha_1 = \beta_1 = 0$
$\alpha\beta_{1j} = 0 \ \forall j$
$\alpha\beta_{21} = 0$

$\text{logit } p_{23} - \text{logit } p_{13}$

$$= \cancel{\mu} + \alpha_2 + \cancel{\beta_3} + (\alpha\beta)_{23} - (\cancel{\mu} + \underset{\uparrow 0}{\alpha_1} + \cancel{\beta_3} + \underset{\uparrow 0}{(\alpha\beta)_{13}})$$

$$= \alpha_2 + (\alpha\beta)_{23}$$

2) $\text{logit } p_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$

$\sum \alpha_i = 0 \quad \sum \beta_j = 0$
$\sum_i (\alpha\beta)_{ij} = 0 = \sum_j (\alpha\beta)_{ij}$

$\text{logit } p_{23} - \text{logit } p_{13}$

$$= \mu + \alpha_2 + \beta_3 + (\alpha\beta)_{23} - (\mu + \alpha_1 + \beta_3 + \alpha\beta_{13})$$

$$= \alpha_2 - \alpha_1 + (\alpha\beta)_{23} - (\alpha\beta)_{13}$$

$$\sum_{t=1}^{T} \operatorname{logit} p_{tj} = \sum_{t} \{\mu + \alpha_t + \beta_j + (\alpha\beta)_{tj}\} \qquad \sum \alpha_t = 0$$

$$= T\mu + 0 + T\beta_j + 0$$

Linear regression

1) $y = X\beta + \varepsilon$

$\underset{\sim}{y} \sim N_n(X\beta, \sigma^2 I)$ $(*)$

§10.1.2 Nonlinear regression

$f(y_j; \eta_j)$ $j = 1, \dots, n$ independent

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \prod_{j=1}^{n} f(y_j; \eta_j)$$

$$\eta_j = \eta_j(\beta) \qquad \beta = (\beta_1 \cdots \beta_p)$$

2) Least squares $\hat{\beta} = (X^T X)^{-1} X^T y$

maximum likelihood

$$\hat{\beta}: \sup_{\beta} L(\beta; \underset{\sim}{y})$$

$$= \sup_{\beta} \underbrace{\prod_{j=1}^{n} f(y_j, \eta_j(\beta))}_{\ell}$$

$$\left.\frac{\partial \ell(\beta)}{\partial \beta}\right|_{\beta=\hat{\beta}} = 0 \quad \text{usually} \quad \left.\frac{\partial \log L(\beta)}{\partial \beta}\right|_{\beta = \hat{\beta}} = \underset{\sim}{0}$$

nonlinear

IWLS

iteratively weighted

3) $E\hat{\beta} = \beta$ $\operatorname{var}\hat{\beta} = \sigma^2 (X^T X)^{-1}$

$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$ under $(*)$

[not necessary to insist $\varepsilon \sim N$]

$$\hat{\beta} \xrightarrow{d} N_p(\beta, \Omega) \quad \checkmark$$

$$\hat{\Omega} = - \left.\frac{\partial^2 \ell(\beta)}{\partial \beta \, \partial \beta^T}\right|_{\beta = \hat{\beta}}$$

$$\hat{\underset{\sim}{\beta}} \mathrel{\dot\sim} N_p(\underset{\sim}{\beta}, -\ell''(\hat{\underset{\sim}{\beta}}))$$

$$E(\hat\beta) = \beta + \frac{1}{n}(\text{big mess}) + \frac{1}{n^2}(\text{bigger mess}) + \cdots$$

not needed yet

4) $\frac{SS_{res}}{n-p}$ estimates σ^2

5) analysis of variance

analysis of deviance

i) $y^T y - n\bar y^2 = \hat\beta^T X^T X \hat\beta + (y-\hat y)^T(y-\hat y)$

ii) <u>reminder</u> a) fit $\underset{\sim}{y} = \beta_0 \underset{\sim}{1} + \underset{\sim}{\varepsilon}$

get $SS_{residual,0}$

b) fit $y = \cancel{\beta_0} X\beta + \underset{\sim}{\varepsilon}$

get $SS_{residual,full}$

$SS_{residual,0} - SS_{residual,full}$

$= SS_{regression}$

a) fit model

$\eta_j = \eta_j(\beta_0)$ β_0 sc.

b) get max'd log-lik

$\ell(\hat\beta_0)$

b) fit $\eta_j = \eta_j(\underset{\sim}{\beta})$

get max'd log-lik

$\ell(\hat{\underset{\sim}{\beta}})$ $\beta = (\beta_0, \ldots, \beta_{p-1})$

$$2\{\ell(\hat{\underset{\sim}{\beta}}) - \ell(\beta_0)\} \xrightarrow{d} \chi^2_{p-1}$$

Quick sketch: $\left.\frac{\partial \ell}{\partial \beta}\right|_{\beta=\hat\beta} = 0 = \ell'(\hat\beta)$

1) $0 = \ell'(\hat\beta) = \ell'(\beta) + (\hat\beta - \beta)\ell''(\beta) + R_n$ $\qquad \ell'(\beta) = \sum_{j=1}^n \ell_j'(\beta)$

$$\hat\beta - \beta \simeq -\frac{\ell'(\beta)}{\ell''(\beta)} \simeq N$$

2) $\ell(\hat\beta) - \ell(\beta) = \ell(\hat\beta) - \{\ell(\hat\beta) + (\beta-\hat\beta)\ell'(\hat\beta) + \frac{1}{2}(\beta-\hat\beta)^2\ell''(\hat\beta) + \cdots\}$

$$2\{\ell(\hat\beta) - \ell(\beta)\} = \underbrace{(\beta-\hat\beta)^2\ell''(\hat\beta)}_{\dot\sim\ \chi^2_p}$$

bec.

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

Analysis of deviance w̄ linear model ?

$y \sim N(X\beta, \sigma^2 I)$

$$L(\beta, \sigma^2) = \frac{1}{(\sqrt{2\pi}\,\sigma)^n} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(y_j - x_j^T\beta)^2}$$

$$\ell(\beta, \sigma^2) = -n\log\sigma - \frac{1}{2\sigma^2}\sum_{j=1}^{n}(y_j - x_j^T\beta)^2 - \frac{n}{2}\log(2\pi)$$

$\frac{\partial \ell}{\partial \beta} = 0$ is a set of linear eq^{n}s $\qquad \sigma^2$ known

$$\ell(\hat\beta) - \ell(\beta) = -\frac{1}{2\sigma^2}\sum(y_j - x_j^T\hat\beta)^2 + \frac{1}{2\sigma^2}\sum(y_j - x_j^T\beta)^2$$

$$= \frac{1}{2\sigma^2}(\hat\beta - \beta)^T X^T X(\hat\beta - \beta) \qquad \text{n.t.b.c.}$$

$$2\{\qquad\qquad\} = \frac{SS_{regr}}{\sigma^2}$$ $\rightarrow$ leads eventually to an F-test

---

R reports scaled deviance $\neq -2\ell(\hat\beta)$

Def: §10.2 : scaled dev. $D = 2\sum_{j=1}^{n}\{\log f(y_j, \tilde\eta_j) - \log f(y_j, \eta_j(\hat\beta))\}$

$$= 2\sum_{j=1}^{n} \ell_j(\tilde\eta_j) - \ell_j(\eta_j(\hat\beta))$$

$\eta_j(\hat\beta)$ is value of η_j at m.l.e. $\hat\beta$

$\tilde\eta_j$ is maximum lik. est. of η_j when no dependence on β (saturated model)

Example i) $y_j \sim N(\eta_j, \sigma^2)$ $\qquad \tilde{\eta}_j = y_j$

ii) $y_j \sim N(\eta_j = x_j^T\beta, \sigma^2)$ $\qquad \hat{\eta}_j = \eta_j(\hat{\beta}) = x_j^T\hat{\beta}$

$$D = 2\sum_{j=1}^{n}\Big\{-\frac{1}{2\sigma^2}\underbrace{(y_j - y_j)^2}_{\overset{|||}{0}} - n\log\sigma + \frac{1}{2\sigma^2}(y_j - x_j^T\hat{\beta})^2 + n\log\sigma\Big\}$$

$$= \frac{1}{\sigma^2}\sum_{j=1}^{n}(y_j - x_j^T\hat{\beta})^2 \qquad \text{scaled deviance} = \frac{SSresid}{\sigma^2}$$

Example $\quad y_j \sim Bin(m_j, p_j)$ $\quad j = 1, \dots, n$ (indep't)

$(m_j, \eta_j/m_j)$ $\qquad \eta_j = m_j p_j$

i) saturated $\quad \tilde{\eta}_j = y_j$ $\qquad f(y_j, \eta_j) = \underline{\binom{m_j}{y_j}}\left(\frac{\eta_j}{m_j}\right)^{y_j}\left(1 - \frac{\eta_j}{m_j}\right)^{m_j - y_j}$

$\hat{p}_j = y_j/m_j$

ii) $\quad p_j = p_j(\beta) = \dfrac{e^{x_j^T\beta}}{1 + e^{x_j^T\beta}}$ $\qquad \hat{\beta} = \dots$ solved iteratively

$$D = 2\sum_{j=1}^{n}\Big[\underbrace{y_j\log\frac{y_j}{m_j} + (m_j - y_j)\log\left(\frac{m_j - y_j}{m_j}\right)}_{\text{not zero}}$$

$$-\{y_j\log p_j(\hat{\beta}) + (m_j - y_j)\log\{1 - p_j(\hat{\beta})\}\}\Big]$$

scaled deviance for binomial

$$= 2\sum_{j=1}^{n}\Big\{y_j\log\left(\frac{y_j}{m_j p_j(\hat{\beta})}\right) + (m_j - y_j)\log\frac{(m_j - y_j)}{(m_j - y_j)(1 - \hat{p}_j(\beta))}\Big\}$$

$O\log\frac{O}{E} \qquad O\log\frac{O}{E}$

$$\simeq \sum \frac{(O-E)^2}{E} \quad \text{Pearson}$$

$\tilde{\eta}_j = y_j$??

i) saturated model $\quad y_j$ density $f(y_j, \eta_j)$

$$L(\eta) = \prod_{j=1}^{n} f(y_j, \eta_j)$$

$$\ell(\eta) = \sum_{j=1}^{n} \ell_j(y_j, \eta_j)$$

$$\frac{\partial \ell}{\partial \eta_k} = 0 \qquad k = 1, \dots, n$$

$$= \frac{\partial \ell_k(y_k, \eta_k)}{\partial \eta_k} = 0 \qquad \tilde{\eta}_k = \tilde{\eta}_k(y_k)$$

$N(\eta_j, \sigma^2)$ $\quad - \quad$ $\ell_k = -\dfrac{(y_k - \eta_k)^2}{2\sigma^2}$ $\quad \tilde{\eta}_k = y_k$

$Bin(m_j, p_j)$ $\quad \ell_k = y_k \log p_k + (m_k - y_k)\log(1-p_k)$ $\quad \tilde{p}_k = \dfrac{y_k}{m_k}$

ii) model of interest $\quad \eta_j = \eta_j(\beta)$ $\quad$ link observations through β

$j = 1, \dots, n \qquad \beta = \beta_1 \cdots \beta_p \qquad p < n$

See §10.2 for analysis of deviance described as

Model A $\quad \eta_j = \eta_j(\beta) \qquad \beta = (\beta_1, \dots, \beta_p)$

Model B $\quad \eta_j = \eta_j(\beta_0) \qquad \beta_0 = (\beta_1, \dots, \beta_q, 0, \dots, 0)$

sc. dev. $D_A = 2\sum\left\{\ell_j(\tilde{\eta}_j) - \ell_j(\eta_j(\hat{\beta}))\right\}$

sc. dev $D_B = 2\sum\left\{\ell_j(\tilde{\eta}_j) - \ell_j(\eta_j(\hat{\hat{\beta}}_0))\right\}$

$$D_B - D_A = 2\sum_{j=1}^{n}\{\ell_j(\eta_j(\hat{\beta})) - \ell_j(\eta_j(\hat{\hat{\beta}}_0))\}$$

$$= 2\{\ell(\hat{\beta}) - \ell(\hat{\hat{\beta}}_0)\} \xrightarrow{d} \chi^2_{p-q} \text{ under}$$

$H_0: \beta_{q+1} = \dots = \beta_p = 0$

p. 473 talks about finding $\hat{\beta}$

... at convergence $\hat{\beta} = (\hat{X}^T\hat{W}\hat{X})^{-1}\hat{X}^T\hat{W}\hat{z}$ weighted LS solution

$$X_{n\times p} = \frac{\partial\eta(\beta)}{\partial\beta^T} = \begin{pmatrix} \frac{\partial\eta_1}{\partial\beta_1} & \cdots & \frac{\partial\eta_1}{\partial\beta_p} \\ \vdots & & \vdots \\ \frac{\partial\eta_n}{\partial\beta_1} & & \frac{\partial\eta_n}{\partial\beta_p} \end{pmatrix}$$

$\underset{n\times n}{W(\beta)}$ = diagonal $W_{jj}(\beta) = E\left(-\frac{\partial^2\ell_j}{\partial\eta_j^2}\right)$

$$z_{n\times 1} = \underset{n\times p}{X(\beta)}\cdot\underset{p\times 1}{\beta} + \underset{n\times n}{W^{-1}(\beta)}\,\underset{n\times 1}{u(\beta)}$$

$$u_j(\beta) = \frac{\partial\ell(\beta)}{\partial\eta_j} \quad j = 1, \dots, n$$