HW # 2 due Thursday Mar 3

GLM : max. lik. eq'ns

$$\sum A_i \left[\frac{y_i - \mu_i(\hat\beta)}{V\{\mu_i(\hat\beta)\}} \right] \frac{x_{ij}}{g'(\mu_i(\hat\beta))} = 0,$$

$j = 1, \dots, p$

(defines $\hat\beta$)

$g(\cdot)$ <u>link</u> function ($\ell(\cdot)$)

$g(\mu_i) = \underline{x}_i^T \underline{\beta}$ $\quad$ ~~g' mean~~ $\frac{\partial}{\partial \beta_j}$

$g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) \quad g' \dots$

simpler: $V \times g' = 1$ $\quad$ (if $\theta_i = x_i^T \beta$)

$$\sum A_i (y_i - \hat\mu_i) x_{ij} = 0$$

$\sum_{j=1}^{p} x_{ij} \beta_j = \underline{x}_i^T \underline{\beta}$

even simpler $\quad \mu_i = x_i^T \beta$

$$\sum A_i (y_i - \underline{x}_i^T \underline{\beta}) x_{ij} = 0$$

<u>wt'd LS</u>

$$
\begin{array}{c}
 \\
i = 1 \\
\vdots \\
\vdots \\
n
\end{array}
\begin{array}{c}
j = 1 \cdots p \\
\begin{bmatrix} x_{11} & \cdots & x_{11} \\ \vdots & & \\ \vdots & & \\ x_{n1} & \cdots & x_{np} \end{bmatrix}
\end{array}
\longrightarrow \underset{1 \times p}{\underline{x}_i}
$$

$\uparrow$

1 (usually)

p covariates

n obs

WLS connection suggests:

new response $\underline{z}_i = \underset{\overset{\shortparallel}{g(\mu_i)}}{\eta_i} + (y_i - \mu_i) g'(\mu_i)$

$$E z_i = g(\mu_i) \qquad \operatorname{var} z_i = \frac{\phi}{A_i} V(\mu_i) \{g'(\mu_i)\}^2$$

use these as weights

WLS algorithm

$$\hat{\mu}_i^{(0)} = \mu_i(\hat{\beta}^{(0)})$$

$$\left[\begin{array}{l} \hat{z}_i^{(t)} = g(\hat{\mu}_i^{(t)}) + \{ y_i - \mu_i(\hat{\beta}^{(t)}) \} g'(\hat{\mu}_i^{(t)}) \\ \hat{w}_i^{(t)} = \dfrac{A_i}{V(\hat{\mu}_i^{(t)}) \{ g'(\hat{\mu}_i^{(t)}) \}^2} \end{array} \right.$$

$\leftarrow$ left out ϕ

$$\hat{\beta}^{(t+1)} = (X^T \hat{W}^{(t)} X)^{-1} X^T \hat{W}^{(t)} \hat{z}^{(t)}$$

continue until "convergence"

(Note ϕ is <u>not needed</u> for computation of $\hat{\beta}$)

Can show: at convergence

~~a$\hat{\beta}$ ... ϕ~~

$$\text{as.var}\,\hat{\beta} = \phi\,(X^T W X)^{-1}$$ (just like WLS)

estimate W using $\hat{W} = \text{diag}(\hat{w}_1, \ldots, \hat{w}_n)$

To get CIs for β, we need an estimate of ϕ

in normal $\phi = \sigma^2$

in gamma $\phi = 1/\beta$

in binomial $\phi = 1$

Poisson $\phi = 1$

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum (y_i - \hat{y}_i)^2 \qquad \frac{SSE}{d.f.}$$

By analogy

$$\hat{\phi} = \frac{1}{n-p} \sum A_i \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$$ "dispersion parameter"

~~gl~~ glm (cbind(r, m-r) ~ rain + rain2 + rain3,
family = binomial, data = tox)

"null deviance" 74.21 ← like full SS_{tot}

"residual deviance" 62.63 ← like SSE

Recall, if we have a log-likelihood fn

$\ell(\theta)$ & a sub-model $\ell(\theta_1, 0)$

$= \ell(\theta_1, \theta_2)$

$2\{\ell(\hat\theta_1, \hat\theta_2) - \ell(\tilde\theta_1, 0)\}$ where $\tilde\theta_1$ is mle when $\theta_2 = 0$

$\xrightarrow{d} \chi^2_{df}$

(likelihood ratio test)

deviances are log-likelihoods for 2 models

1. model fitted ~~r + r + r^2 + r^3~~

$\log\left\{\frac{\text{pr}(+ve)}{\text{pr}(-ve)}\right\} = \beta_0 + \beta_1 \text{rain} + \beta_2 \text{rain}^2 + \beta_3 \text{rain}^3$

" ~~$\ell(\hat\theta_1, \hat\theta_2)$" a ~~for is~~ di~~

resid. dev. is ~~diff fitted and~~ $2[\ell(\text{huge model}) - \ell(\text{fitted model})]$

null. dev. $2[\ell(\text{huge model}) - \ell(\text{only } \beta_0)]$

null dev − resid. dev

$= 2l(\text{fitted model}) - 2l(\text{only } \beta_0 \text{ model})$

this difference ≡ likelihood ratio test of only β_0 model compared to $(\beta_0\ \beta_1\ \beta_2\ \beta_3)$ model

LRT of $H_0: \beta_1 = \beta_2 = \beta_3 = 0$

$74.21 - 62.63 = 11.7(?) \ \dot{\sim}\ \chi^2_3$

"significant"

$\Rightarrow \beta_1, \beta_2, \beta_3$ not all $= 0$

---

l(huge model) huge means $\hat{\mu}_i = y_i$

---

for binomial it means $\hat{\mu}_i = \frac{r_i}{m_i}$ where

r_i = # +ve's in city i
m_i = # tests

(dispersion parameter for binomial taken to be 1)

$\phi = 1$ in binomial

(Same for Poisson)

family = gamma .... disp. par. ... ——

Number of Fisher scoring iterations 3

$\hat{var}\,\hat{\beta} = \hat{\phi}(X^T \hat{W} X)^{-1}$ matrix $p \times p$

jth diagonal element estimates a.$\hat{var}(\hat{\beta}_j)$

approx. test for $\beta_j = 0$ $\hat{\beta}_j / \sqrt{a.\hat{var}(\hat{\beta}_j)}$

> summary (glm)

In general: if model

$$g(\mu_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \dots + \beta_p x_i^p$$

is contemplated, then all terms of lower order than the highest significant one should be retained.

Newton ~~Raphsone~~ Raphson

$$\ell'(\hat{\theta}) = 0 \qquad \text{m.l. eq'n}$$

$$\ell'(\hat{\theta}) \doteq \ell'(\theta_0) + (\hat{\theta} - \theta_0)\,\ell''(\theta_0)$$

1st 2 terms of Taylor

$$\hat{\theta} - \theta_0 = \frac{-\ell'(\theta_0)}{\ell''(\theta_0)}$$

$$\hat{\theta} = \theta_0 - \frac{\ell'(\theta_0)}{\ell''(\theta_0)} \qquad \text{solved.}$$

$$\hat{\theta}^{(t)} = \hat{\theta}^{(t-1)} - \frac{\ell'(\hat{\theta}^{(t-1)})}{\ell''(\hat{\theta}^{(t-1)})}$$

iterate until conv.