GLM: max. lik. eq'n's

\[ \sum A; \left[ \frac{y_i - \mu_i(\hat{\beta})}{\sqrt{\mu_i(\hat{\beta})}} \right] \frac{x_{ij}}{g'(\mu_i(\hat{\beta}))} = 0, \quad j = 1, \ldots, p \]

(defines \( \hat{\beta} \))

\[ g(\cdot) \text{ link function (e.g.,) } \]
\[ g(\mu_i) = x_i^T \beta \quad \text{if } \theta_i = x_i^T \beta \]
\[ g(\mu_i) = \log \left( \frac{\mu_i}{1 - \mu_i} \right) \quad g' = \frac{1}{1 - \mu_i} \]

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

\[ \sum A; (y_i - \hat{\mu}_i) x_{ij} = 0 \]

\[ \frac{1}{j} \sum x_{ij} \beta_j = \hat{\beta} \]

even simpler \( \mu_i = x_i^T \beta \)

\[ \sum A; (y_i - x_i^T \hat{\beta}) x_{ij} = 0 \quad \text{wt'd LS} \]
WLS connection suggests:

new response $z_i = \eta_i + (y_i - \mu_i)g'(\mu_i)\,$

$g(\mu_i)$

$Ez_i = g(\mu_i)$, $\text{var } z_i = \frac{\text{var}(\mu_i) [g'(\mu_i)]^2}{A_i}$

--- WLS algorithm ---

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

$\hat{z}_i^{(t)} = g(\hat{\mu}_i^{(t)}) + (y_i - \hat{\mu}_i^{(t)})g'(\hat{\mu}_i^{(t)})$

$\hat{W}_i^{(t)} = \frac{A_i}{\text{var}(\hat{\mu}_i^{(t)}) [g'(\hat{\mu}_i^{(t)})]^2}$

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

continue until "convergence"

(Note $\phi$ is not needed for computation of $\hat{\beta}$)
Can show: at convergence

$$\hat{\beta} \sim N(\beta, \text{var} \hat{\beta})$$

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

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

To get CIs for $\beta$, we need an estimate of $\phi$

in normal $\phi = \sigma^2$

in gamma $\phi = 1/\beta$

in binomial $\phi = 1$

Poisson $\phi = 1$

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

By analogy

$$\hat{\phi} = \frac{1}{n-p} \sum A_i \left( \frac{y_i - \hat{y}_i}{V(\hat{y}_i)} \right)^2$$

"dispersion parameter"
glm (cbind(r, m-r) ~ rain + rain^2 + rain^3, 
family = binomial, data = tox)

"null deviance"  74.21  ~ like full SStot
"residual deviance"  62.63  ~ like SSE

Recall, if we have a log-likelihood $l(\theta)$ & a sub-model $l(\theta_1, 0) = l(\theta_1, \theta_2)$

$$2 \left( l(\tilde{\theta}_1, \tilde{\theta}_2) - l(\tilde{\theta}_1, 0) \right)$$

where $\tilde{\theta}_1$ is mle when $\theta_2 = 0$

$\frac{d}{df} \xrightarrow{} \chi^2$ (likelihood ratio test)

deviances are log-likelihoods for 2 models
1. model fitted

$$\log\left( \frac{\hat{m}(\text{true})}{\hat{m}(\text{true})} \right) = \beta_0 + \beta_1 m + \beta_2 m^2 + \beta_3 m^3$$

 Neyman's test

resid. dev. is $\sum_{i=1}^{n} (\text{true} - \hat{m}(\text{true}))^2$

null dev. is $\sum_{i=1}^{n} (\text{true} - \hat{m}(\text{true} | m = 0))^2$
null dev - resid. dev

\[ \Lambda = 2l (\text{fitted model}) - 2l (\text{only } \beta_0 \text{ model}) \]

This difference \(\equiv\) likelihood ratio test of only \(\beta_0\) model compared to \((\beta_0, \beta_1, \beta_2, \beta_3)\) model

LRT of \(H_0: \beta_1 = \beta_2 = \beta_3 = 0\)

\[ 74.21 - 62.63 = 11.7(?) \sim \chi^2_3 \]

"significant"

\[ \Rightarrow \beta_1, \beta_2, \beta_3 \text{ not all } = 0 \]

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

For binomial it means \(\hat{\beta}_i = \frac{r_i}{m_i}\) where

\[ r_i = \# \text{ for's in city } i \]

\[ m_i = \# \text{ tests} \]
(dispersion parameter for binomial taken to be 1)
\[ \phi = 1 \quad \text{in binomial} \]

(Same for Poisson)

family = gamma \quad \text{disp. per} \quad \text{---}

Number of Fisher scoring iterations 3

\[ \text{Var} \hat{\beta} = \hat{\phi} (X^T \hat{W} X)^{-1} \quad \text{matrix} \quad p \times p \]

jth diagonal element estimates \( \text{var} (\hat{\beta}_j) \)

approx. test for \( \beta_j = 0 \)

\[ \hat{\beta}_j / \sqrt{\text{var} (\hat{\beta}_j)} \]

> summary (glm)

In general: if model \( g \)

\[ g(\mu_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \cdots + \beta_p x_i^p \]

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

\[ l' (\hat{\theta}) = 0 \quad \text{m.l. eq.} \]

\[ l' (\hat{\theta}) = l'' (\theta_0) + (\hat{\theta} - \theta_0) l'' (\theta_0) \]

\[ \hat{\theta} - \theta_0 = \frac{-l' (\theta_0)}{l'' (\theta_0)} \]

\[ \hat{\theta} = \theta_0 - \frac{l' (\theta_0)}{l'' (\theta_0)} \quad \text{solved.} \]

\[ \hat{\theta}^{(t)} = \hat{\theta}^{(t-1)} - \frac{l' (\hat{\theta}^{(t-1)})}{l'' (\hat{\theta}^{(t-1)})} \]

iterate until conv.