GLM Models

response \( y_i \): not (usually) normally dist'd

linear predictor \( \mathbf{x}_i^T \beta = \eta_i \): just like linear regression

+ link function \( \ell(\mu_i) = \eta_i \): where \( \mu_i = \text{E}y_i \)

\[ g(\cdot) \text{ in HO chapter, but } \ell(\cdot) \text{ in VR} \]

distribution of \( y_i \):

\[
f(y_i) = \exp \left[ \frac{A_i \{ y_i - \gamma(\theta_i) \}}{\phi} + \tau(y_i, A_i/\phi) \right] \]

\( A_i \): known

\( \gamma(\cdot) \) is a known \( f^2 \), so is \( \tau(\cdot, \cdot) \)

\( \phi \): scale parameter

\( \theta_i = \theta_i(\mu_i) = \ldots = \theta_i(\mathbf{x}_i^T \beta) \)

\( \gamma(\cdot) \) inverse

\[
\text{E}y_i = \gamma'(\theta_i) \rightarrow \mu_i;
\]

\[
\text{var } y_i = \frac{\phi}{A_i} \gamma''(\theta_i);
\]

\[
= \frac{\phi}{A_i} \gamma''(\mu_i);
\]

because of (\ref{eq:gamma})
Ex. 1 \( y_i \sim N(\mu_i, \sigma^2) \)  \( \phi = \sigma^2 \) \( A_i = 1 \) \( \Theta_i = \mu_i \) \( V(\mu_i) = 1 \)
\[
\text{var}(y_i) = \sigma^2 = \frac{\sigma^2}{A_i} \cdot 1
\]

\( y_i \sim \text{Poisson}(\mu_i) \)  \( \phi = 1 \) \( A_i = 1 \) \( \Theta_i = \log(\mu_i) \)

\[ V(\mu_i) = \mu_i \]

\( y_i \sim \text{Gamma}(\beta, \mu_i) \)  \( \phi = \frac{1}{\beta} \) \( A_i = 1 \) \( \Theta_i = -\frac{2}{\mu_i} \)

\[ V(\mu_i) = \mu_i^2 \]

\[
\text{var}(y_i) = \frac{\phi}{A_i} V(\mu_i) = \frac{\beta \mu_i^2}{\beta}
\]

\[
\frac{1}{\Gamma(\beta)} \left( \frac{\beta}{\mu} \right)^\beta y_i^{\beta-1} e^{-y_i \beta / \mu}
\]

\[
= \exp -\beta \frac{y_i}{\mu} + \beta \log(\beta / \mu) + (\beta-1) \log y_i
\]

\( y_i \sim \text{Bin}(m_i, p_i) \) then let \( z_i = y_i / m_i \)

\[
\ell(\beta, p_i) = \exp [m_i \log \frac{p_i}{1-p_i} + \log(1-p_i)]
\]

\[
= \exp [m_i \{ \log \frac{p_i}{1-p_i} + \log(1-p_i) \} + \log(m_i)]
\]

\[
(\beta \mu, p_i, m_i) = (\frac{m_i}{n} \mu, \frac{m_i}{n} \frac{1}{m_i} \mu_i, \mu_i)
\]

only so \( A_i = m_i, (\phi = 1) \)

\[ V(\mu_i) = \mu_i(1-\mu_i) \]  \( \Theta_i = \log \left( \frac{\mu_i}{1-\mu_i} \right) \]  \( \mu_i = \text{E}z_i = p_i \)
Inference: \( y_1, \ldots, y_n \) ind't from \( f(y; \beta) \)

Inference for \( \beta \) based on likelihood \( L(\beta; y) = \prod_{i=1}^{n} f(y_i; \beta) = \prod_{i=1}^{n} \exp \left[ -\frac{1}{\phi} \sum_{i=1}^{n} \left[ A: \{ y_i; \theta_i = x(\theta_i) \} + 2(y_i, \theta_i) \right] \right] \)

\( \ell(\beta; y) = \sum_i \left[ \right] \) log-lik.

Max. lik. estimate solves \( \ell'(\hat{\beta}; y) = 0 \) begins as long as \( -\ell''(\hat{\beta}) \) is positive semi-def.

For \( j = 1, \ldots, p \)

\[ \frac{\partial \ell}{\partial \beta_j} = \frac{1}{\phi} \sum_{i=1}^{n} A: \{ y_i - x'(\theta_i) \} \frac{\partial x'(\theta_i)}{\partial \beta_j} \]

Link

\[ \phi (\mu_i) = x_i^T \beta \quad \frac{\partial \phi (\mu_i)}{\partial \beta_j} = x_{ij} \quad \mu_i = x'(\theta_i) \]

\[ \frac{g'(x'(\theta_i)) x'(\theta_i)}{g'(\mu_i)} \frac{\partial \theta_i}{\partial \beta_j} = x_{ij} \]

\[ \frac{1}{\phi} \sum_{i=1}^{n} A: \{ y_i - \mu_i \} \frac{z_{ij}}{g'(\mu_i) v(\mu_i)} \]
At mle we have

\[ \sum_{i=1}^{n} A_{i} \left( y_{i} - \mu_{i}(\beta) \right) x_{ij} g'(\mu_{i}(\beta)) V(\mu_{i}(\beta)) = 0 \quad j = 1, \ldots, p \tag{**} \]

(Sol'n exists (usually) & determines max. (not min.)

- p. 62 of our HO or eq'n (4.1)
- hidden in Ch. 7 on p. 185 in eq' (7.4)
  text A: HO \( A_{i} = \frac{1}{a_{i}} \)
  \( l(\cdot) \)
  \( g(\cdot) \) for link.

If \( \theta_{i} = x_{i}^{T} \beta \quad g(\mu_{i}) = \theta_{i} \quad \theta'(\theta_{i}) = \mu_{i} \)

\[ \Rightarrow g'(\theta'(\theta_{i})) = \theta_{i} \]
\[ g'\left( \theta'(\theta_{i}) \right) \theta''(\theta_{i}) = 1 \]
\[ g'(\mu_{i}) V(\mu_{i}) = 1 \]

Then (**)

\[ \sum_{i=1}^{n} A_{i} E y_{i} - \mu_{i}(\beta) ^{3} \ x_{ij} = 0 \tag{7.4} \]

\[ \sum_{i=1}^{n} A_{i} \ E y_{i} \ x_{ij} = \sum_{i=1}^{n} A_{i} \mu_{i}(\beta) \ x_{ij} \quad j = 1, \ldots, p \]

"obs'd" "expected"
\[ \sum_{i=1}^{n} \frac{A_i \left( y_i - \mu_i(\beta) \right)^3 x_{ij}} {V(\mu_i(\beta)) g'\mu_i(\beta)} = 0 \quad j = 1, \ldots, p \]

Special case: \( V(\mu_i) = 1 \quad g'(\mu_i) = 1 \)

Let

**Weighted Least Squares**

Special case: \( V(\mu_i) = 1 \quad g'(\mu_i) = 1 \quad g(\mu_i) = x_i \beta \)

\[ \sum_{i=1}^{n} A_i \left( y_i - \mu_i(\beta) \right)^3 x_{ij} = 0 \]

**OLS**

\[ y = X\beta + \varepsilon \quad \varepsilon \sim (0, \sigma^2 I) \]

\[ \hat{\beta} = (X^TX)^{-1} X^T y \]

**Generalize** \( \varepsilon \sim (0, \sigma^2 V) \) \( V = \text{diag}(v_1, \ldots, v_n) \) known

\[ y^* = V^{-\frac{1}{2}} y \quad \varepsilon^* = V^{-\frac{1}{2}} \varepsilon \quad X^* = V^{-\frac{1}{2}} X \]

\[ V^{-\frac{1}{2}} y = V^{-\frac{1}{2}} X\beta + V^{-\frac{1}{2}} \varepsilon \]

\[ y^* = X^*\beta + \varepsilon^* \]

\[ \varepsilon^* \sim (0, \sigma^2 I) \]

\[ \hat{\beta} = (X^{*T}X^*)^{-1} X^{*T} y^* \]

\[ = (X^TV^{-1}X)^{-1} X^TV^{-1}y \quad \in \text{weighted LS} \]

\[ = (X^{T}WX)^{-1} X^TWy \quad W = V^{-1} \text{ weights} \]
You can show sol'n to (7.4) [easy and eqs] is a weighted LS sol:

\[ \sum A_i (y_i - \hat{x}_i) x_{ij} = 0 \quad j = 1, \ldots, p \]

Soln: \[ \hat{\beta} = (X'WX)^{-1} X'Wy \]

\[ W = \text{diag}(w_i) = \text{diag}(A_i) \]

\[ \min_{\hat{\beta}} \sum A_i (y_i - \hat{x}_i)^2 \quad \text{gives } \hat{x} \]

\[ \sum A_i \frac{(y_i - \hat{x}_i) x_{ij}}{\sqrt{\hat{x}_i} g'(\hat{x}_i)} = 0 \quad \text{iterative WLS scheme} \]

Alg.

\[ \hat{x}_i^{(0)} = x_i (\hat{x}_i^{(0)}) \]

step: \[ \hat{\eta}_i^{(t)} = \hat{\eta}_i^{(t)} + (y_i - \hat{x}_i^{(t)}) g'(\hat{x}_i^{(t)}) \]

\[ \hat{\beta}^{(t+1)} = (X' \hat{W}^{(t)} X)^{-1} X' \hat{W}^{(t)} \hat{\eta}^{(t)} \]

\[ E \tilde{x}_i^{(t)} = \mu_i^{(t)} \quad \text{var } \tilde{x}_i^{(t)} = \frac{\nu(\mu_i^{(t)}) g'(\mu_i^{(t)})^2}{\mu_i^{(t)} A_i} \]

new problem new \ y(t) new \ \hat{w}_i \ (\hat{w}) \text{ depends on } \hat{\beta} \]
glm (y ~ x_1 + x_2 \ldots + x_p, family = binomial)

y: either 2 cols (success failure)
or it can be a proportion \( r = \text{success/total} \)
but if it is a proportion need to input \( m \) as 'weight' option ...

cbind(r, m-r) ~ 1st way