Lecture 7: GLMs: Score equations, Residuals

Author: Nick Reich / Transcribed by Bing Miu and Yukun Li

Course: Categorical Data Analysis (BIOSTATS 743)

Likelihood Equations for GLMs

- The GLM likelihood function is given as follows:

$$\begin{aligned} L(\overrightarrow{\beta}) &= \sum_i log(f(y_i|\theta_i, \phi)) \\ &= \sum_i \left\{ \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + C(y_i, \phi) \right\} \\ &= \sum_i \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + \sum_i C(y_i, \phi) \end{aligned}$$

- ϕ is a dispersion parameter. Not indexed by i, assumed to be fixed
- θ_i contains β, from η_i
- $C(y_i, \phi)$ is from the random component.

Score Equations

- Taking the derivative of the log likelihood function, set it equal to 0

$$\frac{\partial L(\overrightarrow{\beta})}{\partial \beta_j} = \sum_i \frac{\partial L_i}{\partial \beta_j} = 0, \forall j$$

- Since $\frac{\partial L_i}{\partial \theta_i} = \frac{(y_i - \mu_i)}{a(\phi)}$, $\mu_i = b'(\theta_i)$, $Var(Y_i) = b''(\theta_i)a(\phi)$, and $\eta_i = \sum_j \beta_j x_{ij}$

$$\begin{aligned} 0 = \sum_i \frac{\partial L_i}{\partial \beta_j} &= \sum_i \frac{y_i - \mu_i}{a(\phi)} \frac{a(\phi)}{Var(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} \\ &= \sum_i \frac{(y_i - \mu_i) x_{ij}}{Var(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} \end{aligned}$$

- $V(\theta) = b''(\theta)$, $b''(\theta)$ is the variance function of the GLM.
- $\mu_i = E[Y_i|x_i] = g^{-1}(X_i\beta)$. These functions are typically non-linear with respect to β's, thus require iterative computation solutions.

Example: Score Equation from Binomial GLM (Ch5.5.1)

Y$\sim$ $Binomial(n_i, \pi_i)$

- The joint probability mass function:

$$\prod_{i=1}^{N} \pi(x_i)^{y_i}[1-\pi(x_i)]^{n_i-y_i}$$

- The log likelihood:

$$L(\beta) = \sum_j \Big(\sum_i y_i x_{ij}\Big)\beta_j - \sum_i n_i log\Big[1 + exp\Big(\sum_j \beta_j x_{ij}\Big)\Big]$$

- The score equation:

$$\frac{\partial L(\vec{\beta})}{\partial \beta_j} = \sum_i (y_i - n_i\hat{\pi}_i)x_{ij} \qquad \text{note that } \hat{\pi}_i = \frac{e^{X_i\beta}}{1+e^{X_i\beta}}$$

.

Asymptotic Covariance of $\hat{\beta}$:

- The likelihood function determines the asymptotic covariance of the ML estimate for $\hat{\beta}$.
- Given the information matrix, $\mathcal{I}$ with hj elements:

$$\mathcal{I} = E\Big[\frac{-\partial^2 L(\overrightarrow{\beta})}{\partial \beta_h \beta_j}\Big] = \sum_{i=1}^{N} \frac{x_{ih} x_{ij}}{Var(Y_i)} \Big(\frac{\partial \mu_i}{\partial \eta_i}\Big)^2$$

where w_i denotes

$$w_i = \frac{1}{Var(Y_i)} \Big(\frac{\partial \mu_i}{\partial \eta_i}\Big)^2$$

Asymptotic Covariance Matrix of $\hat{\beta}$:

- The information matrix, $\mathcal{I}$ is equivalent to: $\mathcal{I} = \sum_{i=1}^{N} x_{ih} x_{ij} w_i = X^T W X$
- W is a diagonal matrix with w_i as the diagonal element. In practice, W is evulated at $\hat{\beta}^{MLE}$ and depdent on the link function
- The square root of the main diagonal elements of $(X^T W X)^{-1}$ are estimated standard errors of $\hat{\beta}$

Analogous to SLR

	SLR	GLM
$Var(\hat{\beta}_i)$	$\frac{\hat{\sigma}^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$	the i^{th} main diagnal element of $(X^T W X)^{-1}$
$Cov(\hat{\beta}_i)$	$\hat{\sigma}^2(X^T X)^{-1}$	$(X^T W X)^{-1}$

Residual and Diagnostics

- Deviance Tests
  - Measure of goodness of fit in GLM based on likelihood
  - Most useful as a comparison between models (used as a screening method to identify important covariates)
  - Use the saturated model as a baseline for comparison with other model fits
  - For Poisson or binomial GLM: $D = -2[L(\hat{\mu}|y) - L(y|y)]$.
- Example of Deviance

Model	D($(y, \hat{\mu})$)
Gaussian	$\sum(Y_i - \hat{\mu}_i)^2$
Poisson	$2\sum(y_i \ln(\frac{y_i}{\hat{\mu}_i}) - (y_i - \hat{\mu}_i))$
Bionomial	$2\sum(y_i ln(\frac{y_i}{\hat{\mu}_i}) + (n_i - y_i) ln(\frac{n_i - y_i}{n_i - \hat{\mu}_i}))$

Deviance tests for nested models

- Consider two models, M_0 with fitted values $\hat{\mu}_0$ and M_1 with fitted values $\hat{\mu}_1$:
- M_0 is nested within M_1

$$\begin{aligned} \eta_1^{\mu_1} &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} \\ \eta_0^{\mu_0} &= \beta_0 + \beta_1 X_{11} \end{aligned}$$

- Simpler models have smaller log likelihood and larger deviance: $L(\hat{\mu}_0|y) \leq L(\hat{\mu}_1|y)$ and $D(y|\hat{\mu}_1) \leq D(y|\hat{\mu}_0)$.
- The likelihood-ratio statistic comparing the two models is the difference between the deviances.

$$\begin{aligned} &-2[L(\hat{\mu}_0|y) - L(\hat{\mu}_1|y)] \\ &\qquad = -2[L(\hat{\mu}_0|y) - L(y|y)] - \{-2[L(\hat{\mu}_1|y) - L(y|y)]\} \\ &\qquad = D(y|\hat{\mu}_0) - D(y|\hat{\mu}_1) \end{aligned}$$

Hypothesis test with differences in Deviance

- $H_0 : \beta_{i1} = ... = \beta_{ij} = 0$, fit a full and reduced model
- Hypothesis test with difference in deviance as test statistics. df is the number of parameter different between μ_1 and μ_0

$$D(y|\hat{\mu}_0) - D(y|\hat{\mu}_1) \sim \chi^2_{df}$$

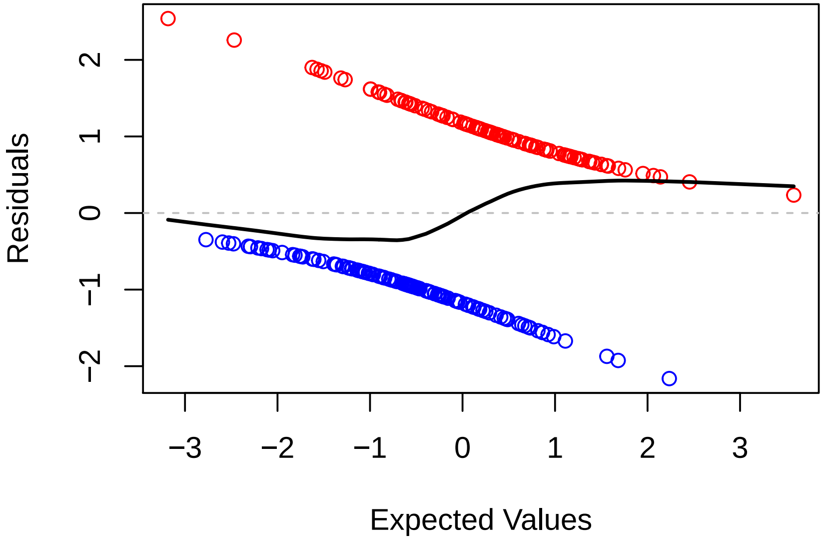
- Reject H_0 if the the chi-square calculated value is larger than $\chi^2_{df,1-\alpha}$, where df is the number of parameters difference between μ_0 and μ_1.

Residual Examinations

- Pearson residuals :
  $e_i^p = \frac{y-\hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$, where $\mu_i = g^{-1}(\eta_i) = g^{-1}(X_i\beta)$
- Deviance residuals :
  $e_i^d = sign(y_i - \hat{\mu}_i)\sqrt{d_i}$, where d_i is the deviance contribution of i_{th} obs. and $sign(x) = \begin{cases} 1 & x > 0 \\ -1 & x \leq 0 \end{cases}$
- Standardized residuals:
  $r_i = \frac{e_i}{\sqrt{(1-\widehat{h_i})}}$, where $e_i = \frac{y-\hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$, $\widehat{h_1}$ is the measure of leverage, and $r_i \cong N(0,1)$

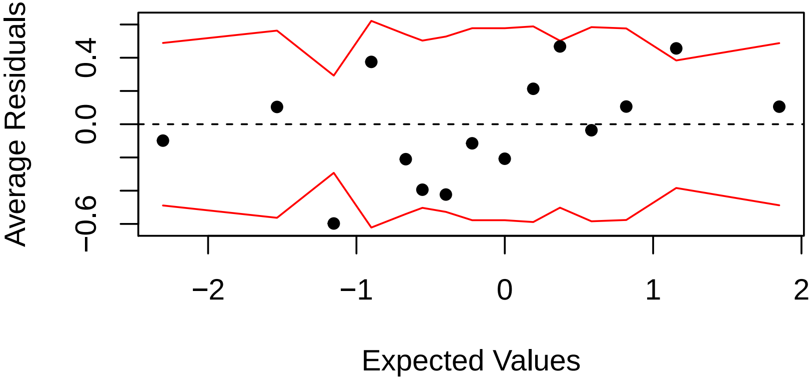
Residual Plot

Problem: Residual plot is hard to interpret for logistic regression



Binned Residual Plot

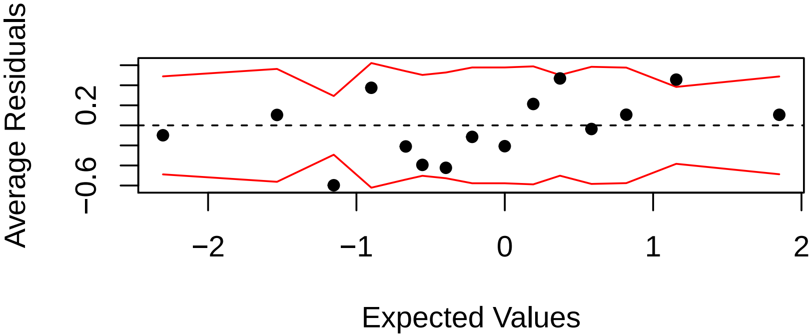
- Group observations into ordered groups (by $x_j, \hat{y}$ or x_{ij}), with equal number of observations per group.
- Compute group-wise average for raw residuals
- Plot the average residuals vs predicted value. Each dot represent a group.



Binned Residual Plot (Part 2)

- Red lines indicate ± 2 standard-error bounds, within which one would expect about 95% of the binned residuals to fall.
- R function avaiable.

```
linrary(arm)
binnedplot(x ,y, nclass...)
# x <- Expected values.  # y <- Residuals values.
# nclass <- Number of bins.
```



Binned Residual Plot (Part 3)

- In practice may need to fiddle with the number of observations per group. Default will take the value of nclass according to the n such that:
  - if n ≥ 100, $nclass = floor(sqrt(length(x)))$;
  - if $10 < n < 100$, $nclass = 10$;
  - if $n < 10$, $nclass = floor(n/2)$.

Ex: Binned Residual Plot with different bin sizes