Lecture 8: GLMs: Residuals, Newton-Raphson Algorithm, CIs

Author: Nick Reich / Transcribed by Yukun Li and Daveed Goldenberg

Course: Categorical Data Analysis (BIOSTATS 743)

Model Diagnostics via Residuals

- Example: Heart disease and blood pressure (Ch.6.2.2)
- Random sample of male residents in Framingham, MA aged 40-57. The response variable is whether they developed coronary heart disease during a six-year follow-up period.
- Let π_i be the probability of heart disease for blood pressure category i.
- Let x_i be the categories of systolic blood pressure

Model Diagnostics via Residuals

- Independent model (Blood Pressure is independent of Heart Disease)

$$logit(\pi_i) = \alpha$$

- Linear logit model (models association between Blood Pressure and Heart Disease)

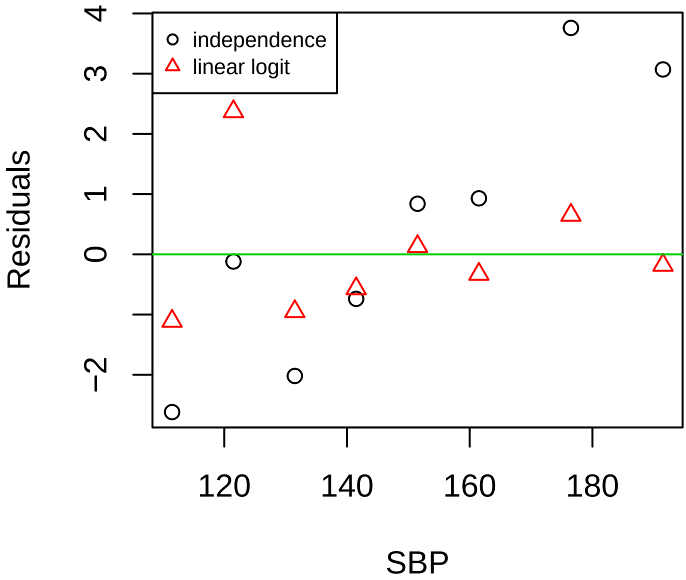
$$logit(\pi_i) = \alpha + \beta x_i$$

Model Diagnostics via Residuals

SBP	Independence	Linear Logit
$<$117	-2.62	-1.11
117-126	-0.12	2.37
127-136	-2.02	-0.95
137-146	-0.74	-0.57
147-156	0.84	0.13
157-166	0.93	-0.33
167-186	3.76	0.65
$>$186	3.07	-0.18

Model Diagnostics via Residuals

- Standardized residual plot of two models



Model Diagnostics via Residuals

- The Independent model has increasing residuals as blood pressure increases, a pattern like this breaks our assumption that residuals are normally distributed with mean of 0 and variance of 1.
- The residuals shows that the Independent model does not seem to be a good model
- However, the Linear Logit model appears to have no pattern throughout the residual plot and appears to be a good model

Newton Raphson Algorithm

- An iterative method for solving nonlinear equations
- General steps:
  - 1. initial guess for the solution
  - 2. approximate the function locallly and find maximum
  - 3. the maximum becomes the next guess
  - 4. repeat steps 2 and 3 until convergence
- Recall the solution to the estimating equations:

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}$$

$$\boldsymbol{\mu}^T = \left(\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_1}, ..., \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_p}\right)$$

$$\boldsymbol{H} = \left(\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_i \partial \beta_j}\right), \forall\ i, j = 1, 2, ..., p$$

- where $\boldsymbol{H}$ is Hessian matrix, also called observed information.

Newton Raphson Algorithm

- We are trying to maximize $L(\beta)$ through this iterative process
- starting with t $= 1$

$$\boldsymbol{\mu}^{(t)} = \boldsymbol{\mu}(\beta^{(t)})$$
$$\boldsymbol{H}^{(t)} = \boldsymbol{H}(\beta^{(t)})$$

- where $\beta^{(t)}$ is our t^{th} guess of β

Newton Raphson Algorithm

- Let $\boldsymbol{\mu}^{(t)}$ and $\boldsymbol{H}^{(t)}$ be $\boldsymbol{\mu}$ and $\boldsymbol{H}$ evaluated at $\beta^{(t)}$.
- According to Taylor series expansion,

$$L(\beta) \approx L(\beta^{(t)}) + \boldsymbol{\mu}^{(t)T}(\beta - \beta^{(t)}) + \frac{1}{2}(\beta - \beta^{(t)})^T \boldsymbol{H}^{(t)}(\beta - \beta^{(t)})$$

- Solve $\partial L(\beta)/\partial\beta \approx \mu^{(t)} + \boldsymbol{H}^{(t)}(\beta - \beta^{(t)}) = \mathbf{0}$, we get

$$\beta^{(t+1)} = \beta^{(t)} - (\boldsymbol{H}^{(t)})^{-1}\boldsymbol{\mu}^{(t)}$$

Fisher Scoring Method

- Fisher scoring is an alternative iterative method for solving likelihood equations.
- Use the expected value of Hessian matrix, called the expected information, denoted as $\boldsymbol{\mathcal{J}}$.

$$\boldsymbol{\mathcal{J}} = (-E(\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_i \partial \beta_j})) \; \forall \; i, j = 1, 2, ..., p$$

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - (\boldsymbol{\mathcal{J}}^{(t)})^{-1} \boldsymbol{\mu}^{(t)}$$

Inference for GLMs

- The curvature of our likelihood function determines the uncertainty/information about a parameter
- Classical/Frequentist inference assumes under certain regularity conditions, that parameters follows Multivariate Normal distribution centered at $\hat{\beta}^{MLE}$ with asymptotic covariance approximately by $\boldsymbol{H}^{-1}$ or $\boldsymbol{\mathcal{J}}$.
- for confidence interval:

$$\hat{\beta} \sim N(\hat{\beta}^{MLE},\ SE(\hat{\beta}^{MLE}))$$

- for hypothesis testing:

$$\hat{\beta}^{MLE} \sim N(\hat{\beta}_0,\ SE(\hat{\beta}^{MLE}))$$

Inference for GLMs

- Bayesian inference
  - Sample directly from posterior multivariate distribution and calculate the credible sets.
- Likelihood based inference
  - Similarly, directly calculate confidence intervals from likelihood function.

Inference for GLMs

- SE and CI for GLMs (Frequentist edition)
- 95% CI for β:

$$\hat{\beta} \pm 1.96 SE(\hat{\beta})$$

- 95% CI for $logit(\mu_i)$:

$$\begin{aligned}
logit(\mu_i) &= \alpha + \beta x_i \\
\Rightarrow Var(logit(\hat{\mu}_i)) &= Var(\hat{\alpha} + \hat{\beta} x_i) \\
&= Var(\hat{\alpha}) + x_i^2 Var(\hat{\beta}) + 2 Cov(\hat{\alpha}, \hat{\beta}) \\
\Rightarrow CI \ : \ logit(\hat{\mu}_i) &\pm 1.96 SE(logit(\hat{\mu}_i))
\end{aligned}$$

- We can get the 95% CI for μ_i with Delta method and the CI for $logit(\hat{\mu}_i)$.