

Logistic regression: Diagnostics

Patrick Breheny

April 12

Introduction

- After a model has been fit, it is wise to check the model to see how well it fits the data
- In linear regression, these diagnostics were build around residuals and the residual sum of squares
- In logistic regression (and all generalized linear models), there are two kinds of residuals (and thus, two kinds of residual sum of squares)

Pearson residuals

- The first kind is called the *Pearson residual*, and is based on the idea of subtracting off the mean and dividing by the standard deviation
- For a logistic regression model,

$$r_i = \frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)}}$$

- Note that, if we ignore the fact that $\hat{\pi}_i$ is an estimate based on y_i , then r_i has mean 0 and variance 1

Pearson residuals

- The preceding approach treats each observation as a single binary outcome
- An alternative approach is to treat all subjects with the same covariate pattern as a single observation, following a binomial distribution:

$$r_i = \frac{y_i - n_i \hat{\pi}_i}{\sqrt{n_i \hat{\pi}_i (1 - \hat{\pi}_i)}}$$

where n_i subjects share the i th covariate pattern, and y_i of them experience the event of interest

- The distinction between the two is important only if a number of subjects share the same covariate pattern (this will not come up if, for example, one of your explanatory variables is continuous)

Deviance residuals

- The other approach is based on the contribution of each point to the likelihood
- For logistic regression,

$$\ell = \sum_i \{y_i \log \hat{\pi}_i + (1 - y_i) \log(1 - \hat{\pi}_i)\}$$

- By analogy with linear regression, the terms should correspond to $-\frac{1}{2}r_i^2$; this suggests the following residual, called the *deviance residual*:

$$d_i = s_i \sqrt{-2 \{y_i \log \hat{\pi}_i + (1 - y_i) \log(1 - \hat{\pi}_i)\}},$$

where $s_i = 1$ if $y_i = 1$ and $s_i = -1$ if $y_i = 0$

Deviance residuals (cont'd)

- If we consider observations with the same covariate patterns to generate a single binomial response, the deviance residual becomes

$$d_i = s_i \sqrt{2 \left\{ y_i \log \frac{y_i}{n_i \hat{\pi}_i} + (n_i - y_i) \log \frac{n_i - y_i}{n_i - n_i \hat{\pi}_i} \right\}},$$

where $s_i = 1$ if $y_i > n_i \hat{\pi}_i$ and $s_i = -1$ if $y_i < n_i \hat{\pi}_i$

Deviance and Pearson's statistic

- Each of these types of residuals can be squared and added together to create an RSS-like statistic
- Combining the deviance residuals produces the *deviance*:

$$D = \sum d_i^2$$

which is, in other words, -2ℓ

- Combining the Pearson residuals produces the *Pearson statistic*:

$$X^2 = \sum r_i^2$$

Goodness of fit tests

- One might think that both statistics could be compared to the χ^2_{n-p} distribution as a rough goodness of fit test
- However, this test does not actually work very well
- Several modifications have been proposed, including an early test proposed by Hosmer and Lemeshow that remains popular and is available in SAS
- Other, better tests have been proposed as well (an extensive comparison was made by Hosmer *et al.* (1997))

The hat matrix for GLMs

- As you may recall, in linear regression it was important to divide by $\sqrt{1 - H_{ii}}$ to account for the leverage that a point had over its own fit
- Similar steps can be taken for logistic regression; here, the projection matrix is

$$\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2},$$

where $\mathbf{W}^{1/2}$ is the diagonal matrix with $\mathbf{W}_{ii}^{1/2} = \sqrt{w_i}$

Properties of the hat matrix

- In logistic regression, $\hat{\boldsymbol{\pi}} \neq \mathbf{H}\mathbf{y}$ – no matrix can satisfy this requirement, as logistic regression does not produce linear estimates
- However, it has many of the other properties that we associate with the linear regression projection matrix:
 - $\mathbf{r} = (\mathbf{I} - \mathbf{H})\mathbf{r}$
 - \mathbf{H} is symmetric
 - \mathbf{H} is idempotent
 - $\mathbf{H}\mathbf{W}^{1/2}\mathbf{X} = \mathbf{W}^{1/2}\mathbf{X}$ and $\mathbf{X}^T\mathbf{W}^{1/2}\mathbf{H} = \mathbf{X}^T\mathbf{W}^{1/2}$

where \mathbf{r} is the vector of Pearson residuals

Standardized residuals

- The diagonal elements of \mathbf{H} are again referred to as the *leverages*, and used to standardize the residuals:

$$r_{si} = \frac{r_i}{\sqrt{1 - H_{ii}}}$$
$$d_{si} = \frac{d_i}{\sqrt{1 - H_{ii}}}$$

- Generally speaking, the standardized deviance residuals tend to be preferable because they are more symmetric than the standardized Pearson residuals, but both are commonly used

Leave-one-out diagnostics

- You may recall that in linear regression there were a number of diagnostic measures based on the idea of leaving observation i out, refitting the model, and seeing how various things changed (residuals, coefficient estimates, fitted values)
- You may also recall that for linear regression, it was not actually necessary to refit the model n times; explicit shortcuts based on \mathbf{H} were available
- The same idea can be extended to generalized linear models, although we cannot take advantage of the explicit-solution shortcuts without making approximations

One-step approximations

- The resulting approximate statistics are said to be *one-step approximations* to the true values
- The issue is that we can quickly calculate the one-step approximations based on the current weights $\{w_i\}$ without refitting anything, but to calculate the exact value, we would need to go through n IRLS algorithms
- The approximations are usually pretty good, although if one point has a very large influence, then the approximation may be quite different from the true value

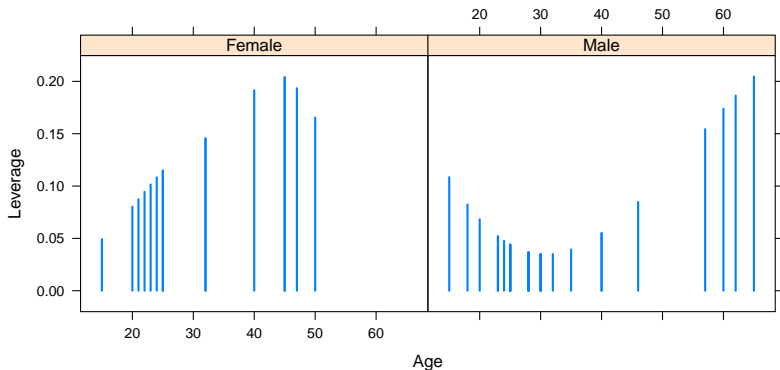
One-step approximations

One-step approximations allow us to quickly calculate the following diagnostic statistics for GLMs:

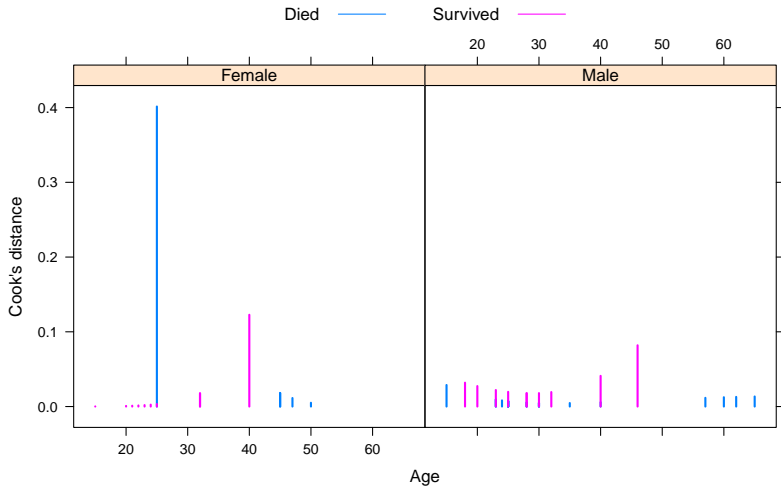
- Studentized deleted residuals
- Δ_{β} (for assessing the change in individual coefficients)
- Cook's distance (for assessing overall influence over the model fit)

Leverage

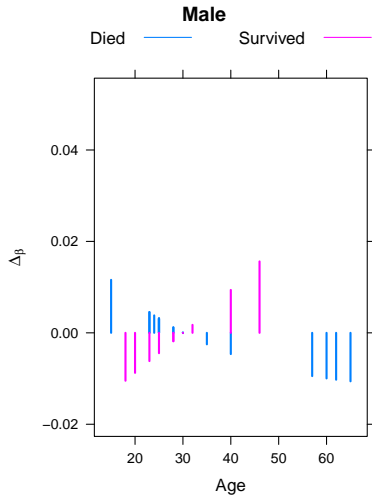
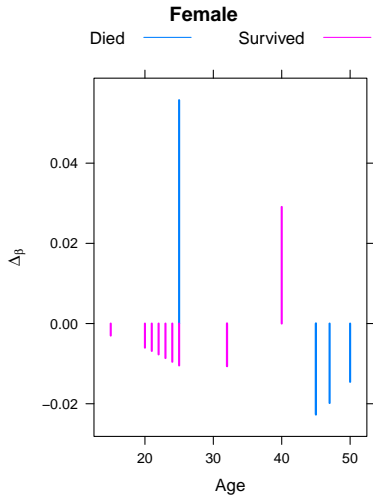
To get a sense of the information these statistics convey, let's look at various plots of the Donner party data, starting with leverage:



Cook's Distance



Delta-beta (for effect of age)



Residuals / proportional influence

