Lecture 39

Logistic Regression - Diagnostics

This lecture will discuss various diagnostic tools that apply to logistic regression without repeats - that is, where there are not multiple responses from each combination of predictor variables.

Analog of the Nested Models F-test: Likelihood Ratio Test

When assessing whether a smaller model fits better than a full model, the appropriate analog of the nested models F-test for least squares regression is the likelihood ratio test.

$$\chi^2 = 2 \log \frac{L(\text{Full model})}{L(\text{Reduced model})} = 2\left(\log L(\text{Full model}) - \log L(\text{Reduced model})\right)$$

This test has approximately a Chi-Square distribution and the degrees of freedom are calculated as the difference in the number of parameters in the two models. The log likelihood values are given as part of the Minitab output when you run a logistic regression.

For instance, to test whether a model with sex alone fits the Donner party data as well as the full model with age and sex and the interaction term does, we run the two logistic regressions and get a log likelihood for the larger model of -23.673 and a log likelihood for the smaller model of -28.643. There are 2 more parameters in the larger model (the slopes for age and age by male interaction) than in the smaller model. So the likelihood ratio test is

$$\chi^2 = 2(-23.673 - (-28.643)) = 9.94$$

and with two degrees of freedom, the p-value is approximately 0.007. Thus the full model fits significantly better than the reduced one.

Goodness of Fit Tests

Without repeated responses from each (or most) combination of predictor variable values, neither the Deviance nor the Pearson goodness of fit test is meaningful. The only meaningful test is Hosmer-Lemeshow test that groups the data into categories and uses them to assess whether the model fits the observed responses. This is not a very powerful test, but it is the only relevant one for the Donner party data.

Other Measures of Fit

Suppose you have N responses in your data. Two distinct responses are concordant if the responses are different and the response with the event 1 has a higher predicted probability of occurring than the non-response 0. They are discordant if the responses and different and the response with the event 1 has a lower probability of occurring than the non-response 0. They are uninformative if they have the same response (both are 1 or both are 0). And they are tied if they have the same probabilities of the event occurring.

There are $\binom{N}{2} = N(N-1)/2$ total possible pairs. If N_0 is the number with response zero and N_1 is the number with response one, then there are $N_0 \times N_1$ informative pairs. Let n_c by the number of concordant pairs, n_d be the number of discordant pairs, and n_t be the number of ties so that $n_c + n_d + n_t = N_0 \times N_1$. Models with more concordant pairs are better at predicting new responses. Minitab spits out a lot of summary statistics based on these notions, but does not give any distributional help for them and does not provide confidence intervals for them. Besides the raw numbers, the statistics include:

$$\text{Sommer's D} = \frac{n_c - n_d}{n_c + n_d + n_t}$$

$$\text{Goodman-Kruskal Gamma} = \frac{n_c - n_d}{n_c + n_d}$$

and

$$\text{Kendall's Tau-a} = \frac{n_c - n_d}{N(N-1)/2}$$

References and Readings

Exercises for Lecture 39

1. –

2. –