Cornell University
Cornell Statistical Consulting Unit

StatNews #82

Bias Adjustment in Logistic Regression Models

February 2012

In logistic regression models, if the sample size is small or if a predictor is strongly associated with one of the possible outcomes (say, 0 or 1), the estimated coefficients may be biased. A similar problem occurs in contingency tables when the sample size is small or when too many cells in the table have low counts (see StatNews #01, “Sparse Contingency Tables”). This is known as “***the phenomenon of separation***”.

One way to address the separation problem is to use “exact tests” (see StatNews #01). However, in the case of logistic regression exact methods are only practical in very simple cases, where there is just one predictor. When additional (continuous) predictors are of interest, these methods are computationally intensive in terms of memory requirements, and are rendered impractical. Furthermore, exact methods adjust only the p-value, not the parameter estimates. An alternative approach is to use ***Firth’s bias-adjusted estimates*** [1]. In addition to being computationally efficient, Firth’s method, which maximizes a *penalized* likelihood function, also guarantees that the parameter estimates will be finite (as opposed to the standard maximum likelihood estimates).

We apply Firth’s method to the famous “O-rings” data set, and compare the estimates with the ones obtained via the standard methods. This data set contains the failures of O-rings in launches of space shuttles (one such failure proved fatal for the crew of the Challenger space shuttle). There are 23 observations, which include the O-ring failure status (1 for failure, 0 for no failure) and the temperature in Fahrenheit at the time of the launch. The data are summarized as follows:

Temperature in cases of failure: 70, 57, 63, 70, 53, 75, 58

Temperature in cases of no-failure: 66, 69, 68, 67, 72, 73, 70, 78, 67, 67, 75, 70, 81, 76, 79, 76.

We suspect that problem of separation might affect the estimates, since all launches that took place when the temperature was below 66^{o}F resulted in failed O-rings.

The results for the parameter of interest (temperature) are summarized in the following table:

	Temperature		
	Estimate	Std.Err.	P-value
Standard logistic	-0.232	0.108	0.032
Firth’s bias adjustment	-0.186	0.092	0.009

The Firth-adjusted parameter estimate is smaller in absolute value, but so is the standard error, and hence the adjusted p-value is quite a bit smaller. Although in this case we conclude that temperature is significant at the 5% level regardless of the estimation method, in general, the conclusion based on the standard logistic regression method may be different than that obtained from the bias-adjustment version.

Currently, JMP, SAS, and R allow the user to obtain Firth's bias-adjusted estimates in logistic regression models. In **SAS version 9.2** this is done via the option FIRTH in the PROC LOGISTIC statement. For conditional logistic regression, a macro is available at http://www.meduniwien.ac.at/cemsiis/kb/wf/software/statistische-software/fllogistf/.

To obtain Firth's bias-adjusted estimates in **JMP 9.0.0**, select "Analyze" in the main menu, then choose Fit Model, and in the pop-up window select the "Generalized Linear Model" personality. Click on the checkbox labeled "Firth Bias-adjusted Estimates" before running your model.[1]

For **R** users, the **logistf** package computes Firth's bias-adjusted estimates.

As always if you would like assistance with this topic or any other statistical consulting question feel free to contact statistical consultants at CSCU.

Author: Haim Bar (hyb2@cornell.edu)

(This newsletter was distributed by the Cornell Statistical Consulting Unit. Please forward it to any interested colleagues, students, and research staff. Anyone not receiving this newsletter who would like to be added to the mailing list for future newsletters should contact us at cscu@cornell.edu. Information about the Cornell Statistical Consulting Unit and copies of previous newsletters can be obtained at http://www.cscu.cornell.edu).

[1] Firth, D. (1993): "Bias reduction of maximum likelihood estimates", Biometrika 80, 27 - 38

[1] CSCU has found a bug in the "Whole Model Test" with the Firth bias adjustment option in JMP9.0.0. This bug will be fixed in the next version.