

Exact logistic regression: an extension of Barnard's approach for continuous covariates. *

W. DANA FLANDERS[†] AND CHIRANJEEV DASH[‡]

Abstract

Logistic regression is commonly used to analyze observational studies when the outcome (dependent variable) is dichotomous. For studies with a large sample size or relatively few parameters, unconditional maximum likelihood is the preferred method for parameter estimation. However, for unbalanced studies, matched studies and those with many parameters, unconditional maximum likelihood methods are not recommended; instead, exact methods are preferred. Two major methodologies are available for comparing binomial proportions in the context of contingency tables, those based on a conditional (e.g., Fisher's exact test) or on an unconditional (Barnard [1]) approach.

Here we propose a generalization of the unconditional approach which allows calculation of exact p -values for logistic regression. The method, as implemented here, is actually conditional on the observed number of cases, but not on the sufficient statistics for parameters other than the intercept. We prove that the test size does not exceed the nominal level under the null hypothesis of no disease-exposure association, when controlling for additional covariates some of which may be continuous and assuming the logistic model is appropriate.

A small simulation study suggests that this unconditional approach can, in at least some circumstances, have better power than the corresponding exact conditional

**Received:* February 8, 2008; *Accepted:* June 18, 2008.

Key words and phrases: logistic regression.

[†]*Mailing Address:* Rollins School of Public Health, Emory University, Departments of Epidemiology and Biostatistics, Atlanta, GA 30345. *E-mail:* flanders@sph.emory.edu.

[‡]*Mailing Address:* Emory University, Department of Epidemiology, Atlanta, GA 30345.

logistic regression. In one example, we tried to implement the approach based on just a sample of the potential outcomes. This modification poorly approximated the exact value, so we cannot recommend implementation based on a sample, even in the context of many potential outcomes. In the discussion, we note the possibilities of generalizations to calculation of confidence intervals, and to an approach that incorporates features of both the conditional and unconditional approaches.

REFERENCES

- [1] Barnard, G. A. (1947). Significance tests for 2×2 tables. *Biometrika* **34**(1/2) 123–138.
- [2] Fisher, R. (1947). The logic of inductive inference. *J. Royal Statist. Soc. Ser. A* **98** 39–54.
- [3] Greenland, S. (1995) Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology* **6** (4)356–65.
- [4] Heinze, G. A. (2006). A comparative investigation of methods for logistic regression with separated or nearly separated data. *Stat Med.* **25** 4216–4226.
- [5] Kleinbaum, D. G. and Klein, M. (1994). *Logistic Regression: a Self-Learning Text*. Springer-Verlag, New York.
- [6] Koch, H. F. and Hothorn, L. A. (1999). Exact unconditional distributions for dichotomous data in many-to-one comparisons. *J. Statist. Plan. Infer.* **82** 83–99.
- [7] Mato, A. S and Andres, A. M. (1997). Simplifying the calculation of the p -value for Barnard’s test and its derivatives. *Statistics and Computing* **7** 137–143.
- [8] Mehta C. R. and Patel, N. R. (1995). Exact logistic regression: theory and examples. *Statistics in Medicine* **14** (19) 2143–2160.
- [9] Mehta, C. R.; Patel, N. R. and Senchaudhuri, P. (2000). Efficient Monte Carlo methods for conditional logistic regression. *Journal of the American Statistical Associatio* **95** (449) 99–108.
- [10] Potter, D. M. (2005). A permutation test for inference in logistic regression with small- and moderate-sized data sets. *Statist. in Med.* **24** 693–708.
- [11] SAS, Version 9. 1. The SAS Institute, 2004, Cary, NC.