Biostatistics (2008), **9**, 4, *pp.* 777–778
doi:10.1093/biostatistics/kxn026
Advance Access publication on August 20, 2008

Letter to the editor

ANTON K. FORMANN

Faculty of Psychology, University of Vienna, Liebiggasse 5, A-1010 Vienna, Austria
anton.formann@univie.ac.at

DANKMAR BÖHNING*

Quantitative Biology and Applied Statistics, School of Biological Sciences, University of Reading, Whiteknights, Reading RG6 6FN, UK

A recent contribution by Pepe and Janes (2007) contains some errors which we summarize here as the following 4 comments.

First, as can be easily seen, the formula (3.3) on page 476 (also occurring in Appendix A on page 482, very last formula) is incorrect.

The correct formula for the prevalence estimate is achieved when the plus sign is replaced by a minus sign under the square root: $\hat{\rho} = 1/2 \pm \sqrt{1/4 - 1/(4 + \hat{V}^2)}$. This is a well-known result which can be found in the book by Lazarsfeld and Henry (1968, p 42) as Theorem 2.

Second, the formulas for the true- and false-positive rates, denoted by ϕ_k and ψ_k in the article by Pepe and Janes (2007), contain an error in the term C_k of the size of the multiplicative factor $1/p_k^2$. The correct formula for C_k is $C_k = (p_{kj} - p_k p_j)(p_{k\ell} - p_k p_\ell)/(p_{j\ell} - p_j p_\ell)$. Note that an equivalent method of parameter estimation was already described by Lazarsfeld (1950a,b).

Third, a problem with these old methods of parameter estimation in latent class models is that parameter estimates may lie outside the interval [0, 1]. Lazarsfeld and Henry (1968, p 33) give such a numerical example. From $\hat{p}_1 = 0.38$, $\hat{p}_2 = 0.40$, $\hat{p}_3 = 0.26$, $\hat{p}_{12} = 0.264$, $\hat{p}_{13} = 0.166$, $\hat{p}_{23} = 0.128$, and $\hat{p}_{123} = 0.1248$, the following estimates were calculated: $\hat{\rho} = 0.2$, $\hat{\phi}_1 = 1.5$, $\hat{\phi}_2 = 0.8$, $\hat{\phi}_3 = 0.5$, $\hat{\psi}_1 = 0.1$, $\hat{\psi}_2 = 0.3$, and $\hat{\psi}_3 = 0.2$. As a consequence of these complications, Goodman (1974) developed a maximum likelihood (ML)–based algorithm forcing the parameter estimates to lie within the admissible interval [0, 1]. Applying this iterative proportional fitting algorithm (later identified as a special case of the expectation–maximization [EM] algorithm) to the above-mentioned example results in the parameter estimates $\hat{\rho} = 0.321$, $\hat{\phi}_1 = 1.$, $\hat{\phi}_2 = 0.782$, $\hat{\phi}_3 = 0.489$, $\hat{\psi}_1 = 0.087$, $\hat{\psi}_2 = 0.219$, and $\hat{\psi}_3 = 0.152$ and the likelihood-ratio statistic 30.01 on df $= 0$, assuming $N = 1000$. Hence, the appropriate ML algorithm finds a so-called boundary solution for $\hat{\phi}_1$. In contrast, Pepe and Janes (2007) claim in part 4 of their paper (see also their Appendix B) that the estimates given in part 3 are identical to the ML estimates conventionally found by the EM algorithm.

Our fourth and final remark refers to the audiology data analyzed by Pepe and Janes (2007, p 479), more precisely to the comparison of the latent class estimates with their analogs obtained from a gold standard measure of disease status. This comparison indicated "that the tests are substantially worse than the latent class analysis suggests," a finding that is not really surprising: As Uebersax (1988) pointed out,

*To whom correspondence should be addressed.

the latent probabilities of the positive (= diseased) class give upper limits of the symptoms sensitivities and one minus the latent probabilities of the negative (= not diseased) class give upper limits of their specificities, since the individuals cannot be assigned deterministically to the latent classes but only in terms of their class membership probabilities.

ACKNOWLEDGMENT

Conflict of Interest: None declared.

REFERENCES

GOODMAN, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61**, 215–231.

LAZARSFELD, P. F. (1950a). The logical and mathematical foundation of latent structure analysis. In: Stouffer, S. A. *and others* (editors), *Measurement and Prediction*. Princeton: Princeton University Press, pp. 362–412.

LAZARSFELD, P. F. (1950b). The interpretation and computation of some latent structures. In: Stouffer, S. A. *and others* (editors), *Measurement and Prediction*. Princeton: Princeton University Press, pp. 413–472.

LAZARSFELD, P. F. AND HENRY, N. W. (1968). *Latent Structure Analysis.* Boston: Houghton Mifflin.

PEPE, M. S. AND JANES, H. (2007). Insights into latent class analysis of diagnostic test performance. *Biostatistics* **8**, 474–484.

UEBERSAX, J. S. (1988). Validity inferences from interobserver agreement. *Psychological Bulletin* **104**, 405–416.

[*Received February 25, 2008; revision March 14, 2008; accepted for publication April 4, 2008*]