DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 8 | Issue 2 Article 24

11-1-2009

A Maximum Test for the Analysis of Ordered Categorical Data

Markus Neuhäeuser
Koblenz University of Applied Sciences, Remagen, Germany, neuhaeuser@rheinahrcampus.de

Journal of Modern Applied Statistical Methods
November 2009, Vol. 8, No. 2, 592-593

1538 – 9472/09/$95.00

BRIEF REPORTS
A Maximum Test for the Analysis of Ordered Categorical Data

Markus Neuhäeuser
Koblenz University of Applied Sciences
RheinAhrCampus, Remagen, Germany

Different scoring schemes are possible when performing exact tests using scores on ordered categorical data. The standard scheme is based on integer scores, but non-integer scores were proposed to increase power (Ivanova & Berger, 2001). However, different non-integer scores exist and the question arises as to which of the non-integer schemes should be chosen. To solve this problem, a maximum test is proposed. To be precise, the maximum of the competing statistics is used as the new test statistic, rather than arbitrarily choosing one single test statistic.

Key words: Exact test, maximum test, ordered categorical data, scores.

Introduction

Ordered categorical data occur often in various applications. For example, Gregoire & Driver (1987) pointed out that such ordinal data frequently result from questionnaire surveys in behavioral science investigations. Sheu (2002) noted that ordered categorical variables play an important role in psychological studies because precise measurement is not always possible. Hence, Likert scales are frequently used in psychological research (Rasmussen, 1989). Moreover, ordered categorical data can be found in medical studies (Rabbee, et al., 2003).

When performing exact tests using scores on ordered categorical data, different scoring schemes are possible. In case of three categories the standard scheme is v_1 = (0 0.5 1); because this scheme corresponds to (0 1 2) it is called integer scoring. Ivanova & Berger (2001) proposed non-integer scores: the middle score should be changed to either 0.49 or 0.51 in order to increase the power.

Senn (2007) criticized these non-integer scores because "there is no substantial reason either in terms of likelihood under an alternative hypothesis or on the basis of some other appeal to logic or experience" (p. 297) to replace the standard scheme. Berger (2007) replied that the standard scheme is also arbitrary in case of ordered categorical data and, therefore, the increased power is a rationale for choosing a non-integer scoring scheme. However, the question is which of the two non-integer schemes should be chosen? Berger wrote: "The existence of two viable replacements creates this controversy... If it helps at all, then always shrink to 0, and use only 0.49" (Berger, 2007, p. 299). The latter proposal is arbitrary and may therefore be regarded as unacceptable. However, using the less powerful test with integer scores may also be regarded as unacceptable. Is there an alternative?

In some areas, statistical genetics for example, it is common to apply a maximum test. That is, the maximum of several competing test statistics is used as a new statistic, and the permutation distribution of the maximum is used for inference (Neuhäuser & Hothorn, 2006). Thus, an alternative is using the maximum of the competing statistics as the new test statistic, rather than arbitrarily choosing one single test statistic. Thus, in the case of three categories with the three scoring schemes v_1 = (0 0.5 1), v_2

Markus Neuhäeuser is Professor of Statistics in the Department of Mathematics and Technology. Email him at: neuhaeuser@rheinahrcampus.de.

$= (0\ 0.49\ 1)$, and $v_3 = (0\ 0.51\ 1)$ the test is performed with the statistic

$$T_{\max} = \max_{i=1,2,3} S(C, v_i)$$

where

$$S(C, v_i) = \frac{\sum_{j=1}^{3} v_{ij} C_{1j}}{n_1} - \frac{\sum_{j=1}^{3} v_{ij} C_{2j}}{n_2}$$

are the individual test statistics with scores $v_i = (v_{i1}, v_{i2}, v_{i3})$, C_{ij} are the frequencies for ordered category j ($j = 1, 2, 3$) in group i ($i = 1, 2$), and n_i is the sample size of group i.

This maximum test has the advantage of a less discrete null distribution and an accompanied increased power as the single tests based on the non-integer scores. The following example was considered by Ivanova & Berger (2001) and discussed by Senn (2007): $C_{11} = 7$, $C_{12} = 3$, $C_{13} = 2$, $C_{21} = 18$, $C_{22} = 4$, $C_{23} = 14$.

In case of this example, the maximum test gives a significant result for the table (C_{11}, C_{12}) = (9, 1), as the scheme v_3 does, in contrast to v_1. For all 76 possible tables with the margins of this example the maximum test's p-value is at least as small as the p-value of the test with the standard scheme. To be precise, the two p-values are identical for 25 tables, but for 51 tables the maximum test's p-value is smaller. Moreover, the maximum test's p-value is always smaller than or equal to the bigger one of the two p-values of the non-integer scoring tests; note that in this example $\max_{i=2,3} S(C, v_i)$ results in an identical test as T_{max}.

Conclusion

The maximum test is a compromise that avoids the arbitrary choice of just one scheme and maintains the advantage of the non-integer scores. Note that the maximum test is not complicated. Because the exact permutation null distribution of T_{max} is used for inference, one does not need to know the correlation between the different $S(C, v_i)$. Thus, when a researcher selects a test based on the trade-off between power and simplicity – as suggested by Ivanova & Berger (2001) – the maximum test is a reasonable choice. In addition, the approach may have some appeal to logic: there is more than one possible test statistic, so combine the competing statistics. Recently, it was shown that a maximum test can be regarded as an adaptive test with the test statistics themselves as selectors (Neuhäuser & Hothorn, 2006). Thus, in a maximum test the data decide and the statistician does not need to arbitrarily choose between different tests.

Note that a multitude of alternative tests applicable to ordered categorical data exist (Liu & Agresti, 2005). However, the discussed score tests as well as the proposed maximum tests have – at a minimum – the advantage that they are easy to apply.

References

Berger, V. W. (2007). [Reply to Senn (2007).] *Biometrics*, 63, 298-299.

Gregoire, T. G., & Driver, B. L. (1987). Analysis of ordinal data to detect population differences. *Psychological Bulletin*, 101, 159-165.

Ivanova, A., & Berger, V. W. (2001). Drawbacks to integer scoring for ordered categorical data. *Biometrics*, 57, 567-570.

Liu, I., & Agresti, A. (2005). The analysis of ordered categorical data: An overview and a survey of recent developments. *Test*, 14, 1-73.

Neuhäuser, M., & Hothorn, L. A. (2006). Maximum tests are adaptive permutation tests. *Journal of Modern Applied Statistical Methods*, 5, 317-322.

Rabbee, N., Coull, B. A., Mehta, C., Patel, N., & Senchaudhuri, P. (2003). Power and sample size for ordered categorical data. *Statistical Methods in Medical Research*, 12, 73-84.

Rasmussen, J. L. (1989). Analysis of Likert-scale data: a reinterpretation of Gregoire and Driver. *Psychological Bulletin*, 105, 167-170.

Senn, S. (2007). Drawbacks to non-integer scoring for ordered categorical data. *Biometrics*, 63, 296-298.

Sheu, C.-F. (2002). Fitting mixed-effects models for repeated ordinal outcomes with the NLMIXED procedure. *Behavior Research Methods, Instruments, & Computers*, 34, 151-157.