CTA *vs*. Chi-Square: Comparing Voter Sentiment in Political Wards

Paul R. Yarnold, Ph.D.

Optimal Data Analysis, LLC

Random samples of 200 registered voters from each of four political wards were asked if they favored a particular candidate.[1] Seven chi-square analyses (one omnibus comparison between all four wards, six follow-up pair-wise comparisons to specify the underlying effect[2]) were used to compare the proportion of voters favoring the candidate between wards. Evaluating results at either the generalized or the experimentwise criterion for statistical significance, chi-square found the omnibus effect ($p<0.013$), and two pairwise comparisons: Ward 1>Ward 2 ($p<0.0025$), and Ward 1>Ward 4 ($p<0.014$). In contrast, a single CTA analysis was conducted predicting voter sentiment (treated as the class variable, and coded as 1 if the voter favors the candidate, or 0 otherwise) with ward (dummy-coded as 1-4) treated as a multicategorical attribute.[3-9] A single model emerged: if Ward=1, then predict the voter favors the candidate; otherwise predict the voter does not favor the candidate ($p<0.036$). The training (total sample) effect was relatively weak (ESS=10.2), and the predictive accuracy declined to levels worse than expected by chance (ESS= -14.8) in jackknife analysis. CTA thus revealed that the most accurate model possible for this application is weak, and that there is evidence that the model may not cross-generalize if it is used to classify independent random samples.

Data investigated herein are given in Table 1.

Table 1: Study Data[1]

Favor Candidate	Ward 1	Ward 2	Ward 3	Ward 4
Yes	76	53	59	48
No	124	147	141	152

In training (total sample) analysis the CTA model correctly classified 440/564 (78.0%) of the voters opposing the candidate, and 76/236 (32.2%) of the voters supporting the candidate. However, in jackknife analysis model sensitivity fell to 299/564 (53.0%) for voters who were against the candidate. ESS cannot ordinarily be computed for findings for multicategorical designs that are based on chi-square, and can never be computed unless the findings are completely specified. For example, here chi-square finds Ward 1>Ward 2 and Ward 1> Ward 3: how do Ward 2 and Ward 3 compare, and what about Ward 4?

References

[1]Mendenhall W, Reinmuth JE (1974). *Statistics for management and economics* (2nd Ed.). North Scituate, MA: Duxbury Press (pp. 482-483).

[2]Grimm LG, Yarnold PR (1995). *Reading and Understanding Multivariate Statistics*. Washington, DC: APA Books.

[3]Yarnold PR (1996). Discriminating geriatric and non-geriatric patients using functional status information: An example of classification tree analysis via UniODA. *Educational and Psychological Measurement*, *56*, 656-667.

[4]Soltysik RC, Yarnold PR (2010). Automated CTA software: Fundamental concepts and control commands. *Optimal Data Analysis, 1*, 144-160.

[5]Yarnold PR (2013). Univariate and multivariate analysis of categorical attributes with many response categories. *Optimal Data Analysis*, *2*, 177-190.

[6]Yarnold PR, Bryant FB (2015). Obtaining an enumerated CTA model via automated CTA software. *Optimal Data Analysis*, *4*, 54-60.

[7]Yarnold PR (2015). UniODA *vs*. chi-square: Voter sentiment and political ward. *Optimal Data Analysis*, *4*, 161-162.

[8]Linden A, Yarnold PR (2016). Using data mining techniques to characterize participation in observational studies. *Journal of Evaluation in Clinical Practice*. DOI: 10.1111/jep.12515

[8]Yarnold PR, Soltysik RC (2016). *Maximizing predictive accuracy*. Chicago, IL: ODA Books. DOI: 10.13140/RG.2.1.1368.3286

Author Notes

Mail: Optimal Data Analysis, LLC
      6348 N. Milwaukee Ave., #163
      Chicago, IL 60646