Optimal Data Analysis
Vol. 8 (January 13, 2019), 11-14

2155-0182/10/$3.00

Value-Added by ODA *vs*. Chi-Square

Paul R. Yarnold, Ph.D.
Optimal Data Analysis, LLC

Beyond identifying the *most accurate* classification model which exists for the sample, and estimating *cross-generalizability* vis-à-vis jackknife, hold-out and/or other validity methods, ODA provides the *exact* one- or two-tailed *P*-value, the *sensitivity* and *predictive value* for each category of the class variable, and the *effect strength* corrected for chance.

Contrasts of inherent theoretical limitations of chi-square analysis *vs*. corresponding strengths of ODA and CTA are well documented.[1-23] It is likewise imperative to emphasize that ODA and CTA are also the analytic methods of choice in seemingly straightforward applications—for example, if data for a truly qualitative attribute satisfy the minimum expectation criterion and one obvious underlying solution exists.[24-27]

This note illustrates the value added by using ODA in analysis for an example of the latter case. For example consider a recent study of gully erosion, in which a complex sampling and interviewing methodology yielded the data in Table 1 on type of adjustment made by the respondent, and if adjustment was individually- or community-motivated.[28] Chi-square=265.0 (df=3, n=720, $p<0.0001$) finds a statistically significant relationship between adjustment and motivation. Chi-square critical value for $p<0.05$ in this application is 7.82: every *cell* in the table contributes a value greater than this to omnibus chi-square. It was concluded: "local population driven by self-responsibility relies on individual as well as community-led adjustment measures to ensure their continued survival and habitation in their ancestral homes" (p. 6).

Table 1: Individual- *vs*. Community-Led Adjustments in Response to Gully Erosion[28]

Adjustment	Individual	Community
Use of Ridges	85	173
Shifting Habitation	65	170
Relocation	172	10
Intensified Cultivation	45	0

MegaODA software[29-32] was employed to discriminate source of motivation (individual *vs*. community, treated as a class variable) as a function of type of adjustment (multicategorical attribute). The resulting ODA model was:

If Adjustment=Use of Ridges or Shifting Habitation, *then predict* Motivation= Community-Led;

Otherwise

If Adjustment=Relocation or Intensified Cultivation, *then predict* Motivation= Individual-Led.

Using the ODA model to classify the sample data yielded the confusion matrix presented in Table 2.

2155-0182/10/$3.00

Table 2: Confusion Matrix for ODA Model, for Training and LOO Results (Sens=Sensitivity; PV=Predictive Value)

Actual Motivation	Predicted Motivation: *Individual*	*Community*	Sens
Individual	217	150	59.1
Community	10	343	97.2
PV	95.6	69.6	

Because 50% correct classification of each motivation category is expected by chance, model classification of community-motivated adjustments is (97.2/50 – 1) x 100% or 94.4% greater than expected by chance, and classification of individually-motivated adjustments is 18.2% greater than expected by chance: this is greater than a 5-fold difference in accuracy. The finding is thus dominated by the near-uniform response to community-based adjustments.

The Effect Strength for Sensitivity or ESS index of classification accuracy, computed as a function of mean sensitivity achieved by the model across class categories, is used to norm classification accuracy: ESS=0 is the level of accuracy expected by chance, ESS=100 is perfect errorless classification, and ESS<0 is accuracy worse than expected by chance.[1,26,27] Here, ESS=56.3. The rule-of-thumb used to qualitatively assess classification accuracy achieved by a model was based on billions of Monte Carlo experiments: ESS<25 is a relatively weak effect; ESS<50 is a moderate effect; ESS <75 is a relatively strong effect; ESS<90 is a strong effect; and greater values of ESS reflect a very strong effect.[26,33]

Finally, ESS indicates the ability of the ODA model to correctly classify all members of all class categories for the entire sample, while the Effect Strength for Predictive Value or ESP index assesses the accuracy of the model when it is used to make point predictions into a class category (ESS is invariant over base-rate, ESP is not[26]). As seen (Table 2), presently the ODA model is >95% accurate when classifying an individual motivation, and almost 70% accurate when classifying a community motivation.[34-36]

References

[1]Yarnold PR (2017). What is optimal data analysis? *Optimal Data Analysis*, *6*, 26-42.

[2]Yarnold PR, Soltysik RC (1991). Theoretical distributions of optima for univariate discrimination of random data. *Decision Sciences*, *22*, 739-752.

[3]Yarnold PR, Soltysik RC (2005). *Optimal data analysis: A guidebook with software for Windows*. Washington, DC, APA Books.

[4]Yarnold PR (2010). UniODA *vs*. chi-square: Ordinal data sometimes feign categorical. *Optimal Data Analysis*, *1*, 62-65.

[5]Yarnold PR (2014). UniODA *vs*. chi-square: Audience effect on smile production in infants. *Optimal Data Analysis*, *3*, 3-5.

[6]Yarnold PR (2014). UniODA *vs*. chi-square: Discriminating inhibited and uninhibited infant profiles. *Optimal Data Analysis*, *3*, 9-11.

[7]Yarnold PR (2015). UniODA *vs*. chi-square: Measures of effect size. *Optimal Data Analysis*, *4*, 137-138.

[8]Yarnold PR (2016). CTA *vs*. *not* chi-square: Differentiating statistical and ecological significance. *Optimal Data Analysis*, *5*, 112-115.

[9]Yarnold PR (2016). CTA *vs*. chi-square: Differentiating statistical and ecological significance. *Optimal Data Analysis*, *5*, 116-117.

[10]Yarnold, PR (2015). UniODA *vs*. chi-square: Deciphering *R* x *C* contingency tables. *Optimal Data Analysis*, *4*, 156-158.

2155-0182/10/$3.00

[11]Yarnold, PR (2015). UniODA *vs*. *not* chi-square: Vaccine administration and flu. *Optimal Data Analysis*, *4*, 159-160.

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

[13]Yarnold, PR (2015). UniODA *vs*. *not* chi-square: Work shift and raw material production quality. *Optimal Data Analysis*, *4*, 168-170.

[14]Yarnold PR (2016). ODA *vs*. undocumented chi-square: Clarity *vs*. confusion. *Optimal Data Analysis*, *5*, 121-123.

[15]Linden A, Yarnold PR (2016). Using data mining techniques to characterize participation in observational studies. *Journal of Evaluation in Clinical Practice*, *22*, 839-847.

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

[17]Yarnold PR (2016). CTA *vs*. *not* chi-square: Fear and specific recommendations *do* synergistically affect behavior. *Optimal Data Analysis*, *5*, 108-111.

[18]Yarnold PR (2016). CTA *vs*. *not* chi-square: Differentiating statistical and ecological significance. *Optimal Data Analysis*, *5*, 112-115.

[19]Yarnold PR (2016). CTA *vs*. chi-square: Differentiating statistical and ecological significance. *Optimal Data Analysis*, *5*, 116-117.

[20]Yarnold PR (2016). CTA *vs*. disintegrated chi-square: Integrated *vs*. piecemeal analysis. *Optimal Data Analysis*, *5*, 118-120.

[21]Yarnold PR (2016). CTA *vs*. chi-square: Comparing voter sentiment in political wards. *Optimal Data Analysis*, *5*, 129-130.

[22]Yarnold PR (2016). CTA vs. non-disentangled omnibus chi-square: Comparing samples (not) selected for study participation. *Optimal Data Analysis*, *5*, 154-157.

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

[24]Yarnold PR (2018). Minimize usage of binary measurement scales in rigorous classical research. *Optimal Data Analysis*, *7*, 3-9.

[25]Yarnold JK (1970). The minimum expectation in χ^2 goodness of fit tests and the accuracy of approximations for the null distribution. *Journal of the American Statistical Association*, *65*, 864-886. URL: http://www.jstor.org/stable/2284594

[26]Yarnold PR, Soltysik RC (2005). *Optimal data analysis: A guidebook with software for Windows*. Washington, DC, APA Books.

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

[28]Okuh D, Osumgborogwu IE (2019). Adjustments to hazards of gully erosion in rural southeast Nigeria: A case of Amacha communities. *Applied Ecology and Environmental Sciences*, *7*, 11-20.

[29]Yarnold PR (2018). Obtaining personal copies of ODA, MegaODA and CTA software. *Optimal Data Analysis*, *7*, 42-43.

[30]Soltysik RC, Yarnold PR (2013). MegaODA large sample and BIG DATA time trials: Separating the chaff. *Optimal Data Analysis*, *2*, 194-197.

[31]Soltysik RC, Yarnold PR (2013). MegaODA large sample and BIG DATA time trials: Harvesting the Wheat. *Optimal Data Analysis*, *2*, 202-205.

2155-0182/10/$3.00

[32]Yarnold PR, Soltysik RC (2013). MegaODA large sample and BIG DATA time trials: Maximum velocity analysis. *Optimal Data Analysis*, *2*, 220-221.

[33]Soltysik RC, Yarnold PR (2013). Statistical power of optimal discrimination with one attribute and two classes: One-tailed hypotheses. *Optimal Data Analysis*, *2*, 26-30.

[34]Yarnold PR (2013). Minimum standards for reporting UniODA findings. *Optimal Data Analysis*, *2*, 63-68.

[35]Yarnold PR (2013). Minimum standards for reporting UniODA findings for class variables with three or more response categories. *Optimal Data Analysis*, *2*, 86-93.

[36]Yarnold PR (2013). Standards for reporting UniODA findings expanded to include ESP and all possible aggregated confusion tables. *Optimal Data Analysis*, *2*, 106-119.

Author Notes

No conflict of interest was reported.