Optimal Data Analysis
Vol. 9 (May 6, 2020), 162-166

2155-0182/10/$3.00

Implementing ODA from Within Stata: Directional Hypothesis, Multicategorical Class Variable, Ordinal Attribute

Paul R. Yarnold, Ph.D. and Ariel Linden, Dr.P.H.
Optimal Data Analysis, LLC Linden Consulting Group, LLC

This paper describes how to assess a confirmatory (directional) hypothesis for a design involving a multicategorical class ("dependent") variable and an ordinal attribute ("independent variable") using the new Stata package for implementing ODA.

Recent papers[1-19] introduce the new Stata package called **oda**[20] for implementing ODA from within the Stata environment. Because this package is a wrapper for the MegaODA software system[21-23], the MegaODA.exe file must be loaded on the computer for the **oda** package to work (MegaODA software is available at https://odajournal.com/resources/). To download the **oda** package, at the Stata command line type: "ssc install oda" (without the quotation marks). This paper demonstrates use of the **oda** package to evaluate the *a priori* hypothesis for a design involving a four-category class variable and equivalent four-category ordinal attribute.

Methods

Data

As an example of a directional hypothesis involving a multicategorical class variable and an ordinal attribute, consider Bishop, Feinberg and Holland's data[24] concerning visual acuity of left and right eyes for a sample of 7,477 women undergoing eye exams. Unaided distant visual acuity was assessed separately for each eye using a 4-point ordinal scale with 1=acuity $\leq 25^{th}$ percentile; 2=acuity $\leq 50^{th}$ percentile; 3=acuity $\leq 75^{th}$ percentile; and 4=acuity $\leq 100^{th}$ percentile. The directional alternative hypothesis being tested is that since a person's eyes tend to have comparable visual acuity, the right eye acuity (randomly selected to serve as class variable) should be directly discriminable on the basis of the left eye acuity.

Analytic Process

The directional ("one-sided") *a priori* hypothesis is left and right eyes have the same visual acuity rating, and the null hypothesis is that this is not true. Exact *p* is estimated by a 25,000-iteration permutation test. For the entire sample, **oda** is implemented with the following syntax (see the help file for **oda** for a complete description of syntax options):

2155-0182/10/$3.00

oda righteye lefteye , pathoda(“C:\ODA\”) store(“C:\ODA”) iter(25000) dir(< 1 2 3 4)

This syntax is explained as follows: “righteye” is the *class* variable and “lefteye” is the *attribute*; “C:\ODA\” is the directory path where the MegaODA.exe file exists on the computer, and where all other files generated in analysis are stored; the number of iterations (repetitions) used to compute a permutation *p*-value is 25,000; the attribute (righteye) is ordered (default); and the directional hypothesis is that visual acuity code assignments of the left and right eye agree. Data for each observation was entered in free format on a separate line in space-delimited text (ASCII) characters.[25,26]

The **oda** package produces an extract of the total output produced by the ODA software (the complete output is stored in the specified directory with the extension “.out”).

```
ODA model:
----------
IF LEFTEYE <= 1.5 THEN RIGHTEYE = 1
IF 1.5 < LEFTEYE <= 2.5 THEN RIGHTEYE = 2
IF 2.5 < LEFTEYE <= 3.5 THEN RIGHTEYE = 3
IF 3.5 < LEFTEYE THEN RIGHTEYE = 4

Summary for Class RIGHTEYE  Attribute LEFTEYE
---------------------------------------------

Performance Index          Train
-----------------         ------
Overall Accuracy           70.83%
PAC RIGHTEYE=1             76.92%
PAC RIGHTEYE=2             67.02%
PAC RIGHTEYE=3             72.15%
PAC RIGHTEYE=4             62.36%
Effect Strength PAC        59.48%
PV RIGHTEYE=1              79.71%
PV RIGHTEYE=2              68.05%
PV RIGHTEYE=3              70.68%
PV RIGHTEYE=4              58.50%
Effect Strength PV         58.98%
Effect Strength Total      59.23%

Monte Carlo summary (Fisher randomization):
-------------------------------------------
Iterations:  25000
Estimated p: 0.000000
```

The effect strength for sensitivity (ESS) is labelled in the output as the “Effect Strength PAC” (Percentage Accurate Classification). As seen, ESS for the *a priori* hypothesis is 59.48%, which corresponds to a relatively strong effect[25] and has permutation $p<0.00001$.

In summary, ODA identified a relatively strong model which supported the *a priori* hypothesis that one’s left and right eyes tend to exhibit comparable visual acuity.

We believe ODA should be considered the preferred statistical approach over other methods because it avoids statistical assumptions required of conventional models, is insensitive to skewed data or outliers, and has the ability to handle any variable metric including categorical, Likert-type integer, and real number measurement scales.[25] In contrast to alternative methods, only ODA can identify the optimal (maximum-accuracy) assignments (categorical attributes) or cutpoints (ordered attributes) that exist for the attribute, which in turn facilitates the use of measures of predictive accuracy.

Furthermore, ODA can evaluate model reproducibility by multiple methods, allowing assessment of potential cross-generalizability of the model applied to classify an independent random sample.[25]

For these reasons we recommend that researchers employ ODA and CTA frameworks to evaluate the statistical hypotheses which are explored in their laboratory and field research endeavors.[27-46]

References

[1]Linden A (2020). Implementing ODA from within Stata: An application to data from a randomized controlled trial (*Invited*). *Optimal Data Analysis, 9*, 9-13.

[2]Linden A (2020). Implementing ODA from within Stata: Implementing ODA from within Stata: An application to estimating treatment effects using observational data (*Invited*). *Optimal Data Analysis, 9*, 14-20.

2155-0182/10/$3.00

[3]Linden A (2020). Implementing ODA from within Stata: An application to dose-response relationships (*Invited*). *Optimal Data Analysis*, *9*, 26-32.

[4]Linden A (2020). Implementing ODA from within Stata: assessing covariate balance in observational studies (*Invited*). *Optimal Data Analysis*, *9*, 33-38.

[5]Linden A (2020). Implementing ODA from within Stata: Evaluating treatment effects for survival (time-to-event) outcomes (*Invited*). *Optimal Data Analysis*, *9*, 39-44.

[6]Linden A (2020). Implementing ODA from within Stata: Evaluating treatment effects in multiple-group interrupted time series analysis (Invited). *Optimal Data Analysis*, *9*, 45-50.

[7]Linden A (2020). Implementing ODA from within Stata: identifying structural breaks in single-group interrupted time series designs (Invited). *Optimal Data Analysis*, *9*, 51-56.

[8]Linden A (2020). Implementing ODA from within Stata: Finding the optimal cut-point of a diagnostic test or index (Invited). *Optimal Data Analysis*, *9*, 74-78.

[9]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Exploratory hypothesis, binary class variable, and binary attribute. *Optimal Data Analysis*, *9*, 94-98.

[10]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Confirmatory hypothesis, binary class variable, and binary attribute. *Optimal Data Analysis*, *9*, 99-103.

[11]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Exploratory hypothesis, binary class variable, and binary attribute. *Optimal Data Analysis*, *9*, 104-108.

[12]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Exploratory hypothesis, binary class variable, and ordinal (rank) attribute. *Optimal Data Analysis*, *9*, 109-113.

[13]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: confirmatory hypothesis, binary class variable, and ordinal attribute. *Optimal Data Analysis*, *9*, 128-132.

[14]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Exploratory hypothesis, binary class variable, categorical ordinal attribute. *Optimal Data Analysis*, *9*, 133-136.

[15]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Nondirectional hypothesis, binary class variable, categorical ordinal attribute. *Optimal Data Analysis*, *9*, 137-140.

[16]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Directional hypothesis, binary class variable, ordinal attribute. *Optimal Data Analysis*, *9*, 141-145.

[17]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Confirmatory hypothesis, binary class variable, continuous attribute. *Optimal Data Analysis*, *9*, 146-151.

[18]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Nondirectional, multicategorical class variable, multicategorical attribute. *Optimal Data Analysis*, *9*, 152-156.

[19]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Directional hypothesis, multicategorical class variable and attribute. *Optimal Data Analysis*, *9*, 157-161.

[20]Linden A (2020). ODA: Stata module for conducting Optimal Discriminant Analysis. *Statistical Software Components S458728, Boston College Department of Economics*.

2155-0182/10/$3.00

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

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

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

[24]Bishop YMM, Fienberg SE, Holland PW (1975). *Discrete multivariate analysis*. Cambridge, England: Cambridge University Press.

[25]Yarnold PR, Soltysik RC (2005). *Optimal data analysis: Guidebook with software for Windows*. Washington, D.C.: APA Books.

[26]Bryant FB, Harrison PR (2013). How to create an ASCII input data file for UniODA and CTA software (*Invited*). *Optimal Data Analysis*, *2*, 2-6.

[27]Linden A, Yarnold PR, Nallomothu BK (2016). Using machine learning to model dose-response relationships. *Journal of Evaluation in Clinical Practice*, *22*, 860-867.

[28]Yarnold PR, Linden A. (2016). Novometric analysis with ordered class variables: The optimal alternative to linear regression analysis, *Optimal Data Analysis*, *5*, 65-73.

[29]Yarnold PR, Linden A (2016). Theoretical aspects of the D statistic. *Optimal Data Analysis*, *22*, 171-174.

[30]Linden A, Yarnold PR (2017). Using classification tree analysis to generate propensity score weights. *Journal of Evaluation in Clinical Practice*, *23*, 703-712.

[31]Linden A, Yarnold PR (2017). Modeling time-to-event (survival) data using classification tree analysis. *Journal of Evaluation in Clinical Practice*, *23*, 1299-1308.

[32]Linden A, Yarnold PR (2018). Identifying causal mechanisms in health care interventions using classification tree analysis. *Journal of Evaluation in Clinical Practice*, *24*, 353-361.

[33]Linden A, Yarnold PR (2017). Minimizing imbalances on patient characteristics between treatment groups in randomized trials using classification tree analysis. *Journal of Evaluation in Clinical Practice*, *23*, 1309-1315.

[34]Linden A, Yarnold PR (2018). Estimating causal effects for survival (time-to-event) outcomes by combining classification tree analysis and propensity score weighting. *Journal of Evaluation in Clinical Practice, 24*, 380-387.

[35]Linden A, Yarnold PR (2016). Using machine learning to assess covariate balance in matching studies. *Journal of Evaluation in Clinical Practice*, 22, 848-854.

[36]Linden A, Yarnold PR (2016). Using machine learning to identify structural breaks in single-group interrupted time series designs. *Journal of Evaluation in Clinical Practice*, *22*, 855-859.

[37]Linden A, Yarnold PR (2016). Combining machine learning and matching techniques to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice*, *22*, 868-874.

[38]Linden A, Yarnold PR (2016). Combining machine learning and propensity score weighting to estimate causal effects in multivalued treatments. *Journal of Evaluation in Clinical Practice*, *22*, 875-885.

[39]Linden A, Yarnold PR (2018). Using machine learning to evaluate treatment effects in multiple-group interrupted time series analysis. *Journal of Evaluation in Clinical Practice, 24,* 740-744.

[40]Rhodes NJ (2020). Statistical power analysis in ODA, CTA and Novometrics (Invited). *Optimal Data Analysis*, *9*, 21-25. [41]Yarnold PR (2010). UniODA *vs*. chi-square: Ordinal data sometimes feign categorical. *Optimal Data Analysis, 1*, 62-65.

[42]Yarnold PR, Brofft GC (2013). ODA range test *vs*. one-way analysis of variance: Comparing strength of alternative line connections. *Optimal Data Analysis, 2*, 198-201.

[43]Yarnold PR (2013). ODA range test *vs*. one-way analysis of variance: Patient race and lab results. *Optimal Data Analysis, 2*, 206-210.

[44]Yarnold PR (2014). How to assess the inter-method (parallel-forms) reliability of ratings made on ordinal scales: Evaluating and comparing the Emergency Severity Index (Version 3) and Canadian Triage Acuity Scale. *Optimal Data Analysis, 3*, 50-54.

[45]Yarnold PR (2016) Causality of adverse drug reactions: The upper-bound of arbitrated expert agreement for ratings obtained by WHO and Naranjo algorithms. *Optimal Data Analysis, 5,* 37-40.

[46]Yarnold PR (2016). Novometric *vs*. ODA reliability analysis *vs*. polychoric correlation with relaxed distributional assumptions: Inter-rater reliability of independent ratings of plant health. *Optimal Data Analysis, 5*, 179-183.

Author Notes

No conflicts of interest were reported.