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 introduce the new Stata package called oda for implementing ODA from within the Stata environment. Because this package is a wrapper for the MegaODA software system, 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 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 ≤25th percentile; 2=acuity ≤50th percentile; 3=acuity ≤75th percentile; and 4=acuity ≤100th 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):
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:

<table>
<thead>
<tr>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>IF LEFTEYE &lt;= 1.5 THEN RIGHTEYE = 1</td>
</tr>
<tr>
<td>IF 1.5 &lt; LEFTEYE &lt;= 2.5 THEN RIGHTEYE = 2</td>
</tr>
<tr>
<td>IF 2.5 &lt; LEFTEYE &lt;= 3.5 THEN RIGHTEYE = 3</td>
</tr>
<tr>
<td>IF 3.5 &lt; LEFTEYE THEN RIGHTEYE = 4</td>
</tr>
</tbody>
</table>

Summary for Class RIGHTEYE Attribute LEFTEYE

<table>
<thead>
<tr>
<th>Performance Index</th>
<th>Train</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Accuracy</td>
<td>70.83%</td>
</tr>
<tr>
<td>PAC RIGHTEYE=1</td>
<td>76.92%</td>
</tr>
<tr>
<td>PAC RIGHTEYE=2</td>
<td>67.02%</td>
</tr>
<tr>
<td>PAC RIGHTEYE=3</td>
<td>72.15%</td>
</tr>
<tr>
<td>PAC RIGHTEYE=4</td>
<td>62.36%</td>
</tr>
<tr>
<td>Effect Strength PAC</td>
<td>59.48%</td>
</tr>
<tr>
<td>PV RIGHTEYE=1</td>
<td>79.71%</td>
</tr>
<tr>
<td>PV RIGHTEYE=2</td>
<td>68.05%</td>
</tr>
<tr>
<td>PV RIGHTEYE=3</td>
<td>70.68%</td>
</tr>
<tr>
<td>PV RIGHTEYE=4</td>
<td>58.50%</td>
</tr>
<tr>
<td>Effect Strength PV</td>
<td>58.98%</td>
</tr>
<tr>
<td>Effect Strength Total</td>
<td>59.23%</td>
</tr>
</tbody>
</table>

Monte Carlo summary (Fisher randomization):

<table>
<thead>
<tr>
<th>iterations: 25000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated p: 0.00000</td>
</tr>
</tbody>
</table>

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 effect25 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

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

2Linden 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.


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.


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.


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.


Author Notes

No conflicts of interest were reported.