Optimal Data Analysis
Vol. 4 (December 26, 2015), 194-196

2155-0182/10/$3.00

UniODA-Based Structural Decomposition *vs*. Legacy Linear Models: Statics and Dynamics of Intergenerational Occupational Mobility

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

Analysis assessed structure underlying the cross-classification of occupational category of N = 3,396 sons and fathers.[1] A plethora of linear legacy models have been developed for these data, without a clearly superior solution emerging.[1] Structural decomposition analysis identified four models that together yielded a very strong effect.

UniODA-based structural decomposition analysis[2,3] first tested the *a priori* hypothesis that the occupational category of father and son (1 = professional and managerial; 2 = clerical, sales, and proprietors; 3 = craftsmen; 4 = operatives and laborers; 5 = farmers and farm laborers) is coincident[2,3] using the following UniODA[2,3] and MegaODA[4-6] software syntax (sons formed the columns, and fathers the rows, of the table):

```
OPEN DATA;
OUTPUT mobility.out;
CATEGORICAL ON;
TABLE 5;
CLASS COLUMN;
DIRECTIONAL < 1 2 3 4 5;
MCARLO ITER 10000;
DATA;
152 66 33 39 4
201 159 72 80 8
138 125 184 172 7
143 161 209 378 17
98 146 207 371 226
END DATA;
GO;
```

The classification accuracy yielded by the *a priori* hypothesis was statistically significant ($p < 0.0001$) but it reflected a weak effect[2,3] (*ESS* = 23.4). Nevertheless, fathers and sons in the farmer and farm laborer category showed very strong consistency (see Table 1).

To prepare the data for the second step of the decomposition analysis, in the initial data the correctly classified observations (i.e., the table cells) were set equal to zero, rendering a *residual table* consisting of all misclassified observations: that is, the original table with all elements of the major diagonal set to zero. In the absence of an *a priori* hypothesis concerning structure underlying the residual data, an exploratory UniODA model was obtained by deleting the DIRECTIONAL command, and substituting the following residual data for the original data, in the syntax given earlier:

```
0 66 33 39 4
201 0 72 80 8
138 125 0 172 7
143 161 209 0 17
98 146 207 371 0
```

The resulting UniODA model was:

if father's class = 1, predict son's class = 3;
if father's class = 2, predict son's class = 1;
if father's class = 3, predict son's class = 2;
if father's class = 4, predict son's class = 5;
if father's class = 5, predict son's class = 4;

This model identified a statistically significant ($p < 0.0001$), relatively weak (*ESS* = 17.3) effect. Considered as a whole the two models identified thus far correctly classify 54.4% of the total sample, and the combined classification accuracy (*ESS* = 46.7) reflects a moderate effect (see Table 1).

In the third step of the decomposition analysis all observations correctly classified by the second model are deleted. In addition, since there are insufficient class = 5 data for sons to provide adequate statistical power, all remaining son = 5 cells are set to zero. The residual data table for the third analysis is:

```
0 66 0 39 0
0 0 72 80 0
138 0 0 172 0
143 161 209 0 0
98 146 207 0 0
```

In UniODA and MegaODA software the TABLE macro requires at least one observation in every class category, so standard data entry must be used.[7] The residual data (residual.txt) are analyzed using the following UniODA and MegaODA command syntax:

```
OPEN residual.txt;
OUTPUT model3.out;
VARS father son;
CLASS son;
ATTR father;
CAT father;
MC ITER 10000;
GO;
```

The resulting *degenerate* (i.e., not all of the class categories are used in the model[2,3]) UniODA model was:

if father's class = 1, predict son's class = 2;
if father's class = 2, predict son's class = 4;
if father's class = 3, predict son's class = 4;
if father's class = 6, predict son's class = 2;
if father's class = 7, predict son's class = 3;

This model identified a statistically significant ($p < 0.0001$), moderate (*ESS* = 30.0) effect. Considered as a whole the three models identified thus far correctly classify 74.6% of the total sample, and the combined classification accuracy (*ESS* = 68.8) constitutes a relatively strong effect (see Table 1).

Using the following residual table (vis-à-vis standard data entry), a fourth and final (third exploratory, second degenerate) UniODA model was extracted:

```
0 0 0 39 0
0 0 72 0 0
138 0 0 0 0
143 0 209 0 0
98 146 0 0 0
```

The UniODA model was:

if father's class = 1, predict son's class = 4;
if father's class = 2, predict son's class = 3;

2155-0182/10/$3.00

if father's class = 3, predict son's class = 1;
if father's class = 4, predict son's class = 3;
if father's class = 5, predict son's class = 2;

This model identified a statistically significant ($p < 0.0001$), strong (*ESS* = 78.8) effect.

Sequential cumulative classification performance of the four models identified in the structural decomposition analysis is summarized in Table 1. The models yield monotonically increasing cumulative *ESS* and percent of total sample correctly classified.

Table 1: Cumulative Classification Performance of the Four UniODA Models Identified in Structural Decomposition Analysis

	Sensitivity as a Function of Sequentially Extracted UniODA Models			
Son's Class	1	2	3	4
1	20.8	48.2	48.2	67.1
2	24.2	43.2	77.8	100
3	26.1	30.8	60.1	100
4	36.4	72.0	96.2	100
5	86.3	92.7	92.7	92.7
% of Sample Correctly Classified	32.4	54.4	74.6	92.3
ESS	23.4	46.7	68.8	90.0

Considered together the four models yielded a very strong effect, correctly classifying 9 of 10 sons in the sample. This level of accuracy is impossible in the absence of nearly perfect reliability of class assignment.

References

[1]Gilbert N (1993). Analyzing tabular data: Log-linear and logistic models for social researchers. London, England: UCL Press (pp. 86-88).

[2]Yarnold PR, Soltysik RC (In Review). *Maximizing predictive accuracy*. Chicago, IL: ODA Books.

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

[4]Soltysik RC, Yarnold PR (2013). MegaODA large sample and BIG DATA time trials: Separating the chaff. *Optimal Data Analysis*, *2*, 194-197. URL: http://optimalprediction.com/files/pdf/V2A29.pdf

[5]Soltysik RC, Yarnold PR (2013). MegaODA large sample and BIG DATA time trials: Harvesting the Wheat. *Optimal Data Analysis*, *2*, 202-205. URL: http://optimalprediction.com/files/pdf/V2A31.pdf

[6]Yarnold PR, Soltysik RC (2013). MegaODA large sample and BIG DATA time trials: Maximum velocity analysis. *Optimal Data Analysis*, *2*, 220-221. URL: http://optimalprediction.com/files/pdf/V2A35.pdf

[7]Bryant FB, Harrison PR (2013). How to create an ASCII input data file for UniODA and CTA software. *Optimal Data Analysis*, *2*, 2-6. URL: http://optimalprediction.com/files/pdf/V2A1.pdf

Author Notes

The study analyzed de-individuated data and was exempt from Institutional Review Board review. No conflict of interest was reported.

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