Optimal Data Analysis
Vol. 4 (December 23, 2015), 188-189

2155-0182/10/$3.00

Occupational Class, Tenure, and Voting

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

A globally-optimal (GO) classification model yields moderate accuracy in modeling the voting behavior of $N = 695$ people as a function of occupational class.[1]

Attributes were occupational *class* (5-point ordered scale) and *tenure* (4-point ordered scale), and the class variable was *vote* (conservative or labour).[1] The minimum-denominator selection algorithm (MDSA) was used to identify the descendant family of all possible enumerated-optimal CTA (EO-CTA) models that exist in this application (all models except the first had sufficient statistical power).[2] The unrestricted initial (most granular) model in the family was identified via the following CTA software[3] syntax:

```
OPEN vote.dat;
OUTPUT vote.out;
VARS vote class tenure;
CLASS vote;
ATTRIBUTE class tenure;
MC ITER 5000 CUTOFF .05 STOP 99.9;
PRUNE .05;
ENUMERATE;
GO;
```

Table 1 summarizes the descendant family of three EO-CTA models: model number indicates order of discovery by MDSA; N_{MIN} is the size of the smallest (least populated) strata (endpoint) in the model; *ESS* is a chance- and maximum-corrected measure of predictive accuracy; and *D* is the number of additional equivalent effects needed to obtain a theoretically ideal statistical model in this application.[2]

Table 1: The Descendant Family

Model	N_{MIN}	Strata	*ESS*	*D*
1	36	5	46.5	5.75
2	45	3	41.9	4.16
3	269	2	41.0	2.87

Model 3 (if class ≤ 2 then predict vote = conservative; otherwise predict vote = labour) has the lowest *D* statistic and so is the globally-optimal (GO) model in this application: *ESS* was a statistically significant ($p < 0.0001$) and moderate effect (the exact discrete 95% CI for *ESS* is 33.2 – 48.7 for the model, and 0.18 – 7.3 for chance). Table 2 presents the confusion table summarizing predictive accuracy of the model.

Table 2: GO Model Predictive Accuracy

		Predicted Vote		
		Con	Lab	Sensitivity
Actual	Con	206	142	59.2%
Vote	Lab	63	284	81.8%
Predictive Value		76.6%	66.7%	

2155-0182/10/$3.00

References

[1]Gilbert N (1993). Analyzing tabular data: Log-linear and logistic models for social researchers. London, England: UCL Press (pp. 52-62).

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

[3]Soltysik RC, Yarnold PR (2010). Automated CTA software: Fundamental concepts and control commands. *Optimal Data Analysis, 1*, 144-160. URL: http://odajournal.com/2013/09/19/62/

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