Optimal Data Analysis
Vol. 4 (December 22, 2015), 186-187

2155-0182/10/$3.00

Globally-Optimal CTA Model of Voting for US Senators

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

A globally-optimal classification tree analysis (GO-CTA) model yields relatively strong predictive accuracy in modeling the behavior of N = 2,843 people voting for Senators in statewide elections.[1]

The three categorical attributes included *party* identification (Democrat, Republican, Independent); if the candidate was an *incum*bent (Democrat incumbent, Republican incumbent, Not an incumbent); and attitude toward party *policy* (Democrat, Republican, Neutral).[1] The class variable was *vote* (Democrat, Republican).

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 four family models had sufficient statistical power).[2] The unrestricted initial (most granular) model in the family was identified via the CTA software[3] syntax below:

```
OPEN senator.dat;
OUTPUT senator.out;
VARS party incum policy vote;
CLASS vote;
ATTRIBUTE party incum policy;
CATEGORICAL party incum policy;
MC ITER 5000 CUTOFF .05 STOP 99.9;
PRUNE .05;
ENUMERATE;
GO;
```

Table 1 is a summary of the descendant family of four 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	68	6	60.7	3.88
2	193	5	60.1	3.32
3	274	5	58.2	3.59
4	1,310	2	56.5	1.54

Model 4 (illustrated in Figure 1, the endpoints give exact discrete 95% confidence intervals in parentheses) has the lowest *D* statistic and thus is the globally-optimal classification tree analysis (GO-CTA) model. *ESS* yielded by the model was statistically significant ($p < 0.0001$) and reflected a relatively strong effect: the exact discrete 95% CI for *ESS* is 52.9 – 60.0 for the model, and is 0.09 – 3.7 for chance.

2155-0182/10/$3.00

Figure 1: GO-CTA Model Predicting Voting Behavior

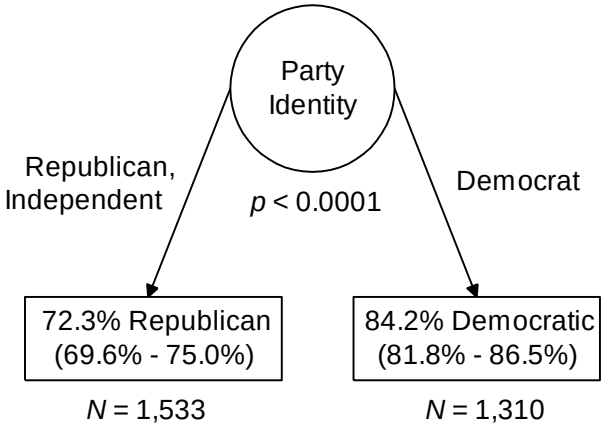

Table 2 presents the confusion table summarizing the predictive accuracy of the GO-CTA model: seven in ten people who identified with the Republican or Independent party voted Republican; and seven in eight people identifying with Democrats voted Democratic.

Table 2: GO-CTA Model Predictive Accuracy

		Predicted Vote		
		Dem	Rep	Sensitivity
Actual	Dem	1,109	207	84.3%
Vote	Rep	424	1,103	72.2%
Predictive Value		72.3%	84.2%	

References

[1]Gilbert N (1993). Analyzing tabular data: Log-linear and logistic models for social researchers. London, England: UCL Press (pp. 131-109).

[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