Optimal Data Analysis
Vol. 5 (November 10, 2016), 194-197

2155-0182/10/$3.00

Novometric Analysis Predicting Voter Turnout: Race, Education, and Organizational Membership Status

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

Prior research modeled voter turnout (“not voted”=0; “voted”=1) as a function of race (“white”=1; “black”=2), education (“less than high school”=1; “high school graduate”=2; “college”=3), and memberships in organizations (“none”=0; “one or more”=1) via log-linear analysis. Results revealed: “…in the absence of a confirmatory analysis with another sample and in the absence of any compelling theoretical argument for expecting the particular three-variable interaction, (our own preference) would be to choose the more parsimonious model 28, {MER}{MV}{EV}. That model gives a satisfactory fit to the full crosstabulation (L^2=4.76, df=5, p<0.45) without resort to a complex three-variable interaction. It also omits the race-turnout effect which is known to be trivial, but which would have to be included in model 34 because it is subsumed in hierarchical relation to the {ERV} term” (p. 40). Exploratory novometric analysis[2-22] is used to model voter turnout (binary class variable) as a function of race (a categorical attribute), education and number of organizational memberships (both treated as ordered attributes measured on categorical ordinal scales).

Data analyzed herein[1] are indicated in SAS™ code used to construct the data set required for novometric analysis (see Appendix), yielding the descendant family (DF) of optimal models summarized in Table 1. Model 4 had the lowest D statistic and thus is the globally-optimal (GO) model in this application.[2,21] The two-strata GO model was: if no organizational memberships then predict not voted; otherwise predict voted.

The confusion matrix for the GO model is given in Table 2 (relatively weak ESS=22.1, p<0.001). The model accurately predicted 1 in 2 of actual non-voters (50% accuracy is expected by chance for each class category in two-category applications that do not involve analytic weighting[2,24-27]), and 7 in 10 of actual voters.

Table 1: Descendant Family of Optimal Models Predicting Voter Turnout

Step	ESS	Strata	Efficiency	D	Minimum Endpoint N
1	23.87	4	5.97	12.76	41
2	23.72	4	5.93	12.86	83
3	23.72	3	7.91	9.65	441
4	22.07	2	11.04	7.06	552

2155-0182/10/$3.00

Table 2: Confusion Matrix for GO Model

		Predicted Non-Voter	Predicted Voter	
Actual	Non-Voter	254	232	52.3%
	Voter	298	689	69.8%

In contrast to the conclusion reached on the basis of the original analysis, the GO model involved only organizational membership as an attribute. Prior to the discovery of novometric statistical analysis, EO-CTA-based structural decomposition analysis was used to identify structure in data ordinarily analyzed using the log-linear model, and the maximum-accuracy solutions always achieved more parsimonious and accurate models than obtained using legacy linear multivariable methods.[2,28-30] Novometrics now enables explicit identification of the model representing the best combination of accuracy and parsimony possible for any application.

References

[1]Knoke D, Burke PJ (1980). *Log-linear models*. Beverly Hills, CA: Sage (pp. 22-40).

[2]Yarnold PR, Soltysik RC (2016). *Maximizing predictive accuracy*. Chicago, IL: ODA Books. DOI: 10.13140/RG.2.1.1368.3286

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

[4]Yarnold PR, Bennett CL (2016). Novometrics *vs*. correlation: Age and clinical measures of PCP survivors, *Optimal Data Analysis*, *5*, 74-78.

[5]Yarnold PR, Bennett CL (2016). Novometrics *vs*. multiple regression analysis: Age and clinical measures of PCP survivors, *Optimal Data Analysis*, *5*, 79-82.

[6]Yarnold PR (2016). Novometrics *vs*. regression analysis: Literacy, and age and income, of ambulatory geriatric patients. *Optimal Data Analysis*, *5*, 83-85.

[7]Yarnold PR (2016). Novometrics *vs*. regression analysis: Modeling patient satisfaction in the Emergency Room. *Optimal Data Analysis*, *5*, 86-93.

[8]Yarnold PR (2016). Matrix display of pairwise novometric associations for ordered variables. *Optimal Data Analysis*, *5*, 94-101.

[9]Yarnold PR, Batra M (2016). Matrix display of pairwise novometric associations for mixed-metric variables. *Optimal Data Analysis*, *5*, 104-107.

[10]Yarnold PR (2016). Novometrics *vs*. ODA *vs*. One-Way ANOVA: Evaluating comparative effectiveness of sales training programs, and the importance of conducting LOO with small samples. *Optimal Data Analysis*, *5*, 131-132.

[11]Yarnold PR (2016). Parental smoking behavior, ethnicity, gender, and the cigarette smoking behavior of high school students. *Optimal Data Analysis*, *5*, 136-140.

[12]Yarnold PR (2016). Using gender of an imaginary rated smoker, and subject's gender, ethnicity, and smoking behavior to identify perceived differences in peer-group smoking standards of American high school students. *Optimal Data Analysis*, *5*, 141-143.

[13]Yarnold PR (2016). Novometric models of smoking habits of male and female friends of American college undergraduates: Gender, smoking, and ethnicity. *Optimal Data Analysis*, *5*, 146-150.

[14]Yarnold PR (2016). Predicting daily television viewing of senior citizens using education, age and marital status. *Optimal Data Analysis*, *5*, 151-152.

2155-0182/10/$3.00

[15]Yarnold PR (2016). Novometric statistical analysis and the Pearson-Yule debate. *Optimal Data Analysis*, *5*, 162-165.

[16]Yarnold PR (2016). Comparing WAIS-R qualitative information for people 75 years and older, with *vs*. without brain damage. *Optimal Data Analysis*, *5*, 166-170.

[17]Yarnold PR (2016). Using novometrics to disentangle complete sets of sign-test-based multiple-comparison findings. *Optimal Data Analysis*, *5*, 175-176.

[18]Yarnold PR (2016). Novometric analysis *vs*. MANOVA: MMPI codetype, gender, setting, and the MacAndrew Alcoholism scale. *Optimal Data Analysis*, *5*, 177-178.

[19]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.

[20]Yarnold PR (2016). Novometrics *vs*. polychoric correlation: Number of lambs born over two years. *Optimal Data Analysis*, *5*, 184-185.

[21]Yarnold PR (2016). Comparing MMPI-2 *F-K* Index normative data among male and female psychiatric and head-injured patients, individuals seeking disability benefits, police and priest job applicants, and substance abusers. *Optimal Data Analysis*, *5*, 186-193.

[22]Yarnold PR, Linden A (2016). Theoretical aspects of the D statistic. *Optimal Data Analysis*, *5*, 171-174.

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

[24]Linden A, Yarnold PR (In Press). Combining machine learning and propensity score weighting to estimate causal effects in multivalued treatments. *Journal of Evaluation in Clinical Practice*.

[25]Linden A, Yarnold PR (In Press). Combining machine learning and matching techniques to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice*.

[26]Linden A, Yarnold PR (In Press). Using machine learning to assess covariate balance in matching studies. *Journal of Evaluation in Clinical Practice*. DOI: 10.1111/jep.12538

[27]Linden A, Yarnold PR (In Press). Using data mining techniques to characterize participation in observational studies. *Journal of Evaluation in Clinical Practice*.

[28]Yarnold PR (2015). UniODA-based structural decomposition *vs*. log-linear model: Statics and dynamics of intergenerational class mobility. *Optimal Data Analysis*, *4*, 179-181.

[29]Yarnold PR (2015). Modeling religious mobility by UniODA-based structural decomposition. *Optimal Data Analysis*, *4*, 192-193.

[30]Yarnold PR (2015). UniODA-based structural decomposition *vs*. legacy linear models: Statics and dynamics of intergenerational occupational mobility. *Optimal Data Analysis*, *4*, 194-196.

Author Notes

This study analyzed publically available data. No conflict of interest was reported.

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

2155-0182/10/$3.00

Appendix

SAS™ Code used to Construct (Reproduce[1]) the Data File for Analysis by ODA Software[2,23]

```
data real;
infile datalines;
input race edu member
vote;
cards;
1 1 1 1
;
Data example;
Do n=1 to 114;
put '1 1 0 1';
end;
Do n=1 to 122;
put '1 1 0 0';
end;
Do n=1 to 150;
put '1 1 1 1';
end;
Do n=1 to 67;
put '1 1 1 0';
end;
Do n=1 to 88;
put '1 2 0 1';
end;
Do n=1 to 72;
put '1 2 0 0';
end;
Do n=1 to 208;
put '1 2 1 1';
end;
Do n=1 to 83;
put '1 2 1 0';
end;
Do n=1 to 58;
put '1 3 0 1';
end;
Do n=1 to 18;
put '1 3 0 0';
end;
Do n=1 to 264;
put '1 3 1 1';
end;
Do n=1 to 60;
put '1 3 1 0';
end;
Do n=1 to 23;
put '2 1 0 1';
end;
Do n=1 to 31;
put '2 1 0 0';
end;
Do n=1 to 22;
put '2 1 1 1';
end;
Do n=1 to 7;
put '2 1 1 0';
end;
Do n=1 to 12;
put '2 2 0 1';
end;
Do n=1 to 7;
put '2 2 0 0';
end;
Do n=1 to 21;
put '2 2 1 1';
end;
Do n=1 to 5;
put '2 2 1 0';
end;
Do n=1 to 3;
put '2 3 0 1';
end;
Do n=1 to 4;
put '2 3 0 0';
end;
Do n=1 to 24;
put '2 3 1 1';
end;
Do n=1 to 10;
put '2 3 1 0';
end;
Output;
Run;
```