Optimal Data Analysis
Vol. 9 (July 4, 2020), 209

2155-0182/10/$3.00

The Novometric Descendant Family as a Maximum-Accuracy Cluster Analysis

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

This note interprets a novometric *descendant family* (DF) as an optimal "cluster" analysis indicating number of discriminable groups and strength of their differences at every differentiable point identified for the sample.

Axiom three[1] of novometric theory discusses employing the minimum denominator selection algorithm (MDSA) to identify the DF—the set of all models existing in the sample that satisfy sequential Sidak *p*-value criteria[2] and that have a monotonically increasing minimum strata N (model endpoint sample size).[3] For exposition, Table 1 is the DF obtained when parsing SEER-reported eye and orbit cancer-incidence data for national samples of black *vs*. white patients.[4]

Table 1: MDSA Parsing of Eye and Orbit Cancer Incidence by Race: Blacks *vs*. Whites[4]

Strata	MinD	ESS	Efficiency	D
7	27	71.1	10.2	2.85
		64.3-77.6	9.19-11.1	2.02-3.89
		0-7.89	0-1.13	Division by 0
6	47	68.1	11.3	2.81
		61.3-74.6	10.2-12.4	2.04-3.79
		0.33-7.57	0.06-1.26	1812-73.3
5	51	62.8	12.6	2.96
		55.6-70.1	11.1-14.0	2.13-3.99
		0.33-7.57	0.07-1.51	1510-61.1
2	202	62.5	31.2	1.20
		55.8-69.0	27.9-34.5	0.90-1.58
		0-7.24	0-3.62	Division by 0

Novometric analysis found four different patient groupings indicated by number of strata. Point estimates and 95% exact CIs of 2-6-strata models were relatively strong effects ($50<ESS\leq75$), and the 7-strata 95% CI upper bound met criterion for a strong effect ($ESS\leq90$).[2] Chance was always very weak. Results justify further evaluation of 2-, 5-, 6- and 7-strata solutions, for example examining magnitude and pattern of overlap of subjects over the different solutions.

References

[1]Yarnold PR (2020). What is novometric data analysis? *Optimal Data Analysis*, *9*, 195-206.

[2]Yarnold PR, Soltysik RC (2005). *Optimal data analysis: Guidebook with software for Windows*. Washington, D.C.: APA Books.

[3]Yarnold PR (2016). How many EO-CTA models exist in my sample and which is the best model? *Optimal Data Analysis*, *5*, 62-64.

[4]Yarnold PR, Soltysik RC (2014). Globally optimal statistical classification models, I: Binary class variable, one ordered attribute. *Optimal Data Analysis*, *3*, 55-77.