Optimal Data Analysis
Vol. 7 (March 24, 2018), 26-27

2155-0182/10/$3.00

ANOVA with Two Between-Groups Factors *vs*. Novometric Analysis

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

ANOVA with two between-groups factors is used to compare an ordered attribute (score) between four independent categories of class variable “A”, and three independent categories of class variable “B”. The novometric multiple regression analogue is demonstrated.

Described elsewhere[1], data were: “…72 scores [integers from 1 to 14] obtained from 72 different subjects, organized into twelve groups of six. Columns represent the four levels of factor A; rows represent the three levels of factor B.” (p. 71). Prior analysis of these data using a 4 (Factor A) x 3 (Factor B) between-groups factorial ANOVA revealed statistically significant main effects of Factors A [$F(3,60)=2.94$, $p<0.0403$] and B [$F(2,60)=5.78$, $p<0.0051$], but there was no statistically significant interaction between Factors A and B [$F(6,60)=0.46$, $p<0.8339$] (p. 80). Discussion did not consider follow-up analyses required to disentangle main effects, nor consider strength (“ecological significance”) of the effects or potential cross-generalizability of the findings.[2]

For univariable main effect (one-way ANOVA) applications, the between-subjects factor may be treated as a (multi)categorical class variable to be discriminated using ODA: scores are treated as an ordered attribute, and the objective is to accurately model observations’ class membership.[3-6]

For applications involving two or more between-subjects factors (and also for one-way designs), novometric multiple regression analogue methodology is used: score is treated as an ordered class variable, and is discriminated by GO-CTA using Factors A and B (multicategorical attributes) and their interaction .[7-9] In this approach the objective is to accurately model observations’ scores.

Summarized in Table 1, statistically significant models were obtained for only three scores, and all three CTA models involved only a single node (models were constrained to have identical training and LOO accuracy).

Table 1: Findings of CTA Analyses Predicting Score

		p≤	
Model	ESS	Training	LOO
Predict Score≤3 if Factor B=1	70.6	0.030	0.011
Predict Score≤8 if Factor B=1	31.9	0.017	0.005
Predict Score≤12 if Factor A=2, 3, or 4	79.4	0.012	0.003

2155-0182/10/$3.00

As seen, the third model in Table 1 has the greatest ESS (which is classified as a strong effect[3]) of all three statistically tenable, equally complex ODA models, and thus is the globally optimal (GO) model in this application.

Table 2: Confusion Matrix for GO Model Predicting Score≤12

		Predicted Score ≤12	Predicted Score >12	Sensitivity
Actual Score	≤12	54	14	79.4
	>12	0	4	100.0
Predictive Value		100.0	22.2	

In this application scores >12 occur in 5.6% of the sample, yet the GO model yields a positive predictive value of 22.2% when predicting that an individual's score will be greater than 12 (Table 2). If the cost or return of such high scores is important then such a model may have real-world value.

The confusion matrix for the model with second-highest ESS (70.6), predicting score≤3, is presented in Table 3.

Table 3: Confusion Matrix for Model Predicting Score≤3

		Predicted Score <3	Predicted Score >3	Sensitivity
Actual Score	≤3	4	0	100.0
	>3	20	48	70.6
Predictive Value		16.7	100.0	

In this application scores ≤3 occur in 5.6% of the sample, yet the GO model yields a negative predictive value of 16.7% when predicting that an individual's score will be 3 or less. If the cost or return of such low scores is important then such a model may have real-world value.

References

[1]Collyer CE, Enns JT (1987). *Analysis of variance: The basic designs*. Chicago, IL: Nelson-Hall.

[2]Weinfurt KP (2005). Multivariate analysis of variance. In: LG Grimm & PR Yarnold (Eds.), *Reading and Understanding Multivariate Statistics*. Washington, DC: APA Books, pp. 245-276.

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

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

[5]Yarnold PR (2017). What is optimal data analysis? *Optimal Data Analysis*, *6*, 26-42.

[6]Yarnold PR (2018). ANOVA with one between-groups factor *vs*. novometric analysis. *Optimal Data Analysis*, *7*, 23-25.

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

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

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

Author Notes

This article uses publically-available data and is exempt from IRB review. No conflict of interest was reported.