Optimal Data Analysis
Vol. 7 (March 18, 2018), 23-25

2155-0182/10/$3.00

ANOVA with One Between-Groups Factor *vs*. Novometric Analysis

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

An integer attribute (dependent) variable is compared between four categories of a class (independent) variable using ANOVA with one between-groups factor, as well as novometric analogues to ANOVA (to predict class status) and to multiple regression (to predict score).

Described elsewhere[1], data used herein were: "…72 numerical pieces of data [integers from 1 to 14], each obtained from a different subject, arranged in four groups…" (p. 52). Prior analysis of these data by omnibus (all four groups) one-factor between-groups ANOVA revealed $F(3,68)=2.69$, $p<0.0531$ (p. 59). Follow-up analyses required to disentangle an omnibus effect, strength ("ecological significance") of the effect, and potential cross-generalizability of the finding were not discussed.[2]

Using ODA the four-category between-subjects factor is treated as a multicategorical class variable which is to be discriminated, and the "numerical data" is treated as an ordered attribute to be used to discriminate (i.e., to be compared between) class categories.[3,4] Novometric analysis requires the identification of all of the statistically viable models which exist for the sample.[4,5] Table 1 outlines the analyses which are needed to identify the globally-optimal model in this application.

Table 1

Non-Directional Comparisons of Four Nominal Class Categories (1, 2, 3, 4)

Three Inequalities	Two Inequalities	One Inequality
$1 \neq 2 \neq 3 \neq 4$	$(1 = 2) \neq 3 \neq 4$	$(1 = 2) \neq (3 = 4)$
	$(1 = 3) \neq 2 \neq 4$	$(1 = 3) \neq (2 = 4)$
	$(1 = 4) \neq 2 \neq 3$	$(1 = 4) \neq (2 = 3)$
	$(2 = 3) \neq 1 \neq 4$	$1 \neq (2 = 3 = 4)$
	$(2 = 4) \neq 1 \neq 3$	$2 \neq (1 = 3 = 4)$
	$(3 = 4) \neq 1 \neq 2$	$3 \neq (1 = 2 = 4)$
		$4 \neq (1 = 2 = 3)$

2155-0182/10/$3.00

Table 2: Performance Statistics for ODA Models Described in Table 1

	Training		LOO	
Model	ESS	$p <$	ESS	$p <$
$1 \neq 2 \neq 3 \neq 4$	20.4	0.422	---	---
$(1 = 2) \neq 3 \neq 4$	22.2	0.257	---	---
$(1 = 3) \neq 2 \neq 4$	19.4	0.418	---	---
$(1 = 4) \neq 2 \neq 3$	11.1	0.983	---	---
$(2 = 3) \neq 1 \neq 4$	23.6	0.177	---	---
$(2 = 4) \neq 1 \neq 3$	19.4	0.437	---	---
$(3 = 4) \neq 1 \neq 2$	20.8	0.329	---	---
$(1 = 2) \neq (3 = 4)$	25.0	0.119	---	---
$(1 = 3) \neq (2 = 4)$	22.2	0.209	---	---
$(1 = 4) \neq (2 = 3)$	11.1	0.878	---	---
$1 \neq (2 = 3 = 4)$	24.1	0.201	---	---
$2 \neq (1 = 3 = 4)$	9.3	0.963	---	---
$3 \neq (1 = 2 = 4)$	7.4	0.999	---	---
$4 \neq (1 = 2 = 3)$	29.6	0.091	29.6	0.019

Table 2 gives ODA model performance statistics for each analysis indicated in Table 1. Except for the final entry in Table 2, all models evaluated had relatively weak ESS (defined[3,4] as ESS<25), and were statistically insignificant (i.e., p>0.10) so LOO analysis was not conducted. For the final model in Table 2, which contrasted group 4 *vs*. the other three groups combined, the ODA model was:

If Score≤6.5 then predict Group=4;

If Score>6.5 then predict Group=1, 2, or 3.

Based on 1,000,000 Monte Carlo experiments, estimated p<0.09721, confidence for p< 0.10=100%. Table 3 presents the confusion matrix for this model, which correctly classified 50% of Group 4 observations (exactly what is expected by chance[3]) and 79.6% of others. This level of accuracy reflects moderate ESS=29.6. The model yielded identical performance in LOO analysis (p<0.019), suggesting findings may cross-generalize to other groups involving different subjects.

Table 3: Confusion Matrix for ODA Model

		Predicted Group 4 or Groups 1-3		
		1-3	4	Sensitivity
Actual Group	1-3	43	11	79.6
Membership	4	9	9	50.0
Predictive Value		82.7	45.0	

Novometric Regression Analogue

Respecting tradition[2], the novometric ANOVA analogue discussed above discriminates group membership status (class category) on the basis of score (ordered attribute). In this approach the objective is to accurately model the group to which individual observations belong.

The mirror image of this approach is the novometric multiple regression analogue, which discriminates score (class category) on the basis of group membership status (multicategorical attribute).[4-8] In this approach the objective is to accurately model individual's scores.

2155-0182/10/$3.00

Table 4: Findings of Novometric Multiple Regression Analogue Analysis

	Training		LOO	
Model	ESS	$p\leq$	ESS	$p\leq$
Predict Score≤4 if Group=4	34.9	0.176	---	---
Predict Score≤5 if Group=3 or 4	29.5	0.224	---	---
Predict Score≤6 if Group=2 or 4	27.7	0.206	---	---
Predict Score≤7 if Group=3 or 4	26.0	0.163	---	---
Predict Score≤8 if Group=3 or 4	16.9	0.587	---	---
Predict Score≤9 if Group=3 or 4	13.1	0.828	---	---
Predict Score≤10 if Group=4	14.7	0.740	---	---
Predict Score≤11 if Group=2, 3, or 4	35.6	0.340	---	---
Predict Score≤12 if Group=2, 3, or 4	79.4	0.012	79.4	0.003

Table 5: Confusion Matrix for ODA Model

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

In the present sample scores greater than 12 occur in 4/(4+68) or 5.6% of the sample, but the final model in Table 4 has a positive predictive value of 22.2% when it predicts that an individual's score will be greater than 12. If the weight (cost or return) of extremely high scores (i.e., scores>12 represent the upper 5.6% of the distribution presently) were important in this application, then this model may have merit.

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

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

[8]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.