

UniODA vs. Doubly Incomplete Three-Factor ANOVA: Production Failure Attributable to Acid Corrosion

Paul R. Yarnold, Ph.D.

Optimal Data Analysis, LLC

Production units for concentrating dilute acid are subject to failure as a result of corrosion.¹ Seven units were selected from each of nine factories representing three groups according to the type of acid being concentrated. Productivity before failure was compared between type of acid and factory using doubly incomplete three-factor ANOVA that considers mean productivity across units, versus using UniODA that evaluates productivity of individual units.

The ANOVA unfolded in two steps. First, it was assessed whether differences exist between the factories within an acid group: no statistically significant effect emerged. Second, therefore, data were pooled across factory within each super-block (type of acid). ANOVA revealed that the mean number of units produced before unit failure differed as a function of the type of acid being concentrated: $F(2,60) = 45.6$, $p < 0.0001$. Follow-up tests required to elucidate the exact nature of the effect weren't reported.

This two-step procedure was conducted using UniODA to assess the actual production of individual units, rather than mean production.

First a nondirectional UniODA model was conducted to discriminate factory on the basis of production separately for each acid type (of course, directional hypotheses may be tested and individual weights—for example, cost of failure measured in dollars, may also be used).^{2,3} The UniODA model discriminating factory by production within each acid type was obtained

using the UniODA^{2,3} and MegaODA⁴⁻⁶ software syntax shown below:

```
OPEN acid.dat;  
OUTPUT acid.out;  
VARS acid factory units;  
CLASS factory;  
ATTRIBUTE units;  
MCARLO ITER 25000;  
IN acid=1;GO;  
IN acid=2;GO;  
IN acid=3;GO;
```

As for ANOVA analysis, no statistically significant between-factory effects emerged (p 's > 0.31) within the different acid types. However the effects were of moderate strength ($35.7 \leq ESS \leq 42.9$): the $N = 7$ units sampled within each factory failed to provide sufficient statistical power in this application.^{2,7}

As in the ANOVA analysis, data were pooled within each super-block (across factory), and the UniODA model discriminating acid type by production was obtained by appending the UniODA and MegaODA software syntax above as shown below:

CLASS acid;
 ATTRIBUTE units;
 LOO;
 GO;

The UniODA model was: if productivity ≤ 47.5 units, then predict that acid = 1; if $47.5 < \text{productivity} \leq 69.0$ units, predict that acid = 2; and if productivity > 69.0 units, then predict that acid = 3. As for ANOVA analysis, a statistically significant effect emerged: $p < 0.0001$. While the effect was relatively strong effect in training analysis ($ESS = 61.9$), effect size declined (but remained relatively strong) in LOO analysis ($ESS = 54.8$). Table 1 gives the confusion table for the training model.^{2,3}

Table 1: Confusion Table for UniODA Model Discriminating Acid Type by Unit Productivity

		Predicted Acid Type			Sensitivity
		<u>1</u>	<u>2</u>	<u>3</u>	
<u>Actual</u>	<u>1</u>	19	2	0	90.5%
<u>Acid</u>	<u>2</u>	9	8	4	38.1%
<u>Type</u>	<u>3</u>	0	1	21	95.2%
<u>Predictive</u>	<u>Value</u>	67.9%	72.3%	83.3%	

As seen, model accuracy was close to perfect (sensitivity = 100%) for prediction of acid types 1 and 3, but was only marginally greater than expected by chance (33.3%) for acid type 2. The productivity in acid types 1 and 3 is extremely homogeneous with respect to the UniODA-defined thresholds. In contrast, the productivity of individual production units is heterogeneous for acid type 2, and is skewed toward lower productivity.

Finally, two exploratory UniODA analyses were conducted to discriminate acid types 1 and 2, and acid types 2 and 3, and were accomplished using the following appended UniODA and MegaODA software syntax:

EX acid=3;GO;
 EX acid=1;GO;

For the comparison of acid types 2 and 3 the UniODA model was: if productivity ≤ 71.0 , then predict acid = 2; otherwise predict acid = 3 ($p < 0.0001$, ESS in training = 81.0, ESS in LOO = 76.2). For the comparison of acid types 1 and 2 the UniODA model was: if productivity ≤ 47.5 , then predict acid = 1; otherwise predict acid = 2 ($p < 0.013$, ESS in training = 47.6, ESS in LOO = 38.1). Clearly the productivity for acid type 3 was greatest, and for acid type 1 was lowest, although the productivity for acid types 1 and 2 showed a moderate degree of overlap.

References

¹Brownlee KA (1953). *Industrial experimentation*. New York, NY: Chemical Publishing Company (pp. 209-111).

²Yarnold PR, Soltysik RC (In Review). *Maximizing predictive accuracy*. Chicago, IL: ODA Books.

³Yarnold PR, Soltysik RC (2005). *Optimal data analysis: A guidebook with software for Windows*. Washington, DC: APA Books.

⁴Soltysik RC, Yarnold PR (2013). MegaODA large sample and BIG DATA time trials: Separating the chaff. *Optimal Data Analysis*, 2, 194-197. URL: <http://optimalprediction.com/files/pdf/V2A29.pdf>

⁵Soltysik RC, Yarnold PR (2013). MegaODA large sample and BIG DATA time trials: Harvesting the Wheat. *Optimal Data Analysis*, 2, 202-205. URL:
<http://optimalprediction.com/files/pdf/V2A31.pdf>

⁶Yarnold PR, Soltysik RC (2013). MegaODA large sample and BIG DATA time trials: Maximum velocity analysis. *Optimal Data Analysis*, 2, 220-221. URL:
<http://optimalprediction.com/files/pdf/V2A35.pdf>

⁷Soltysik RC, Yarnold PR (2013). Statistical power of optimal discrimination with one attribute and two classes: One-tailed hypotheses. *Optimal Data Analysis*, 2, 26-30. URL:
<http://optimalprediction.com/files/pdf/V2A4.pdf>

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