Optimal Data Analysis
Vol. 2, Release 2 (November 19, 2013), 198-201

2155-0182/10/$3.00

ODA Range Test *vs*. One-Way Analysis of Variance: Comparing Strength of Alternative Line Connections

Paul R. Yarnold, Ph.D., and Gordon C. Brofft, BS
Optimal Data Analysis, LLC Marine and Water Consultant

Among the most popular conventional statistical methods, Student's *t*-test is used to compare the means of two groups on a single dependent measure assessed on a continuous scale. When three or more groups are compared, *t*-test is generalized to one-way analysis of variance (ANOVA). If the F statistic associated with the overall or "omnibus" effect is statistically reliable, then a range test that is more efficient than performing all possible comparisons is used to ascertain the exact nature of interclass differences. In contrast, the dependent measure may be compared between classes via UniODA, to assess if thresholds on the dependent measure separate the classes. If the resulting ESS accuracy statistic for the omnibus effect is statistically reliable, then a recently-developed optimal range test is used to assess the exact nature of interclass differences. ANOVA and UniODA are used to compare three methods commonly used in competitive big-game sport fishing for connecting segments of fishing line. Similarities and differences of parametric GLM and non-parametric ODA methods are demonstrated.

Recent research compared the strength of the Double Uni knot used to attach 40-pound-test monofilament line to: a) 50-pound-test solid spectra; b) 60-pound-test hollow spectra; and c) 65-pound-test solid spectra (modal reel-backing selections).[1] The three different combinations a-c were dummy-coded as being from class 1, 2, and 3, respectively. The free end of monofilament line was pulled with increasing force until the connection failed, at which point pounds of force (the attribute) was recorded. The resulting data are presented in Table 1.

Table 1: Pounds of Force Required to Break Each Type of Line-to-Line Connection

	Class 1	Class 2	Class 3
n	14	20	13
Mean	23.4	22.5	25.6
SD	4.13	2.35	3.25
Median	23	23	25
Skewness	0.13	-0.32	0.72
Kurtosis	-1.72	-0.44	0.08

2155-0182/10/$3.00

One-Way ANOVA

Eyeball analysis suggests classes 1 and 2 are essentially the same with respect to pounds of force. Though the mean for class 3 is greater, mean differences are less than the SD and the number of knots evaluated is modest, so any statistical effect identified must be weak.

Evaluated using the general linear model (GLM) paradigm the *post hoc* hypothesis that the three different connections fail at different levels of force is tested via one-way analysis of variance (ANOVA), which compares mean differences between the three classes.[2] Using this method all classes are simultaneously compared using an "omnibus" test. If the omnibus effect is statistically significant a multiple-comparisons procedure (MCP) is subsequently employed to understand all the interclass differences.[2] MCPs disentangle omnibus effects efficiently when compared to conducting all possible pairwise comparisons, and therefore have lower experimentwise Type I error rates, especially when the number of class categories is large (at least three class categories are needed to conduct a MCP).[3]

Presently the omnibus effect of class was statistically significant[4] at the generalized[5] per-comparison criterion: $F(2,44)=3.8$, $p<0.032$. The $R^2=14.6$ indicates the GLM model accounts for one-seventh of the variation in knot strength that was observed between classes.[2] Conducting all possible pairwise comparisons adds three tests of statistical hypotheses (the Sidak criterion[5] for experimentwise $p<0.05$ with a total of four tests of statistical hypotheses is $p<0.01275$). Planned pairwise contrasts[4] found class 1 and 2 means were not significantly different ($t=0.8$, $p<0.42$); class 1 and 3 means differed marginally by the generalized criterion ($t=1.8$, $p<0.09$); and class 2 and 3 means were significantly different at the experimentwise criterion ($t=2.7$, $p<0.0092$). In summary, analysis using one-way ANOVA and all possible pairwise comparisons to disentangle the statistically reliable omnibus effect indicated that class 3 knots required greater mean force to break than class 1 and class 2 knots, which had statistically comparable mean knot strength.

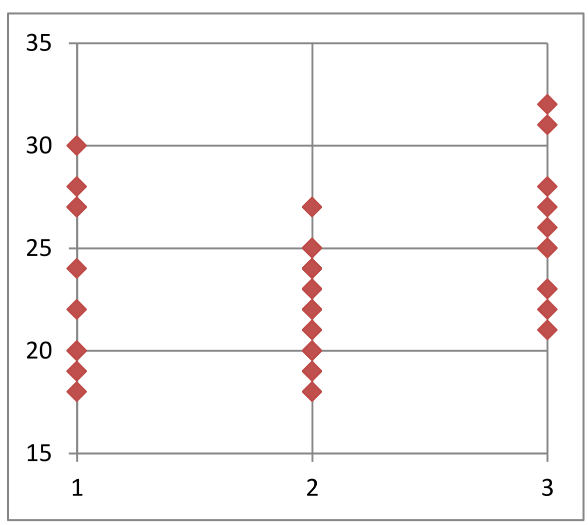
The omnibus effect was disentangled by several MCPs. The MCPs that examined overlapping 95% confidence intervals (CIs) reported no significant pairwise effects (including 95% CIs for Sheffe, *t*-test, the studentized maximum modulus, and Sidak *t*-test). All range test-based MCPs reported class 3 means are greater than class 2 means, with class 1 means intermediate and not significantly different than class 2 or 3 means (including Sidak *t*-tests, Sheffe test, LSD *t*-test, Duncan multiple range test; Ryan-Einot-Gabriel-Welsch multiple range test, and Tukey's studentized HSD range test). MCPs index *p* but not the strength of observed differences.[2]

While determining the appropriate MCP to use in a specific application is a problem for researchers using parametric analyses, no such ambiguity exists in the ODA paradigm.

Optimal Range Test

Eyeball examination of descriptive data in Table 1 presents little information concerning the potential discriminability of the three class categories. Instead raw data presented as seen in Figure 1 helps visualize interclass differences.

Figure 1: Raw Knot Strength Data by Class



As seen relatively few class 2 data points are 25 pounds or greater, versus many class 3 data points, indicating these two class categories are discriminable. A similar but weaker pattern exists for the comparison of classes 1 and 3.

Presently the omnibus effect of class was statistically significant[6] at the generalized per-comparison criterion: estimated $p<0.015$, with confidence for $p<0.05$ of >99.9% in 400 Monte Carlo iterations. The model achieved moderate accuracy in classifying the classes (ESS=36.0), and yielded moderate accuracy when making classification decisions (ESP=35.6).[7] The model was: if strength≤20 pounds then predict class=1; if 20 pounds<strength≤24.5 pounds then predict class=2; and if strength>24.5 pounds then predict class=3. The resulting confusion model is presented in Table 2.

Table 2: Initial Omnibus ODA Model Confusion Table

		Predicted Class			
		1	2	3	
Actual	1	6	2	6	42.9%
Class	2	5	12	3	60.0%
	3	0	4	9	69.2%
		54.6%	66.7%	50.0%	

The effect is symbolically represented as follows: 1 ≤ 2 ≤ 3: as seen, resolving remaining ambiguity requires determining if either equality sign should be changed to strict inequality. The initial step of this optimal range test starts at the left-hand side of the current model, and the first pairwise comparison 1≤ 2 is performed with the Monte Carlo simulation parameterized for two tests of statistical hypotheses (the initial analysis plus the present analysis).[3,6] For this analysis: estimated $p<0.41$ with >99.99% confidence for $p>0.10$; ESS=17.9; ESP=19.8. The symbolic notation is updated: (1 = 2) ≤ 3.

The final step of this optimal range test involves resolving the remaining equality. This is accomplished by combining class categories 1 and 2 (presently this was done making a copy of the data set, and manually converting the class 1 codes into class 2 codes) and then running the directional UniODA analysis assessing if class 3 is greater than the combined class 1 and class 2. For this analysis: estimated $p<0.0123$; >99.9% confidence for $p<0.017$ (the Sidak criterion for three tests of statistical hypotheses and experimentwise $p<0.05$); ESS=42.8; ESP=36.2. The symbolic representation is finalized: (1=2)<3: that is, class 1 and 2 are statistically comparable and have knot strengths significantly lower than knot strengths observed in class 3. The model was: if strength≤24.5 pounds then predict class= 1 and 2; and if strength>24.5 pounds then predict class=3. The resulting confusion model is presented in Table 3.

Table 3: Final ODA Model Confusion Table

		Predicted Class		
		1&2	3	
Actual	1&2	25	9	73.5%
Class	3	4	9	69.2%
		86.2%	50.0%	

Note that exactly the same conclusion would be reached had the range test initiated from the right-hand side of the initial model. In the first step of the range test the right-hand inequality would be evaluated (here, estimated $p<0.0001$), and the symbolic notation would be amended as: (1≤2) < 3. The second and final step of the range test would evaluate the remaining inequality, yielding the identical final model. Also, the same conclusion would have been reached in the present example if, instead of the range test, all possible comparisons were performed and then integrated.[3] The efficiency of range tests versus all possible comparisons in this application begins significant divergence if at least four class categories are to be compared. In this application all possible comparisons and the optimal range test yielded the same results. However, as the number of tests of statistical hypotheses increases, the Sidak criterion for experimentwise $p<0.05$ also increases, and thus different results can accrue as a consequence of divergent statistical power.[3]

References

[1]Yarnold PR. Brofft GC (2013). Comparing knot strength using UniODA. *Optimal Data Analysis*, *2*, 54-59.

[2]Levine DM, Krehbiel TC, Berenson ML (2010). *Business statistics: A first course*. New Jersey: Prentice Hall.

[3]Yarnold PR (2013). Univariate and multivariate analysis of categorical attributes having many response categories. *Optimal Data Analysis*, 2, 177-190.

[4]SAS™ code used to conduct this analysis was (control commands indicated in red):

```
data q;
infile 'c: knots.txt';
input knot pounds;
proc glm;
  class knot;
    model pounds=knot;
      means knot / sidak regwq
      scheffe tukey cldiff clm
      duncan gabriel gt2 lsd
      welch lines;
    estimate 'c1 vs c2' knot 1 -1 0;
    estimate 'c1 vs c3' knot 1 0 -1;
    estimate 'c2 vs c3' knot 0 1 -1;
run;
```

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

[6]UniODA™ and MegaODA™ code employed to conduct this analysis was:

```
open knots.dat;
output knots.out;
vars knot pounds;
class knot;
attr pounds;
mc iter 10000 target .05 sidak 1 stop 99.9;
go;
```

The following code is used to conduct all possible comparisons:

```
ex knot=3;
dir < 1 2;
mc iter 10000 target .05 sidak 4 stop 99.9;
go;
ex knot=1;
dir < 2 3;
go;
ex knot=2;
dir < 1 3;
go;
```

[7]Yarnold PR (2013). Standards for reporting UniODA findings expanded to include ESP and all possible aggregated confusion tables. *Optimal Data Analysis*, 2, 106-119.

Author Notes

E-mail: Journal@OptimalDataAnalysis.com

Address: 1220 Rosecrans St., #330
San Diego, CA 82106

ODA Blog: http://odajournal.com