Optimal Data Analysis
Vol. 4 (December 26, 2014), 6-8

2155-0182/10/$3.00

UniODA *vs.* Mann-Whitney *U* Test: Comparative Effectiveness of Laxatives

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

Compared to the Mann-Whitney *U* test, UniODA explicitly maximizes model accuracy for a specific sample and hypothesis; identifies the optimal threshold discriminating the groups; yields invariant findings over monotonic transformations of the data; indexes model effect strength on an absolute scale ranging from 0 (accuracy expected for the sample and hypothesis by chance) to 100 (perfect accuracy); and estimates cross-generalizability of the findings if the model is applied to classify an independent random sample.

The Mann-Whitney *U* test is also known as the Mann–Whitney–Wilcoxon (MWW) test, the Wilcoxon rank-sum test, the Wilcoxon–Mann–Whitney test, and Kendall's S, and in the presence of ties *U* is equivalent to a chi-square test for trend. *U* is a non-parametric test for which the null hypothesis is that the two populations are the same with respect to values obtained on an ordered variable.[1]

In the present paper the use of *U* and of UniODA is illustrated in an application comparing the effectiveness rating (Likert-type score ranging from 0="very ineffective" to 10="very effective") of two different types of laxatives (the class variable).[2] The *U*-test was judged appropriate because of the small sample sizes (Table 1). Identical effectiveness ratings share the same rank and were assigned the mean of the corresponding rank values.

Table 1: Laxative Type, Effectiveness Rating, and Corresponding Rank Score

Type	Rating	Rank
1	3	3
1	4	4
1	2	1.5
1	6	7.5
1	2	1.5
1	5	5.5
2	9	11
2	7	9
2	5	5.5
2	10	12
2	6	7.5
2	8	10

2155-0182/10/$3.00

Findings of analysis using *U* with these data indicated: "The difference that we have found between the ratings for the two laxatives is unlikely to have occurred by chance ($p<0.05$). It looks as if participants' assessments of the laxative's effectiveness do indeed differ. Inspection of the medians suggests one brand is rated as being more effective than the other brand. However we initially predicted that there would be *some kind of difference* between the two laxatives, and not which brand would be better than the other brand: we have therefore conducted a non-directional, two-tailed test, and strictly speaking all we can conclude from it is that the two laxatives differ in rated effectiveness" (p. 5). No estimate of the cross-generalizability of the model if used to classify an independent random sample was conducted.

Exploratory ("two-tailed") analysis via UniODA was conducted next, comparing the effectiveness ratings of the two types of laxatives.[3-6] The UniODA model was: if rating $\leq$4.5 then predict that type=1; otherwise predict that type=2. This model correctly classified all 6 (100%) of type 2 laxatives and 4 of 6 (66.7%) of type 1 laxatives. This level of accuracy was NOT statistically significant ($p<0.29$; 100% certainty that $p>0.10$), the effect was relatively strong[7] on the basis of the Effect Strength for Sensitivity (ESS) statistic of 66.7. Model performance declined in LOO validity analysis (4/6 of type 2 laxatives were correctly classified, ESS=33.3), suggesting that the finding is unlikely to cross-generalize with comparable strength if the model is used to classify an independent random sample.

Rank score was compared between the two groups next. The UniODA model was: if rank score $\leq$4.75 then predict that type=1; otherwise predict that type=2. This model yielded identical classification accuracy for each laxative type, ESS, Type I error, and LOO performance as was obtained for the analysis of effectiveness ratings: UniODA is invariant over a monotonic transformation of the attribute (for UniODA no transformation of raw data into ranks is necessary, as is required by *U*).[1,3]

A directional (confirmatory) analysis was conducted next[8] (for expository purposes, laxative type 2 was hypothesized to have greater effectiveness ratings and rankings than laxative type 1), and revealed a statistically marginal effect ($p<0.073$), with comparable accuracy, ESS, and LOO (the model is unlikely to cross-generalize to an independent random sample) results obtained by the exploratory model.

While *U* indicated that the median effectiveness ranking of laxative types 1 and 2 differ, failure to conduct cross-generalizability analysis did NOT reveal that this effect is unlikely to hold-up in an attempted replication. UniODA identifies statistically reliable effects in small samples when the classification accuracy is very strong and is validated in cross-generalizability analysis—and thus is likely to reflect a reproducible finding.[1,9-11]

References

[1]Yarnold PR (2014). UniODA *vs*. Mann-Whitney *U* test: Sunlight and Petal Width. *Optimal Data Analysis*, *4*, 3-5.

[2]http://www.sussex.ac.uk/Users/grahamh/RM1web/MannWhitneyHandout%202011.pdf

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

[4]Yarnold PR, Soltysik RC (2010). Optimal data analysis: A general statistical analysis paradigm. *Optimal Data Analysis*, *1*, 10-22.

[5]Bryant FB, Harrison PR (2013). How to create an ASCII input data file for UniODA and CTA software. *Optimal Data Analysis*, *2*, 2-6.

[6]The exploratory ("two-tailed") UniODA program[3] used to find and evaluate the optimal (maximum-accuracy) models for rating and rank

2155-0182/10/$3.00

data (using 25,000 Monte Carlo experiments to estimate exact Type I error), including their leave-one-out (LOO, a one-sample jackknife) validity analysis, was:

```
VARS type rating rank;
CLASS type;
ATTR rating rank;
MC ITER 25000;
LOO;
GO;
```

[7]By convention, ESS<25 is a relatively weak effect; ESS<50 is a moderate effect; ESS<75 is a relatively strong effect; and ESS≥75 is a very strong effect.[6]

[8]The confirmatory ("one-tailed") UniODA program[3] was identical to the exploratory model[6], except that the directional hypothesis was specified:

```
DIR < 1 2;
```

[9]Soltysik RC, Yarnold PR (2013). Statistical power of optimal discrimination with one attribute and two classes: One-tailed hypotheses. *Optimal Data Analysis, 2*, 26-30.

[10]Yarnold PR (2013). Percent oil-based energy consumption and average percent GDP growth: A small sample UniODA analysis. *Optimal Data Analysis, 2*, 60-61.

[11]Yarnold PR (2013). UniODA and small samples. *Optimal Data Analysis, 2*, 71.

Author Notes

E-mail: Journal@OptimalDataAnalysis.com.

Mail Address: Optimal Data Analysis LLC; 6238 N. Milwaukee Ave.; Chicago, IL 60646.