Optimal Data Analysis
Vol. 4 (December 9, 2014), 3-5

2155-0182/10/$3.00

UniODA *vs*. Mann-Whitney *U* Test: Sunlight and Petal Width

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

The Mann-Whitney *U* test is a nonparametric statistical test used to compare samples of scores obtained on an ordered variable between two independent groups. In contrast to *U* the use of UniODA in this application *explicitly maximizes model accuracy* for each specific sample and hypothesis; *explicitly identifies the optimal threshold* for discriminating the two groups; yields findings that are *invariant over monotonic transformations* of the data; and expresses effect strength *on an absolute scale* ranging from 0 (accuracy expected for the sample and hypothesis by chance) to 100 (perfect accuracy).

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. A non-directional (exploratory, two-sided) alternative hypothesis is that one population tends to have different values than the other, and a directional (confirmatory, one-sided) alternative hypothesis is that one population tends to have larger (or smaller) values than the other. *U* is more efficient than *t*-test for non-normal distributions, and is nearly as efficient for normal distributions. Assumptions underlying use of the *U* test include random samples from populations; all observations from both groups are independent of each other; responses are ordinal (i.e. one can say, of any two observations, which is greater); distributions of both groups are equal under the null hypothesis, so the probability of an observation from one population exceeding an observation from the second population is symmetric between populations; and under the alternative hypothesis, the probability of an observation from one population exceeding an observation from the other population (after exclusion of ties) is not 0.5 (an exploratory hypothesis) or is significantly greater (or lesser) than 0.5 (a confirmatory hypothesis).[1,2] The theta statistic [U/(n1*n2)] is used as an index of effect size that is equivalent to the area under the receiver operating characteristic (ROC) curve.[3,4]

Use of *U* and UniODA is demonstrated for an application comparing size (cm at widest point) of bramble bush leaves growing in full sunlight *versus* growing in the shade (the class variable). The Mann-Whitney *U*-test was judged as being appropriate because the sample sizes

2155-0182/10/$3.00

are too small to determine if the samples reflect normally distributed data (Table 1).[5] Note that identical petal width values share the same rank and are identified using the mean of the corresponding rank values.

Table 1: Number of Sunlight and Shade Leaves per Petal Width: Raw (cm) and Rank Score

Petal Width	N Sunlight	N Shade	Rank Score
4.1	1		1
4.5	1		2
4.8	1		3
5.1	2		4.5
5.3	1		6
5.5	1	3	8.5
5.9		1	11
6.0	1		12
6.3		1	13
6.5		1	14
6.8		1	15
7.2		1	16

Findings of analysis using *U* with these data indicated: "Reject the Null Hypothesis if the smallest value of $\mathbf{U_1}$ (58.5) or $\mathbf{U_2}$ (5.5) is below $\mathbf{U_{crit}}$. In this case $\mathbf{U_2}$ is below 13 so we reject the Null Hypothesis and accept the Alternative Hypothesis. The difference between the size of the bramble leaves in the light and the dark is significant for P>0.05."[5] Thus the difference found in leaf size for the two light conditions is unlikely to have occurred by chance. Inspecting medians suggests that shade leaves are larger than sunlight leaves. However it was initially predicted there would be some kind of difference between the size of the two types of leaves, not that shade leaves would be larger than sunlight leaves. Having conducted a non-directional, two-tailed test, strictly speaking all one can conclude from it is that the two types of leaves differ in ranked width.

Analysis using UniODA requires a reconfigured data matrix (Table 2).[6-9]

Table 2: Number of Sunlight and Shade Leaves per Petal Width: UniODA Data Set

Class (Group)	Petal Width	Rank Score
1	4.1	1
1	4.5	2
1	4.8	3
1	5.1	4.5
1	5.1	4.5
1	5.3	6
0	5.5	8.5
0	5.5	8.5
0	5.5	8.5
1	5.5	8.5
0	5.9	11
1	6.0	12
0	6.3	13
0	6.5	14
0	6.8	15
0	7.2	16

Petal width was compared between the two groups first. The UniODA model was: if width ≤5.4 cm then predict that group=sunlight; otherwise if width >5.4 cm then predict that group=shade. This model correctly classified all 8 (100%) of the leaves from the Shade group and 6 of 8 (75%) of the leaves from the Sunshine group. This level of accuracy was statisti-

2155-0182/10/$3.00

cally significant (p<0.012; 100% certainty that p<0.05; 99.5% certainty that p>0.01), and corresponded to a very strong effect[10] on the basis of the Effect Strength for Sensitivity (ESS) statistic of 75.0 (0=accuracy expected by chance; 100= perfect accuracy).[6] Model performance was stable in LOO validity analysis, suggesting the finding may cross-generalize if the model is used to classify an independent random sample of bramble bush leaves.

Rank score was compared between the two groups next. The UniODA model was: if rank score ≤7.25 then predict that group=Sunlight; otherwise if rank score >7.25 then predict that group=Shade. This model correctly classified all 8 (100%) of the leaves from the Shade group and 6 of 8 (75%) of the leaves from the Sunshine group. This level of accuracy was statistically significant (p<0.013; 100% certainty that p<0.05; 99.9% certainty that p>0.01), and corresponded to a very strong effect: ESS= 75.0. Stable LOO validity analysis performance suggests the model may cross-generalize if used to classify an independent random sample of bramble bush leaves.

The performance of the UniODA models for width and rank score attributes was identical because UniODA is invariant over a monotonic transformation of the attribute.[6] Therefore, in these applications, when using UniODA no transformation of raw data into ranks is necessary, as is required by *U*.

References

[1]Mann HB; Whitney DR (1947). On a test of whether one of two random variables is stochastically larger than the other". *Annals of Mathematical Statistics*, *18*, 50–60.

[2]http://www.statsdirect.com/help/default.htm#nonparametric_methods/mann_whitney.htm

[3]http://onlinelibrary.wiley.com/doi/10.1002/sim.2323/abstract

[4]Yarnold PR (2014). UniODA vs. ROC analysis: Computing the "optimal" cut-point. *Optimal Data Analysis*, *3*, 117-120.

[5]http://www.saburchill.com/IBbiology/stats/002.html

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

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

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

[9]The UniODA program[6] used to find and evaluate the optimal (maximum-accuracy) models for width and rank 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 group width rank;
CLASS group;
ATTR width rank;
MC ITER 25,000;
LOO;
GO;
```

[10]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]

Author Notes

E-mail: Journal@OptimalDataAnalysis.com.

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