Practical Assessment, Research, and Evaluation

Volume 23 *Volume 23, 2018* Article 3

2018

A Note on the Assumption of Identical Distributions for Nonparametric Tests of Location

David W. Nordstokke

S. Mitchell Colp

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 23 Number 3, April 2018 ISSN 1531-7714

A Note on the Assumption of Identical Distributions for Nonparametric Tests of Location

David W. Nordstokke, Ph.D., *University of Calgary*
S. Mitchell Colp, Ph.D., *University of Calgary*

Often, when testing for shift in location, researchers will utilize nonparametric statistical tests in place of their parametric counterparts when there is evidence or belief that the assumptions of the parametric test are not met (i.e., normally distributed dependent variables). An underlying and often unattended to assumption of nonparametric tests of location is that of identical distributions. The assumption of identical distributions requires that distributions conform to one another in terms of variability and shape (i.e., variance, skew and kurtosis). The purpose of the current study is to demonstrate, via the use of Monte Carlo simulation, the assumption of identical distribution using the Wilcoxon-Mann-Whitney (WMW) test and the Student t-test for comparison. For each of the conditions, there are several levels of sample size, variance ratio, group sample size ratio, and degree of skew in the parent distribution. Empirical Type I error rates are compared to nominal Type I error rates to determine the validity of the result for each run of the simulation. Violation of the assumption of identical distributions lead to bias in the result of the WMW test and the Student t-test. Practical implications are also discussed.

The purpose of the current paper is to bring to bear an issue that occurs in current statistical practices related to educational, behavioral and social science research. Often, when testing for shift in location (i.e., differences in means or medians), researchers will utilize rank-based nonparametric statistical tests in place of their parametric counterparts when there is evidence or belief that the assumptions of the parametric test have not been met (i.e., normally distributed dependent variables). The utilization of rank-based nonparametric tests have been widely recommended to replace parametric tests for at least 50 years (e.g., Siegel, 1956) and these practices have become entrenched in current statistical methodologies. There is, however, a potential for bias in nonparametric analyses when distributional forms of data are not attended to appropriately in terms of meeting the statistical assumptions related to the test used.

Statistical hypothesis testing models were developed using a location shift model. That is, differences between groups are viewed as differences between the central tendency of the distributions (e.g., means or medians). In order to determine the effectiveness of treatments, an additive shift model was adopted. In an additive shift model, treatment effect, or alternatively effect size, is determined via calculating the distance between the central tendencies of the distributions that are being compared. For example, if X1 represents the mean or median of the control or comparison group and X2 represents the mean or the median of the treatment group, the treatment effect is determined by calculating X1 +/- c; where c represents the distance between X1 and X2. It is implied that the only difference between the groups is their mean or median. Figure 1 provides an illustration of this, where the distribution on the left would be X1 (control or comparison group) and the distribution on the right would be X2 (treatment group). The vertical dotted line in the middle of each of the distributions represents the central tendency (e.g., mean or median – in the case of a normal distribution, the mean and median would be the same). The horizontal arrow indicates the shift in location (i.e., difference between means or medians) and represents the treatment effect.

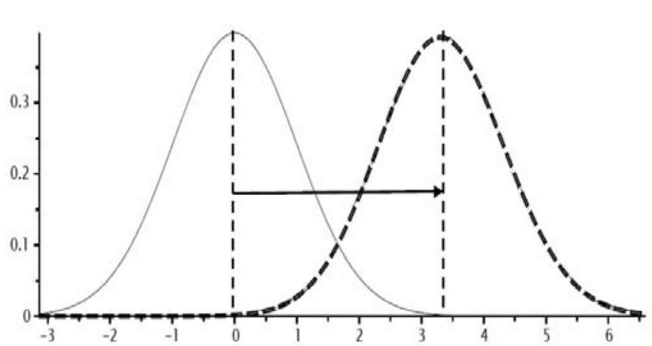

Figure 1. Illustration of shift in location for the two-group case in hypothesis testing

When testing for differences between groups, nonparametric tests come with a set of assumptions related to the nature of the data from where samples are drawn. For nonparametric tests the assumptions are that data are independent and identically distributed. The assumption of independence requires that any data point will have an influence on no other data point. This assumption is a requirement of both parametric and nonparametric tests and will not be discussed further. The assumption of identical distributions will be the focus of this paper. The assumption of identical distributions for nonparametric tests is inherently related to the nature of the distributions of all groups involved and is often misunderstood or not attended to by researchers and data analysts. The misunderstanding relates to the requirements of the distributional features of the data used in the analysis and is often misrepresented or ignored. The assumption of identical distributions states that when comparing samples for differences in their central tendency (i.e., means or medians) they must have identically shaped population distributions in terms of variance, skew, and kurtosis. This assumption applies to all nonparametric test that test for group differences; however, for the purpose of the current simulation study, the Wilcoxon-Mann-Whitney (WMW) test will be used. Traditionally, the WMW is a test of identical distributions; however, if the assumption of identical distributions is applied in conjunction with this test, it can be used as a test of difference in medians, as a rejection of the null hypotheses indicates a difference in the central tendencies of the distributions.

In research, situations can arise where distributional forms are different between groups that are to be compared. In experimental designs, groups are formed by sampling from a single population, thus it is likely that groups will possess very similar distributions; however, this is not guaranteed as treatments may interact with the dependent variable resulting in a change in the treatment group's distribution. This is generally referred to as the Behrens-Fisher problem (Scheffé, 1970). In non-experimental designs, data are not sampled from a single population but instead from two or more populations (e.g., gender), where there may exist substantial domain differences between groups that are sampled from these populations. When using non-experimental approaches, there is an increased likelihood that distributions have different shapes.

In the case where distributions possess equal shapes, the WMW test is generally robust against non-normality and demonstrates high power when compared with other tests (Fagerland & Sandvik, 2009). A potential issue arises when the two distributions used in the WMW test are not identical in terms of shape (i.e., variance, skew, or kurtosis) as this violates the assumption of identical distributions, thus entering bias (i.e., increased Type I and Type II error rates) into the between-groups analysis resulting in tests being either too conservative or too liberal. To exacerbate this issue, many statistical textbooks reviewed by the authors do not state the necessity of the assumption of identical distributions (e.g., Brace et al., 2012; Corder & Foreman, 2009; Aron et al., 2013; Pagano, 2001; Primavera, 2012). Many of these textbooks are designed to cater towards psychology and the behavioral sciences, thus creating generations of researchers who do not fully understand the application or implications of using and interpreting the result of nonparametric tests that may be biased leading to inappropriate inferences based on the result of these tests. It is difficult to determine how frequently the WMW is used in lieu of the Student t-test in daily statistical practice, but it can certainly be argued that it is widely recommended in textbooks.

Thus, the purpose of the current study is to demonstrate, using Monte Carlo simulation, the robustness of the WMW test under conditions where the assumption of identical distribution has been violated to varying degrees. In addition, the parametric counterpart of the WMW, the Student t-test, will be included as a comparison as parametric tests of central tendency are impacted by unequal distributions as well (i.e., equality of variances). It is expected that when both the WMW test and Student t-test are applied to groups with varied distributional forms, the Type I error rate of the test will be impacted.

Method

Data Generation

Standard simulation methodology was utilized to perform the current Monte Carlo simulation (e.g., Nordstokke & Zumbo, 2010; Nordstokke & Colp, 2014; Zimmerman, 1987; 2004). Population distributions were generated and the statistical procedures were carried out using the statistical software package for the social sciences, SPSS 24. A pseudo random number sampling strategy with the initial seed selected randomly was utilized to generate $\chi 2$ distributions. The design of the current investigation is a 7 X 2 X 3 completely crossed design with: (a) seven levels of skew (-3, -2, -1, 0, 1, 2, 3), (b) two levels of total sample size (24, 48), and (c) three levels of sample size ratios (1-1, 2-1, and 3-1). These sample sizes and sample size ratios were selected to provide an array of conditions that attempt to represent real world research conditions. Clearly, every eventuality cannot be simulated; however, utilizing the current design will at least give an indication of a variety of research possibilities. The outcome of interest in this simulation study is the proportion of rejections of the null hypothesis in each cell of the design. That is, the Type I error rate of the test for each given cell in the design. Variances in the current simulation were fixed to be equivalent throughout all of the simulations as the Type I error rates of the two tests under investigation are the outcome of interest.

Statistical tests

The WMW test used in the current set of simulations is calculated as follows. The combined data from both groups are sorted and ranks assigned to all cases, with average rank being used in the case of ties. The test statistic is:

$$Z = \frac{(U - n_1 n_2/2)}{\sqrt{\frac{n_1 n_2}{N(N-1)}\left(\frac{N^3-N}{12} - \sum_i T_i\right)}}$$

The Student t-test used in the current set of simulations is calculated using the following equation:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Distributional forms

Four levels of skew (0, 1, 2 & 3) are investigated in the current simulation. It should be noted that the population skew was determined empirically for large sample sizes of 240,000 and 480,000 values with 1,000, 7.4, 2.2, and .83 degrees of freedom resulting in skew values of .03, 1.03, 1.92 and 3.06 respectively; because the degrees of freedom are not whole numbers and the resulting distributions are simply approximations. The mathematical relation is $\gamma 1 = \sqrt{8/df}$. The negative skew population distribution was created by 1) simulating a positive skew population distribution, 2) finding the largest score in the positive skew population distribution, and 3) subtracting the largest score, plus 1, from each score in the positive skew population distribution (Tabachnik & Fidell, 2013). Once the negative skew population distribution was created, a correction was applied to ensure the reflection procedure did not change the simulated population mean. This correction process involved 1) subtracting the negative skew population distribution mean by the positive skew population mean, 2) obtaining the difference score, and 3) subtracting each value of the negative skew population distribution by the obtained difference score. This minor correction resulted in negative skew and positive skew population distributions having identical mean values and reflected skew values. Population distributions were generated for each cell of the design then manipulated to fit the condition for that given cell.

Determining Type I error rates

To assess the performance of each of the two tests used in the current simulation, the frequency of the Type I errors for each cell in the design was used. To briefly review the methodology used, the following will describe one cell in the design. In this example, samples will be drawn from two distributions with skew values in the opposite direction (e.g., -3; 3). To begin, two distributions are generated with respective skew values (e.g., 3). Once distributions are generated, one of the distributions (e.g., the second distribution) are reflected using the technique described above resulting in distributions whose skew values are opposite from one another (i.e., -3; 3). After the distributions have been generated, samples are drawn from each of the distributions. For this example, each group consisted of 12. This results in 10,000 sets of groups that each test will be conducted upon. That is, each set consist of one group from each of the distributions, and each of these sets represents one draw of the distributions. Hence, there were 10,000 draws used in each cell of the design in the current simulation. Once the groups have been determined, the WMW and Student t-test are performed on each of the sets and the Type I error rates for each test is recorded. To determine whether tests are operating within an acceptable level, a nominal alpha level of .05 (± .025) was selected. The nominal Type I error rate was determined based on Bradley's (1978) description of liberal robustness. The reason for using the most liberal criterion is that this represents the situation where the tests have the best chance of not being biased and thus may possess some utility in non-experimental research designs. In the current study, power is not reported as the Type I error is used solely to determine the validity of the two tests under the simulated conditions.

Results

The Type I error rate for the Student t-test and WMW for the first simulation condition where N = 24, and sample size ratio for the two groups was 1/1 are reported in Table 1. Table 1 has seven rows and columns that represent the various sample size ratios that were simulated in the design. As well, within each of the column/row combinations there are two columns that report the Type I error rate of the Student t-test and the WMW respectively. For example, the first cell of Table 1 illustrates the condition where the skew of each distribution is -3, N = 24 and the sample size ratio is 1/1. The Type I error rate of the Student t-test in this cell is .041 and the Type I error rate of the WMW is .057. When taking the entire set of results for this condition into consideration, it should be noted that when the parent distributions were both normal (i.e., skew = 0) and the variances of the distributions were equal, the Type I error rates of both tests were maintained. In fact, when distributions were similar (e.g., both have a skew = -3), the Type I error rates of both tests were maintained within an acceptable level. When the distributions differed and the variances were equal, both tests had elevated Type I error rates with the WMW possessing higher degrees of error. For example, when the distributions possessed opposite skew (i.e., 3/-3), the Type I error rates of the Student t-test and WMW were .108 and .253 respectively.

The Type error rates of the two tests for the second simulation condition where N = 48, and the sample size ratio of the two groups are 1/1 are reported in Table 2. When distributions are similar (i.e., equal amount of

Table 1. Type I Error Rates of Student t-test and WMW for the First Simulation Condition

N=24 (12/12), Variance Ratio = 1/1														
Skew	-3		-2		-1		0		1		2		3	
N1/N2	TT	MW	TT	MW	TT	MW	TT	MW	TT	MW	TT	MW	TT	MW
-3	.041	.057	.058	.124	.066	.108	.060	.135	.071	.120	.087	.171	.108	.253
-2	.057	.126	.045	.050	.055	.074	.068	.128	.061	.091	.075	.128	.089	.171
-1	.065	.112	.053	.075	.049	.053	.062	.108	.061	.082	.070	.101	.072	.118
0	.059	.134	.059	.121	.062	.112	.049	.051	.058	.113	.061	.125	.060	.133
1	---	---	---	---	---	---	.061	.107	.048	.052	.055	.076	.065	.104
2	---	---	---	---	---	---	.060	.124	.057	.080	.049	.054	.053	.122
3	---	---	---	---	---	---	.067	.137	.068	.110	.060	.125	.042	.051

* Note TT = Student t-test; MW = Wilcoxon-Mann-Whitney test

Table 2. Type I Error Rates of Student t-test and WMW for the Second Simulation Condition

N=48 (24/24), Variance Ratio = 1/1														
Skew	-3		-2		-1		0		1		2		3	
N1/N2	TT	MW	TT	MW	TT	MW	TT	MW	TT	MW	TT	MW	TT	MW
-3	.044	.050	.053	.177	.059	.137	.056	.114	.063	.151	.070	.259	.087	.413
-2	.052	.174	.048	.050	.051	.089	.055	.107	.057	.117	.063	.188	.072	.263
-1	.058	.135	.051	.086	.053	.052	.057	.098	.057	.091	.055	.121	.061	.145
0	.054	.111	.055	.109	.058	.094	.050	.048	.059	.098	.059	.105	.057	.124
1	---	---	---	---	---	---	.052	.094	.046	.045	.054	.088	.055	.132
2	---	---	---	---	---	---	.058	.108	.051	.083	.046	.049	.052	.175
3	---	---	---	---	---	---	.057	.121	.060	.131	.052	.179	.042	.047

* Note TT = Student t-test; MW = Wilcoxon-Mann-Whitney test

skew), the Type I error rates of both tests are generally maintained. For example, in the condition where the skew of each distribution was equal to -3, the Type I error rates of the Student t-test and the WMW were .044 and .05 respectively. Similar to the conditions where N=24 and the group sizes were equal when distributions were skewed in opposite directions, the Type I error rate of both tests were inflated. For example, in the case where the skew was 3/-3 and the variances were equal, the Type I error rates of the Student t-test and WMW were .087 and .413 respectively.

Table 3 illustrates the third simulation conditions where N=24, the group ratios were 2 to 1 (i.e., 8/16). Again, when the distributions were similar, the Type I error rates of both tests was generally maintained but perhaps slightly conservative. For example, in the condition where N=24 and skew values were -3/-3, the Type I error of the Student t-test and WMW were .036 and .040 respectively. Notably, when the distributional form of the two sampling distributions are not identical, the Type I error rates deviate from the nominal Type I error rate in terms of being both liberal and conservative. For example, referring to Table 3, in the condition where N=24, variance ratio 1/1, and skew values were -3/0, the Type I error rates of the Student t-test and WMW were conservative with values of .009 and .024 respectively. Conversely, when N=24, variance ratio 1/1, and skew values were -3/3, the Type I error rate of the Student t-test and WMW were liberal with values of .104 and .215 respectively.

Table 4 illustrates the fourth simulation condition where N=48 and the group sample size ratios 2/1 (i.e., 8/16). In line with the other results, when the distributional forms were similar, the Type I error rates of both tests were maintained. For example, where N=48 and the skew values were -3/-3, the Type I error rates for the Student t-test and WMW were .042 and .057 respectively, but as the distributional forms deviated from one another the Type I error rate increased for both tests. For example, when N=48 and the skew for each of the parent distributions are 3/-3, the Type I error rates of the Student t-test and WMW are .085 and .373

Table 3. Type I Error Rates of Student t-test and WMW for the Third Simulation Condition

N=24 (8/16), Variance Ratio = 1/1														
Skew	-3		-2		-1		0		1		2		3	
N1/N2	TT	MW	TT	MW	TT	MW	TT	MW	TT	MW	TT	MW	TT	MW
-3	.036	.040	.031	.070	.019	.049	.010	.027	.019	.056	.050	.119	.104	.215
-2	.104	.141	.044	.044	.025	.039	.009	.026	.030	.056	.073	.113	.128	.163
-1	.160	.139	.112	.099	.046	.044	.009	.030	.058	.063	.119	.108	.158	.137
0	.199	.082	.196	.094	.198	.106	.048	.043	.191	.106	.198	.095	.194	.081
1	---	---	---	---	---	---	.009	.029	.051	.046	.109	.099	.164	.141
2	---	---	---	---	---	---	.010	.028	.023	.040	.045	.045	.104	.148
3	---	---	---	---	---	---	.011	.028	.019	.050	.030	.067	.037	.043

* Note TT = Student t-test; MW = Wilcoxon-Mann-Whitney test

Table 4. Type I Error Rates of Student t-test and WMW for the Fourth Simulation Condition

N=48 (16/32), Variance Ratio = 1/1														
Skew	-3		-2		-1		0		1		2		3	
N1/N2	TT	MW	TT	MW	TT	MW	TT	MW	TT	MW	TT	MW	TT	MW
-3	.042	.057	.028	.131	.014	.083	.006	.047	.016	.102	.042	.224	.085	.373
-2	.105	.208	.045	.048	.025	.055	.009	.045	.025	.090	.060	.175	.108	.260
-1	.145	.165	.106	.119	.051	.050	.008	.045	.057	.092	.116	.150	.148	.182
0	.183	.198	.179	.169	.177	.160	.048	.050	.178	.160	.172	.175	.178	.198
1	---	---	---	---	---	---	.008	.041	.050	.050	.101	.114	.149	.173
2	---	---	---	---	---	---	.007	.044	.019	.048	.049	.051	.097	.206
3	---	---	---	---	---	---	.009	.053	.014	.084	.025	.126	.043	.047

* Note TT = Student t-test; MW = Wilcoxon-Mann-Whitney test

respectively. Once again, the findings demonstrate the impact of differential distributions on the Type I error rates of the two tests.

Table 5 illustrates the results from the fifth simulation condition where N=24 and the group sample size ratio is 1/3 (i.e., 6/18). Once again, when samples are drawn from similar distributions the Type I error rates of both tests are maintained. For example, when N=24 and skew values are -3/-3, the Type I error rate of the Student t-test and WMW tests are .041 and .042 respectively, but once the distributional forms are different, Type I error rates are elevated. For example, when N=24 and skew values are 3/-3, the Type I error rates of the Student t-test and WMW are .102 and .197 respectively.

Table 6 shows the Type I error rates from the sixth simulation condition where N=48 and the group sample size ratio is 1/3 (i.e., 6/18). When data were drawn from distributions that were the same, the Type I error rates of both tests were once again maintained. For example, when N=48 and skew values are 3/-3, the Type one error rates of the Student t-test and WMW are .043 and .052 respectively. Once again, as distributional forms become more disparate, the Type I error rates of both tests were elevated. For example, when N=48 and skew is 3/-3, the Type I error rates of the Student t-test and WMW are .081 and .33 respectively.

Discussion

The main purpose of the current series of simulations was to investigate the impact of conducting the Student t-test and WMW on data that have been sampled from populations with differing distributional forms as can occur in many non-experimental designs. It is evident from the results of the simulation that when distributional forms are not identical, the Type I error rates of both the WMW test and Student t-test can become either elevated or conservative far beyond what is considered acceptable. Elevated Type I error rate result in very liberal tests where the null hypothesis is incorrectly rejected. Alternatively, conservative tests reject the null hypothesis at a very low frequency, thus resulting in tests with very low power to detect differences in the central tendency of the distributions.

Table 5. Type I Error Rates of Student t-test and WMW for the Fifth Simulation Condition

N=24 (6/18), Variance Ratio = 1/1														
Skew	-3		-2		-1		0		1		2		3	
N1/N2	TT	MW	TT	MW	TT	MW	TT	MW	TT	MW	TT	MW	TT	MW
-3	.041	.042	.024	.048	.010	.027	.002	.025	.009	.037	.039	.105	.102	.197
-2	.132	.155	.043	.044	.013	.024	.002	.020	.017	.041	.067	.104	.157	.167
-1	.236	.164	.157	.123	.047	.047	.002	.017	.055	.063	.160	.125	.243	.178
0	.304	.215	.302	.206	.291	.192	.050	.049	.292	.195	.302	.211	.310	.221
1	---	---	---	---	---	---	.003	.018	.051	.049	.154	.121	.234	.165
2	---	---	---	---	---	---	.002	.022	.016	.027	.048	.050	.140	.156
3	---	---	---	---	---	---	.002	.026	.009	.027	.024	.042	.043	.046

* Note TT = Student t-test; MW = Wilcoxon-Mann-Whitney test

Table 6. Type I Error Rates of Student t-test and WMW for the Sixth Simulation Condition

	N=48 (12/36), Variance Ratio = 1/1													
Skew	-3		-2		-1		0		1		2		3	
N1/N2	TT	MW	TT	MW	TT	MW	TT	MW	TT	MW	TT	MW	TT	MW
-3	.043	.052	.017	.083	.009	.048	.001	.024	.007	.064	.023	.177	.081	.330
-2	.125	.197	.046	.047	.012	.034	.001	.021	.014	.066	.064	.157	.143	.249
-1	.214	.187	.138	.129	.050	.051	.002	.016	.055	.079	.137	.138	.223	.194
0	.291	.159	.278	.160	.282	.164	.051	.051	.276	.160	.287	.162	.282	.158
1	---	---	---	---	---	---	.001	.019	.051	.048	.149	.129	.219	.186
2	---	---	---	---	---	---	.001	.020	.015	.033	.051	.049	.122	.200
3	---	---	---	---	---	---	.002	.025	.006	.048	.021	.085	.046	.049

* Note TT = Student t-test; MW = Wilcoxon-Mann-Whitney test

It is imperative that researchers attend to the nature of the distributions of their data when using nonparametric statistical approaches to investigating group differences. A potential source of this problem relates to the nomenclature often used to refer to nonparametric statistics, that is that they are 'distribution free'. These techniques are often referred to as 'distribution free' because they do not require the assumption of normality; however, as Kendall and Sundrum (1953) point out that,

> …the term distribution free is not perfect for this reason: it means that the test is free of a parent distribution, but not that the test-statistic itself has no distribution. The statistic, of course, must have a distribution because otherwise no probabilistic inference would be possible. Provided that this point is understood, the term gives rise to no difficulty. (p. 127)

This quote illustrates the potential problem with using a term such as 'distribution free' when referring to rank-based techniques in that 'distribution free' suggests that there are no distributional requirement for utilizing the test. This is not the case, and data analysts and researchers must acknowledge that there are distributional requirements associated with using nonparametric tests. Therefore, it is important to keep in mind that even though there is no requirement for a specific distributional form, like in parametric tests, there is the requirement that whatever the distributional form is, it must be the same for each group used in the analysis.

Implications for Data Analysis

Researchers and data analysts must recognize the importance of gaining insight and understanding pertaining to the nature of the distributions of the populations that they are working with, as well as an understanding that this is equally as important when using either parametric or nonparametric methods. With an understanding of the shape of population distributions, researchers and data analysts will be able to make informed decisions related to the type of statistical approach they choose to use for their analysis. A major point that needs highlighting here relates to experimental versus nonexperimental designs. When using an experimental design, groups are sampled from the same population, thus there is an increased likelihood that the distributions utilized in the analysis will have quite similar shapes; whereas, when a nonexperimental design is utilized, groups are sampled from different populations (e.g., gender), and thus the likelihood of distributions being equivalent between the groups is greatly diminished. Researchers and data analysts must recognize this distinction between experimental and nonexperimental designs and take steps to address this. When utilizing nonexperimental designs, the use of nonparametric tests is not recommended unless there is evidence that the assumption of identical distributions has been satisfied.

Future research should investigate techniques for demonstrating the equivalence of distributional forms. One suggestion is to employ a combination of graphical and descriptive statistical techniques to help determine whether distributional forms are sufficiently identical to conduct analysis with the data; however, prior to employing this type of approach, simulation studies that explore the bounds of the validity of nonparametric tests when the assumption of identical distributions is violated must be conducted. The current study provides some insight into this, but further simulations are needed to explore these bounds in more depth. Another suggestion is to perform a test of equivalence (e.g., Levene's median test) as a preliminary test prior to employing a nonparametric analysis of group differences. This would involve a two-step procedure, much like when testing for equality of variances when utilizing parametric tests for groups differences (e.g., a Student t-test). This approach might provide some protection against elevated Type I errors when testing for group differences; however, simulation studies should be conducted to test this prior to implementing this strategy in daily research practices.

The take away message from this simulation study is that researchers and data analysts need to be cautious when utilizing nonparametric tests for group differences as violations of the assumption of identical distributions, as demonstrated in the current simulation study, lead to seriously inflated Type I error rates. It is important to note that both parametric and nonparametric tests are impacted by non-identical distributions and the 'nonparametric' nature of the test does not make up for this assumption. What this means is that researchers and data analysts need to attend to the nature of the population distributions and the research design that they employ (i.e., experimental versus nonexperimental) to reduce biasing the analyses they are conducting.

References

Aron, A., Coups, E., & Aron, E. (2013). *Statistics for Psychology* 6th edition. New York: Pearson.

Brace, N., Kemp, R., & Sneglar, R. (2012). *SPSS for Psychologists* 5th edition. New York: Palgrave Macmillan.

Bradley, J.V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*, 144-152.

Corder, G., & Foreman, D. (2009) *Nonparametric Statistics for Non-Statisticians*. Hoboken: Wiley.

Fagerland, M., & Sandvik, L. (2009). The Wilcoxon-Mann-Whitney test under scrutiny. *Statistics in Medicine*, *28*, 1487-1497.

Kendall, M.G. & Sundrum, R.M. (1953). Distribution-free methods and order properties. *Review of the International Statistical Institute*, *21*(3), 124-134.

Nordstokke, D.W., & Zumbo, B.D. (2010). A new non-parametric test for equal variances. *Psicologica*, *31*, 401-430.

Nordstokke, D.W., & Colp, S.M. (2014). Investigating the robustness of the nonparametric Levene test with more than two groups. *Psicologica*, *35*, 339-361.

Pagano, R., (2001). *Understanding Statistics for the Behavioral Sciences* 6th edition. Belmont: Wadsworth.

Privitera, G. (2012*). Statistics for the Behavioral Sciences*. Thousand Oaks: Sage.

Scheffe, H. (1970). Practical solutions of the Behrens-Fisher problem. *Journal of the American Statistical Association*, *65*, 1501-1508.

Siegel, S. (1956). *Nonparametric Statistics*. New York: McGraw-Hill Book Company

Tabachnik, B. & Fidell, L. (2013). *Using Multivariate Statistics* 6th edition. New York: Pearson

Tomarken, A.J. & Serlin, R. C. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin*, *99*(1), 90-99.

Zimmerman, D.W. (1987). Comparative power of Student t test and Mann-Whitney U test for unequal sample sizes and variances. *Journal of Experimental Education*, *55*, 171-174.

Zimmerman, D.W. (2004). A note on preliminary test of equality of variances. *British Journal of Mathematical and Statistical Psychology*, *57*, 173-181.

Citation:

Nordstokke, David W., & Colp, S. Mitchell. (2018) A Note on the Assumption of Identical Distributions for Nonparametric Tests of Location. *Practical Assessment, Research & Evaluation*, 23(3). Available online: http://pareonline.net/getvn.asp?v=23&n=3

Corresponding Author

David W. Nordstokke
University of Calgary
2500 University Drive NW
Calgary, Alberta, Canada

email: dnordsto [at] calgary.ca