Journal of Educational Research & Practice
2024, Volume 14, Issue 1, Pages 309–322
DOI: 10.5590/JERAP.2024.14.1.21

Original Research
Walden University

A Rank-Order Alternative for Nonparametric Analysis With the General Linear Model

Jeffry White, PhD
National University, San Diego, California, United States
https://orcid.org/0009-0004-3017-0171

Contact: *jwhite2@nu.edu*

Abstract

Violations of normality and homogeneity are common in educational data. When this occurs, the use of parametric statistics may be inappropriate. A generalized form of nonparametric analyses based on the Puri and Sen *L* statistic provides an alternative approach. Using a chi-square distribution, this technique is easy to apply and has significant power. Another advantage of the *L* statistic is its utility to link nonparametric tests with their parametric counterparts. After rank-ordering and analyzing the data, an adjustment is made by calculating *L* instead of relying on the parametric test statistic. This permits the researcher to choose between parallel strategies based on the distribution characteristics of the data. In this paper, I present case illustrations to demonstrate the approach and offer guidance to faculty, students, and staff researchers interested in nonparametric options when working with educational data that possess asymmetrical and/or heteroscedastic qualities.

Keywords: *normality, homoscedastic, nonparametric, Puri and Sen test*
Date Submitted: January 17, 2024 | Date Published: December 19, 2024

Recommended Citation

White, J. (2024). A rank-order alternative for nonparametric analysis with the general linear model. *Journal of Educational Research and Practice, 14*, 309–322. https://doi.org/10.5590/JERAP.2024.14.1.21

Introduction

In this paper, I begin by presenting an overview of the assumptions of parametric statistics, criticisms of the use of nonparametric statistics, and reproval of the robustness of the general linear model. I present an alternative statistical approach with rank-ordered data and case examples to demonstrate the technique. The purpose is to illustrate how educational researchers can use nonparametric and parametric statistics in tandem to address the assumptions and ensure the authenticity of their results.

The use of parametric statistics with normally distributed data possessing asymmetrical qualities is not uncommon in educational research, as is the use of comparison data with unequal variances possessing heteroscedastic characteristics. What is surprising, however, is that nonparametric techniques do not frequently appear in the literature (Keselman et al., 1998; Skidmore & Thompson, 2010). There are myriad reasons for this (Harwell, 1988), such as limited exposure to nonparametric techniques in graduate curricula (Capraro &

Thompson, 2010) and faculty perceptions of the statistical training required for doctoral students (Henson & Williams, 2006). The consequences of this inattention have been well-debated (Glass et al., 1972, as cited in Lix et al., 1996; Huynh & Finch, 2000; Blanca et al., 2013) and led to a call for greater use of nonparametric statistics in educational research (Leech & Onwuegbuzie, 2002).

A nonparametric technique, the Puri and Sen *L* test (1969, 1985) can be used with many parametric statistics, such as when the assumptions of normality and homogeneity have been violated. The null hypothesis for the general linear model is that there is no relationship when the dependent variable is regressed on the independent. Similarly, the *L* statistic tests the same relationship. In contrast to the *F* distribution, the *L* employs a chi-square approximation that does not require a normal distribution when the variables are transformed to ranked data.

Parametric Assumptions

Many of the classic parametric statistics used in social science research (e.g., *t* test, analysis of variance [ANOVA], linear regression) are predicated on the data meeting the normality and homoscedasticity assumptions (Corder & Foreman, 2014; Trochim & Donnlley, 2008). The normality assumption is based on the central limit theorem, which posits that in a random and independently drawn sample from a population, a distribution of sample means will be approximately normal regardless of the original shape of the population distribution (Pearson, 2010). The symmetrical shape of the distribution of the data is one of the key concepts in statistics that allows researchers to draw inferences about the population from the sample data. However, when the distribution is skewed (asymmetrical), the means cannot accurately reflect the central tendency. Attempting to make inferences about the population from this type of data can be paramount to interjecting bias drawn from the sample statistics.

The importance of the normality assumption has been well-debated (Howell, 2012; Grice, 2011; Nussbaum, 2014). Probability simulations have demonstrated that normal distribution-based tests are less sensitive to normality violations, and the consequences are not as severe as previously thought (Hill & Lewicki, 2007). Some researchers have chosen to ignore the assumption altogether and proceed with parametric analyses (e.g., Keselman et al., 1998), citing support from Monte-Carlo studies (e.g., Glass et al., 1972, as cited in Lix et al., 1996; Blanca et al., 2013). Others have warned against this practice and called for the use of more nonparametric statistics in educational research (Leech & Onwuegbuzie, 2002).

Another key assumption is homogeneity of variances. It is referenced in the literature as homoscedasticity, equality of variances, homogeneity of regression slopes, sphericity, or homogeneity of the variance–covariance matrix. This assumption calls for the variances to be relatively equal across all levels of the independent variables (Warner, 2013). It means that the variances of different groups should have an approximate ratio of 1:1. In reality, the variance ratio will frequently drift significantly from that required to satisfy the assumption. In a review of educational and child psychology journals, Keselman et al. (1998) discovered variance ratios as high as 566:1. Grissom (2000) identified some as high as 281:1 and Erceg-Hurn and Mirosevich (2008) found that most exceeded 20:1, with some as high as 121:1. Even after accounting for sampling error, the frequency in reporting ratios of these magnitudes indicates that heteroscedastic data are common. Any inferences drawn from this kind of data can lead to inaccurate estimations of the population, making any meaningful comparison about the groups questionable.

Given the nature of educational data, scholars should not be surprised that it often fails to meet the parametric assumptions. At times, researchers are interested in comparing the performance of preexisting groups on the basis of demographic variables like race/ethnicity, gender, socioeconomic status, and similar characteristics. Given the uniqueness of many groups, the differences found in the variances should not be unexpected.

Violations of these assumptions can affect the results of parametric tests by distorting the Type I, or false positive, error rate (Erceg-Hurn & Mirosevich, 2008). When parametric tests are used in the analysis of asymmetrical and heteroscedastic data, the risk of making a Type I error can be greater than what is reported by the *p*-value (Müller et al., 2015). For example, consider data with equal sample sizes but heterogeneous variances. Using any statistical software to perform an ANOVA and subsequently reporting an associated probability, or *p*-value of $\leq .05$, thus rejecting the null hypothesis, would typically be interpreted as less than a 5% chance of making a Type I error. In practice, the truth may be closer to 30% (Wilcox et al., 1986). Contrary to popular belief, equal samples provide little relief when the variances are not homogeneous (Brunner, 2021; Harwell et al., 1992). For instance, the probability of a Type I error in linear regression analysis can exceed 50% ($\alpha = .05$) when data are heteroscedastic and not normally distributed (Wilcox, 2016). This highlights the importance of understanding the assumptions that underpin the analyses rather than relying on the software (White, 2013).

Another implication of violating the parametric assumptions is the effect on power. Small deviations from normality can significantly lower the post-hoc power of classic parametric tests. Wilcox (1998) cited a case where a small departure from normality reduced the power of a *t* test from .96 to .28. He summarized the impact of violating the normality and homoscedasticity assumptions, saying:

> As hundreds of articles in statistical journals have pointed out and for reasons summarized in several books. .... small departures from normality can result in lower power; even when distributions are normal, heteroscedasticity can seriously lower the power of standard ANOVA and regression methods. (p. 300)

Wilcox (1998) also pointed out that the importance of understanding the assumptions cannot be understated. Researchers should inspect the normality and homogeneity of the data before making the choice to use parametric or nonparametric statistics to recognize the substantive implications of that decision.

Nonparametric Criticisms

Because rank-ordered and nominal-level data are used, nonparametric statistics do not require satisfaction of the normality assumption (Higgins, 2004). In fact, Gibbons (1993) argued that the Likert-type scales commonly found in educational data possess more ordinal than interval qualities, making them better suited for nonparametric procedures. Yet, the dearth of nonparametric statistics from the empirical literature suggests there are still critics, which likely leads to their less-than-desirable popularity.

The first criticism about nonparametric statistics is that, because they do not make strong assumptions about the population, little if any inference can be made from the sample data. Second, contrary to a common misconception, they still must satisfy the homoscedasticity assumption (Nordstokke et al., 2011). Third, a degree of precision will be lost when transforming Likert-scaled to ranked data (Edgington et al., 2007). With a few exceptions, nonparametric statistics will generally have lower power (Freidlin & Gastwirth, 2000). They are also denounced for being incapable of answering focused research questions (Johnson, 1995). For example, the Mann-Whitney and Wilcoxon procedures examine whether two distributions are different but cannot indicate how they differ in mean, variance, or shape. Finally, there can be a false sense of security that they are immune from parametric assumptions such as outliers and unequal variances (Zimmerman, 2000).

Given these shortcomings, it might seem there would be little support for nonparametric analyses. But despite the reprovals, they are still recommended in many situations. For example, Skovlund and Fenstad (2000) compared the Type I error rates of the independent samples *t* test and the Mann-Whitney *U* with different variances, distributions, and sample sizes. The Mann-Whitney was more robust when the data had highly skewed distributions, heterogeneous variances, and unequal groups. Educational data with similar issues would likely increase the Type I error rate (Wilcox, 1998; Lix & Keselman, 1998).

Robustness Argument

Erceg-Hurn and Mirosevich (2008) contended that parametric statistics are robust to some of the assumption violations. The term *robustness* refers to procedures that maintain the Type I error rate at its nominal level while preserving statistical power (Wilcox, 2005). This argument can be traced back to classic parametric methods and include studies that have made their way into textbooks and come to be widely accepted (Boneau, 1960, as cited in Delacre et al., 2019; Box, 1953, as cited in Koenker, 1981; Lindquist, 1953, as cited in Richardson, 2018; Glass et al., 1972, as cited in Lix et al., 1996; Blanca et al., 2013). Subsequently, investigators have found that they only examined the impact of small deviations in normality and homogeneity (e.g., Bradley, 1978 as cited in Migdadi, 2015; Harwell, 1992). When Bradley (1978) and Harwell revisited the research, they came to very different conclusions and pointed out that some authors ignored evidence not supporting the robustness argument and extended their assertions beyond the capacity of the data.

Supporters of the robustness argument typically focus their attention on the Type I error and fail to account for the power of parametric statistics when the data are not normally distributed or violate homogeneity (Erceg-Hurn & Mirosevich, 2008). Akritas and colleagues (1997) found that even when robust to Type I error, parametric statistics are less powerful and generalizability becomes problematic. This presents challenges for the researcher who must transform, resample, or apply multilevel modeling to the data to overcome these problems. There are modern statistical tests that provide methods for testing the homogeneity (e.g., Levene's, Bartlett's, Hartley's Fmax) and normality (e.g., Anderson-Darling, Shapiro-Wilk, Kolmogorov-Smirnov) assumptions. Yet some prominent statisticians criticize them as flawed and suggested they should not be used (D'Agostino et al., 1990; Glass & Hopkins, 1996). Therein lies the conundrum when researchers are confronted with unbalanced, asymmetrical, and heteroscedastic data. There is, however, an alternative that links nonparametric tests with their parametric counterpart (Harwell, 1988; Lawson, 1983 as cited in Wilcox, 2016) to overcome this problem.

Puri and Sen L Statistic

The Puri and Sen (1969, 1985) approach for hypothesis testing uses two parallel forms of the data. The first entails a parametric approach with the original data followed by a nonparametric analysis with the data rank-ordered (Harwell & Serlin, 1989). These two forms, raw and rank-ordered, are used to link a number of parametric and nonparametric tests that share common hypotheses, such as those associated with the general linear model family of statistics (e.g., *t* tests, ANOVA, multivariate analysis of variance [MANOVA], linear regression, etc.). The major advantage of the *L* statistic is that it does not require normally distributed data and is expressed as $L = (N - 1)\,\theta$, where *L* is the Puri and Sen test and theta (θ) is the measure of explained variance. Because the *L* is calculated using rank-ordered data, theta represents a ratio of the explained variance in the ranks of the total variance. The *L* is then compared to a chi-square distribution at the desired level of significance where the degrees of freedom (*df*) are calculated by the number of dependent variables (*p*) multiplied by the number of independent variables (*q*) and $df = pq$. The null hypothesis is rejected when the *L* statistic exceeds the chi-square critical value. Next, the results are compared to the parametric test results derived from the raw (unranked) data. When both tests are consistent, the researcher notes the parallelism and has greater confidence in the parametric results. Conversely, if the parametric results are significant and there are nonsignificant nonparametric findings, there is a greater likelihood of a Type I error (false positive) in the parametric statistics (Thomas et al., 1999). The following case scenarios illustrate the Puri and Sen test applied to a one-way ANOVA, two-way ANOVA, multiple linear regression analysis, and MANOVA.

Computational Illustrations

In the first scenario, educational researchers explored the differences in SAT scores based on the gender of the high school student (National Center for Educational Statistics, 2021). Data from a sample of high school seniors (N = 106) indicated that the male students (M = 984.87, SD = 58.520) seemed to do better than females (M = 967.50, SD = 47.317). A closer inspection revealed an unbalanced design with the number of males (n = 88) significantly outnumbering the number of female students (n = 18). Ranking the SAT data also revealed a smaller difference between the males (MR = 54.41, SD = 30.243) and females (MR = 44.17, SD = 32.322). Using this illustration, the Puri and Sen (1969, 1985) test takes the form as indicated in Equation 1:

$$L = (N - 1)\ (\text{Sums of Squares Between Groups}/\text{Sums of Squares Total}) \quad (1)$$

Keeping in mind that a two-sample t test is equivalent to a one-way ANOVA (see Note 1), the between groups sums of squares and the sum of squares total are obtained by using the ranked data in a one-way ANOVA (Casella, 2008). The F-ratio test indicated no statistically significant differences (p > .05). When the L statistic was compared to the critical value from a chi-square distribution with df = 1, similar to the one-way ANOVA, no gender differences were detected in the SAT ranked data (χ^2 = 3.841, α > .05). Consequently, the researcher could now be more assured when examining the results in Table 1 that a Type II error had not been committed due to the unbalanced design and unequal variances.

Table 1. *Comparison of* F *and* L *Statistics in One-Way ANOVA*

SAT (raw data)	*SS*	*df*	*MS*	*F*	*p*
Between groups.	4493.586	1	4493.586	1.391	ns
Within groups	335996.273	104	3230.733		
Total	340489.858	105			
SAT (ranked data)	*SS*	*df*	*MS*	*L**	*p*
Between groups.	1888.727	1	1888.727	1.9987	ns
Within groups	97333.273	104	935.897		
Total	99222.000	105			

* L = (106-1) (1888.727/99222.000).

To demonstrate the L in a two-way factorial ANOVA (see Note 2), consider the same scenario with the researcher now interested in the effect of teacher gender and years of teaching experience on their attitude about the inclusion of students with autism in general education classrooms (Spirko, 2015). Data was collected from a sample of teachers (N = 223) in which women appeared to have higher attitudinal scores than men. Inspection of the raw data revealed a highly skewed distribution with unequal groups. It was next rank-ordered and the means, ranked means, and standard deviations were examined. A cursory analysis of the teaching experience categories indicated a decline in attitude based on the number of years. A comparison of the means, mean ranks, and standard deviations is presented in Table 2.

Table 2. *Descriptive Statistics for the Puri & Sen Factorial ANOVA Illustration*

Variable		Teachers attitude toward inclusion	
		M(SD) raw data	*MR(SD) ranked data*
Gender			
	Male	67.17(20.302)	10141(70.718)
	Female	71.49(17.513)	117.77(62.947)
Teaching experience			
	0–5 years	72.27(16.820)	120.09(62.526)
	6–10 years	70.98(17.921)	115.98(64.920)
	10+ years	66.29(20.080)	98.66(67.567)

Because an expansion of the *L* computations is needed to account for the effect of the two factors and their interaction, Equations 2 and 3 express the simple main effects of gender and teaching experience, and Equation 4 the interaction effects:

$$L_1 = (N - 1) \text{ (Sums of Squares Gender/Sums of Squares Total)} \quad (2)$$

$$L_2 = (N - 1) \text{ (Sums of Squares Years Teaching/Sums of Squares Total)} \quad (3)$$

$$L_3 = (N - 1) \text{ (Sums of Squares Gender x Years Teaching/Sums of Squares Total)} \quad (4)$$

The rank-ordered data were next analyzed, and the *L* statistic calculated for each simple main and interaction effect and compared with the *F* statistic results presented in Table 3.

Table 3. *Comparison of F and L statistics in Factorial ANOVA*

Attitude (raw data)	*SS*	*df*	*MS*	*F*	*p*
Gender	426.240	1	426.240	1.280	ns
Years teaching	781.604	2	390.802	1.173	ns
Gender x years teaching	102.721	2	51.360	.154	ns
Error	72271.467	217	333.048		
Total	1170340.000	223			
Attitude (ranked data)	*SS*	*df*	*MS*	*L**	*p*
Gender	7051.556	1	7051.556	.4218	ns
Years teaching	11641.578	2	5820.789	.6882	ns
Gender x years teaching	645.654	2	322.827	.0444	ns
Error	918249.959	217	4231.567		
Total	3774880.500	223			

*$L_{Gender} = (N - 1)$ *SS* gender/*SS* total.
$L_{Years\ Teaching} = (N - 1)$ *SS* years teaching/*SS* total.
$L_{Gender\ x\ Years\ Teaching} = (N - 1)$ *SS* gender x *SS* years teaching/*SS* total.

Based on a chi-square distribution with two degrees of freedom (*df*), the nonparametric data indicated that neither gender nor years of experience had an effect on teacher attitudes regarding the inclusion of students with autism in general education classes (χ^2critical = 5.99, $\alpha > .05$). There was also no evidence of a significant interaction in the *L* test. These findings complemented those from the parametric data. Similar to the previous example, the researcher can be reasonably confident the sample results are not the product of a Type II error.

In the next illustration, an educational researcher was interested in determining if the Basic Educational Skills (BEST) Pretest, education level, and age are predictors of a student's English proficiency score (Grafals, 2013). In Table 4, a correlation matrix of the criterion and predictor variables supported the linearity assumption and their use in the regression model (see Note 3).

Table 4. *Descriptive Statistics for the Puri & Sen Linear Regression Illustration*

Variable	Age	Education	BEST pretest	English language competency	*M(SD)* raw data	*M(SD)* ranked data
Age	1.00	-.142*	.038	-.205**	-	171.98 (98.570)
Education		1.00	.240**	.326**	-	171.50 (98.826)
BEST pretest scores			1.00	.532**	397.96 (109.413)	138.50 (79.806)
English language competency scores				1.00	3.38 (.996)	171.50 (98.826)

*$p < .01$ **$p < .001$.

After reviewing the descriptive statistics, Grafals (2013) discovered the skew and kurtosis of the distribution exceeded the boundaries of normal limits. When confronted with nonnormal data the researcher must choose between data transformation or exploring nonparametric regression models (Balanda & MacGillivray, 1988 as cited in DeCarlo, 1997). However, the Puri and Sen (1969, 1985) approach can also be applied.

Before proceeding, an adjustment to the Puri and Sen (1969, 1985) test was required. Because the coefficient of determination (R^2) represents the proportion of variance in the dependent variable that is accounted for (shared) with the predictor variables in the model (Kelly & Preacher, 2012), it is substituted for theta (θ) in the computation of the *L* statistic. After first running the linear regression with the raw data, the researcher can interpret the regression results based on the normality assumption. Next, the predictor and criterion variables could be ranked and compared to the *L* statistics results. The inferential statistics indicate that the three variables were three significant predictors in the model. The results are presented in Table 5.

Table 5. *Comparison of F-to-Enter and L-to-Enter Statistics in Stepwise Multiple Linear Regression*

English language competency (raw data)	*R*	R^2	β	*df*	*F-to-enter*
BEST pretest	.526	.276	.488	1, 273	104.240**
Education level	.562	.316	.182	2, 272	62.844**
Age group	.585	.343	-.165	3, 271	47.118**
English language competency (ranked data)	*R*	R^2	β	*df*	*L-to-enter**
BEST pre-test	.526	.277	.506	1, 273	76.175**
Education level	.573	.328	-.203	2, 272	90.20**
Age group	.588	.345	.139	3, 271	94.875**

* *L-to-enter* = (276-1) R^2 ** $p < .001$.

When using a stepwise approach, the highest *F-to-enter* probability (*p*-value) of each predictor is input at each successive step. In this example, only the standardized beta weights for Model 3 are reported because they represent all the predictor variables. The *L* statistics for each step of the regression model are expressed in Equation 5:

$$L\text{-to-enter} = (N - 1)\, R^2 \quad (5)$$

Analogous to the results obtained from the raw data, the *L* tests indicated all variables are predictors of English language competency scores based on a chi-square distribution with 3 *df* (χ^2critical = 16.266, $\alpha < .001$).

In the next scenario, the researcher's interest was in the multivariate effect (see Note 4) that education level and native language have collectively on the BEST posttest and English language proficiency scores (Grafals, 2013). In Table 6, the descriptive statistics for the raw and ranked data are presented with variability evident in the BEST posttest and English language competency data across the different levels of the two factors.

Table 6. *Descriptive Statistics for the Puri and Sen MANOVA Illustration*

Variable	BEST Posttest		English language competency	
	M(SD) (raw)	*MR(SD) (ranked)*	*M(SD) (raw)*	*MR(SD) (ranked)*
Education				
Less than high school	443.46(60.778)	98.84(65.226)	3.014(1.099)	140.98(102.437)
High school	457.50(56.011)	114.27(62.114)	3.538(.773)	182.86(85.953)
College	477.26(60.778)	138.11(60.951)	3.790(.782)	211.89(86.790)
Native language				
French	490.00(81.479)	148.36(66.047)	3.717(.849)	205.75(90.797)
Arabic	463.30(82.743)	111.10(71.856)	3.283(1.057)	163.53(106.006)
Chinese	466.47(49.406)	122.53(63.265)	2.978(1.051)	131.25(87.572)
Spanish	458.09(46.249)	111.26(58.384)	3.323(.868)	160.50(92.765)
Other native language	436.29(96.350)	100.50(68.232)	3.359(1.153)	176.35(107.076)

After ensuring that the dependent variables had satisfied the linearity requirement ($r = .560$, $p < .001$), the researcher found the BEST posttest scores were leptokurtic and highly skewed and opted to use the Puri and Sen (1969, 1985) test. Before proceeding with the MANOVA, an adjustment needed to be made to compute the *L* statistic. Pillai's Trace (*V*) was substituted for the sum of the total variance that is accounted for by all the variates in the model (Anderson, 2003). Anderson found that Pillai's Trace is the most robust of the multivariate analysis of variance tests of significance. It is calculated using Roy's Root (Θ) and Hotelling Trace (T^2) test statistics and written in Equation 6 as:

$$\Sigma\, (\Theta / 1 + \Theta) + (\Theta - T^2 / 1 + \Theta - T^2) \quad (6)$$

After adjustment of Pillai's Trace, the *L* statistic is denoted as $L = (N - 1)$ for each of the simple main effects of each factor as well as their interaction effect. As illustrated in the previous case scenarios, the multivariate analysis of variance is first computed with the raw data and is then rank-ordered and rerun to obtain the nonparametric values presented in Table 7.

Table 7. *Comparison of F and L Statistics in MANOVA*

Raw data	*Pillai's Trace*	*df*	*F*	*p*
Education	.052	4, 398	2.678	< .05
Native language	.173	16, 398	2.356	< .01
Education x Native language	.207	28, 398	1.637	< .05
Ranked data	*Pillai's Trace*	*df*	*L**	*p*
Education	.046	4	15.686	< .01
Native language	.173	16	58.647	< .001
Education x Native language	.165	28	55.935	< .001

**L = (N-1) Pillai's Trace.*

After contrasting the *L* statistic using a chi-square distribution with six degrees of freedom (*df* = 6), both the raw and rank-ordered data indicate that there are simple main effects and interaction effects across the two dependent variables. Noticeably different, though, are their respective *p*-values. The chi-square distribution suggests there is a more significant native language effect and education x native language interaction than the *F* distribution. That is because the relationship between the distributions is the same after normalizing the data. As the degrees of freedom in the denominator increase, the *p*-value of the *F* statistic will move toward the value reported by chi square (Gould, 2006). The implication is that a greater likelihood of inflating the alpha error exists by using asymmetrical data with the *F* distribution. Conversely, with normality not assumed in the chi-square distribution, the *L* test results decrease the likelihood of Type I error.

Discussion

Though not illustrated here, the *L* statistic is applicable with repeated measures ANCOVA, MANCOVA, and the gamut of statistical techniques in the general linear model. Because of its flexibility and simple computations, it is well-suited for a variety of research situations and warrants consideration when working with data that possess the characteristics described in this paper. While others have recommended Puri and Sen's (1965, 1985) techniques (Kepner & Wackerly, 1996; Marden & Muyot, 1995; Sun, 1997, as cited in Kössler, 2005), this paper has proffered case scenarios using applied data from contemporary research to demonstrate their utility in education.

In the context of the real world, much of the data derived from elementary, secondary, and postsecondary schools and colleges is moderately to highly skewed. For example, instruments that employ Likert-type scales are often skewed due to polarization of many responses at the end of scales. This type of data typically has multiple modes and distributions with long tails that deviate significantly from the normality requirements of parametric statistics. When this happens and the assumptions are violated, nonparametric tests, such as the *L* statistic, should be considered.

While nonparametric tests can be less efficient than their parametric counterparts and prone to inaccurate results (Rasmussen & Dunlap, 1991, as cited in Bishara & Hittner, 2012), they also can be less complicated for students and novice researchers as they often do not require complex mathematical formulas (Okoye & Hosseini, 2024). The major advantage of the Puri and Sen (1969, 1985) technique is that the *L* test can easily be hand-calculated to bring rank-ordered nonparametric procedures into parallel form with their parametric counterparts using any statistical software.

Recommendations

It is important to use the appropriate statistical test that maximizes the power of a study to reject the null hypothesis when it should be rejected. All too often though we scholars think that researchers are limited to either a parametric or nonparametric approach. This is a misperception, and the rank-ordered technique proffered by Puri and Sen (1969, 1985) affords us an alternative that combines the benefits of both. Educational researchers should give serious consideration to using the *L* statistic when they find their data has violated some or all of the assumptions or parametric statistics.

In a similar vein, student and novice researchers should be exposed to more than the basic concepts and principles of parametric and nonparametric statistics. Teachers of statistics need to include instruction on how they can be synthesized to work together. Rather than being taught and viewed as two distinctly different statistical families, they should be encouraged to contemplate the approach presented in this paper.

Conclusions

In summary, my objective is to provide educational researchers with a way to analyze data that violates the assumptions of normality and homogeneity using corresponding parametric and nonparametric statistical methods. Ancillary to that is to make clear that rather than regarding these assumptions as absolute, the educational research community should view them as companion ways to analyze data with skewed distribution and/or heterogenous variances. The Puri and Sen (1969, 1985) method presented in this paper is an alternative that affords researchers the flexibility of examining data in such a manner.

Notes:

1: For more information about one-way analysis of variance (ANOVA), see https://resources.nu.edu/statsresources/One-WayANOVA

2: For more information about two-way factorial analysis of variance (ANOVA), see https://www.statstest.com/factorial-anova/

3: For more information about linear regression, see https://resources.nu.edu/statsresources/simplelinear

4: For more information about multivariate analysis of variance (MANOVA), see https://www.msicertified.com/multivariate-analysis-of-variance-manova/

References

Akritas, M. G., Arnold, S. F., & Brunner, E. (1997). Nonparametric hypotheses and rank statistics for unbalanced factorial designs. *Journal of the American Statistical Association, 92*(437), 258–265. https://doi.org/10.2307/2291470

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis* (3rd ed.). John Wiley & Sons.

Bishara, A. J., & Hittner, J. B. (2012). Testing the significance of a correlation and nonnormal data: Comparison of Pearson, Spearman, transformation and resampling approaches. *Psychological Methods, 17*(3), 399–417. https://doi.org/10.1037/a0028087

Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 9*(2), 78–84. https://doi.org/10.1027/1614-2241/a000057

Bradley, J. V., (1978). Robustness? *British Journal of Mathematical and Statistical Psychology, 31*(2), 144–152. https://doi.org/10.1111/j.2044-8317.1978.tb00581.x

Brunner, E., Konietschke, F., Bathke, A. C., & Pauly, M. (2021). Ranks and pseudo-ranks: Surprising results of certain rank tests in unbalanced designs. *International Statistical Review, 89*(2), 349–366. https://doi.org/10.1111/insr.12418

Casella, G. (2008). *Statistical design.* Springer Publishing Co.

Capraro, R. M., & Thompson, B. (2008). The educational researcher defined: What will future researchers be trained to do? *Journal of Educational Research, 101*(4), 247–253. https://doi.org/10.3200/JOER.101.4.247-253

Corder, G. W., & Foreman, D. I. (2014). *Nonparametric statistics: A step-by-step approach.* John Wiley & Sons.

D'Agostino, R. B., Belanger, A., D'Agostino, R. B., Jr. (1990). A suggestion for using powerful and informative tests of normality. *The American Statistician, 44*(4), 316–321. https://doi.org/10.2307/2684359

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods, 2*(3), 292–307. https://doi.org/10.1037/1082-989X.2.3.292

Delacre, M., Leys, C., Mora, Y. L., & Lakens, D. (2019). Taking parametric assumptions seriously: Arguments for the use of Welch's *F*-test instead of the classical *F*-test in one-way ANOVA. *International Review of Social Psychology, 32*(1), 1–12. https://doi.org/10.5334/irsp.198

Edgington, E., Edgington, E. S., & Onghena, P. (2007). *Randomization tests (*4th ed.*).* Taylor & Francis Group. https://doi.org/10.1201/9781420011814

Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and po of your research. *American Psychologist, 63*(7), 591–601. https://doi.org/10.1037/0003-066X.63.7.591

Freidlin, B., & Gastwirth, J. L. (2000). Should the median test be retired from general use? *American Statistician, 54*(3), 161–164. https://doi.org/10.2307/2685584

Gibbons, J. D. (1993). *Nonparametric statistics: An introduction.* Sage Publications.

Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology* (3rd ed.). Allyn & Bacon.

Gould, W. (2006). Mata matters: Interactive use. *The Strata Journal, 6*(3), 387–396. https://doi.org/10.1177/1536867X0600600308

Grafals, Z. (2013). *English learning predictors of listening and speaking self-efficacy for adult second language learners* (Publication number 3572507) [Doctoral Dissertation, Trident University International]. ProQuest LLC.

Grice, J. G. (2011). *Observation oriented modeling: Analysis of cause in the behavioral sciences.* Elsevier.

Grissom, R. J. (2000). Heterogeneity of variances in clinical data. *Journal of Counseling and Clinical Psychology, 68*(1), 155–165. https://doi.org/10.1037//0022-006x.68.1.155

Harwell, M. R. (1988). Choosing between parametric and nonparametric tests. *Journal of Counseling and Development, 67*(1), 35–38. https://doi.org/10.1002/j.1556-6676.1988.tb02007.x

Harwell, M. R. (1992). Summarizing Monte Carlo results in methodological research. *Journal of Educational Statistics, 17*(4), 297–313. https://doi.org/10.2307/1165126

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo results in methodological research: The one- and two-factor fixed effects ANOVA cases. *Journal of Educational Statistics, 17*(4), 315–339. https://doi.org/10.2307/1165127

Harwell, M. R., & Serlin, R. C. (1989). A nonparametric test statistic for the general linear model. *Journal of Education Statistics, 14*(4), 351–371. https://doi.org/10.3102/10769986014004351

Henson, R. K., & Williams, C. S. (2006, April). Doctoral training in research methodology: A national survey of education-related degrees [Paper presentation]. Annual meeting of the American Educational Research Association, San Francisco, CA.

Higgins, J. J. (2004). *Introduction to modern nonparametric statistics.* Brooks/Cole Cengage Learning.

Hill, T., & Lewicki, P. (2007). *Statistics: Methods and application.* StatSoft, Inc.

Howell, D. C. (2012). *Statistical methods for psychology* (8th ed.). Wadsworth Cengage Learning.

Huynh, H., & Finch, H. (2000, April). Robust/resistant statistical procedures with application to multiple regression analysis and analysis of variance [Paper presentation]. Annual meeting of the American Educational Research Association, New Orleans, LA.

Johnson, D. H. (1995). Statistical sirens: The allure of nonparametrics. *Ecology, 76*(6), 1998–2000. https://doi.org/10.2307/1940733

Kelly, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods, 17*(2), 137–152. https://doi.org/10.1037/a0028086

Kepner, J. L., & Wackerly, D. D. (1996). On ranked transformation techniques for balanced and incomplete repeated measures. *Journal of American Statistical Association, 91*(436), 1619–1625. https://doi.org/10.1080/01621459.1996.10476730

Koenker, R. (1981). A note on studentizing a test for heteroscedasticity. *Journal of Econometrics, 17*(1), 107–112. https://doi.org/10.1016/0304-4076(81)90062-2

Kössler, W. (2005). Some *c*-sample rank tests of homogeneity against ordered alternatives based on U-statistics. *Journal of Nonparametric Statistics, 17*(7), 777–795. https://doi.org/10.1080/10485250500077254

Keselman, H. J., Huberty, C., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., Kowalchuk, R. K., Lowman, L. L., Petoskey, M. D., Keselman, J. C., & Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research, 68*(3), 350–386. https://doi.org/10.3102/0034654306800335

Leech, N., & Onwuegbuzie, A. (2002, October). A call for greater use of nonparametric statistics [Paper presentation]. Annual meeting of the Mid-South Educational Research Association, Chattanooga, TN.

Lix, L. M., Keselman, J. C., & Keselman, H. L. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance "F" test. *Review of Educational Research, 66*(4), 579–619. https://doi.org/10.3102/00346543066004579

Lix, L. M., & Keselman, H. J. (1998). To trim or not to trim: Tests of location equality under heteroscedasticity and nonnormality. *Education and Psychological Measurement, 58*(3), 409–429. https://doi.org/10.1177/0013164498058003004

Marden, J. L., & Muyot, M. E. (1995). Rank tests for main and interaction effects in analysis of variance. *Journal of American Statistical Association, 90*(432), 1388–1398.

Migdadi, H. S. (2015). On the power performance of test statistics for the generalized Rayleigh interval grouped data. *Open Journal of Statistics, 5*(5), 474–482. https://doi.org/10.4236/ojs.2015.55049

Müller, P., Quintana, F. A., Jara, A., & Hanson, T. (2015). *Bayesian nonparametric data analysis.* Springer. https://doi.org/10.1002/bimj.201600006

National Center for Educational Statistics (2021). *SAT mean scores of college-bound seniors by sex.* U.S. Department of Education, Institute of Educational Sciences. Washington, DC. https://nces.ed.gov/programs/digest/d14/tables/dt14_226.20.asp

Nordstokke, D. W., Zumbo, B. D., Cains, S. L., & Saklofske, D. H. (2011). The operating characteristics of the nonparametric Levene's test for equal variances with assessment and evaluation data. *Practical Assessment, Research & Evaluation, 16*(5), 1–8. https://doi.org/10.7275/5t99-zv93

Nussbaum, E. M. (2014). *Categorical and nonparametric data analysis: Choosing the best statistical technique.* Routledge Taylor & Francis Group. https://doi.org/10.4324/9780203122860

Okoye, K., & Hosseini, S. (2024). Choosing between parametric and non-parametric tests in statistical data analysis. In *R programming: Statistical data analysis* (pp. 87–98). Springer. https://doi.org/10.1007/978-981-97-3385-9_4

Pearson, R. W. (2010). *Statistical persuasion: How to collect, analyze, and present data accurately, honestly and persuasively.* Routledge Taylor & Francis Group.

Puri, M. L., & Sen, P. K. (1969). A class of rank order tests for a general linear hypothesis. *The Annals of Mathematical Statistics, 40*(4), 1325–1343.

Puri, M. L., & Sen, P. K. (1985). *Nonparametric methods in general linear models.* John Wiley & Sons.

Richardson, J. T. E. (2018). The use of Latin-square designs in educational and psychological research. *Educational Research Review, 24,* 84–97. https://doi.org/10.1016/j.edurev.2018.03.003

Skidmore, S. T., & Thompson, B. (2010). Statistical techniques used in published articles: A historical review of reviews. *Educational and Psychological Measurement, 70*(5), 777–795. https://doi.org/10.1177/0013164410379320

Spirko, C. L. (2015). *Teacher attitudes: Factors contributing to teacher attitudes toward students with autism spectrum disorders* (Publication number 3732449) [Doctoral Dissertation, Capella University].

Skovlund, E., & Fenstad, G. U. (2001). Should we always choose a nonparametric test when comparing two apparently nonnormal distributions? *Journal of Clinical Epidemiology*, *54*(1), 86–92. https://doi.org/10.1016/s0895-4356(00)00264-x

Thomas, J. R., Nelson, J. K., & Thomas, K. T. (1999). A generalized rank-order method for nonparametric analysis of data from exercise science: A tutorial. *Research Quarterly for Exercise and Sport, 70*(1), 11–23. https://doi.org/10.1080/02701367.1999.10607726

Trochim, W., & Donnlley, J. P. (2008). *The research methods knowledge base* (3rd ed.). Atomic Dog/Cengage Learning.

Warner, R. M., (2013). *Applied statistics: From bivariate through multivariate techniques* (2nd ed.). Sage Publications.

White, J. L. (2013). Logistic regression model effectiveness: Proportional chance criteria and proportional reduction in error. *Journal of Contemporary Research in Education, 2*(1), 4–10.

Wilcox, R. R. (1998). Can tests for treatment group equality be improved? The bootstrap and trimmed means conjecture. *British Journal of Mathematical and Psychological Psychology, 51*(1), 123–134.

Wilcox, R. R. (2003). *Applying contemporary statistical techniques.* Academic Press.

Wilcox, R. R. (2016). *Introduction to robust estimation and hypothesis testing* (4th ed.). Academic Press.

Wilcox, R. R. (2016). ANOVA: A global test based on a robust measure of location or quartiles when there is curvature. *Journal of Modern Applied Statistical Methods, 15*(1), 12–31. https://doi.org/10.22237/jmasm/1462075320

Wilcox, R. R., Charlin, V. L., & Thompson. K. L. (1986). New Monte Carlo results on the robustness of the ANOVA F, W and F statistics. *Communication in Statistics: Simulation and Computation, 15*(4), 933–943.

Zimmerman, D. W. (2000). Statistical significance levels of nonparametric tests biased by heterogeneous variances of treatment groups. *Journal of General Psychology, 127*(4), 354–364. https://doi.org/10.1080/00221300009598589

| JERAP

The *Journal of Educational Research and Practice* is a peer-reviewed journal that provides a forum for studies and dialogue about developments and change in the field of education and learning. The journal includes research and related content that examine current relevant educational issues and processes. The aim is to provide readers with knowledge and with strategies to use that knowledge in educational or learning environments. *JERAP* focuses on education at all levels and in any setting, and includes peer-reviewed research reports, commentaries, book reviews, interviews of prominent individuals, and reports about educational practice. The journal is sponsored by The Richard W. Riley College of Education and Human Sciences at Walden University, and publication in *JERAP* is always free to authors and readers.