Perspectives in Practice

Publishing Nutrition Research: A Review of Nonparametric Methods, Part 3

JEFFREY E. HARRIS, DrPH, RD; CAROL BOUSHEY, PhD, MPH, RD; BARBARA BRUEMMER, PhD, RD; SUJATA L. ARCHER, PhD, RD

ABSTRACT

This is the third article in a periodic five-part series on publishing nutrition research. These monographs are designed to assist in the interpretation of the *Journal of the American Dietetic Association* author guidelines and provide guidance in publishing and interpreting nutrition-related research articles. This installment focuses on the use of nonparametric statistical methods. The rationale for their use, their advantages and disadvantages, nonparametric alternatives to parametric tests, nonparametric statistical analysis, examples of their use, and helpful resources for further study are topics and issues addressed in this article.

J Am Diet Assoc. 2008;108:1488-1496.

This is the third installment in a series of articles on research design and data analysis. These articles are designed to assist in the interpretation of the *Journal of the American Dietetic Association*'s author guidelines (1) and provide guidance in publishing and interpreting nutrition-related research articles. The purpose of this article is to explain the use of nonparametric statistical tests. As a result of this article, the reader will be able to:

J. E. Harris is an associate professor and didactic program director, Department of Health, Sturzebecker Health Sciences Center, West Chester University, West Chester, PA. C. Boushey is an associate professor and director of Coordinated Program in Dietetics, Purdue University, West Lafayette, IN. B. Bruemmer is a senior lecturer, Department of Epidemiology, director, Didactic Program in Dietetics, and technical advisor, Center for Public Health Nutrition, University of Washington, Graduate Program in Nutritional Sciences, Seattle. S. L. Archer is a research assistant professor, Department of Preventive Medicine, Northwestern University, Feinberg School of Medicine, Chicago, IL.

Address correspondence to: Jeffrey E. Harris, DrPH, MPH, RD, Department of Health, Sturzebecker Health Sciences Center, #302, West Chester University, West Chester, PA 19383. E-mail: jharris@wcupa.edu

Manuscript accepted: March 7, 2008.

0002-8223/08/10809-0008$34.00/0

doi: 10.1016/j.jada.2008.06.426

- Describe the rationale for using nonparametric statistical tests;
- Contrast parametric and nonparametric tests;
- Identify nonparametric alternatives to parametric tests;
- Match appropriate nonparametric tests to univariate and bivariate situations; and
- Analyze elements of nonparametric tests.

It will be helpful to review the first two monographs in the series to most effectively read this monograph (2,3). For assistance in following the progression of this article, refer to Figure 1 for definitions of statistical terms used throughout this article (4).

Most registered dietitians are familiar with parametric statistical tests, such as *t* tests, analysis of variance (ANOVA), Pearson's correlation, etc. These are what most people learn in basic and intermediate statistics courses. These tests are used to determine if there are relationships and differences between variables. So what are nonparametric statistical tests and when are they used?

RATIONALE FOR NONPARAMETRIC TESTS

There are three general situations in which nonparametric tests are used: when assumptions of parametric tests are violated; analyzing ordinal or nominal data; and analyzing data derived from small samples. These situations are discussed in detail here.

When the variable or variables are quantitative and assumptions of parametric tests are violated, nonparametric tests are applied. Parametric statistical tests are those that depend on a defined distribution (such as the normal distribution shown in Figure 2A) and statistics such as means, variances, and standard deviations. Parametric tests have assumptions about the data that must be met in order for the tests to apply effectively. For example, the two-sample *t* test (also called the independent *t* test), which compares two independent groups on a quantitative variable requires the data for each group to be normally distributed (ie, the data, when compiled into a histogram, present as approximately a bell-shaped curve) and the group variances approximately equal. In a sense, the two-sample *t* test compares the means of the data from two groups to see if they are statistically significantly different.

If the data in the groups are not normally distributed, then the means are not useful or meaningful. For instance, if a sample of people had blood urea nitrogen levels of 5, 7, 7, 8, 8, 8, 10, 10, 12, 43, and 54 mg/dL (Figure 2B). From these data it is evident that the distribution is skewed to the right and nonnormal. The ex-

Parametric tests	Statistical tests that concern population parameters (such as means and standard deviations) and require assumptions about these parameters.
Nonparametric tests	Nonparametric tests are statistical tests which make no assumptions about the distribution of a statistical population. Also, no assumptions are made about any statistical parameter.
Nominal variable	A set of data is said to be nominal if the values belonging to it can be assigned a code in the form of a number where the numbers are simply labels. You can count but not order or measure nominal data. For example, in a data set males could be coded as 0, females as 1.
Ordinal variable	A set of data is said to be ordinal if the values belonging to it can be ranked (put in order) or have a rating scale attached. You can count and order, but not measure, ordinal data. The categories for an ordinal set of data have a natural order, for example, suppose a group of people were asked to taste varieties of biscuit and classify each biscuit on a rating scale of 1 to 5, representing strongly dislike, dislike, neutral, like, strongly like. A rating of 5 indicates more enjoyment than a rating of 4, for example, so such data are ordinal. However, the distinction between neighboring points on the scale is not necessarily always the same. For instance, the difference in enjoyment expressed by giving a rating of 2 rather than 1 might be much less than the difference in enjoyment expressed by giving a rating of 4 rather than 3.
Quantitative variable	A variable that takes numerical values for which arithmetic makes sense, for example, counts, temperatures, weights, amounts of money, etc. For some variables that take numerical values, arithmetic with those values does not make sense; such variables are not quantitative. For example, adding and subtracting social security numbers does not make sense. Quantitative variables typically have units of measurement, such as inches, people, or pounds.
Normal distribution	The normal curve is the familiar "bell curve." The normal curve is equally distributed about the mean.
Skewed distribution	A skewed distribution describes a population whose values are not equally distributed about the mean. In a positive skew there are a small number of very large values; this means that when the curve is drawn there is a long tail after the peak. In a negative skew the reverse occurs; there are a small number of small values. The tail appears before the peak.
Variance	The variance of a list is the square of the standard deviation of the list, that is, the average of the squares of the deviations of the numbers in the list from their mean.
Standard deviation	The standard deviation is a calculated number that represents the variation of a set of data.
Mean	The sum of a list of numbers, divided by the number of numbers.
Median	The "middle value" of a list of numbers. The smallest number such that at least half the numbers in the list are no greater than it. If the list has an odd number of entries, the median is the middle entry in the list after sorting the list into increasing order. If the list has an even number of entries, the median is the smaller of the two middle numbers after sorting. The median can be estimated from a histogram by finding the smallest number such that the area under the histogram to the left of that number is 50%.
Likert scale	A Likert scale is a type of psychometric response scale often used in questionnaires, and is the most widely used scale in survey research. When responding to a Likert questionnaire item, respondents specify their level of agreement to a statement. The Likert technique presents a set of attitude statements. Subjects are asked to express agreement or disagreement on a five-point scale. Each degree of agreement is given a numerical value from one to five. Thus a total numerical value can be calculated from all the responses.
Post hoc testing	Statistical testing that is done to determine specific difference between groups after one has run an analysis of variance (ANOVA) or its nonparametric version.
Test statistic	A statistic used to test hypotheses. A hypothesis test can be constructed by deciding to reject the null hypothesis when the value of the test statistic is in some range or collection of ranges. It is the calculated value that is used to decide if a result is statistically significant or not. Calculated t values, r's, and F's are examples of test statistics.
Statistical significance	The probability of a false rejection of the null hypothesis in a statistical test. It is the probability that we have wrongly concluded that there is a statistically significant difference between groups or a statistically significant relationship between variables.
Level of significance	The significance level of a hypothesis test is the chance that the test erroneously rejects the null hypothesis when the null hypothesis is true. It is usually set at less than 5%, meaning that there is a less than 5% chance that we are erroneously rejecting the null hypothesis.
P value	It is the actual probability of wrongly rejecting the null hypothesis if it is in fact true. The *P* value is compared with the actual significance level of our test and, if it is smaller, the result is significant. That is, if the null hypothesis were to be rejected at the 5% significance level, this would be reported as "$P<0.05$". Small *P* values suggest that the null hypothesis is unlikely to be true. The smaller it is, the more convincing is the rejection of the null hypothesis.
Univariate	Having or having to do with a single variable. Some univariate techniques and statistics include the histogram, mean, median, percentiles, and standard deviation.

(continued)

Figure 1. Definitions of key statistical terms.

Bivariate	Having or having to do with two variables. For example, bivariate data are data where we have two measurements of each "individual." These measurements might be the heights and weights of a group of people (an "individual" is a person), the heights of fathers and sons (an "individual" is a father-son pair), the pressure and temperature of a fixed volume of gas (an "individual" is the volume of gas under a certain set of experimental conditions), etc. Scatterplots, the correlation coefficient, and regression make sense for bivariate data but not univariate data.
Null hypothesis	In hypothesis testing, the hypothesis we wish to falsify on the basis of the data. The null hypothesis is typically that something is not present, that there is no effect, there is no relationship, or that there is no difference between treatment and control. For example, "there is no significant relationship between body mass index and serum cholesterol," is a null hypothesis.
Histogram	A histogram is a kind of plot that summarizes how data are distributed. Starting with a set of class intervals, the histogram is a set of rectangles sitting on the horizontal axis. The bases of the rectangles are the class intervals, and their heights are such that their areas are proportional to the fraction of observations in the corresponding class intervals. That is, the height of a given rectangle is the fraction of observations in the corresponding class interval, divided by the length of the corresponding class interval. A histogram does not need a vertical scale, because the total area of the histogram must equal 100%. The units of the vertical axis are percent per unit of the horizontal axis. The horizontal axis of a histogram needs a scale.
Natural log	The natural logarithm is the logarithm to the base e, where e is a certain constant approximately equal to 2.718281828459. The logarithms most people are familiar with are those with base 10. For natural logarithms, ln (x), the base is e.
Range	The range is the difference between the lowest value and the highest value in a data set.
Central limit theorem	The central limit theorem states that given a distribution with a mean μ and variance σ^2, the sampling distribution of the mean approaches a normal distribution with a mean (μ) and a variance σ^2/N as N, the sample size, increases (greater than or equal to 30).
Binomial distribution	The theoretical frequency distribution of events that have two possible outcomes.

Figure 1. Continued

treme values will substantially alter the mean so it will not be a valid representation of a measure of central tendency. Without the values of 43 and 54 mg/dL, the data would be normally distributed with a mean of 8.3 mg/dL. With those extreme values the mean is 15.6 mg/dL, which is not as representative a measure of central tendency as the median, which is 8 mg/dL. Measures of central tendency and their relevance to nonparametric tests are discussed in greater detail later in this monograph. So what can be done to analyze nonnormal data?

Options exist to deal with nonnormally distributed data (3). One approach is to apply a mathematical transformation to the data to attempt to normalize the data distribution. For example, in a study of the relationship of ghrelin and leptin to body mass index and waist circumference, Monti and colleagues (5) found that serum ghrelin and leptin were not normally distributed. These authors calculated the natural log (see Figure 1 for a definition of natural log) of these two variables, attempting to correct for the nonnormality. By calculating the natural log of each data value the distribution of the variable shifted from skewed to normally distributed. This correction then permitted the authors to apply parametric statistical tests to the data. However, at times this method may not be successful in that the transformation is inadequate to produce a normal distribution. It is in these situations that nonparametric methods are used.

Suppose right-skewed blood urea data from another sample were to be compared to the sample described above to determine if there was a statistically significant difference between samples. If the serum urea data were normally distributed then the parametric two-sample *t* test would be used for the comparison. However, if the serum urea data could not be normalized with a transformation, then the parametric two-sample *t* test would not be applied. If this is the case, how should the data be analyzed?

Violations of parametric test assumptions necessitate the use of nonparametric tests. Nonparametric tests are not dependent on a defined distribution (that is why they are often called distribution-free tests) or on statistical parameters such as means, standard deviations, and variances. Therefore the next step is to choose the nonparametric alternative to the parametric method (two-sample *t* test) to conduct the statistical analysis. In the previous example, the nonparametric test, the Mann-Whitney U test (Figure 3 contains a decision rubric for choosing appropriate nonparametric tests), is more appropriate in that it does not rely on the assumptions of normally distributed data and equal variances between the two groups being compared.

The second general situation in which nonparametric statistical tests are used is if the data are ordinal or nominal, such as rating scales (eg, Likert scales; rating of taste intensity on a 5-point scale) or categories (eg, sex; male, female) (2). In this situation parametric tests generally do not apply. For instance, if study subjects are rating their like for skim milk on a 5-point Likert scale from 1 "significantly dislike" to 5 "significantly like," a rating of 5 is not considered to be 5 times a rating of 1. Also, the preference distance from 1 to 2 is not necessarily the same distance from 2 to 3. In this case of ordinal data, the mean is not meaningful and the variation of responses is not sufficient to see a meaningful normal dis-

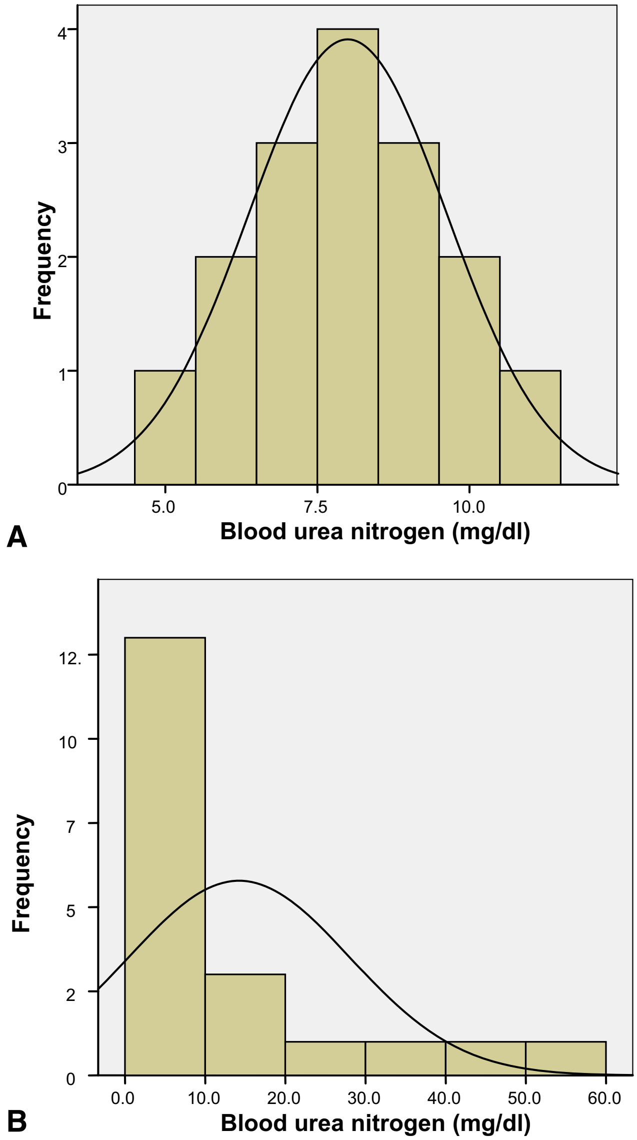

Figure 2. Normal vs right-skewed distribution for blood urea nitrogen. (A) An illustration of a normal distribution as evidenced by the bell-shaped curve. (B) An illustration of a right-skewed distribution.

tribution. In addition, means and standard deviations are meaningless for nominal variables. There is no such thing as a mean sex. Hence, nonparametric tests are preferred for use with ordinal and nominal data. For example, suppose both males and females were asked on a questionnaire to rate their dietary preference for skim milk on the Likert scale mentioned here. To compare males and females as to their preferences for skim milk, a nonparametric test would be used. In this situation, the relationship between a nominal variable (sex) and an ordinal variable (preference for skim milk) would be tested. When this type of relationship is being tested, the nonparametric χ^2 test would be appropriate (Figure 3 contains a decision rubric for deciding on the relevant nonparametric test).

The third general situation is when sample sizes are small. If the data are not normally distributed for a quantitative variable, but the sample size is significantly large ($n \geq 30$), the Central Limit Theorem (see Figure 1 for a definition) will apply so that parametric tests can be used. The Central Limit Theorem cannot be applied to small sample sizes; therefore, it is then appropriate to use nonparametric tests.

ADVANTAGES AND DISADVANTAGES OF NONPARAMETRIC TESTS

There are three advantages to the use of nonparametric tests. First, they are not dependent on a type of distribution (eg, normal). Second, they are not dependent on the mean, standard deviation, or variance. Finally, they provide useful statistical test options for ordinal and nominal data.

However, there are also disadvantages to using nonparametric tests. They do not use all the characteristics of the data (eg, means and standard deviations), but often use ranks and directions (positive or negative) of the data. Also, because they do not use all the characteristics of the data, the results of the tests tend to be more conservative than parametric tests. This means that if a null hypothesis for a study is false, the nonparametric test is less likely to reject it than a parametric test. Suppose the null hypothesis, "there is no statistically significant difference in diastolic blood pressure between those receiving evening primrose oil vs those receiving corn oil," is false, the appropriate nonparametric test when applied to data from samples is less likely to reject this null hypothesis than the corresponding parametric test.

NONPARAMETRIC ALTERNATIVES TO PARAMETRIC TESTS

Most parametric tests have nonparametric versions. So if the assumptions for a parametric test application are violated, a nonparametric alternative can be used to analyze the data. For example, the nonparametric alternative to one-way ANOVA is the Kruskal-Wallis test. Figure 3 presents the type of variable situation and parametric tests and their nonparametric counterparts. This rubric is helpful in selecting the appropriate nonparametric alternatives to parametric statistical tests. The following sections address nonparametric statistical analysis.

MEASURE OF CENTRAL TENDENCY FOR NONPARAMETRIC SITUATIONS

The most common measures of central tendency are the mean, median, and mode. When data are normally distributed these statistics are equal. When data are skewed they are not equal and the median is the best measure of central tendency because it is least affected by extreme values. The serum urea nitrogen measurement example given previously in this article illustrates this phenomenon. When using the nonparametric test, the Mann-Whitney U test, a statistically significant difference between medians is tested, rather than means, as used in the parametric two-sample *t* test. When reporting a measure of central tendency in the context of a nonparametric situation, it is most appropriate to report the medians rather than means. The best measure of variation to report with the

Variables	Parametric test	Nonparametric test
Quantitative data from a sample compared to a reference or standard	One-sample t test	One-sample Signed Rank test
Paired quantitative data (eg, before-after, matched pairs)	Paired t test	Wilcoxon Signed-Rank test McNemar's test
Two independent quantitative samples (difference between samples)	Two-Sample t test	Mann-Whitney U test Wilcoxon Rank-Sum test Sign test Kolmogorov-Smirnov test Median test χ^2 test
Greater than two independent quantitative samples (difference between samples)	One-Way Analysis of Variance	Kruskal-Wallis test
Two quantitative variables (relationship)	Pearson's Product Moment Correlation	Spearman's Rank Correlation
Randomized Block or Repeated Measures	Multiway Analysis of Variance	Friedman's test
Two Qualitative Variables (relationship)		χ^2 test
Paired Qualitative Variables		McNemar's test

Figure 3. Decision-making rubric for choosing appropriate parametric and nonparametric tests given the types of variables.

median is the range (the difference between the lowest and highest value), rather than the standard deviation, which is reported with means.

With ordinal and nominal data there is no appropriate measure of central tendency to report. Counts and percentages for each category are appropriate for presenting the data.

EXAMPLE OF A NONPARAMETRIC PROCEDURE CALCULATION

Before examining the application of different nonparametric tests, an example of a nonparametric procedure calculation is presented in this section. This will provide an understanding of how a nonparametric statistical test differs from a parametric one.

Suppose there are two independent samples of quantitative data and there is a desire to see if there is a statistically significant difference between these samples. For example, five subjects with the metabolic syndrome are compared to five who do not have the metabolic syndrome for hemoglobin A1c. For the metabolic syndrome group, the hemoglobin A1c values are 6.5%, 7.0%, 7.8%, 7.0%, and 13%. For the other group the values are 3.0%, 2.5%, 4.0%, 1.0%, and 4.0%. For this example, if there were larger sample sizes and the data in each group were normally distributed the parametric, two-sample t test would be used. However, with small sample sizes and nonnormal data, the nonparametric Mann-Whitney U test would be chosen from Figure 3. Figure 4 illustrates the calculations for the Mann-Whitney U test. As illustrated, nonparametric tests often use ranks rather than the raw data. Also, medians are the best measures of central tendency to report and compare for this example, and ranges are the best measures of variation to report, rather than standard deviations.

n_a = *number of results in metabolic syndrome group (MSG)* = 5

n_b = *number of results in comparison group (CG)* = 5

T_a = *sum of ranks for MSG group*

T_b = *sum of ranks for CG group*

Take data from both groups and order them from lowest to highest

Rank the data from lowest to highest starting with 1; *for ties give the mean rank*

Next, assign the appropriate group label to the ranks

1.0%	2.5%	3.0%	4.0%	4.0%	6.5%	7.0%	7.0%	7.8%	13.0%
1	2	3	4.5	4.5	6	7.5	7.5	9	10
CG	*CG*	*CG*	*CG*	*CG*	*MSG*	*MSG*	*MSG*	*MSG*	*MSG*

Add the ranks for each group which gives you T_a *and* T_b

$T_a=40;\ T_b=15$

The appropriate U statistic is given by the smaller of:

$$T_a-\frac{n_a(n_a+1)}{2} \text{ or } T_b-\frac{n_b(n_b+1)}{2}$$

$$40-\frac{5(6)}{2}=40-15=25$$

$$15-\frac{5(6)}{2}=15-15=0$$

Look up critical value in a table for $\alpha=.05$

Critical Value = 1

$0<1$

There is a statistically significant difference between groups for HgB−A1c.

$Median_a=7.8\%$

$Median_b=3.0\%$

Those with metabolic syndrome have a higher HgB−A1c than those without metabolic syndrome.

Figure 4. Example of Mann-Whitney U test calculations and interpretation.

SPECIFIC NONPARAMETRIC TESTS

The following sections give illustrations of the choice and application of nonparametric statistical tests. Please refer back to Figure 3 as a reference for choosing the appropriate tests. In these situations it is assumed that basic assumptions for the conduct of parametric tests are violated, necessitating the use of a nonparametric test.

ONE-SAMPLE SIGNED RANK TEST

Suppose a researcher wishes to see if the calcium intakes of a sample of 50 postmenopausal female faculty and staff members are <1,200 mg/day. In this case, assume that dietary calcium intake, which is a quantitative variable, is not normally distributed. The One-Sample signed rank test would be used rather than the parametric One-Sample *t* test to compare the median of the sample and the standard to assess if there is a statistically significant difference. If the median intake of the sample is 800 mg/day this test will identify if this value is statistically significantly lower than 1,200 mg/day. This test can also be used when the variable measured for the sample is ordinal. The outcome of the test is the sum of negative ranks (T_-) and the sum of positive ranks (T_+). The smaller of these ranks is reported as the test statistic in publications. If there is no statistically significant difference between the sample data and the standard, T_- is approximately equal to T_+.

WILCOXON SIGNED-RANK TEST

The Wilcoxon signed-rank test applies in situations where two sets of quantitative or ordinal values are not independent, and assumptions for parametric analysis are violated. Situations that would warrant the use of this test are pretest-posttest comparisons and comparison of data between matched pairs (eg, identical twins). The parametric counterpart to this test is the paired *t* test.

To determine whether supplementary potassium affects diastolic blood pressure in a sample of 100 subjects, this test would be used. Diastolic blood pressure would be measured before and after supplementing potassium to assess statistically significant change. Suppose the median diastolic blood pressures before and after were 110 mm Hg and 85 mm Hg, respectively. The Wilcoxon signed-rank test would be the appropriate nonparametric test to determine if there is a statistically before-and-after difference.

Another situation that would warrant this test is one in which groups of identical twins are compared (with one twin in one group and the other in the second group). If one group is exposed to a treatment and the other to control conditions, this can be treated as a matched pair situation because the twins are genetically alike. For example, suppose one group of twins is exposed to a diet high in soluble fiber and the other to a control diet for 6 weeks, and the groups are compared on serum low-density lipoprotein levels after the intervention. Because one twin is in the control group and the other in the intervention group, the control and intervention groups are considered matched. In this situation, the Wilcoxon signed-rank test is used. Later in this article, an example will be given from this *Journal*, describing the use of this test as a post-hoc analysis strategy for conducting Friedman's test.

The outcome of the Wilcoxon signed-rank test for paired data is a positive rank sum and a negative rank sum as reported in the section on the One-Sample Wilcoxon signed-rank test. Again, the smaller of the two ranks sums is reported as the test statistic, and these sums are approximately equal if there is no statistically significant difference.

MANN-WHITNEY U/WILCOXON RANK-SUM TEST

The nonparametric version of the two-sample *t* test most often used currently is known as the Mann-Whitney U test. It is also known by another name, the Wilcoxon rank-sum test.

An example of the application of this test has already been given previously in this article. This test is used when a comparison is being made between two independent groups on a quantitative or ordinal variable. Essentially, it is used to examine if there is a statistically significant difference between the medians of the two samples.

For instance, Darmon and colleagues (6) developed a scoring system to estimate the nutritional adequacy of fruits and vegetables, and found the three scores (nutrient adequacy, nutrient density, and nutrient-to-price) to be nonnormally distributed. They appropriately reported median scores, rather than the parametric equivalent. When comparing the medians for each of the three scores between two groups, fruits and vegetables and other foods, they used the Mann-Whitney U to determine if there were statistically significant differences. They found that there were statistically significant differences between fruits and vegetables and other foods on all three scores (nutrient adequacy, nutrient density, and nutrient-to-price). The test statistic to be reported in publications is the U value. As can be seen in Figure 4, the U value is the smaller of the two sum of ranks. In the case of the example given in Figure 4, U=0 (vs the higher of the two rank sums, 25). If there is not a difference, the rank sums will be approximately equal.

KRUSKAL-WALLIS TEST

If more than two independent groups are compared on a quantitative or ordinal variable, and assumptions for the parametric one-way ANOVA are violated, the Kruskal-Wallis test is warranted. For example, if the taste satisfaction for various types of milk is compared among four samples of similar people, this test could be used. Suppose 80 children were randomly assigned to four groups (20 in each group). Groups were given either skim milk, 1% milk, 2% milk, or whole milk. Each group consumes their assigned type of milk and then must rate it on an ordinal scale from 1 "greatly dislike" to 5 "like very much." The median taste satisfaction for each group was 2 (skim), 3 (1%), 4 (2%), and 4 (whole). The Kruskal-Wallis test is the appropriate statistical test to determine statistically significant differences between multiple groups. Statistical significance for this test indicates that a difference exists somewhere between the groups, but nonparametric post-hoc testing is necessary to identify what specific groups are different. The test statistic to report in publications is the H value.

A study by Wiecha and colleagues (7) examining the potential relationship between school vending machine and fast-food restaurant use and sugar-sweetened beverage intake in youth provides a practical example of the application of the Kruskal-Wallis test. The main outcome variable in this study, daily intake of sugar-sweetened beverages, was not normally distributed. Nonparametric statistical tests were used in the analysis of the data. Wiecha and colleagues used the Kruskal-Wallis test to determine if there was a statistically significant difference in daily consumption of sugar-sweetened beverages based on frequency of eating at fast-food restaurants in the past week. The former variable was a continuous variable and the latter a categorical variable with four categories; none, one, one to three, and greater than or equal to four. The Kruskal-Wallis test was statistically significant, meaning that the investigators found that those eating more frequently at fast-food restaurants had higher daily intakes of sugar-sweetened beverages.

SPEARMAN'S RANK CORRELATION

Researchers calculate correlations to test for potential relationships between two quantitative, two ordinal, or an ordinal and quantitative variable. When testing the potential relationship for the latter two conditions the nonparametric Spearman's rank correlation is calculated. Also, if the two quantitative variables are not normally distributed Spearman's correlation is calculated, rather than the parametric Pearson's correlation.

Suppose managers of a particular fast-food chain are surveyed as to their knowledge and attitudes about food safety. Attitude is rated on a 10-point Likert scale, and knowledge is scored from 0 to 20. Because attitude is an ordinal variable and an association is being examined with a quantitative variable (knowledge), the Spearman's correlation is the appropriate statistical test. Spearman's correlation (r_s) ranges from -1.0 to $+1.0$. The closer the value is to -1.0 or $+1.0$ the stronger the relationship. The sign indicates if there is a direct or inverse relationship. The test statistic for this test is the r_s.

For the example just given, an r_s approaching 1.0 would indicate that it is highly likely that as knowledge increases about food safety, so does attitude. An $r_s$approaching -1.0 would indicate that as knowledge increases about foods safety, attitude decreases.

For example, Delahanty and colleagues (8) sought to identify predictors of physical activity in the Diabetes Prevention Program. They calculated Spearman's correlation to identify the direction and strength of potential relationships between baseline physical activity (activity before participating in Diabetes Prevention Program) and psychological variables such as exercise efficacy, anxiety, depression, and perceived stress. These psychological variables were all ordinal. They also used this statistical test to examine potential relationships between physical activity 1 year and 2 to 3 years after having receiving the Diabetes Prevention Program, and the same psychological variables mention here. At all three times, physical activity was directly associated with exercise efficacy and inversely associated with anxiety, depression, and perceived stress.

Table 1. Hypothetical data for illustrating the use of Friedman's test[a]

Subjects	Soy	Low-fat cheese	Regular cheese
1	2	3	1
2	2	3	1
3	3	0	2
4	3	2	1
5	3	2	1
6	3	2	1
7	2	3	1
8	1	3	2
9	3	1	2
10	3	1	2
11	3	2	1
12	3	2	1
13	3	1	2
14	3	1	2
15	3	1	2
16	3	1	2
17	2	3	1
18	3	2	1
19	2	3	1
20	3	1	2

[a]The number in each cell for a subject represents preference for macaroni and cheese made with different types of cheese (soy, low-fat, regular). 1=most preferred macaroni and cheese and a 3=least preferred. The Friedman's test is used to determine whether there is a statistically significant difference in preference for type of macaroni and cheese.

FRIEDMAN'S TEST

The appropriate nonparametric test to use when analyzing the data from a repeated measure or randomized block study design is Friedman's test. Typically, multiway ANOVA would be used in parametric applications. Friedman's test is used when assumptions about normality of the data cannot be made. The Friedman's test follows the χ^2 distribution, so the test statistic is χ^2. The following is an example of its application: A food company recruits 20 subjects to test preference for macaroni and cheese. Three types of macaroni and cheese are prepared, one with soy cheese, one with lowfat cheese, and the final one with full-fat cheddar cheese. Each subject is presented with the three types of macaroni and cheese in random order. They rank their preference from one through three. Table 1 illustrates this type of data. Because the type of data is ordinal, Friedman's test is used to determine if there is a statistically significant difference in preference for macaroni and cheese. The company may use this data to decide which product to mass produce and market.

Peterson and colleagues (9) assessed growth and seizure reduction in epileptic children using the ketogenic diet as a treatment for intractable epilepsy. Growth was assessed at baseline, 6 months, and 12 months in a group exhibiting high ketosis and another group with moderate ketosis. Because it was a repeated-measure design and data were not normally distributed, the Friedman's test was used rather than the parametric repeated measures ANOVA to determine if there was a statistically significant difference in growth over time in both groups. The Wilcoxon signed-rank test was used to make pairwise-comparisons between different time periods (eg, baseline vs 6 months, 6 months vs 12 months) as follow-up post hoc analysis because this test is used to examine paired within-group measurements. Subjects on the moderate ketosis diet did not experience decreases in growth over time, however, those on the high ketosis diet did.

χ^2 TEST

The χ^2 test is used to examine the relationship between two nominal variables, two ordinal variables with limited categories, or a nominal and an ordinal variable. The closest parametric alternative is the two-sample t test. The χ^2 test follows the binomial distribution. The test statistic is χ^2.

Suppose a relationship between two nominal variables, religious preference (Hindu, Buddhist, Catholic, Protestant, Moslem) and preference for soy products (Do you like soy products, yes or no?), is examined. Do Hindus prefer soy products beyond what you would expect by chance? Are there some groups that have less preference than you would expect by chance? In this situation, the χ^2 test is the appropriate test to examine these questions.

Fulkerson and colleagues (10) examined the family mealtime environment from the perspectives of adolescents and parents. Parents were surveyed by telephone interview while adolescents completed a school-based survey. The investigators looked for associations between family meal environment categorical variables (frequency, priority, atmosphere, and structure) and adolescents and parents responses to the survey (this was a categorical variable with three categories; parents, younger adolescents, and older adolescents). Each family meal environment variable was a multiple choice question on both the parents' and adolescents' surveys. The χ^2 test demonstrated statistically different reports by adolescents and parents on nearly all family meal environment variables. Parents were more likely than adolescents to report eating five or more family meals per week, the importance of eating together, and scheduling difficulties.

MCNEMAR'S TEST

In situations where samples are matched, or there is a before-after design, and there are two nominal variables with two categories each, McNemar's test is used. The Wilcoxon signed-rank test was discussed previously as an appropriate nonparametric test for matched samples or before-and-after designed, but it applies when two dependent samples are compared on an ordinal or quantitative variable, rather than a nominal variable. The closest parametric alternative is the paired t test. The test statistic for McNemar's test is χ^2. Table 2 presents a 2×2 table comparing two nominal variables, saturated fat intake (high or low) before a myocardial infarction (MI) and saturated fat intake (high or low) after an MI. As can be seen from Table 2, a disproportionate number of subjects switched from a high to a low saturated fat diet after having an MI. McNemar's test is used to determine if there is a statistically significant change in category frequencies from before to after an MI.

Table 2. Hypothetical data presented in a 2×2 table for calculating McNemar's test[a]

Before MI[b]	After MI Saturated Fat Intake		
Saturated fat intake	Low	High	Totals
Low	15	5	20
High	35	15	50
	50	20	70

[a]In this case, McNemar's test is used to determine whether there is a statistically significant change in proportion of subjects eating a low saturated fat before and after subjects have had a myocardial infarction.
[b]MI=myocardial infarction.

In a study to characterize patterns of dairy intake among a cohort of girls progressing from 5 to 11 years of age, Fiorito and colleagues (11) used McNemar's test to analyze whether there was a change in percentage of girls consuming milk and other dairy products over time. The test revealed a statistically significant reduction in percentage of girls consuming milk and yogurt as they progressed from 5 to 11 years of age, but a statistically significant increase in the consumption of cheese and dairy desserts.

REPORTING NONPARAMETRIC STATISTICS IN JOURNAL ARTICLES

In journal articles, it is vital when reporting nonparametric tests to describe the specific variables involved, the specific test, and the hypothesis to be tested. Also, when reporting the test results, present the test statistic (eg, χ^2, r_s), the level of significance (eg, P=0.036), and the appropriate conclusion.

RESOURCES FOR FURTHER STUDY

There are several excellent books with more specific information about nonparametric tests (4,12-14). Most notably, *Introduction to Modern Nonparametric Statistics* by Higgins (15) provides a comprehensive presentation of the subject. In the previous article in this series, useful computer statistical packages were listed. All of these packages have the capability to conduct nonparametric analyses.

SUMMARY

When there is a choice, it is preferable to use parametric over nonparametric statistical tests. However, when there are ordinal or nominal variables, or the assumptions for parametric tests are violated, nonparametric statistical tests can be used.

The authors received no funding for this project. There are no financial conflicts of interest.

References

1. American Dietetic Association. *Journal of the American Dietetic Association.* Author guidelines. *J Am Diet Assoc.* 2006;106:140-147.

2. Boushey C, Harris J, Bruemmer B, Archer SL, Van Horn L. Publishing nutrition research: A review of study design, statistical analyses, and other key elements of manuscript preparation, Part 1. *J Am Diet Assoc.* 2006;106:89-96.
3. Boushey C, Harris J, Bruemmer B, Archer SL. Publishing nutrition research: A review of sampling, sample size, statistical analysis, and other key elements of manuscript preparation, Part 2. *J Am Diet Assoc.* 2008;108:679-688.
4. Rosner B. *Fundamentals of Biostatistics.* 6th ed. Clifton Park, NY: Cengage Learning; 2006.
5. Monti V, Carlson JJ, Hunt SC, Adams TD. Relationship of ghrelin and leptin hormones with body mass index and waist circumference in a random sample of adults. *J Am Diet Assoc.* 2006;106:822-828.
6. Darmon N, Darmon M, Maillot M, Drewnowski A. A nutrient density standard for vegetables and fruits: nutrients per calorie and nutrients per unit cost. *J Am Diet Assoc.* 2005;105:1881-1887.
7. Wiecha JL, Finkelstein D, Troped PJ, Fragala M, Peterson KE. School vending machine use and fast-food restaurant use are associated with sugar-sweetened beverage intake in youth. *J Am Diet Assoc.* 2006; 106:1624-1630.
8. Delahanty LM, Conroy M, Nathan DM, Diabetes Prevention Program Research Group. Psychological predictors of physical activity in the Diabetes Prevention Program. *J Am Diet Assoc.* 2006;106:698-705.
9. Peterson SJ, Tangney CC, Pimentel-Zablah EM, Hjelmgren B, Booth G, Berry-Kravis E. Changes in growth and seizure reduction in children on the ketogenic diet as a treatment for intractable epilepsy. *J Am Diet Assoc.* 2005;105:718-725.
10. Fulkerson JA, Neumark-Sztainer D, Story M. Adolescent and parent views of family meals. *J Am Diet Assoc.* 2006;106:526-532.
11. Fiorito LM, Mitchell DC, Smicklas-Wright H, Birch LL. Dairy and dairy-related nutrient intake during middle childhood. *J Am Diet Assoc.* 2006;106:534-542.
12. Kuzma JW, Bohnenblust SE. *Basic Statistics for the Health Sciences.* 5th ed. New York, NY: McGraw Hill; 2005.
13. D'Agostino RB, Sullivan LM, Beiser AS. *Introductory Applied Biostatistics.* 1st ed. Clifton Park, NY: Cengage Learning; 2006.
14. Pagano M, Gauvreau K. *Principles of Biostatistics.* 2nd ed. Clifton Park, NY: Cengage Learning; 2000.
15. Higgins JJ. *Introduction to Modern Nonparametric Statistics.* 1st ed. Clifton Park, NY: Cengage Learning; 2003.