Non- Parametric Statistics: A Set of Statistical Techniques to Compare Two or More Independent Populations

Athanasios Vasilopoulos
St. John's University

Nonparametric methods are a powerful research tool used by investigators in practically every field of human activity. Nonparametric methods are useful alternatives to the parametric methods (when parametric methods are not available) and their use and application is made much easier when statistical tools, like MINITAB, are used to solve problems completely or partially. The techniques of we discuss in nonparametric statistics fall in the following 5 categories:

I) Tests for Randomness
II) Chi-Square Tests
III) Tests for Matched Pairs
IV) Tests to Compare 2 or More Independent Populations
V) Spearman Rank Correlation Test

MINITAB examples are given for: Tests of Randomness, Tests for Matched Pairs (Wilcoxon Sign Rank Test, Friedman Test) and Tests to Compare 2 or More Independent Populations (Mann-Whitney Test, Kruskal-Wallis H Test).

Keywords: Parametric Methods, Non-Parametric Methods, Tests for Randomness, Chi-square tests, Tests for Matched Pairs, Tests to Compare 2 or more Independent Populations, and the Spearman Rank Correlation Test

Problem Statement

When researchers are testing the validity of claims during their research, and no Parametric methods exists to perform such tests, what do they do? The answer is to use Non-Parametric methods which may take many forms.

Approach

We divide the many non-parametric techniques in five (5) categories, discuss representative techniques in each category and, where possible, compare the manual Non-Parametric methods with the MINITAB Non-Parametric methods, if they are available. But in this paper, we discuss the Non-Parametric methods of category IV above, (i.e. Tests to Compare two or more Independent Populations), having discussed the techniques of the other categories in previous papers.

Result

The results indicate that Non-Parametric methods are very powerful and useful alternatives to the Parametric Methods, when Parametric Methods are not available. Many examples, throughout the paper,

show how the various tests are applied and the use and application of the statistical tool MINITAB shows how easily these tests can be performed.

Conclusions/Recommendations

When parametric methods are not available, a useful alternative approach is to us Non-Parametric methods which, as shown, provide the solution to many research problems. Use the Statistical too MINITAB, to make the application of these Non-Parametric techniques procedural.

INTRODUCTION

Most of the statistical methods that people are familiar with are referred to as Parametric Statistics and the term is used to indicate that we have knowledge of the nature of the population from which the sample data set: $x_1, x_2, x_3, , x_n$, which we are about to analyze, came from. For example, when we use the EMPIRICAL RULE, we make the assumption that our sample data came from a normal distribution. However, when we use CHEBYCHEV'S INEQUALITY, namely:

$$P[\bar{x} - k\hat{s} \leq X \leq \bar{x} + k\hat{s}] \geq 1 - \frac{1}{k^2}, \text{ for k>1} \tag{1}$$

which states that the probability that a random variable X is between k standard deviations of the mean

$(\bar{x})$ is at least $1 - \frac{1}{k^2}$, the result is valid for all possible distributions of the random variable X [12].

Such a "distribution-free" result, and similar other tests that we will discuss in this paper, are called non-parametric statistics. In these nonparametric tests the parameters of the distribution continue to be important. What is not important is the nature of the distribution of the population from which the sample came from. That is these tests are valid whether the population distribution is Normal, Binomial, Uniform, Exponential, etc. As we discuss the availability and development of nonparametric statistics, we will discuss several tests which are equivalent to familiar parametric tests. When this is the case it is natural to ask: Which test is preferable, the parametric or nonparametric one? The answer is: the parametric test, because for the non-parametric tests to achieve the same results (i.e. the same power) require a much larger sample (i.e. nonparametric methods are less efficient than parametric methods).

But there are many situations in which the form (or nature) of the population distribution is not well known and the nonparametric method is the only meaningful alternative. This is also the case when no parametric alternative exists.

However, even though the computations on nonparametric statistics are usually less complicated than those for parametric statistics, the calculations for many nonparametric statistics can become very tedious, when the samples are large.

Another disadvantage for most nonparametric methods is the fact that the null hypothesis (H_0) being tested is less precise than in the parametric methods, and the conclusions drawn may be somewhat vague. But, even with these drawbacks, nonparametric statistics are very useful, and it is important to know when and where they can be used, and the conclusions which can be drawn from their application.

DISCUSSION

Before starting our discussion on Non- Parametric Statistics, let us do a Parametric Statistics example to make the comparison easier to understand.

Example 1: - The claim is made that "the mean Carbon dioxide level of air pollution in New York City is 4.9 particles per cubic foot".

Does a random sample of 25 readings (which produced an $\overline{X}=5.1$ and $\hat{S}=2.1$) present sufficient evidence to accept or reject this claim at $\alpha= 0.05$?

Also construct a 95% Confidence Interval on μ and compare the results of the 2 solutions. Then, the hypothesis Test solution is:

1) $H_o : \mu = 4.9$ vs $H_1: \mu \neq 4.9$

2) $\alpha= 0.05$(and $\alpha/2 = 0.025$)

3) The estimator for μ is $\overline{X}$ and, because σ is not known and n= 25<30,

$\frac{\overline{X}-\mu_0}{\hat{S}(\overline{X})} = t_{n-1} = t_{24}$ distributed.

4) Rejection region/Acceptance Region

-2.064 $t_{24}(0.025) = 2.064$
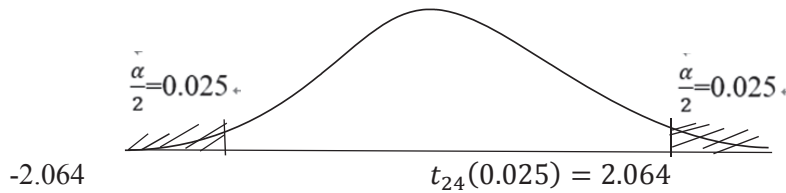
5) Value of test statistic $= t^* = \frac{\overline{X}-\mu_0}{\hat{S}(\overline{X})} = \frac{5.1-4.9}{\frac{2.1}{\sqrt{25}}} = \frac{0.2}{\frac{2.1}{5}} = \frac{1.0}{2.1} = 0.476$

6) Since t^* falls outside of RR, We do not reject H_o: $\mu = 4.9$

7) Therefore, the mean Carbon Dioxide level of air pollution in New York City is
 $\mu= 4.9$.

We can also construct a Confidence Interval on μ using the equation :

$P[\overline{X} - t_{n-1}(\alpha/2)\hat{S}(\overline{X}) <= \mu <= \overline{X} + t_{n-1}(\alpha/2)\hat{S}(\overline{X})] = 1-\alpha$ which , because $\overline{X} = 5.1$, $t_{n-1}(\alpha/2) = t_{24}(0.025) = 2.064$,

$\hat{S}(\overline{X}) = \hat{s}/\sqrt{n} = 2.1/\sqrt{25} = 2.1/5 = 0.42$ becomes upon substitution:

$$P [4.23 \leq \mu \leq 5.97] = 0.95$$

Then, since the hypothesis value of $\mu = 4.9$ falls INSIDE of this confidence interval, H_o: $\mu = 4.9$, is not rejected. This is the same conclusion we reached from the hypothesis Test solution to the problem, above. The nonparametric statistics available are usually grouped in 5 categories, as shown in the table below, which also shows corresponding parametric tests, when they exist.

TABLE 1
EQUIVALENCY OF NON-PARAMETRIC AND PARAMETRIC TESTS

NON-PARAMETRIC TESTS	CORRESPONDING PARAMETRIC TESTS

I) TESTS FOR RANDOMNESS
   1) Runs Tests
     a) Small Sample Runs test
     b) Large Sample Runs test
   2) Runs Above and Below the Median
   3) Runs Up and Down
   4) The Von Newmann Ratio Test for Independence
II) CHI-SQUARE TESTS
   1) Tests on Frequencies
     a) Tests on the Frequency of 2 classes
     b) Tests on the Frequency of more than 2 classes
   2) Tests on Independence – Contingency Tables
   3) Tests on Homogeneity - Are 2 or more Independent random samples drawn from the same population?
   4) Tests on Goodness of Fit
     a) For the Uniform Distribution
     b) For the Normal Distribution
   5) Kolmogorov-Smirnoff Statistic - A more appropriate goodness of Fit test (than χ^2)
III) TESTS FOR MATCHED PAIRS
   1) The Sign test - Uses only the sign of the difference
   2) The Wilcoxon Signed test - Uses both sign of difference and rank
     a) Small sample ($n < 16$)
     b) Large Sample ($n \geq 16$)
   3) Friedman test for k paired samples
IV) TESTS TO COMPARE 2 (OR MORE) INDEPENDENT POPULATIONS
   1) The Wald-Wolfowitz test
   2) The Mann-Whitney U test
     a) Small Sample Case
     b) Large Sample Case
   3) The Kruskal-Wallis H test for comparing k(k>2) independent random samples
V) CORRELATION
   1) Spearman's Rank Correlation Coefficient (for ordinal data
	
	t-test for paired data
	
	Randomized block design
	
	t-test to compare 2 independent samples, from 2 different populations
	
	One-Way ANOVA
	
	Pearson Product-Moment Correlation Coefficient

Note: Because we have previously discussed "TESTS FOR RANDOMNESS", "CHI-SQUARE TESTS" and "TESTS FOR MATCHED PAIRS" in other papers/presentations, the topic of this paper will be the Non-Parametric Methods available for "TESTS TO COMPARE TWO OR MORE INDEPENDENT POPULATIONS" and how they compare to the corresponding parametric methods.

Example 2: A parametric method for testing the equality of two means ($\mu_1 = \mu_2$),
Problem: The claim is made that the mean reported Annual Earnings of Carpenters and House painters is the same.

I)   Use the data of the table to test the validity of this claim, at α=0.05

CARPENTERS	PAINTERS
n_1=12	n_2=15
$\bar{x}_1$=16,000	$\bar{x}_2$=15,400
$\hat{S}_1^2$=565,000	$\hat{S}_2^2$=362,500
$\hat{S}_1$=751.6698189	$\hat{S}_2$=602.0797

II)   Also construct a 95% confidence Interval on Δμ and compare the two solutions.

Since n_1, n_2 < 30 & σ_1, σ_2 are not known, $\Delta\bar{x} = \bar{x}_1 - \bar{x}_2$ is $t_{n1+n2-2} = t_{25}$, if we can show that $\sigma_1^2 = \sigma_2^2$

Solution: First we need to test H_0: $\frac{\sigma_1^2}{\sigma_2^2}$=1 vs H_1: $\frac{\sigma_1^2}{\sigma_2^2}\neq1$ to see if $\sigma_1^2=\sigma_2^2$, because only then the problem H_0: Δμ=0 vs H_1: Δμ≠0 has a valid solution.

SOLUTION OF THE HYPOTHESIS TESTING PROBLEM

To test the hypothesis: H_0: Δμ=0 vs H_1: Δμ≠0 (when n_1 and n_2 are not simultaneously large), we must first test the Hypothesis.

$$H_0: \frac{\sigma_1^2}{\sigma_2^2}=1 \text{ vs } H_1: \frac{\sigma_1^2}{\sigma_2^2}\neq1$$

Then:
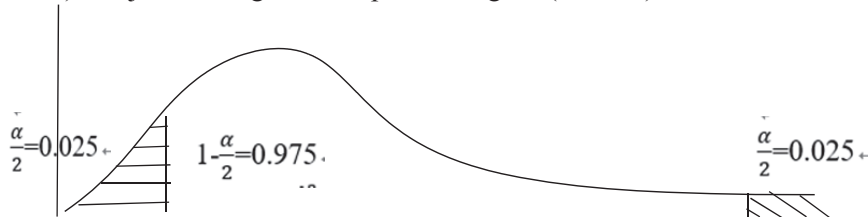1) If H_0: $\frac{\sigma_1^2}{\sigma_2^2}$=1 is accepted, the problem: H_0: Δμ=0 has a solution

2) If H_0: $\frac{\sigma_1^2}{\sigma_2^2}$=1 is rejected, the problem: H_0: Δμ=0 has no solution!
The test to check $\sigma_1^2=\sigma_2^2$ follows:
- H_0: $\frac{\sigma_1^2}{\sigma_2^2}$=1 vs H_1: $\frac{\sigma_1^2}{\sigma_2^2}\neq1$
- α=0.05 & $\frac{\alpha}{2}$=0.025
3) The estimator for $\frac{\sigma_1^2}{\sigma_2^2}$ is $\frac{\hat{S}_1^2}{\hat{S}_2^2}$ which is $F_{n_2-1}^{n_1-1} = F_{14}^{11}$ distributed.
4) Rejection Region/Acceptance Region (RR/AR)

$\frac{\alpha}{2}$=0.025 $1-\frac{\alpha}{2}$=0.975 $\frac{\alpha}{2}$=0.025

$F_{n_2-1}^{n_1-1}\left(1-\frac{\alpha}{2}\right) = F_{14}^{11}(0.975)$ $F_{14}^{11}(0.025) \approx F_{14}^{12}(0.025)$=3.05

$=\dfrac{1}{F_{11}^{14}{}_{(0.025)}} \approx \dfrac{1}{F_{11}^{15}{}_{(0.025)}}$

$= \dfrac{1}{3.33} = \dfrac{1}{10/3}$=0.3

5) $F^* = \frac{\hat{S}_1^2}{\hat{S}_2^2} = \frac{565000}{362500} = 1.559$

6) Do not reject H_0: $\frac{\sigma_1^2}{\sigma_2^2}$ b/c F^* falls outside RR

7) Therefore $\frac{\sigma_1^2}{\sigma_2^2} = 1$ & $\sigma_1^2 = \sigma_2^2$, and the Problem H_0: $\Delta\mu=0$ has a valid solution, as shown below:
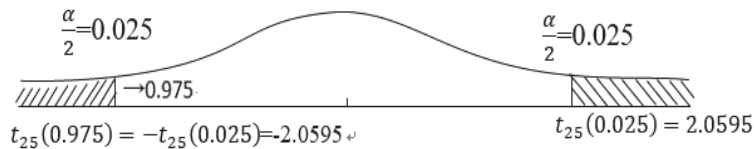
(1) H_0: $\Delta\mu=0$ VS H_1: $\Delta\mu\neq0$

(2) $\alpha=0.05$ (and $\alpha/2=0.025$)

(3) Estimator of $\Delta\mu$ is $\Delta\bar{x}$ ($\Delta\bar{x}=\bar{x}_1-\bar{x}_2=16000-15400=600$) which ,b/c:$n_1, n_2<30, \sigma_1, \sigma_2$ are unknown but $\sigma_1^2=\sigma_2^2$(shown above with an F-test) makes $\Delta\bar{x}/\sigma(\Delta\bar{x})$ $t_{n1+n2-2}$ = $t_{12+15-2}$ = $t_{25,}$ distributed.

with: $\sigma(\Delta\bar{x})=\sqrt{\left(\frac{12+15}{12\times15}\right)\left[\frac{(11)(565000)+(14)(362500)}{12+15-2}\right]}=260.26,$

and E $(\Delta\bar{x})= \Delta\mu=0$

(4) REJECTION REGION/ACCEPTANCE REGION(RR/AR)

$\frac{\alpha}{2}=0.025$ $\frac{\alpha}{2}=0.025$

$\rightarrow 0.975$

$t_{25}(0.975) = -t_{25}(0.025)=-2.0595$ $t_{25}(0.025) = 2.0595$

(5) Value of test statistic= $t^* = \frac{\Delta\bar{x}}{\sigma(\Delta x)} = \frac{600}{260.26} = 2.305$

(6) Since t^* is in the Rejection Region, we reject H_0: $\Delta\mu=0$

(7) Conclusion: The average (Reported) Income of Carpenters and House painters is not the same. ($\mu_1\neq\mu_2$)

III)   SOLUTION OF THE CONFIDENCE INTERVAL PROBLEM

Since: $\Delta\bar{x}= \bar{x}_1- \bar{x}_2=16000-15400=600,$

$t_{n1+n2-2}(\alpha/2) =t_{25}(0.025) =2.0595$, $\sigma(\Delta\bar{x}) =260.26$, and $1-\alpha=0.95$, the CI is obtained from:

P $[\Delta\bar{x}- t_{n1+n2} -2(\alpha/2)\sigma(\Delta\bar{x}))\leq\Delta\mu\leq\Delta\bar{x} + t_{n1+n2} -2(\alpha/2) \sigma(\Delta\bar{x}))] =1-\alpha$

When $\Delta\bar{x}- t_{25} (0.025), \sigma(\Delta\bar{x})$, and $1-\alpha=0.95$ are substituted in the equation above,

we obtain: P $[600-2.0595(260.26) \leq\Delta\mu\leq600+2.0595(260.26)] =0.95$

or: P$[63.86\leq\Delta\mu\leq1,136.14] =0.95$

Since the Hypothesized value, $\Delta\mu=\Delta\mu_0 =0$, is outside of this interval, H_0 is again REJECTED, and we conclude that the Average (Reported) income of carpenters and House Painters is not the same

Non-Parametric Tests to Compare 2 or More Independent Populations

In this section we discuss the following 3 nonparametric tests:
1) The Wald-Wolfowitz Test,
2) The Mann-Whitney U test, and
3) The Kruskal-Wallis Test

The first 2 tests are non-parametric equivalent to the t-test for testing the equality of the means of two independent populations (see example 2 above), while the third test is capable of testing the equality of 2 or more populations and, as such, is equivalent to the one-way ANOVA procedure.

The Wald- Wolfowitz Test

Assume that $x_1, x_2,...,x_{n1}$ and $y_1,y_2,...,y_{n2}$ are independent random samples from 2 populations. For this test the null hypothesis (H_0) is that the 2 independent random samples come from populations with identical distributions, or:

$$H_0: f_1(x)=f_2(x) \quad \text{vs.} \quad H_1:f_1(x)\neq f_2(x), \tag{2}$$

where $f_1(x)$ and $f_2(x)$ are the corresponding probability functions. The test begins by ranking the observations in the two samples, in a single array, where we represent an observation by X if it comes from sample 1 and by Y if it comes from sample 2. In such an array we then count the number of runs (i.e. the number of X and Y alternating sequences). In such an ordering of observations from 2 samples, the minimum number of runs is 2 and the maximum is n_1+n_2. If the 2 samples are drawn independently from 2 populations with identical distributions the observations, when ranked, will be well mixed, and the number of runs should be large. But if the samples are drawn from 2 populations with different distributions, the number of runs should be small. If we let R represent the number of runs in an ordered sequence of n_1+n_2 observations, R is a random variable with a rather-complicated. Probability function P(R), with possible values for R=2,3,4,...,(n_1+n_2),

$$E(R) = \frac{2n_1 n_2}{n_1 + n_2} + 1 \tag{3}$$

and:

$$\sigma(R) = \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}} \tag{4}$$

To test H_0 (given by equation (2)), for some α value, we need to find an integer R_0 such that:

$$\sum_{R=2}^{R_0} P(R) = \alpha, \text{ as closely as possible} \tag{5}$$

H_0 will be rejected if R=observed number of runs$\leq R_0$. Therefore, the rejection region is the lower one-sided region because H_0 should be rejected when the number of runs is small. When $n_1 \geq 10$ and $n_2 \geq 10$, R is approximately normally distributed with E(R) and σ (R) given by (3) and (4) respectively. Therefore, when the samples are large, the Wald-Wolfowitz test can be performed using the test statistic

$$z^* = \frac{R - E(R)}{\sigma(R)} \tag{6}$$

Note: For small samples we need to use the "complicated" P(R) function which is not given in this paper. [1,3]

Example 3

Suppose we have the 2 samples, X and Y, and: X= sample 1: 18, 25, 30, 35, 38, 40, 45, 52, 60, 68, 74, 80 ($n_1=12$) and Y= sample 2: 11, 16, 22, 23, 27, 39, 42, 44, 55, 65 ($n_2=10$) and wish to test to see if these observations come from populations with identical distributions. Combining these observations into a single ranked array of X and Y observations, we obtain:

YY X YY X Y XXX Y X YY XX Y X Y XXX

The observed number of runs is R=14 and, with n_1=12 and n_2=10, E(R)=11.9 and σ(R)=2.269. Then, because n_1=12≥10 and n_2=10≥10, R is approximately normally distributed, and:

$z^* = \dfrac{R - E(R)}{\sigma(R)} = \dfrac{14 - 11.9}{2.269} = 0.926$. If α=0.05, the rejection region is in the lower tail and $z_{critical}$= $-z_α$=-

$z_{0.05}$=-1.65; Since z^*> $z_{critical}$, H_0 is not rejected and we conclude that the X and Y observations come from identical distributions.

<u>Note</u>: We can also state the rejection of H_0 directly in terms of R, by stating:

Reject H_0 if: $z^* = \dfrac{R - E(R)}{\sigma(R)} < -z_α$,

from which we obtain: R≤E(R)-$z_α$ σ(R)   (7)

or: R≤8.16, since E(R)=11.9, and σ(R)=2.269, and -$z_α$= -1.65. Since the observed R=14>8.16, H_0 is not rejected.

The Mann-Whitney U Test

This test, even though it is in many aspects similar to the Wald-Wolfowitz test, also differs from it in the following 3 points:

1) A rank is assigned to the ordered sequence of the n_1+n_2 observations coming from the X and Y observations.
2) The alternative hypothesis (H_1) may be either 2-sided or 1-sided, and
3) The alternative hypothesis implies only a difference in the central tendency of one distribution relative to the other, and does not suggest a difference in Dispersion, or Skewness, or Kurtosis, as is the case with the Wald-Wolfowitz test.

<u>Note</u>: Therefore, the Mann-Whitney test should be used if we wish to compare only the central tendencies of 2 distributions, and the Wald-Wolfowitz test if broader comparisons are to be made. The 2-tailed hypotheses being tested here are: H_0: The 2 populations are identical   vs. H_1: The 2 populations are not identical. To implement this test, we combine the 2 sets of observations to form a single ranked set, consisting of n_1+n_2 observations. Then a rank is assigned to each observation, starting with 1 and ending with n_1+n_2. If ties occur in the ordered sequence, the average of the ranks is assigned to the tied observations. For example, if the values for the 4th, 5th, and 6th observations are the same, then

each of the observations is assigned the rank=$\dfrac{4+5+6}{3} = 5$.

The sum of the ranks of values from populations 1 and 2 are designated by R_1 and R_2 respectively, and this test determines if an observed arrangement of ranks can lead to the conclusion that the 2 sets of observations (X,Y) come from populations which have the same central tendency (i.e. do not reject H_0) or different central tendency (i.e. Reject H_0).This test is implemented differently for small and large samples: If n_1≤10 and n_2≤10, the samples are considered small while, if both n_1>10 and n_2>10 the samples are considered large and the variable U (i.e. the Mann-Whitney test statistic) is approximately normally distributed, with:

$E(U) = (n_1 n_2)/2$,   (8)

and $\sigma(U) = \sqrt{\dfrac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$,   (9)

and the normal distribution can be used to test the validity of H_0 (the variable U is defined below). The next step in the procedure, after R_1 and R_2 have been obtained, is to calculate:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \quad , \tag{10}$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 \quad , \tag{11}$$

and $\quad U = \text{minimum}\,(U_1, U_2)$ \hfill (12)

a) Small Sample Case
The exact distribution of U has been derived and is tabulated, but the table is very extensive and is not given here. This table contains p-values for U. To determine the p-value from this table for a calculated U_{value}, we let: n_1= minimum (sample1, sample 2), n_2= maximum (sample1, sample 2), locate the U value under the U_0 column, and at the intersection of $U = U_0$ and n_1, for the specified n_2 value, read the p-value for a one tailed test. For a 2-tailed test the value shown in the table should be doubled. [2,3,5,8]

Example 4
Suppose a large corporation is conducting a study to determine if male and female employees (of similar backgrounds) receive equivalent compensation, and 8 female and 9 male employees were selected. Suppose from the combined set of ranked observations we then obtained: R_1=57, R_2=96, U_1=51, U_2=21, and U=min (U_1, U_2) =21 Then from the table, with n_1=8 and U=21 and n_2=9, we obtain the p-value of 0.0836 which, for a 2-tailed test becomes: 2(0.0836) =0.1672. Because p=0.1672>α=0.05, H_0 is not rejected, and we conclude that the two random samples come from populations with identical central tendencies.

b) Large Sample Case
When n_1>10 and n_2>10, the statistic U is approximately normally distributed with E(U) and σ(U), given by equations (8) and (9) respectively. Therefore, when the 2 samples are large,

the Mann-Whitney test can be performed using the z^* test $\left(i.e: \quad z^* = \dfrac{U - E(U)}{\sigma(U)} \right)$ and, for

a 2-tailed test, H_0 will be rejected if: $z^* > z_{\alpha/2}$, or $z^* < -z_{\alpha/2}$, for a given α value.

Example 5
A two-year study was conducted to determine whether there is a difference in the number of colds experienced by smokers and nonsmokers. The random sample selected consisted of 14 nonsmokers and 12 smokers and the recorded data represent the number of colds observed during the two-year study period:

| Nonsmokers | 1 | 0 | 2 | 7 | 3 | 1 | 2 | 2 | 4 | 3 | 5 | 0 | 2 | 1 |
| Smokers | 4 | 2 | 6 | 5 | 8 | 10 | 8 | 7 | 6 | 4 | 9 | 3 | | |

Use the Mann-Whitney U statistic to determine whether there is reason to believe that these random samples come from populations with different distributions, at α = 0.05. (Hint: Use the large-sample method because $n_1 > 10$ and $n_2 > 10$. Form a single array of the observations and find:

$R_1 = 4 + 1.5 + 8 + 21.5 + 12 + 4 + 8 + 8 + 15 + 12 + 17.5 + 1.5 + 8 + 4 = 125.$

$R_2 = 15 + 8 + 19.5 + 17.5 + 23.5 + 26 + 23.5 + 21.5 + 19.5 + 15 + 25 + 12 = 226$

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 = (14)(12) + \frac{14(15)}{2} - 125 = 168 + 105 - 125 = 273 - 125 = 148$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 = (14)(12) + \frac{12(13)}{2} - 226 = 168 + 78 - 226 = 246 - 226 = 20$$

and $U = \min(U_1, U_2) = \min(148, 20) = 20$; $E(U) = \frac{n_1 n_2}{2} = \frac{(14)(12)}{2} = 84$ and

$$\sigma(U) = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{(14)(12)(27)}{12}} = \sqrt{(14)(27)} = \sqrt{378} = 19.44, \text{ and}$$

$Z^* = \dfrac{U - E(U)}{\sigma(U)} = \dfrac{20 - 84}{19.44} = \dfrac{-64}{19.44} = -3.29$. Since for $\alpha = 0.05$, the rejection regions are (for a 2-tailed

test): $Z_{cr1} = -1.96$ and $Z_{cr2} = 1.96$, H_0 is rejected because Z^* falls in the rejection region, and we conclude that the random samples came from different populations. The MINITAB output for this example is shown below:

MTB > set c1
DATA> 1 0 2 7 3 1 2 2 4 3 5 0 2 1
DATA> end
MTB > set c2
DATA> 4 2 6 5 8 10 8 7 6 4 9 3
DATA> end
MTB > Mann-Whitney 95.0 c1 c2;
SUBC> Alternative 0.

Mann-Whitney Test and CI: C1, C2
C1   N = 14   Median =   2.000
C2   N = 12   Median =   6.000
Point estimate for ETA1-ETA2 is   -4.000
95.2 Percent CI for ETA1-ETA2 is (-6.000,-2.001)
W = 125.0
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.0011
The test is significant at 0.0010 (adjusted for ties))
The two solutions (hand and MINITAB) are the same ($R_1 = W = 125$ for both and, because

$p = 0.0011 < \alpha = 0.05$, H_0 is rejected.)

The Kruskal-Wallis H-test for Comparing k (k≥2) Independent Samples
    The Kruskal-Wallis test is the nonparametric equivalent to the one-way analysis of variance (ANOVA) where we are interested in testing the null hypothesis

H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$
vs   (13)
H_1: The means are not all equal, based on k mutually independent random samples, drawn from k populations which are assumed to be normally distributed. The Kruskal-Wallis test can be used to analyze k ordinal data sets and does not depend on any assumption about the shape of the populations from which the data was drawn from. But it does depend on the assumption that the k groups (or data sets) are independent and that the individual items of each group are obtained randomly.

The Kruskal-Wallis procedure tests the following H_0 / H_1 hypotheses:

H_0: the k populations are identical

vs   (14)

H_1: At least one of the k populations is different

However, the Kruskal-Wallis test is particularly sensitive to differences in central tendency and, as such, it is regarded as an extension of the Mann-Whitney U test, which compares the "identicality" of 2 populations while the Kruskal-Wallis H test is capable of comparing the "identicality" of k populations, where $k > 2$. We let n_1 be the sample size from population 1, n_2 the sample size of population 2, …, n_i the sample size of population i, and n_k the sample size of population k, with:

$$N = n_1 + n_2 + \dots n_i + \dots + n_k = \sum_{i=1}^{k} n_i \; ,$$ (15)

where N is the total number of all observations from the k populations. The procedure for computing a Kruskal-Wallis H statistic begins by combining all observations in the k samples to form a single set of N observations, which are then ranked in order of increasing magnitude. The smallest value is given the rank of 1, and the highest the rank of N. When ties occur in ranking, we assign the average of the adjoining ranks to each of the tied observations. For example if 2 observations are tied and their ranks should be 9 and 10, we assign to both of these observations the rank of $(9 + 10)/2 = 9.5$. Similarly, if 3 observations are tied and their ranks should be 12, 13, and 14, we assign to all 3 observations the rank of $\dfrac{12 + 13 + 14}{3} = 13$.

Under the assumption that the k samples come from populations with identical distributions, the Kruskal Wallis H statistic is defined by:

$$H = \frac{12}{N(N+1)} \left[\sum_{i=1}^{k} \frac{T_i^2}{n_i} \right] - 3(N+1)$$ (16)

where: k = number of populations being compared
         N = total number of observations from the k populations
         T_i = total (sum) of ranks in population (group) i
         n_i = number of observations in population i
         H = Kruskal-Wallis Test statistic

The H, defined by (16), is approximated by a chi-square distribution with DOF = Degrees of Freedom = k – 1, as long as $n_i \geq 5$ for all populations. In testing the validity of (14), for a specified value of α, the rejection region is the upper tail of the chi-square distribution and H_0 will be rejected if H > . $\chi_{k-1}^2(\alpha)$ If the number of tied observations is large (in the combined single set of N observations), an adjusted value of H, defined by (16) should be calculated and used. If we let H_{adj} represent the adjusted H value, we define H_{adj} by:

$$H_{adj} = \frac{H}{M}$$ (17)

where $M = 1 - \dfrac{\sum_{j=1}^{c} t_j^3 - t_j}{N^3 - N}$, (18)

and: c = number of tied sets
       t_j = number of tied observations in set j
       N = total number of observations from the k populations
       M = correction factor

As can be seen from (18) and (17) the effect of the correction factor M is to always increase the value of H defined in (16) because M is always less than 1. But this effect is usually small, even if there is a large number of ties, as the following situation illustrates:

Suppose in a given problem there are 4 sets of 2 tied observations, 2 sets of 3 tied observation, and one set of 4 tied observations, and N = 40.

Then, from (18) $M = 1 - \left[\dfrac{4(2^3 - 2) + 2(3^3 - 3) + 1(4^3 - 4)}{40^3 - 40} \right]$

$$= 1 - \frac{132}{63{,}960} = 1 - 0.002 = 0.998 \ ,$$

and $H_{adj} = \dfrac{H}{M} = \dfrac{H}{0.998} \approx 1.002H$

The H test compares well with the F test in the Analysis of Variance. However, if the samples are large, the ranking of the samples is a tedious job, unless a computer is used to sort the data. [1,2,3,5,9]

Example 6

To test the claim that there is no significant difference in the amount of customer's initial deposits when they open savings accounts according to geographic region in the USA, an analyst selects savings and loan offices of equal size from 4 regions of the United States, and located in areas having similar economic and population characteristic. The following data represent the dollar amounts of adult customers selected randomly:

Region 1	Region 2	Region 3	Region 4
$ 1200 (23)	$ 225 (5)	$ 675 (15)	$ 1075 (21)
450 (12)	950 (19)	500 (13)	1050 (20)
110 (2)	100 (1)	1100 (22)	750 (17)
800 (18)	350 (9)	310 (7)	180 (3)
375 (10)	275 (6)	660 (14)	330 (8)
200 (4)			680 (16)
			425 (11)
$n_1 = 6$	$n_2 = 5$	$n_3 = 5$	$n_4 = 7$
T_1	$T_2 = 5+19+1+9+6=40$	$T_3 = 15+13+22+7+$	T_4
$=23+12+2+18+10+4=69$		$14=71$	$=21+20+17+3+8+16+11=96$

Use the Kruskal-Wallis H test to test the claim in H_0 that there is no significant difference in the initial amount of deposit because of geographic region, at α = 0.05.

<u>Solution</u>: Form a single array of ordered values, assign ranks to the ordered values and then find $T_j =$ sum of ranks in each region j, n_j, and $N = \sum_{i=1}^{k} n_j$. The ranks are shown in the table above in parentheses. Also shown are n_j and T_j. Then calculate:

$$H = \frac{12}{23(24)} \left[\frac{69^2}{6} + \frac{40^2}{5} + \frac{71^2}{5} + \frac{96^2}{7} \right] - 3 \, (24) = 2.745 \text{ and compare to } \chi^2_{k-1(\alpha)} = \chi_3^2 (\alpha = 0.05) = 7.815;$$

Since $H < \chi_3^2$ ($\alpha = 0.05$) = 7.815, do not reject H_0, and we conclude that there is no difference in the initial amount of deposit because of geographic region. The MINITAB output for this example is shown below:

MTB > set c1
DATA> 1200 450 110 800 375 200 225 950 100 350 275 675 500 1100　310 660 1075　1050　750 180 330 680 425
DATA> end
MTB > set c2
DATA> 1　1　1　1　1　1　2　2　2　2　2　3　3　3　3　3　4　4　4　4　4　4　4
DATA> end
MTB > Kruskal-Wallis c1 c2.

Kruskal-Wallis Test: C1 versus C2
Kruskal-Wallis Test on C1

C2	N	Median	Ave Rank	Z
1	6	412.5	11.5	-0.21
2	5	275.0	8.0	-1.49
3	5	660.0	14.2	0.82
4	7	680.0	13.7	0.80
Overall	23		12.0	

H = 2.75 DF = 3 P = 0.433)
The two solutions are the same (H = 2.75 in both and, because p = 0.433 > α = 0.05, H_0 is not rejected.)

V) Spearman Rank Correlation Coefficient
　　The rank correlation coefficient, r_s, was developed by C. Spearman in 1904, and is used to determine the amount of association between two variables X and Y when their values are ordinal rank numbers, with 1 for first, 2 for second, ..., and n for the last.
If we let X_i = rank of individual i on variable X

Y_i = rank of individual i on variable Y

and　　$d_i = X_i - Y_i$　　　　　　　　　　　　　　　　　　　　　　　　　(19)

then r_s is defined as:

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \tag{20}$$

where $-1 \le r_s \le 1$

The sampling distribution of r_s is t_{n-2} and is used to test hypotheses about ρ_s, the population parameter for which r_s can be considered as its estimator.

The interpretation of r_s values is similar to the interpretation of r values. If r_s is near +1, high positive correlation exists between the variables and indicates that high values of one variable are associated with high values of the other variable. If r_s is near -1, high negative correlation exists between the variables and indicates that high values of one variable are associated with low values of the other variable. If r_s is near 0 the 2 variables have little association between them. The process of calculating r_s begins by assigning ranks to the X and Y values, and then using equation (20) above, which is derived from the definitional equation of r, assuming that there are no ties in the ranks.

Example 7

Compute Spearman's rank correlation coefficient for the following variables to determine the degree of association between the 2 variables, and then test $H_0: \rho_s = 0$ vs $H_1: \rho_s \neq 0$,

at $\alpha = 0.05$

X values	23 (3)	41 (10.5)	37 (8)	29 (6)	25 (4)	17 (1)	33 (7)	41 (10.5)	40 (9)	28 (5)	19 (2)
Y values	201 (2)	259 (11)	234 (7)	240 (8)	231 (6)	209 (3)	229 (5)	246 (9)	248 (10)	227 (4)	200 (1)

First rank the X and Y values separately (the ranks are shown in parentheses) and then take the difference of the corresponding ranks, square this difference and add them to obtain:

$$\sum_{i=1}^{n} d_i^2 = 1^2 + (-0.5)^2 + 1^2 + (-2)^2 + (-2)^2 + (-2)^2 + 2^2 + (1.5)^2 + (-1)^2 + 1^2 + 1^2$$

$$= 1 + 0.25 + 1 + 4 + 4 + 4 + 4 + 2.25 + 1 + 1 + 1 = 23.50.$$

Then, $r_s = 1 - \dfrac{6\sum d_i^2}{n(n^2 - 1)} = 1 - \dfrac{6(23.50)}{11(121 - 1)} = 1 - \dfrac{141}{1320} = 1 - 0.1068 = 0.893.$ If we calculate

$$\bar{r}_s = 1 - \frac{\sum d_i^2}{\frac{1}{6}\left[n(n^2 - 1)\right] + 1} = 1 - \frac{23.5}{\frac{1}{6}(1320) + 1} = 1 - \frac{23.5}{220 + 1} = 1 - 0.1063 = 0.8937;$$

Then: $\dfrac{\bar{r}_s \sqrt{n - 2}}{\sqrt{1 - \bar{r}_s^2}}$ is $t_{n-2} = t_9$. To complete the testing of the hypotheses: $H_0: \rho_s = 0$ vs

$H_1: \rho_s \neq 0$, we let t* $= \dfrac{\bar{r}_s \sqrt{n - 2}}{\sqrt{1 - \bar{r}_s^2}} = \dfrac{0.8937(3)}{\sqrt{1 - (0.8937)^2}} = \dfrac{2.6811}{0.448665} = 5.9757$ and compare to $t_9(0.025) =$

2.262.

Since t* $> t_9(0.025)$ we reject $H_0: \rho_s = 0$ and conclude that the degree of association between the 2 variables X and Y is very strong. [1,2,3,5,8,9,12]

CONCLUSIONS

We have discussed many nonparametric tests in which the parameters of the distribution continue to be important but the nature of the distribution, from which the sample data we used in our analysis came from, is not important and is not needed to perform these tests. This is in contrast to the parametric tests which depend very much on the nature of the population from which out data set came from.

- The nonparametric tests that we have discussed fall into the following 5 categories
    - I. Tests of Randomness
    - II. Chi-Square Tests
    - III. Tests for Matched Pairs
    - IV. Tests to Compare 2 of more Independent Populations
    - V. Spearman Rank Correlation Test

But in this paper, we have discussed only the technique of parts IV &V.

- Some of these non-parametric tests have corresponding parametric tests; others do not!
- Nonparametric methods can solve the same type of problems that parametric methods can solve (but with reduced efficiency) and can solve additional problems when no parametric methods are available.
- Many of the nonparametric tests we discussed depend on the number of runs in a sequence of observations.
- The use of a Statistical Software tool, like MINITAB, simplifies the application of these tests considerably. Unfortunately, not every nonparametric test is supported by MINITAB yet.
- We have shown in several examples, where MINITAB was applicable, that the hand and
- MINITAB solutions are identical. Therefore, MINITAB should be used in nonparametric applications, where applicable.

MINITAB uses the p-value to decide whether to reject or not reject a hypothesis.

REFERENCES

Berenson, M.L., Levine, D.M., & Krehbiel, T.C. (2004). *Basic Business Statistics*. Pearson-Prentice Hall, 9th Edition. ISBN 0-13-047784-2, p/424-432.

Black, K. (1984). *Business Statistics*, 4th Edition. Wiley. ISBN 0-471-42983-X, p/656-703.

Canavos, G.C. (1984). Applied Probability and Statistical Methods. Little, Brown. ISBN 0-316-12778-7, p/530-546.

Carlson, W.L., & Thorne, B. (1997). Applied Statistical Methods. Prentice Hall. ISBN 0-13-570847-8, p/566-584.

Chou, Y-L. (1992). *Statistical Analysis for Business and Economics*. Elsevier. ISBN 0-444-07301-6, p/1047-1065.

Freund, J.E., & Williams, F.J. (1969). Modern Business Statistics. Revised by: Perles, Benjamin, and Sullivan, Charles; Prentice-Hall. ISBN 13-589580-4, p/257-260.

Freund, J.E., & Williams, F.J. (1982). *Elementary Business Statistics: The Modern Approach*. Prentice-Hall. ISBN 0-13-253120-8, p/409-430.

Johnson, R. (1973). *Elementary Statistics*. Duxbury Press. ISBN 0-87872-043-X, p/377-391.

McClave, J.T., Benson, G.P., & Sincich, T. (2001). *Statistics for Business and Economics*, 8th Edition. Prentice Hall. ISBN 0-13-027293-0, p/887-924.

Salvatore, D. (1982). Theory and Problems of Statistics and Econometrics. SCHAUM'S OUTLINE SEERIES: McGraw-Hill. ISBN 0-07-054505-7, p /120, 134-135.

Steel, R.G.D., & Torrie, J.H. (1976). *Introduction to Statistics*. McGraw-Hill. ISBN 0-07-060918-7, p/182-190.

Vasilopoulos, A. (2007). *Business Statistics- A Logical Approach Theory, Models, Procedures and Applications, Including Computer (MINITAB) Solutions*. Pearson Custom Publishing, Boston. ISBN 0-536-29714-2, p/754-820.