**Notes – Ch19 Non-Parametric Tests**

Nonparametric methods which can be used to make inferences about a population without requiring an assumption about the specific form of the population’s probability distribution. For this reason, these nonparametric methods are also called distribution-free methods.

Most of the statistical methods referred to as parametric methods require quantitative data, while nonparametric methods allow inferences based on either categorical or quantitative data. However, the computations used in the nonparametric methods are generally done with categorical data. Whenever the data are quantitative, we will transform the data into categorical data in order to conduct the nonparametric test.

The following conditions must be satisfied for a test to classified as non-parametric:

* The method can be used with nominal data
* The method can be used with ordinal data
* The method can be used with interval or ration data when no assumptions can be made about the population probability distribution

1. **Sign Test:** It is a test for identifying differences between two populations based on the analysis of nominal data.
2. **Sign Test for Small Sample**

We assign a + sign whenever the data is success or the preference selected and a - sign whenever the data is a failure or the preference is not selected. Any data not showing a preference should be discarded from the sample size. We determine the sample size n = Total no of + and – signs. We use the binomial distribution with p = 0.5 as the sampling distribution whenever n , where p is the proportion.

**Step 1: Setting up the hypothesis**

**Two tailed test:** A two-tailed test of a hypothesis will reject the null hypothesis if the sample proportion is significantly higher than or lower than 0.5

H0 : p = 0.5

Ha : p 0.5

**Left-tailed test**: A left-tailed test of a hypothesis will reject the null hypothesis in favour of alternative hypothesis if the sample proportion is significantly lower than 0.5

H0 : p 0.5

Ha : p < 0.5

**Right-tailed test:** A right-tailed test of a hypothesis will reject the null hypothesis in favour of alternative hypothesis if the sample proportion is significantly higher than 0.5

H0 : p 0.5

Ha : p> 0.5

**Step 2: Set the criteria for a decision.**

Identify the value of α

**Step 3: Compute the test statistic.**

The test statistics is a binomial variable x (indicating no of + signs), p = 0.5 and

n = Total no of + and – signs (ignoring values with no preferences)

The rejection region will be in either one or both tails, depending on the null hypothesis being tested. Determine the p-value.

**Step 4: Make a decision.**

If p-value , then reject H0.

1. **Sign Test for Large Sample**

We assign a + sign whenever the data is success or the preference selected and a - sign whenever the data is a failure or the preference is not selected. Any data not showing a preference should be discarded from the sample size. We determine the sample size n = Total no of + and – signs. We use the normal approximation whenever n >

**Step 1: Setting up the hypothesis**

**Two tailed test:** A two-tailed test of a hypothesis will reject the null hypothesis if the sample proportion is significantly higher than or lower than 0.5

H0 : p = 0.5

Ha : p 0.5

**Left-tailed test**: A left-tailed test of a hypothesis will reject the null hypothesis in favour of alternative hypothesis if the sample proportion is significantly lower than 0.5

H0 : p 0.5

Ha : p < 0.5

**Right-tailed test:** A right-tailed test of a hypothesis will reject the null hypothesis in favour of alternative hypothesis if the sample proportion is significantly higher than 0.5

H0 : p 0.5

Ha : p> 0.5

**Step 2: Set the criteria for a decision.**

Identify the value of α

**Step 3: Compute the test statistic.**

The test statistics is no of + signs

Standard deviation:

The rejection region will be in either one or both tails, depending on the null hypothesis being tested. Determine the p-value.

**Step 4: Make a decision.**

If p-value , then reject H0.

1. **Sign Test for hypothesis test about a Median**

Sign test can be used to conduct hypothesis tests about a population median. Median splits the population in such a way that 50% of the values are at median or above and 50% of the values are at median or below. We assign a + sign whenever the data in the sample is above the hypothesized value of the median and a - sign whenever the data in the sample is below the hypothesized value of the median. Any data exactly equal to the median should be discarded from the sample size. We determine the sample size n = Total no of + and - signs

**Step 1: Setting up the hypothesis**

**Two tailed test:** A two-tailed test of a hypothesis will reject the null hypothesis if the sample median is significantly higher than or lower than the hypothesized population median.

H0 : Median = hypothesized median

Ha : Median hypothesized median

**Left-tailed test**: A left-tailed test of a hypothesis will reject the null hypothesis if the sample median is significantly lower than the hypothesized population median.

H0 : Median hypothesized median

Ha : Median < hypothesized median

**Right-tailed test:** A right-tailed test of a hypothesis will reject the null hypothesis if the sample median is significantly higher than the hypothesized population median.

H0 : Median hypothesized median

Ha : Median > hypothesized median

**Step 2: Set the criteria for a decision.**

Identify the value of α

**Step 3: Compute the test statistic.**

The test statistics is no of + signs

Standard deviation:

The rejection region will be in either one or both tails, depending on the null hypothesis being tested. Determine the p-value.

**Step 4: Make a decision.**

If p-value , then reject H0.

**2. The Wilcoxon Signed Rank Test:** In the matched-sample situation, each experimental unit generated two paired or matched observations, one from population A and one from population B. The difference between the matched observations provide insight about the difference between the two populations.

The Wilcoxon Signed Rank Test requires the ranking of the absolute values of the differences between the two populations. We discard any differences of zero and then rank the remaining absolute differences from the lowest to the highest. Tied differences are assigned the average ranking of their positions in the combined data set. Once the ranks of the absolute differences have been determined, the ranks are given the sign of the original difference in the data.

If the two populations are identical, then the positive and the negatives ranks will cancel each other, so the sum of the signed ranked values would be approximately zero. Thus the test of significance involves determining whether the computed sum of signed ranks is significantly different from zero.

**Step 1: Setting up the hypothesis**

H0 : The populations are identical

Ha : The populations are not identical

**Step 2: Set the criteria for a decision.**

Identify the value of α

**Step 3: Compute the test statistic.**

Let T denote the sum of the signed-rank values. If the two populations are identical and the number of matched pairs of data is 10 or more, the sampling distribution of T can be approximated by normal distribution.

Mean: μT = 0

Standard deviation: provided n

The rejection region will be in both tails. Determine the p-value.

**Step 4: Make a decision.**

If p-value , then reject H0.

**3. Mann-Whitney-Wilcoxon(MWW) Test:** This test can be used to determine whether a difference exists between two populations. This test, unlike the signed-rank test, is not based on a matched sample. Two independent samples, one from each population, are used. The test was developed jointly by Mann, Whitney, and Wilcoxon. It is sometimes called Mann-Whitney test and sometimes the Wilcoxon rank-sum test. The only requirement of this test is that the measurement scale of the data is at least ordinal.

**A. Sign Test for Small Sample:** It should be used whenever the sample sizes for both populations are less than or equal to 10.

**Step 1: Setting up the hypothesis**

H0 : The populations are identical

Ha : The populations are not identical

**Step 2: Set the criteria for a decision.**

Critical value TL(lower limit of T) is provided in Table: TL values for Mann-Whitney-Wilcoxon(MWW) Test. In this table n1 refers to the sample size corresponding to the sample whose rank sum(T) is being used in the test. n2 refers to the sample size of the sample whose rank sum is not being used in the test. TU(upper limit of T) is computed using the equation

TU = n1(n1 + n2 + 1) – TL

**Step 3: Compute the test statistic.**

Rank the combined data from the two samples from low to high. If one or more data values are the same, the tied values are given the average ranking of their position in the combined data set. Sum the ranks for each sample separately. The MWW procedure can use the sum of ranks for either sample. Denote this sum by the symbol T.

**Step 4: Make a decision.**

The null hypothesis should be rejected only if T is strictly less than TL or strictly greater than TU. Reject H0 if T < TL or T > TU

**B. Sign Test for Large Sample:** Normal approximation of T distribution can be used whenever the sample sizes for both populations are greater than or equal to 10.

**Step 1: Setting up the hypothesis**

H0 : The populations are identical

Ha : The populations are not identical

**Step 2: Set the criteria for a decision.**

Identify the value of α

**Step 3: Compute the test statistic.**

Rank the combined data from the two samples from low to high. If one or more data values are the same, the tied values are given the average ranking of their position in the combined data set.

Sum the ranks for each sample separately. The MWW procedure can use the sum of ranks for either sample. Denote this sum by the symbol T. Here, n1 refers to the sample size corresponding to the sample whose rank sum(T) is being used in the test. n2 refers to the sample size of the sample whose rank sum is not being used in the test. We can use normal approximation to the sampling distribution of the rank sum T.

Mean: μT =( ½ )n1(n1+n2+1)

Standard deviation: provided n1  and n2

The rejection region will be in both tails. Determine the p-value.

**Step 4: Make a decision.**

If p-value , then reject H0.

**4. Kruskal-Wallis:** The test can be used to test if 3 or more populations are identical (k 3 where k is no of populations. It can be used with ordinal data as well as with interval or ratio data. The test does not require the assumption that the k populations are normally distributed.

**Step 1: Setting up the hypothesis**

H0 : All populations are identical

Ha : Not all populations are identical

**Step 2: Set the criteria for a decision.**

Identify the value of α

**Step 3: Compute the test statistic.**

Rank the combined data from the two samples from low to high. If one or more data values are the same, the tied values are given the average ranking of their position in the combined data set. Sum the ranks for each sample separately. Compute the Kruskal-Wallis test statistics W given by

Where

k = the number of population

ni = the number of items in the sample i

nT = = total number of items in all samples

Ri = the sum of ranks for sample i

Kruskal and Wallis were able to show that under the null hypothesis in which the populations are identical, the sampling distribution of W can be approximated by Chi-square distribution with k – 1 degrees of freedom. This approximation is generally acceptable if each of the sample sizes is greater than or equal to five. The procedure used an upper tail test.

**Step 4: Make a decision.**

If p-value , then reject H0.

**5. Rank Correlation Test:** The Spearman Rank correlation coefficient is based on rank-ordered data for two variables. The Spearman Rank correlation coefficient ranges from -1 to +1 and its interpretation is similar to the Karl Pearson correlation coefficient in that positive values near 1 indicate a strong positive association between ranking; as one rank increases, the other rank also increases. Rank correlation values near -1 indicate a strong association negative between ranking; as one rank increases, the other rank also decreases.

**Step 1: Setting up the hypothesis**

We want to make an inference about the population rank correlation (. The null hypothesis (indicates that there is no rank correlation between the populations and the rankings are independent.

H0 :

Ha :

**Step 2: Set the criteria for a decision.**

Identify the value of α

**Step 3: Compute the test statistic.**

Rank the data from the two samples separately from low to high. If one or more data values are the same, the tied values are given the average ranking of their position in the combined data set. Compute the difference in the ranks and the square of the differences.

Sum the square of the differences. The Spearman rank correlation coefficient rs given be

The distribution can be approximated using normal if n

Standard deviation:

The rejection region will be in either one or both tails, depending on the null hypothesis being tested. Determine the p-value.

**Step 4: Make a decision.**

If p-value , then reject H0.