# TOPIC 01: WRITTEN ASSIGNMENT

## Non-parametric Tests


Submitted by:

**Atheek Mohamed Rafi**

**Matriculation: 9212873**

Enrolled in:

**B.Sc. Applied Artificial Intelligence**

**Inferential Statistics**


Written assignment submitted to the

IU International University

In fulfilment of the requirements for the course of

Inferential Statistics | DLBDSSIS01

Instructor Name: Hashem Zarafat

Date: 28/09/2023

## Table of Contents

## 1. Introduction

Statistical analysis and research heavily rely on the use of statistical testing which plays a major role in evaluating hypotheses and making informed decisions (Conn et al., 2014 p. 291). The purpose of this paper is to delve into the framework of testing and its interpretation.

The reason for selecting this topic is its relevance in research. Statistical tests form the foundation of inquiry, by validating hypotheses and providing evidence to guide decision making. This paper aims to clarify the core principles, methodologies and interpretation of testing, equipping readers with skills for effective data analysis.

To maintain focus we will primarily concentrate on principles of testing while introducing two non-parametric tests: the Mann Whitney U Test and Kruskal-Wallis Test. These tests are particularly useful when parametric assumptions do not hold (Field, 2013, p. 214). Furthermore, this paper will conduct a non-parametric test to identify if there is any significant difference in number of cancer cases before and after Chernobyl incident.

Throughout the following sections we will systematically explore testing from basics to advance levels. Emphasis will be placed on result interpretation as it is essential in transforming data into insights.

## 2. General Framework for Statistical Testing

Statistical testing is a systematic approach used to assess hypotheses, measure the significance of observed effects, and make data driven decisions (Hung et al., 1997, p.11). It plays a role, in research enabling us to evaluate the validity of assumptions and draw conclusions from data. This term paper explores the framework of statistical testing, delves into the significance of the p value, provides comprehensive interpretation of results, and considers including detailed formulas.

### 2.1. Steps Involved in Statistical Testing

Statistical testing follows a defined set of steps that are essential for its application. These steps include:

1. Formulating Hypotheses: Every statistical test starts with establishing both null ($H_0$) and alternative ($H_1$) hypotheses. The null hypothesis represents the default assumption while the alternative hypothesis suggests a claim or effect (Erdogmus, 2022, p. 2).
2. Collecting Data: Data collection is a stage that involves gathering information for hypothesis testing. The data should be collected systematically and accurately to ensure the reliability of the test.

3. Choosing a Test: The selection of a test depends on factors such as data nature and research question. Selecting the test whether it be a t-test, chi-square test or non-parametric test is extremely important.
4. Setting Significance Level (α): Researchers must determine the significance level (α) which acts as a threshold, for considering results as statistically significant. Typically values like 0.05 or 0.01 are commonly used.
5. Calculating the Test Statistic: To quantify the observed effect or difference between groups researchers calculate the test statistic using values obtained from the data (Erdogmus, 2022, p. 5).
6. Obtaining the p-value: The p value is crucial as it measures the probability of obtaining results as extreme as, or more extreme than, the observed data under the assumption that the null hypothesis true. It plays a major role, in interpreting results (Conn et al., 2014, p. 291).
7. Making a Decision: Based on both the p value and chosen significance level (α), researchers make decisions to either reject or fail to reject the null hypothesis. These decisions guide result interpretation.

## 2.2. Significance of the p-value

“The P-value is a random variable derived from the distribution of the test statistic used to analyze a data set and to test a null hypothesis” (Hung et al., 1997, p.11). Formally,

$$p - value \ = \ \mathbb{P}(\mathrm{U} \ > \ u_{obs} |\ \mu \ = \ \mu_0)$$

where U is test statistic and $u_{obs}$ is observed value

The significance of the p value cannot be emphasized enough; it holds significance importance in statistical testing. The p value acts, as a quantifiable measure of evidence against the null hypothesis. If the p value is low (less than α), it suggests strong evidence against the null hypothesis indicating that the observed results are unlikely to happen by random chance alone (Hung et al., 1997, p.11). On the other hand, a high p value implies weak evidence against the null hypothesis suggesting that the observed results could reasonably be due to random variability.

The p value is a tool for researchers to determine whether to accept or reject the hypothesis. It plays a major role in drawing conclusions from data and evaluating the significance of observed effects. A smaller p value indicates stronger evidence against the null hypothesis supporting the validity of alternative hypotheses and supporting researcher’s claims or hypotheses.

### 2.3. Interpreting Results

Interpreting statistical test results is a part of the process that involves translating the findings into meaningful insights and actionable conclusions. When interpreting results consider:

a. Decision on $H_0$: Based on the significance level and p value decide whether to "reject" or "fail to reject" the null hypothesis. This decision plays a major role in result interpretation.

b. Practical significance: While statistical significance holds importance it is equally essential to consider the practical significance.

c. Effect Size: Understanding the effect size—how substantial the observed effect is—provides valuable information. It quantifies the difference or relationship enabling researchers to assess the relevance of their findings.

d. Confidence Intervals: Incorporating confidence intervals in result interpretation enhances understanding. These intervals offer a range of values within which the true population parameter's likely to fall, providing a comprehensive perspective on the data (McCluskey & Lalkhen, 2007, p. 209).

Proper result interpretation should encompass considerations such as decisions regarding null hypothesis, practical significance, effect size, and confidence intervals (McCluskey & Lalkhen, 2007, p. 208). By adhering to these principles and appropriately interpreting results, researchers can transform data into insights.

## 3. Introduction to Non-Parametric Tests in Inferential Statistics

Nachar (2008, p. 13) states that in inferential statistics non-parametric tests are utilized when the data does not meet the assumptions required for parametric tests. These assumptions typically include a distribution or homogeneity of variances. Non-parametric tests are often referred to as distribution-free tests because they make assumptions about the underlying population distribution. They are commonly employed when dealing with ordinal, nominal, or skewed data.

### 3.1. Detailed Explanation of Mann-Whitney U Test

The Mann Whitney U test, also known as the Wilcoxon Rank Sum Test, is a technique that falls under non-parametrical methods (Nachar, 2008, p. 14). It allows for a comparison between two independent samples. It helps us determine whether they likely originate from different populations. This test is particularly valuable when working with small sample sizes and data that do not adhere to a normal distribution pattern. This method focuses less on the values of the data and concentrates more on comparing their ranks, which makes it a robust choice applicable, in various scenarios.

### 3.1.1. Steps to Conduct the Mann-Whitney U Test

Mann-Whitney U Test follows a certain set of steps that are essential for its application. These steps include:

**Step 1: Establish Hypotheses and Significance Level (α):**

- Null Hypothesis ($H_0$): This states that the two populations from which the samples are drawn are equal.
- Alternative Hypothesis ($H_1$): This suggests that the two groups are not equal and there is a difference, between them.
- Choose a significance level (α), usually set at 0.05. This represents the threshold for statistical significance; you will compare your test statistic to critical values to make decisions.

**Step 2: Select the Test Statistic (U):**

Mann & Whitney (1947, p. 51) states that value of U can determined using the provided formula below.

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} \tag{1}$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} \tag{2}$$

$U_1$ represents the sum of ranks for the first group and $U_2$ represents the sum of ranks for the second group. The test statistic, U is determined as whichever value is smaller between $U_1$ and $U_2$. This helps measure the evidence supporting or opposing the null hypothesis.

**Step 3: Find the Critical Value to Establish Decision Rule:**

To make a decision about whether to reject the null hypothesis, you need to find the critical value of U based on your sample sizes and chosen significance level (α) (Mann & Whitney, 1947, p. 51). This critical value can be obtained from tables of critical values specific to the Mann-Whitney U test.

- The decision rule is simple: Reject $H_0$ if the observed U is less than or equal to the critical value. If U is greater then you should not reject the null hypothesis.

**Step 4: Calculation the Test Statistic (U):**

Assign ranks to the data, within each group. In case of tied values assign them the mean rank. This ensures that tied values receive a rank on average. Calculate the sum of ranks for each group. Label them as $R_1$ and $R_2$ respectively.

**Step 5: Results:**

Compare the computed value of U with the value obtained critical value in Step 3. Nachar (2008, p.16) states that If the computed U is less than or equal to the critical value it means that you reject the hypothesis ($H_0$) in favor of the alternative hypothesis ($H_1$). This indicates that there is statistical evidence suggesting a difference between the two populations. On the other hand, if U is greater than the critical value you cannot reject the null hypothesis. In this scenario there isn't evidence to conclude that there is a difference, between the two populations.

### 3.1.2. Interpreting the Mann-Whitney U Test Results

Nachar (2008, p.16) states that rejecting the null hypothesis implies that there is statistical evidence to suggest that the two populations are different, or that the samples were drawn from populations with different distributions.

Failing to reject the null hypothesis suggests that there is not enough evidence to conclude that the two populations differ significantly. This does not necessarily mean the populations are identical; it simply means you couldn't detect a significant difference with the given data and significance level.

## 3.2. Detailed Explanation of Kruskal Wallis Test

The Kruskal Wallis test is a test that doesn't rely on assumptions and is used to determine if there are significant differences, between three or more independent groups by comparing their medians (Niedoba et al., 2023, p. 4). It's an extension of the Mann Whitney U test, which is used for two groups. It comes into play when the assumptions of analysis of variance (ANOVA) cannot be met (Feir-Walsh & Toothaker, 1974, p. 791).

### 3.2.1 Steps to conduct the Kruskal Wallis Test

Kruskal Wallis Test follows a defined set of steps. These steps include:

**Step 1: Establish Hypotheses and Significance Level (α):**

- Null Hypothesis ($H_0$): This is the default assumption stating that there are no differences in the population medians among all the groups being compared.
- Alternative Hypothesis ($H_1$): This is what we aim to find evidence for suggesting that one groups population median differs from others.
- Significance Level (α): This is the predetermined threshold for statistical significance. A commonly used value is 0.05, which corresponds to a 5% chance of making a Type I error (incorrectly rejecting a null hypothesis).

**Step 2: Select the Test Statistic (H):**

The Kruskal Wallis test uses the H as the test statistic. It's calculated using the following formula:

$$H = \left( \frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} \right) - 3(N+1) \tag{3}$$

Where:

- H is test statistic
- N is the total number of observations in all groups
- $R_j$ is sum of the ranks in group j
- $n_j$ is the number of observations in group j

**Step 3: Establish Decision Rule:**

To determine whether to reject the null hypothesis, you need to compare the calculated H statistic to a critical value from the chi-square distribution with degrees of freedom (df) equal to $k - 1$, where k is the number of groups.

For example, if you have 3 groups, you will look up the critical value from a chi-square distribution with 2 degrees of freedom for your chosen significance level (α). If the calculated H exceeds than this critical value, you reject the null hypothesis.

**Step 4: Compute the Test Statistic (H):**

Here's how you calculate the H statistic:

1. Combine all the data from the different groups into one large dataset.
2. Rank the combined data in ascending order.
3. Calculate the sum of ranks for each group by adding up ranks assigned to observations, within each group.
4. Use these values to compute the H statistic using the formula mentioned in Step 2.

**Step 5: Results:**

- If the calculated H statistic is greater than the critical value from the chi-square distribution, you reject the null hypothesis.
- If the calculated H statistic is lower than the critical value, you do not reject the null hypothesis.

### 3.2.2 Interpretation of Kruskal-Wallis Test Results

let's expand on the interpretation of the results for the Kruskal Wallis test:

- Rejecting the null hypothesis indicates that there is statistical evidence suggesting that at least one groups population median differs from others.

- Failing to reject the null hypothesis implies that there is not enough evidence to conclude a significant difference in population medians, among groups.

## 4. Self-Developed Real-Life Example: Cancer Rates Before and After Chernobyl

In this section, one of the non-parametric tests that we discussed before will be used to get insights about real-life problems.

### 4.1. Problem Statement

The Chernobyl nuclear disaster that occurred in 1986 had an impact on the health of individuals residing in the affected areas (Reste et al., 2016, p. 257). One aspect of concern is whether there was a significant change in cancer rates, specifically kidney and thyroid cancer, in Belarus before and after the Chernobyl accident. This example aims to investigate whether there is a statistically significant difference in cancer rates between these two-time periods.

### 4.2. Data Collection and Description

- Data Source: Dataset from Kaggle sourced from the Byelorussian Center of Medical Technologies.
- Data Periods: Two periods were considered: before the accident (1977 - 1985) and after the accident (1986 - 1994).
- Geographic Areas: Data collected for the cities of Gomel, Mogilev, and the whole country of Belarus.
- Measurement: Cancer cases per 100,000 people were recorded for each period and location.

The data are shown below.

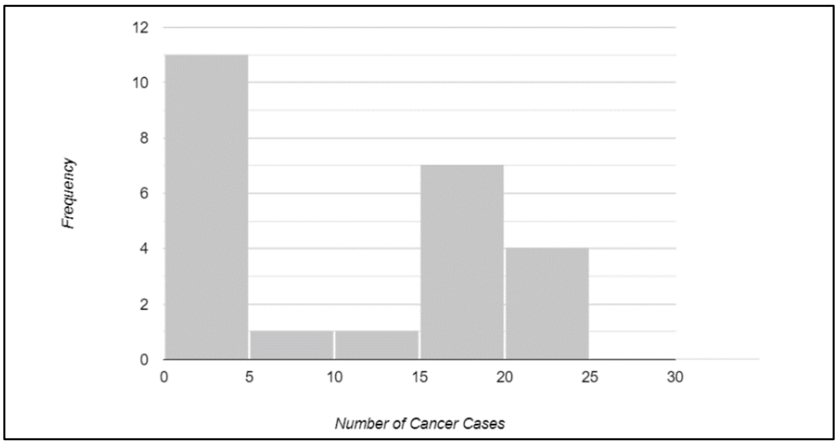
**Table 1: Cancer Cases Before and After the Chernobyl Disaster**

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Before Chernobyl | 1.5 | 1.9 | 1.7 | 2.8 | 3.1 | 3.5 | 1.6 | 1.3 | 1.5 | 0.7 | 0.5 | 0.8 |
| After Chernobyl | 3 | 3.1 | 3.7 | 6.3 | 6.8 | 7.5 | 6.8 | 3.5 | 4.5 | 2.5 | 1.1 | 1.4 |

Is there a significant difference in the incidence rates of kidney and thyroid cancer before and after the Chernobyl disaster in Belarus? By observation, it appears that there is a change in cancer rates following the Chernobyl disaster, but is this change statistically significant?

Frequency Histogram of number of cancer cases per 100,000 people is given below.

**Figure 1: Frequency Histogram Cancer Cases**




The histogram above shows a non-normal distribution as it is not favouring the bell-shaped structure of normal distribution.

## 4.3. Steps to Conduct the Test

The steps to conduct the test are given below.

**Step 1: Establish hypotheses and determine level of significance.**

- Null Hypothesis ($H_0$): There is no significant difference in cancer rates between the periods before and after the Chernobyl accident.
- Alternative Hypothesis ($H_1$): There is a significant difference in cancer rates between the two periods.
- Significance Level ($\alpha$) = 0.05

**Step 2: Choice of Non-Parametric Test**

Considering the sample sizes (n1 = 12 n2 = 12) and the objective of comparing two independent groups (before and, after Chernobyl) with non-normal distributions, we have chosen to employ the Mann Whitney U test as it aligns with our requirements.

**Step 3: Find critical value.**

The appropriate critical value can be found using the table of Critical values of the Mann-Whitney U (Two- tailed testing). To determine the critical value, sample size and level of significance are required. The critical value for this test with $n_1 = 12$, $n_2 = 12$ and $\alpha = 0.05$ is 37.

- The decision rule: Reject H0 if $U \leq 37$.

**Step 4: Calculate the Test Statistic**

First step is to assign ranks. The process of assigning can be done by making a table with relevant data as show below.

**Table 2: Sum of Ranks Calculation**

| | | Sample in Ascending order | | Ranks | |
|---|---|---|---|---|---|
| Before | After | Before | After | Before | After |
| 1.5 | 3 | 0.5 | | 1 | |
| 1.9 | 3.1 | 0.7 | | 2 | |
| 1.7 | 3.7 | 0.8 | | 3 | |
| 2.8 | 6.3 | | 1.1 | | 4 |
| 3.1 | 6.8 | 1.3 | | 5 | |
| 3.5 | 7.5 | | 1.4 | | 6 |
| 1.6 | 6.8 | 1.5 | | 7.5 | |
| 1.3 | 3.5 | 1.5 | | 7.5 | |
| 1.5 | 4.5 | 1.6 | | 9 | |
| 0.7 | 2.5 | 1.7 | | 10 | |
| 0.5 | 1.1 | 1.9 | | 11 | |
| 0.8 | 1.4 | | 2.5 | | 12 |
| | | 2.8 | | 13 | |
| | | | 3 | | 14 |
| | | | 3.1 | | 15.5 |
| | | 3.1 | | 15.5 | |
| | | | 3.5 | | 17.5 |
| | | 3.5 | | 17.5 | |
| | | | 3.7 | | 19 |
| | | | 4.5 | | 20 |
| | | | 6.3 | | 21 |
| | | | 6.8 | | 22.5 |
| | | | 6.8 | | 22.5 |
| | | | 7.5 | | 24 |
| | | | | $R_1 = 102$ | $R_2 = 198$ |

Sum the ranks in each group.

- Before Chernobyl total rank ($R_1$) = 102
- After Chernobyl total rank ($R_2$) = 198

Calculate $U_1$ and $U_2$ as follows:

- $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} = 12(12) + \frac{12(13)}{2} - 102 = 72$
- $U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} = 12(12) + \frac{12(13)}{2} - 198 = 24$

. The smaller test statistic is $U_2$. Therefore, U = 24.

**Step 5: Results:**

Upon conducting this analysis, we obtained a calculated test statistic U which is 24. Based on our selected significance level (α = 0.05) we determined that the critical value is 37.

Based on the calculated U statistic (24) being lower than the critical value (37) we can confidently reject the null hypothesis. This finding offers statistical evidence indicating a significant disparity in cancer rates between the periods prior to and following the Chernobyl incident.

### 4.4. Limitations and Future Research

It's important to note that this analysis has its limitations as it solely focuses on kidney and thyroid cancer rates in affected regions without examining types of cancer or individual risk factors. For future research it is recommended to broaden the scope by including various types of cancer, radiation levels, socioeconomic factors and access to healthcare. Conducting long term studies would provide a comprehensive understanding.

In summary through the implementation of the Mann Whitney U test we were able to identify a discrepancy in kidney and thyroid cancer rates before and after Chernobyl. However further investigation is necessary to evaluate its impact on public health.

## 5. Conclusion

In conclusion, this term paper has provided a comprehensive overview of statistical testing and its role in research. We have explored principles and methodologies while emphasizing their practical implications. By presenting a real-life case study on cancer rates pre and post the Chernobyl incident we have showcased the usefulness of non-parametric tests such as the Mann Whitney U Test in analysing intricate data.

Statistical testing is not merely a set of calculations rather it serves as a tool, for making well informed decisions using data. As we conclude we recognize that research is a journey that constantly presents new questions. We leave with a commitment to continue exploration, knowing that statistical testing will remain a valuable companion in our pursuit of knowledge.

## 6. List of Tables and Figures



## 7. List of Equations

## 8. Bibliography


Conn, V. S., Chan, K. C., & Cooper, P. S. (2014). The Problem With p. *Western Journal of Nursing Research, 36*(3), 291–293. https://doi.org/10.1177/0193945913492495

Erdogmus, H. (2022). Bayesian Hypothesis Testing Illustrated: An Introduction for Software Engineering Researchers. *ACM Computing Surveys, 55*(6), 1–28. https://doi.org/10.1145/3533383

Feir-Walsh, B. J., & Toothaker, L. E. (1974). An Empirical Comparison of the Anova F-Test, Normal Scores Test and Kruskal-Wallis Test Under Violation of Assumptions. *Educational and Psychological Measurement, 34*(4), 789–799. https://doi.org/10.1177/001316447403400406

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics, 4th Ed* (4th Ed.). SAGE.

Hung, H. M. J., O'Neill, R. T., Bauer, P., & Kohne, K. (1997). The Behavior of the P-Value When the Alternative Hypothesis is True. *Biometrics, 53*(1), 11. https://doi.org/10.2307/2533093

McCluskey, A., & Lalkhen, A. G. (2007). Statistics IV: Interpreting the results of statistical tests. *Continuing Education in Anaesthesia Critical Care & Pain, 7*(6), 208–212. https://doi.org/10.1093/bjaceaccp/mkm042

Nachar, N. (2008). The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution. *Tutorials in Quantitative Methods for Psychology, 4*(1), 13–20. https://doi.org/10.20982/tqmp.04.1.p013

Niedoba, T., Surowiak, A., Hassanzadeh, A., & Khoshdast, H. (2023). Evaluation of the Effects of Coal Jigging by Means of Kruskal–Wallis and Friedman Tests. *Energies, 16*(4), 1600. https://doi.org/10.3390/en16041600

Reste, J., Zvagule, T., Kurjāne, N., Šķesters, A., Silova, A., Eglīte, M., Cīrule, J., Gabruševa, N., Zīverts, A., & Čurbakova, E. (2016, October 1). Investigations on Health Conditions of Chernobyl Nuclear Power Plant Accident Recovery Workers from Latvia in Late Period after Disaster. *Proceedings of the Latvian Academy of Sciences. Section B. Natural, Exact, and Applied Sciences., 70*(5), 257–265. https://doi.org/10.1515/prolas-2016-0040