Journal of Advanced Zoology

ISSN: 0253-7214

Volume 45 Special Issue 02 Year 2023 Page 1636:1644

# Non parametric statistical tools in biological research: A Review

**Yaser Qureshi [1], Dharmender Singh [2]**

1. **Govt. College Khertha, Distt. Balod, Chhattisgarh, India**
2. **Govt. N.C.J. College Dallirajhara, Distt. Balod, Chhattisgarh, India**

***Corresponding author – E-mail = dryaserqureshi@gmail.com**

***Article History***

Received: 12 March 2023
Revised: 21 August 2023
Accepted: 09 October 2023



Nonparametric statistical tools, also known as distribution-free methods, are a set of techniques used to analyze data when certain assumptions about the underlying population distribution are not met or when little is known about the population parameters. Unlike parametric methods, nonparametric methods do not rely on specific assumptions about the shape or parameters of the population distribution.

Nonparametric statistical tools are useful in situations where the data may not follow a specific distribution, have outliers, or exhibit nonlinearity. They are also valuable when dealing with small sample sizes or ordinal or categorical data. Nonparametric methods can provide robust and reliable results in such cases.

Here are a few commonly used nonparametric statistical tools:

1. Mann-Whitney U test: This test is used to compare two independent groups to determine if there is a statistically significant difference between their distributions. It does not assume normality and is applicable to ordinal or continuous data.
2. Wilcoxon signed-rank test: Similar to the Mann-Whitney U test, this test is used to compare two related or paired samples. It assesses whether there is a significant difference between the paired observations, without assuming a specific distribution.
3. Kruskal-Wallis test: When comparing three or more independent groups, this test determines if there are significant differences among the groups. It is a nonparametric alternative to the one-way analysis of variance (ANOVA) test.
4. Friedman test: This nonparametric test is used when comparing three or more related samples. It examines if there are significant differences among the related groups. It is the analogous nonparametric version of repeated measures ANOVA.
5. Spearman's rank correlation coefficient: This nonparametric measure assesses the strength and direction of the monotonic relationship between two variables. It is used when the data is ordinal or when the relationship is not linear.

6. Kendall's tau correlation coefficient: Similar to Spearman's rank correlation, Kendall's tau measures the strength and direction of the monotonic relationship between two variables. It is also suitable for ordinal data.

These are just a few examples of nonparametric statistical tools. There are many more available depending on the specific analysis and data characteristics. Nonparametric methods provide flexibility and robustness in analyzing data without making strong assumptions about the underlying distribution, making them valuable in a wide range of research fields and applications.

Nonparametric statistical tests have several advantages and limitations compared to their parametric counterparts. Here are some of the key advantages and limitations of nonparametric tests:

Advantages of nonparametric tests:

1. Distribution-free: Nonparametric tests do not rely on assumptions about the shape or parameters of the population distribution. They are applicable when the data does not follow a specific distribution, or when the distribution is unknown or cannot be assumed.
2. Robustness: Nonparametric tests are generally more robust to violations of assumptions than parametric tests. They can handle outliers and non-normal data more effectively, making them suitable for data with extreme values or when the assumptions of normality are not met.
3. Suitable for small sample sizes: Nonparametric tests can provide reliable results even with small sample sizes. They do not require a large sample size to make valid inferences, unlike some parametric tests.
4. Applicable to ordinal and categorical data: Nonparametric tests can be used with ordinal or categorical data, where parametric tests would not be appropriate. They allow for analysis of data with ranked or grouped categories without the need for transformation.
5. Less stringent assumptions: Nonparametric tests have fewer assumptions compared to parametric tests. This can be advantageous when the underlying assumptions of parametric tests, such as normality or homogeneity of variances, are not met.

Limitations of nonparametric tests:

1. Less statistical power: Nonparametric tests generally have less statistical power compared to their parametric counterparts, especially when the assumptions of the parametric tests are met. Nonparametric tests may require larger sample sizes to achieve the same level of power as parametric tests.
2. Limited in scope: Nonparametric tests are not suitable for all types of statistical analyses. They have specific applications and may not be applicable in situations where parametric tests are better suited or when specific assumptions can be reasonably met.
3. Less precise estimation: Nonparametric tests often provide less precise estimation of population parameters compared to parametric tests. This is because nonparametric methods do not make assumptions about the underlying distribution, leading to wider confidence intervals and less precise estimates.

4. Reduced ability to detect subtle differences: Nonparametric tests may be less sensitive to detecting small or subtle differences between groups or variables. They are generally designed to detect larger differences or relationships and may have limited power to identify small effects.
5. Complexity: Nonparametric tests can be more complex and computationally intensive compared to parametric tests. They may require more computational resources and specialized software for implementation.

It is important to carefully consider the advantages and limitations of nonparametric tests in relation to the specific research question, data characteristics, and available resources before selecting the appropriate statistical analysis approach.

Nonparametric statistical tests are applicable to a wide range of biological data types, particularly when the data does not meet the assumptions of parametric tests or when specific distributional assumptions cannot be made. Here are some examples of biological data suitable for nonparametric statistical tests:

1. Ordinal data: Nonparametric tests are well-suited for analyzing data that can be ranked but may not have equal intervals between categories. This includes data such as Likert scale ratings, subjective assessments, or ordered categories.
2. Categorical data: Nonparametric tests can handle categorical data where observations fall into distinct groups or categories. For example, analyzing the distribution of genotypes or the presence/absence of a particular trait among different groups.
3. Count data: Nonparametric tests can be applied to count data, where the outcome represents the number of occurrences of an event within a fixed period or region. This includes analyzing ecological counts, microbial abundances, or the frequency of rare events.
4. Ranked data: Nonparametric tests are useful for analyzing data that has been explicitly ranked or sorted based on a particular criterion. This could include rankings of performance, preference, or effectiveness, such as ranking the efficacy of different drugs or treatments.
5. Skewed or non-normally distributed data: Nonparametric tests are appropriate when the data exhibits a non-normal distribution, including skewed or heavy-tailed distributions. For instance, analyzing gene expression levels, protein concentrations, or metabolite concentrations that often follow non-normal distributions.
6. Survival data: Nonparametric tests, such as the Kaplan-Meier estimator and the log-rank test, are commonly used in survival analysis. They are suitable for analyzing time-to-event data, such as studying survival rates, disease progression, or time until relapse.
7. Matched or paired data: Nonparametric tests can be employed when dealing with paired data or when each observation in one group is directly related to an observation in another group. For instance, analyzing pre- and post-treatment measurements in the same individuals or comparing data from twins or siblings.

It's important to note that the appropriateness of nonparametric tests for specific biological data types depends on the research question, study design, and the specific characteristics of the data. It is advisable to consult with a statistician or data analyst to ensure the correct choice and interpretation of nonparametric statistical tests for a particular biological study.

**Popular non parametric tests used in biological research**

In biological research, several nonparametric tests are commonly used to analyze data when the assumptions of parametric tests are not met or when dealing with specific types of data. Here are some popular nonparametric tests frequently used in biological research, along with a detailed explanation of each:

1. **Mann-Whitney U test (also known as Wilcoxon rank-sum test):** The Mann-Whitney U test is used to compare two independent groups to determine if there is a statistically significant difference between their distributions. It does not assume normality and is applicable to ordinal or continuous data. Here's how it works:

- The test ranks all the observations from both groups combined, from lowest to highest.
- It then calculates the sum of the ranks for each group separately (U1 for group 1 and U2 for group 2).
- The test statistic, U, is the smaller of U1 and U2.
- The significance of U is determined by comparing it to the critical value from the Mann-Whitney U distribution or by using a p-value.

The Mann-Whitney U test is widely used in biological research to compare variables between different treatment groups or conditions.

2. **Wilcoxon signed-rank test:** The Wilcoxon signed-rank test is used to compare two related or paired samples. It assesses whether there is a significant difference between the paired observations without assuming a specific distribution. Here's how it works:

- The test ranks the absolute differences between paired observations.
- It then calculates the sum of the ranks of the positive differences (W+) and the sum of the ranks of the negative differences (W-).
- The test statistic, W, is the smaller of W+ and W-.
- The significance of W is determined by comparing it to the critical value from the Wilcoxon signed-rank distribution or by using a p-value.

The Wilcoxon signed-rank test is commonly used in biological research when studying paired data, such as pre- and post-treatment measurements or left-right comparisons.

3. **Kruskal-Wallis test:** The Kruskal-Wallis test is a nonparametric alternative to the one-way analysis of variance (ANOVA) test. It is used to compare three or more independent groups and determine if there are significant differences among the groups. Here's how it works:

- The test ranks all the observations from all the groups combined, from lowest to highest.
- It then calculates the sum of the ranks for each group separately (R1, R2, R3, etc.).

- The test statistic, H, is calculated using the formula: H = [12 / (N(N+1))] * [(ΣR^2_i / n_i) - 3(N+1)], where N is the total number of observations and n_i is the number of observations in each group.
- The significance of H is determined by comparing it to the critical value from the chi-square distribution or by using a p-value.

The Kruskal-Wallis test is commonly used in biological research when comparing multiple groups, such as analyzing the effect of different treatments or interventions on a particular outcome.

4. **Friedman test:** The Friedman test is the nonparametric equivalent of the repeated measures ANOVA and is used when comparing three or more related samples. It examines if there are significant differences among the related groups. Here's how it works:

- The test ranks the observations within each group separately.
- It then calculates the sum of the ranks for each group (W1, W2, W3, etc.).
- The test statistic, χ^2, is calculated using the formula: χ^2 = [12 / (kN(N+1))] * [ΣW^2_j - 3N(N+1)], where k is the number of groups, and N is the number of observations per group.
- The significance of χ^2 is determined by comparing it to the critical value from the chi-square distribution or by using a p-value.

The Friedman test is commonly used in biological research when analyzing repeated measures or when comparing multiple treatments or interventions over time.

5. **Spearman's rank correlation coefficient:** Spearman's rank correlation coefficient is a nonparametric measure that assesses the strength and direction of the monotonic relationship between two variables. It is used when the data is ordinal or when the relationship is not linear. Here's how it works:

- The test ranks the observations of both variables.
- It calculates the differences between the ranks for each pair of observations.
- It computes the correlation coefficient, rho (ρ), which ranges from -1 to 1. Positive values indicate a direct relationship, negative values indicate an inverse relationship, and zero indicates no monotonic relationship.
- The significance of ρ is determined by comparing it to the critical value from the t-distribution or by using a p-value.

Spearman's rank correlation coefficient is commonly used in biological research to assess associations between variables that may not have a linear relationship or when the data is ranked or ordinal.

These nonparametric tests are just a few examples of the many available techniques used in biological research. They provide valuable tools for analyzing data when specific distributional assumptions cannot be made or when dealing with specific types of data, such as ranked, categorical, or skewed data.

Following are a few case studies and examples of nonparametric analysis in biological research:

1. Case Study: Comparison of gene expression levels Research Question: Is there a significant difference in gene expression levels between two treatment groups?

Nonparametric Test: Mann-Whitney U test

Description: In a study investigating the effect of a drug on gene expression, researchers collected RNA-seq data from two groups: a control group and a treatment group. The gene expression levels were not normally distributed. To compare the two groups, the researchers performed a Mann-Whitney U test. The test revealed a significant difference in gene expression levels between the two groups, indicating that the drug had an impact on gene expression.

2. Case Study: Association between biomarkers and disease severity Research Question: Is there an association between the levels of two biomarkers and the severity of a disease?

Nonparametric Test: Spearman's rank correlation coefficient

Description: A study aimed to investigate the relationship between the levels of two biomarkers (Biomarker A and Biomarker B) and the severity of a particular disease. The researchers collected ordinal data representing disease severity and measured the levels of the two biomarkers in a cohort of patients. To assess the association, they used Spearman's rank correlation coefficient. The analysis revealed a significant positive correlation between Biomarker A levels and disease severity, indicating that higher levels of Biomarker A were associated with more severe disease.

3. Case Study: Comparison of treatment response in clinical trials Research Question: Is there a significant difference in treatment response among three different treatment groups?

Nonparametric Test: Kruskal-Wallis test

Description: In a clinical trial comparing the efficacy of three different treatments for a specific condition, researchers measured a continuous outcome variable related to treatment response. The data did not meet the assumptions of normality and homogeneity of variances. To determine if there were differences among the treatment groups, the researchers performed a Kruskal-Wallis test. The test results showed a significant difference among the treatment groups, indicating that the treatments had varying effects on treatment response.

4. Case Study: Survival analysis in cancer research Research Question: Are there differences in survival rates among different cancer treatment groups?

Nonparametric Test: Log-rank test

Description: In a study evaluating the survival outcomes of patients with different types of cancer undergoing different treatments, researchers collected time-to-event data, specifically the survival times of patients. The survival data did not follow a normal distribution. To compare the survival rates among the treatment groups, they used the log-rank test, a nonparametric test commonly used in survival analysis. The test revealed a significant difference in survival rates among the treatment groups, indicating that the treatments had varying effects on patient survival.

These examples demonstrate how nonparametric tests are utilized in various biological research scenarios, such as gene expression analysis, biomarker associations, treatment response comparisons, and survival analysis. Nonparametric methods provide valuable insights and robust statistical analyses in situations where parametric assumptions are not met or when dealing with specific types of data.

**Comparison of parametric and non parametric tests in biological research.**

In biological research, both parametric and nonparametric tests are used for statistical analysis, but they differ in their assumptions and applications. Here's an elaboration on the comparison of parametric and nonparametric tests in biological research:

**Assumptions:** Parametric tests make specific assumptions about the population distribution, typically assuming normality and equal variances. They rely on these assumptions to estimate parameters and calculate p-values accurately. On the other hand, nonparametric tests do not assume a specific population distribution and are distribution-free. They are based on fewer or weaker assumptions, such as independence, random sampling, and exchangeability.

**Data Types:** Parametric tests are suitable for continuous data that follow a specific distribution, whereas nonparametric tests are more flexible and applicable to a wider range of data types. Nonparametric tests can handle ordinal, categorical, count, and non-normally distributed data, making them useful for analyzing biological data that may not meet parametric assumptions.

**Sample Size:** Parametric tests generally require larger sample sizes to produce reliable results, especially when testing assumptions like normality and equal variances. Nonparametric tests, however, are often robust with smaller sample sizes. They can provide valid inferences even with limited data, making them suitable for biological studies with smaller sample sizes or rare events.

**Statistical Power:** Parametric tests generally have higher statistical power when the underlying assumptions are met. They can detect smaller effects and differences between groups more easily than nonparametric tests. Nonparametric tests, while robust, may have slightly lower power compared to parametric tests, particularly when assumptions of the parametric tests hold.

**Flexibility:** Nonparametric tests offer greater flexibility in terms of data analysis. They can be applied to a variety of study designs, including paired or matched samples, repeated measures, and multiple group comparisons. Parametric tests may have more limited applications and may not be suitable or provide accurate results for certain study designs or data types.

**Interpretation:** Parametric tests provide estimates of population parameters, such as means or regression coefficients, which can be interpreted directly. Nonparametric tests, being distribution-free, provide estimates that are often less precise or harder to interpret in terms of population parameters. They focus more on ranks, medians, or non-parametric effect sizes, making interpretation slightly different.

Overall, the choice between parametric and nonparametric tests depends on the nature of the data, assumptions, research question, and study design. While parametric tests have advantages in terms of power and precise estimation under appropriate assumptions, nonparametric tests offer flexibility, robustness, and wider applicability to various types of biological data. Researchers should carefully consider the characteristics of their data and the assumptions of the tests when selecting the appropriate analysis approach.

**Future prospects for non parametric statistical tools in biological research**

Nonparametric statistical tools have a promising future in biological research due to several factors. Here are some future prospects for nonparametric statistical tools in this field:

1. **Handling Complex Data Structures:** Biological research often involves complex data structures, such as longitudinal data, clustered data, or data with high dimensionality. Nonparametric methods can handle these complexities effectively and provide robust analysis approaches. As biological research continues to generate increasingly complex data, nonparametric tools will play a crucial role in extracting meaningful insights from such data structures.
2. **Omics Data Analysis:** The advent of high-throughput technologies has led to the generation of large-scale omics data, such as genomics, transcriptomics, proteomics, and metabolomics. Nonparametric methods can be particularly valuable in analyzing these types of data, which may not adhere to parametric assumptions or exhibit non-normal distributions. Nonparametric tools can help identify differentially expressed genes, detect associations between variables, and uncover patterns or clusters within omics datasets.
3. **Integration of Multidimensional Data:** Biological research often requires the integration of multiple data modalities, such as combining genetic data with clinical data or imaging data. Nonparametric methods can provide flexible approaches to integrate and analyze such multidimensional data, accommodating the different data types and handling the heterogeneity of the data sources.
4. **Personalized Medicine and Precision Biology:** Nonparametric methods can contribute to the advancement of personalized medicine and precision biology by analyzing individual-level data. Nonparametric tools can help identify biomarkers, assess treatment response, and stratify patients into subgroups based on their individual characteristics. This can lead to improved personalized treatment strategies and better understanding of disease mechanisms.
5. **Nonlinear Relationships and Complex Interactions:** Biological systems often involve nonlinear relationships and complex interactions between variables. Nonparametric methods, such as nonparametric regression or decision tree-based methods, are well-suited for capturing and modeling such complexities. They can uncover nonlinear associations, detect interaction effects, and provide insights into complex biological processes.
6. **Non-Euclidean Data Analysis:** Biological data, such as microbiome data or protein-protein interaction networks, are often non-Euclidean in nature and require specialized analysis approaches. Nonparametric tools, including permutation-based methods or rank-based methods, can handle non-Euclidean data structures and provide reliable statistical inferences.
7. **Robustness and Reproducibility:** Nonparametric methods are known for their robustness against violations of assumptions and their ability to provide reliable results

even with small sample sizes. This robustness contributes to the reproducibility of research findings, as nonparametric tools can deliver consistent results across different studies or datasets.

Given these prospects, nonparametric statistical tools will continue to be widely used in biological research, complementing parametric methods and providing valuable insights into complex biological phenomena. As research advances and generates more diverse and intricate data, the versatility, flexibility, and robustness of nonparametric methods will play an increasingly vital role in analyzing and interpreting biological data.

**Reference**

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association, 32(200), 675-701.

Kendall, M. (1938). A new measure of rank correlation. Biometrika, 30(1/2), 81-89.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association, 47(260), 583-621.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics, 18(1), 50-60.

Spearman, C. (1904). The proof and measurement of association between two things. American Journal of Psychology, 15(1), 72-101.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics Bulletin, 1(6), 80-83.