International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Comparative Study of Some Parametric Tests and Non-Parametric Tests in Statistics

Dr. Mohinder Pal

Associate Professor, Department of Statistics, Govt. Degree College Udhampur

Abstract

This paper provides a comparative analysis of parametric and non parametric statistical tests, evaluating their underlying assumptions, statistical power and robustness. This paper presents a comparative study of selected parametric and non-parametric statistical tests, discussing their assumptions, characteristics, applications, advantages, limitations, and appropriate areas of use. The study concludes that selecting the appropriate statistical test depends on the nature of the data, sample size, measurement scale, and research objectives rather than on the complexity or popularity of the method.

Keywords: Statistics, Parametric Tests, Non-Parametric Tests, Hypothesis Testing, Normal Distribution, Statistical Inference.

1. Introduction

This paper provides a comparative overview of parametric and non-parametric statistical tests, outlining when to use each. It highlights that parametric tests, such as t-tests and ANOVA, are more powerful but require strict assumptions like normal distribution. Conversely, non-parametric tests, like the Mann-Whitney U test, are robust alternatives for skewed or ordinal data.

Statistical tests are broadly divided into two main categories: parametric and non-parametric. Parametric tests make specific assumptions about the parameters of the population distribution from which the data is drawn. They typically require continuous data, homogeneity of variances, and a normally distributed population. Because they utilize the inherent information of the raw data, these tests are generally more powerful and sensitive.

However, real-world data often violates these stringent assumptions. Datasets may be heavily skewed, contain unmanageable outliers, or consist of ordinal variables rather than continuous measurements. In such scenarios, applying parametric tests can lead to highly misleading or erroneous conclusions. This is where non-parametric tests become essential. Non-parametric tests—such as the Mann-Whitney U test, Wilcoxon signed-rank test, and Kruskal-Wallis test—are "distribution-free" and make no strict assumptions about the underlying population parameters

This paper aims to thoroughly compare these two families of statistical methods. While parametric tests remain the gold standard when assumptions are met, non-parametric tests offer a robust, flexible alternative for non-normal or ordinal data. The primary objective of this study is to evaluate the theoretical foundations, assumptions, advantages, and limitations of both approaches. By systematically comparing their statistical power and practical applications, this paper will serve as a guide for researchers in selecting the most appropriate analytical tool for their specific dataset requirements.

2. Objectives

The objective of a comparative study on parametric and nonparametric tests is to evaluate how each handles different data types, identify when to use them based on underlying assumptions, and compare their statistical power. Ultimately, the goal is to guide researchers in selecting the most accurate test for their specific dataset. It guides researchers in selecting the correct analytical tool based on their specific data distributions.

2.1 Primary Objectives

Evaluate Underlying Assumptions: Determine how data distribution impacts the performance of parametric tests.

Identify Appropriate Applications: Match specific data types—like interval/ratio data for parametric tests versus ordinal/nominal data for nonparametric tests—to the correct analytical methods.

Compare Statistical Power: Assess the probability of avoiding Type II errors, showing that parametric tests are generally more powerful when assumptions are met, while nonparametric tests are preferred for non-normal or skewed data.

3. Parametric Tests

3.1 Definition

Most of the statistical tests, we perform are based on a set of assumptions. When these assumptions are violated the results of the analysis can be misleading or completely erroneous

3.2 Assumptions of Parametric Tests

Before running a parametric test, your data must typically meet three strict criteria:

1. **Normality:** The data or the errors must be approximately normally distributed.
2. **Homogeneity of Variance:** The variance between the groups being compared should be roughly equal.
3. **Independence:** Observations must be independent of one another .

3.3 Z-Test

A Z-test is a statistical method used to determine whether the average of a sample significantly differs from a known population mean or another sample's mean when the population variance is known and the sample size is large ($n > 30$).

3.3.1 Applications

1. **Testing population means:** evaluates whether a sample mean significantly differs from a known population mean
2. **Comparing two population means**;. It is ideal when you have large sample sizes and the population standard deviations are known.
3. **Quality Control & Manufacturing**: to determine if a production process mean significantly deviates from a target standard. It relies on large sample sizes ($n \geq 30$) and a known population standard deviation.
4. **Education & Standardized Testing**: it is used to determine if a sample of student test scores is significantly different from a known population mean.
5. **Healthcare & A/B Testing**: it is invaluable for evaluating pilot programs, improving patient experiences, and measuring the success of health interventions.

3.4 Student's t-Test

The discovery of 't' is regarded as a landmark in the history of statistical inference. Before Student gave his 't', it was customary to replace σ^2 in $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ by its unbiased estimate to give $t = \frac{\bar{x}-\mu}{S/\sqrt{n}}$ and then normal test was applied even for small samples. It has been found that although the distribution of t is asymptotically normal for large samples it is far from normality for small samples. The Student's t ushered in an era of exact sample distribution (and tests) and since its discovery many important contributions have been made towards the development and extension of small (exact) sample theory.

3.4.1 Graph of t-distribution. The p.d.f. of t-distribution with n d.f is:

$$f(t) = C\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} \qquad -\infty < t < \infty$$

Where $c = \dfrac{1}{\sqrt{n}\, B\left(\frac{1}{2}, \frac{n}{2}\right)}$

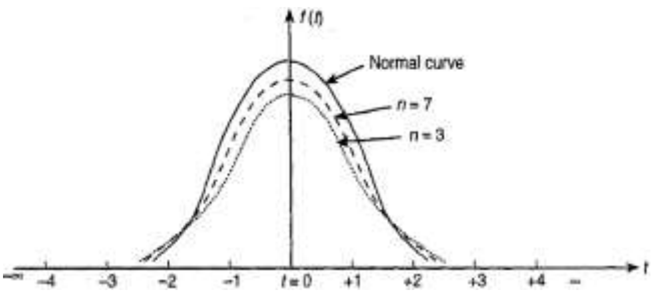
Since f(-t)=f(t), the probability curve is symmetrical about the line t = 0. As t increases, f(t) decreases rapidly and tends to zero as t $\to \infty$, so that t-axis is an asymptote to the curve. We have shown that

$$\mu_2 = \frac{n}{n-2} \qquad n > 2 \;; \qquad \beta_2 = \frac{3(n-2)}{(n-4)}, \quad n > 4$$

Hence for n> 2, μ_2>1 i.e., the variance of t-distribution is greater than that of standard normal distribution and for n > 4, β_2> 3 and thus t-distribution is more flat on the top than the normal curve. In fact, for small n, we have

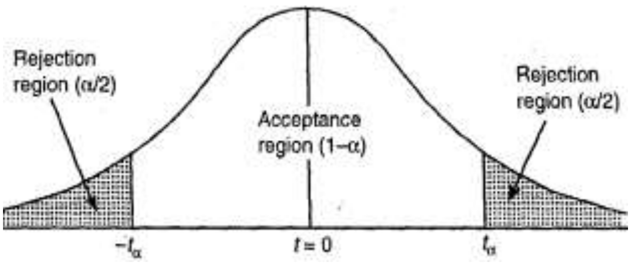
$$P[|t| \geq t_0] \geq P[|Z| \geq t_0], \qquad Z \sim N(0,1)$$

i.e., the tails of the t-distribution have a greater probability (area) than the tails of standard normal distribution.



3.4.2 Critical Values of t. The critical (or significant) values of t at level of significance α and d.f ν for two-tailed test are given by the equation

$$P[|t| > t_\nu(\alpha)] = \alpha \qquad \Rightarrow \qquad P[|t| \leq t_\nu(\alpha)] = 1-\alpha$$



Since t-distribution is symmetric about t = 0,so we have

$P[t > t_\nu(\alpha)] + P[t < -t_\nu(\alpha)] = \alpha \quad \Rightarrow \quad 2P[t > t_\nu(\alpha)] = \alpha$

$\Rightarrow \quad P[t > t_\nu(\alpha)] = \alpha/2 \quad$ Therefore $\quad P[t > t_\nu(2\alpha)] = \alpha$

$t_\nu(2\alpha)$ (from the Tables) gives the significant value of t for a single-tail test [Right-tail or Left-tail-since the distribution is symmetrical], at level of significance α and ν d.f. Hence the significant values of t at level of significance 'α'for a single-tailed test can be obtained from those of two-tailed test by looking the values at level of significance 2α .

3.4.3 APPLICATIONS

1. To test if the sample mean differs significantly from the hypothetical value μ of the population mean
2. To test the significance of the difference between two sample means.
3. To test the significance of an observed sample correlation coefficient and sample regression coefficient.
4. To test the significance of observed partial correlation coefficient.

3.5 Analysis of Variance (ANOVA)

Analysis of variance is a technique for testing the simultaneous significance of the difference among the mean of several categories (often called classes or groups) as developed by Prof. R.A. fisher in 1920's. According to him "Separation of variance ascribable to one group of causes from the variance ascribable to other groups was the form of this concept." Formally analysis of variance may be defined as:

Analysis of variance is a technique of separating the total variation observed in a set of observed data into its non-negative components, each of which measures the variability ascribable to different specific sources (or factors) and then estimating the variance of the population by independent estimates followed by a comparison of variation due to each assignable factor with that of the chance factor."

3.5.1 Types of ANOVA

a. One-Way ANOVA

Used to determine if there are any statistically significant differences between the means of three or more independent groups based on *one* categorical independent variable (or factor).

b. Factorial ANOVA (Two-Way, Three-Way, etc.)

Used when a study involves *two or more* categorical independent variables. This allows researchers to examine not only the individual effect of each independent variable but also how they interact with one another.

c. Repeated Measures ANOVA

Used when the same subjects are tested multiple times across different conditions or time points. This method removes individual baseline differences, increasing statistical power.

d. Mixed ANOVA

Combines between-subjects and within-subjects designs. It evaluates at least one independent variable that is tested across independent groups, alongside at least one independent variable that is measured repeatedly for each participant.

3.5.2 Applications

The main application of the analysis of variance is to test the homogeneity of the observations. In fact, it is an improvement upon student's t-test for testing the significance of several means taken together as a group. In analysis of variance all the sample means are studied simultaneously and the population variance is estimated through the pooled variance. Furthermore, some main applications/utility in medical science, agriculture science, business as well as education sectors are hereby mentioned below:

1. **Medical & Healthcare:** Used to test the efficacy of multiple medications or treatments.
2. **Manufacturing & Operations:** Used to compare defect rates across various production lines, test the durability of different materials, or evaluate performance from multiple suppliers.
3. **Agriculture & Biology:** Helps researchers compare crop yields across different fertilizer treatments or soil types to optimize agricultural output.
4. **Business & Marketing:** Used to analyze how different pricing tiers, promotional strategies, or product features impact sales volumes across different demographics.
5. **Education & Psychology:** Assists educators in evaluating the effectiveness of different teaching methodologies and testing the cognitive impacts of therapies or age differences.

3.5.3 Assumptions:

1. The observations are independent and are distributed about a true but unknown mean μ, say.
2. The parent population of distributions are normal with common but unknown variance σ^2.
3. Each observation is composed of a general (inherent) effect and the relevant effects under study are in additive form.
4. The errors attached to each observation are independently and normally distributed with mean zero and variance σ^2.
5. The variances are homogeneous. When all the subpopulations have the same variance they are said to be homoplastic, homoscedastic.
6. There should be at least two observations within each class otherwise analysis of variance technique cannot be applied.

3.6 Pearson's Correlation Coefficient

The Pearson correlation coefficient, denoted as 'r' measures the strength and direction of the linear relationship between two variables.

3.6.1 Characteristics

Quantifies Association: It helps researchers identify how changes in one variable correspond to changes in another.

No Causation: It only measures association and does not establish a cause-and-effect relationship between variables.

3.6.2 Interpretation Guide:

Direction: A positive *r* indicates variables move together, while a negative *r* indicates they move in opposite directions.

Strength: Generally, values closer to 1 indicate a stronger linear relationship, while values closer to 0 represent a weaker one.

3.6.3 Critical Assumptions

For the Pearson correlation coefficient to be mathematically valid, researchers must verify that their data meets the following assumptions:

Continuous Variables: Data must be measured on an interval or ratio scale.

Linearity: The relationship between the two variables must be linear, which is typically verified using a scatter plot.

Normality: Variables should follow an approximately normal distribution.

Homoscedasticity: The variance of the residuals should be constant across all values of the independent variable.

Independence: The observations must be independent of one another.

3.6.4 Applications

1. **Healthcare & Medicine:** Medical researchers use it to determine the linear association between continuous metrics, such as evaluating if increases in patient age strongly correlate with higher Body Mass Index (BMI) or blood pressure.
2. **Finance & Economics:** Analysts employ it to construct correlation matrices, helping them examine how two different assets or stock prices move together to diversify portfolios and manage risk.
3. **Machine Learning & Data Science:** It is frequently used for feature selection in data pipelines. By dropping features that are highly correlated with one another, developers can reduce redundancy and multicollinearity in predictive models.
4. **Social & Behavioral Sciences:** Psychologists use the coefficient to evaluate test results—for instance, measuring if a student's attitude toward standardized testing strongly correlates with their final test scores.

4. Non-Parametric Tests

4.1 Definition

Non-parametric test does not make any assumption regarding the form of the population, it is a "distribution-free" statistical method used for hypothesis testing that does not assume your data follows a specific probability distribution. They are often referred to as **distribution-free tests** and are particularly useful for ordinal, ranked, or non-normal data. These tests are also preferred when sample sizes are small or when outliers make parametric assumptions invalid.

4.2 Characteristics

No Distributional Assumptions: They make no assumptions about the shape or parameters of the population distribution

Focus on Medians: Instead of evaluating parameters like the population mean, they generally analyze medians, ranks, or signs.

Handles Any Data Type: They are ideal for nominal and ordinal categorical scales, and work well when numerical values are just ranks.

Robust to Outliers: Because they often rely on data ranking rather than absolute numerical values, extreme outliers have minimal impact on the results.

Lower Statistical Power: If your data does meet the strict assumptions of parametric tests, non-parametric tests are generally less powerful.

4.3 Chi-Square Test (χ^2)

The distribution of the random variable χ_k^2 which was first discovered by Helmert in 1876 and later independently given by Karl Pearson in 1900 when Karl Pearson used it for frequency data classified into k-mutually exclusive categories.

Another way to understand chi-square is: if X_1, X_2, . . .Xn are n independent normal variates with mean zero and variance unity, the sum of squares of these variates is distributed as chi-square with n degrees of freedom.

4.3.2 Applications

1. **χ^2 test for Inferences About a Population Variance:** Under the null hypothesis that the population variance is $\sigma^2 = \sigma_0^2$, the statistic

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(x_i-\bar{x})^2}{\sigma_0^2}\right] = \frac{1}{\sigma_0^2}\left[\sum_{i=1}^{n}x_i^2 - \frac{(\sum x_i)}{n}\right] = \frac{ns^2}{\sigma_0^2}$$

follows chi-square distribution with (n -1) d.f.

By comparing the calculated value with the tabulated value of χ^2 for (n -1) d.f at certain level of significance (usually 5%), we may retain or reject the null hypothesis.

If the sample size n is large (>30), then we can use Fisher's approximation and apply Normal Test.

$\sqrt{2\chi^2} \sim N(\sqrt{2n-1},\ 1)$ so that $Z = \sqrt{2\chi^2} - (\sqrt{2n-1}) \sim N(0,1)$

2. **χ^2 test for Goodness of Fit Test**: this test is used for testing the significance of the discrepancy between theory and experiment was given by Prof. Karl Pearson.
3. **χ^2 Test of Independence of Attributes**: to analyze frequency data in contingency tables, comparing observed counts with what would be expected by random chance.

4.4 Mann–Whitney U Test

The Mann-Whitney U test is a non-parametric test used to determine if there is a statistically significant difference between two independent groups and alternative to the independent t-test and is used when your data is ordinal or continuous but violates the assumption of normality.

4.4.1 Applications

Medical & Healthcare: Comparing the effectiveness of two different treatments or medications on patient recovery times when the data violates normality assumptions.

Psychology & Social Sciences: Analyzing Likert scale survey responses or attitude scores between two different demographic groups or control/experimental conditions.

Business & Marketing: Comparing consumer preferences, purchase amounts, or engagement scores between two different target segments or website design variations.

Education: Comparing test scores or grading distributions between two different teaching methods, especially with small sample sizes.

4.5 Wilcoxon Signed-Rank Test

It is the go-to alternative to the paired *t*-test when your data is ordinal or the differences between pairs are not normally distributed.

4.5.1 Applications

1. **Medical & Clinical Trials:** Comparing patient vitals before and after administering a new drug.

2. **Education & Training:** Testing student exam scores or athlete performance metrics before and after completing a specific training program.

4.6 Kruskal–Wallis Test

The Kruskal-Wallis test extends the Mann-Whitney U test to more than two independent groups. This test is used when the assumptions for a one-way analysis of variance are not met. Because the Kruskal-Wallis test is a non-parametric test, the data used do not have to be normally distributed.

4.6.1 Applications

1. **Comparing more than two independent groups**: to used when your continuous or ordinal data violates normality or homogeneity of variance assumptions.
2. **Medical and social science research** : it is used to compare the medians of three or more independent groups when the dependent variable is ordinal or continuous, but the data fails to meet normality assumption.

4.7 Spearman's Rank Correlation

The Spearman rank correlation examines the relationship between two variables and is the non-parametric counterpart of Pearson's correlation. Spearman correlation uses ranks rather than the original values, hence the name rank correlation.

4.7.1 Applications

1. **Survey and Questionnaire** : to measures the strength and direction of a monotonic relationship between two variables,
2. **Education and Human Resources**: commonly used to analyze non-normally distributed data, such as comparing survey responses, performance ratings, or skill rankings.
3. **Ranked observations**: It evaluates monotonic relationship, where variables increase or decrease together without necessarily forming a straight line.

5. Comparative Study of Parametric and Non-Parametric Tests

Basis of Comparison	**Parametric Tests**	**Non-Parametric Tests**
Population Distribution	Assumes to follow a specific probability distribution most commonly the **normal distribution**	Do not assume a specific population distribution
Measurement Scale	Interval or Ratio scale	Nominal or Ordinal scale
Sample Size	Moderate to large preferred	Suitable for both small and large samples
Statistical Power	Higher when their strict assumptions are hold	Lower compared to parametric tests
Sensitivity to Outliers	Highly sensitive to outliers because they rely on parameters like mean and standard deviation to estimate population statistics	Highly robust to outliers
Data Requirements	Strict assumptions and generally follow a normal distribution	Few assumptions and it must not require the estimation of specific population parameters, observations must be independent

Mathematical Complexity	More Complex and if the assumptions are violated, the mathematical integrity and p- value accuracy of the test degrade rapidly	Comparatively simple and distribution free mathematical frameworks and it lies in advanced combinatorics, handling tied data and calculating exact p-values.
Examples	Z-test, t-test, ANOVA, Pearson Correlation	Chi-Square, Mann–Whitney, Wilcoxon, Kruskal–Wallis, Spearman Correlation

6. Advantages and disadvantage of Parametric Tests

6.1 Advantages

1. It does not require much data that could be converted in some order or format of ranks.
2. It can calculate everything so easily. In short, you will be able to find software much quicker.
3. They give you real information regarding the population which is in terms of the confidence intervals as well as the parameters.
4. It can perform quite well when they have spread over and each group happens to be different.

6.2 Disadvantage

1. Not valid when it comes to small data sets.
2. The size of the sample is always very big, it makes a little difficult to carry out the whole test.
3. Typical tests will only be able to assess data that is continuous and the result will be affected by the outliers at the same time.
4. Performance decreases when assumptions are violated.

7. Advantages and Disadvantages of Non-Parametric Tests

7.1 Advantages

1. It can be used in such situation where the normality assumption is not true.
2. It can be used for all types of data, whether it is ‘qualitative, quantitative or ranked data.
3. Suitable for small samples.
4. Suitable for studies in behavioural sciences.
5. It has fewer and less rigid requirements and also applied for a wide variety of situations.

7.2 Disadvantages

1. N.P test are wasteful of time and data, if all the assumptions of a statistical models are satisfied.
2. To test statistical hypothesis only and not for estimating the parameters.
3. Ranking procedures may reduce information.
4. It is less efficient than their parametric counterparts.

References;

1. Gupta, S. C., & Kapoor, V. K. *Fundamentals of Mathematical Statistics*. Sultan Chand & Sons.
2. Gibbons, J. D., & Chakraborti, S. *Nonparametric Statistical Inference*. CRC Press.
3. Conover, W. J. *Practical Nonparametric Statistics*. John Wiley & Sons.
4. Siegel, S., & Castellan, N. J. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill.
5. Agresti, A. *An Introduction to Categorical Data Analysis*. John Wiley & Sons.
6. Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. *Probability and Statistics for Engineers and Scientists*. Pearson.

7. Montgomery, D. C. *Design and Analysis of Experiments*. John Wiley & Sons.Daniel, W. W., & Cross, C. L. *Biostatistics: A Foundation for Analysis in the Health Sciences*. Wiley.