DOI: https://doi.org/10.56815/IJMRR.V3I3.2024/184-192

ISSN: 2945-3135

Peer Reviewed -Referred-Open Access-Scholarly Journal

Pirani, S. (2024). *Simplifying statistical Decision Making: A Research Scholar's Guide to parametric and Non-Parametric Methods, International Journal of Multidisciplinary Research & Reviews, Vol 03, No. 03, pp. 184-192.*

INTERNATIONAL JOURNAL OF
MULTIDISCIPLINARY RESEARCH & REVIEWS

journal homepage: www.ijmrr.online/index.php/home

SIMPLIFYING STATISTICAL DECISION MAKING: A RESEARCH SCHOLAR'S GUIDE TO PARAMETRIC AND NON-PARAMETRIC METHODS

Sohil Altaf Pirani

Senior Research Fellow, Dr. Babasaheb Ambedkar Marathwada University Postgraduate Diploma in Statistics University of Mumbai, India

Submitted: July 30, 2024 **Revised: Aug 12, 2024** **Accepted: Aug 20, 2024**

How to Cite the Article: *Pirani, S. (2024). Simplifying statistical Decision Making: A Research Scholar's Guide to parametric and Non-Parametric Methods, International Journal of Multidisciplinary Research & Reviews, Vol 03, No. 03, pp. 184-192.*

Keywords

Parametric test, Non-parametric Test

Abstract

The basis of any research is proper application of Research tools and techniques. A statistical technique is the foundation of any research, the correct utilization of these techniques will foster results that are valid and useful for the world to ponder upon. Through several literatures it has been observed that researchers are not well-versed with the utilisation of parametric and non-parametric Techniques. The purpose of conducting this study is to ease the confusion or lack of knowledge among the researchers in application of various parametric and non-parametric technique. This paper is an attempt to simplify the statistical decision for future research scholar. This study is useful as it converts the complex theoretical concepts of parametric and non-parametric techniques into simplified summarized content. This will eventually lead to the research fraternity fostering effective solutions to the social issues.

1. Introduction

1.1 Definition of Parametric Vs Non-Parametric

DOI: https://doi.org/10.56815/IJMRR.V3I3.2024/184-192

ISSN: 2945-3135

Peer Reviewed -Referred-Open Access-Scholarly Journal

Pirani, S. (2024). *Simplifying statistical Decision Making: A Research Scholar's Guide to parametric and Non-Parametric Methods, International Journal of Multidisciplinary Research & Reviews, Vol 03, No. 03, pp. 184-192.*

A statistical test known as a parametric test is one that makes assumptions regarding the characteristics (such as mean and variance) of the population distribution that the sample is taken from. The data used in these tests are typically assumed to have known or estimated parameters and to follow a normal distribution. (McCrum-Gardner, E. 2008).

A statistical test that does not assume a particular distribution for the data is known as a non-parametric test. When the sample size is small or the data is ordinal, for example, these tests are utilised when the data does not match the assumptions needed for parametric testing. (Gibbons, J.D., & Chakraborti, S. 2014).

Table 1: Simplified Comparison of application of Parametric and Non-Parametric Test

	Parametric	**Non-Parametric**
Conclusions	Conclusions are applicable to population and thus can Generalise the Result.	Conclusions are limited to the sample and thus cannot Generalise the Result.
Distribution	Can apply parametric Test only if the data is Normally Distributed (Can examine with the help of **Kolmogrov-smirnov, Shapiro Wilcoxon or by computing Skewness and Kurtosis**)	Distribution free test (Can also apply in case of Positively skewed and Negatively Skewed data)
Homogeneity of variance	Can apply parametric Test only if there is Homogeneity of variances of various Groups (Can examine with the help of **Levene's Test, F-Test**)	Not Required
Sampling	Random Sampling is preferred	Probability or non-probability sampling method can be used.
Variables	At least one scale variable is needed (Interval or Ratio scale variable)	Nominal and Ordinal Data set is used to apply Non-

		parametric Test
Sample Size	Large Sample size is Required to apply parametric Test	Can apply in case of small sample size (n < 30)
Outliers	There must be no outliers in the data set	No assumption of No outliers to be Fulfilled.
Observation	Independent observation assumption must be fulfilled.	No assumption of independent observation to be Fulfilled.

2. Types of Parametric Test

Table 2: Comparison between Z-test and T-test

Z-Test	T test
Test the difference of mean of two groups, which may be: sample mean & population mean, two sample means, proportions of sample and population, two sample proportions etc.	Test the difference of mean of 2 groups, which may be sample mean & population mean, two sample means, repeated sample means, paired sample means etc.
Finite Population	***Infinite population***
Population mean & Variance is known with certainty	***Population Mean & Variance is Unknown.***
Large Sample and its variance is same as population variance	*Sample is small as compared to population size.*

So practically as the population mean and variance is always unknown, so T-test is applied in case of Z-test.

DOI: https://doi.org/10.56815/IJMRR.V3I3.2024/184-192

ISSN: 2945-3135

Peer Reviewed -Referred-Open Access-Scholarly Journal

Pirani, S. (2024). *Simplifying statistical Decision Making: A Research Scholar's Guide to parametric and Non-Parametric Methods, International Journal of Multidisciplinary Research & Reviews, Vol 03, No. 03, pp. 184-192.*

Table 3: Types of Parametric T-tests

Purpose	Type of t test	Formula
To test the Difference between Sample mean and population mean	One Sample t Test	$\frac{\lvert \bar{X} - \mu \rvert}{\frac{S}{\sqrt{n}}}$
To test the Difference between two independent sample means	Independent sample t test	$\frac{\frac{\lvert \overline{X_1} - \overline{X_2} \rvert}{S}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$
To test the Difference between pre and post event outcomes	Paired/Repeated Sample t test	$\frac{\bar{d}}{\frac{S}{\sqrt{n}}}$

2.1 ANOVA (Analysis of Variance)

"ANOVA is commonly used when comparing the means of multiple groups to identify any significant differences" (Lindman, 1993). The means of three or more groups can be compared using the ANOVA statistical method to see if there are any statistically significant differences between them. It compares the alternative hypothesis—that at least one group mean differs—with the null hypothesis, which states that all group means are equal. The homogeneity of variances and normal distribution of the data are presumptions made by ANOVA.

$$\frac{Mean\ sum\ of\ squares\ between\ samples}{mean\ sum\ of\ squares\ within\ samples}$$

DOI: https://doi.org/10.56815/IJMRR.V3I3.2024/184-192

ISSN: 2945-3135

Peer Reviewed -Referred-Open Access-Scholarly Journal

Pirani, S. (2024). *Simplifying statistical Decision Making: A Research Scholar's Guide to parametric and Non-Parametric Methods, International Journal of Multidisciplinary Research & Reviews, Vol 03, No. 03, pp. 184-192.*

Table 4: Family of ANOVA

Dependent Variable	*Independent variable*	*Test*	*Example of Application*
One Scale variable	One nominal variable	One way ANOVA	Effect of Education Background (nominal variable) on Income (scale variable)
One scale variable	Two nominal variables	Two-way ANOVA	Effect of Education Background (First nominal variable) and Level of experience (second nominal variable) on Income (scale variable)
One scale variable	One nominal variable and one scale variable	One-way ANCOVA	Effect of Education Background (nominal variable) and age in years (scale variable) on Income (scale variable)
Two scale Variable	One nominal variable	One-way MANOVA	Effect of Education Background (nominal variable) on the Performance score (first scale variable) and Income (second scale variable)
Two scale Variable	Two nominal variables	Two Way MANOVA	Effect of Education Background (First nominal variable) and Level of experience (second nominal variable) on the Performance score (first scale variable) and Income (second scale variable)
Two scale Variable	One nominal variable and one scale variable	One-way MANCOVA	Effect of Education Background (nominal variable) and age in years (scale variable) on the Performance score (first scale variable) and Income (second scale variable)

2.2 Karl Pearson Correlation

"The Pearson correlation coefficient is widely used to assess the degree of linear association between two continuous variables" (Rodgers and Nicewander, 1988). The linear link between two continuous

DOI: https://doi.org/10.56815/IJMRR.V3I3.2024/184-192

ISSN: 2945-3135

Peer Reviewed -Referred-Open Access-Scholarly Journal

Pirani, S. (2024). *Simplifying statistical Decision Making: A Research Scholar's Guide to parametric and Non-Parametric Methods, International Journal of Multidisciplinary Research & Reviews, Vol 03, No. 03, pp. 184-192.*

variables is measured by the Karl Pearson Correlation coefficient, which also indicates the direction of the association. A perfect positive linear relationship is represented by a coefficient of +1, a perfect negative linear relationship by a value of -1, and no linear relationship by a coefficient of 0.

$$\frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2} \times \sqrt{n\sum y^2 - (\sum y)^2}}$$

2.3 Regression

According to Kutner, Nachtsheim, and Neter (2004), "Regression analysis is essential for modelling and predicting the relationship between dependent and independent variables". A statistical method for determining the relationship between a dependent variable and one or more independent variables is regression analysis. Modelling the relationship in order to forecast or deduce causal linkages is the aim. Multiple regression uses two or more independent variables, whereas simple linear regression only uses one. The regression model is as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e$$

3. Types of Non-Parametric Test

Table 5: Types of Chi-Square Test

Purpose	Type of Chi-square Test	Formula
To examine association between Two categorical variables	Chi-square Test of independence of Attributes	$\frac{\sum(O-E)^2}{E}$
To examine significant difference in Expected probability distribution (Theory) and observed sample distribution.	Chi-square Test of Goodness of fit	$\frac{\sum(O-E)^2}{E}$
To examine proportion of categories are same in two or more population.	Chi-square Test of Homogeneity (One way Chi-square)	$\frac{\sum(O-E)^2}{E}$
To compare sample variance with population variance (Assumed population follows Normal Distribution)	Chi-square Test of inference about population variance (Parametric)	$\frac{(n-1)\times S^2}{\sigma^2}$

3.1 Krushkal-Wallis Test/H test

An approach to nonparametric variance analysis is the Kruskal-Wallis test. The analysis looks for differences between the median values of three or more independent samples. Given that they both rank the initial data values, the Mann-Whitney and Kruskal-Wallis tests can be considered equivalent. Stated differently, it compiles each instance of data from the samples and arranges them in ascending order. When two scores are equal, the result is determined by taking the average of the two ranks. (Nahm, F. S. 2016).

$$H = \frac{12}{N\,(N+1)}\ x\left(\frac{R_1^2}{n_1} + \frac{R_1^2}{n_2} + \frac{R_3^2}{n_3}\right) - 3\,(N+1)$$

3.2 Wilcoxon Sign Rank Test

The Wilcoxon rank sum test evaluates the differences in the rank sums after ranking all data points in order and calculating each sample's rank sum. The rank sums of two groups will be similar if their scores are similar; however, if one group's score is greater or lower than the other groups, the rank sums of the two groups will differ more. (Nahm, F. S. 2016).

$$Z = \frac{|R - mean|}{Standard\ Deviation}$$

$$mean = \frac{n\,(n+1)}{4}$$

$$Standard\ Deviation = \sqrt{\frac{n\,(n+1)(2n+1)}{24}}$$

3.3 Wilcoxon Rank Sum Test/Mann Whitney/ U test

"The Mann-Whitney U test is a non-parametric method for comparing two independent groups when data do not follow a normal distribution" (McKnight and Najab,2010). For comparing differences between two independent groups when the data does not match the assumptions of parametric tests, the Wilcoxon Rank Sum Test, also called the Mann-Whitney U Test, is a non-parametric test that is utilised. It does not assume a normal distribution of the data; instead, it evaluates whether one group tends to have larger or lower values than the other.

$$Z = \frac{|U - mean|}{Standard\ Deviation}$$

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

$$Where\ n_1 < n_2$$

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

$$Where\ n_1 > n_2$$

$$mean = \frac{n_1 + n_2}{2}$$

$$Standard\ deviation = \sqrt{\frac{n_1\ n_2\ \times (n_1 + n_2 + 1)}{12}}$$

3.4 Spearman Correlation:

"Spearman's rank correlation is a non-parametric measure that quantifies the association between two ranked variables" (Hauke and Kossowski, 2011). The strength and direction of the monotonic link between two variables are evaluated using the non-parametric Spearman correlation, which is a measure of rank correlation. It does not require a linear relationship or a normal distribution of the data, in contrast to the Pearson correlation. Spearman correlation is especially helpful in situations with non-linear connections or ordinal data.

$$\frac{1 - 6\sum d^2}{n\ (n^2 - 1)}$$

4. Conclusion

This paper effectively addresses the crucial requirement to make parametric and non-parametric statistical approaches easier for researchers to grasp and apply. The goal of this study is to increase the accuracy and dependability of research findings by illuminating the variations, presumptions, and suitable applications of these tests. Researchers will be able to make better decisions and produce higher quality studies if they have a better understanding of these statistical techniques. Ultimately, this will help generate solid research that can guide industry practices, policy, and societal advancements to put it simply, it will guide the researcher's productive ways to address social problems.

5. Authors Contribution

The writers affirm that they have no connections to, or engagement with, any group or body that provides financial or non-financial assistance for the topics or resources covered in this manuscript.

6. Conflict Of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

7. Plagiarism Policy

All authors declare that any kind of violation of plagiarism, copyright and ethical matters will taken care by all authors. Journal and editors are not liable for aforesaid matters.

8. Sources Of Funding

The authors received no financial aid to support for the research.

References

[1] Gibbons, J.D., & Chakraborti, S. (2014). Nonparametric Statistical Inference. Boca Raton: CRC Press.

[2] Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. Quaestiones Geographicae, 30(2), 87-93.

[3] Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). Applied Linear Regression Models (4th ed.). McGraw-Hill Irwin.

[4] Lindman, H. R. (1992). Analysis of Variance in Complex Experimental Designs. Springer.McKnight, P. E., & Najab, J. (2010). Mann-Whitney U Test. The Corsini Encyclopedia of Psychology, 1-1.

[5] McCrum-Gardner, E. (2008). Which is the correct statistical test to use? British Journal of Oral and Maxillofacial Surgery, 46(1), 38-41.

[6] Nahm, F. S. (2016). Nonparametric statistical tests for the continuous data: The basic concept and the practical use. Korean Journal of Anesthesiology, 69(1), 8-14. https://doi.org/10.4097/kjae.2016.69.1.8

[7] Pirani, S. (2024). Navigating Research Ethics: Strategies for preventing and Addressing Research Misconduct, International Journal of Multidisciplinary Research & Reviews, Vol 03, No. 02, PP.96-104.

[8] Pirani, S. (2024). Navigating the complexity of sample size determination for Robust and Reliable Results, International Journal of Multidisciplinary Research & Reviews, Vol 03, No. 02, PP.73-86.

[9] Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. The American Statistician, 42(1), 59-66.