International Journal of Health Sciences
Available online at www.sciencescholar.us
E-ISSN: 2550-696X
Int. J. of Health Sci.
DOI: 10.53730

How to Cite:
Ilyas, M. (2026). The importance and efficacy of advanced parametric and non-parametric tests commonly used in health sciences research. *International Journal of Health Sciences*, *10*(S1), 207–231. https://doi.org/10.53730/ijhs.v10nS1.15956

The importance and efficacy of advanced parametric and non-parametric tests commonly used in health sciences research

Muhammad Ilyas
Lincoln University College, Petaling Jaya, Malaysia

***Abstract*---Background:** Advanced parametric and non-parametric statistical tests are important tools in modern health sciences research because they allow researchers to correctly and methodically assess complicated clinical and epidemiological data. The scientific foundation for interpreting healthcare outcomes, assessing the effectiveness of treatments, finding disease predictors, and promoting evidence-based medical practice is provided by these statistical techniques. The use of sophisticated statistical approaches has become essential for guaranteeing the validity, reliability, and reproducibility of research findings as the amount and complexity of biomedical data continue to rise. Parametric tests make the assumption that variables satisfy certain requirements, such as homogeneity of variance and interval or ratio-level measurement, and that data have a certain normal distribution. Parametric approaches are frequently used in clinical and biological research to compare therapy groups, investigate risk factors, assess illness outcome predictions, and track changes in patient health indicators over time. Non-parametric approaches are particularly useful in nursing, public health, psychology, and medical research where data may not be normally distributed since they do not necessitate rigid assumptions about population distribution. The application and interpretation of both parametric and non-parametric analyses have been greatly improved by contemporary statistical software like SPSS and SAS. Parametric tests include t-tests, Analysis of Variance (ANOVA) and Pearson correlation while non-parametric tests include Wilcoxon's

---

***Corresponding author**: Ilyas, M.; Email: ilyas.phdscholar@lincoln.edu.my*
Submitted: 09 March 2026, Revised: 18 April 2026, Accepted: 24 May 2026

Signed Rank test, Mann-Whitney U test, Kruskal-Wallis test, Spearman's Rank Correlation, and Chi-square test that play their key roles and highlighting their fundamental importance in health sciences research. **Objective:** To evaluate and assess the importance and efficacy of advanced statistical tests used in health sciences research. **Approach:** This review article systematically examines and synthesizes the existing literature on the importance and efficacy of advanced statistical tests, used in health sciences research. **Significance:** The use of advanced statistical tests in health sciences research is highly significant as they enhance the accuracy, reliability, and validity of research findings. These methods enable researchers to analyze complex and high-dimensional data, which are increasingly common in modern healthcare due to the growth of electronic health records, clinical trials, and large population-based studies. **Conclusion:** Advanced statistical tests are indispensable in health sciences research as they improve analytical precision, support informed decision making, and ultimately contribute to better health outcomes and more efficient healthcare systems.

***Keywords*---**T-tests, ANOVA, Pearson correlation, Wilcoxon's Signed Rank test, Mann-Whitney U test, Kruskal-Wallis test, Spearman's Rank Correlation, Chi-square test

Introduction

Statistical tests are key tools in health sciences research because they allow researchers to objectively analyze data, assess hypotheses, and reach evidence-based conclusions about diseases, treatments, interventions, and healthcare outcomes (Smeltzer M.P. & Ray M.A. 2022). Statistical tests are used in contemporary scientific and public health research to assess whether observed trends, relationships, or differences are likely the result of chance or actually reflect population impacts. Clinical trials, epidemiological studies, laboratory investigations, and efforts to improve healthcare quality are all supported by frequently used statistical tests (Guglielmetti, L.C. et al., 2022).

Descriptive statistics and inferential statistics are the two main types of statistics that are frequently used in health sciences research. Measures like mean, median, standard deviation, percentages, and frequency distributions are used in descriptive statistics to organize and summarize data (Thomas E., 2005). Conversely, inferential statistics employ statistical tests to draw inferences about populations from sample data as in hypothesis testing, where researchers look for meaningful correlations or differences between variables or groups (Bensken W.P. et al., 2021).

The Student's t-test, Chi-square test, Analysis of Variance (ANOVA), correlation analysis, regression analysis, and non-parametric tests like the Mann–Whitney U test and Kruskal–Wallis test are some of the most commonly used statistical tests in health sciences research. When comparing the means of two groups, such as blood pressure values between treatment and control groups, the Student's t-test is frequently used (Ranganathan P., 2021). When examining relationships between categorical variables, such as gender and disease prevalence, the Chi-square test is frequently utilized. While correlation and regression analyses assess the direction and strength of correlations between variables, ANOVA expands the comparison of means to more than two groups. When data do not fit the presumptions needed for parametric testing, like a normal distribution, non-parametric tests might be helpful (Chicco, D. et al., 2025).

The type of data gathered, study design, sample size, number of comparison groups, and distribution of the data all influence the choice of an acceptable statistical test (Ranganathan P., 2021). If normality assumptions are met, parametric tests are typically used to examine continuous variables; however, non-parametric alternatives are frequently needed for ordinal or non-normally distributed data. Choosing the right tests is crucial because improper statistical analysis can lead to false results and undermine the quality of research (Guglielmetti, L.C. et al., 2022).

P-values should not be the only factor used in statistical testing, as evidenced by the recent literature and it is becoming more and more common for researchers to publish clinical significance, effect sizes, and confidence intervals in addition to statistical significance which promotes biomedical research transparency and interpretation. Additionally, improvements in statistical tools like R, Stata, and SPSS have improved the accuracy and accessibility of statistical analysis for health researchers (Silva-Ayçaguer, L.C. et al., 2010).

In general, evidence-based practice in health sciences research is based on statistical tests as understanding and using these tests correctly enhances the validity, reliability, and scientific quality of study findings, which eventually improves patient outcomes and healthcare decision-making (Guglielmetti, L.C. et al., 2022). This article will help the researchers in concluding all the advanced statistical tests commonly used in health sciences research.

Statistical Tests Commonly used in Health Sciences Research

Common statistical tests in health sciences research include both parametric and nonparametric methods, chosen, based on the data type and distribution assumptions. With variations like the independent sample t-test and paired sample t-test, the Student's t-test is frequently used to compare means between two groups; however, it is not suitable for comparing more than two groups or multifactor designs and requires assumptions of normality, independence, and homogeneity of variance (Liang, G. et al., 2019 & Chicco, D. et al., 2025). Nonparametric tests, such as the Mann-Whitney U test and the Kruskal-Wallis test, provide alternatives that do not require a normal distribution when normality assumptions are failed (Nahm, F., 2016 & Chicco, D. et al., 2025).

Chi-square test is a vital nonparametric test used for analyzing categorical data because they can evaluate correlations or distributional differences across groups without supposing normality (Valarmathi S. et al., 2024). Before using statistical tests, it is crucial to test for normality, techniques like the Shapiro-Wilk and D'Agostino tests can be used to evaluate whether parametric tests are appropriate or whether nonparametric alternatives should be employed (Mishra, P. et al., 2019 & Kamath, A. et al., 2025).

Furthermore, as breaches might compromise the validity of t-tests and ANOVA, evaluating homogeneity of variance is crucial in clinical trials; specific tests like Cochran's or Jackknife can efficiently identify variance differences (Zhou, Y., 2023). In general, trustworthy findings in health sciences research are ensured by the appropriate selection and use of these statistical tests based on data characteristics.

PARAMETRIC TESTS

Parametric tests assume that the data are on a quantitative (numerical) scale, with a normal distribution of the underlying population. The samples have the same variance (homogeneity of variances). The samples are randomly drawn from the population, and the observations within a group are independent of each other. Commonly used parametric tests are the t-tests, Analysis of Variance (ANOVA) and Pearson correlation (Ali Z. & Bhaskar. B., 2016).

T-Tests (One-sample t-test, Independent (two-sample) t-test and Paired t-test)

T-tests are the most used statistical techniques for comparing group means in health sciences research. They are of three types; one-sample t-test compares a sample mean to a known population mean; the independent (two-sample) t-test compares means between two independent groups; and the paired (dependent) t-test is employed when there are repeated measures or matched observations (Skaik, Y., 2015, Jankowski, K. et al., 2018 & Liang, G. et al., 2019). T-tests make the assumptions that observations are independent, variances are homogeneous, and data are normally distributed; violating these presumptions can result in abuse and inaccurate conclusions (Schober, P., & Vetter, T., 2019, & Liang, G. et al., 2019).

One-sample t-test

One-sample t-test is a fundamental tool in health and medical research for determining if the mean of a sample deviates from a known or proposed value, such as a clinical norm or target level and it compares the mean of a single group to a known value or population mean (Jankowski, K. et al., 2018 & Liang, G. et al., 2019). The frequency of its use, as well as its presumptions, frequent abuses, and more recent extensions, are all highlighted in recent studies. Its typical applications in health sciences include:

Assessing whether student or patient scores e.g., GPA, clinical scales, deviate from an expected level (Al-Kassab, M., 2022), comparing patient measurements to published normative values e.g., heart rate variability vs. age norms (Hart, J., 2021), and comparing internal quality control (IQC) results to a predetermined laboratory target values (Gerald, B., 2018). According to Liang et al. (2019) and Al-Kassab (2022), this test is predicated on independent observations, roughly normal data, and a single factor design.

Common Assumptions for One-sample t-test

1). Measurement scale and variable type: Data must be continuous and measured using an interval or ratio scale (Kim, T., & Park, J., 2019 & Al-Kassab, M., 2022).
2). Random sampling and independence: Simple random sampling must be used to acquire observations, and they must be independent to one another (Kim, T., & Park, J., 2019 & Al-Kassab, M., 2022).

3). Normality of the population: For valid one sample t tests, it is frequently emphasized that the population (or the distribution of sample values) must be normally distributed (Kim, T., & Park, J., 2019 & Sharma, L., & Jha, S., 2023).
4). One sample t test is utilized when the population standard deviation (SD) is unknown, especially with small samples, the small to moderate "n" (Al-Kassab, M., 2022 & Sharma, L., & Jha, S., 2023).

Independent (two-sample) t-test

The independent (two-sample) t-test, assumes independence and equal or unequal variances depending on the test type, is used in health sciences research to compare means between two unrelated groups, such as treatment against control or male versus female (Manfei et al., 2017 & Gerald, B., 2018). There are two variations of the independent t-test based on whether or not equal variances may be assumed and when variances are different, Welch's t-test is frequently chosen (Kim, H., 2019 & Kelter, R., 2020). In randomized clinical trials and medical and psychological research, independent (two-sample) t-tests are one of the most popular methods *(Kelter, R., 2020 &* Chatzi, A., 2025*).*

Common Assumptions for Independent (two-sample) t-test

1). Samples must originate from two independent populations and there must not be any coupling or association among groups (Bourget, M., & Rakovski, C., 2025) and these observations should be made independently within each group using a straightforward random sampling technique (Zhou, Y. et al., 2023 & Thomas, M. et al., 2025).
2). This classical test is not appropriate for categorical data; instead, the outcome variable should be continuous or metric i.e. interval or ratio (Pradubsri, W., & Suphirat, C., 2024).
3). Distribution normality: The traditional Student two sample t test assumes that every group is selected from a normal distribution (Zhou, Y. et al., 2023 & Thomas, M. et al., 2025) and it is rather resilient for moderate or large samples, although skewed or heavy-tailed distributions may alter power and increase type-I error, particularly when heteroscedasticity is present (Medugu, P. et al., 2023).
4). The student's t test necessitates homoscedasticity, or equal population variances, between groups (*Zhou, Y. et al., 2023 &* Thomas, M. et al., 2025), recent research suggests using Welch's t test as a default when this is uncertain because equal variances are frequently implausible in practical and psychological/clinical data (*Delacre, M., 2019 &* Vankelecom, L. et al., 2024).

Paired t-test

Paired t-test is utilized to account for correlation between pairs when matching observations or repeated measurements are obtained from the same participants.

These tests are essential to the health sciences because many studies use matched designs or measure the same patients repeatedly, such as before and after therapy and employing the appropriate paired analysis boosts power, enhances accuracy, and prevents drawing false findings (Manfei et al., 2017 & Gerald, B., 2018). When every observation in one condition naturally corresponds to one in another e.g., the same patient before and after the intervention, or matched pairings based on age, etc., is an appropriate design for paired t-test and here the independent (unpaired) t-tests are ineffective and invalid (Liang, G. et al., 2019 & Albassam, M., & Aslam, M., 2021).

These tests focus on the mean change, which is essentially a one-sample t-test on the differences, determines whether the mean difference between paired measurements is zero (Ross, A. & Willson, V.L., 2017 & Albassam, M., & Aslam, M., 2021). Assumptions require roughly normal difference scores and proper pairing; when parametric assumptions, design type, or pairing structure are disregarded, misuse occurs (Liang, G. et al., 2019 & Albassam, M., & Aslam, M., 2021). Step-by-step SPSS guidelines and tutorials are globally accessible which represent that paired t-tests are commonly employed in medical and educational research, but they are also frequently misused (Vetter, T., & Mascha, E., 2017 & Liang, G. et al., 2019).

Common Assumptions for Paired t-test

Paired t-test is a parametric test, it depends on a number of statistical presumptions; if these are broken, the findings may be deceptive, its common assumptions are:

1). Each observation in one condition must be connected to precisely one observation in the other i.e. the same person, twin, matched by covariates, or recurrent measure, which is known as paired or dependent observations. A paired t-test is not acceptable when groups are independent; instead, an independent t-test or alternative techniques should be employed (Manfei et al., 2017 & Liang, G. et al., 2019).

2). Paired t-test is applicable to numerical e.g. interval or ratio variables and a single factor design, means single quantitative outcome (Liang, G. et al., 2019 & Chicco, D. et al., 2025).

3). In order for results to correctly generalize, pairs should be produced by a procedure that approximates simple random sampling or random assignment within matched pairs (Grabchak, M., 2022 & Chicco, D. et al., 2025).

4). Differences between paired observations, rather than the individual raw variables, should roughly follow a normal distribution (Kim, T., & Park, J., 2019 & Ghadhban, G., & Rasheed, H., 2021) and the typical paired t-test may lose validity and robustness when this is broken; robust or nonparametric alternatives, such as Wilcoxon signed rank, robustified, or bootstrap-based t tests, are then advised (Ghadhban, G., & Rasheed, H., 2021).

5). Pairs should be independent of one another; there should be no complex relationship or clustering and mixed effects or longitudinal models are recommended when data are clustered or involve repeated measures across time (Stellato, R. et al., 2025).

Analysis of Variance (ANOVA)

Analysis of variance (ANOVA) is the most popular statistical technique in health sciences research for comparing means across three or more groups, and it addresses the issue of inflated Type-I error that results from multiple t-tests. Using the F statistic to determine if group means differ significantly, it divides total variance into between-group and within-group components (Kim, T., 2017 & Schober, P., & Vetter, T., 2020). ANOVA examines differences in group means, just like t-tests but it provides greater versatility than a t-test, even though the researcher may use more than two groups in the analysis, additionally, ANOVA can take into account the effects of several independent variables and ascertain whether they interact or how one independent variable might influence another (Barkus, 2014).

ANOVA can be used in a variety of designs, such as one-way ANOVA for a single factor and two-way ANOVA for factorial designs that evaluate interaction effects across factors and it extends the two-sample t-test to several groups (Larson, M., 2007 & Schober, P., & Vetter, T., 2020). After a significant ANOVA, post hoc tests like Tukey's, Bonferroni, or Dunnett's are frequently employed to determine whether particular groups differ while accounting for multiple comparisons (Schober, P., & Vetter, T., 2020 & Varalakshmi, G., 2025). This approach is crucial for assessing treatment effects, accounting for confounding variables, and supporting evidence-based decision-making in the health sciences in clinical trials, epidemiological investigations, and experimental research (McCloskey, J., 1988, Kaufmann, J., & Schering, A., 2014 & Chatzi, A., & Doody, O., 2023).

A repeated measures ANOVA looks at whether the means of three or more groups are the same, much like an ANOVA and it is used when every variable in a sample is examined under different conditions or at different periods (Ali Z. & Bhaskar. B., 2016). By taking into consideration within-subject correlations and lowering error variance, repeated measures ANOVA provides more statistical power than one-way ANOVA, improving sensitivity to identify treatment effects in longitudinal or crossover investigations (Otaibi, A., 2024 & Strale, F., 2024). However, nonparametric options like the Friedman test may be more effective and suitable than parametric ANOVA when data are extremely skewed or defy normalcy assumptions (Otaibi, A., 2024).

Common Assumptions for ANOVA

ANOVA is based on certain data and design assumptions and compares mean outcomes across many groups, for frequentist one- and two-way ANOVA, the majority of recent reviews and implementations agree on the following assumptions:

1). Continuous outcome is that dependent variable on a ratio or interval scale at minimum (Adeniran, A. T. et al., 2021 & Bongbeebina, C., & Rahman, M., 2025).
2). Every observation is independent; unless repeated measures ANOVA is used, there is no inadvertent clustering or repeated measurements (Adeniran, A. T. et al., 2021 & Bongbeebina, C., & Rahman, M., 2025).
3). For every treatment or cell, the residuals i.e. the experimental errors, are distributed normally around zero (Marantika, A. et al., 2020 & Adeniran, A. T. et al., 2021).

4). The variances within a group or cell are the same for all groups and situations i.e. homoscedasticity or equal variances (Marantika, A. et al., 2020 & Adeniran, A. T. et al., 2021).
5). Correct model form and additivity which means in planned trials, treatment and block effects are additive, while factor effects accumulate linearly (Adeniran, A. T. et al., 2021 & Ahmed, H., 2024).
6). Assigning treatments and blocks randomly i.e. randomization, ensures that errors have a common variance and are independent (Miller, R. et al., 2021 & Soto-Rodríguez, I. et al., 2024).

Pearson Correlation

In health sciences research, Pearson's correlation is a fundamental parametric tool for measuring linear connections between continuous health variables. On interval or ratio scales, it assesses the degree and direction of a linear relationship between two variables (Rohwer, D., 2022). Many meta-analyses are based on this key effect-size parameter. The straightforward formula $(1-r^2)/(N-3)$, which is accurate for $N \geq 40$ and helpful in meta-analytic weighting, is suggested by recent work that improves standard error estimation (Gnambs, T., 2023). Adequate significance testing is provided by parametric tests for r e.g. t test, Fisher z, and software like SPSS (Suresh, P., & Raju, K., 2022).

In health research, Pearson's correlation is frequently used for its advantages like:

i). Correlate novel “Quality of Life (QoL) scales with World Health Organization Quality of Life-BREF (WHOQOL-BREF) domains to establish construct validity (Odhaib, S. et al., 2022).
ii). The evaluation scores' test-retest reliability i.e. reliability/agreement (Seo, J. et al., 2024).
iii). Connect health determinants to illness; intrinsic capability to functional decline i.e. clinical association (Sánchez-Sánchez, J. et al., 2024).
iv). Link performance and satisfaction to logistics or Electronic Medical Record (EMR) perceptions i.e. health services and systems (Taupiq, B. et al., 2025).
v). When combined with permutation testing, Pearson is still the most effective or nearly the most effective technique for identifying linear connections, even in the presence of noise (Karch, J. et al., 2024).
vi). Compared to a number of more recent metrics, Pearson proved more resilient to measurement noise in noisy drug-screening data (Smirnov, P. et al., 2021).
vii). Used in time-series association with modified parametric significance approaches, brain connection networks, factor analysis, and reliability estimates (Metsämuuronen, J., 2022).

Common Assumptions for Pearson Correlation

1). For appropriate measurement and data, both variables must be continuous i.e. ratio or interval (Schober, P.et al., 2018).
2). Observation pairs (x,y) should be independent of one another; special techniques are required for repeated measurements (Tabatabai, M. et al., 2021).
3). Both variables are taken to be jointly normally distributed i.e. bivariate normal in the population for the purposes of traditional t-tests and confidence intervals (Schober, P.et al., 2018 & Tabatabai, M. et al., 2021).

4). Two useful tests of bivariate normality are that each variable is roughly normally distributed and that the scatterplot's linear connection indicates a violation (Schober, P.et al., 2018 & Tabatabai, M. et al., 2021).
5). Pearson's "r" measures the linear component of connection; nonlinear associations may yield a deceptively low "r," and scatterplots are strongly advised before utilizing "r" (Schober, P.et al., 2018 & Tabatabai, M. et al., 2021).
6). Lack of significant outliers i.e. when outliers or heavy tails are present, robust or rank-based correlations such as Spearman, Kendall, Winsorized, frequently perform better and relevant or extreme outliers can significantly skew "r", particularly with small samples (Schober, P.et al., 2018).

NON-PARAMETRIC TESTS

Distributed parametric tests can produce inaccurate results when the sample means are not normal and the assumptions of normality are not satisfied. Because they do not require the normalcy assumption, non-parametric tests, also known as distribution-free tests, are employed in such circumstances. When compared to parametric tests, non-parametric tests might not be able to identify a significant difference means, they typically have less power (Ali Z. & Bhaskar. B., 2016). The most popular nonparametric tests include Wilcoxon's Signed Rank Test, Mann-Whitney U Test, Kruskal-Wallis Test, Spearman's Rank Correlation, and Chi-square Test. These tests may be more effective when normality is broken, particularly with skewed distributions or small samples *(Bridge, P., & Sawilowsky, S., 1999).*

Wilcoxon's Signed Rank (WSR) Test

Wilcoxon's signed-rank test is a fundamental nonparametric technique in health research for comparing paired or pre-post measures when normality is in doubt is. Recent work extends it to complicated designs, improves its application and timing, and integrates it into contemporary health-data situations. It's importance and key roles in health sciences research are:

i). When data are ordinal, skewed, or parametric assumptions are questionable, this alternative to the paired t test is utilized; the power loss in comparison to the t test is typically negligible and may even be reversed in some circumstances (Rosenblatt, J., & Benjamini, Y., 2018 & Garren, S., & Davenport, G., 2022).
ii). Classic applications include matched-pair health data and pre-post interventions, such as blood pressure before and after therapy (Divine, G. et al., 2013).
iii). Wilcoxon signed-rank test can be more potent than the t test under mixture alternatives where just a subgroup reacts, which is important for medical imaging and tailored medication (Rosenblatt, J., & Benjamini, Y., 2018).
iv). This test improves type-I error control and power in ophthalmologic and nutritional research by accounting for clustering for example, both eyes in ophthalmology, repeated measures within patient (Rosner, B. et al., 2006).
v). A grouped sequential version uses variance formulae and sample size planning to provide one-sample/paired ordinal outcomes in stroke and other trials (Zhang, B., & Wu, Y., 2025).

vi). For small-sample clinical trials when asymptotic approximations are insufficient, Stata's exact Wilcoxon techniques are prioritized (Harris, T., & Hardin, J., 2013).

vii). Kernel-based transformations extend use to asymmetric data while maintaining excellent operational characteristics; redesigned tests explicitly reflect zero or tied differences and can be more effective when ties or zeros are common (Bagkavos, D., & Patil, P., 2021).

Common Assumptions for Wilcoxon's Signed Rank (WSR) Test

Wilcoxon Signed Rank (WSR) test has significant underlying assumptions that impact validity and interpretation, as recent studies make clear, which are:

1). Data must be matched pairs for example, pre and post, or a single sample compared to a constant for paired or one-sample differences to be analyzed (Imam, A. et al., 2014).

2). Standard WSR assumes that changes across subjects are independent; specific clustered versions are needed for clustered data e.g. two eyes per person (Rosner, B. et al., 2006).

3). The traditional WSR assumes that the distribution of differences is symmetric around the median, which is frequently zero. When symmetry is broken, power and interpretability are primarily impacted, and the test may become "not good and applicable" in cases of severe asymmetry (Tapio, R., 2025).

4). Standard theory makes the assumption that perfect zero differences and ties are uncommon; zeros are usually eliminated and ties are treated cautiously, however, its modified versions ease this assumption and specifically include zero and tied differences (Imam, A. et al., 2014).

Mann-Whitney U Test

Mann-Whitney U test also known as Wilcoxon rank-sum, is a fundamental nonparametric technique for comparing two independent groups when data are skewed, ordinal, or sample sizes are limited. Its widespread clinical application as well as a number of methodological extensions are demonstrated by recent work in health sciences research is as under:

i). When the two-sample t test's normality or equal-variance assumptions are questioned, particularly for skewed or ordinal outcomes in clinical investigations, a nonparametric substitute for the t test is employed (Park, Y., 2025).

ii). Frequently chosen since many medical variables such as risk indices, scores, and biomarkers are ordinal or non-normal; the test focuses on variations in distributions or medians and the likelihood that one group produces higher values than other (Kelter, R., 2020).

iii). Two-stage group-sequential trial designs for ordinal outcomes and hierarchical composite endpoints (DOOR) in clinical studies, therapy comparison, and infectious disease trials (Park, Y., 2025).

iv). Assessing the disease risk e.g. triglyceride-glucose index (TyG index) vs metabolic syndrome components and genetic risk scores for malaria among populations (Tai, K. et al., 2022).

v). Validation of the Patient-reported outcome measures (PROMs) and other questionnaire e.g. the new Obstructive sleep apnea quality of life (OSA QoL) instrument (SLEAP) distinguishes between patients and controls;

psychiatric measures based on age, sex, and diagnosis (Wang, J. et al., 2022).

vi). Observational clinical studies e.g. colorectal cancer care before and during COVID-19, emergency versus non-emergency cancer diagnoses (Redha, H. et al., 2024).

Common Assumptions for Mann-Whitney U Test

Current literature clarifies the following assumptions for Mann-Whitney U Test:

1). Independence is a basic assumption of Mann-Whitney U Test therefore each group's observations must be mutually independent and random samples (Kitchen, C., 2009 & Ye, J., & Lai, D., 2022).

2). Data must be at least ordinal (able to be ranked) or continuous i.e. Suitable outcome scale be used (Kitchen, C., 2009 & Park, Y., 2025).

3). Since classical theory presumes that results originate from continuous distributions, precise links are uncommon. Mid-ranks or other tie-handling methods are necessary when there are several ties, discrete or ordinal scores, but they alter the null distribution and may have an impact on type-I error and power (Kitchen, C., 2009 & McGee, M., 2018).

4). The test can be interpreted as a test of equal distributions under equal-shape distributions, or as a test that the chance of a random value from group 1 exceeding one from group 2 is 0.5 i.e. Mann–Whitney parameter (Kitchen, C., 2009).

5). It can be deceptive to use traditional power/sample-size formulas that imply simple location shift if forms change significantly e.g. J- or U-shaped bounded scales (Lesaffre, E. et al., 1993).

Kruskal-Wallis Test

Kruskal–Wallis (KW) test is a crucial nonparametric substitute for one-way ANOVA when comparing three or more independent groups and when normality or equal variance assumptions are questioned. Its usefulness for skewed, ordinal, or complex data is highlighted by recent methodological and applied work in health sciences research. Its key roles and advantages are as under:

i). Kruskal–Wallis test is particularly useful for ordinal or non-normally distributed continuous outcomes since it uses ranks to compare distributions and medians across at least three independent groups (Chicco, D. et al., 2025).

ii). It is frequently suggested as the preferred test in biostatistics when ANOVA assumptions are not met, group sizes are different, and researchers are more interested in medians than means (Kitchen, C., 2009 & Chicco, D. et al., 2025).

iii). KW and other rank tests maintain excellent efficiency with relatively slight power loss under perfect normalcy in biomedical contexts, where data are frequently skewed and samples are small however, they achieve significant power improvements when distributions are heavily skewed or heavy-tailed (Kitchen, C., 2009).

iv). KW power techniques that use Bernstein fits perform well on medical datasets such as dialysis blood pressure, diabetic sleep hours, HDL versus marital status and offer more accurate sample size forecasts than ANOVA, even when ANOVA assumptions exist (Clark, J. et al., 2023).

v). Comparative research reveals that KW has asymptotic relative effectiveness ~0.955 vs. ANOVA under normality, but it is significantly more potent with skewed or heavy-tailed biomedical data, giving it a reliable default when assumptions are questionable (Kitchen, C., 2009).
vi). Statistical Analysis System (SAS) macro provides practitioners with nonparametric pairwise group comparisons up to 20 groups by implementing numerous comparison procedures after a significant KW test (Elliott, A., & Hynan, L., 2011).
vii). KW test expands rank approaches to correct censored outcomes with potentially unequal censoring distributions are directly inspired by medical follow-up research (Breslow, N., 1970).
viii). When dealing with interval-valued, indeterminate data such as ICU occupancy by age in COVID-19, a standard neutrosophic KW test produces interval-valued results that are more appropriate for ambiguous or uncertain biological measures (Sherwani, R. et al., 2021).
ix). A generalized KW with group uncertainty, maintains robustness and power while incorporating probabilistic group membership, especially for genotype uncertainty in genetic association studies (Acar, E., & Sun, L., 2012).

Common Assumptions for Kruskal-Wallis Test

Recent literature shows the following assumptions for Kruskal-Wallis test:

1). Using rankings rather than raw data, the original formulation determines whether several independent samples originate from the same underlying distribution i.e. identical cumulative distribution functions (Kruskal, W., & Wallis, W., 1952).
2). It is essentially a test of stochastic homogeneity, according to contemporary explanations, as every group has the same distribution rather than just the same mean or median (Vargha, A., & Delaney, H., 1998 & Kroeger, C. et al., 2021).
3). Independent findings both within and between the groups (Vargha, A., & Delaney, H., 1998 & Kroeger, C. et al., 2021).
4). A randomly generated sampling from each population (Kruskal, W., 1952 & Ostertagová, E. et al., 2014).
5). Data must be at least ordinal i.e. the values must be able to be ranked (Ostertagová, E. et al., 2014).
6). If it is to be taken as a test of central tendency (medians), the distributional shape across groups must be same; otherwise, it examines more general distributional disparities (Vargha, A., & Delaney, H., 1998 & Kroeger, C. et al., 2021).
7). The original theory assumes that distributions are continuous; ties are permitted but call for standard corrections (Kruskal, W., 1952 & Ostertagová, E. et al., 2014).
8). It is troublesome to use Kruskal-Wallis as a means or medians test when heteroscedasticity i.e. unequal variances are present, particularly when sample sizes are uneven, simulations exhibit increased type I error and false conclusions (Kroeger, C. et al., 2021).
9). Even when central tendencies are identical, variance heterogeneity alone can produce significance since Kruskal–Wallis is sensitive to any distributional difference, including variance (Kroeger, C. et al., 2021).

10). For interval-valued, undetermined data, a neutrosophic Kruskal–Wallis generalization adds the following presumptions: samples are random and mutually independent, ties should not be significantly concentrated in one area, and data has quantified uncertainty (Sherwani, R. et al., 2021).

Spearman’s Rank Correlation

Spearman's rank correlation is a non-parametric statistical test used to measure monotonic relationships between two variables, particularly in cases when the data are skewed, ordinal, or contain outliers. Spearman's approach has been extended in recent health sciences research to handle complex data structures, censoring, clustering, and covariate correction, its common applications in health sciences research are:

i). Spearman is the recommended test for ordinal data, nonlinear but monotonic relationships, abnormal continuous data, and data containing pertinent outliers (Schober, P. et al., 2018 & Alsaqr, A., 2021).
ii). It gauges how steadily one variable alters in tandem with the other, though not always in a straight line (Schober, P. et al., 2018).
iii). When choosing between Pearson and Spearman in health sciences research, factors including data type, linearity, outliers, and parametric assumptions should be taken into consideration (Schober, P. et al., 2018 & Alsaqr, A., 2021).
iv). When variables can only be sorted, such as severity scores or Likert scales, rank-based coefficients are helpful since they are distribution-free (Nikitina, M., & Chernukha, I., 2023).
v). Used in medical communities to evaluate correlations in biomarker research, cancer cell proliferation, ophthalmology outcomes, pain and health ratings, and conference/Twitter metrics (Schober, P. et al., 2018 & Alsaqr, A., 2021).
vi). Extensively used for large diagnostic panels with skewed, heterogeneous distributions that are restricted by detection limitations (Liu, Q. et al., 2018).
vii). Spearman is used in clinical laboratory medicine to assess the correlation between measurements made by various devices; new variations are employed to deal with imprecise i.e. interval or fuzzy data (Aslam. M., 2021).
viii). New estimators for Spearman's correlation in limited survival time regions e.g. time to viral failure vs. regimen change in HIV, using right censored survival data (Eden, S. et al., 2021).
ix). Definitions and estimators for Spearman correlations across and within clusters in clustered or longitudinal data, connecting them to covariate adjusted partial Spearman measures (Tu, S. et al., 2024).
x). Applicable to both continuous and ordinal health variables, general definitions of partial and conditional Spearman employing probability scale residuals (Liu, Q. et al., 2018).
xi). A neutrosophic Spearman test enhances the management of imprecise clinical lab values for indeterminate or interval data (Aslam. M., 2021).
xii). Correct interpretation, reporting of normality, r, p values, and shared variance, and refraining from misusing correlation as agreement are all emphasized in methodological evaluations (Schober, P. et al., 2018 & Alsaqr, A., 2021).

Common Assumptions for Spearman’s Rank Correlation Test

Recent literature shows the following assumptions for Spearman’s Rank Correlation Test:

1). This test evaluates the X–Y relationship is described by a monotonic function i.e. consistently increasing or decreasing, which is not always linear (Alsaqr, A., 2021 & Bocianowski, J. et al., 2024).

2). This test is frequently chosen when Pearson's assumptions are not valid or the data is ordinal or significantly skewed and it does not presume normality or any particular frequency distribution (Liu, Q. et al., 2018 & Bocianowski, J. et al., 2024).

3). This test is described as the Pearson correlation between X and Y rankings (Liu, Q. et al., 2018 & Bocianowski, J. et al., 2024).

4). In this test one value for X and one for Y i.e. bivariate data are assigned to each subject or experimental unit (Liu, Q. et al., 2018 & Bocianowski, J. et al., 2024).

5). Pairs are taken to be independent; clustered or repeated-measures data require specific techniques (Tu, S. et al., 2024).

6). In order for ranks to have meaning, variables must be in ordinal scale (Liu, Q. et al., 2018 & Bocianowski, J. et al., 2024).

7). Simulation work reveals that Spearman's ρ performs similarly to Pearson test under various distributions for tests of no monotonic relationship; however, inference is best accurate when there are no ties as ties necessitate corrections or different CI approaches (Ruscio, J., 2008 & Puth, M. et al., 2015).

8). Redefining Spearman correlations between and within clusters and modeling conditional distributions relax independence with clustered data; accurate model specification is then an additional assumption (Liu, Q. et al., 2018 & Tu, S. et al., 2024).

9). Nonparametric estimation of Spearman-type measures for right-censored survival data frequently limits the area where correlation is established and requires assumptions about the censoring method (Eden, S. et al., 2021).

Chi-square (x^2) Test

Chi-square (x^2) test is a basic nonparametric statistical method that is frequently used in health sciences research to examine relationships between categorical variables without assuming a normal distribution and it comprises three primary types: the Test of Homogeneity, which compares distributions of a categorical variable among various populations; the Test of Independence, which examines relationships between two categorical variables within a sample; and the Goodness-of-Fit test, which determines whether observed data fit an expected distribution (Valarmathi S. et al., 2024, Montaña, R. et al., 2024 & Dizaji, P. et al., 2026). This test is frequently used to assess risk factors, treatment outcomes, nursing interventions, and health behavior trends by examining contingency tables of frequency counts *(Rana, R., & Singhal, R., 2015, Schober, P., & Vetter, T., 2019 & Valarmathi S. et al., 2024).* Chi-square test is supplemented by effect size measures such as Phi and Cramér's V, which quantify the strength of connections in 2x2 or bigger tables (Dizaji, P. et al., 2026).

Additionally, researchers can construct studies with sufficient power to identify significant differences in categorical data by using sample size calculation tools specifically designed for Chi-square tests. (Rahman, H. et al., 2025). Chi-square test is still fundamental to testing hypotheses using categorical data in clinical and epidemiological research, and when used appropriately, it supports evidence-based decision-making.

Common Assumptions for Chi-square Test

According to the recent literature chi square tests are frequently employed for categorical data, their validity is dependent on a number of important assumptions which are frequently misinterpreted or broken, that results in wrong conclusions (McHugh, M., 2013 & Gurvich, V., & Naumova, M., 2025), they are:

1). Random sampling and proper design must be used to gather data, convenience sampling invalidate the inferential evidence (McHugh, M., 2013 & Gurvich, V., & Naumova, M., 2025).
2). Cell entries for categorical data must be frequencies i.e. counts, not percentages or converted values (McHugh, M., 2013 & Gurvich, V., & Naumova, M., 2025).
3). Only one category of each variable may be assigned to each subject (McHugh, M., 2013 & Gurvich, V., & Naumova, M., 2025).
4). No paired or repeated measures without additional techniques; each subject only contributes to one cell (McHugh, M., 2013 & Gurvich, V., & Naumova, M., 2025).
5). The sampling schemes and nulls for homogeneity, independence, and goodness of fit are distinctly different; mislabeling them causes misinterpretation (Franke, T. et al., 2012 & Turhan, N., 2020).
6). Expected cell counts ≥ 5 in almost all cells in general tables (McHugh, M., 2013 & Gurvich, V., & Naumova, M., 2025).
7). A relaxed rule for chi square test is that, no more than 20% of cells <5 and none <1 (Pandis, N., 2016 & Gurvich, V., & Naumova, M., 2025).
8). Fisher's exact or likelihood ratio should be used for small samples (McHugh, M., 2013 & Pandis, N., 2016).

According to simulation and review studies, these guidelines are practical rather than rigorously mathematical, although breaking them can increase Type-I or Type-II errors (Bolboacă, S. et al., 2011 & McHugh, M., 2013).

Following are the common misinterpretations and limitations about x^2 test:

Even when the test is determined correctly, confusion between independence and homogeneity results in incorrect verbal conclusions (Franke, T. et al., 2012 & Turhan, N., 2020).

Error rates can be distorted by using x^2 when expected counts are too low or when data is paired or clustered (Bolboacă, S. et al., 2011 & McHugh, M., 2013).

An effect size such as Cramer's V for the strength of association should be included with the chi-square (x^2) test (McHugh, M., 2013 & Pandis, N., 2016).

CHALLENGES AND LIMITATIONS

Advanced parametric and non-parametric statistical tests are essential to evidence-based health sciences research, but their effectiveness is hampered by a number of methodological, practical, and interpretational issues. as if assumptions are broken or statistical techniques are chosen incorrectly, these constraints may affect the validity, reliability, and generalizability of results.

Clinical data are frequently skewed, contain outliers, or have disproportionate variances because of biological variability in health sciences research. Assumptions like linearity, homogeneity of variance, independence of data, and normal distribution are the foundation of parametric tests and may yield skewed results, deceptive p-values, and incorrect conclusions when assumptions are broken.

Non-parametric tests typically have less statistical power particularly in small samples, lower power raises the possibility of missing real differences or treatment effects. This restriction may make it harder to find significant intervention outcomes or illness connections in clinical and epidemiological research.

In the health sciences, choosing the appropriate statistical technique is still a significant difficulty due to insufficient statistical expertise or an excessive dependence on software results, researchers often abuse complex tests. For instance, after noticing non-normality, researchers could instinctively transition to non-parametric tests without taking sample size, study goals, or the robustness of parametric approaches into account.

Results from advanced statistical tests could be challenging for researchers and physicians to understand. Clinical interpretation is less obvious when non-parametric tests compare rank distributions instead of means. Coefficients, interactions, and confidence intervals in sophisticated parametric models also need to be carefully interpreted.

For both parametric and non-parametric tests, sample size is a significant constraint while very large samples can yield statistically significant but clinically irrelevant results, small samples may not meet assumptions of normalcy and limit power. Additionally, large datasets make them more sensitive to insignificant departures from assumptions, which may lead to the needless application of non-parametric techniques.

For high-dimensional datasets, repeated measurements, correlated observations, and missing or censored data, traditional parametric and non-parametric tests may not be sufficient. This complexity frequently calls for sophisticated modeling techniques, such as multilevel models, survival analysis, or machine learning-assisted analytics.

Extreme observations are often found in health data due to biological heterogeneity, disease severity, or measurement errors. Because parametric tests are so susceptible to outliers, mean estimates and standard deviations may be distorted. Although non-parametric tests are more resilient to outliers, the conversion of data into ranks may cause them to lose precise quantitative information.

Numerous variables are frequently tested concurrently in health sciences investigations e.g. biomarkers, treatment objectives, the likelihood of a Type I error rises with repeated testing. Results may become unreliable or irreproducible if proper adjustments such as Bonferroni or false discovery rate modifications are not made.

Researchers who apply tests mechanically without comprehending assumptions or interpretation may become overly dependent on statistical software due to its widespread use. Validity may be jeopardized by improper settings, incorrect code, or software abuse. Although statistical software increases productivity, methodological knowledge cannot be replaced by it.

Reproducibility issues in medical literature are caused by inconsistent statistical reporting, inadequate assumption checking, selective analysis, and test abuse. Inappropriate statistical techniques may provide results that are not generalizable to different populations.

CONCLUSION

Advanced parametric and non-parametric statistical tests are essential tools in health sciences research because they allow researchers to rigorously analyze data, test hypotheses, assess treatment outcomes, and produce trustworthy scientific evidence for clinical and public health decision-making. While non-parametric tests offer flexible and reliable alternatives for skewed, ordinal, non-normal, or small-sample datasets commonly found in biomedical and clinical investigations, parametric tests are especially valuable for their efficiency, precision, and statistical power when assumptions like normality and homogeneity of variance are sufficiently satisfied. However, choosing the best test based on study objectives, research design, sample size, measurement scale, and data characteristics is crucial to the effectiveness of these statistical techniques. The validity and repeatability of results in health sciences research may be jeopardized by improper application, neglecting to evaluate assumptions, relying too heavily on software-generated outputs, and misinterpreting statistical significance. To ensure reliable and clinically significant results, statistical approaches should be used in conjunction with meticulous methodological design, assumption checking, and open reporting procedures.

REFERENCES

[1]. Acar, E., & Sun, L. (2012). A Generalized Kruskal–Wallis Test Incorporating Group Uncertainty with Application to Genetic Association Studies. *Biometrics, 69*. https://doi.org/10.1111/biom.12006.

[2]. Adeniran, A. T., Olilima, J. O., Akano, R. O., (2021). Analysis of Variance: The Fundamental Concepts and Application with R. *International Journal of Mathematics and Computer Research*. https://doi.org/10.47191/ijmcr/v9i10.04.

[3]. Ahmed, H. (2024). Inferential statistics for cardiothoracic surgeons: Part 3 - drawing valid conclusions from clinical data. *Indian Journal of Thoracic and Cardiovascular Surgery, 41*, 233 - 247. https://doi.org/10.1007/s12055-024-01867-7.

[4]. Albassam, M., & Aslam, M. (2021). Testing Internal Quality Control of Clinical Laboratory Data Using Paired t-Test under Uncertainty. *BioMed Research International, 2021*. https://doi.org/10.1155/2021/5527845.

[5]. Ali, Zulfiqar; Bhaskar, S Bala1. Basic statistical tools in research and data analysis. Indian Journal of Anaesthesia 60(9):p 662-669, September 2016. | DOI: 10.4103/0019-5049.190623

[6]. Al-Kassab, M. (2022). The Use of One Sample t-Test in the Real Data. *JOURNAL OF ADVANCES IN MATHEMATICS*. https://doi.org/10.24297/jam.v21i.9279.

[7]. Alsaqr, A. (2021). Remarks on the use of Pearson's and Spearman's correlation coefficients in assessing relationships in ophthalmic data. *African Vision and Eye Health*. https://doi.org/10.4102/aveh.v80i1.612.

[8]. Aslam, M. (2021). Clinical laboratory medicine measurements correlation analysis under uncertainty. *Annals of Clinical Biochemistry, 58*, 377 - 383. https://doi.org/10.1177/00045632211006453.

[9]. Bagkavos, D., & Patil, P. (2021). Improving the Wilcoxon signed rank test by a kernel smooth probability integral transformation. *Statistics & Probability Letters, 171*, 109026. https://doi.org/10.1016/j.spl.2020.109026.

[10]. Barkus, E. (2014). *Understanding and Using Advanced Statistics. January 2009*.

[11]. Bensken WP, Ho VP, Pieracci FM. Basic Introduction to Statistics in Medicine, Part 2: Comparing Data. *Surgical Infections*. 2021;22(6):597-603. doi:10.1089/sur.2020.430

[12]. Bocianowski, J., Wrońska-Pilarek, D., Krysztofiak-Kaniewska, A., Matusiak, K., & Wiatrowska, B. (2024). Comparison of Pearson's and Spearman's correlation coefficients for selected traits of Pinus sylvestris L. *Biometrical Letters, 61*, 115 - 135. https://doi.org/10.2478/bile-2024-0008.

[13]. Bolboacă, S., Jäntschi, L., Sestras, A., Sestras, R., & Pamfil, D. (2011). Pearson-Fisher Chi-Square Statistic Revisited. *Inf., 2*, 528-545. https://doi.org/10.3390/info2030528.

[14]. Bongbeebina, C., & Rahman, M. (2025). ASSESSING HOMOGENEITY: A COMPARATIVE STUDY FOR ROBUST STATISTICAL ANALYSIS. *Far East Journal of Mathematical Sciences (FJMS)*. https://doi.org/10.17654/0972087125009.

[15]. Bourget, M., & Rakovski, C. (2025). P-value or False Discovery Rate When Two-Sample t-Test is Employed Instead of Paired t-Test. *Journal of Student Research*. https://doi.org/10.47611/jsrhs.v14i1.8553.

[16]. Breslow, N. (1970). A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika, 57*, 579-594. https://doi.org/10.1093/biomet/57.3.579.

[17]. Bridge, P., & Sawilowsky, S. (1999). Increasing physicians' awareness of the impact of statistics on research outcomes: comparative power of the t-test and Wilcoxon Rank-Sum test in small samples applied research. *Journal of clinical epidemiology, 52 3*, 229-35. https://doi.org/10.1016/s0895-4356(98)00168-1.

[18]. Chatzi, A. (2025). Understanding the independent samples t test in nursing research. *British journal of nursing, 34 1*, 56-62 . https://doi.org/10.12968/bjon.2024.0133.

[19]. Chatzi, A., & Doody, O. (2023). The one-way ANOVA test explained. *Nurse researcher*. https://doi.org/10.7748/nr.2023.e1885.

[20]. Chicco, D., Sichenze, A., & Jurman, G. (2025). A simple guide to the use of Student's t-test, Mann-Whitney U test, Chi-squared test, and Kruskal-

Wallis test in biostatistics. *BioData Mining, 18*. https://doi.org/10.1186/s13040-025-00465-6.

[21]. Clark, J., Kulig, P., Podsiadło, K., Rydzewska, K., Arabski, K., Białecka, M., Safranow, K., & Ciechanowicz, A. (2023). Empirical investigations into Kruskal-Wallis power studies utilizing Bernstein fits, simulations and medical study datasets. *Scientific Reports, 13*. https://doi.org/10.1038/s41598-023-29308-2.

[22]. Delacre, M., Leys, C., Mora, Y., & Lakens, D. (2019). Taking Parametric Assumptions Seriously: Arguments for the Use of Welch's F-test instead of the Classical F-test in One-Way ANOVA. *International Review of Social Psychology*. https://doi.org/10.5334/irsp.198.

[23]. Divine, G., Norton, J., Hunt, R., & Dienemann, J. (2013). Statistical grand rounds: a review of analysis and sample size calculation considerations for Wilcoxon tests. *Anesthesia and analgesia, 117 3*, 699-710 . https://doi.org/10.1213/ane.0b013e31827f53d7.

[24]. Dizaji, P., Fico, A., & PhDc, R. (2026). Chi-square test applications. *Medical hypothesis, discovery & innovation in optometry*. https://doi.org/10.51329/mehdioptometry234.

[25]. Eden, S., Li, C., & Shepherd, B. (2021). Nonparametric estimation of Spearman's rank correlation with bivariate survival data. *Biometrics, 78*, 421 - 434.https://doi.org/10.1111/biom.13453.

[26]. Elliott A.C, Hynan L.S., (2011), A SAS® macro implementation of a multiple comparison post hoc test for a Kruskal–Wallis analysis, Computer Methods and Programs in Biomedicine, Volume 102, Issue 1, 2011, Pages 75-80, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2010.11.002.

[27]. Franke, T., Ho, T., & Christie, C. (2012). The Chi-Square Test. *American Journal of Evaluation, 33*, 448 - 458. https://doi.org/10.1177/1098214011426594.

[28]. Garren, S., & Davenport, G. (2022). Using Kurtosis for Selecting One-Sample T-Test or Wilcoxon Signed-Rank Test. *Current Journal of Applied Science and Technology*. https://doi.org/10.9734/cjast/2022/v41i1831737.

[29]. Gerald, B. (2018). A Brief Review of Independent, Dependent and One Sample t-test*, 4*, 50. https://doi.org/10.11648/j.ijamtp.20180402.13.

[30]. Ghadhban, G., & Rasheed, H. (2021). Robust Tests for the Mean Difference in Paired Data by Using Bootstrap Resampling Technique. *Ibn AL- Haitham Journal For Pure and Applied Sciences*. https://doi.org/10.30526/34.3.2680.

[31]. Grabchak, M. (2022). How Do We Perform a Paired t-Test When We Don't Know How to Pair? *The American Statistician, 77*, 127 - 133. https://doi.org/10.1080/00031305.2022.2115552.

[32]. Guglielmetti, L.C., Faber-Castell, F., Fink, L. *et al.* Statistics decrypted—a comprehensive review and smartphone-assisted five-step approach for good statistical practice. *Langenbecks Arch Surg* **407**, 529–540 (2022). https://doi.org/10.1007/s00423-021-02360-0

[33]. Gurvich, V., & Naumova, M. (2025). Critical issues with the Pearson's chi-square test. *Modern Mathematical Methods*. https://doi.org/10.64700/mmm.75.

[34]. Hart, J. (2021). Comparison of p-value results between one versus two sample t testing: A case study. *Journal of Medical Statistics and Informatics*. https://doi.org/10.7243/2053-7662-9-1.

[35]. Imam, A., Usman, M., & Chiawa, M. (2014). On Consistency and Limitation of paired t-test, Sign and Wilcoxon Sign Rank Test. *IOSR Journal of Mathematics, 10*, 01-06. https://doi.org/10.9790/5728-10140106.

[36]. Jankowski, K., Flannelly, K., & Flannelly, L. (2018). The t-test: An Influential Inferential Tool in Chaplaincy and Other Healthcare Research. *Journal of Health Care Chaplaincy, 24*, 30 - 39. https://doi.org/10.1080/08854726.2017.1335050.

[37]. Kamath, A., Poojari, S., & Varsha, K. (2025). Assessing the robustness of normality tests under varying skewness and kurtosis: a practical checklist for public health researchers. *BMC Medical Research Methodology, 25*. https://doi.org/10.1186/s12874-025-02641-y.

[38]. Karch, J., Perez-Alonso, A., & Bergsma, W. (2024). Beyond Pearson's Correlation: Modern Nonparametric Independence Tests for Psychological Research. *Multivariate Behavioral Research, 59*, 957 - 977. https://doi.org/10.1080/00273171.2024.2347960.

[39]. Kaufmann, J., & Schering, A. (2014). Analysis of Variance ANOVA.10-25. https://doi.org/10.1002/9781118445112.stat06938.

[40]. Kelter, R. (2020). Analysis of Bayesian posterior significance and effect size indices for the two-sample t-test to support reproducible medical research. *BMC Medical Research Methodology, 20*. https://doi.org/10.1186/s12874-020-00968-2.

[41]. Kim, H. (2019). Statistical notes for clinical researchers: the independent samples t-test. *Restorative Dentistry & Endodontics, 44*. https://doi.org/10.5395/rde.2019.44.e26.

[42]. Kim, T. (2017). Understanding one-way ANOVA using conceptual figures. *Korean Journal of Anesthesiology, 70*, 22 - 26. https://doi.org/10.4097/kjae.2017.70.1.22.

[43]. Kim, T., & Park, J. (2019). More about the basic assumptions of t-test: normality and sample size. *Korean Journal of Anesthesiology,72*, 331 - 335. https://doi.org/10.4097/kja.d.18.00292.

[44]. Kitchen, C. (2009). Nonparametric vs parametric tests of location in biomedical research. *American journal of ophthalmology, 147 4*, 571-2. https://doi.org/10.1016/j.ajo.2008.06.031.

[45]. Kroeger, C., Ejima, K., Hannon, B., Halliday, T., McComb, B., Téran-García, M., Dawson, J., King, D., Brown, A., & Allison, D. (2021). Persistent confusion in nutrition and obesity research about the validity of classic nonparametric tests in the presence of heteroscedasticity: evidence of the problem and valid alternatives. *The American journal of clinical nutrition*. https://doi.org/10.1093/ajcn/nqaa357.

[46]. Kruskal, W. (1952). A Nonparametric test for the Several Sample Problem. *Annals of Mathematical Statistics, 23*, 525-540. https://doi.org/10.1214/aoms/1177729332.

[47]. Kruskal, W., & Wallis, W. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association, 47*, 583-621. https://doi.org/10.1080/01621459.1952.10483441.

[48]. Larson, M. (2007). Analysis of Variance. *Circulation, 117*, 115-121. https://doi.org/10.1161/circulationaha.107.654335.

[49]. Lesaffre, E., Scheys, I., Fröhlich, J., & Bluhmki, E. (1993). Calculation of power and sample size with bounded outcome scores. *Statistics in medicine, 12 11*, 1063-78. https://doi.org/10.1002/sim.4780121106.

[50]. Liang, G., Fu, W., & Wang, K. (2019). Analysis of t-test misuses and SPSS operations in medical research papers. *Burns & Trauma, 7*. https://doi.org/10.1186/s41038-019-0170-3.

[51]. Liu, Q., Li, C., Wanga, V., & Shepherd, B. (2018). Covariate-adjusted Spearman's rank correlation with probability-scale residuals. *Biometrics, 74*. https://doi.org/10.1111/biom.12812.

[52]. Manfei., Fralick, D., Zheng, J., Wang, B., Tu, X., & Feng, C. (2017). The Differences and Similarities Between Two-Sample T-Test and Paired T-Test. *Shanghai Archives of Psychiatry, 29*, 184 - 188. https://doi.org/10.11919/j.issn.1002-0829.217070.

[53]. Marantika, A., Fithriani, I., & Nurrohmah, S. (2020). Estimating parameter in two-way analysis of variance when variance between cells is heterogeneous. *Journal of Physics: Conference Series, 1442*. https://doi.org/10.1088/1742-6596/1442/1/012043.

[54]. McCloskey, J. (1988). Analysis of Variance in Sports Injury Research. *The American Journal of Sports Medicine, 16*, S-63 - S-64. https://doi.org/10.1177/03635465880160s115.

[55]. McGee, M. (2018). Case for omitting tied observations in the two-sample t-test and the Wilcoxon-Mann-Whitney Test. *PLoS ONE, 13*.https://doi.org/10.1371/journal.pone.0200837.

[56]. McHugh, M. (2013). The Chi-square test of independence. *Biochemia Medica, 23*, 143 - 149. https://doi.org/10.11613/bm.2013.018.

[57]. Medugu, P., Pwalakino, C., Mutah, Y., & Gandada, D. (2023). An Empirical Comparison of Power of Two Independent Population Tests under Different Underlined Distributions. *Asian Journal of Probability and Statistics*. https://doi.org/10.9734/ajpas/2023/v24i1514.

[58]. Metsämuuronen, J. (2022). Directional nature of the product–moment correlation coefficient and some consequences. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.988660.

[59]. Miller, R., Acton, C., Fullerton, D., Maltby, J., & Campling, J. (2021). Analysis of Variance (Anova). *The SAGE Encyclopedia of Research Design*. https://doi.org/10.1016/b978-0-12-397025-1.00319-5.

[60]. Mishra, P., Pandey, C., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive Statistics and Normality Tests for Statistical Data. *Annals of Cardiac Anaesthesia, 22*, 67 - 72. https://doi.org/10.4103/aca.aca_157_18.

[61]. Montaña, R., Roco-Videla, Á., Nieves, A., & Flores, S. (2024). [Chi-square test of homogeneity in clinical studies: A tool for analyzing differences between treatments]. *Semergen, 51 1*, 102332. https://doi.org/10.1016/j.semerg.2024.102332.

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

[63]. Nikitina, M., & Chernukha, I. (2023). Nonparametric statistics. Part 3. Correlation coefficients. *Theory and practice of meat processing*. https://doi.org/10.21323/2414-438x-2023-8-3-237-251.

[64]. Odhaib, S., Amiri, F., Altemimi, M., Imran, H., Alidrisi, H., Mohammed, M., & Mansour, A. (2021). Development of the First Health-Related Quality of Life Questionnaires in Arabic for Women With Polycystic Ovary Syndrome (Part II): Dual-Center Validation of PCOSQoL-47 and PCOSQoL-42 Questionnaires. *Cureus, 13*. https://doi.org/10.7759/cureus.18060.

[65]. Ostertagová, E., Ostertag, O., & Kováč, J. (2014). Methodology and Application of the Kruskal-Wallis Test. *Applied Mechanics and Materials, 611*, 115 - 120. https://doi.org/10.4028/www.scientific.net/amm.611.115.

[66]. Otaibi, A. (2024). Parametric Test Efficiency One-Way Repeated Measures Analysis of Variance and its Nonparametric Alternative Friedman Test. *Journal of Umm Al-Qura University for Educational and Psychological Sciences*.https://doi.org/10.54940/ep52668865.

[67]. Pandis, N. (2016). The chi-square test. *American journal of orthodontics and dentofacial orthopedics: official publication of the American Association of Orthodontists, its constituent societies, and the American Board of Orthodontics, 150 5*, 898-899. https://doi.org/10.1016/j.ajodo.2016.08.009.

[68]. Park, Y. (2025). Optimal two-stage group sequential designs based on Mann-Whitney-Wilcoxon test. *PLOS ONE, 20*. https://doi.org/10.1371/journal.pone.0318211.

[69]. Pradubsri, W., & Suphirat, C. (2024). Comparison of Test Statistics for Mean Difference Testing Between Two Independent Populations. *International Journal of Analysis and Applications*. https://doi.org/10.28924/2291-8639-22-2024-4.

[70]. Puth, M., Neuhäuser, M., & Ruxton, G. (2015). Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. *Animal Behaviour, 102*, 77-84. https://doi.org/10.1016/j.anbehav.2015.01.010.

[71]. Rahman, H., Noraidi, ., Khalid, ., Mohamad-Adam, ., Zahari, N., & Tuming, N. (2025). Practical guide to calculate sample size for chi-square test in biomedical research. *BMC Medical Research Methodology, 25*. https://doi.org/10.1186/s12874-025-02584-4.

[72]. Rana, R., & Singhal, R. (2015). Chi-square test and its application in hypothesis testing. *Journal of the Practice of Cardiovascular Sciences, 1*, 69 - 71. https://doi.org/10.4103/2395-5414.157577.

[73]. Ranganathan P. An Introduction to Statistics: Choosing the Correct Statistical Test. Indian J Crit Care Med. (2021) May;25(Suppl 2): S184-S186. doi: 10.5005/jp-journals-10071-23815. PMID: 34345136; PMCID: PMC8327789.

[74]. Redha, H., Hatmi, K., Al-Ghaithi, S., Zeedy, K., & Alawi, A. (2024). Emergency admission preceding malignancy diagnosis: Insights from a study at a tertiary care hospital. *Journal of Family & Community Medicine, 31*, 295 - 304. https://doi.org/10.4103/jfcm.jfcm_93_24.

[75]. Rohwer, D. (2022). Interpreting Correlations. *Inquiry in Music Education*. https://doi.org/10.4324/9781003057703-14.

[76]. Rosner, B., Glynn, R., & Lee, M. (2006). The Wilcoxon Signed Rank Test for Paired Comparisons of Clustered Data. *Biometrics, 62*. https://doi.org/10.1111/j.1541-0420.2005.00389.x.

[77]. Rosenblatt, J., & Benjamini, Y. (2018). On Mixture Alternatives and Wilcoxon's Signed-Rank Test. *The American Statistician, 72*, 344 - 347. https://doi.org/10.1080/00031305.2017.1360795.

[78]. Ross, A., Willson, V.L. (2017). Paired Samples T-Test. In: Basic and Advanced Statistical Tests. SensePublishers, Rotterdam. https://doi.org/10.1007/978-94-6351-086-8_4.

[79]. Ruscio, J. (2008) "Constructing Confidence Intervals for Spearman's Rank

Correlation with Ordinal Data: A Simulation Study Comparing Analytic and Boots trap Methods," Journal of Modern Applied Statistical Methods: Vol. 7: Iss.2, Article 7. DOI:10.22237/jmasm/1225512360

[80]. Sánchez-Sánchez, J., Lu, W., Gallardo-Gómez, D., Del Pozo Cruz, B., De Souto Barreto, P., Lucia, A., & Valenzuela, P. (2024). Association of intrinsic capacity with functional decline and mortality in older adults: a systematic review and meta-analysis of longitudinal studies. *The lancet. Healthy longevity, 5 7*, e480-e492. https://doi.org/10.1016/s2666-7568(24)00092-8.

[81]. Seo, J., Choi, D., Kim, T., Cha, W., Kim, M., Yoo, H., Oh, N., Yi, Y., Lee, K., & Choi, E. (2024). Evaluation Framework of Large Language Models in Medical Documentation: Development and Usability Study. *Journal of Medical Internet Research, 26*. https://doi.org/10.2196/58329.

[82]. Schober, P., & Vetter, T. (2019). Chi-square Tests in Medical Research. *Anesthesia & Analgesia*. https://doi.org/10.1213/ane.0000000000004410.

[83]. Schober, P., & Vetter, T. (2019). Two-Sample Unpaired t Tests in Medical Research. *Anesthesia & Analgesia*. https://doi.org/10.1213/ane.0000000000004373.

[84]. Schober, P., & Vetter, T. (2020). Analysis of Variance in Medical Research. *Anesthesia & Analgesia*. https://doi.org/10.1213/ane.0000000000004839.

[85]. Schober, P., Boer, C., & Schwarte, L. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia, 126*, 1763–1768. https://doi.org/10.1213/ane.0000000000002864.

[86]. Sharma, L., & Jha, S. (2023). Applying Major Parametric Tests Using SPSS in Research. *International Research Journal of MMC*. https://doi.org/10.3126/irjmmc.v4i2.56017.

[87]. Sherwani, R., Shakeel, H., Awan, W., Faheem, M., & Aslam, M. (2021). Analysis of COVID-19 data using neutrosophic Kruskal Wallis H test. *BMC Medical Research Methodology, 21*. https://doi.org/10.1186/s12874-021-01410-x.

[88]. Silva-Ayçaguer, L.C., Suárez-Gil, P. & Fernández-Somoano, A. The null hypothesis significance test in health sciences research (1995-2006): statistical analysis and interpretation. *BMC Med Res Methodol* **10**, 44 (2010). https://doi.org/10.1186/1471-2288-10-44

[89]. Skaik, Y. (2015). The bread and butter of statistical analysis “t-test”: Uses and misuses. *Pakistan Journal of Medical Sciences, 31*, 1558 - 1559. https://doi.org/10.12669/pjms.316.8984.

[90]. Smeltzer MP, Ray MA. Statistical considerations for outcomes in clinical research: A review of common data types and methodology. *Experimental Biology and Medicine*. 2022;247(9):734-742. doi:10.1177/15353702221085710

[91]. Smirnov, P., Smith, I., Safikhani, Z., Ba-Alawi, W., Khodakarami, F., Lin, E., Yu, Y., Martin, S., Ortmann, J., Aittokallio, T., Hafner, M., & Haibe-Kains, B. (2021). Evaluation of statistical approaches for association testing in noisy drug screening data. *BMC Bioinformatics, 23*. https://doi.org/10.1186/s12859-022-04693-z.

[92]. Soto-Rodríguez, I., Nina, F., Gallardo, J., Villena, F., & Ruiz, R. (2024). Survey on the Compliance of the Assumptions of Variance Analysis of

Experimental Designs in Undergraduate Theses of Peruvian Universities. *Journal of Educational and Social Research*. https://doi.org/10.36941/jesr-2024-0043.

[93]. Stellato, R., Van Den Bor, R., Schipper, M., Lindeboom, M., & Eijkemans, M. (2025). Cohort data with dropout: a simulation study comparing five longitudinal analysis methods. *BMC Medical Research Methodology, 25*. https://doi.org/10.1186/s12874-025-02506-4.

[94]. Strale, F. (2024). Partitioning for Enhanced Statistical Power and Noise Reduction: Comparing One-Way and Repeated Measures Analysis of Variance (ANOVA). *Cureus, 16*. https://doi.org/10.7759/cureus.75322.

[95]. Suresh, P., & Raju, K. (2022). Study of Test for Significance of Pearson's Correlation Coefficient. *International Journal of Science and Research (IJSR)*. https://doi.org/10.21275/sr22915140002.

[96]. Tabatabai, M., Bailey, S., Bursac, Z., Tabatabai, H., Wilus, D., & Singh, K. (2021). An introduction to new robust linear and monotonic correlation coefficients. *BMC Bioinformatics, 22*. https://doi.org/10.1186/s12859-021-04098-4.

[97]. Tai, K., Dhaliwal, J., & Balasubramaniam, V. (2022). Leveraging Mann–Whitney U test on large-scale genetic variation data for analysing malaria genetic markers. *Malaria Journal, 21*. https://doi.org/10.1186/s12936-022-04104-x.

[98]. Tapio, R. (2025). The Role of Data Assumptions in Selecting Between Parametric and Nonparametric Tests. *Asian Journal of Probability and Statistics*. https://doi.org/10.9734/ajpas/2025/v27i11830.

[99]. Taupiq, B., Razak, A., Darmawansyah, D., Arifin, M., Rahman, A., Mallongi, A., & Nurhayani, N. (2025). Statistical Analysis of Logistics Management Impact on Medical Device Indicators in Indonesian Island Clinics. *International Journal of Statistics in Medical Research*. https://doi.org/10.6000/1929-6029.2025.14.32.

[100]. Thomas E. An introduction to medical statistics for health care professionals: basic statistical tests. Musculoskeletal Care. 2005;3(4):201-12. doi:10.1002/msc.11. PMID: 17042008.

[101]. Thomas, M., Dachepalli, S., Gutjahr, G., & Sankaran, R. (2025). Comparison of Two Independent Samples. *Amrita Journal of Medicine*. https://doi.org/10.4103/amjm.amjm_58_25.

[102]. Tu, S., Li, C., & Shepherd, B. (2024). Between- and Within-Cluster Spearman Rank Correlations. *Statistics in Medicine, 44*. https://doi.org/10.1002/sim.10326.

[103]. Turhan, N. (2020). Karl Pearsons chi-square tests. *Educational Research and Reviews*. https://doi.org/10.5897/err2019.3817.

[104]. Valarmathi S., Hemapriya A.S and Jasmine S. Sundar (2024); CHI-SQUARE TESTS: A QUICK GUIDE FOR HEALTH RESEARCHERS, *Int. J. of Adv. Res.*, 12 (10), 1214-1222, ISSN 2320-5407. DOI URL: https://dx.doi.org/10.21474/IJAR01/19746

[105]. Vankelecom, L., Loeys, T., & Moerkerke, B. (2024). How to Safely Reassess Variability and Adapt Sample Size? A Primer for the Independent Samples t Test. *Advances in Methods and Practices in Psychological Science, 7*. https://doi.org/10.1177/25152459231212128.

[106]. Varalakshmi, G. (2025). Application of Analysis of Variance (ANOVA) in Biological Science Research. *Journal of Experimental Agriculture*

International. https://doi.org/10.9734/jeai/2025/v47i93772.

[107]. Vetter, T., & Mascha, E. (2017). Unadjusted Bivariate Two-Group Comparisons: When Simpler is Better. *Anesthesia & Analgesia, 126*, 338–342. https://doi.org/10.1213/ane.0000000000002636.

[108]. Vargha, A., & Delaney, H. (1998). The Kruskal-Wallis Test and Stochastic Homogeneity. *Journal of Educational Statistics, 23*, 170 - 192. https://doi.org/10.3102/10769986023002170.

[109]. Wang, J., Zhu, E., Ai, P., Liu, J., Chen, Z., Wang, F., Chen, F., & Ai, Z. (2022). The potency of psychiatric questionnaires to distinguish major mental disorders in Chinese outpatients. *Frontiers in Psychiatry, 13*. https://doi.org/10.3389/fpsyt.2022.1091798.

[110]. Ye, J., & Lai, D. (2022). Empirical Power and Type I Error of Covariate Adjusted Nonparametric Methods. *Mathematics and Statistics*. https://doi.org/10.13189/ms.2022.100516.

[111]. Zhou, Y., Zhu, Y., & Wong, W. (2023). Statistical tests for homogeneity of variance for clinical trials and recommendations. *Contemporary Clinical Trials Communications, 33*. https://doi.org/10.1016/j.conctc.2023.101119.