

Classical and Modern Statistical Tests for Two-Sample Location Problems for Continuous Data: A Comprehensive Review

Israa Abdulameer Resen ¹ 

¹ Department of Mobile Communications and Computing Engineering, College of Engineering, University of Information Technology and Communications (UoITC), Baghdad, Iraq.

Email: israa.resen@uoitc.edu.iq

Abstract

The area of statistical analysis has been evolving steadily over time to cope with the complexity of comparing two samples, especially in location problems. This comprehensive review attempts to discuss both classical and modern statistical methods for comparing two samples and analyse their strengths and weaknesses. Classical methods have been successfully used to evaluate mean differences between two samples for a long period of time. However, the assumption of normality of the distribution of data and equality of variances makes them less usable in practice. Modern methods that do not make such assumptions should be considered when comparing two samples of data. The Mann-Whitney U test is one of the most popular among nonparametric statistics. Another example of modern statistics is the permutation method used to estimate the significance of the difference in the mean values between two groups. Advantages of these methods include the opportunity to analyse heterogeneous data or small samples, that are the reason for using nonparametric statistics. The use of advanced computational techniques and new software in the area of statistical testing is also worth discussing. Resampling methods can be used for generating more precise confidence intervals and p-values. Machine learning algorithms and other computational methods can be used in addition to classical statistical theory to improve the prediction of results. For example, the use of support vector machines in the area of two-sample comparison will increase the accuracy of results. Numerous studies address classical and modern approaches to comparing two samples; however, there is a lack of systematic comparison regarding the advantages and disadvantages of these methods. Most studies rely on the results of a single type of test to compare two data samples in real-world contexts, without considering whether the underlying assumptions are met. Furthermore, the evolution of statistical theory necessitates a discussion of new data types and their impact on the selection of statistical methods for comparison; indeed, comparing classical approaches with modern computational methods can contribute to the future advancement of statistical theory. Therefore, this review outlines the key frameworks that integrate classical and modern approaches for comparing two samples.

ARTICLE INFO

Review paper

Received: 13 June 2026

Accepted: 02 October 2026

Published: 05 October 2026

DOI: 10.58970/JSR.1246

CITATION

Resen, I. A. (2026). Classical and Modern Statistical Tests for Two-Sample Location Problems for Continuous Data: A Comprehensive Review, *Journal of Scientific Reports*, 15(1), 180-191.

COPYRIGHT

Copyright © 2026 by author(s)
Papers published by IJSAB
International are licensed
under a Creative Commons
Attribution-NonCommercial 4.0
International License.



Keywords: Statistical Tests, Location of Two Samples, Resampling Methods, Normality Assumption, Bootstrapping Methods, Permutation Methods, Bayesian Methods, Machine Learning.

Introduction

Statistical analysis is clearly an integral part of various fields that provide the tools necessary for data analysis. The need for these techniques is particularly evident when comparing the distribution locations (or means) of two statistical populations. The fundamental issue lies in determining whether there is a significant difference between the means of the two populations. Tests addressing this matter constitute a cornerstone of statistical testing. In addition, to having practical applications in fields such as medicine, quality control, and the social sciences (Adejare, 2025). Traditional statistical techniques, including the t-test and the Mann-Whitney U test, have long been established standards for assessing this type of difference. Although these methods differ in their underlying assumptions and parameters, they serve as effective tools in both exploratory and inferential research (Rogerson, 2020; Tybl, 2016; Navidi, 2008; Morgan, 2017). The t-test relies on assumptions of normal distribution and equal variances between samples. These requirements may not hold in real-world scenarios (Nesbitt et al., 2014). In contrast, non-parametric methods, such as the Mann-Whitney U test, are particularly valuable because they do not require prior assumptions regarding the distribution. This makes them a popular and suitable choice for handling skewed data or data that do not follow a normal distribution (Naimi et al., 2014; Risan et al., 2024; Adam, 2011; Mathews, 2004). In recent times, there has been considerable growth in the statistical tests associated with the two-sample location problem. As a result of the advances made in methodology as well as the increasing complexity of the datasets used across various research settings. There have been developments in modern statistical tests that involve sophisticated methods like bootstrapping and Bayesian tests, and these not only increase the flexibility of testing hypotheses but also offer additional information compared to traditional approaches (Caraffini & Iacca, 2020; Kim et al., 2019). Through bootstrapping methods, it is possible to get an estimate of the sampling distribution of the statistic through resampling and hence reach stronger conclusions even in cases of small samples or violated assumptions. Increased computing capabilities have also helped in the application of these sophisticated methods within statistics (Couch et al., 2019). Moreover, the availability of large datasets has necessitated the creation of methods which allow for more complex models. Thus, there have been some considerable advances in the development of multivariate testing techniques, beyond the simple pairwise analysis (Guerrier et al., 2019). With further advancement in this sphere, there is an absolute necessity to compare and critically evaluate various classical and modern approaches in use. The goal of the current review paper is to provide a literature synthesis and summarise the main advantages and limitations of each type of statistical method in question. Such a discussion will point to the way forward for further research and innovation in this sphere through the development of hybrid models, which would allow for combining classical and modern approaches (Batić et al., 2012). The further evolution in the field of statistics depends upon such innovations. It is also important to consider the implications that different test selection can have upon the conclusions of the research (Cao et al., 2024; Rahman et al., 2022). In contrast to other existing reviews, which separate classical and modern statistical tests into silos. This paper provides an innovative approach that helps to analyse and critique the shift from classical parametric/non-parametric testing methods to more advanced computational techniques such as resampling, Bayesian statistics, and machine learning classification. The paper additionally provides an evidence-based solution to one of the key problems in two-sample location cases, namely real-world data violations. In conclusion, it is important to note that connecting classical and modern methods of statistical testing makes the analysis of data much more understandable and sophisticated. It is worth mentioning that future research should pay particular attention to interdisciplinary collaboration in order to generate outstanding conclusions through the application of all these methods on a multidisciplinary database (Kelly, 2019). In general, the development of statistical testing in relation to two-sample location problems can be viewed as a transition from classical methods to modern methods, and therefore, there is a need for an objective evaluation of current approaches. This review aimed to provide such an evaluation as a basis for further discussions (Fu et al., 2025; Ajegbile et al., 2024; Ewusie et al., 2020; Guetterman, 2019; Zhang et al., 2025). In order to bridge this gap and formulate a logical framework, this review will test two hypotheses. The first one is (H_1) Modern

computational and robust statistical techniques outperform classical techniques with respect to accuracy and reliability while working on non-normal or non-homoscedastic data. The second one is (H_2) Combining the classical testing techniques with modern computational techniques. Modern techniques such as machine learning and Bayesian inference will lead to more adaptable methodologies for solving the two-sample location problem.

Fundamentals of the Two-Sample Location Problem

Two-sample location test theory has evolved through more than a century of developments in methodology. It began with the development of pioneering parametric methods, with Student's t-test developed by Gosset in 1908, establishing the basis for rigorous inference under strong assumptions of normality (Rogerson, 2020; Adejare, 2025). To deal with the rigidity of parametric models, non-parametric approaches were developed during the mid-century, notable among which are the development of the Wilcoxon rank-sum (1945) and Mann-Whitney U (1947) tests, which focused on distribution-free, rank-based methods (Caraffini & Iacca, 2020). In the latter part of the twentieth century, advances in computation brought about the introduction of powerful re-sampling methods like Efron's bootstrap (1979) and permutation approaches (Kim et al., 2019). In more recent times, modern statistical theory has made use of Bayesian probability models and machine learning approaches in two-sample tests (Guerrier et al., 2019; Saari, 2025). The discussion of the statistical tests in respect of two-sample location problems calls for a deep knowledge of the key terms in this case. The essence of the two-sample location problem lies in the assessment of the equality of the location parameters in terms of the populations from which two independent samples are taken. Location parameters usually mean the equality of their means or medians. It is one of the most crucial issues that is relevant for a great number of areas of science and engineering (Montgomery et al., 2007). These tests can help make decisions based on the comparison of data (Adejare, 2025). If the samples are believed to come from normal distributions, then the traditional approach is to use t-tests to make conclusions about the difference between the means (Rogerson, 2020). The reliability of t-tests depends on the homogeneity of variances and the normality of the distributions (Tybl, 2016). If the assumptions are not met, then the non-parametric alternative Mann-Whitney U test can be used instead. The importance of knowledge of sample sizes and power calculation is because it is a direct determinant of the reliability of the results of the test. A test should take into consideration both type 1 and type 2 errors, as shown in Figure 1, because they could lead to skewed results if neglected by the researchers (Nesbitt et al., 2014). Recent studies have pointed out the need for power calculation at the beginning stage of the experiment planning in order to be able to choose an appropriate sample size that will detect the difference between the groups, if any (Naimi et al., 2014). At the same time, the connection between the effect sizes and statistical power makes one understand the need to not only pay attention to the significance levels but also vary in the effect sizes (Caraffini & Iacca, 2020).

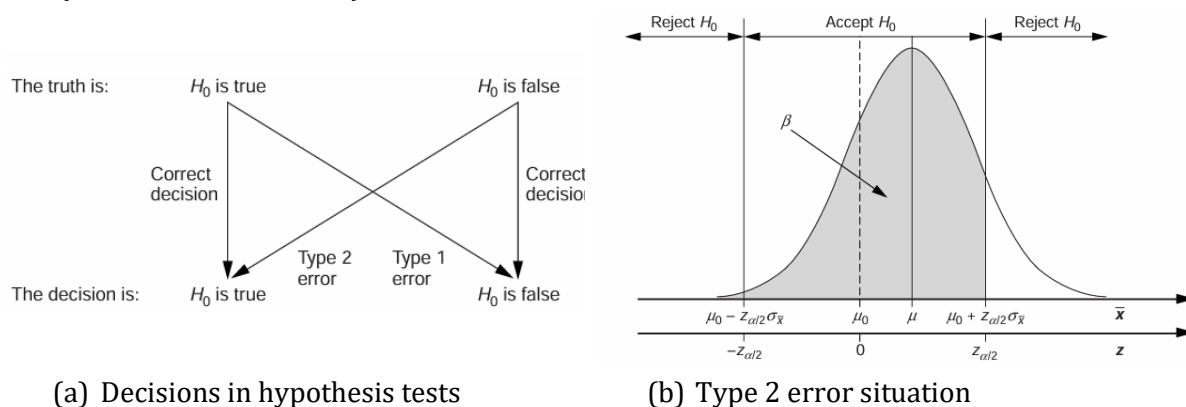


Figure 1: Type 1 and Type 2 Errors (Mathews, 2004)

Computational statistics improvements have also led to new approaches being applied when it comes to analysing two-sample problems. For example, the development of bootstrapping

methods has allowed researchers to apply non-parametric tests as alternatives to classical parametric procedures since the former approach involves resampling to compute the distribution of a certain statistic (Kim et al., 2019). The advantage of this method is obvious since, in cases where normality is not assumed, more types of data can be analysed (Couch et al., 2019). Besides that, the introduction of Bayesian methods in analysis has been one of the major advances since it enables the researcher to take into account prior information when conducting an analysis and gives a probabilistic interpretation of the outcomes. These approaches have proved to be successful when working with two-sample location problems (Guerrier et al., 2019; van Doorn et al., 2021). Classification of two-sample tests also brings about the crucial issue of parametric and non-parametric approaches to testing. Although the first approach remains more efficient in some respects on the grounds of normality assumptions, the second becomes necessary in situations where the assumptions are not met (Bati   et al., 2012). Awareness of the situations that require the application of one method or the other allows for making better decisions concerning the use of appropriate methodology (Cao et al., 2024). Research on the connections between various statistical tests also serves to emphasise the significance of developing methodologies that would be applicable in specific research situations (Rahman et al., 2022). Considering the constant changes that occur within statistical methodologies, there arises a necessity to carry out further studies regarding the two-sample location problem. With the advent of new types of data and ways of collecting them, future investigations should aim at constructing hybrid methodologies, combining classical and new statistical approaches and making them applicable and reliable (Kelly, 2019). Also, engaging interdisciplinary collaboration might help achieve a broader understanding of two-sample tests and ways they can be adapted under different conditions of use (Ewusie et al., 2020). Identifying benefits and drawbacks of current methodologies will become a valuable base for overcoming the problems that arise within the field of statistical inference (Fu et al., 2025). In sum, studying the two-sample location problem is important for revealing inconsistencies within scientific research and developing new techniques for data analysis (Fu et al., 2025; Ajegbile et al., 2024; Ewusie et al., 2020; Guetterman, 2019; Taylor et al., 2014).

Classical Statistical Tests

There are different methods used by researchers in statistics. But one of the most effective is the t-test. Its purpose is to analyse the means of two independent samples. The test was developed by Student at the beginning of the 20th century. It requires that both samples have a normal distribution in their populations. Thus, it is the standard method by which other approaches are compared (Adejare, 2025). Using the independent samples t-test, researchers are able to identify whether there are differences between the means of different populations. This is done based on assumptions related to distribution and variance homogeneity (Rogerson, 2020). Moreover, the paired samples t-test broadens the applicability of classical testing to the case of correlated observations which are connected through the principle of pairing as shown in Figure 2. This test relies on the differences between paired observations and makes use of the increased statistical power in the test by accounting for the within-subject variability (Tybl, 2016). However, some assumptions behind the t-test have prompted researchers to develop other tests which can work better in cases of non-normal distributions or heteroscedasticity. Welch's t-test is the solution to those cases, which provides the researcher with an opportunity to get more accurate estimates of the significance levels (Nesbitt et al., 2014). Its resistance to heteroscedasticity became the reason for its increased popularity, overcoming some limitations of the classical t-test (Naimi et al., 2014). As an alternative to t-tests, the Mann-Whitney U test provides another method which can be applied when there is no possibility of meeting the normality assumption, as shown in Table 1. This test is based on ranking all observations from both samples and finding out if there is any significant difference between the distributions of these samples (Caraffini & Iacca, 2020). Research has proved that it works effectively in the case of ordinal data and outliers, which distort the results of parametric tests (Kim et al., 2019). Further research confirmed that in the case of small sample sizes or skewed distributions, non-parametric methods prove their high level of competitiveness. By ensuring the relevance of the classical framework under various

circumstances (Couch et al., 2019). As an extension of the non-parametric framework to the case of two or more samples, the Kruskal-Wallis test helps in finding possible differences between several independent samples during the exploratory phase of research (Guerrier et al., 2019).

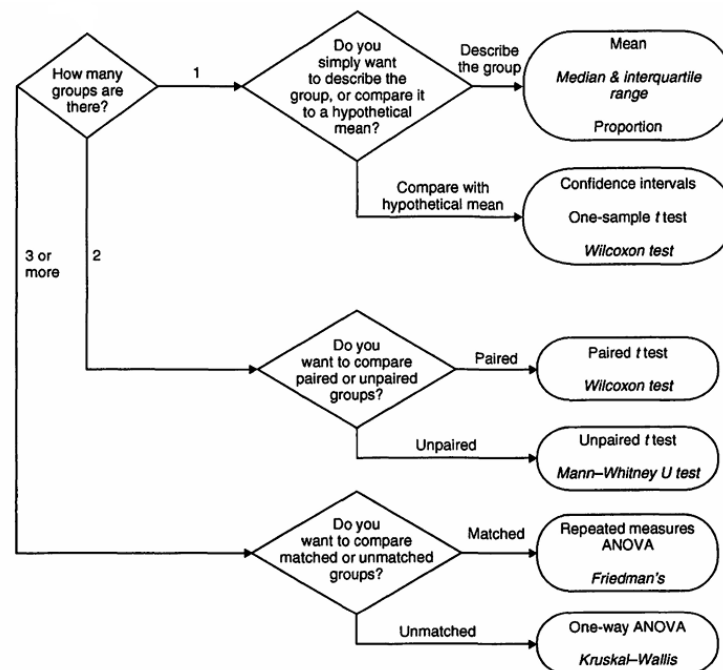


Figure 2: Flow chart for choosing statistical tests (Risan et al., 2024)

Table 1: Tests of Hypothesis for Location in Two-Sample Problems

Test Name	Purpose
One-Sample T-Test	Used for comparing a sample mean with the mean of the population, whether known or assumed.
Two-Sample T-Test (Independent)	Used for comparing the means of two independent groups to see if there is any significant difference between them.
Paired T-Test	Used for comparing the means of two dependent groups, like taking measurements before and after a certain treatment in the same subjects.
Mann-Whitney U Test	Used for comparing the medians of two independent groups.

Nevertheless, it is important to understand that alongside classical hypothesis testing methods. There are also some Type I and Type II errors associated with classical statistical tests that have to be carefully considered in scientific research (Cao et al., 2024). The likelihood of these errors might affect the validity of the research outcomes. Especially in studies where the sample size is relatively small. It is also important to mention that even though classical hypothesis testing techniques are quite powerful and comprehensible. They are also based on some assumptions that need to be confirmed beforehand through additional analysis. This issue has caused scientists to pay more attention to providing descriptive statistics and graphic representation along with hypothesis testing in scientific research (Rahman et al., 2022). This way, the connection between classical statistical procedures and exploratory data analysis proves the constant development of the process. To conclude, despite the popularity of modern statistical test procedures, classical statistical tests are essential in dealing with the location problem of two samples of continuous data. This provides valuable foundations for further development of methodology. The discussion of classical and modern statistical approaches emphasises the importance of strict following of the methodological principles. At the same time, adaptation to changing requirements implies the necessity for future research to consider the effectiveness of classical statistical tests in different scientific fields (Kelly, 2019). Striving for increasing accuracy and depth of interpretation, researchers are provided with an excellent tool to use classical approaches in combination with modern ones (Ewusie et al., 2020b). This way, classical tests

provide scientists with valuable instruments to make a useful contribution to scientific discussion appropriately (Fu et al., 2025; Ajegbile et al., 2024; Guetterman, 2019; Taylor et al., 2014).

Modern and Robust Statistical Tests

In terms of the development of statistical methods. One trend that is notable is the transition to advanced and more powerful statistical tests, which take into account the weaknesses of traditional methods. Especially when applied to two-sample location problems that use continuous variables. Although it is true that traditional methods like the t-test are popular and have been used extensively in statistics due to their simplicity. They may not perform well in the case of non-normality and heteroscedasticity in data sets. On the other hand, statistical tools have been developed in modern times in order to strengthen robustness against such violations. In other words, modern tools such as bootstrapping and permutation methods have been developed to help derive confidence intervals. Further to conduct hypothesis testing without relying on the strict normality assumption used by parametric tools (Adejare, 2025). The main strength of these tools is that they use the empirical distribution of the data in deriving the sampling distribution and hence are flexible enough. Non-parametric tests have been introduced as an advance in statistics. Tests such as the Mann-Whitney U test and the Wilcoxon rank-sum test, as shown in Table 2, serve as robust alternatives to the t-test, which is used for comparing two independent samples. This test is used when dealing with ordinal data or continuous data where there is no need to assume normality (Rogerson, 2020). Such tests are essential since they enable one to get valid conclusions despite some of the common data problems. Also, developments in computational power have helped in adopting more advanced models. Like generalised additive models (GAM) and mixed-effects models that consider fixed as well as random effects in the case of two-sample comparison. Thereby adding further depth to the analysis by including hierarchy that usually exists in real data (Tybl, 2016). Furthermore, the inclusion of Bayesian methodology in the realm of two-sample testing has been a paradigm shift in itself. Bayesian statistics, which involves using prior information along with making decisions based on the probability of hypotheses (Rigdon et al., 2024). This is considered more intuitive and helpful when there is less information available in a complicated situation (Nesbitt et al., 2014). In addition, Bayesian methods help deal with uncertainty more effectively. So it is considered an important tool to solve one of the main drawbacks of classical methods, where the complexity of real data distributions was ignored (Naimi et al., 2014; Stoudt et al., 2021; Saari, 2025).

The latest findings point to the effectiveness of machine learning algorithms that present an additional perspective from which contemporary statistical methods can be applied. Methods like support vector machines and random forests proved to be highly effective in situations where classification was required, and the regular statistical methods were not able to perform (Caraffini & Iacca, 2020). In particular, in the two-sample location problem setting, machine learning methods, such as support vector machine (SVM), redefine the problem of distributions' comparison in terms of classification. Rather than comparing means of samples, SVM finds the optimal hyperplane separating the two samples and maximising the margin between them. The size of the margin and the resulting classification accuracy become non-parametric, high-dimensional measures of location difference between continuous distributions that are particularly useful in the presence of non-linear structure of data or high-dimensionality of data. The fact is that machine learning algorithms are capable of handling nonlinear relationships between the variables, thereby pushing the boundaries of analysis in the two-sample location problem. Through the use of big data and computer technology, machine learning algorithms present the prospect not only of confirming the findings of classical statistical methods but also of discovering new patterns in the data that might produce new knowledge. However, one should note that the variety of improvements notwithstanding, there remain some issues related to modern statistical tests that should be considered in this context. The use of robust and non-parametric tests can result in high statistical power in some cases. However, the situation calls for caution regarding the use of these tests since misunderstanding is possible if researchers start using these tests without proper knowledge of their features (Couch et al., 2019). It is important

for specialists to develop a good understanding of different statistical tests and use the most suitable test in each particular case. Therefore, the combination of both classical and modern statistical tests will allow for developing a more comprehensive approach towards solving two-sample location problems for continuous data (Guerrier et al., 2019).

Table 2: Robust Nonparametric Two-Sample Location Tests

Test Name	Description
Wilcoxon-Mann-Whitney Test	Non-parametric test for comparing two independent samples to determine if they belong to the same distribution.
Yuen's Test	Robust procedure that employs trimmed means to examine differences between the locations of two samples, resulting in increased efficiency in the presence of outliers.
Hodges-Lehmann Estimator Test	Robust estimator for measuring the difference between the locations of two samples, and is commonly employed in conjunction with permutation tests.
Hotelling's T^2 Test (Robust Versions)	Multivariate analogue of the t-test, including robust variants.

Comparative Analysis

Following the critical analysis of both classical and modern tests used for comparing means of two samples. The comparative analysis of these approaches has led to some important findings regarding their respective advantages and disadvantages. There are several research studies confirming the widespread use of classical statistical approaches, including the t-test, due to their simplicity and straightforwardness. Especially if the sample size is large enough and the data distribution is normal and equal variance conditions are met (Adejare, 2025; Rogerson, 2020). At the same time, the necessity to meet the stated assumptions may be a serious problem in cases of real-life data that tend to depart from normal distribution or equal variances. The above-mentioned issues have become a driving force behind the creation of more flexible statistical approaches. For example, nonparametric tests like the Mann-Whitney U test have been recognised as a good alternative. Especially in cases of non-normal distribution and skewed data. Allowing drawing some reliable conclusions about group differences (Tybl, 2016; Nesbitt et al., 2014). In addition, the new developments in modern statistical tests, such as Bayesian tests and permutation tests, increase the robustness of these analyses. They provide additional flexibility concerning model assumptions and reveal data structures that traditional methods cannot see (Naimi et al., 2014; Caraffini & Iacca, 2020). Nowadays, there are increasing numbers of analysts who utilise these advanced statistical tools in different areas of research. Acknowledging their ability to extract meaningful results despite the complexity of the data used (Kim et al., 2019; Couch et al., 2019). Moreover, modern statistical tests, as shown in Table 3, are widely compared with traditional approaches. Demonstrating some examples where modern methods are superior not only statistically but also in terms of the information content due to the inclusion of prior information (Guerrier et al., 2019; BatiĆ et al., 2012). The choice of whether classical or modern methods are used will depend on the details of the data that have been collected. For instance, classical tests would be adequate for small sample sizes and populations that have been characterised well enough. Large data sets that show complex patterns will demand the use of more sophisticated analysis procedures to capture the complexity of the data (Cao et al., 2024; Rahman et al., 2022). Interpretability is one aspect of classical tests that is very important because some practitioners do not possess sufficient statistical knowledge to incorporate more modern methodologies into practice (Kelly, 2019; Ewusie et al., 2020). A combination of classical and modern methods seems to indicate that hybrid techniques could be possible. Classical methods could be used initially to guide which advanced techniques ought to be incorporated. This way, an analytical process will be enhanced and made more interpretable (Fu et al., 2025). The effectiveness of either classical or modern methods has to be examined in terms of applicability in particular cases. Where the use of computational and temporal resources is a limiting factor (Fu et al., 2025; Ajegbile et al., 2024). In this regard, future research might involve the development of software that combines all these different methodologies. Making it possible for researchers in different areas to effortlessly move from classical methods to modern approaches

depending on data features and research aims (Ewusie et al., 2020a). In addition to this, empirical research, which compares the application of classical and modern methodologies in different data sets. It can offer more detailed information about their advantages and disadvantages (Guetterman, 2019; Taylor et al., 2014). It is worth noting that taking into consideration the development of statistical science. It is necessary to establish communication between classical and modern methodologies. This can not only enhance empirical research but also open new horizons for methodological research. Table 4 is provided in order to give a systematic overview of the methodologies mentioned above by describing the basic underlying assumptions, their respective strengths and weaknesses, and possible applications.

Table 3: Power Comparison between Two-Sample Location Tests

Test Type	Power
Permutation Test	High
Hotelling's T^2 Test	Moderate
U-Statistic Based Test	High
Max-Type Rank Test	Moderate

Table 4: Two-Sample Location Test Categories: Comprehensive Comparison

Test Type	Key Assumptions	Main Advantages	Limitations	Recommended Use Cases
Classical Parametric (e.g., Student's t-test)	Homogeneity of variance, Normality, Independence	High power when assumptions are met, Simple calculations and interpretation	Sensitive to the presence of non-normality, heteroskedasticity	Data sets that are well-characterised, with moderately large sample sizes and
Classical Non-Parametric (e.g., Mann-Whitney U)	Independent samples, Ordinal or continuous data	Robust against non-normal distributions and outliers, Applicable to small samples	Less power compared to t-test under conditions of normality	Skewed continuous data, ordinal data, or failure of parametric assumption requirements
Resampling / Robust (e.g., Permutation, Yuen's, Bootstrap)	Distribution assumptions minimum, Observations independent	High efficiency, Deals with heteroscedasticity and heavy skewness	Computationally expensive, Less standard reporting	Small to moderate sample size, skewness, heavy-tail distribution
Modern Advanced (e.g., Bayesian, ML-based)	Prior distribution specifications (Bayesian)	Uses existing knowledge, Detects complicated/curvilinear relationships	Needs high computing power and expertise in statistics.	Large or high-dimensional data sets that need predictive or probabilistic analysis

Practical Guidelines and Future Directions

The present era of statistical testing for two-sample location tests requires an insight into both traditional and current approaches. Due to the fact that researchers are increasingly facing different types of data. In such circumstances, the formulation of recommendations on how to use the approaches becomes essential when analysing continuous data. For users of these approaches, the first thing to consider is the assumptions of the test selected for data analysis. For instance, in the case of classical approaches such as the t-test. It is assumed that there is homogeneity of variances. However, current tests such as the Mann-Whitney U test can be used even when this assumption does not hold (Adejare, 2025). It is also important to perform preliminary analysis of the data by conducting tests of normality and homogeneity of variance before hypothesis testing (Rogerson, 2020). It is also important to comprehend the implications associated with the effects of sample size and power analysis. Power analysis gives insights into the chances of finding a true effect. It is expected to be included in the study design routine as a standard procedure (Tybl, 2016). In small samples, variance is likely to be underestimated. Thus, increasing the risk of obtaining a Type I error. From this perspective, modern simulation methods, such as bootstrap and permutation tests, can become an effective substitute for

theoretical distribution-based procedures in terms of boosting power (Nesbitt et al., 2014). The development of computational statistics helps to adopt simulation methods and allows one to apply particular analytical tools to specific datasets while being accurate enough. Finally, the use of machine learning methods opens perspectives for further research in the two-sample location problem. Unlike classical statistical methods oriented at hypothesis testing procedures. Machine learning can help to detect associations and patterns ignored by hypothesis tests (Naimi et al., 2014). This combination of classical statistics and modern computational methods requires collaboration between a statistician and a data scientist for making predictive models statistically sound and practically applicable (Caraffini & Iacca, 2020). For example, algorithms such as support vector machines in combination with traditional test statistics can become quite useful in analysing data. Speaking of prospects, one of the key challenges that exists relates to the issue of standardisation of reporting practices in two-sample tests. Having a variety of different methodologies available does not necessarily mean that inconsistency in the reporting of the results is out of the picture; therefore, it would be reasonable to promote standardised guidelines on how to report metrics, effect sizes, and confidence intervals to facilitate the comparison between various studies (Couch et al., 2019). It is also important to highlight the fact that focusing on data sharing and collaboration can significantly contribute to establishing transparency within the scientific community and equal access to analytical techniques for all researchers (Guerrier et al., 2019). Additionally, the increased availability of various technologies used for collecting data poses certain challenges for researchers in this field. In light of the emerging big data phenomenon, it becomes necessary to update conventional statistical techniques in order to analyse the huge and multidimensional sets of data that exist today (Batić et al., 2012). Therefore, scholars should not only adopt traditional statistical techniques in this environment but also work on developing new ones. In that, considering domain-specific characteristics in the development of statistical models would significantly increase the relevance and usability of these techniques (Cao et al., 2024). Speaking of prospects, one of the key challenges that exists relates to the issue of standardisation of reporting practices in two-sample tests. Having a variety of different methodologies available does not necessarily mean that inconsistency in the reporting of the results is out of the picture. Therefore, it would be reasonable to promote standardised guidelines on how to report metrics, effect sizes, and confidence intervals in order to facilitate the comparison between various studies (Couch et al., 2019). It is also important to highlight the fact that focusing on data sharing and collaboration can significantly contribute to establishing transparency within the scientific community and equal access to analytical techniques for all researchers (Guerrier et al., 2019). Additionally, the increased availability of various technologies used for collecting data poses certain challenges for researchers in this field. In light of the emerging big data phenomenon, it becomes necessary to update conventional statistical techniques in order to analyse the huge and multidimensional sets of data that exist today (Batić et al., 2012). Therefore, scholars should not only adopt traditional statistical techniques in this environment but also work on developing new ones, as shown in Table 5.

Table 5: Two-Sample Location Problem Tests: Suggestions and the Description

Test Type	Description
Rank-Based Tests	The non-parametric tests used for the comparison of ranks from data belonging to two different samples for assessing location differences.
U-Tests	The non-parametric tests of U statistics, which are appropriate for testing two independent samples with non-normally distributed data.
Max-Type Tests	These tests emphasise the largest difference between observations in the two samples.
Adaptive Tests	These are tests whose approach depends on data characteristics.

In that, considering domain-specific characteristics in the development of statistical models would significantly increase the relevance and usability of these techniques (Cao et al., 2024). However, using only p -values in statistical significance tests is inadequate for measuring practical significance because the p -values are dependent on the sample size and do not indicate the size of the difference. Hence, it is important for researchers to complement the hypothesis test with a proper estimation of the effect size to measure the practical significance of the difference in

location. In the case of parametric analysis, it is appropriate to use the standardised mean difference like Cohen's d or Hedges' g . However, when the assumption of normality or equality of variances is violated, non-parametric measures of effect size like Cliff's δ can be used.

Theoretical Contributions

One contribution this review makes to statistical theory is in terms of developing a taxonomy that incorporates both classical and computational perspectives into one coherent theory. This theory sheds light on the development from parametric assumptions to resampling and non-parametric methods in addressing two-sample location problems.

Managerial Implications and Recommendations

In terms of decision-making processes, the application of automatic preliminary testing tools (for instance, Shapiro-Wilk for normality and Levene's test for homoscedasticity) before the test choice is advised by Risan et al. (2024). Regarding the management of violation of assumptions, in cases where there is heteroscedasticity or skewness, management should opt for Yuen's test or bootstrapping procedures to mitigate the occurrence of Type I error, according to Couch et al. (2019). For high-dimensional decision making, in the context of industrial quality control and high-dimensional data mining, classification algorithms (for example, SVM margin analysis) need to be used along with classical t-tests for evaluating shifts in location.

Conclusion

The study of various statistical tests in the area of the two-sample location problem has provided substantial insights regarding the field of statistics and its development throughout time. In particular, it has become clear that there are many approaches to conducting statistical inference on continuous data that can provide numerous opportunities for researchers in terms of choosing the most appropriate technique for performing a certain type of analysis. Traditional methods like the t-test and Mann-Whitney U test were compared with more recent approaches like bootstrapping and Bayesian methods. The review of these methodologies has shown how the field of statistical inference has evolved in accordance with changes that appeared in data sets in recent times. It is important to note that modern datasets are becoming more and more complex due to the presence of different issues like non-normality and heteroscedasticity, which make classical approaches obsolete in some cases. That is why there is a constant exchange of ideas about the use of various statistical techniques that need to be considered by scientists depending on the features of their data. The importance of the use of modern tools can be easily explained by the fact that nowadays many sciences require highly accurate and reliable statistical conclusions in order to advance further. Some studies prove that the use of modern statistical tests allows overcoming certain problems of ordinary methods in complicated experimental design. However, the change from classical methods to modern approaches has faced some barriers, such as issues concerning the violation of model assumptions, computational capacity, and required knowledge about statistical techniques for making correct conclusions. Furthermore, many gaps in the area require further attention. For instance, it has been revealed that some advanced computational methods have not been widely used in practice because people do not have sufficient training for applying these methods properly.

Despite being exhaustive in its coverage, this literature review has two major limitations: (i) it concentrates mainly on continuous variables in the two-sample location problem, and (ii) it reviews computational approaches theoretically without conducting any empirical simulations with different sample sizes. To further this area, the below-mentioned research questions need to be addressed in future studies:

- Research Question 1 (RQ_1): How do margins of machine learning classification compare to adaptive non-parametric tests in high-dimensional, heavy-tailed data?
- RQ_2 : Which unified guidelines would allow for the standardisation of the incorporation of prior distributions into the Bayesian two-sample tests framework?

Another issue that should be addressed in future research is the exploration of the efficiency of different statistical techniques depending on certain scenarios. In particular, in cases when sample sizes are relatively small, or data is highly dimensional. Finally, it should be stated that the review of classical and modern statistical tests has shown that there is a great number of problems that need to be solved in the area in order to make sure that these tests will allow advancing empirical research and improving the decision-making process in various domains. It is necessary to develop certain protocols for using statistical tests that will promote best practices in the area. Moreover, the incorporation of advancements in the area of machine learning and artificial intelligence may help to overcome the current gaps and problems. In conclusion, ongoing discussions regarding the application of statistical tests for the "two-sample location problem" reveal a significant interplay between traditional and modern approaches to this type of data analysis. There remain numerous research gaps concerning the integration of traditional methods with modern techniques.

References

- Adam, C. (2011). *Essential mathematics and statistics for forensic science*. John Wiley & Sons.
- Adejare, T. (2025). CONSTRAINTS IN SOCIAL SCIENCE RESEARCH: DATA COLLECTION AND SAMPLING TECHNIQUES FOR HARD-TO-REACH POPULATION. *JOURNAL OF SOCIAL SCIENCES*. <https://api.semanticscholar.org/CorpusID:275923871>
- Ajebgile, M. D., Olaboye, J. A., Maha, C. C., Igwama, G. T., & Abdul, S. (2024). The role of data-driven initiatives in enhancing healthcare delivery and patient retention. *World Journal of Biology Pharmacy and Health Sciences*, 19(1), 234–242.
- BatiĈ, M., Paganoni, A. M., Pfeiffer, A., Pia, M. G., & Ribon, A. (2012). A new development cycle of the Statistical Toolkit. *Journal of Physics: Conference Series*, 396(5), 052010.
- Cao, Y., Chen, R. C., & Katz, A. J. (2024). Why is a small sample size not enough? *The Oncologist*, 29, 761–763. <https://api.semanticscholar.org/CorpusID:270761386>
- Caraffini, F., & Iacca, G. (2020). The SOS platform: Designing, tuning and statistically benchmarking optimisation algorithms. *Mathematics*, 8(5), 785.
- Couch, S., Kazan, Z., Shi, K., Bray, A., & Groce, A. (2019). Differentially private nonparametric hypothesis testing. *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, 737–751.
- Ewusie, J. E., Soobiah, C., Blondal, E., Beyene, J., Thabane, L., & Hamid, J. S. (2020a). Methods, applications and challenges in the analysis of interrupted time series data: a scoping review. *Journal of Multidisciplinary Healthcare*, 411–423.
- Ewusie, J. E., Soobiah, C., Blondal, E., Beyene, J., Thabane, L., & Hamid, J. S. (2020b). Methods, applications and challenges in the analysis of interrupted time series data: a scoping review. *Journal of Multidisciplinary Healthcare*, 411–423.
- Fu, A., Shen, T., Roberts, S. B., Liu, W., Vaidyanathan, S., Marchena-Romero, K.-J., Lam, Y. Y. P., Shah, K., & Mak, D. Y. F. (2025). Optimizing the efficiency and effectiveness of data quality assurance in a multicenter clinical dataset. *Journal of the American Medical Informatics Association*, 32(5), 835–844.
- Guerrier, S., Karemera, M., Orso, S., & Victoria-Feser, M.-P. (2019). A simple recipe for making accurate parametric inference in finite sample. *ArXiv Preprint ArXiv:1901.06750*.
- Guetterman, T. C. (2019). Basics of statistics for primary care research. *Family Medicine and Community Health*, 7(2), e000067.
- Kelly, M. (2019). *The standard errors of persistence*. UCD Centre for Economic Research Working Paper Series.
- Kim, I., Lee, A. B., & Lei, J. (2019). *Global and local two-sample tests via regression*.
- Mathews, P. G. (2004). *Design of Experiments with MINITAB*. Quality press.
- Montgomery, D. C., Runger, G. C., & Hubele, N. F. (2007). *Engineering statistics*. John Wiley & Sons. Inc: Hoboken, NJ, USA.
- Morgan, C. J. (2017). Use of proper statistical techniques for research studies with small samples. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 313(5), L873–L877.

- Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography*, 37(2), 191–203.
- Navidi, W. C. (2008). *Statistics for engineers and scientists* (Vol. 20). McGraw-Hill Higher Education New York.
- Nesbitt, R. C., Gabrysch, S., Laub, A., Soremekun, S., Manu, A., Kirkwood, B. R., Amenga-Etego, S., Wiru, K., Höfle, B., & Grundy, C. (2014). Methods to measure potential spatial access to delivery care in low- and middle-income countries: a case study in rural Ghana. *International Journal of Health Geographics*, 13(1), 25. <https://doi.org/10.1186/1476-072X-13-25>
- Rahman, M. H., Islam, M. H., & Neema, M. N. (2022). GIS-based compactness measurement of urban form at neighborhood scale: The case of Dhaka, Bangladesh. *Journal of Urban Management*, 11(1), 6–22.
- Rigdon, S. E., Ronald D. Fricker, J., & Montgomery, D. C. (2024). *Introduction to Probability and Statistics for Data Science*. <https://api.semanticscholar.org/CorpusID:274725436>
- Risan, H. K., Al-Azzawi, A. A., & Al-Zwainy, F. M. S. (2024). *Statistical Analysis for Civil Engineers*. Elsevier.
- Rogerson, P. A. (2020). *Statistical methods for geography: a student's guide*. SAGE Publications Ltd.
- Saari, P.-L. (2025). *Bayesian and Bootstrap Methods for Estimating Population Attributable Risk: Development of an R Package for Epidemiological Inference*.
- Stoudt, S., Pintar, A., & Possolo, A. (2021). Uncertainty evaluations from small datasets. *Metrologia*, 58(1), 015014.
- Taylor, M. J., McNicholas, C., Nicolay, C., Darzi, A., Bell, D., & Reed, J. E. (2014). Systematic review of the application of the plan–do–study–act method to improve quality in healthcare. *BMJ Quality & Safety*, 23(4), 290–298.
- Tybl, A. J. (2016). An overview of spatial econometrics. *ArXiv Preprint ArXiv:1605.03486*.
- van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.-J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, 28(3), 813–826. <https://doi.org/10.3758/s13423-020-01798-5>
- Zhang, C., Sang, P., & Qin, Y. (2025). Two-sample inference for sparse functional data. *Electronic Journal of Statistics*, 19(1). <https://doi.org/10.1214/25-EJS2348>

Published by

IJSAB
International

