

When to Use t-Test, ANOVA, ANCOVA, MANOVA, and MANCOVA: A Step-by-Step Guide for Novice Educational Researchers

Emmanuel Nestory Kayuni ^{1*}

¹ Department of Education, School of Science Education, Mwalimu Nyerere University of Agriculture and Technology (MNUAT), Mara, Tanzania

ABSTRACT

Selecting an appropriate statistical test remains a challenge for novice educational researchers, particularly when distinguishing among parametric procedures with different requirements and analytical purposes. This paper provides a practical guide for selecting the t-test, ANOVA, ANCOVA, MANOVA, and MANCOVA in quantitative educational research. The guide begins with hypothesis testing and variable classification, followed by normality assessment and consideration of data transformation before selecting the appropriate statistical procedure. Each test is explained according to its purpose, assumptions, research question, types, decision criteria, and interpretation. The paper emphasizes the necessity of choosing statistical tests based on the research hypothesis structure, number of groups, dependent variables, and covariates, rather than by familiarity with statistical software. The guide is designed to assist novice education researchers to make sound statistical choices and improve the validity and interpretability of their quantitative research results.

ARTICLE HISTORY

Received 29 August 2026
Revised 12 September 2026
Accepted 25 October 2026
Online First 25 October 2026

KEYWORDS

ANOVA; ANCOVA; MANOVA;
MANCOVA; Novice
Educational Researchers;
t-test

1. Introduction

Statistical analysis is one of the key components of educational research because quantitative studies require researchers to select analytical procedures that correspond appropriately to their research questions, hypotheses, and data structures (Mijar, 2024). At the postgraduate level, this requirement is particularly important because master's and doctoral students are expected to justify the statistical procedures used to analyze and interpret their findings (Najmi et al., 2021). However, choosing an appropriate parametric statistical procedure can be challenging for novice educational researchers particularly when deciding which statistical test to use. The common parametric statistical tests include the t test, Analysis of Variance (ANOVA), Analysis of Covariance (ANCOVA), Multivariate Analysis of Variance (MANOVA), and Multivariate Analysis of Covariance (MANCOVA) (Mijar, 2024). The choice

among these parametric statistical tests depends on the nature of the research question, the independent and dependent variables and the presence of covariates. In earlier methodological studies, shortcomings in the application and reporting of t-test, ANOVA, ANCOVA, MANOVA and MANCOVA in educational research have been reported, especially regarding statistical assumptions and their reporting practices (Kim et al., 2024; Ranganathan, 2021).

The problem is not simply that statistical information is unavailable. Novice educational researchers can access explanations of statistical procedures through journal articles, textbooks, research-methods courses, and statistical software, but access to information does not guarantee a clear pathway from a specific educational research question to the appropriate statistical test. Choice of a particular statistical test to suit a particular research data depends on several considerations. For example, the choice of statistical test depends on whether a researcher compares independent groups or related observations and whether the analysis involves one or several dependent variables (Naskar & Das, 2018). Similarly, the current literature on educational statistics points out the development of statistical literacy through meaningful engagement with substantive questions and data rather focusing exclusively on statistical computation (Yusof et al., 2021). The growing emphasis on data literacy and data science competencies in higher education also exemplifies the importance of developing the capacity to use and interpret quantitative information appropriately.

Consequently, a practical guide is needed to help novice education researchers connect their research questions directly to appropriate statistical procedures. This paper aims to provide such a guide by progressively examining when to use the t test, ANOVA, ANCOVA, MANOVA, and MANCOVA based on the nature of the research question and the characteristics of the data. The study sought to help novice educational researchers at master's and doctoral level to choose appropriate statistical test and build their confidence during data analysis and data reporting.

2. Method

The present study employed a narrative literature review approach, as outlined by Sukhera (2022), to provide practical guidance for selecting appropriate parametric statistical tests in quantitative educational research. A literature search was conducted in Google Scholar, ResearchGate, PubMed, ERIC (Education Resources Information Centre), Scopus and Web of Science between July and August 2026. Search terms included “parametric statistical tests”, “statistical test selection”, “t-test”, “ANOVA”, “ANCOVA”, “MANOVA”, “MANCOVA”, “normality testing” and “data transformation”. Titles and abstracts were screened to find publications that met the review objectives. Sources were selected based on their relevance to the selection of statistical tests, assumptions, applications, and interpretation of statistical tests in quantitative research. All the publications that did not meet these criteria were excluded. Additional relevant sources were collected from the reference sections of the selected publications. The literature was synthesized narratively and thematically, encompassing aspects such as statistical hypothesis testing, variable classification, assessment of normality, the selection of appropriate parametric tests, and data transformation.

3. Practical Guide to Selecting Parametric Statistical Tests

3.1 Statistical Hypothesis Testing

A methodical process by which a researcher assesses a study claim using evidence derived from sample data is known as statistical hypothesis testing (Rahayu et al., 2024). There are five primary steps to follow when testing a hypothesis. (i) The researcher first develops the hypothesis that will be investigated. (ii) After that, a significance level is established as the standard for reaching the statistical conclusion. (iii) The relevant test statistic is computed from the gathered data. (iv) The computed test statistic is interpreted by comparing it with the specified rejection region or by determining its corresponding P value. (v) Based on the statistical data gathered, the researcher concludes with regard to the null hypothesis (Tsushima, 2022).

If the p-value is greater than the predetermined significance level (α), this means that there is insufficient evidence to reject the null hypothesis, as the evidence is not sufficient to claim a statistically significant result (Kuzma, 1998; Samuels & Witmer, 2003; Sokal & Rohlf, 2009). On the other hand, if the p-value is less than or equal to predetermined significance level (α), this indicates that there is sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis, given the assumptions of the chosen test (Gravetter & Wallnau, 2017; Gupta & Kapoor, 2002; Rashid, 2016). For instance, if the computed P value is .001 and the significance threshold is set at .05, the null hypothesis is rejected because .001 is less than .05. Similarly, the null hypothesis is rejected if the computed P value is .08 and the significance threshold is set at .10. Since .08 is less than .10, there is sufficient evidence to reject H_0 . This systematic approach provides a framework for testing research hypotheses and drawing conclusions from statistical data.

3.2 Types of Variables

Choosing an appropriate parametric statistical test requires an understanding of the variables involved. Variables provide a powerful hint for choosing the right statistical test. According to Panos and Boeckler (2023), a variable is a feature of an individual, a group, or an educational phenomenon that can have varying values during observations. The variables in education research may include student achievement scores, age, motivation, type of school, method of instruction, and level of satisfaction. Kim et al. (2024) divided variables into quantitative variables and categorical variables. This classification is important because the characteristics of the variables influence which statistical techniques are appropriate.

Instead of using quantities to explain observations, categorical variables use groups or categories. They could be ordinal or nominal. Nominal variables are variables that do not have any meaningful order such as “school type” (public or private), “teaching method” (lecture, cooperative learning, or inquiry-based learning), and “students’ program of study” (Kim et al., 2024). For example, students’ responses to a Likert-type item range from strongly disagree to strongly agree. Ordinal variables have categories that follow a meaningful order (Kim et al., 2024). Quantitative variables are numerical representations of attributes that indicate the quantity or extent of the property being measured (Naskar & Das, 2018). They can be divided into two categories, i.e., continuous and discrete. Discrete variables, like the number of students enrolled in a class or the number of courses a student has passed, have separate countable values. Continuous variables can take a value in a range and can be seen as quantitative measurements. Examples include the height of pupils, the time taken to complete a test and exam scores (Naskar & Das, 2018).

3.3 Normality Test

After identifying and classifying the variables, the next step is to determine whether the assumptions relevant to the proposed statistical procedure are met to a reasonable extent, including normality when relevant (Di Leo & Sardanelli, 2020). Normality assessment is one component of assumption checking and helps researchers evaluate whether the distributional assumptions relevant to a proposed statistical procedure are reasonably satisfied. The specific distribution to be assessed depends on the procedure and research design, rather than requiring all observed variables to follow a normal distribution (Habibzadeh, 2024). For instance, in a study to determine whether teaching methods affect students’ math achievement, the researcher may first look at the distribution of the achievement scores and, if appropriate, examine the residuals from the statistical model under consideration. For procedures such as the t-test and ANOVA, the normality assumption generally concerns the distribution of errors or residuals within groups, rather than requiring all observed scores across the entire dataset to follow a normal distribution. The goal is not to achieve perfect normality but to determine if violations of normality are so large that one should be concerned about the proposed statistical analysis.

Three methods plot can be used to assess normality: the Shapiro-Wilk test (S-W test), the Kolmogorov-Smirnov test (K-S test) and the Q-Q plot. The Shapiro-Wilk test evaluates whether the data provide sufficient evidence to reject the null hypothesis that they follow a normal distribution (Kim & Park, 2019). The Kolmogorov-Smirnov test assesses the difference between the observed and expected normal distribution. The Q-Q plot is a graphical tool to compare the observed values with the theoretical normal values (Kim & Park, 2019). Decisions based on the Shapiro–Wilk and

Kolmogorov–Smirnov tests are commonly made using a significance level of .05. If $p > .05$, the researcher fails to reject the null hypothesis because there is insufficient evidence to conclude that the data depart significantly from normality. This finding, however, does not show that the data are normally distributed. Furthermore, normality testing should not be treated as an automatic prerequisite for every parametric test; instead, researchers should assess the assumptions relevant to the specific statistical procedure being considered. For instance, if the Shapiro-Wilk test gives $p = .184$, the researcher fails to reject the null hypothesis but also needs to look at the Q-Q plot and other relevant diagnostic information. In contrast, if the test gives $p = .012$, the researcher rejects the null hypothesis because the result shows statistically significant evidence of departure from normality (Schoder et al., 2006). For the Q-Q plot, the observations that generally follow the diagonal reference line indicate approximate normality, while substantial systematic departures from the line indicate possible non-normality. The three approaches should be interpreted together, particularly because formal normality tests can be influenced by sample size (Kim & Park, 2019).

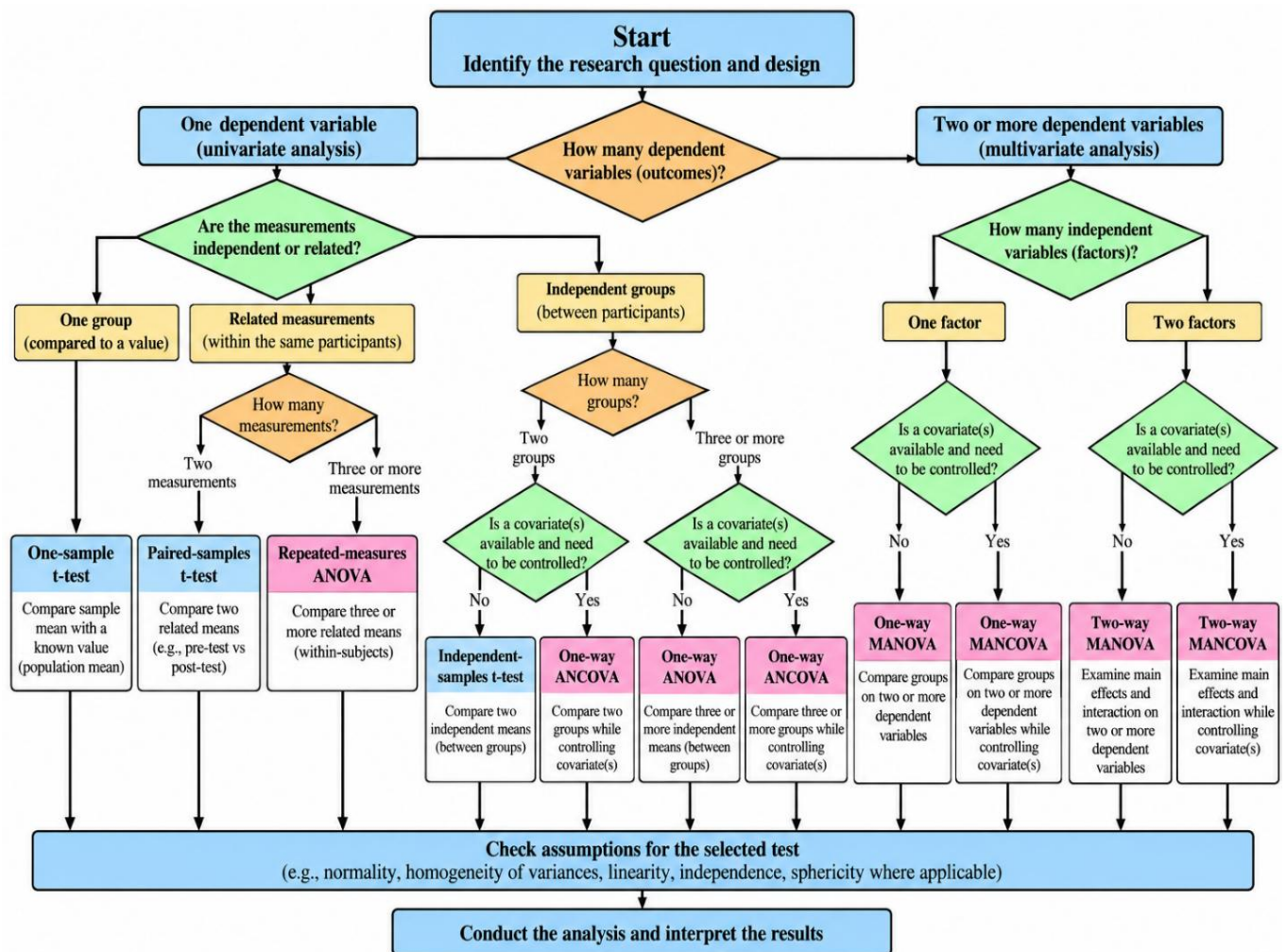
Based on this assessment, the researcher can determine whether the normality assumption relevant to the proposed statistical analysis is reasonably satisfied, while recognizing that other assumptions may also need to be examined (Schoder et al., 2006). A non-significant normality test should not be interpreted as evidence of normality, and a significant result does not automatically mean that a parametric procedure is inappropriate. Thus, normality assessment contributes to evaluating the suitability of a proposed statistical procedure but does not, by itself, determine whether a parametric or non-parametric test should be selected (Mijar, 2024). The choice of a particular test such as t-test, ANOVA, ANCOVA, MANOVA and MANCOVA depends on the research question, the design of the study, the number of groups, the number of dependent variables, and the presence of a covariate.

3.4 Selection of the Appropriate Parametric Test

Parametric statistical tests are used to test hypotheses about population parameters when the characteristics of the data and assumptions of the statistical model justify their use (Ranganathan, 2021). In educational research, the choice of procedure depends on the structure of the research question, specifically the number of groups being compared, the relationship between observations, the number of dependent variables, and whether a covariate is part of the analysis (Schmidt, 2025), as described in Figure 1. The present section focuses on five parametric procedures commonly encountered in quantitative educational research, namely the t test, ANOVA, ANCOVA, MANOVA and MANCOVA and the conditions under which each procedure can be used to address a particular research hypothesis.

Figure 1

Analytical Framework for Selecting Appropriate Parametric Statistical Tests (Researcher's Own Construct).



3.4.1 t-Test

The t test is a parametric procedure used to examine hypotheses concerning population means, particularly when the researcher seeks to determine whether a sample mean differs from a specified value or whether two means differ significantly (Najmi et al., 2021; Vrbín, 2022). Its appropriate use depends on several assumptions. (i) The dependent variable should be quantitative and measured at an appropriate continuous scale. (ii) The observations should be independent when an independent-samples design is used. (iii) The relevant scores should be approximately normally distributed, particularly when sample sizes are small. (iv) For the conventional independent-samples t test, the two populations should have approximately equal variances, although a separate-variance version can be used when this assumption is not satisfied. When this assumption is violated, Welch's t-test can be an appropriate alternative because it does not assume equal population variances. (v) Lastly, the data must be gathered by way of a proper sampling technique in order for the statistical inference to be meaningful (Gravetter & Wallnau, 2017; Kuzma, 1998; Sokal & Rohlf, 2009). These assumptions should be considered before interpreting the t statistic, as violations of relevant assumptions can affect the validity of statistical inference.

There are three major versions of the t test, and the choice of a particular t test depends on the research question, dependent variables, independent groups and the relationship between the observations. First version of the t test is a one sample t test which compares the group's mean with some specified or hypothesized population mean (Panos & Boeckler, 2023). For example, an education researcher might test whether the average mathematics score of a group of students differs from a national benchmark of 60 marks. Second version is the independent-sample t test which compares the means of two independent groups (Kim et al., 2024); for example, comparing the mean achievement scores of students taught through a traditional method and those taught through a flipped-classroom approach (Almasi, 2026; Majeed & Ahmad, 2025; Nawastheen, Sisindra, et al., 2026). The third version is the sample t test, which is appropriate when the two sets of observations are related; for example, comparing the pre-test and post-test achievement scores of the same students following an instructional intervention (Frane, 2021). However, if three or more than three sets of observations are related, repeated measures ANOVA can be appropriate than a paired sample t test which is limited to only two sets of related observations.

In every instance, the researcher formulates H_0 and H_1 , computes the t statistic, and analyzes the associated P value. In addition to the p-value, researchers should also report a suitable effect size such as Cohen's d to indicate the magnitude of the observed difference. At a predetermined significance level of $\alpha = .05$, a P value below .05 leads to the rejection of H_0 , whereas a P value above .05 leads to the failure to reject H_0 . The appropriate test is therefore determined first by the structure of the research question, not by the statistical software being used. However, statistical significance does not necessarily indicate that the observed difference is practically or educationally meaningful. Researchers should therefore interpret the p-value alongside the effect size and the educational context to determine the practical importance of the difference.

3.4.2 Analysis of Variance (ANOVA)

Analysis of variance (ANOVA) is a parametric procedure used when an educational researcher wants to determine whether the means of three or more groups differ on a quantitative outcome. Conducting several independent-samples t tests to compare multiple groups is not advisable because repeated testing increases the probability of making a Type I error, leading to rejection of a true null hypothesis (Gravetter & Wallnau, 2017). ANOVA is a statistical test that helps solve this problem by providing a single overall test of the group means (Kim et al., 2024). Before conducting ANOVA, the researcher should consider the following assumptions: (i) independence of observations, meaning that observations from one participant or group should not influence those from another, (ii) approximate normality of the dependent variable in each group, (iii) homogeneity of variance, meaning that the population variances across the groups are reasonably similar. When the assumption of homogeneity of variance is violated, Welch's ANOVA may be used as a robust alternative to conventional one-way ANOVA. (iv) a quantitative dependent variable that is appropriate for mean comparison (Ranganathan, 2021). For example, suppose a researcher wants to know if students taught through lecture, cooperative learning, and inquiry-based learning differ in mathematics achievement. One-way ANOVA would be appropriate because the study contains one categorical independent variable with three levels and one quantitative dependent variable.

The researcher then must determine whether the research question is answered by one-way or two-way ANOVA. If the study examines one categorical independent variable with three or more groups, the researcher would choose one-way ANOVA (Kim et al., 2024). If the study examines two categorical independent variables simultaneously, the researcher would choose two-way ANOVA (Gravetter & Wallnau, 2017; Sokal & Rohlf, 2009). For example, a researcher asking whether teaching method affects students' achievement would use one-way ANOVA, whereas a study examining whether teaching method and gender influence achievement would use two-way ANOVA (Hu & Zainudin, 2026; Tangkui, 2025). In the latter case, the researcher examines the main effect of teaching method, the main effect of gender, and the interaction between teaching method and gender. After the analysis, the researcher evaluates the F statistic and its corresponding P value. If $p < .05$, the relevant null hypothesis is rejected, and if $p > .05$, the null hypothesis cannot be rejected. Researchers should also report an appropriate effect size, such as eta-squared (η^2) or partial eta-squared (ηp^2), to indicate the magnitude of the observed group differences. If the results of a one-way ANOVA are significant, the researcher needs to conduct an appropriate post-hoc comparison to determine which groups differ. Where the assumption of equal variances is satisfied, Tukey's HSD may be used for pairwise comparisons, whereas the Games–Howell test is appropriate when variances are unequal (Kuzma, 1998).

For a two-way ANOVA, a significant interaction means the researcher needs to further examine the pattern of interaction using appropriate simple-effects or pairwise comparisons (Mijar, 2024). The novice researcher should therefore select one-way ANOVA when one independent variable is being examined and two-way ANOVA when two independent variables are examined simultaneously while ensuring that the assumptions of the selected procedure have been adequately considered.

3.4.3 Analysis of Covariance (ANCOVA)

Analysis of covariance (ANCOVA) is a parametric procedure used to compare the means of two or more groups on a quantitative dependent variable while statistically controlling for the influence of another quantitative variable, referred to as a covariate (Kim et al., 2024). ANCOVA differs from ANOVA in one key respect: ANOVA looks at group differences in the dependent variable without controlling for a covariate, while ANCOVA looks at those group differences after controlling for the variation associated with the covariate (Tsushima, 2022). ANCOVA is useful in educational research when groups differ on a variable that is related to the outcome and the researcher wants to compare them in a more controlled way. For example, when comparing students' post-test achievement across different teaching methods, students' pre-test achievement may be included as a covariate to control for differences in achievement that existed prior to the intervention. Prior to using ANCOVA, the researcher should consider (i) independence of observations, (ii) approximate normality of the model residuals, (iii) homogeneity of variance, (iv) a linear relationship between the covariate and the dependent variable, and (v) homogeneity of regression slopes across the groups (Panos & Boeckler, 2023). The covariate should also have a meaningful theoretical relationship with the dependent variable rather than simply being included because it is present in the dataset (Kim et al., 2024).

The researcher must then decide if ANCOVA is appropriate for the research question and design of the study. For example, ANCOVA can be used when there is one categorical independent variable (e.g., comparing three teaching methods while controlling for students' pre-test scores), or it can be extended to situations with two or more categorical independent variables (e.g., examining the effects of teaching method and gender on students' post-test achievement while controlling for pre-test achievement). The decision is therefore different from ANOVA because the research question contains an additional quantitative variable that the researcher intends to control statistically. The resulting analysis provides adjusted group means, and the researcher examines the F statistic and corresponding P value for the relevant group effect. Researchers should also report an appropriate effect size, such as partial eta-squared (η^2), to indicate the magnitude of the group effect after controlling for the covariate. If $p < .05$, there is evidence of a statistically significant group difference after adjustment for the covariate; if $p > .05$, there is insufficient evidence of such a difference. Statistical significance should be interpreted alongside effect size and educational context, because a statistically significant adjusted group difference does not necessarily indicate practical or educational significance. However, if the covariate-by-grouping variable interaction is statistically significant, then homogeneity of regression slopes assumption may not hold. The model should therefore be examined further before the results are interpreted (Kim et al., 2024). Thus, ANOVA is selected when the research question concerns group differences alone, whereas ANCOVA is selected when the same group comparison requires control of a relevant quantitative covariate.

3.4.4 Multivariate Analysis of Variance (MANOVA)

Multivariate analysis of variance (MANOVA) extends ANOVA when an educational researcher needs to examine differences between groups on two or more quantitative dependent variables simultaneously. The distinction is therefore based primarily on the number of dependent variables: ANOVA evaluates group differences on one dependent variable, whereas MANOVA evaluates whether groups differ on a combination of dependent variables (Zimmermann et al., 2020). MANOVA is particularly appropriate when the dependent variables are conceptually related, because analyzing them jointly allows the researcher to examine the overall multivariate pattern rather than treating each dependent variable as an entirely separate analysis (Mijar, 2024). For example, an education researcher may investigate whether three teaching methods produce differences simultaneously in students' academic achievement, learning motivation, and classroom engagement. Before selecting MANOVA, the researcher should consider (i) independence of observations, (ii) quantitative dependent variables, (iii) multivariate normality, (iv) homogeneity of the variance-covariance matrices across groups, and (v) the absence of serious multivariate outliers.

(Kim et al., 2024). These assumptions are important because violations can affect the validity of the multivariate test and the conclusions drawn from it. Where the homogeneity of variance-covariance matrices assumption is violated, Pillai's Trace may be preferred because it is generally more robust to certain assumption violations

The researcher must then determine whether one-way or two-way MANOVA corresponds to the research question. One-way MANOVA is selected when one categorical independent variable is used to compare groups on two or more quantitative dependent variables (Kim et al., 2024). For example, the question "Do lecture, cooperative learning, and inquiry-based learning produce differences in students' achievement, motivation, and engagement?" calls for a one-way MANOVA because there is one independent variable, the teaching method, and several dependent variables (Nawastheen, Liyanage, et al., 2026). Two-way MANOVA is appropriate when two categorical independent variables are examined simultaneously across the same set of dependent variables (Kim et al., 2024). For example, a researcher may examine whether teaching method and gender influence students' achievement, motivation, and engagement simultaneously (Zimmermann et al., 2020). The overall MANOVA result is evaluated using a multivariate test statistic such as Pillai's Trace, Wilks' Lambda, Hotelling-Lawley Trace, or Roy's Largest Root (Najmi et al., 2021). If the selected multivariate test produces $p < .05$, the researcher rejects the null hypothesis and concludes that the groups differ significantly on the combined dependent variables. The researcher should then examine suitable follow-up univariate tests to identify which dependent variables are responsible for the significant multivariate effect. If $p > .05$, the researcher fails to reject the null hypothesis, indicating insufficient evidence of a statistically significant multivariate difference. Thus, one-way MANOVA is appropriate when one independent variable is examined across multiple dependent variables, whereas two-way MANOVA is appropriate when two independent variables are examined simultaneously across multiple dependent variables.

3.4.5 Multivariate Analysis of Covariance (MANCOVA)

Multivariate analysis of covariance (MANCOVA) extends MANOVA by allowing the researcher to examine differences between groups on two or more quantitative dependent variables simultaneously while statistically controlling for one or more quantitative covariates (Zimmermann et al., 2020). The main distinction is therefore the inclusion of a covariate: MANOVA examines whether groups differ on a combination of dependent variables, whereas MANCOVA examines whether those groups differ on the combined dependent variables after accounting for variation associated with the covariate(s) (Kim et al., 2024). MANCOVA is particularly useful in educational research when an intervention is expected to influence several related dependent variables and the researcher also needs to adjust for an additional quantitative variable. For example, a researcher may examine whether three teaching methods differ in students' achievement, motivation, and engagement while controlling for students' pre-test achievement. Before applying MANCOVA, the researcher should consider (i) independence of observations, (ii) quantitative dependent variables and covariates, (iii) multivariate normality, (iv) homogeneity of variance-covariance matrices, (v) absence of serious multivariate outliers, (vi) linear relationships between the covariates and dependent variables, and (vii) homogeneity of regression slopes across groups (Kim et al., 2024). Where the homogeneity of variance-covariance matrices assumption is violated, Pillai's Trace may be preferred because it is generally more robust to certain assumption violations.

The choice of MANCOVA should therefore be based on the structure of the research question. MANCOVA is appropriate when the researcher has multiple quantitative dependent variables, one or more categorical independent variables, and a theoretically relevant quantitative covariate that needs to be statistically controlled. For example, a study asking whether three instructional approaches differ in students' achievement, motivation, and engagement after controlling for pre-test achievement requires MANCOVA rather than MANOVA because the analysis includes both multiple dependent variables and a covariate. Similar to MANOVA, MANCOVA provides a multivariate test of the overall group effect using statistics such as Pillai's Trace or Wilks' Lambda. If the multivariate P value is less than .05, the researcher rejects the null hypothesis and concludes that the groups differ significantly on the combined dependent variables after adjusting for the covariate. Follow-up analyses can then be used to determine which dependent variables contribute to the multivariate effect. If $p > .05$, the researcher fails to reject the null hypothesis, indicating insufficient evidence of a statistically significant multivariate difference. Researchers should also report an appropriate multivariate effect size and interpret the findings alongside the follow-up analyses, as statistical significance does not necessarily indicate practical or educational significance. Thus, the decision

pathway can be summarized as follows: MANOVA is used when comparing groups across multiple dependent variables, whereas MANCOVA is used when the same multivariate comparison additionally requires statistical adjustment for one or more quantitative covariates.

3.5 Data Transformation

When the normality assessment indicates a substantial departure from the distributional assumptions required by a planned parametric analysis, the researcher has to decide whether the original data can reasonably be analyzed as they are or whether an alternative strategy is necessary. Data transformation provides one possible response to such violations. It involves applying a mathematical function to the observed values to modify characteristics such as skewness and thereby improve the suitability of the data for the intended statistical model (Vrbin, 2022). Common transformations include logarithmic, square-root, and Box-Cox transformations, and the choice of the suitable transformation depends on the data characteristics and the type of the distributional problem (Ranganathan, 2021). For the purpose of this paper, transformation is therefore presented as a decision option after assumption checking, rather than as an automatic procedure whenever a normality test produces $p < .05$. The researcher should first determine whether the departure from normality is substantial enough to threaten the planned analysis and whether transformation is theoretically and statistically justified.

The primary aim of transformation is to make the data more compatible with the assumptions of the proposed statistical model while retaining as much of the substantive information in the original variable as possible. For instance, if the students' achievement scores are highly positively skewed, the researcher can consider of a suitable transformation and then reapply the normality test to the transformed data. If the transformed data exhibit an enhanced normality and the remaining assumptions of the intended t test, ANOVA, or ANCOVA are reasonably satisfied, the researcher can proceed with the parametric analysis. If the transformation does not adequately overcome the problem, the researcher should contemplate whether a robust or nonparametric alternative is more suitable than repeatedly transforming the data until a desired result is obtained (Schoder et al., 2006). The researcher should also report the transformation used because changing the scale of measurement can affect the interpretation and presentation of the results.

4. Conclusion

Selecting an appropriate statistical test in quantitative educational research requires a clear connection between the research hypothesis, characteristics of the variables, assumptions, and structure of the data. This paper has shown that the t test, ANOVA, ANCOVA, MANOVA, and MANCOVA serve different analytical purposes, and their selection should therefore be guided by the research question rather than by familiarity with statistical software. By following the sequence of formulating the hypotheses, identifying the variables involved, assessing normality and other relevant assumptions, and determining the number of groups, dependent variables, and covariates, novice education researchers can make more appropriate statistical decisions and produce more defensible interpretations of their findings.

Declarations

Acknowledgments

Not applicable.

Ethical Statement

This study was exempted from the ethics approval requirements since it does not involve human subjects or collection of sensitive personal data. The research is based on narrative review approach using scholarly published works. As such, it adheres to institutional guidelines that classify this type of study as low-risk and not subject to formal ethics approval.

Competing Interests

None.

Author's Contribution

ENK conceptualized the study, collected reviewed articles and books, prepare manuscript, editing, grammar checking and submitting manuscript.

References

- Almasi, M. (2026). From Classroom to Microlearning: Comparing Learner Perceptions of Traditional and Innovative Approaches in Lifelong and Adult Education in Tanzania. *Journal of Research, Innovation, and Strategies for Education (RISE)*, 3(4), 120–135. <https://doi.org/10.70148/rise.v3i4.8>
- Di Leo, G., & Sardanelli, F. (2020). Statistical significance: p value, 0.05 threshold, and applications to radiomics—reasons for a conservative approach. *European Radiology Experimental*, 4(1). <https://doi.org/10.1186/s41747-020-0145-y>
- Frane, A. V. (2021). Experiment-Wise Type I Error Control: A Focus on 2×2 Designs. *Advances in Methods and Practices in Psychological Science*, 4(1). <https://doi.org/10.1177/2515245920985137>
- Gravetter, F. J., & Wallnau, L. B. (2017). *Statistics for the Behavioral Sciences (10th ed)*. Cengage Learning.
- Gupta, S. C., & Kapoor, V. K. (2002). *Fundamentals of Mathematical Statistics (A Modern Approach) (10th ed.)*. Sultan Chand & Son Educational Publishers.
- Habibzadeh, F. (2024). Data Distribution: Normal or Abnormal? *Journal of Korean Medical Science*, 39(3), 1–8. <https://doi.org/10.3346/jkms.2024.39.e35>
- Hu, X., & Zainudin, Z. (2026). Assessing Student Stress Across Admissions Models: Mixed-Methods Evidence Comparing Portfolio and Standardized Testing Systems in Malaysia and China. *Journal of Research, Innovation, and Strategies for Education (RISE)*, 3(2), 1–11. <https://doi.org/10.70148/v3i2.1>
- Kim, J., Kim, D. H., & Kwak, S. G. (2024). Comprehensive guidelines for appropriate statistical analysis methods in research. *Korean Journal of Anesthesiology*, 77(5), 503–517. <https://doi.org/10.4097/kja.24016>
- Kim, T. K., & Park, J. H. (2019). More about the basic assumptions of t-test: Normality and sample size. *Korean Journal of Anesthesiology*, 72(4), 331–335. <https://doi.org/10.4097/kja.d.18.00292>
- Kuzma, J. W. (1998). *Basic Statistics for the Health Sciences (3rd ed.)*. Mayfield Publishing Company.
- Majeed, M. K., & Ahmad, T. B. (2025). Leveraging Digital Technologies for Informal Learning And Teacher Competency Development in Southern Pakistan. *Journal of Research, Innovation, and Strategies for Education (RISE)*, 2(3), 47–64. <https://doi.org/10.70148/rise.v2i3.4>
- Mijar, P. (2024). Selection of Statistical Test in Educational Research and Assessment. *Educational Review Journal*, 1, 140–150. www.ero.gov.np
- Najmi, A., Sadasivam, B., & Ray, A. (2021). How to choose and interpret a statistical test? An update for budding researchers. *Journal of Family Medicine and Primary Care*, 10(7), 2763–2767. [10.4103/jfmpe.jfmpe_433_21](https://doi.org/10.4103/jfmpe.jfmpe_433_21)
- Naskar, S., & Das, P. (2018). Application of different statistical tests in educational research: an overview. *Journal of Emerging Technologies and Innovative Research*, 5(5), 129–137. www.jetir.org
- Nawastheen, F. M., Liyanage, N. G. L. S. J., Sisindra, K. S. P., Fernando, W. M. A. P. S., & Sandamali, K. A. D. (2026). Learning Management Systems and Self-Regulated Learning Among Postgraduate Teacher-Students in Open and Distance Learning in the Sri Lankan Context. *Journal of Research, Innovation, and Strategies for Education (RISE)*, 3(6), 156–171. <https://doi.org/10.70148/rise.v3i4.10>
- Nawastheen, F. M., Sisindra, K. S. P., Liyanage, N. G. L. S. J., Fernando, W. M. A. P. S., & Dilakshan, R. . (2026). Student Teachers' Perspectives on Academic Support Services in Open and Distance Learning. *Journal of Research, Innovation, and Strategies for Education (RISE)*, 3(4), 156–171. <https://doi.org/10.70148/v3i1.2>
- Panos, G. D., & Boeckler, F. M. (2023). Statistical Analysis in Clinical and Experimental Medical Research: Simplified Guidance for Authors and Reviewers. *Drug Design, Development and Therapy*, 17(July), 1959–1961. <https://doi.org/10.2147/DDDT.S427470>
- Rahayu, N. I., Muktiarni, M., & Hidayat, Y. (2024). An Application of Statistical Testing: A Guide to Basic Parametric

- Statistics in Educational Research Using SPSS. *ASEAN Journal of Science and Engineering*, 4(3), 569–582. <https://doi.org/10.17509/ajse.v4i3.76092>
- Ranganathan, P. (2021). An introduction to statistics: Choosing the correct statistical test. *Indian Journal of Critical Care Medicine*, 25(S2), S184–S186. <https://doi.org/10.5005/JP-JOURNALS-10071-23815>
- Rashid, M. A. (2016). *Statistics for Scientists and Engineers*. University of Dar Es Salaam.
- Samuels, M. L., & Witmer, J. A. (2003). *Statistics for the life sciences* (3rd ed.). Prentice Hall.
- Schmidt, A. (2025). Addressing common inferential mistakes when failing to reject the null-hypothesis. *F1000Research*, 13, 1488. <https://doi.org/10.12688/f1000research.158434.3>
- Schoder, V., Himmelmann, A., & Wilhelm, K. P. (2006). Preliminary testing for normality: Some statistical aspects of a common concept. *Clinical and Experimental Dermatology*, 31(6), 757–761. <https://doi.org/10.1111/j.1365-2230.2006.02206.x>
- Sokal, R. R., & Rohlf, F. J. (2009). *Introduction to biostatistics* (2nd ed.). Dover Publications . Inc
- Sukhera, J. (2022). Narrative reviews: Flexible, rigorous, and practical. *Journal of Graduate Medical Education*, 14(4), 414–417. <https://doi.org/10.4300/JGME-D-22-00480.1>
- Tangkui, R. (2025). Integrasi Model Pembelajaran Kolb dengan Minecraft: Kesan Terhadap Pembelajaran Perimeter dan Luas. *Journal of Research, Innovation, and Strategies for Education (RISE)*, 2(5), 36–52. <https://doi.org/10.70148/rise.v2i5.3>
- Tsushima, E. (2022). Interpreting Results from Statistical Hypothesis Testing: Understanding the Appropriate P-value. *Physical Therapy Research*, 25(2), 49–55. <https://doi.org/10.1298/ptr.r0019>
- Vrbin, C. M. (2022). Parametric or nonparametric statistical tests: Considerations when choosing the most appropriate option for your data. *Cytopathology*, 33(6), 663–667. <https://doi.org/10.1111/cyt.13174>
- Yusof, I. J., Latif, A. A., & Supie, H. S. M. (2021). Assessing Statistical Literacy Level of Postgraduate Education Research Students in Malaysian Research Universities. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(5), 1318–1324. <https://doi.org/10.17762/turcomat.v12i5.1800>
- Zimmermann, G., Pauly, M., & Bathke, A. C. (2020). Multivariate analysis of covariance with potentially singular covariance matrices and non-normal responses. *Journal of Multivariate Analysis*, 177, 104594. <https://doi.org/10.1016/j.jmva.2020.104594>