METHODOLOGICAL REVIEW

A Decision Framework for Selecting, Applying, and Reporting Statistical Methods in Senior High School Quantitative Research

Mohamad T. Simpal and Arjey B. Mangakoy

Esperanza National High School, Sultan Kudarat, Philippines

Abstract

Statistical analysis is often taught as a sequence of software commands, yet defensible quantitative research depends first on alignment among the research question, study design, measurement level, sampling process, assumptions, and intended inference. This methodological review synthesizes the instructional content of Statistics Made Easy into a decision framework for senior high school research. The framework organizes analysis into six linked decisions: define the inferential target; identify the design and dependence structure; classify variables and measurement levels; select descriptive summaries; choose an inferential procedure; and report estimates, uncertainty, assumptions, and substantive meaning. It distinguishes frequency-based description, summaries of Likert-type responses, Pearson correlation, independent- and paired-samples t-tests, and one-way analysis of variance with multiplicity-controlled post-hoc comparisons. Particular attention is given to recurring errors: treating ordinal categories as automatically interval-scaled, selecting tests from software menus without checking design assumptions, reporting $p = 0.000$, equating statistical significance with practical importance, and interpreting association as causation. A compact selection matrix and reporting templates translate the framework into classroom and manuscript practice. The article argues that statistical literacy at the secondary level is strengthened when computation follows design reasoning and when results are communicated through effect estimates, confidence intervals, transparent assumption checks, and context-sensitive interpretation.

Keywords: quantitative research; statistical literacy; Likert-type data; Pearson correlation; t-test; analysis of variance; secondary education

1. Introduction

Quantitative research converts observations into evidence through a chain of decisions. A numerical output is not persuasive merely because it was produced by spreadsheet or statistical software; it becomes meaningful only when the analytic procedure corresponds to the research question, the way observations were obtained, the scale on which variables were measured, and the assumptions required for interpretation. For novice researchers, the most consequential difficulty is therefore not arithmetic but alignment.

Senior high school research courses frequently ask learners to describe survey responses, examine relationships, compare two groups, or compare three or more conditions. These tasks map to a small set of commonly used procedures, including frequency distributions, measures of central tendency and variability, correlation, t-tests, and analysis of variance (ANOVA). The apparent simplicity of this menu can be misleading. The same data set may call for different summaries depending on whether the goal is description or inference; two columns of scores may require an independent or paired comparison depending on how participants were sampled; and a significant omnibus ANOVA does not identify which means differ.

The instructional text Statistics Made Easy was developed to reduce anxiety around these procedures through stepwise explanations and computer applications (Simpal & Mangakoy, 2022). The present article recasts that material as a scholarly decision framework. Its contribution is not another list of tests, but an explicit reasoning sequence that begins with design and ends with transparent reporting. This orientation is consistent with long-standing statistical guidance that emphasizes estimation, uncertainty, and scientific context rather than mechanical thresholding of p-values (Wasserstein & Lazar, 2016; Wasserstein et al., 2019).

The article has three objectives: to integrate common secondary-level statistical tools into one coherent selection framework; to identify interpretive and reporting practices that improve methodological defensibility; and to provide concise templates that teachers and student-researchers can apply when writing results. The intended outcome is a bridge between accessible classroom instruction and the expectations of scholarly quantitative reporting.

2. Method of Synthesis

This article uses a structured conceptual synthesis of the seven instructional domains presented in Statistics Made Easy: statistical foundations and measurement; research design and sampling; descriptive statistics; weighted means and Likert-type responses; Pearson product-moment correlation; independent- and paired-samples t-tests; and one-way ANOVA with post-hoc analysis. The source material was reorganized around decisions that occur before, during, and after computation.

The synthesis followed three principles. First, software steps were translated into design-independent reasoning so that the framework remains applicable across spreadsheet and statistical packages. Second, procedural guidance was reconciled with widely accepted reporting principles: distinguish description from inference, state assumptions, report exact p-values where possible, quantify effect magnitude, and avoid causal language for nonexperimental associations. Third, the framework was bounded to methods commonly encountered in introductory quantitative projects. It is therefore a methodological and pedagogical article rather than an empirical evaluation; it makes no claims about observed student outcomes.

3. The Six-Decision Framework

Decision 1: Define the inferential target

Every analysis should begin with the claim the study is designed to support. A descriptive question asks what was observed in the sample. An associational question asks whether two variables vary together. A comparative question asks whether outcomes differ across groups or conditions. A causal question requires a design that supports causal identification; a correlation or group difference alone does not supply that warrant. Translating the research question into a statistical target prevents researchers from choosing a familiar test simply because it appears in a software menu.

Decision 2: Identify the design and dependence structure

The unit of analysis, sampling method, grouping structure, and repeated measurement pattern determine whether observations can reasonably be treated as independent. Scores from different participants in two groups typically support an independent comparison. Pretest and posttest scores from the same participants are paired. Multiple subject scores obtained from the same learners are repeated measures, not independent

groups. This distinction is fundamental because standard errors and test statistics depend on the covariance structure of observations.

Decision 3: Classify variables and measurement levels

Nominal variables encode categories without rank; ordinal variables encode order without guaranteeing equal distances; interval variables have interpretable equal units but no true zero; and ratio variables additionally possess a meaningful zero. These properties constrain valid summaries. Counts and proportions are natural for nominal data. Medians and ordered distributions are defensible for single ordinal items. Means and standard deviations are most readily interpreted for approximately continuous interval or ratio measurements. Multi-item Likert scales may sometimes be summarized with means when scale construction, reliability, and distributional behavior justify treating the composite as approximately continuous; that judgment should be stated rather than assumed (Sullivan & Artino, 2013).

Decision 4: Select descriptive summaries before testing

Descriptive analysis is not a preliminary formality. It reveals coding errors, sparse categories, outliers, skewness, ceiling or floor effects, and differences in spread that may alter subsequent choices. Categorical variables should be summarized with frequencies and percentages. Continuous variables should generally be described with the mean and standard deviation when distributions are reasonably symmetric, or the median and interquartile range when they are strongly skewed. Graphical inspection should accompany numerical summaries whenever possible.

Decision 5: Choose the inferential procedure

Research objective	Typical structure	Core procedure	Primary cautions
Describe categories	One categorical variable	Frequencies and percentages	Report denominator and missing data.
Summarize ordered ratings	Single Likert-type item	Distribution, median, and mode	Do not assume equal category intervals.
Summarize a composite scale	Multiple coherently scored items	Mean/SD or median/IQR	Justify scoring; assess reliability and shape.
Examine association	Two approximately continuous variables	Pearson correlation	Check linearity and influential points; no causal claim.
Compare two independent groups	Continuous outcome; distinct participants	Independent-samples t-test; Welch variant when appropriate	Check independence, distribution, outliers, and unequal variances.
Compare two paired measurements	Same or matched participants measured twice	Paired-samples t-test	Analyze difference scores; pairing must be preserved.
Compare three or more independent groups	One continuous outcome; one categorical factor	One-way ANOVA; Welch ANOVA when appropriate	Omnibus result precedes multiplicity-controlled contrasts.
Compare three or more repeated conditions	Same participants across conditions	Repeated-measures ANOVA or suitable alternative	Account for within-person dependence and sphericity.

Decision 6: Report estimates, uncertainty, and substantive meaning

A complete result communicates more than whether p is below .05. It identifies the sample and variables, reports descriptive statistics, gives the effect estimate, provides a confidence interval when available, names the test and relevant degrees of freedom, reports the exact p-value, and interprets the finding in relation to the research question. Values smaller than .001 should be reported as $p < .001$ rather than $p = .000$. Because statistical significance is influenced by sample size, practical interpretation should consider effect size, measurement quality, and context (Lakens, 2013).

4. Application to Common Statistical Procedures

4.1 Likert-Type Responses and Weighted Summaries

Likert-type items represent ordered response categories such as strongly disagree through strongly agree. The response codes indicate order, but the psychological distance between adjacent categories is not known to be equal. For a single item, the full distribution is often the most informative display because two groups can share the same mean while having very different response patterns.

Weighted means can be computed by multiplying each response value by its frequency, summing the products, and dividing by the total number of valid responses. The computation is straightforward, but its interpretation requires care. A mean based on numeric labels is most defensible when the categories are plausibly treated as equally spaced or when several items form a coherent composite scale. Researchers should disclose the anchors, scoring direction, number of items, handling of missing responses, and evidence supporting aggregation. Verbal interpretation bands should be established before examining outcomes and should not imply a precision greater than the instrument supports.

4.2 Pearson Correlation

Pearson's r quantifies the direction and strength of a linear relationship between two variables and ranges from -1 to +1. Its sign indicates direction; its absolute value indicates the degree of linear co-variation. Interpretation should be based on the scatterplot, the estimate, its confidence interval, and the measurement context—not on a universal adjective alone.

Pearson correlation assumes that the intended relationship is linear and can be distorted by influential observations or restricted ranges. Independence also matters: repeated measurements from the same participants cannot be treated as unrelated pairs. Most importantly, r does not establish that one variable causes the other. Reverse direction, omitted variables, measurement artifacts, and selection processes may account for the observed association.

4.3 Independent- and Paired-Samples t-Tests

The independent-samples t-test addresses a difference between the means of two distinct groups. The classical pooled-variance form assumes equal population variances. Welch's t-test relaxes that requirement and is generally preferable when group variances or sample sizes differ. A preliminary variance test should not be treated as an automatic gate that selects the test; the analytic choice should reflect design knowledge and the robustness of the procedure.

The paired-samples t-test evaluates the mean of within-pair differences. Its inferential unit is the difference score, so assumptions should be evaluated on those differences rather than on each measurement

separately. Correct pairing can substantially improve precision by removing stable between-person variation. Conversely, breaking the pairing produces an incorrect standard error and discards design information.

For both forms, a journal-style report includes group or time-point means and standard deviations, the estimated mean difference, a confidence interval, t, degrees of freedom, p, and an effect-size measure. The conclusion should describe the observed difference and its uncertainty rather than merely stating that the null hypothesis was “accepted” or “rejected.” Failure to reject is not proof of equality.

4.4 One-Way ANOVA and Post-Hoc Comparisons

One-way ANOVA evaluates whether a set of group means is compatible with a common population mean under the model. The omnibus F statistic compares between-group variation with within-group variation. A significant omnibus result indicates that the mean structure is not fully equal; it does not identify the groups responsible for the difference.

Planned contrasts derived from the research question are preferable when specific comparisons were defined in advance. When all pairwise comparisons are explored after the omnibus test, a multiplicity-control procedure such as Tukey’s method should be used under its applicable assumptions. If homogeneity of variance is doubtful, researchers may consider Welch’s ANOVA and a heteroscedastic post-hoc approach. For repeated observations, a repeated-measures model is required because ordinary one-way ANOVA assumes independent cases.

Reporting should include group descriptives, F with both degrees of freedom, the exact p-value, an effect-size estimate such as eta squared or omega squared, and confidence intervals where feasible. Post-hoc results should provide adjusted p-values and mean differences with uncertainty. Statements such as “all means were different” must be supported by the specific adjusted comparisons, not inferred from the omnibus test alone.

5. Assumptions as Evidence, Not Ritual

Assumption checking should answer whether the model is a reasonable approximation for the design and data, not whether a collection of preliminary tests crossed arbitrary thresholds. Independence is primarily a design property and cannot be repaired by a normality test. Linearity and influential points are best examined visually for correlation. For t-tests and ANOVA, the behavior of residuals or difference scores is more relevant than the marginal distribution of every raw variable.

Formal normality tests can be overly sensitive in large samples and underpowered in small samples. They should be interpreted alongside plots, sample size, balance, and the robustness of the chosen method. Similarly, unequal variances do not always invalidate mean comparisons, but severe heteroscedasticity combined with unequal group sizes can distort conventional tests. Robust variants, transformations justified by the measurement scale, or carefully chosen nonparametric procedures may be appropriate. Any deviation from the prespecified analysis should be documented transparently.

6. From Software Output to Scholarly Results

Software output is an intermediate artifact, not a results section. A publishable narrative selects only the quantities needed to answer the research question and explains what those quantities mean. Screenshots of dialog boxes and raw output tables should normally be replaced by editable, purpose-built tables. Variable labels, units, valid sample sizes, and missing-data rules should be explicit.

Procedure	Compact reporting template
Pearson correlation	X and Y were [positively/negatively] associated, r(df) = value, 95% CI [lower, upper], p = value. The association was interpreted as noncausal.
Independent t-test	Group A (M = , SD =) and Group B (M = , SD =) differed by ΔM = , 95% CI [,], Welch's t(df) = , p = , effect size = .
Paired t-test	Scores changed from Time 1 (M = , SD =) to Time 2 (M = , SD =); mean paired difference = , 95% CI [,], t(df) = , p = , effect size = .
One-way ANOVA	Mean outcomes differed across groups, F(df1, df2) = , p = , effect size = . Adjusted comparisons showed [specific contrasts], with mean differences and confidence intervals.

Good interpretation separates three questions: What was estimated? How uncertain is the estimate? Why does the magnitude matter in the study context? This sequence protects against the common mistake of converting $p < .05$ into a claim of educational importance. It also enables readers to evaluate findings when an estimate is imprecise or when a nonsignificant result remains compatible with effects that matter.

7. Pedagogical Implications

A design-first framework changes the order of instruction. Learners first classify the question, variables, and dependence structure; then predict the appropriate output; only afterward do they use software. This sequence makes errors visible before computation. Worked examples should include near-miss cases: paired data incorrectly treated as independent, an ordinal item summarized only by a mean, a curved relationship summarized by Pearson's r, and an omnibus ANOVA followed by unadjusted pairwise tests.

Assessment should reward reasoning as well as numerical accuracy. A student who selects a robust procedure, states assumptions, reports uncertainty, and uses cautious language demonstrates deeper competence than one who reproduces a software table. Teachers can reinforce this competence through analysis plans completed before data collection, annotated output exercises, reproducible calculation files, and peer review focused on the connection between question, design, method, and claim.

The framework is also useful for research advising. Rather than asking "Which test should I use?" in isolation, advisers can ask six sequential questions: What is the target claim? How were observations generated? What are the variable types? What do the data look like? Which model matches those features? What evidence must be reported? The resulting conversation is both more teachable and more scientifically defensible.

8. Conclusion

Statistical literacy in senior high school research is best developed as disciplined decision-making rather than software operation. Measurement, sampling, design, descriptive analysis, inferential procedure, and reporting form one connected chain. Frequencies, weighted summaries, correlation, t-tests, and ANOVA remain valuable introductory tools when their assumptions and inferential boundaries are explicit. The proposed six-decision framework converts these procedures into a coherent workflow: define the target, identify dependence, classify variables, describe the data, select the procedure, and report estimates with uncertainty and substantive interpretation. Applied consistently, this workflow can improve the clarity, reproducibility, and scholarly credibility of student quantitative research.

References

American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. Educational Researcher, 35(6), 33–40. https://doi.org/10.3102/0013189X035006033

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Frontiers in Psychology, 4, Article 863. https://doi.org/10.3389/fpsyg.2013.00863

Simpal, M. T., & Mangakoy, A. B. (2022). Statistics made easy: Commonly used statistical tools and their computer applications in quantitative research for senior high school. [Instructional book].

Sullivan, G. M., & Artino, A. R., Jr. (2013). Analyzing and interpreting data from Likert-type scales. Journal of Graduate Medical Education, 5(4), 541–542. https://doi.org/10.4300/JGME-5-4-18

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. The American Statistician, 70(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond “p < 0.05.” The American Statistician, 73(sup1), 1–19. https://doi.org/10.1080/00031305.2019.1583913

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. American Psychologist, 54(8), 594–604. https://doi.org/10.1037/0003-066X.54.8.594