

Interpreting t-Test Results: A Practical Comparison of Manual Computation and SPSS Output for Researchers

NAZIRU AMINU¹, ISHAKU AKISSANI²

¹*Department of Mathematics and Computer Science Education, Umaru Musa Yar'adua University, Katsina, Katsina State, Nigeria.*

²*Department of Mathematics, School of Secondary Education Sciences, Isa Kaita College of Education, Dutsin-ma, Katsina State, Nigeria.*

Abstract- *Accurate interpretation of statistical results is essential for credible research, particularly when employing inferential tools like the t-test to compare group means. Despite the increasing use of statistical software such as SPSS, many researchers and supervisors struggle to fully understand how these results are generated and what they imply. This seminar paper, titled "Interpreting t-Test Results: A Practical Comparison of Manual Computation and SPSS Output for Researchers", provides a detailed comparison between manual t-test computation and SPSS-generated results. It explores the underlying statistical concepts, outlines step-by-step procedures for both methods, and demonstrates how to apply decision rules using critical values and p-values. By highlighting the similarities and differences in output interpretation, this paper aims to improve the statistical literacy of researchers and enhance their ability to critically evaluate SPSS outputs. Practical examples and recommendations are provided to support better research practices and supervision.*

Keywords: *T-test, Manual Computation, SPSS Output.*

I. INTRODUCTION

Background to the Study

The t-test, introduced by William Sealy Gosset under the pseudonym "Student," is a fundamental statistical tool used to determine if there is a significant difference between the means of two groups. It is particularly useful when dealing with small sample sizes and unknown population variances. In educational research, the t-test is invaluable for assessing the effectiveness of teaching methods, interventions, and policies. Akpan and Clark (2023) assert, "The t-test is a crucial statistical instrument that provides a methodical and rigorous way to assess the effects of interventions, teaching strategies, and

educational policy." Its application allows researchers to make evidence-based decisions that can enhance educational outcomes. With the advent of statistical software like IBM SPSS, conducting t-tests has become more accessible. SPSS automates complex calculations, providing users with outputs that include test statistics, p-values, and confidence intervals. Pallant (2020) notes, "SPSS provides a quick, user-friendly way to run complex statistical analyses with large datasets." However, this convenience may lead to a superficial understanding of the underlying statistical concepts, potentially resulting in misinterpretation of results. Manual computation of t-tests, though more time-consuming, offers researchers a deeper insight into the mechanics of the test. Gravetter and Wallnau (2017) emphasize that manual calculations help users appreciate the influence of sample size, variance, and mean differences on the test outcome. This hands-on approach reinforces the importance of meeting test assumptions and understanding the derivation of critical values and decision rules. Despite the efficiency of SPSS, a balanced approach that combines both manual and software-based methods is recommended. Understanding both approaches ensures that researchers can validate software outputs, recognize potential anomalies, and make informed decisions based on a comprehensive understanding of the statistical procedures involved.

This paper aims to bridge the knowledge gap by providing a practical comparison between manual t-test computations and SPSS-generated outputs. By elucidating the similarities and differences in procedures, interpretations, and decision-making processes, the paper seeks to enhance the statistical

literacy of researchers and supervisors, ultimately leading to more robust and credible research outcomes.

II. IMPORTANCE OF STATISTICAL ANALYSIS IN RESEARCH

Statistical analysis is a critical component of the research process, serving as the bridge between raw data and meaningful conclusions. It helps researchers summarize data, test hypotheses, and make evidence-based decisions that are essential for the credibility and generalizability of research findings. According to Creswell and Creswell (2018), statistical analysis enables researchers to “make sense of the data by identifying patterns, relationships, and trends that support or refute research hypotheses.” It provides the tools necessary to handle both descriptive and inferential tasks such as summarizing data distributions, comparing group outcomes, and determining the likelihood that findings occurred by chance.

In educational research, for instance, statistical techniques like the t-test, ANOVA, regression analysis, and correlation are used to assess the effectiveness of interventions, compare teaching methods, or measure student performance across different groups (Fraenkel, Wallen, & Hyun, 2019). Without statistical analysis, researchers would rely on subjective judgments, which could lead to biased or inaccurate conclusions. Statistical tools also help to control for confounding variables and establish the validity and reliability of results. As noted by Gravetter and Wallnau (2017), “statistical reasoning allows researchers to distinguish meaningful effects from random variability.” This ensures that findings are not only mathematically sound but also practically significant. Moreover, in an era of data-driven decision-making, statistical literacy is increasingly regarded as a key competency in both academic and professional contexts. Pallant (2020) emphasizes that statistical analysis empowers researchers to communicate their findings more convincingly to policymakers, stakeholders, and other academic audiences.

In sum, statistical analysis is indispensable in research as it transforms data into actionable

knowledge. It promotes objectivity, supports scientific inquiry, and contributes to the development of evidence-based practices across disciplines.

III. PURPOSE OF THE STUDY

The primary purpose of this seminar paper is to explore and clarify the differences between manual computation and SPSS-generated output of the t-test, with the goal of enhancing the statistical literacy of researchers, postgraduate students, and research supervisors. In many academic and professional settings, statistical software such as SPSS is widely used to perform data analysis due to its speed and ease of use. However, reliance on software without a clear understanding of the underlying procedures may lead to misinterpretation of results or incorrect application of statistical techniques.

This seminar aims to demystify the process of conducting a t-test by comparing the manual step-by-step method with the output generated by SPSS. Through this comparison, the seminar will:

1. Demonstrate the computation of t-values, degrees of freedom, and decision rules using both manual and software methods.
2. Highlight the similarities and differences between manual results and SPSS output in terms of interpretation and reporting.
3. Promote a deeper understanding of key concepts such as p-values, critical values, significance levels, and decision making in hypothesis testing.
4. Empower researchers and supervisors to validate software results by understanding the logic behind statistical conclusions.
5. Encourage a balanced approach that combines theoretical knowledge with the practical advantages of statistical tools.

Ultimately, the seminar seeks to bridge the gap between statistical computation and interpretation, enabling participants to make more informed, accurate, and confident decisions in their research work.

IV. SCOPE AND LIMITATIONS

Scope of the Study

This paper focuses specifically on the interpretation and application of the independent samples t-test as conducted through manual computation and SPSS software. The comparison is limited to the independent t-test because it is one of the most commonly used statistical tests in educational and social science research, especially when comparing the means of two independent groups.

V. LIMITATIONS OF THE STUDY

While the paper aims to offer a comprehensive comparison, there are certain limitations:

1. **Focus on One Type of t-Test:** The discussion is limited to the independent samples t-test. Other variants, such as the paired samples t-test and one-sample t-test, are not covered in detail.
2. **Software Limitation:** SPSS is the only software considered in this paper. Alternative platforms like R, STATA, or Excel are not included in the comparison.
3. **Mathematical Simplification:** The manual computation process is simplified for instructional purposes. Advanced mathematical derivations and proofs are not the focus.
4. **Sample Size and Complexity:** The examples used in the comparison involve small, hypothetical datasets. The challenges and nuances of working with large or real-world datasets may not be fully captured.
5. **Assumption Testing:** While basic assumptions of the t-test are mentioned, the seminar does not provide an in-depth analysis of how assumption violations affect results in both manual and software-based methods.

Despite these limitations, the study offers valuable insights into how t-tests function, how SPSS presents results, and how to interpret those results accurately and responsibly.

VI. CONCEPTUAL CLARIFICATIONS

Definition of t-Test

The t-test is a statistical method used to determine whether there is a significant difference between the means of two groups, which may be related in certain features. It is particularly useful when sample sizes are small and the population standard deviation is unknown. The t-test helps to infer whether observed differences in sample means reflect true differences in the populations or are due to random chance. According to Gravetter and Wallnau (2017), the t-test "is designed to test hypotheses about the difference between two population means when the values for the population standard deviation are unknown and must be estimated from the sample data." Similarly, Field (2018) describes the t-test as "a parametric test that compares the means of two groups and determines whether any observed difference is statistically significant."

The formula for the t-test statistic generally involves the difference between sample means divided by the estimated standard error of the difference, reflecting how much difference is expected by chance (Bluman, 2018).

Types of t-Tests (with Emphasis on Independent Samples t-Test)

The t-test encompasses several variants; each suited to specific research designs and data characteristics. The three main types of t-tests are:

1. One-Sample t-test

This test compares the mean of a single sample to a known population mean or a hypothesized value. It determines whether the sample mean significantly differs from the population mean (Gravetter & Wallnau, 2017). This is useful in cases where the researcher wants to assess if a group's performance deviates from a standard or benchmark.

2. Paired Samples t-Test (Dependent t-Test)

Also known as the dependent t-test, this type compares means from the same group under two different conditions or at two time points. It accounts for the fact that the two sets of scores are related or "paired" (Field, 2018). Examples include pre-test/post-test designs or measuring the effect of an intervention on the same participants.

3. Independent Samples t-Test

The independent samples t-test is used to compare the means of two distinct, unrelated groups to determine if there is a statistically significant difference between them (Creswell & Creswell, 2018). It assumes that the two samples are independent, meaning the scores in one group do not affect the scores in the other.

Assumptions for Applying Parametric Statistics

1. The data is assumed to be normally distributed (Bell shape, symmetrical about the mean, skewed, Kolmogorov Smirnov \geq or Shapiro $<=$) (mesocurtic)
2. The data is assumed to be continuous hence the measurement is on interval or ratio scale (data that can take fractions)
3. The measure of central tendency of the data is mean
4. The variances of the groups should be approximately the same
5. The data sample should have been drawn from the random sampling
6. Needs to investigate if there is significant difference between two independent means.

Manual Computation

7.5. Independent Samples t-test

In using independent samples t-test, the researcher must have data from two and only two separate groups, e.g., Males vs. Females, SS1 vs. SS2, Experimental group vs. Control group, Public schools vs. Private schools, etc. The mean scores of the two groups can then be compared using the independent samples t-test. Manual computations of independent samples t-test can be done using the formula:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SEM}$$

Where \bar{x}_1 is the mean of group 1, \bar{x}_2 is the mean of group 2, and SEM is the standard error of the mean, which is given by the formula:

$$SEM = \sqrt{\left(\frac{sd^2x_1}{n_1}\right) + \left(\frac{sd^2x_2}{n_2}\right)}$$

Where sd^2x_1 is the variance of group 1, sd^2x_2 is the variance of group 2, n_1 is the number of subjects in group 1 and n_2 is the number of subjects in group 2.

When we combine these two formulae, we have the universal formula as follows:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{sd^2x_1}{n_1}\right) + \left(\frac{sd^2x_2}{n_2}\right)}}$$

The degree of freedom (df) is given by $(n_1 + n_2 - 2)$.

Example 23

Suppose a researcher want to investigate the effect of Virtual Laboratory on students' academic achievement in Chemistry practical, he randomly assigned twenty (24) subjects to one of the two groups; control and experimental; those that will be taught Chemistry practical using Virtual Laboratory for three days a week for six weeks, and those that will be taught same concepts with conventional method (lecture). Before completing the study, one of the subjects in the experimental group and two from the control group dropped out. At the end of the treatment the two groups were tested, and the scores for the remaining subjects were recorded as follows:

| | | | | | | | | | | | |
|--------------|----|----|----|----|----|----|----|----|----|----|----|
| Experimental | 27 | 18 | 25 | 28 | 15 | 32 | 27 | 33 | 29 | 30 | 20 |
| Control | 25 | 22 | 19 | 25 | 24 | 26 | 24 | 27 | 28 | 26 | - |

- a. State formally the hypothesis necessary to conduct a non-directional test of no difference between the two group means.
- b. Compute the t-test analysis at 5% level of significance.
- c. State your conclusion (Take $t_{crit} = 1.766$)

| Exp. Group (x_1) | Ctrl. Group (x_2) | $x_1 - \bar{x}_1$ | $(x_1 - \bar{x}_1)^2$ | $x_2 - \bar{x}_2$ | $(x_2 - \bar{x}_2)^2$ |
|----------------------|-----------------------|-------------------|-----------------------|-------------------|-----------------------|
| 27 | 25 | 1.18 | 1.39 | 0.40 | 0.16 |
| 18 | 22 | -7.82 | 61.15 | -2.60 | 6.76 |
| 25 | 19 | -0.82 | 0.67 | -5.60 | 31.36 |
| 28 | 25 | 2.18 | 4.75 | 0.40 | 0.16 |
| 15 | 24 | -10.82 | 117.07 | -0.60 | 0.36 |
| 32 | 26 | 6.18 | 38.19 | 1.40 | 1.96 |
| 27 | 24 | 1.18 | 1.39 | -0.60 | 0.36 |
| 33 | 27 | 7.18 | 51.55 | 2.40 | 5.76 |
| 29 | 28 | 3.18 | 10.11 | 3.40 | 11.56 |
| 30 | 26 | 4.18 | 17.47 | 1.40 | 1.96 |
| 20 | - | -5.82 | 33.87 | - | - |
| $\Sigma = 284$ | $\Sigma = 246$ | | $\Sigma = 337.61$ | | $\Sigma = 60.40$ |

We get the means and variances for the two groups as follows:

$$\bar{x}_1 = \frac{\Sigma x_1}{n_1} = \frac{284}{11} = 25.82$$

$$\bar{x}_2 = \frac{\Sigma x_2}{n_2} = \frac{246}{10} = 24.60$$

Also, we have:

$$sd^2x_1 = \frac{\Sigma(x_1 - \bar{x}_1)^2}{n_1 - 1} = \frac{337.61}{11 - 1} = \frac{337.61}{10} = 33.76$$

$$sd^2x_2 = \frac{\Sigma(x_2 - \bar{x}_2)^2}{n_2 - 1} = \frac{60.40}{10 - 1} = \frac{60.40}{9} = 6.71$$

We now apply the formula as follows:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{sd^2x_1}{n_1}\right) + \left(\frac{sd^2x_2}{n_2}\right)}} = \frac{25.82 - 24.60}{\sqrt{\left(\frac{33.76}{11}\right) + \left(\frac{6.71}{10}\right)}}$$

$$t = \frac{1.22}{\sqrt{(3.069) + (0.671)}} = \frac{1.22}{\sqrt{(3.740)}} = \frac{1.22}{1.934} = 0.631$$

- c. Now since t_{cal} is less than t_{crit} (i.e. $t_{cal} = 0.631 < t_{crit} = 1.766$) we accept the null hypothesis. Therefore, we conclude that there is no significant difference in students' mean academic achievement scores between those in the experimental group and those in the control group.

SPSS Computation

| GROUPS | SCORES |
|--------|--------|
| 1.00 | 27.00 |
| 1.00 | 18.00 |
| 1.00 | 25.00 |
| 1.00 | 28.00 |
| 1.00 | 15.00 |
| 1.00 | 32.00 |
| 1.00 | 27.00 |
| 1.00 | 33.00 |
| 1.00 | 29.00 |
| 1.00 | 30.00 |
| 1.00 | 20.00 |
| 2.00 | 25.00 |
| 2.00 | 22.00 |
| 2.00 | 19.00 |
| 2.00 | 25.00 |
| 2.00 | 24.00 |
| 2.00 | 26.00 |
| 2.00 | 24.00 |
| 2.00 | 27.00 |
| 2.00 | 28.00 |
| 2.00 | 26.00 |

| GROUPS | N | Mean | Std. Deviation | Std. Error Mean |
|---------------------|----|---------|----------------|-----------------|
| SCORES EXPERIMENTAL | 11 | 25.8182 | 5.81065 | 1.75198 |
| CONTROL | 10 | 24.6000 | 2.59058 | .81921 |

Independent Samples Test

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|--------|-----------------------------|---|------|------------------------------|--------|-----------------|-----------------|-----------------------|---|---------|
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| | | | | | | | | Lower | Upper | |
| SCORES | Equal variances assumed | 5.625 | .028 | .609 | 19 | .550 | 1.21818 | 1.99985 | -2.96756 | 5.40392 |
| | Equal variances not assumed | | | .630 | 14.102 | .539 | 1.21818 | 1.93405 | -2.92712 | 5.36349 |

VII. IMPLICATIONS FOR RESEARCHERS AND SUPERVISORS

Understanding the differences between manual t-test calculations and SPSS outputs holds several important implications for both researchers and supervisors involved in empirical studies.

1. Enhanced Statistical Literacy

Researchers gain a deeper appreciation of the t-test procedure when they comprehend how values like the t-statistic, degrees of freedom, and p-values are derived manually. This foundational knowledge equips them to critically evaluate SPSS outputs rather than accepting results at face value. Supervisors, in

turn, are better positioned to guide students and junior researchers toward sound statistical reasoning and proper interpretation of results.

2. Improved Accuracy in Interpretation

The manual calculation process clarifies key concepts such as the decision rule based on critical values and significance levels. Recognizing these principles helps avoid common errors like misreading p-values or misunderstanding assumptions underlying the test. Supervisors can thus ensure that research conclusions are drawn on a firm statistical basis, improving the validity and reliability of final reports.

3. Validation of Software Results

While SPSS streamlines statistical analysis, it may sometimes produce output that appears confusing or contradictory due to default settings (e.g., equal variances assumed vs. not assumed). Researchers knowledgeable in manual computation can cross-check software results to detect discrepancies or anomalies, reducing the risk of incorrect conclusions.

4. Informed Decision-Making

Understanding both manual and software methods fosters informed decision-making about research design, data analysis, and result reporting. Supervisors can better evaluate whether researchers have met test assumptions and chosen appropriate statistical techniques, thus enhancing the overall quality of academic work.

5. Bridging Theory and Practice

This dual understanding promotes a balanced approach where theoretical knowledge supports practical application. For researchers, it transforms statistical analysis from a mechanical task into an insightful process that enriches research interpretation. Supervisors can thus cultivate more competent and confident researchers capable of conducting rigorous quantitative analysis.

VIII. CONCLUSION

This study has highlighted the critical differences and complementary roles of manual t-test computation and SPSS output in research analysis. While SPSS offers efficiency and ease in performing statistical tests, understanding the underlying manual calculations is essential for accurate interpretation, validation, and application of results. Researchers and supervisors equipped with both theoretical knowledge and practical skills can make more informed decisions, ensure the integrity of their analyses, and confidently communicate their findings.

The comparison demonstrates that reliance solely on software without grasping the fundamental principles may lead to misinterpretation or misuse of statistical results. Therefore, fostering a strong foundation in manual computation alongside proficiency in statistical software is key to enhancing research quality and rigor.

IX. RECOMMENDATIONS

1. Integrate Statistical Theory with Software Training

Educational institutions and research supervisors should emphasize teaching both the conceptual foundations of statistical tests and the practical use of software like SPSS. This dual approach ensures comprehensive understanding and reduces errors in analysis.

2. Encourage Manual Verification of Software Results

Researchers should be encouraged to manually compute critical values or at least understand the process behind the computations to cross-verify SPSS outputs, especially when results appear unexpected or ambiguous.

3. Develop Workshops and Training Sessions

Institutions should organize regular workshops focusing on statistical concepts, manual calculations, and software interpretation. These sessions can build capacity among students and early-career researchers.

4. Promote Critical Thinking in Statistical Interpretation

Supervisors should foster an environment where questioning and critically evaluating software outputs is standard practice. This will improve the overall quality of research outputs.

5. Update Curriculum to Include Practical Comparisons

Curriculum designers in research methodology courses should incorporate exercises that require students to perform manual calculations and compare them with software results to deepen understanding.

REFERENCES

- [1] Akpan, E. E., & Clark, J. L. (2023). Independent t-test statistics: Its relevance in educational research. *GASPRO International Journal of Eminent Scholars*, 10(1). Retrieved from <https://globalacademicstar.com/article/indepen>

dent-t-test-statistics-it-s-relevance-in-
educational-researchGlobal Academic Star

- [2] Bluman, A. G. (2018). *Elementary Statistics: A Step by Step Approach* (10th ed.). McGraw-Hill Education.
- [3] Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.
- [4] Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5th ed.). SAGE Publications.
- [5] Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2019). *How to design and evaluate research in education* (10th ed.). McGraw-Hill Education.
- [6] Gravetter, F. J., & Wallnau, L. B. (2017). *Statistics for the Behavioral Sciences* (10th ed.). Cengage Learning.
- [7] Pallant, J. (2020). *SPSS Survival Manual: A Step-by-Step Guide to Data Analysis Using IBM SPSS* (7th ed.). Open University Press.
- [8] Ruth, M. (2024). Student's t-test. EBSCO Research Starters. Retrieved from <https://www.ebsco.com/research-starters/social-sciences-and-humanities/students-t-tes>