Applied Medical Informatics
Original Research Vol. 47, No. 4/2025, pp: 139-149

Parametric Post-Hoc Tests: A Refined Examination of Significance and Errors in Small Sample Sizes

Alexandru D. COSTIN* **and Andreea D. ONA**

Department of Genetics and Plant Breeding, University of Agricultural Sciences and Veterinary Medicine, 3-5 Mănăştur Street, 400372, Cluj-Napoca, Romania.
E-mails: alexdan.costin@gmail.com; andreea.ona@usamvcluj.ro

* Author to whom correspondence should be addressed.

Received: 11 July 2025/Accepted: 15 November 2025/ Published online: 12 December 2025

Abstract
Introduction: Post-hoc analysis is the process of examining data after an initial statistical test conducted on more than two groups, such as ANOVA or Kruskal-Wallis tests. *Aim*: This study aimed to perform a comparative analysis of ten parametric post-hoc tests, used in eight different scenarios, with a focus on small sample sizes, regarding the occurrence of type I and type II errors. *Methods*: An analytical study was conducted in which we evaluated 10 parametric post-hoc tests, namely Duncan, Tukey HSD, Bonferroni, Fisher's LSD, Scheffé, Dunnett, Šidák, Gabriel, Games-Howell and Tamhane T2, using a personal database. The dataset contained synthetic values based on experimental data, and a Monte Carlo-like simulation approach was used to simulate different experimental scenarios. For each set of variables, data was randomly modified based on real values, with adjustments made (either by adding or subtracting) to meet the specified criteria of interest (normality, equal/unequal variance and group size, difference/no difference between groups, with ANOVA tests p-value distant or closer to the significance threshold). *Results*: According to the scenario where no differences between groups were found, variance was equal, and the p-value from ANOVA tests was close to the significance threshold, both Duncan and Fisher's LSD post-hoc tests presented type I errors in both equal and unequal group sizes. In contrast, the Dunnett test showed type I errors only in the case of equal group sizes. For the same scenario, but in cases where variance was unequal, all tests presented type I error except for Tamhane T2. *Conclusion*: For small sample sizes, equal variances and p-value close to the significance threshold, Tukey HSD and Bonferroni presented good control of type I errors, while Tamhane T2 was appropriate for unequal variance.

Keywords: Post-hoc analysis; Multiple groups comparison; Post-hoc tests; Test efficiency, Errors.

Introduction

In statistical analysis, a further exploration of group differences typically follows an initial test that indicates or not a significant outcome. Post hoc analysis represents the process of examining data after an initial statistical test, such as ANOVA for parametric or Kruskal-Wallis for non-parametric categories, which indicates significant differences between more than two groups [1,2]. The term "post-hoc", meaning "after the fact", was used when there were no predefined hypotheses regarding which groups might differ, before the analysis [1]. The ANOVA or Kruskal-Wallis (KW) test only shows whether there is at least one significant difference, but not where that difference occurred [3,4]. When significant results are obtained from these tests (ANOVA and KW), parametric (tests like Tukey or Dunnett) or non-parametric post-hoc tests (tests like Nemenyi or Dunn – Figure 1) are used to perform pairwise comparisons between the groups included in the study.

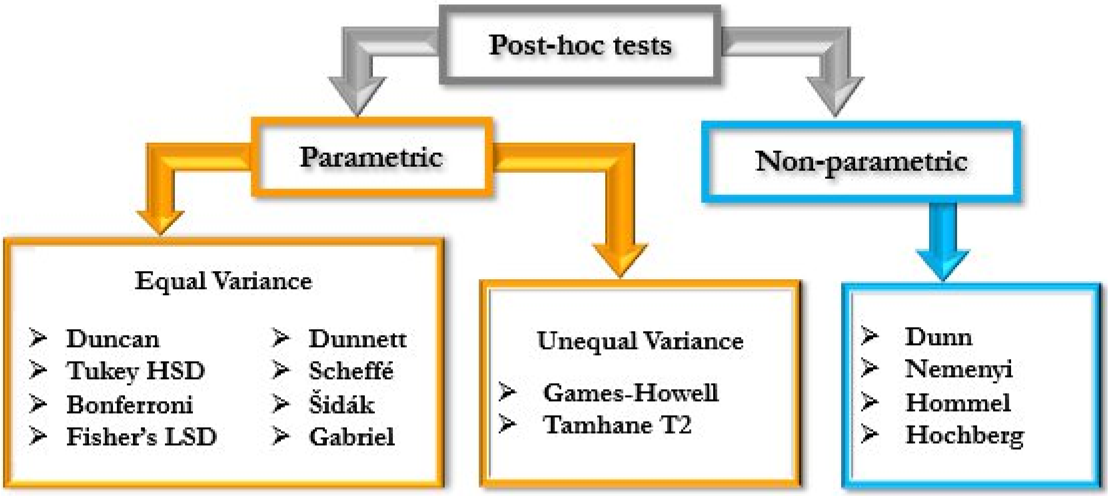

Figure 1. Parametric post-hoc tests

(Original design)

Post-hoc analysis is an exploratory approach, enabling researchers to investigate the data further and highlight the specific group comparisons that contributed to the overall effect observed [4]. Following the main steps of the analysis is essential for ensuring the validity and accuracy of the results, as shown in Figure 2. Moreover, careful consideration of post-hoc test selection based on sample size, variance equality, and error control is critical to avoid misleading conclusions and to improve the robustness of the findings.

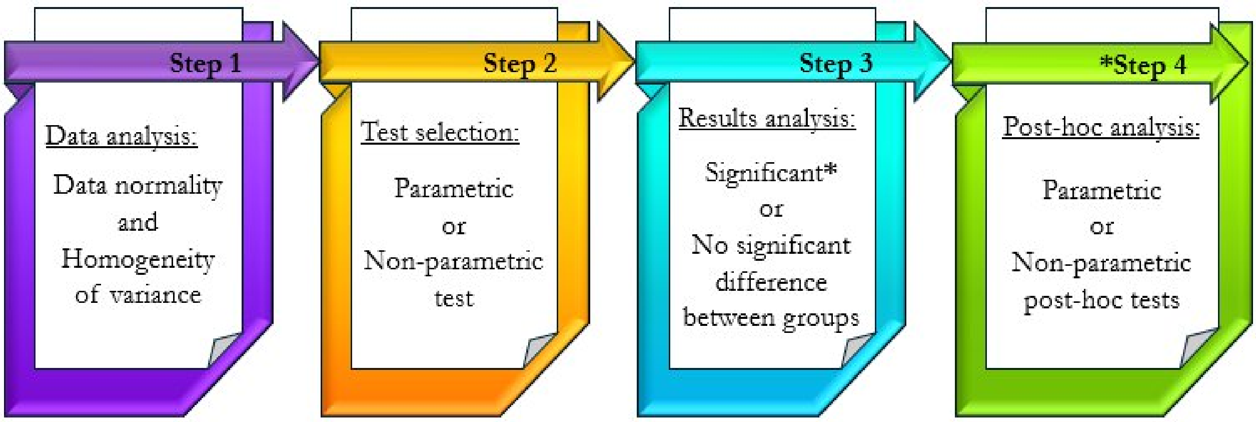

Figure 2. The analysis main steps.

(Original design)

Correct application of post-hoc tests is essential for a valid analysis [5,6]. Improper application of these tests can affect the integrity of the research and lead to inaccurate conclusions. Considering that post-hoc tests often involve multiple comparisons, they increase the risk of type I errors (known as false positives), and if adjustments are not made to control these errors, the results can become invalid [3,7]. To reduce this risk, post-hoc methods such as the Tukey test (Table 1) are commonly used [8,9]. These methods control the inflation of error rates, ensuring that the conclusions are valid and reliable.

Post-hoc tests vary in their characteristics, with some prioritizing conservatism to minimize false positives (e.g., Scheffé's test), while others, like Tukey's HSD, offer a more liberal approach that balances error control with statistical power. Understanding these differences is crucial for selecting the most appropriate test based on the specific goals and data structure of the study [7].

Table 1. Post-hoc tests characteristics

No.	Post-Hoc Test	Tests' characteristics	Reference
1.	Duncan	• Higher type I error risk. • Affected by unequal group sizes.	[10,12]
2.	Tukey HSD	➢ Lower type I error risk. ➢ Robust to unequal group sizes.	[3,10,12]
3.	Bonferroni	• Lower type I error risk. • Sensitive to unequal group size.	[10,12,13]
4.	Fisher's LSD	➢ Higher type I error risk. ➢ Affected by unequal group sizes.	[10,12,14]
5.	Scheffé	• Lower type I error risk. • Less affected by unequal group sizes.	[3,10,12]
6.	Dunnett	➢ Lower type I error risk. ➢ Sensitive to unequal group sizes.	[3,10]
7.	Šidák	▪ Medium type I error risk. ▪ Affected by unequal group sizes.	[11,12,15]
8.	Gabriel	➢ Susceptible to type I errors. ➢ Less affected by unequal group sizes.	[10,11,16]
9.	Games-Howell	• Lower type I error risk. • Less affected by unequal group sizes.	[10,11,12]
10.	Tamhane T2	➢ Lower type I error risk. ➢ Affected by unequal group sizes.	[10,11]

Overall, post-hoc tests are an essential tool for refining initial findings and providing a more detailed understanding of group interactions. Post-hoc tests are also characterized by essential elements such as error type (I – false positive or II – false negative) and significance. Different findings were reported by researchers, highlighting the complexity of the issue (Table 2).

Table 2. Studies reporting findings after post-hoc test analyses

First author et al. [ref]	Sample size	Post-hoc test used	Activity reported
Juarros-Basterretxea et al. [4]	Multiple sample sizes (both small and large)	Fisher's LSD, Bonferroni, Šidák, Scheffé, Tukey HSD, Gabriel, Hochberg's GT2	In cases of violated homoscedasticity and balanced groups, the Bonferroni, Šidák, Hochberg GT2, and Gabriel tests showed great accuracy, exceeding the Tukey HSD and Scheffé tests, presenting a lower error rate.
Shingala and Rajyaguru [7]	Large sample sizes	Games-Howell, Tamhane T2, Dunnett T3, Dunnett C	The Games-Howell method outperformed Tamhane's T2 when the confidence interval ratio was below 1, also showing narrower confidence intervals compared to those from Tamhane's T2, Dunnett's T3 and Dunnett's C

Type I error is represented by the rejection of a true null hypothesis, meaning an incorrect conclusion that there is a significant difference when, in fact, no difference exists [10,11]. Type II error is represented by the acceptance of a false null hypothesis, meaning that the procedure failed to detect a significant difference when, in reality, there is a difference [10,11]. Significance refers to the probability that the differences between groups are not due to chance. All post-hoc tests provide a p-value that indicates significance. If the p-value is less than the chosen significance level (usually 0.05), the differences are considered statistically significant [11]. In the existing literature, few analyses regarding post-hoc tests, particularly in the context of their application to small sample sizes and the nuanced examination of significance and errors were performed. While many studies have explored post-hoc testing in general, there is a lack of in-depth analysis focused specifically on how these tests perform under the constraints of small sample sizes. Some research has concentrated on larger datasets, leaving the challenges of small sample sizes underexplored, especially when it comes to refining the choice of tests based on error rates and type I/II errors. Given the complexity of small sample environments, a more targeted investigation into the efficacy and accuracy of various post-hoc tests is needed, particularly to understand their behavior in conditions of unequal variances and when traditional assumptions no longer hold.

This study aimed to perform a comparative analysis of ten parametric post-hoc tests, used in eight different scenarios, with a focus on small sample sizes, regarding the occurrence of type I and type II errors.

Materials and Methods

Study Design

An analytical study was conducted in which we evaluated 10 parametric post-hoc tests (Duncan, Tukey HSD, Bonferroni, Fisher's LSD, Scheffé, Dunnett, Šidák, Gabriel, Games-Howell, and Tamhane T2) (Table 1), using a personal database containing synthetic values based on experimental data for polyphenol quantities in red clover. These tests were selected due to their frequent use in research and familiarity among researchers. A total of 30 values per variable were selected from the 40 available values and distributed into three groups. As some post-hoc tests may perform better with groups of equal sizes, while others are more robust with groups of unequal sizes, both equal and unequal group sizes were randomly composed and used in the analysis of small samples.

Data Simulation

A Monte Carlo-like simulation approach was used to simulate different experimental scenarios and assess the robustness of the statistical tests. For each set of variables, data was randomly modified based on real values, with adjustments made (either by adding or subtracting) to meet the specified criteria of interest. To each scenario from the study, we assigned an individual variable.

The variables selected for the study were determined based on the following criteria:

- Normally distributed data.
- Difference/no difference between groups.
- Equal/unequal variance.
- Equal/unequal group size.

The scenarios were structured according to Figure 3. For all scenarios, two cases (a and b) were followed using variables with p-value distant or closer to the significance threshold, both for significant difference or no difference between groups (One-Way ANOVA or Welch ANOVA).

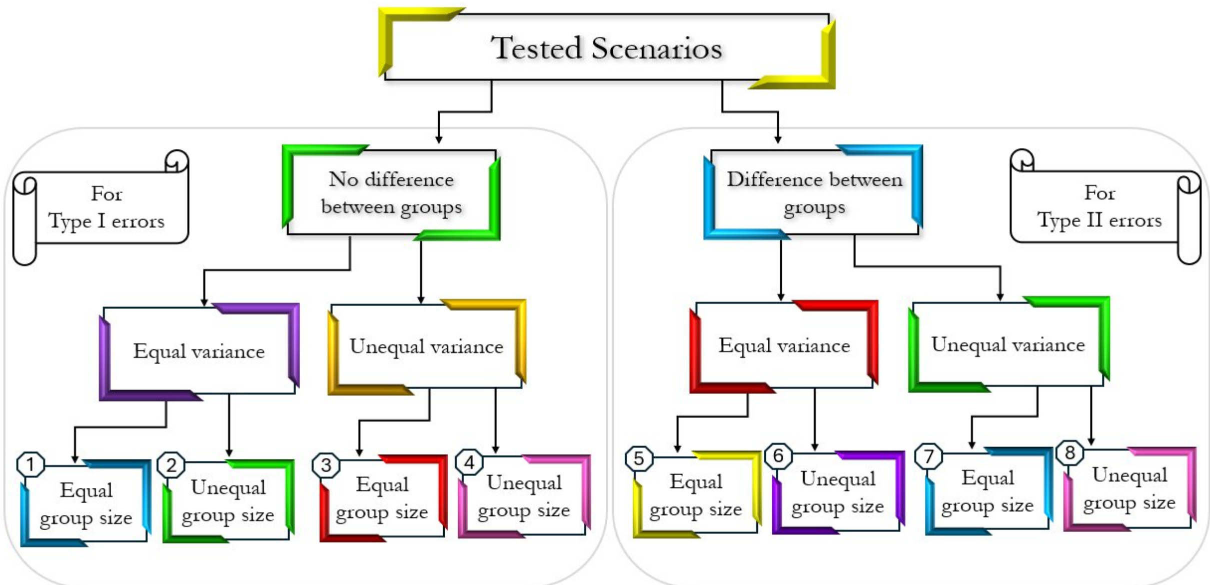

Figure 3. Study scenarios.

(Original design)

Statistical Methods

All data were analyzed using R 4.4.2 ('Pile of Leaves') software version [17] and IBM SPSS 26.0 (demo version) [18], with the following packages from R: "agricolae" [19] and "DescTools" [20]. Data normality was assessed using descriptive statistics, Q-Q (Quantile-Quantile) plot, and the Shapiro-Wilk test with Holm correction for multiple comparisons. The homogeneity of variances across different groups was evaluated with Bartlett test.

The significance cutoff was set to 0.05 (at 95% confidence level). One-Way ANOVA and Welch ANOVA tests were performed to determine the presence of a significant difference between groups. Duncan, Tukey HSD, Bonferroni, Fisher's LSD, Scheffé, Dunnett, Šidák, Gabriel, Games-Howell, or Tamhane T2 post-hoc tests for multiple pairwise comparisons were used to detect the presence of type I and type II errors.

Results

In this study, all groups from each scenario were normally distributed (p>0.05, Shapiro-Wilk with Holm correction for multiple comparison) and One-Way ANOVA or Welch ANOVA test was performed according to variance p-value obtained with Bartlett test (Table 3 – for both categories of variables, with p-values from the One-Way ANOVA/Welch ANOVA test that are either distant "a" or closer "b" to the threshold value).

Table 3. Characteristics of the scenario variables.

Variable		Shapiro-Wilk test with Holm correction (data normality) p-value			Bartlett test (variance) p-value	One-Way ANOVA/ Welch ANOVA/ p-value
		Group A	**Group B**	**Group C**		
Equal Group Size		**(n=10)**	**(n=10)**	**(n=10)**		
Scenario 1	a	>0.999	>0.999	0.718	0.289	0.383
	b	>0.999	>0.999	>0.999	0.784	0.056
Scenario 3	a	0.174	0.187	0.376	0.012*	0.535
	b	0.209	0.209	0.376	0.023*	0.089
Scenario 5	a	>0.999	>0.999	>0.999	0.619	<0.001***
	b	>0.999	>0.999	>0.999	0.281	0.047*
Scenario 7	a	0.092	0.163	0.337	0.035*	0.007**
	b	0.193	0.193	0.337	0.012*	0.024*
Unequal Group Size		**(n=7)**	**(n=12)**	**(n=11)**		
Scenario 2	a	>0.999	>0.999	0.194	0.538	0.217
	b	0.212	0.408	0.408	0.131	0.073
Scenario 4	a	0.667	0.962	0.962	0.012*	0.245
	b	>0.999	>0.999	>0.999	0.016*	0.068
Scenario 6	a	0.191	0.118	0.276	0.155	<0.001***
	b	0.401	0.118	0.401	0.807	0.012*
Scenario 8	a	0.142	0.799	0.799	<0.001***	<0.001***
	b	>0.999	>0.999	>0.999	<0.001***	0.015*

Note: */**/*** - significance according to p-value.

Analysis of Cases with no Significant Difference Between Groups

In the case of no significant difference between groups and both equal and unequal variance with p-value distant from the significance threshold, no errors were determined (Table 4 and Table 6). In the case of no significant difference between groups and equal variance with closer p-value to the significance threshold, Duncan, Fisher's LSD and Dunnett tests presented type I error (even though One-Way ANOVA did not show presence of difference between groups), for the defined sample size (counting 30 values/variable) (Table 5). In the case of no significant difference between groups and unequal variance with p-value closer to the significance threshold, the only post-hoc test that did not present type I errors was Tamhane T2 (Table 7).

Table 4. Testing on **equal variance**, group size type and p-value distant from the significance threshold

Post Hoc Test	Group size type	One-Way ANOVA p-value	Error type I/II	Post-hoc group comparison		
				Group A & B	Group A & C	Group B & C
				p-value	p-value	p-value
Duncan	Equal	0.383	No	0.923	0.242	0.252
	Unequal	0.217		0.152	0.064	0.582
Tukey HSD	Equal	0.383	No	0.995	0.425	0.481
	Unequal	0.217		>0.999	0.358	0.246
Bonferroni	Equal	0.383	No	>0.999	0.647	0.756
	Unequal	0.217		>0.999	0.523	0.338
Fisher's LSD	Equal	0.383	No	0.923	0.216	0.252
	Unequal	0.217		0.983	0.174	0.113
Scheffé	Equal	0.383	No	0.995	0.458	0.513
	Unequal	0.217		>0.999	0.391	0.277
Dunnett	Equal	0.383	No	0.993	0.355	0.408
	Unequal	0.217		>0.999	0.279	0.197
Šidák	Equal	0.383	No	>0.999	0.517	0.582
	Unequal	0.217		>0.999	0.437	0.301
Gabriel	Equal	0.383	No	>0.999	0.508	0.572
	Unequal	0.217		>0.999	0.424	0.294
Games-Howell	Equal	0.383	No	0.996	0.417	0.407
	Unequal	0.217		>0.999	0.344	0.236
Tamhane T2	Equal	0.383	No	>0.999	0.514	0.501
	Unequal	0.217		>0.999	0.431	0.290

Note: */**/*** - significance according to p-value obtained from post-hoc test for groups comparison.

Table 5. Testing on **equal variance**, group size type and p-value closer to the significance threshold.

Post Hoc Test	Group size type	One-Way ANOVA p-value	Error type I/II	Post-hoc group comparison		
				Group A & B	Group A & C	Group B & C
				p-value	p-value	p-value
Duncan	Equal	0.056	Yes	0.091	0.026*	0.484
	Unequal	0.073		0.796	0.098	0.039*
Tukey HSD	Equal	0.056	No	0.204	0.052	0.760
	Unequal	0.073		0.963	0.219	0.076
Bonferroni	Equal	0.056	No	0.272	0.061	>0.999
	Unequal	0.073		>0.999	0.295	0.092
Fisher's LSD	Equal	0.056	Yes	0.091	0.020*	0.484
	Unequal	0.073		0.796	0.098	0.031*
Scheffé	Equal	0.056	No	0.232	0.065	0.779
	Unequal	0.073		0.966	0.249	0.093
Dunnett	Equal	0.056	Yes	0.158	0.038*	0.703
	Unequal	0.073	No	0.947	0.163	0.057
Šidák	Equal	0.056	No	0.248	0.060	0.863
	Unequal	0.073		0.991	0.267	0.089
Gabriel	Equal	0.056	No	0.242	0.059	0.857
	Unequal	0.073		0.991	0.257	0.087
Games-Howell	Equal	0.056	No	0.236	0.070	0.728
	Unequal	0.073		0.969	0.186	0.076
Tamhane T2	Equal	0.056	No	0.292	0.083	0.837
	Unequal	0.073		0.994	0.236	0.091

Note: */**/*** - significance according to p-value obtained from post-hoc test for groups comparison.

Table 6. Testing on **unequal variance**, group size type and p-value distant from the significance threshold.

Post Hoc Test	Group size type	Welch ANOVA p-value	Error type I/II	Post-hoc group comparison		
				Group A & B	Group A & C	Group B & C
				p-value	p-value	p-value
Duncan	Equal	0.535	No	0.397	0.220	0.644
	Unequal	0.245		0.911	0.560	0.514
Tukey HSD	Equal	0.535	No	0.669	0.392	0.887
	Unequal	0.245		0.675	0.969	0.436
Bonferroni	Equal	0.535	No	>0.999	0.586	>0.999
	Unequal	0.245		>0.999	>0.999	0.668
Fisher's LSD	Equal	0.535	No	0.397	0.195	0.644
	Unequal	0.245		0.402	0.812	0.223
Scheffé	Equal	0.535	No	0.694	0.425	0.897
	Unequal	0.245		0.700	0.971	0.469
Dunnett	Equal	0.535	No	0.602	0.324	0.855
	Unequal	0.245		0.954	0.589	0.370
Šidák	Equal	0.535	No	0.781	0.479	0.955
	Unequal	0.245		0.786	0.993	0.530
Gabriel	Equal	0.535	No	0.773	0.470	0.953
	Unequal	0.245		0.775	0.993	0.521
Games-Howell	Equal	0.535	No	0,715	0.508	0.771
	Unequal	0.245		0.821	0.983	0.192
Tamhane T2	Equal	0.535	No	0.829	0.618	0.874
	Unequal	0.245		0.916	0.998	0.235

Note: */**/*** - significance according to p-value obtained from post-hoc test for groups comparison.

Table 7. Testing on **unequal variance**, group size type and p-value closer to the significance threshold.

Post Hoc Test	Group size type	Welch ANOVA p-value	Error type I/II	Post-hoc group comparison		
				Group A & B	Group A & C	Group B & C
				p-value	p-value	p-value
Duncan	Equal	0.089	Yes	0.065	0.012*	0.390
	Unequal	0.068	No	0.138	0.953	0.097
Tukey HSD	Equal	0.089	Yes	0.151	0.024*	0.661
	Unequal	0.068	No	0.293	0.998	0.184
Bonferroni	Equal	0.089	Yes	0.194	0.028*	>0.999
	Unequal	0.068	No	0.412	>0.999	0.243
Fisher's LSD	Equal	0.089	Yes	0.065	0.009*	0.390
	Unequal	0.068	No	0.137	0.952	0.081
Scheffé	Equal	0.089	Yes	0.176	0.032*	0.686
	Unequal	0.068	No	0.325	0.998	0.212
Dunnett	Equal	0.089	Yes	0.115	0.017*	0.593
	Unequal	0.068	No	0.224	0.997	0.144
Šidák	Equal	0.089	Yes	0.181	0.028*	0.773
	Unequal	0.068	No	0.358	>0.999	0.223
Gabriel	Equal	0.089	Yes	0.178	0.027*	0.765
	Unequal	0.068	No	0.344	>0.999	0.219
Games-Howell	Equal	0.089	No	0.213	0.072	0.449
	Unequal	0.068	Yes	0.543	0.999	0.044*
Tamhane T2	Equal	0.089	No	0.267	0.087	0.551
	Unequal	0.068		0.666	>0.999	0.051

Note: */**/*** - significance according to p-value obtained from post-hoc test for groups comparison.

Analysis of Cases with Significant Difference Between Groups

In the case of differences and both equal and unequal variance, no errors were determined (Table 8-11).

Table 8. Testing on **equal variance**, group size type and p-value distant from the significance threshold.

Post Hoc Test	Group size type	One-Way ANOVA p-value	Error type I/II	Post-hoc group comparison		
				Group A & B	Group A & C	Group B & C
				p-value	p-value	p-value
Duncan	Equal	<0.001	No	<0.001***	<0.001***	<0.001
	Unequal	<0.001		0.002**	<0.001***	0.414
Tukey HSD	Equal	<0.001	No	<0.001***	<0.001***	<0.001
	Unequal	<0.001		0.004**	0.001***	0.689
Bonferroni	Equal	<0.001	No	<0.001***	<0.001***	<0.001
	Unequal	<0.001		0.005**	0.001***	>0.999
Fisher's LSD	Equal	<0.001	No	<0.001***	<0.001***	<0.001
	Unequal	<0.001		0.002**	<0.001***	0.414
Scheffé	Equal	<0.001	No	<0.001***	<0.001***	<0.001
	Unequal	<0.001		0.006**	0.001***	0.712
Dunnett	Equal	<0.001	No	<0.001***	<0.001***	<0.001
	Unequal	<0.001		0.003**	0.001***	0.629
Šidák	Equal	<0.001	No	<0.001***	<0.001***	<0.001
	Unequal	<0.001		0.005**	0.001***	0.799
Gabriel	Equal	<0.001	No	<0.001***	<0.001***	<0.001
	Unequal	<0.001		0.004**	0.001***	0.792
Games-Howell	Equal	<0.001	No	<0.001***	<0.001***	<0.001
	Unequal	<0.001		0.055	0.024*	0.579
Tamhane T2	Equal	<0.001	No	<0.001***	<0.001***	<0.001
	Unequal	<0.001		0.069	0.029*	0.691

Note: */**/*** - significance according to p-value obtained from post-hoc test for groups comparison.

Table 9. Testing on **equal variance**, group size type and p-value closer to the significance threshold.

Post Hoc Test	Group size type	One-Way ANOVA p-value	Error type I/II	Post-hoc group comparison		
				Group A & B	Group A & C	Group B & C
				p-value	p-value	p-value
Duncan	Equal	0.047	No	0.198	0.204	0.019*
	Unequal	0.012		0.023*	0.005**	0.343
Tukey HSD	Equal	0.047	No	0.397	0.407	0.037*
	Unequal	0.012		0.058	0.009**	0.605
Bonferroni	Equal	0.047	No	0.594	0.613	0.043*
	Unequal	0.012		0.069	0.010**	>0.999
Fisher's LSD	Equal	0.047	No	0.198	0.204	0.014*
	Unequal	0.012		0.023*	0.003**	0.343
Scheffé	Equal	0.047	No	0.430	0.440	0.047*
	Unequal	0.012		0.072	0.013*	0.633
Dunnett	Equal	0.047	No	0.328	0.338	0.026*
	Unequal	0.012		0.041*	0.006**	0.540
Šidák	Equal	0.047	No	0.484	0.496	0.042*
	Unequal	0.012		0.068	0.010**	0.717
Gabriel	Equal	0.047	No	0.475	0.487	0.041*
	Unequal	0.012		0.064	0.010**	0.708
Games-Howell	Equal	0.047	No	0.396	0.492	0.023*
	Unequal	0.012		0.110	0.033*	0.579
Tamhane T2	Equal	0.047	No	0.489	0.598	0.026*
	Unequal	0.012		0.136	0.039*	0.691

Note: */**/*** - significance according to p-value obtained from post-hoc test for groups comparison.

Table 10. Testing on **unequal variance**, group size type and p-value distant from the significance threshold.

Post Hoc Test	Group size type	Welch ANOVA p-value	Error type I/II	Post-hoc group comparison		
				Group A & B	Group A & C	Group B & C
				p-value	p-value	p-value
Duncan	Equal	0.007	No	0.002	<0.001	0.418
	Unequal	<0.001		<0.001***	<0.001***	0.408
Tukey HSD	Equal	0.007	No	0.005**	0.001***	0.693
	Unequal	<0.001		<0.001***	0.001***	0.682
Bonferroni	Equal	0.007	No	0.005**	0.001***	>0.999
	Unequal	<0.001		<0.001***	0.001***	>0.999
Fisher's LSD	Equal	0.007	No	0.002**	<0.001***	0.418
	Unequal	<0.001		<0.001***	<0.001***	0.408
Scheffé	Equal	0.007	No	0.007**	0.001***	0.716
	Unequal	<0.001		<0.001***	0.001***	0.706
Dunnett	Equal	0.007	No	0.003**	<0.001***	0.628
	Unequal	<0.001		<0.001***	<0.001***	0.622
Šidák	Equal	0.007	No	0.005**	0.001***	0.803
	Unequal	<0.001		<0.001***	0.001***	0.793
Gabriel	Equal	0.007	No	0.005**	0.001***	0.796
	Unequal	<0.001		<0.001***	0.001***	0.786
Games-Howell	Equal	0.007	No	0.017*	0.006**	0.521
	Unequal	<0.001		0.001***	0.004**	0.667
Tamhane T2	Equal	0.007	No	0.019*	0.006**	0.632
	Unequal	<0.001		0.001***	0.004**	0.784

Note: */**/*** - significance according to p-value obtained from post-hoc test for groups comparison.

Table 11. Testing on **unequal variance**, group size type and p-value closer to the significance threshold.

Post Hoc Test	Group size type	Welch ANOVA p-value	Error type I/II	Post-hoc group comparison		
				Group A & B	Group A & C	Group B & C
				p-value	p-value	p-value
Duncan	Equal	0.024	No	0.007**	0.001**	0.456
	Unequal	0.015		0.001**	0.006**	0.417
Tukey HSD	Equal	0.024	No	0.018*	0.003**	0.732
	Unequal	0.015		0.002**	0.017*	0.692
Bonferroni	Equal	0.024	No	0.021*	0.003**	>0.999
	Unequal	0.015		0.003**	0.019*	>0.999
Fisher's LSD	Equal	0.024	No	0.007**	0.001***	0.456
	Unequal	0.015		0.001***	0.006**	0.417
Scheffé	Equal	0.024	No	0.025*	0.004**	0.753
	Unequal	0.015		0.004**	0.022*	0.715
Dunnett	Equal	0.024	No	0.013*	0.002**	0.671
	Unequal	0.015		0.002**	0.011*	0.633
Šidák	Equal	0.024	No	0.021*	0.003**	0.839
	Unequal	0.015		0.003**	0.019*	0.802
Gabriel	Equal	0.024	No	0.020*	0.003**	0.832
	Unequal	0.015		0.002**	0.018*	0.795
Games-Howell	Equal	0.024	No	0.049*	0.018*	0.521
	Unequal	0.015		0.015*	0.026*	0.667
Tamhane T2	Equal	0.024	No	0.059	0.021*	0.632
	Unequal	0.015		0.018*	0.033*	0.784

Note: */**/*** - significance according to p-value obtained from post-hoc test for groups comparison.

Discussion

Following the study, a series of post-hoc tests were identified as presenting type I errors. According to the scenario where no differences between groups were found, variance was equal, and the p-value from One-Way ANOVA tests was close to the significance threshold, both Duncan and Fisher's LSD post-hoc tests presented Type I errors in both equal and unequal group sizes. In contrast, the Dunnett test showed Type I errors only in the case of equal group sizes (Table 5). The Duncan and Fisher's LSD tests were more susceptible to Type I errors compared to the other post-hoc tests, with a higher risk, as confirmed by scientists [10,12]. Additionally, both post-hoc tests were influenced by the group size. In contrast, the Dunnett test, compared to the other two, was affected only by equal group sizes. According to scientists, it was recommended for cases where the comparison groups had equal sizes [11,12], but a small group size affected this post-hoc test.

Similarly, for the same scenario, but in cases where variance was unequal, the Duncan, Tukey HSD, Bonferroni, Fisher's LSD, Scheffé, Dunnett, Šidák, and Gabriel tests presented Type I errors when group sizes were equal (Table 7). The majority of these tests were recommended by scientists as suitable for groups with unequal sizes [7,10]. The Games-Howell test, compared to the others, showed Type I errors only when group sizes were unequal. Of this category, the only test that did not present Type I errors was Tamhane T2 post-hoc test.

For the scenario where no differences were determined between groups, variance was either equal or unequal, group sizes were equal or unequal, and the p-value of both ANOVA tests was far from the significance threshold, none of the post-hoc tests showed type I errors (Table 4 and Table 6).

In the other scenario where differences between groups were present, variance was either equal or unequal, group sizes were equal or unequal, and the p-value of both ANOVA tests was either distant or close to the significance threshold, none of the post-hoc tests showed type II errors, all returning good values (Table 8-11).

Some tests like Tukey HSD and Bonferroni are more appropriate for larger samples ($n>30$) [10], and using a small sample size influenced the post-hoc tests negatively.

Study Limitations

A small sample size (n=30/variable) may reduce the statistical power of the analysis and limit the applicability of the findings to the larger population. Another limitation was represented by the fact that this study focused only on parametric post-hoc tests. Also, using only variables that follow a normal distribution, the study may limit the ability to determine the behavior of all relevant variables. Future research is recommended to investigate this aspect, with the goal of providing clearer guidance on the selection of appropriate post-hoc tests for various experimental conditions. Additionally, studies using a wider range of scenarios and larger sample sizes will offer a more comprehensive understanding of the findings and enhance the applicability of the results.

Conclusions

For analysis, in the case of small sample size, presenting equal variance and p-value closer to the significance threshold (from One-Way ANOVA tests), post-hoc test like Tukey HSD, Bonferroni and others (except for Duncan, Fisher's LSD and Dunnett) controlled the type I errors. In the case of unequal variance, a proper test to be used was Tamhane T2 since it controlled the type I error, better than the other post-hoc tests. For cases where the p-value from One-Way ANOVA tests was distant to the significance threshold, every post-hoc performed well, no errors occurred during the process.

List of Abbreviations: ANOVA – Analysis of Variance; Tukey HSD – Tukey's Honest Significant Difference; Fisher's LSD – Fisher's Least Significant Difference.

Author Contributions: ADC defined the research's aim and the experiment's design. ADC and ADO participated in the study design. ADC conducted experiments and performed statistical analysis. ADC and ADO coordinated and helped to draft the manuscript. All authors read and approved of the final manuscript.

Funding: This research received no funding.

Ethics Statement: Not applicable.

Data Availability Statement: The data supporting this research have been uploaded to Fighare and can be accessed at: doi: 10.6084/m9.figshare.30479504.

Conflict of Interest: The authors declare no conflict of interest.

References

1. Curran-Everett D, Milgrom H. Post-hoc data analysis: benefits and limitations. Curr Opin Allergy Clin Immunol. 2013;13(3):223-4, https://doi.org/10.1097/ACI.0b013e3283609831.
2. Drugan T, Bolboacă SD, Leucuţa D, Bondor C, Călinici T, Văleanu M, Colosi H, Iancu M, Istrate D. Curs de Biostatistică Medicală. Cluj-Napoca: Editura medicală Universitară ”Iuliu Haţieganu”. 2018. pp 127.
3. Brown AM. A new software for carrying out one-way ANOVA post hoc tests. Comput Methods Programs Biomed. 2005;79(1):89-95. https://doi.org/10.1016/j.cmpb.2005.02.007.
4. Juarros Basterretxea J, Aonso Diego G, Postigo Gutiérrez Á, Montes Álvarez P, Menéndez Aller Á, García Cueto E. Post-hoc tests in one-way ANOVA: The case for normal distribution. Methodol. 2024;20(2):84–99. https://doi.org/10.5964/meth.11721.
5. Kucuk U, Eyuboglu M, Kucuk HO, Degirmencioglu G. Importance of using proper post hoc test with ANOVA. Int J Cardiol, 2015;209:346. http://dx.doi.org/10.1016/j.ijcard.2015.11.061.
6. Grinde KE, Arbet J, Green A, O'Connell M, Valcarcel A, Westra J, Tintle N. Illustrating, quantifying, and correcting for bias in post-hoc analysis of gene-based rare variant tests of association. Front Genet. 2017;8:117. https://doi.org/10.3389/fgene.2017.00117.
7. Shingala MC, Rajyaguru A. Comparison of post hoc tests for unequal variance. Int J New Technol Sci Eng. 2015;2(5):22-33.
8. Cramer AOJ, van Ravenzwaaij D, Matzke D, Steingroever H, Wetzels R, Grasman RPPP, Waldorp LJ, Wagenmakers EJ. Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. Psychon Bull Rev. 2016;23:640–647. https://doi.org/10.3758/s13423-015-0913-5.
9. McHugh ML. Multiple comparison analysis testing ANOVA. Biochem Med. 2011;21(3):203–209. https://doi.org/10.11613/BM.2011.029.
10. Freeman J, Obasohan P. Multiple testing including post hoc tests. The university of Sheffield. 2020. [Internet]. https://www.sheffield.ac.uk/media/30898/download?attachment. [accessed on 27 June 2025].
11. Sauder D. Examining the type I error and power of 18 common post-hoc comparison tests. *Graduate Psychology:* James Madison University (Master thesis). James Madison University https://commons.lib.jmu.edu/master201019/524. 2017. pp: 16-70.
12. Agbangba CE, Aide ES, Honfo H, Kakai RG. On the use of post-hoc tests in environmental and biological sciences: A critical review. Heliyon. 2024;10(3):3e25131. https://doi.org/10.1016/j.heliyon.2024.e25131.
13. Sarstedt M, Mooi E. Hypothesis testing and ANOVA. In: Sarstedt M, Mooi E. (Eds.), A concise guide to market research: The process, data, and methods using IBM SPSS statistics. 2014; pp. 141-192. Springer. https://doi.org/10.1007/978-3-662-56707-4_6.
14. Meyers LS, Gamst G, Guarino AJ. Applied multivariate research: Design and interpretation (3rd ed.). SAGE. 2016; pp: 2-50.
15. Lee S, Lee DK. What is the proper way to apply the multiple comparison test? Korean J Anesthesiol, 2020; 71(5):353–60. https://doi.org/10.4097/kja.d.18.00242.
16. Keppel G, Wickens TD. Design and analysis: A researcher's handbook. Pearson Prentice-Hall. 2004; pp: 120-130. http://libraries.najah.edu/node/13381.
17. R Core Team R: a language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing). 2024. [Internet]. Available at: https://www.R-project.org/ [Accessed 15 June 2025).
18. IBM Corp. Released 2019. IBM SPSS Statistics for Windows, version 26.0, IBM Corp, Armonk, NY, USA.
19. de Mendiburu F. _agricolae: Statistical Procedures for Agricultural Research_. R package version 1.3-7. [Internet]. <https://CRAN.R-project.org/package=agricolae>. 2023. [accessed on 25 June 2025].
20. Signorell A. _DescTools: Tools for Descriptive Statistics_. R package version 0.99.60. [Internet]. <https://CRAN.R-project.org/package=DescTools>. 2025. [accessed on 25 June 2025].