

Learn How to Perform a Two-Way ANOVA in R

Authored by
Mohammed loot

November 9, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Learn How to Perform a Two-Way ANOVA in R*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=14237>

The [Analysis of Variance](#) (ANOVA) is a powerful statistical technique used to compare the means of different groups. Specifically, a **Two-Way ANOVA** extends this concept, allowing researchers to determine if there is a statistically significant difference in a continuous dependent variable based on two independent categorical [factors](#). This method is essential when investigating the simultaneous effects of two distinct variables--and crucially, whether these factors interact with each other--on a measured outcome. Understanding how to execute this analysis correctly in [R](#) is a fundamental skill for advanced data analysis.

This comprehensive tutorial will guide you step-by-step through the process of conducting a Two-Way ANOVA using the [R](#) programming environment. We will cover everything from setting up the experimental data and performing essential exploratory data analysis to fitting the model, rigorously checking statistical assumptions, and finally, interpreting the complex results, including post hoc comparisons.

Setting Up the Two-Way ANOVA Example in R

To illustrate the Two-Way ANOVA process, consider a hypothetical clinical trial designed to assess factors influencing **weight loss**. Our primary objective is to determine if two distinct factors--exercise intensity and gender--impact the amount of weight lost over a one-month period. In this design, the two **independent variables** (or [factors](#)) are *exercise* (with levels: None, Light, Intense) and *gender* (Male, Female), while the **response variable** is *weight loss*, measured in pounds.

The power of the Two-Way ANOVA lies in its ability to test not only the **main effects** of exercise and gender individually but also the potential **interaction effect**. An interaction effect occurs if the impact of exercise on weight loss is different for males than it is for females, meaning the effect of one factor is dependent on the level of the other factor. This holistic approach provides a much richer understanding of the underlying phenomena than analyzing each factor in isolation.

For our experiment, we recruit a total of 60 participants (30 men and 30 women). Participants are randomly assigned to one of three exercise regimens: no exercise, light exercise, or intense exercise, ensuring equal group sizes (10 participants per group) across all combinations of gender and exercise intensity. The following [R](#) code generates the reproducible data frame that simulates the results of this one-month study.

The code uses the `set.seed()` function to ensure that the random data generated remains consistent every time the script is run, a crucial step for **reproducible research**. It then constructs a data frame containing the categorical factors (gender and exercise) and the continuous outcome (weight loss). We conclude this setup by inspecting the structure and group counts of our synthesized data.

Ensure the analysis is reproducible

set.seed(10)

```
# Create the data frame for the two-way ANOVA
data <- data.frame(gender = rep(c("Male", "Female"), each = 30),
  exercise = rep(c("None", "Light", "Intense"), each = 10, times = 2),
  weight_loss = c(runif(10, -3, 3), runif(10, 0, 5), runif(10, 5, 9),
  runif(10, -4, 2), runif(10, 0, 3), runif(10, 3, 8)))

# View the structure of the data frame (first six rows)
head(data)

# gender exercise weight_loss
#1 Male None 0.04486922
#2 Male None -1.15938896
#3 Male None -0.43855400
#4 Male None 1.15861249
#5 Male None -2.48918419
#6 Male None -1.64738030

# Verify the number of participants in each treatment cell
table(data$gender, data$exercise)

# Intense Light None
# Female 10 10 10
# Male 10 10 10
```

Exploratory Data Analysis and Visualization

Before proceeding with the formal statistical modeling, conducting **Exploratory Data Analysis** (EDA) is crucial. EDA allows us to gain an intuitive understanding of the data distribution, identify potential outliers, and observe initial trends that may influence the [ANOVA](#) results. Our first step is to calculate the descriptive statistics--the mean and standard deviation--of the weight loss outcome for each of the six unique treatment groups (the combinations of gender and exercise level). This insight helps us anticipate the outcomes of the formal hypothesis tests.

We utilize the popular [dplyr](#) package for efficient data manipulation in R. By employing the `group_by()` and `summarise()` functions, we can quickly generate a comprehensive summary table detailing the central tendency and variability within each experimental condition. A review of this table immediately highlights differences in average weight loss across the different exercise intensity levels and genders.

Load the essential data manipulation package**library(dplyr)**

Calculate the mean and standard deviation of weight loss for all six treatment groups

data %>%

group_by(gender, exercise) %>%

summarise(mean = mean(weight_loss),

sd = sd(weight_loss))

A tibble: 6 x 4

Groups: gender

gender exercise mean sd

#

#1 Female Intense 5.31 1.02

#2 Female Light 0.920 0.835

#3 Female None -0.501 1.77

#4 Male Intense 7.37 0.928

#5 Male Light 2.13 1.22

#6 Male None -0.698 1.12

Beyond numerical summaries, visualization is a critical component of EDA. We generate a **boxplot** for each of the six treatment groups to visually inspect the distribution of weight loss, assess potential symmetry, and compare the spread (variability) across groups. The boxplot visualization is particularly effective for identifying group differences and checking for consistency in variances, which is a key assumption for [ANOVA](#). Note that we adjust the plot margins using `par(mar=c(...))` to prevent axis labels from being truncated in the resulting visualization.

Set margins to ensure x-axis labels are fully visible**par(mar=c(8, 4.1, 4.1, 2.1))**

Create boxplots to visualize weight loss distribution across groups

boxplot(weight_loss ~ gender:exercise,

data = data,

main = "Weight Loss Distribution by Group",

xlab = "Group",

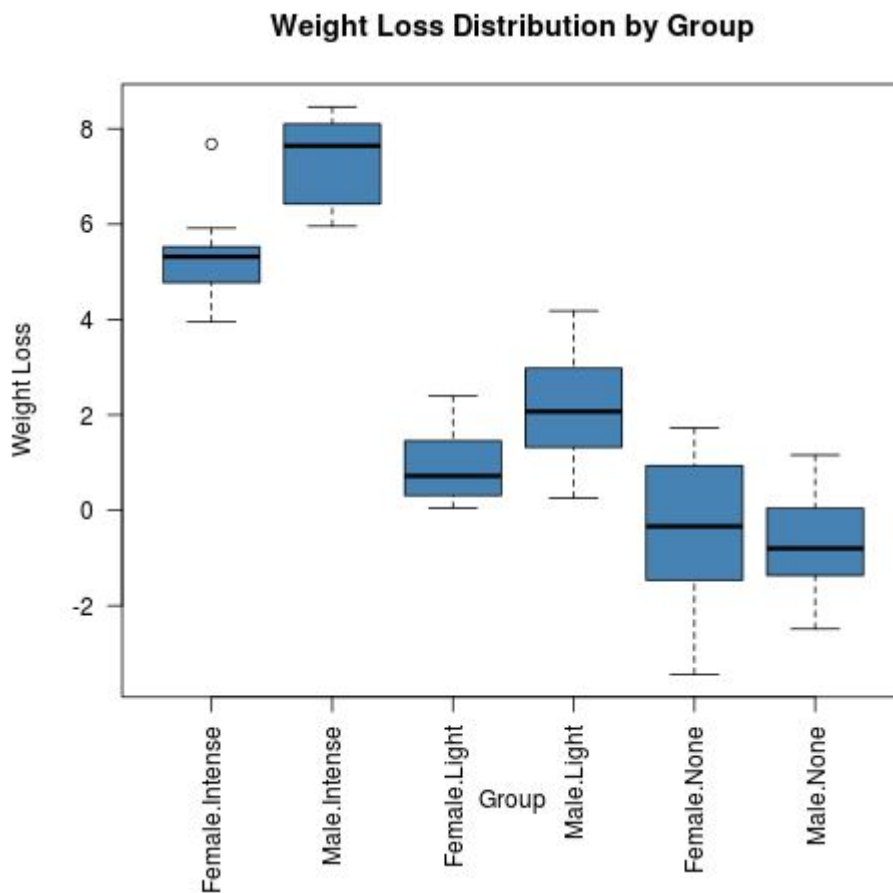
ylab = "Weight Loss",

col = "steelblue",

border = "black",

las = 2 # Orient x-axis labels vertically for readability

)



The boxplots immediately reveal several important preliminary findings. Firstly, both males and females in the *intense* exercise group exhibit substantially greater weight loss compared to the other groups. Secondly, there appears to be a difference in the magnitude of weight loss between genders, particularly in the *intense* and *light* exercise categories, where males tend to show higher weight loss values than females. These visual observations suggest that both main effects (exercise and gender) and the interaction effect may be **statistically significant**, prompting us to proceed with fitting the formal Two-Way [ANOVA](#) model to test these hypotheses rigorously.

Implementing and Interpreting the Two-Way ANOVA Model

To formally test our hypotheses regarding the main effects and the interaction effect, we employ the `aov()` function in [R](#). The standard syntax for a Two-Way ANOVA explicitly includes the interaction term, denoted by the asterisk (*) operator between the predictor variables. This operator signifies that we are interested in partitioning the variance attributable to Factor A, Factor B, and the combined interactive effect of A and B.

The general structure for fitting a Two-Way ANOVA model in R is as follows:

```
aov(response variable ~ predictor_variable1 * predictor_variable2, data = dataset)
```

In our specific example, we define *weight_loss* as the dependent variable and *gender* and *exercise* as our two independent predictor [factors](#). The inclusion of the `*` operator--as opposed to a plus sign (`+`), which would only test for main effects--ensures that the model also calculates the effect of the interaction term (`gender:exercise`). Once the model is fitted, the `summary()` function provides the official ANOVA table, displaying the partitioned Sums of Squares, Mean Squares, F-statistics, and associated p-values.

```
# Fit the Two-Way ANOVA model, including the interaction term
```

```
model <- aov(weight_loss ~ gender * exercise, data = data)
```

```
# View the ANOVA summary table output
```

```
summary(model)
```

```
# Df Sum Sq Mean Sq F value Pr(>F)
```

```
#gender 1 15.8 15.80 11.197 0.0015 **
```

```
#exercise 2 505.6 252.78 179.087 <2e-16 ***
```

```
#gender:exercise 2 13.0 6.51 4.615 0.0141 *
```

```
#Residuals 54 76.2 1.41
```

```
#---
```

```
#Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The interpretation of the ANOVA table is straightforward: we examine the p-value ($P_{r(>F)}$) for each row against our established significance level, typically $\alpha = 0.05$. The results indicate that the effect of **gender** is statistically significant ($p = 0.0015$), as is the effect of **exercise** ($p < 2e-16$). Most critically, the **interaction term** (`gender:exercise`) is also statistically significant ($p = 0.0141$). The finding of a significant interaction suggests that the effect of exercise intensity on weight loss is fundamentally different across the two genders, necessitating further investigation through post hoc testing to pinpoint exactly where these differences lie.

Validating the Model: Checking ANOVA Assumptions

The reliability of the results derived from the Two-Way ANOVA model is entirely dependent on meeting several core statistical assumptions. Before interpreting the significant effects, it is mandatory to verify these assumptions. Failure to meet these criteria can lead to invalid p-values and unreliable conclusions. The three primary assumptions for ANOVA are Independence, Normality, and Homogeneity of Variances.

Independence of Observations: This assumption requires that the measurements taken from one participant are independent of measurements taken from another participant. Since our study

used a **randomized design** where participants were independently assigned to treatment groups, this assumption is generally considered to be met through experimental design, provided no external factors influenced participants collectively.

Normality of Residuals: The dependent variable (weight loss) should be approximately normally distributed within each combination of the two factors. A practical way to assess this assumption is by examining the distribution of the model's **residuals**. If the residuals--the differences between the observed and predicted values--are roughly normally distributed around a mean of zero, the normality assumption is generally considered acceptable.

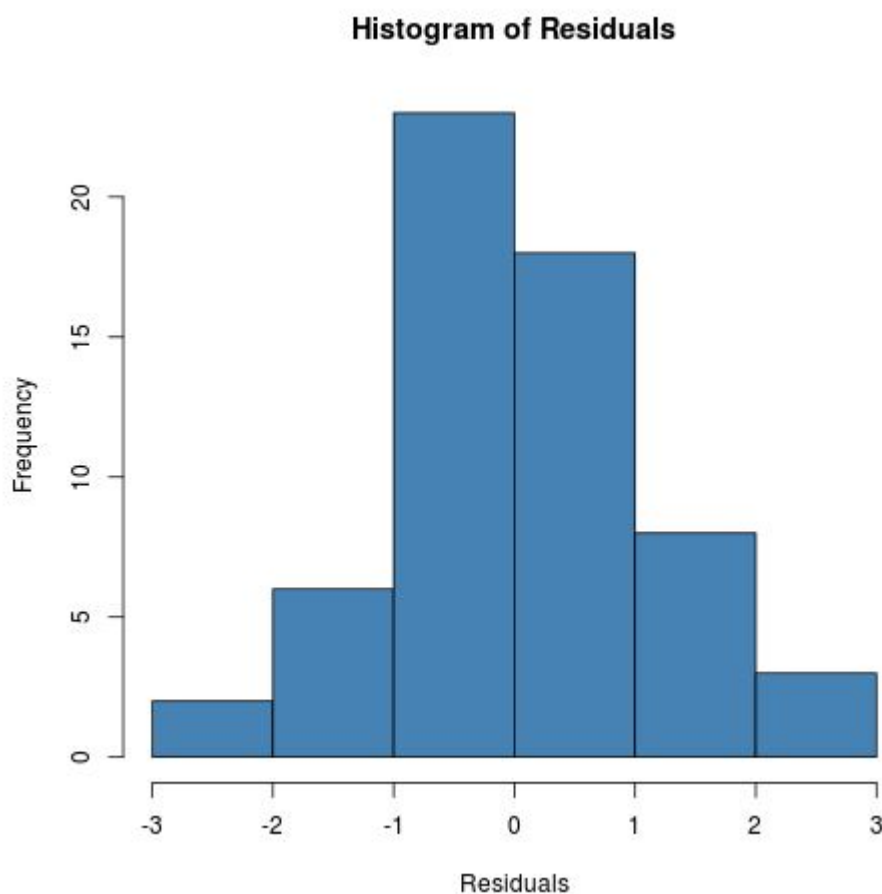
We check the Normality assumption by extracting the residuals from our fitted model and generating a histogram. A bell-shaped curve suggests that the residuals are adequately normally distributed, which supports the validity of our ANOVA calculation.

Define the model residuals

```
resid <- model$residuals
```

Create a histogram to visualize the distribution of residuals

```
hist(resid, main = "Histogram of Residuals", xlab = "Residuals", col = "steelblue")
```



The histogram displays a distribution that is reasonably symmetrical and mound-shaped, suggesting that the model residuals are **roughly normally distributed**. Therefore, we can comfortably conclude that the assumption of Normality is satisfied for this analysis.

Homogeneity of Variances (Equal Variance): This crucial assumption, often referred to as [Homoscedasticity](#), requires that the variance of the dependent variable is equal or approximately equal across all levels of the independent factors. We test this formally using the **Levene's Test**, which is less sensitive to deviations from normality than Bartlett's test. We must use the `car` package in R to perform this check.

The null hypothesis for Levene's test is that the variances are equal. If the p-value is greater than $\alpha = 0.05$, we fail to reject the null hypothesis and conclude that the assumption of [Equal Variance](#) is met. Conversely, a significant p-value ($p < 0.05$) would indicate unequal variances, potentially requiring data transformation or the use of a non-parametric alternative.

Load the 'car' package, which contains Levene's Test
library(car)

```
# Conduct Levene's Test for equality of variances
leveneTest(weight_loss ~ gender * exercise, data = data)

#Levene's Test for Homogeneity of Variance (center = median)
# Df F value Pr(>F)
#group 5 1.8547 0.1177
# 54
```

The output of Levene's test shows a p-value of 0.1177. Since this value is considerably greater than the standard significance threshold of 0.05, we do not have sufficient evidence to reject the null hypothesis. Therefore, we confidently assume that the variances among our six treatment groups are equal, and the assumption of [Homogeneity of Variances](#) is met. With all assumptions validated, we can proceed to interpret the statistically significant differences identified in the ANOVA table.

Identifying Specific Group Differences with Post Hoc Tests

While the ANOVA summary tells us that at least some group means are different (i.e., the main effects and the interaction are significant), it does not specify which particular pairs of groups are significantly different from one another. To determine the precise location of these treatment differences, we must conduct a **post hoc analysis** (Latin for "after the fact"). Given that we have equal sample sizes and have met the assumption of equal variances, **Tukey's Honest Significant Difference (TukeyHSD) test** is the most appropriate method for pairwise comparisons.

The [TukeyHSD](#) function performs all possible pairwise comparisons between the means of the groups, while simultaneously controlling the **Family-Wise Error Rate (FWER)**. This control is essential because running multiple t-tests without correction dramatically inflates the probability of making a Type I error (finding a significant difference when one does not truly exist). The function takes our fitted ANOVA model as input and calculates the mean difference (`diff`), the confidence interval (`lwr` and `upr`), and the adjusted p-value (`p_adj`) for every comparison.

Perform Tukey's Test for all multiple comparisons, using a 95% confidence level

TukeyHSD(model, conf.level=.95)

```
# Tukey multiple comparisons of means
# 95% family-wise confidence level
#
#Fit: aov(formula = weight_loss ~ gender * exercise, data = data)
#
# $gender
# diff lwr upr p adj
#Male-Female 1.026456 0.4114451 1.641467 0.0014967
#
# $exercise
# diff lwr upr p adj
#Light-Intense -4.813064 -5.718493 -3.907635 0.0e+00
#None-Intense -6.938966 -7.844395 -6.033537 0.0e+00
#None-Light -2.125902 -3.031331 -1.220473 1.8e-06
#
# $ `gender:exercise`
# diff lwr upr p adj
#Male:Intense-Female:Intense 2.0628297 0.4930588 3.63260067 0.0036746
#Female:Light-Female:Intense -4.3883563 -5.9581272 -2.81858535 0.0000000
#Male:Light-Female:Intense -3.1749419 -4.7447128 -1.60517092 0.0000027
#Female:None-Female:Intense -5.8091131 -7.3788841 -4.23934219 0.0000000
#Male:None-Female:Intense -6.0059891 -7.5757600 -4.43621813 0.0000000
#Female:Light-Male:Intense -6.4511860 -8.0209570 -4.88141508 0.0000000
#Male:Light-Male:Intense -5.2377716 -6.8075425 -3.66800066 0.0000000
#Female:None-Male:Intense -7.8719429 -9.4417138 -6.30217192 0.0000000
#Male:None-Male:Intense -8.0688188 -9.6385897 -6.49904786 0.0000000
#Male:Light-Female:Light 1.2134144 -0.3563565 2.78318536 0.2185439
#Female:None-Female:Light -1.4207568 -2.9905278 0.14901410 0.0974193
#Male:None-Female:Light -1.6176328 -3.1874037 -0.04786184 0.0398106
#Female:None-Male:Light -2.6341713 -4.2039422 -1.06440032 0.0001050
```

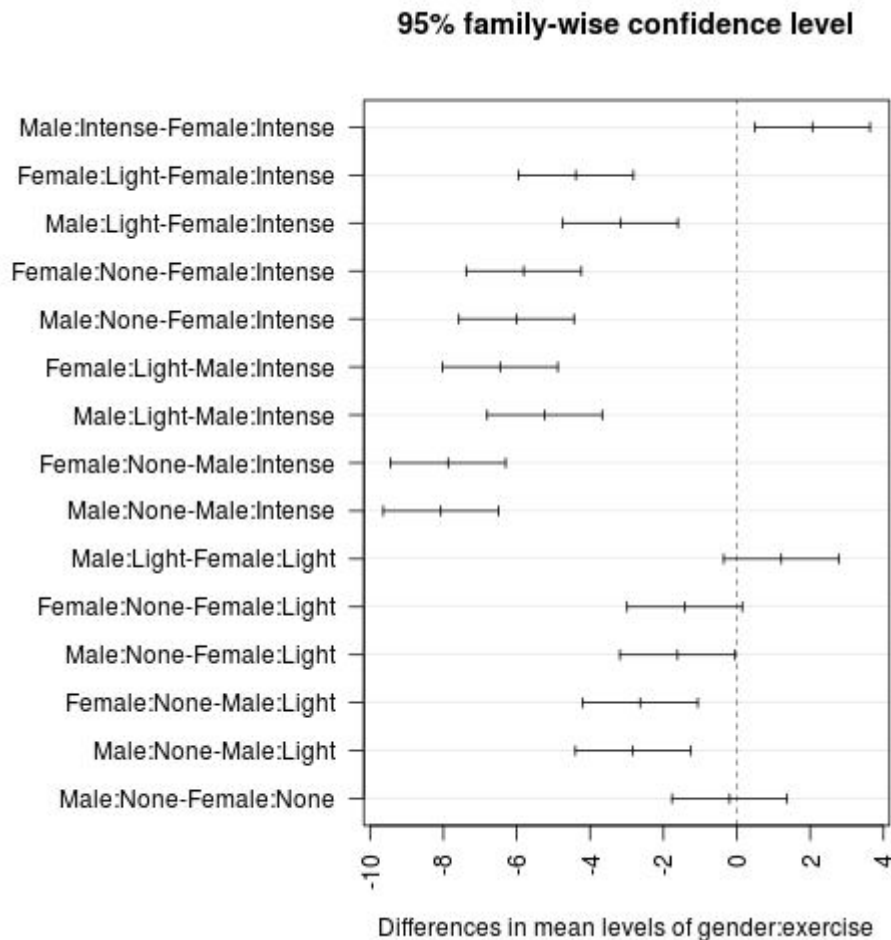
```
#Male:None-Male:Light -2.8310472 -4.4008181 -1.26127627 0.0000284  
#Male:None-Female:None -0.1968759 -1.7666469 1.37289500 0.9990364
```

Each line in the output represents a comparison between two specific groups. A comparison is considered statistically significant if the `p adj` value is less than 0.05. For instance, examining the `$gender:exercise` section, we see that the difference in weight loss between 'Male:Intense' and 'Female:Intense' is significant ($p = 0.0036746$), indicating that males lost significantly more weight than females when both groups engaged in intense exercise. Conversely, comparing 'Male:None' and 'Female:None' yields a p-value of 0.9990364, confirming that there is no statistically detectable difference in weight loss between genders when no exercise is performed.

A powerful way to summarize these complex comparisons is through visualization. The `plot()` function applied to the [TukeyHSD](#) results generates a plot of the 95% **confidence intervals** for the mean difference of each comparison. If the confidence interval for a specific comparison does not cross the zero line, the difference between those two groups is considered statistically significant at the 0.05 level. We adjust the plot margins again to ensure the comparison labels are fully visible on the y-axis.

```
# Adjust axis margins to accommodate lengthy comparison labels  
par(mar=c(4.1, 13, 4.1, 2.1))
```

```
# Create confidence interval plot for all pairwise comparisons  
plot(TukeyHSD(model, conf.level=.95), las = 2)
```



Formal Reporting of Two-Way ANOVA Findings

The final and most crucial step in any statistical analysis is clearly and concisely reporting the findings in a standard scientific format, ensuring that the results are accessible and contextualized. The report must include the type of test conducted, the variables involved, the test statistics, degrees of freedom, and the p-values for all main effects and the interaction effect. It must also reference the results of the post hoc tests to detail the specific differences found.

A **Two-Way ANOVA** was conducted to examine the simultaneous effects of gender (*male, female*) and exercise regimen (*none, light, intense*) on the dependent variable of weight loss (measured in pounds). The analysis confirmed a statistically significant interaction between the effects of gender and exercise on weight loss ($F(2, 54) = 4.615, p = 0.0141$). This significant interaction implies that the relationship between exercise intensity and weight loss is not constant across the two genders. Additionally, both main effects were highly significant: gender ($F(1, 54) = 11.197, p = 0.0015$) and exercise ($F(2, 54) = 179.087, p < 0.001$).

Due to the significance of the interaction term, [Tukey's HSD post hoc tests](#) were performed on the

simple main effects to dissect the specific group differences. The results showed substantial differences based on exercise intensity within each gender. For males, the *intense* exercise regimen led to significantly higher weight loss compared to both a *light* regimen ($p < .0001$) and a *no exercise* regimen ($p < .0001$). Furthermore, the *light* regimen resulted in significantly higher weight loss compared to the *no exercise* regimen ($p < .0001$).

Similarly, for females, the *intense* exercise regimen resulted in significantly higher weight loss when compared to both the *light* regimen ($p < .0001$) and the *no exercise* regimen ($p < .0001$). Importantly, the post hoc comparisons also revealed that males in the *intense* group lost significantly more weight than females in the *intense* group ($p = 0.0037$). All model assumptions, including Normality of residuals and the [Homogeneity of Variances](#) (Levene's test $p = 0.1177$), were successfully verified prior to the interpretation of the results.