

How to Perform a Two-Way ANOVA in R: A Step-by-Step Guide

Authored by
stats writer

December 30, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Perform a Two-Way ANOVA in R: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=110031>

The Two-Way ANOVA (Analysis of Variance) is a powerful statistical technique used to evaluate the simultaneous impact of two independent categorical factors (or "factors") on a single continuous dependent variable (or "response variable"). This method is essential for researchers looking to determine if there is a statistically significant difference between the means of three or more groups, where those groups are defined by the cross-classification of two factors.

To begin conducting a Two-Way ANOVA in R, the data must first be correctly structured into an R **data frame**. The primary tool for model fitting is the aov() function, which creates the ANOVA model object. The **summary() function** is then crucial for viewing the resulting ANOVA table and interpreting the F-statistics and P-values to draw reliable conclusions.

A Two-Way ANOVA is used to determine whether or not there is a statistically significant difference between the means of three or more independent groups that have been split based on two independent factors.

This tutorial explains, step-by-step, how to perform a Two-Way ANOVA in R, covering data preparation, model fitting, assumption checking, and post hoc analysis.

Setting Up the Example and Creating the Data

Suppose we want to determine if **exercise intensity** and **gender** independently impact **weight loss**. In this simulated experiment, the two independent factors we are studying are *exercise* (None, Light, Intense) and *gender* (Male, Female), and the continuous response variable is *weight loss*, measured in pounds.

We conduct a Two-Way ANOVA to assess both the individual impact of exercise and gender on weight loss (main effects) and to determine if there is a crucial ****interaction effect****--meaning the effect of exercise intensity depends on the participant's gender.

For this simulation, we recruit 30 men and 30 women (N=60 total participants). We randomly assign 10 of each gender to follow a program of either no exercise, light exercise, or intense exercise for one month, creating six distinct treatment groups.

The following code creates the necessary data frame in R that we will use for the subsequent analysis:

```
#make this example reproducible
```

```
set.seed(10)
```

```
#create data frame
```

```
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 first six rows of data frame
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
```

```
#see how many participants are in each group
table(data$gender, data$exercise)
```

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

Exploratory Data Analysis (EDA)

Before fitting any formal statistical model, it is best practice to perform Exploratory Data Analysis (EDA). This step helps us gain a crucial preliminary understanding of the data distribution by calculating descriptive statistics, such as the mean and standard deviation of weight loss for each of the six distinct treatment groups, using the **dplyr** package in R.

These descriptive summaries provide immediate context, showing which combination of factors appears to yield the highest or lowest response variable values before we test for statistical significance.

```
#load dplyr package
library(dplyr)
```

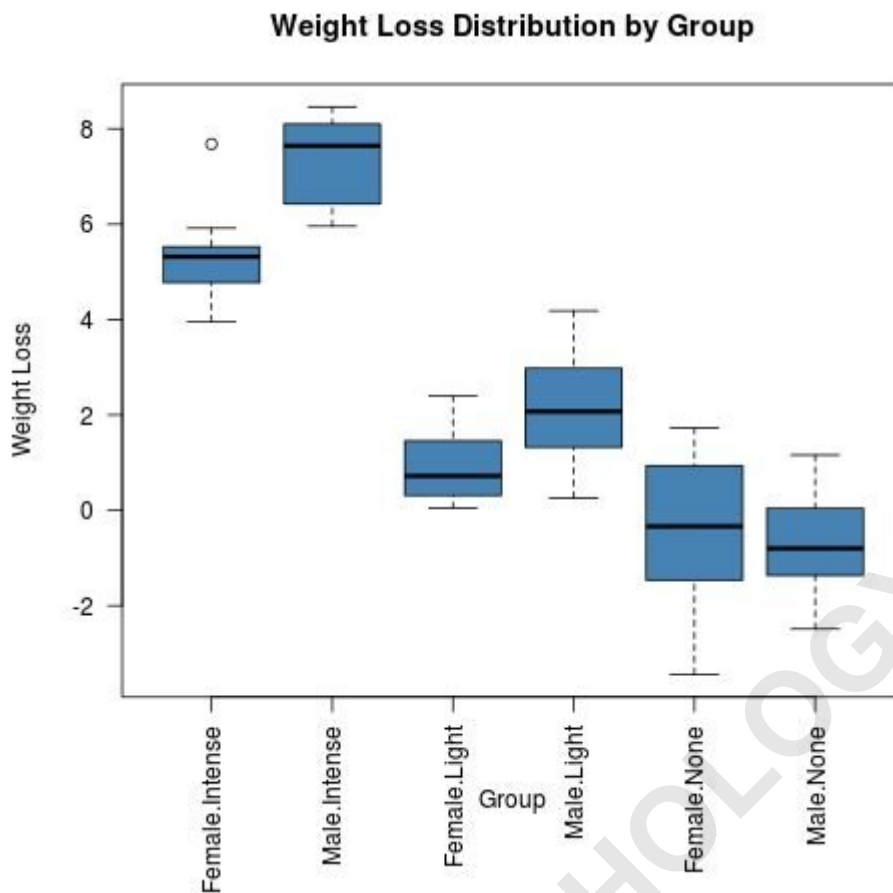
```
#find mean and standard deviation of weight loss for each treatment group
data %>%
group_by(gender, exercise) %>%
summarise(mean = mean(weight_loss),
sd = sd(weight_loss))
```

```
# 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
```

We can further visualize the distribution of weight loss for each of the six treatment groups by creating **boxplots**. Boxplots are excellent for quickly comparing the median, spread (interquartile range), and potential outliers across different categories.

```
#set margins so that axis labels on boxplot don't get cut off  
par(mar=c(8, 4.1, 4.1, 2.1))
```

```
#create boxplots  
boxplot(weight_loss ~ gender:exercise,  
data = data,  
main = "Weight Loss Distribution by Group",  
xlab = "Group",  
ylab = "Weight Loss",  
col = "steelblue",  
border = "black",  
las = 2 #make x-axis labels perpendicular  
)
```



Based on the visualization, the groups participating in *intense* exercise clearly show greater weight loss values compared to light or no exercise. Furthermore, there is a visual indication that males tend to have higher weight loss compared to females specifically in the *intense* and *light* exercise categories. We now fit the formal Two-Way ANOVA model to test if these visual differences are indeed statistically significant.

Fitting and Interpreting the Two-Way ANOVA Model

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

The asterisk (*) between the two ****predictor variables**** is essential, as it instructs the `aov()` function to test for the ****interaction effect**** between the two factors, in addition to testing their main effects.

In our example, we fit the `aov()` function using `weight_loss` as the response variable and `gender` and `exercise` as our two predictor variables. We then use the `summary()` function to generate and view the standard ANOVA output table:

#fit the two-way ANOVA model

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

```
#view the model 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
```

Interpreting the output, we find that the main effect of *gender*, the main effect of *exercise*, and the interaction between the two variables (*gender:exercise*) are all statistically significant ($P < 0.05$). Since the interaction term is significant ($p = 0.0141$), we must interpret the results by looking at how the factors combine, which will be achieved through a post hoc analysis.

Checking the Model Assumptions (Part 1: Independence & Normality)

Before relying on the P-values derived from the ANOVA model, we must verify that its underlying statistical assumptions are met. If these assumptions are violated, the results may be unreliable. A Two-Way ANOVA primarily assumes:

1. Independence: The observations in each group must be independent of one another. Since we used a **randomized experimental design**, participants' scores are unrelated, and this assumption is met by design.

2. Normality: The dependent variable (weight loss) should be approximately **normally distributed** within each combination of the two factors. This is typically assessed by examining the distribution of the **model residuals**.

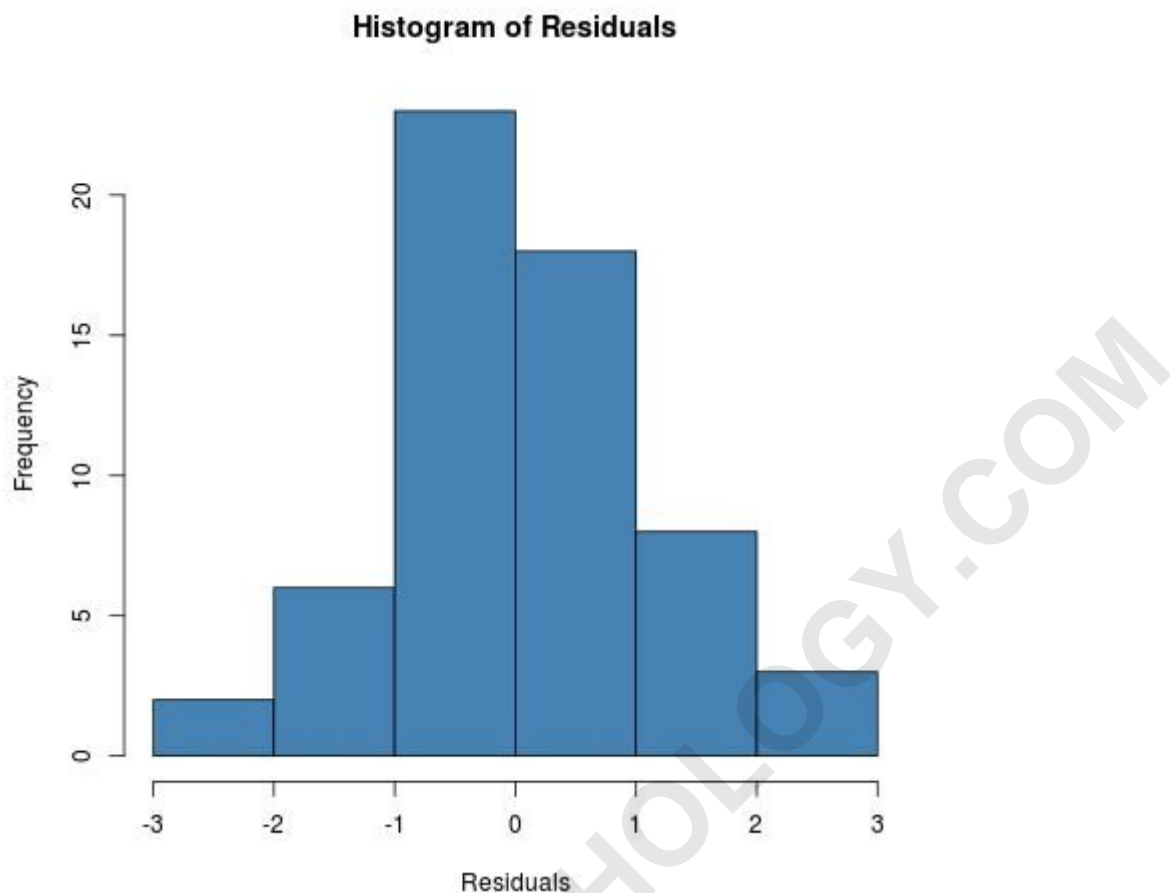
We check the **normality** assumption by extracting the model residuals and plotting them in a histogram. If the histogram displays a bell-shaped curve that is roughly centered around zero, the normality assumption is generally considered satisfied.

#define model residuals

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

#create histogram of residuals

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



The distribution of the **residuals** is relatively symmetrical and follows a mound shape, suggesting that the **normality** assumption is adequately met for this model.

Checking the Model Assumptions (Part 2: Homogeneity of Variance)

3. Equal Variance (Homogeneity of Variance): The variances (the spread of scores) for the dependent variable must be equal, or at least approximately equal, across all groups defined by the factors.

To formally check this assumption, we conduct the **Levene's Test for equality of variances**, which is robust to non-normality. This test requires the installation and loading of the **car** package in R:

```
#load car package
```

```
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 null hypothesis for Levene's Test assumes that the variances are equal (Equal Variance). Since the resulting p-value (0.1177) is greater than our typical significance level of 0.05, we fail to reject the null hypothesis. Thus, we conclude that the assumption of ****Homogeneity of Variance**** among groups is met, and our ANOVA results are reliable.

Analyzing Treatment Differences via Post Hoc Test

Given that the interaction effect in our ANOVA model was statistically significant, we must now conduct a post hoc test to precisely locate which of the six combined treatment groups differ from one another. Performing pairwise comparisons helps us fully understand the nature of the interaction.

For this multiple comparison procedure, we utilize the **TukeyHSD()** function, which performs Tukey's Honestly Significant Difference (HSD) test. This method effectively controls the Family-Wise Error Rate (FWER):

#perform Tukey's Test for multiple comparisons

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`
```

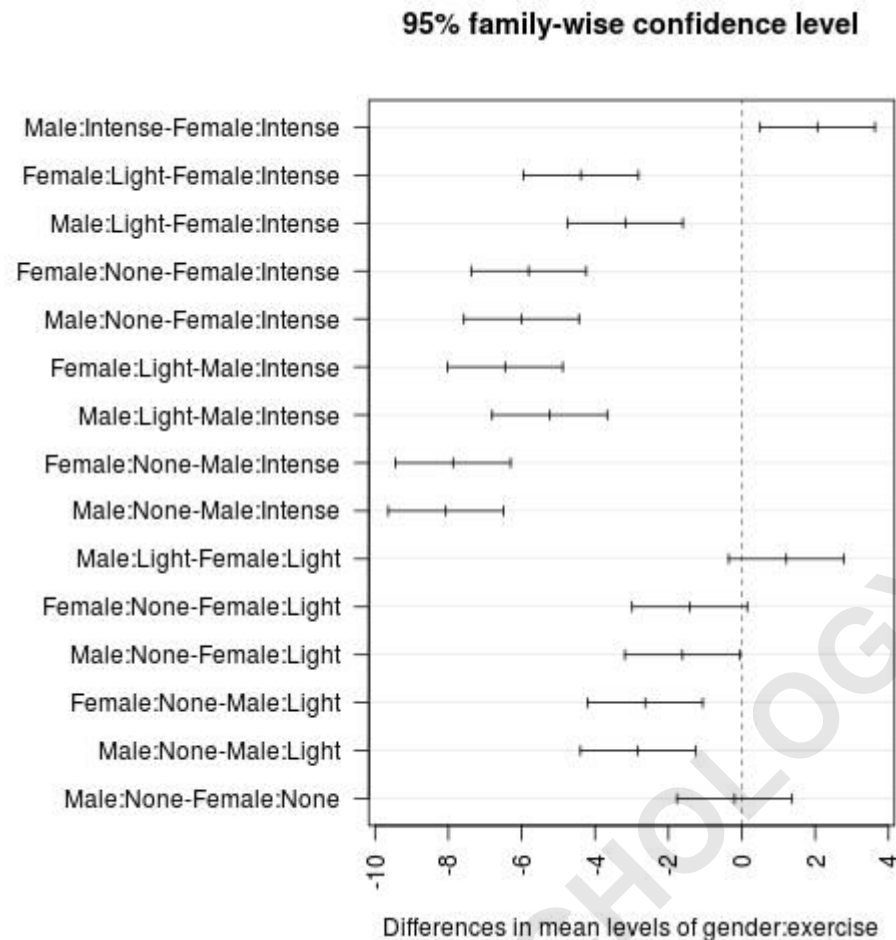
```
# diff lwr upr p adj
#Male:Intense-Female:Intense 2.0628297 0.4930588 3.63260067 0.0036746
#Female:Light-Female:Intense -4.3883563 -5.9581272 -2.81858535 0.0000000
#Male:Light-Female:Intense -3.1749419 -4.7447128 -1.60517092 0.0000027
#Female:None-Female:Intense -5.8091131 -7.3788841 -4.23934219 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
```

The crucial column in this output is `p adj` (adjusted P-value), which determines whether the mean difference (`diff`) between two groups is statistically significant. If `p adj` is less than 0.05, a significant difference exists. For example, the final row shows that "Male:None" compared to "Female:None" yields a P-value of 0.9990364, indicating no statistically significant difference in weight loss for individuals performing no exercise, regardless of gender.

We can also visualize the 95% **confidence intervals** resulting from the Tukey's HSD Test. Any interval in the plot that does not overlap with the vertical zero line represents a significant difference between the two groups compared:

```
#set axis margins so labels don't get cut off
par(mar=c(4.1, 13, 4.1, 2.1))
```

```
#create confidence interval for each comparison
plot(TukeyHSD(model, conf.level=.95), las = 2)
```



Reporting the Final Results of the Two-Way ANOVA

Finally, we compile the statistical findings into a cohesive report that summarizes the effects found, ensuring we highlight the significant interaction as the primary finding:

A Two-Way ANOVA was conducted to examine the effects of gender (*male, female*) and exercise regimen (*none, light, intense*) on weight loss (measured in lbs). Normality checks and Levene's Test confirmed that the model assumptions were met. Critically, there was a statistically significant interaction between the effects of gender and exercise on weight loss ($F(2, 54) = 4.615, p = 0.0141$). Tukey's HSD post hoc tests were carried out to explore this interaction.

For males, an *intense* exercise regimen led to significantly higher weight loss compared to both a *light* regimen ($p < .0001$) and *no exercise* regimen ($p < .0001$). Additionally, for males, a *light* regimen lead to significantly higher weight loss compared to *no exercise* regimen ($p < .0001$).

For females, an *intense* exercise regimen also led to significantly higher weight loss compared to both a *light* regimen ($p < .0001$) and *no exercise* regimen ($p < .0001$).

The significant interaction indicates that the benefit of intense exercise was stronger for males than for females (Male: Intense vs Female: Intense, $p = 0.0037$). These results provide evidence that both factors, and their combination, significantly influence the outcome of weight loss.

ARABPSYCHOLOGY.COM