

Learning to Calculate Eta Squared for ANOVA in R

Authored by
Mohammed looti

November 6, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Learning to Calculate Eta Squared for ANOVA in R*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=11522>

Understanding Eta Squared and Effect Size

[Eta Squared](#) (η^2) is a fundamental measure of [effect size](#) widely utilized in statistical analysis, particularly within [Analysis of Variance \(ANOVA\)](#) models. Its primary purpose is to move beyond mere statistical significance (p-values) by providing critical insight into the **practical significance** of research findings. By quantifying the magnitude of an effect, η^2 allows researchers to assess the real-world importance of the relationships observed in their data.

More precisely, [Eta Squared](#) is defined as the proportion of the total variability in the dependent variable that can be directly attributed to a specific main effect or an interaction effect within the broader [ANOVA](#) framework. This ratio reveals how much of the overall observed variation is explained by the independent variable under investigation. Understanding this proportion is essential for accurately gauging the strength of the association between the factors and the measured outcome, providing a clearer picture than significance testing alone.

The calculation of [Eta Squared](#) is inherently based on the partitioning of variance, making the formula intuitive and directly linked to the core principles of ANOVA. It is calculated as the ratio of the variation explained by the factor in question to the total variation across the entire model:

$$\text{Eta squared } (\eta^2) = \text{SS}_{\text{effect}} / \text{SS}_{\text{total}}$$

In this formula, the components are defined meticulously to isolate the variance contributions. **SS_{effect}** represents the [Sum of Squares](#) associated with the specific factor (e.g., a single independent variable or interaction term), quantifying the variability explained by that factor. Conversely, **SS_{total}** is the overall [Sum of Squares](#) for the entire [ANOVA](#) model, which captures all observed variability in the dataset.

The resulting value for [Eta Squared](#) spans a range from 0 to 1. A score closer to 1 signifies that a substantial proportion of the total variance is successfully accounted for by the variable under scrutiny, indicating a powerful [effect size](#). Conversely, values approaching 0 suggest that the variable explains very little of the outcome's variability. Researchers typically rely on established conventions, such as those popularized by Cohen, to categorize the practical magnitude of the calculated η^2 :

.01: Defined as a **Small effect size**, implying a minor or negligible influence on the dependent measure.

.06: Characterized as a **Medium effect size**, suggesting a moderate and potentially noticeable influence.

.14 or higher: Classified as a **Large effect size**, representing a substantial and practically important influence on the outcome.

Preparing the R Environment for Effect Size Calculation

The statistical computing environment known as [R](#) is indispensable for fitting sophisticated statistical models and calculating essential measures of [effect size](#). While the base [R](#) distribution is fully capable of executing the primary [ANOVA](#) procedures, specialized extension packages are typically required to efficiently extract and compute advanced effect size metrics like Eta Squared, streamlining the analytical workflow significantly.

For the purposes of this tutorial, we will rely on the `lsr` package, an abbreviation for "Learning Statistics with R." This package is highly recommended for beginners and experts alike, as it contains the crucial `etaSquared()` function. This dedicated function is designed to quickly and accurately compute η^2 values directly from fitted ANOVA model objects, eliminating the need for tedious manual calculations derived from the standard ANOVA table output.

This step-by-step guide will walk you through the entire process: defining an experimental design, generating reproducible data, fitting the model, and finally, utilizing the `lsr` package to calculate and rigorously interpret [Eta Squared](#) for factors within your ANOVA model within the versatile [R](#) ecosystem.

Step 1: Defining the Experimental Design and Data Creation

To effectively demonstrate the calculation procedure for Eta Squared, we must first establish a robust, hypothetical study dataset. Our central research inquiry focuses on assessing whether two distinct independent variables--exercise intensity and gender--exert a significant and measurable impact on the dependent variable, which we define as weight loss. This specific structure mandates the application of a two-way [ANOVA](#) design to analyze the resulting variance.

Our experimental framework involves a total of 60 participants, evenly split between 30 men and 30 women. Crucially, 10 participants of each gender are randomly assigned to one of three distinct exercise conditions: a control group receiving no exercise, a light exercise regimen, or an intense exercise regimen. This results in six total experimental cells (2 genders \times 3 intensity levels). The outcome measure, weight loss, is recorded in kilograms over a standardized period of one month.

The following R code block illustrates the process of generating a synthetic data frame that accurately mirrors this factorial experimental design. To ensure that our results are completely reproducible, we initialize the process using the command `set.seed(10)`. The `runif()` function is then strategically employed to simulate weight loss data, generating numerical values that reflect the hypothesized differences across the varying exercise and gender groups.

#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
```

Step 2: Fitting the Two-Way ANOVA Model in R

Following the successful construction of the data frame, the next crucial step in the analytical pipeline involves fitting the two-way ANOVA model using R's fundamental function, `aov()`. This process requires the explicit specification of the dependent variable (`weight_loss`) and the independent factor variables (`gender` and `exercise` intensity) that are hypothesized to influence the outcome.

The model formula, written as `weight_loss ~ gender + exercise`, instructs R to model the variation in weight loss as a function of the primary, or main, effects of both gender and exercise intensity. It is worth noting that, for the sake of obtaining pure η^2 values, we have intentionally excluded the interaction term (`gender $times$ exercise`) in this specific formulation. However, incorporating interaction effects is straightforward should the research question necessitate investigating combined factor influences.

Executing the model fit and subsequently reviewing the `summary()` output generates the classic ANOVA table. This output is critical because it meticulously partitions the total observed [Sum of Squares](#) into components uniquely attributable to each factor and the remaining residual error. This standard table is the primary source for determining **statistical significance** (via the F-statistic and P-value) before moving forward to the assessment of **practical significance** through [Effect Size](#) measures.

#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 9.916 0.00263 **
exercise 2 505.6 252.78 158.610 < 2e-16 ***
Residuals 56 89.2 1.59
```

The initial model summary immediately reveals that both independent factors are statistically significant predictors of weight loss, as evidenced by P-values falling well below the conventional α threshold of 0.05. However, the strikingly large F-value and the near-zero P-value associated with the `exercise` variable strongly suggest that its influence is vastly greater than that of `gender`. This initial disparity in statistical evidence must now be rigorously quantified and confirmed by calculating standardized [effect size](#) metrics.

Step 3: Calculating Eta Squared Using the `lsr` Package

Although the standard ANOVA summary table furnishes all the necessary [Sum of Squares](#) (SS) values required for manual Eta Squared calculation, modern statistical practice in [R](#) advocates for leveraging dedicated functions. We utilize the highly efficient `etaSquared()` function, sourced from the `lsr` package, to automate this complex computation. This function is designed to accept the fitted `aov` object directly as its input, providing clean and labeled output instantly.

A significant benefit of employing `etaSquared()` is its ability to simultaneously calculate two critical effect size measures: the traditional [Eta Squared](#) (η^2 , outputted as `eta.sq`) and Partial Eta Squared (η_p^2 , outputted as `eta.sq.part`). Traditional η^2 quantifies the variance explained by a factor relative to the total variance in the entire model, whereas η_p^2 calculates the variance explained by a factor relative only to the variance of that factor plus the error variance, effectively excluding other factors. Because these two measures offer slightly different perspectives on effect magnitude, researchers must be clear about which statistic they are

reporting. For this tutorial, we focus on the interpretation of the traditional η^2 .

Before running the calculation, the package must be loaded into the R session using the `library()` command. The output below clearly presents the results for both metrics across the factors of gender and exercise:

#load lsr package

```
library(lsr)
```

```
#calculate Eta Squared
```

```
etaSquared(model)
```

```
eta.sq eta.sq.part
```

```
gender 0.0258824 0.1504401
```

```
exercise 0.8279555 0.8499543
```

Interpreting the Practical Significance of the Results

The resulting output from the `etaSquared()` function provides the concrete quantitative evidence necessary to assess the practical significance of our two factors. We now isolate the traditional [Eta Squared](#) values (`eta.sq`) for interpretation against Cohen's established guidelines:

Eta squared (η^2) for gender: **0.0258824**

Eta squared (η^2) for exercise: **0.8279555**

Applying the standard interpretation framework to these derived values yields clear conclusions regarding the strength and importance of each factor's influence on weight loss:

The η^2 value calculated for **gender** (approximately 0.026) falls squarely within the range designated for a **small effect size** (0.01 to 0.06). Despite achieving statistical significance ($P = 0.00263$), gender explains only about 2.6% of the total variance observed in weight loss. This suggests that while gender differences exist, their contribution to overall variability is minor in the context of this study.

In stark contrast, the η^2 value for **exercise intensity** (approximately 0.828) is dramatically higher than the large effect threshold of 0.14. This result mandates classification as an **extremely large effect size**. Exercise intensity successfully accounts for nearly 83% of the total variance in weight loss across the entire sample group, indicating that it is the overwhelmingly dominant factor driving the observed outcome.

These effect size calculations serve to rigorously confirm and significantly strengthen the initial, qualitative observations drawn from the standard ANOVA table. The factor exhibiting the smallest

p-value (exercise) is also demonstrably the factor exhibiting the vastly larger practical effect size, measured by η^2 . This powerful example underscores the critical importance of integrating standardized effect size measures, such as Eta Squared, alongside traditional hypothesis testing to deliver a comprehensive, robust, and complete interpretation of research findings.

Further Resources for Advanced ANOVA in R

Mastering the calculation and interpretation of Eta Squared represents a key milestone toward conducting comprehensive and reproducible statistical analysis within [R](#). For analysts and researchers eager to expand their capabilities to handle more intricate experimental designs, such as complex factorial layouts or models involving repeated measures, further exploration of R's statistical modeling functions is highly recommended.

The following resources provide guidance on advanced techniques for fitting various ANOVA models and extracting detailed statistical information necessary for advanced reporting: