

How to calculate Eta Squared in R?

Authored by
stats writer

December 15, 2025

RECOMMENDED CITATION

stats writer (2025). *How to calculate Eta Squared in R?*. PSYCHOLOGICAL SCALES.
Retrieved from <https://scales.arabpsychology.com/?p=107564>

Eta squared (η^2) is a crucial measure of effect size widely employed in statistical analysis, particularly within ANOVA (Analysis of Variance) models. Researchers utilize η^2 to quantify the strength of the relationship between predictors (independent variables) and the outcome (dependent variable). Unlike p-values, which only indicate statistical significance, eta squared provides a practical measure of the magnitude of the observed effect, helping practitioners understand the real-world importance of their findings.

Specifically, η^2 measures the proportion of total variance in the dependent variable that can be attributed to a specific factor (a main effect or an interaction effect) within the ANOVA model. This value allows for standardized comparison across different studies, providing context often missing when relying solely on null hypothesis significance testing (NHST). Understanding effect size is paramount for robust research interpretation and power analysis, ensuring that statistically significant findings also possess practical relevance.

The Calculation and Formal Definition of Eta Squared

The calculation of Eta squared relies on partitioning the variability observed in the data. This involves determining the variability explained by the experimental factors relative to the overall variability present in the dataset. This metric is a fundamental component of variance accountability in statistical modeling. The fundamental formula for calculating Eta squared is defined by the ratio of the Sum of Squares for the effect being tested to the Total Sum of Squares in the model:

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

Where the components are rigorously defined as follows:

SS_{effect}: This represents the Sum of Squares attributable to a specific effect (e.g., the main factor for "Gender" or an interaction term). It quantifies the amount of variability in the outcome explained by that particular factor.

SS_{total}: This is the Total Sum of Squares, representing the total variability in the dependent variable across all observations in the ANOVA model.

Since η^2 is inherently a proportion of variance explained, its value is bounded between 0 and 1. A calculated value approaching 1 signifies that the tested variable accounts for nearly all the variability in the outcome, indicating a remarkably powerful effect. Conversely, values closer to 0 suggest that the variable provides minimal explanatory power relative to the total variance.

Interpreting the Magnitude of Practical Significance

While the quantitative value of Eta squared is mathematically precise, interpreting its practical

significance requires established benchmarks. The commonly accepted rules of thumb, derived primarily from guidelines proposed by Jacob Cohen, allow researchers to categorize the strength of an effect size, thereby providing a standardized qualitative assessment.

These benchmarks are essential for contextualizing results across different studies and disciplines. The following rules of thumb are typically used to interpret the magnitude of calculated Eta squared values in behavioral and social sciences:

.01: Corresponds to a **Small effect size**. The factor explains 1% of the total variance.

.06: Corresponds to a **Medium effect size**. The factor explains 6% of the total variance.

.14 or higher: Corresponds to a **Large effect size**. The factor explains 14% or more of the total variance.

It is crucial for analysts to remember that these guidelines serve as general references. In fields where experimental control is difficult and inherent variability is high (such as clinical psychology), even a small effect size might represent a significant theoretical breakthrough or clinical relevance. Conversely, in highly controlled physical sciences, researchers might expect and require larger effect sizes.

Prerequisites and Calculation Environment in R

The standard R distribution includes the necessary functions to fit ANOVA models (e.g., `aov()`). However, calculating effect size measures like η^2 is not a default output of the base `summary(aov())` function. To streamline the calculation and ensure accuracy, we rely on specialized packages within the statistical programming language R.

For this tutorial, we will utilize the `lsr` package (Learning Statistics with R). This package provides the dedicated `etaSquared()` function, which efficiently computes both standard Eta squared and Partial Eta squared directly from the fitted ANOVA model object. Before proceeding, ensure that the `lsr` package is installed and loaded into your current R session.

Step 1: Setting up the Experimental Data Frame in R

To demonstrate the calculation process, we simulate data from a hypothetical two-factor experiment designed to assess the impact of Exercise Intensity and Gender on Weight Loss. This setup requires generating a balanced dataset suitable for a two-way factorial ANOVA.

The experimental structure involves 60 participants: 30 males and 30 females. Each gender group is equally distributed across three levels of Exercise Intensity ("None," "Light," and "Intense"). This results in 10 participants per cell (Gender \times Exercise level), ensuring orthogonality and statistical balance. The outcome, `weight_loss`, is generated using the `runif()` function,

simulating different ranges of weight loss based on the intensity level, thereby mimicking a plausible experimental outcome where intense exercise leads to greater weight loss.

The following code block demonstrates the creation of this structure, utilizing `set.seed(10)` for reproducibility and the `rep()` function to generate the categorical factor levels efficiently. We then inspect the resulting data frame structure using `head()` and confirm the cell counts using `table()`.

#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

The confirmation via the `table()` output is vital, as it guarantees that the structure necessary for a standard two-way ANOVA is satisfied, with equal sample sizes across all treatment combinations (10 subjects per group).

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

With the data frame correctly configured, the subsequent step is to execute the statistical test using

R's base statistical functions. We employ the `aov()` function to fit a linear model, specifying `weight_loss` as the outcome variable and `gender` and `exercise` as the main factors contributing to the variance.

The syntax `aov(weight_loss ~ gender + exercise, data = data)` defines the model structure. By assigning the result to the object `model`, we store all the necessary statistical information, including the crucial Sum of Squares values required for the calculation of Eta squared. Reviewing the `summary()` output gives us the standard inferential statistics:

#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 summary table confirms that both independent variables are statistically significant at a conventional $\alpha = 0.05$ level. The p-value for gender (0.00263) and the extremely low p-value for exercise ($< 2 \times 10^{-16}$) indicate that the observed differences in weight loss across levels of both factors are highly unlikely to have occurred by random chance. The Sum of Squares column (Sum Sq) contains the raw variability measures (15.8 for Gender, 505.6 for Exercise, and 89.2 for Residuals), which are the building blocks for Eta squared.

Step 3: Calculating Eta Squared Using the lsr Package

To move beyond statistical significance and quantify the practical relevance of the findings, we now calculate Eta squared. This is achieved by loading the `lsr` package and applying its specialized function, `etaSquared()`, to our fitted ANOVA model object.

Loading the library ensures that the required effect size computation functions are available in the R environment. The `etaSquared()` function automates the process of extracting the necessary Sum of Squares values from the model output and performing the division required for the effect size calculation.

```
#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
```

The output displays two distinct types of Eta squared measures: `eta.sq` and `eta.sq.part`.

Eta Squared (η^2 , represented by `eta.sq`): This is calculated as $SS_{\text{effect}} / SS_{\text{total}}$. It accounts for the variance explained by the factor relative to all variance in the model, including the variance explained by other factors.

Partial Eta Squared (η_p^2 , represented by `eta.sq.part`): This is calculated as $SS_{\text{effect}} / (SS_{\text{effect}} + SS_{\text{error}})$. It expresses the variance explained by the factor relative only to the variance not explained by that factor (excluding other factors from the denominator), often resulting in higher values.

For the purpose of evaluating the proportion of total variance explained across the entire study, the standard Eta squared (η^2) is often preferred, particularly when reporting findings consistent with the overall structure of the ANOVA model.

Interpreting the Practical Findings

Focusing on the standard η^2 values provided by the `lsr` package, we can now draw definitive conclusions about the practical impact of our variables:

Eta squared (η^2) for gender: **0.0258824**
Eta squared (η^2) for exercise: **0.8279555**

For the factor **Gender**, the η^2 value of approximately 0.026 falls between Cohen's small (0.01) and medium (0.06) effect size categories, leaning toward small. This indicates that while gender differences in weight loss exist and are statistically significant, they account for only about 2.6% of the total variability observed in the dependent measure. This is a small practical contribution.

In sharp contrast, the factor **Exercise** yielded an extraordinarily high η^2 value of approximately 0.828. This value far exceeds the threshold for a large effect size (0.14) and confirms that exercise intensity is the primary driver of weight loss in this experimental design. Exercise explains nearly 83% of the total variance in weight loss.

Conclusion: Combining Significance and Effect Size

The synthesis of the ANOVA p-values and the Eta squared effect sizes provides a complete picture of the experimental results. Both Gender and Exercise were statistically significant predictors. However, the η^2 values clarified that the effect of exercise is overwhelmingly dominant in magnitude compared to the effect of gender.

These results confirm the initial hypothesis implied by the data simulation: the level of exercise intensity is a highly impactful factor in predicting weight loss, holding massive practical significance. While the gender difference is statistically robust, its contribution to the overall observed variability is minor. Reporting both the p-value and the Eta squared ensures transparent and comprehensive reporting of the research findings, adhering to best practices in quantitative statistics.

For those seeking to explore further advanced statistical modeling in R, numerous resources are available detailing the fitting of other complex ANOVA designs.