

Conduct a MANOVA in R

Authored by
Mohammed looti

November 9, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Conduct a MANOVA in R*. PSYCHOLOGICAL STATISTICS.
Retrieved from <https://statistics.arabpsychology.com/?p=14249>

Understanding the Foundations: The Analysis of Variance (ANOVA)

Before diving into the complexity of multivariate statistics, it is crucial to establish a strong understanding of the standard [ANOVA](#) (Analysis of Variance). An [ANOVA](#) is a powerful inferential statistical technique used to determine whether or not there is a [statistically significant difference](#) between the means of three or more independent groups. It serves as the foundational stepping stone toward understanding its multivariate counterpart. The core limitation of the standard [ANOVA](#), however, is that it can only handle a single response (dependent) variable at a time.

Consider a scenario where researchers are investigating the effectiveness of different studying techniques on student performance. They randomly assign a class of students into three distinct groups, each utilizing a unique method for a month to prepare for a standardized exam. The primary outcome measure is the score achieved on that single exam. In this design, we have one categorical independent variable (Studying Technique, with three levels) and one continuous response variable (Exam Score).

To analyze this data and determine if the studying technique truly impacts the results, we would conduct a one-way [ANOVA](#). This analysis would test the null hypothesis that the mean exam scores across all three groups are equal. If the resulting F-statistic is large enough, yielding a small p-value, we can reject the null hypothesis and conclude that there is a [statistically significant difference](#) in mean scores attributable to the studying technique.



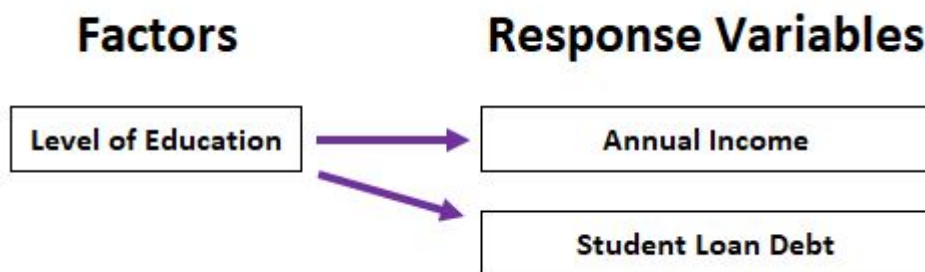
Transitioning to Multivariate Analysis: MANOVA Defined

While [ANOVA](#) is suitable for single outcomes, many real-world research questions involve assessing the simultaneous impact of a factor on several interrelated dependent variables. This is where the **MANOVA** (Multivariate Analysis of Variance) becomes essential. The primary distinction between the two techniques lies in the number of dependent variables: [ANOVA](#) uses one response variable, whereas **MANOVA** utilizes multiple response variables, measured concurrently.

By analyzing multiple dependent variables simultaneously, **MANOVA** offers two significant advantages over running multiple individual ANOVAs. First, it controls the overall Type I error rate (the risk of falsely rejecting a true null hypothesis) that would inflate if multiple separate tests were

conducted. Second, and more importantly, **MANOVA** considers the correlations between the dependent variables. It tests whether the independent variable has a [statistically significant difference](#) on the group means of the dependent variables, considered collectively as a composite variable.

For instance, suppose we want to investigate how the level of education (e.g., High School Diploma, Associate's Degree, Bachelor's Degree, Master's Degree) impacts both annual income and the accumulated amount of student loan debt. In this scenario, we have one factor (Level of Education) and two distinct, yet potentially related, continuous response variables (Annual Income and Student Loan Debt). Because we are examining the effect of a single factor on two or more dependent variables, a one-way **MANOVA** is the appropriate statistical methodology.



Case Study: Implementing MANOVA in R

To illustrate the practical application of **MANOVA**, we will conduct an analysis using the statistical programming language [R](#). We will utilize the built-in **iris** dataset, a classic dataset in statistics and machine learning, which contains measurements in centimeters of the sepal length, sepal width, petal length, and petal width for 150 iris flowers, sampled from three distinct species: *setosa*, *virginica*, and *versicolor*. This dataset is ideal for a one-way **MANOVA** because it features one categorical independent variable (Species) and multiple continuous dependent variables (the four measurements).

Our specific research question is: Does the species of the iris flower have a [statistically significant difference](#) on its sepal measurements (Sepal Length and Sepal Width) collectively? By using *Species* as the independent variable and treating *Sepal Length* and *Sepal Width* as our two response variables, we can proceed with the analysis. The first step, as always in data analysis using [R](#), is to familiarize ourselves with the structure of the data:

```
#view first six rows of iris dataset  
head(iris)
```

```
# Sepal.Length Sepal.Width Petal.Length Petal.Width Species
#1 5.1 3.5 1.4 0.2 setosa
#2 4.9 3.0 1.4 0.2 setosa
#3 4.7 3.2 1.3 0.2 setosa
#4 4.6 3.1 1.5 0.2 setosa
#5 5.0 3.6 1.4 0.2 setosa
#6 5.4 3.9 1.7 0.4 setosa
```

To perform the **MANOVA** in [R](#), we utilize the native **manova()** function. This function requires a specific syntax to handle the simultaneous inclusion of multiple response variables. These dependent variables must be combined using the **cbind()** function, which binds the specified columns together for the multivariate test. The structure of the function is defined clearly below, demonstrating how the response variables (rvs) are modeled against the independent variable (iv).

manova(cbind(rv1, rv2, ...) ~ iv, data)

The arguments used in the **manova()** function are defined as follows:

rv1, rv2: These represent the response variable 1, response variable 2, and so on--the dependent measures being analyzed.

iv: This is the independent variable, or factor, whose levels are being compared (e.g., Species).

data: This specifies the name of the data frame containing all the necessary variables.

Applying this syntax to our iris dataset, where *Sepal.Length* and *Sepal.Width* are the response variables and *Species* is the independent variable, we fit the model and then immediately generate the summary output to examine the results:

#fit the MANOVA model

```
model <- manova(cbind(Sepal.Length, Sepal.Width) ~ Species, data = iris)
```

```
#view the results
```

```
summary(model)
```

```
# Df Pillai approx F num Df den Df Pr(>F)
```

```
#Species 2 0.94531 65.878 4 294 < 2.2e-16 ***
```

```
#Residuals 147
```

```
#---
```

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

Interpreting the MANOVA Output and Key Statistics

The primary output of the **MANOVA** provides critical information regarding the overall effect of the independent variable on the set of dependent variables. The first row, labeled "Species," contains the results of the multivariate test. By default, R's `manova()` function uses the **Pillai** test statistic, specifically [Pillai's trace](#) (or Pillai-Bartlett Trace), which is generally considered the most robust test, especially when assumptions regarding homogeneity of variance-covariance matrices may be violated.

In our output, the [Pillai's trace](#) value is 0.94531. Since the exact distribution of this multivariate test statistic is mathematically complex, the output also provides an approximate F value (65.878) along with its corresponding degrees of freedom (num Df = 4, den Df = 294) to facilitate easier interpretation. The most crucial piece of information is the p-value, denoted by **Pr(>F)**. In this case, the p-value is extremely small ($< 2.2e-16$).

Since our p-value is far less than the conventional significance level of 0.05, we confidently reject the null hypothesis. The conclusion is that there is a [statistically significant difference](#) in the combined mean sepal measurements (length and width) among the three iris species. In essence, the species classification does influence the overall sepal morphology.

Technical Note: While [Pillai's trace](#) is the default and often preferred test statistic due to its robustness, the `manova()` function allows researchers to specify other common multivariate test statistics, including "Roy," "Hotelling-Lawley," or "Wilks." This can be achieved by modifying the summary function call, such as: `summary(model, test = 'Wilks')`.

Deeper Dive: Univariate Follow-up Analyses

The overall significant **MANOVA** result tells us that the group means are different across the dependent variables taken as a set, but it does not specify which particular response variable(s) contributed to this multivariate effect. To uncover the specific relationship between the independent variable (Species) and each individual dependent variable (Sepal Length and Sepal Width), we must perform follow-up univariate ANOVAs.

In R, we can easily extract these individual [ANOVA](#) results from the fitted **MANOVA** model using the `summary.aov()` function. This method is computationally efficient because it reuses the variance partitions already calculated within the **MANOVA** framework. The output will provide an [ANOVA](#) summary table for each dependent variable included in the original model.

`summary.aov(model)`

```
# Response Sepal.Length :  
# Df Sum Sq Mean Sq F value Pr(>F)  
#Species 2 63.212 31.606 119.26 < 2.2e-16 ***
```

```
#Residuals 147 38.956 0.265
#---
#Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Response Sepal.Width :
# Df Sum Sq Mean Sq F value Pr(>F)
#Species 2 11.345 5.6725 49.16 < 2.2e-16 ***
#Residuals 147 16.962 0.1154
#---
#Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results of the univariate ANOVAs clearly indicate that the factor *Species* exerts a highly [statistically significant difference](#) on both response variables independently. For **Sepal.Length**, the F-value is 119.26 with a p-value of $< 2.2e-16$. Similarly, for **Sepal.Width**, the F-value is 49.16, also yielding an extremely low p-value. Therefore, we can conclude that the species of the iris flower affects both the length and the width of the sepal measurements individually. Had one of these p-values been non-significant (i.e., > 0.05), we would conclude that the overall **MANOVA** effect was primarily driven by the other variable.

Visualizing and Confirming Group Differences

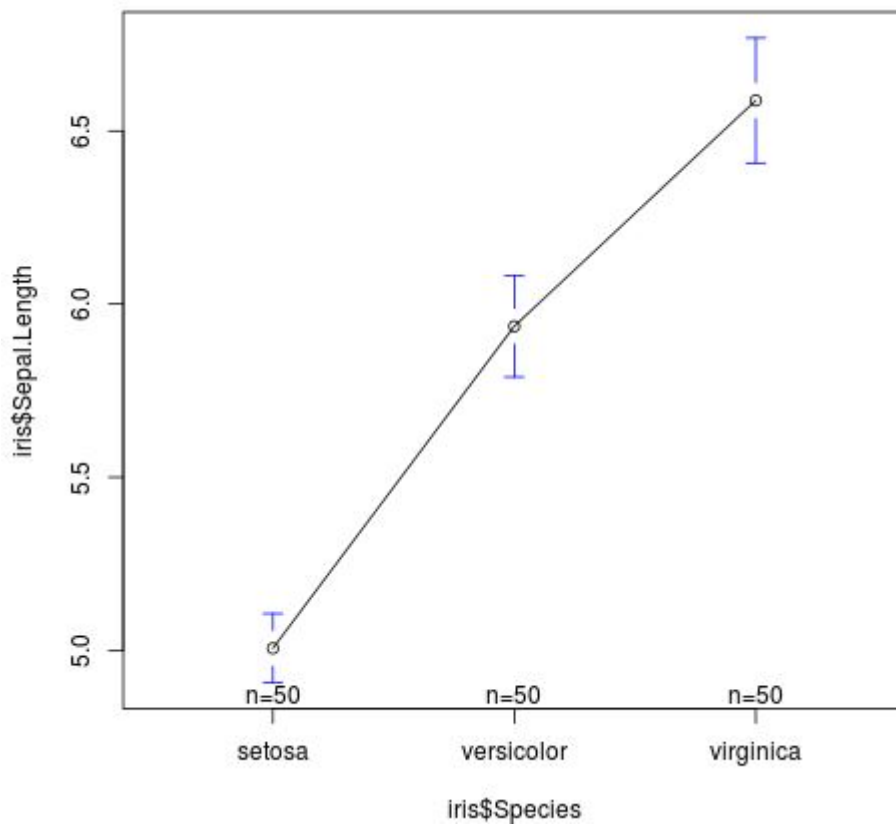
While statistical tests confirm significance, visualizing the data is crucial for gaining an intuitive and comprehensive understanding of the magnitude and direction of the differences between groups. Plotting the group means for each dependent variable helps to connect the statistical results back to the observed data patterns.

To facilitate this visualization, we can use the **gplots** library in [R](#), which offers the convenient **plotmeans()** function. This function generates a plot showing the mean of a continuous variable for each level of a factor, often including confidence intervals to illustrate variability and precision.

We begin by visualizing the mean *Sepal Length* across the three species:

```
#load gplots library
library(gplots)

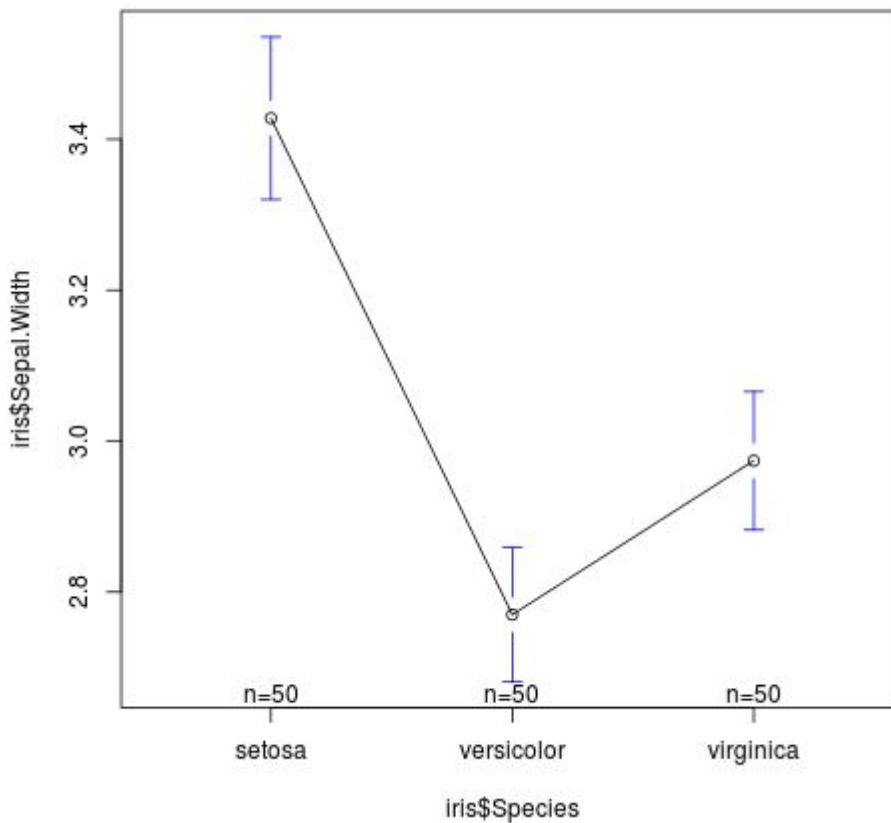
#visualize mean sepal length by species
plotmeans(iris$Sepal.Length ~ iris$Species)
```



The plot clearly illustrates substantial differences in mean sepal length among the three species. The *setosa* species has the shortest mean sepal length, while *virginica* exhibits the longest. The lack of overlap in the confidence intervals visually reinforces the finding from the **MANOVA** and the univariate [ANOVA](#) that species significantly impacts this measure.

Next, we perform the same visualization for the second response variable, *Sepal Width*:

```
plotmeans(iris$Sepal.Width ~ iris$Species)
```



For sepal width, the pattern is slightly different: *setosa* has the widest sepal, while *versicolor* and *virginica* have narrower widths, with *versicolor* being the narrowest. Again, the distinct separation of the mean points confirms that *Species* has a strong, [statistically significant difference](#) on sepal width. These visualizations serve as robust confirmation for the calculated multivariate and univariate test results, demonstrating that the morphological measurements of the iris flower are indeed highly dependent on its species classification.

For advanced users seeking more detailed parameter options and syntax specifics, consulting the official R documentation for the **manova()** function is highly recommended.