

# When should I use `aov()` versus `anova()` in R?

Authored by  
**stats writer**

June 27, 2024

## RECOMMENDED CITATION

stats writer (2024). *When should I use `aov()` versus `anova()` in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=155018>

The `aov()` and `anova()` functions in R are used for conducting analysis of variance (ANOVA) tests. While both functions serve the same purpose, there are slight differences in their usage. The `aov()` function is typically used for conducting ANOVA tests on data with a single factor, while the `anova()` function is more suitable for data with multiple factors. Additionally, the `aov()` function allows for more flexibility in the model specification, while the `anova()` function requires a specific model formula. Therefore, it is important to carefully consider the nature of the data and the desired model before deciding whether to use `aov()` or `anova()` for ANOVA testing in R.

## When to Use `aov()` vs. `anova()` in R

**The `aov()` and `anova()` functions in R seem similar, but we actually use them in two different scenarios.**

**We use `aov()` when we would like to fit an ANOVA model and view the results in an ANOVA summary table.**

**We use `anova()` when we would like to compare the fit of nested regression models to determine if a regression model with a certain set of coefficients offers a significantly better fit than a model with only a subset of the coefficients.**

**The following examples show how to use each function in practice.**

### Example 1: How to Use `aov()` in R

**Suppose we would like to perform a to determine if three different exercise programs impact weight loss**

**differently.**

**We recruit 90 people to participate in an experiment in which we randomly assign 30 people to follow either program A, program B, or program C for one month.**

**The following code shows how to use the aov() function in R to perform this one-way ANOVA:**

```
#make this example reproducible  
set.seed(0)  
  
#create data frame  
df <- data.frame(program = rep(c("A", "B", "C"),  
each=30),  
weight_loss = c(runif(30, 0, 3),  
runif(30, 0, 5),  
runif(30, 1, 7)))  
  
#fit one-way anova using aov()  
fit <- aov(weight_loss ~ program, data=df)  
  
#view results  
summary(fit)  
  
Df Sum Sq Mean Sq F value Pr(>F)
```

```
program 2 98.93 49.46 30.83 7.55e-11 ***
Residuals 87 139.57 1.60
```

```
---
```

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

From the model output we can see that the p-value for program (.0000000000755) is less than .05, which means there is a statistically significant difference in mean weight loss between the three programs.

Example 2: How to Use anova() in R

Suppose we would like to use number of hours studied to predict exam score for students at a certain college. We may decide to fit the following two regression models:

**Full Model:  $\text{Score} = \beta_0 + B_1(\text{hours}) + B_2(\text{hours})^2$**

**Reduced Model:  $\text{Score} = \beta_0 + B_1(\text{hours})$**

The following code shows how to use the anova() function in R to perform a lack of fit test to determine if the full model offers a significantly better fit than the reduced model:

```
#make this example reproducible
```

```
set.seed(1)
```

```
#create dataset
```

```
df <- data.frame(hours = runif(50, 5, 15), score=50)
```

```
df$score = df$score + df$hours^3/150 +  
df$hours*runif(50, 1, 2)
```

```
#view head of data
```

```
head(df)
```

```
hours score
```

```
1 7.655087 64.30191
```

```
2 8.721239 70.65430
```

```
3 10.728534 73.66114
```

```
4 14.082078 86.14630
```

```
5 7.016819 59.81595
```

```
6 13.983897 83.60510
```

```
#fit full model
```

```
full <- lm(score ~ poly(hours,2), data=df)
```

```
#fit reduced model
```

```
reduced <- lm(score ~ hours, data=df)
```

```
#perform lack of fit test using anova()
```

## anova(full, reduced)

### Analysis of Variance Table

Model 1: score ~ poly(hours, 2)

Model 2: score ~ hours

Res.Df RSS Df Sum of Sq F Pr(>F)

1 47 368.48

2 48 451.22 -1 -82.744 10.554 0.002144 \*\*

---

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

Since the p-value in the output table (.002144) is less than .05, we can reject the null hypothesis of the test and conclude that the full model offers a statistically significantly better fit than the reduced model.