

Performing ANCOVA in R: A Step-by-Step Tutorial

Authored by
Mohammed looti

November 9, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Performing ANCOVA in R: A Step-by-Step Tutorial*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=14246>

This comprehensive tutorial demonstrates how to perform and interpret an [ANCOVA](#) (Analysis of Covariance) within the [R](#) statistical environment. ANCOVA is a powerful statistical tool that combines elements of ANOVA and regression, allowing researchers to test the differences between group means while statistically controlling for the effects of an extraneous continuous variable, known as the **covariate**. Understanding how to properly structure, fit, and interpret this model is essential for ensuring robust research findings.

Understanding the ANCOVA Framework in R

The primary purpose of using [ANCOVA](#) is to increase statistical power by accounting for variability in the dependent variable that is explained by the covariate. By removing this variance, the model can provide a cleaner assessment of the effect of the primary independent variable (or treatment) on the response variable. This adjustment is particularly useful in observational studies or quasi-experimental designs where strict random assignment cannot fully control for baseline differences among groups.

In the context of statistical modeling, the ANCOVA framework requires careful definition of three distinct types of variables. The fidelity of the resulting analysis relies heavily on correctly identifying these roles before proceeding with data preparation and model fitting. Our example will illustrate how these variables interact to assess a specific research hypothesis.

For this tutorial, we will investigate whether different **studying techniques** have a significant impact on students' **exam scores**, while simultaneously controlling for the influence of the students' existing academic proficiency, measured by their **current grade** in the class.

Defining the Study Variables and Dataset Creation

To conduct our analysis, we must clearly define the variables involved and ensure the dataset is properly structured. The ANCOVA model requires one categorical independent variable, one continuous response variable, and at least one continuous covariate.

Here are the variables we will utilize in this demonstration:

Studying technique: This is our **categorical independent variable** (or treatment factor), representing the groups we are comparing (A, B, or C). We hypothesize that this variable drives differences in the outcome.

Student's current grade: This is the **covariate**, a continuous variable that we want to statistically control for. We include this covariate to adjust the exam scores based on prior student performance, thereby isolating the true effect of the studying technique.

Exam score: This is the **response variable** (or dependent variable), the continuous outcome we are interested in analyzing.

The simulated dataset used for this example contains information for 90 students who were randomly assigned to one of three study groups (30 students per technique). The data includes the technique used (A, B, or C), the student's baseline grade recorded before the intervention, and the final exam score received after one month of using the assigned technique. The following R code demonstrates how to create this reproducible dataset and view the initial records:

```
#make this example reproducible
```

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

```
#create dataset
```

```
data <- data.frame(technique = rep(c("A", "B", "C"), each = 30),
```

```
current_grade = runif(90, 65, 95),
```

```
exam = c(runif(30, 80, 95), runif(30, 70, 95), runif(30, 70, 90)))
```

```
#view first six lines of dataset
```

```
head(data)
```

```
# technique current_grade exam
```

```
#1 A 80.22435 87.32759
```

```
#2 A 74.20306 90.67114
```

```
#3 A 77.80723 88.87902
```

```
#4 A 85.79306 87.75735
```

```
#5 A 67.55408 85.72442
```

```
#6 A 71.76310 92.52167
```

Initial Data Exploration and Descriptive Statistics

Before fitting any complex model, a crucial first step involves exploring the data. This process allows us to gain a foundational understanding of the distributions, identify potential **outliers**, and confirm that the data conforms to expected patterns. We begin by examining a statistical summary of all variables in the dataset using the built-in R function:

```
summary(data)
```

```
# technique current_grade exam
```

```
# A:30 Min. :65.43 Min. :71.17
```

```
# B:30 1st Qu.:71.79 1st Qu.:77.27
```

```
# C:30 Median :77.84 Median :84.69
```

```
# Mean :78.15 Mean :83.38
```

```
# 3rd Qu.:83.65 3rd Qu.:89.22
```

```
# Max. :93.84 Max. :94.76
```

The summary confirms that the categorical variable, `technique`, is balanced, with 30 observations in each of the three levels (A, B, and C). The summaries for the continuous variables, `current_grade` and `exam`, show reasonable ranges and means, suggesting no immediate data entry errors or extreme values that might drastically skew the results. For instance, the mean current grade is 78.15, and the mean exam score is 83.38.

To further investigate the data structure, particularly how the covariate and response variables vary across the treatment groups, we utilize the **dplyr** package. This package simplifies the process of calculating group-wise descriptive statistics, such as the mean and standard deviation for both current grades and final exam scores, stratified by studying technique:

```
#load dplyr
library(dplyr)

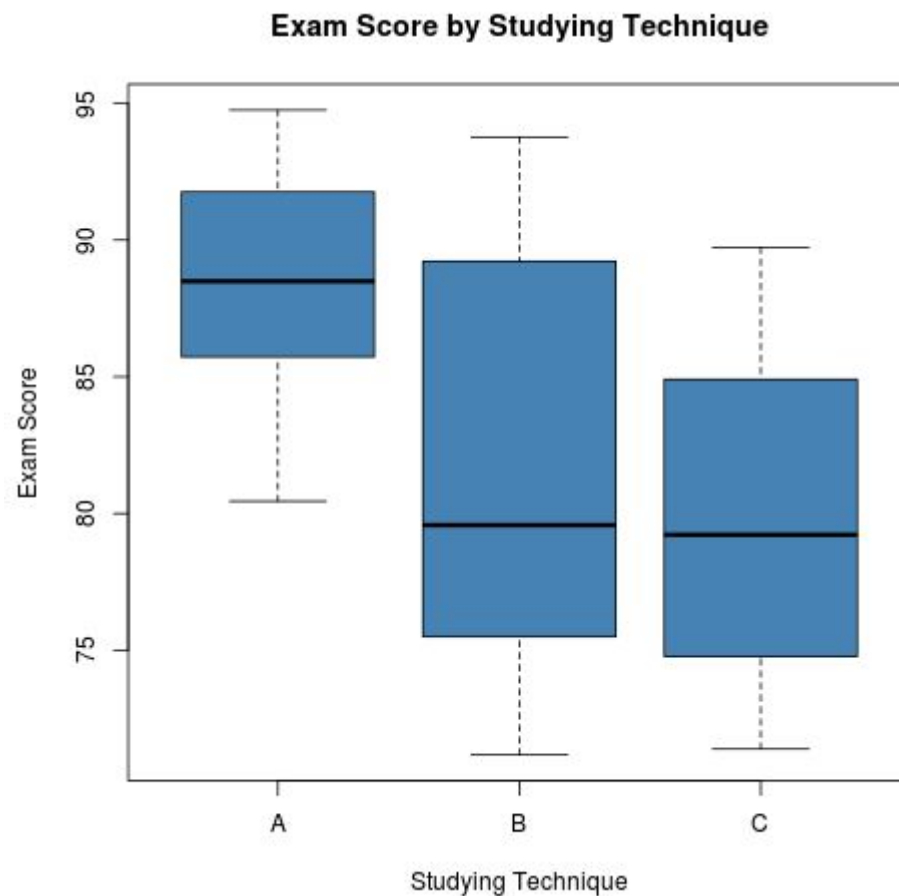
data %>%
  group_by(technique) %>%
  summarise(mean_grade = mean(current_grade),
            sd_grade = sd(current_grade),
            mean_exam = mean(exam),
            sd_exam = sd(exam))

# A tibble: 3 x 5
# technique mean_grade sd_grade mean_exam sd_exam
#1 A 79.0 7.00 88.5 3.88
#2 B 78.5 8.33 81.8 7.62
#3 C 76.9 8.24 79.9 5.71
```

This output reveals that the mean and standard deviation of the baseline covariate (`current_grade`) are quite similar across all three techniques (A, B, and C). This similarity is expected and desirable in an experimental design where subjects are randomly assigned to groups, minimizing initial systematic bias. Conversely, the mean exam scores (`mean_exam`) show some noticeable differences between the groups, particularly technique A compared to B and C, hinting at a potential treatment effect that the ANCOVA will formally test. Visualization through **boxplots** is also critical for assessing the distributional properties of the response variable across the treatment groups:

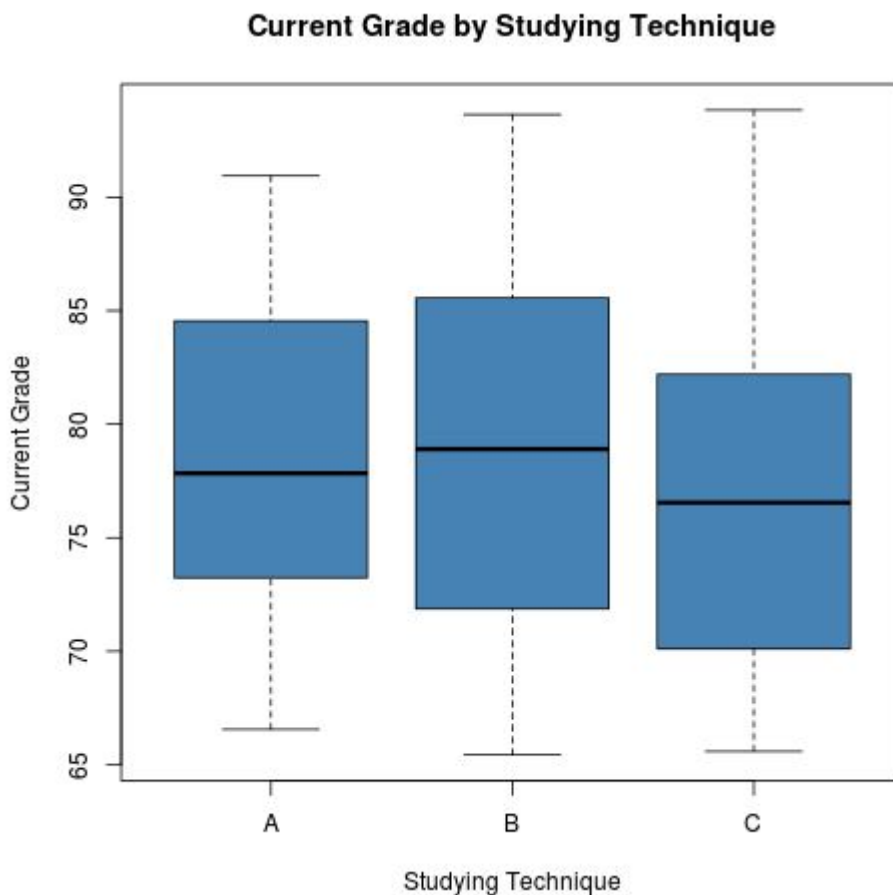
```
boxplot(exam ~ technique,
data = data,
main = "Exam Score by Studying Technique",
xlab = "Studying Technique",
```

```
ylab = "Exam Score",  
col = "steelblue",  
border = "black"  
)
```

We also generate boxplots for the covariate, `current_grade`, stratified by the studying technique. Visualizing the covariate helps confirm that the initial distribution of academic proficiency is similar across the different treatment arms, reinforcing the validity of the randomization process and the subsequent ANCOVA adjustment.

```
boxplot(current_grade ~ technique,  
data = data,  
main = "Current Grade by Studying Technique",  
xlab = "Studying Technique",  
ylab = "Current Grade",  
col = "steelblue",  
border = "black"  
)
```



Verifying Model Assumptions

Before proceeding to fit the [ANCOVA](#) model, we must verify that the critical statistical assumptions are met. Violations of these assumptions can lead to unreliable p-values and biased parameter estimates. Two key assumptions specific to ANCOVA must be assessed: the independence of the covariate and treatment, and the homogeneity of variances.

The first critical assumption is the **independence of the covariate and the treatment**. This means that the covariate (current grade) should not be affected by the treatment factor (studying technique). If the groups were randomly assigned, this assumption is usually met. To statistically verify this, we run a one-way [ANOVA](#) using the covariate as the response variable and the treatment factor as the predictor:

```
#fit anova model
```

```
anova_model <- aov(current_grade ~ technique, data = data)
```

```
#view summary of anova model
```

```
summary(anova_model)
```

```
# Df Sum Sq Mean Sq F value Pr(>F)
```

```
#technique 2 74 37.21 0.599 0.552  
#Residuals 87 5406 62.14
```

The resulting p-value (0.552) is substantially greater than the conventional significance level of 0.05. This non-significant result confirms that there is no statistical difference in the mean current grades across the three studying techniques, suggesting that the covariate and the treatment are indeed **independent**, which is a strong prerequisite for the ANCOVA model.

The second essential assumption is the **homogeneity of variance**, which stipulates that the variance of the response variable (exam score) must be equal across all treatment groups, after adjusting for the covariate. We formally test this using **Levene's Test**, often implemented via the `car` package in R:

#load car library to conduct Levene's Test

```
library(car)
```

```
#conduct Levene's Test
```

```
leveneTest(exam~technique, data = data)
```

```
#Levene's Test for Homogeneity of Variance (center = median)
```

```
# Df F value Pr(>F)
```

```
#group 2 9.4324 0.0001961 ***
```

```
# 87
```

The p-value yielded by Levene's Test is 0.0001961. Since this value is less than 0.05, we must reject the null hypothesis of equal variances. This indicates a statistically significant violation of the **homogeneity of variance** assumption among the groups. While a severe violation often necessitates data transformation or the use of non-parametric methods, in practice, ANCOVA models are often considered relatively robust to minor violations, especially if the sample sizes are equal (as they are here, N=30 per group). For the purpose of this demonstration, we will proceed with the ANCOVA, acknowledging this limitation in the interpretation of the results.

Fitting the ANCOVA Model using Type III Sum of Squares

Having completed the exploratory data analysis and assumption checks, we proceed to fit the ANCOVA model. We define the model with `exam` as the response variable, `technique` as the primary factor of interest, and `current_grade` as the covariate we wish to control.

When fitting ANCOVA or complex ANOVA models, it is crucial to select the correct type of sum of squares. We employ **Type III Sum of Squares** because it tests the effect of each predictor after

accounting for all other predictors in the model, providing a robust test for balanced or unbalanced designs where interaction terms might be present. We use the `Anova()` function from the `car` package (note the capital 'A') to specify the required Type III calculation, thereby overcoming the default behavior of the base R `aov()` function (which uses Type I, where results depend on the order of terms entered).

#load car library

```
library(car)
```

```
#fit ANCOVA model
```

```
ancova_model <- aov(exam ~ technique + current_grade, data = data)
```

```
#view summary of model using Type III Sum of Squares
```

```
Anova(ancova_model, type="III")
```

```
#Response: exam
```

```
# Sum Sq Df F value Pr(>F)
```

```
 #(Intercept) 7161.2 1 201.4621 < 2.2e-16 ***
```

```
 #technique 1242.9 2 17.4830 4.255e-07 ***
```

```
 #current_grade 12.3 1 0.3467 0.5576
```

```
 #Residuals 3057.0 86
```

The output provides crucial insights into the adjusted effects of the variables. We focus on the rows corresponding to `technique` and `current_grade`. The row for `current_grade` shows a p-value of 0.5576, indicating that the covariate itself does **not** have a statistically significant relationship with the exam score in this specific model. However, the row for `technique` yields a p-value of 4.255e-07, which is extremely small (far below 0.05). This result leads us to reject the null hypothesis and conclude that **studying technique has a statistically significant effect on exam scores**, even after accounting for the students' initial academic performance (current grade).

Conducting and Interpreting Post Hoc Tests

The significant F-test for the `technique` factor tells us that at least one group mean is significantly different from the others, but it does not specify which pairs are different. To pinpoint the exact location of these differences, we must perform **Post Hoc Tests** (also known as multiple comparisons). These tests compare all possible pairs of means while controlling the **Family-Wise Error Rate** (FWER) to prevent inflated Type I errors.

We will use **Tukey's Honestly Significant Difference (HSD) test**, which is appropriate for pairwise comparisons following a significant ANCOVA result. This test is implemented using the `glht()` function (General Linear Hypothesis Test) within the `multcomp` package in R:

#load the *multcomp* library**library(multcomp)**

```
#fit the ANCOVA model (re-run for clarity, though it's the same model)
ancova_model <- aov(exam ~ technique + current_grade, data = data)
```

```
#define the post hoc comparisons to make (Tukey contrasts)
postHocs <- glht(ancova_model, linfct = mcp(technique = "Tukey"))
```

```
#view a summary of the post hoc comparisons
summary(postHocs)
```

```
#Multiple Comparisons of Means: Tukey Contrasts
```

#

```
#Fit: aov(formula = exam ~ technique + current_grade, data = data)
```

#

```
#Linear Hypotheses:
```

```
# Estimate Std. Error t value Pr(>|t|)
```

```
#B - A == 0 -6.711 1.540 -4.358 0.000109 ***
```

```
#C - A == 0 -8.736 1.549 -5.640 < 1e-04 ***
```

```
#C - B == 0 -2.025 1.545 -1.311 0.393089
```

```
#view the confidence intervals associated with the multiple comparisons
confint(postHocs)
```

```
# Simultaneous Confidence Intervals
```

#

```
#Multiple Comparisons of Means: Tukey Contrasts
```

#

```
#Fit: aov(formula = exam ~ technique + current_grade, data = data)
```

#

```
#Quantile = 2.3845
```

```
#95% family-wise confidence level
```

#

```
#Linear Hypotheses:
```

```
# Estimate lwr upr
```

```
#B - A == 0 -6.7112 -10.3832 -3.0392
```

```
#C - A == 0 -8.7364 -12.4302 -5.0426
```

```
#C - B == 0 -2.0252 -5.7091 1.6588
```

The summary of the **Tukey Contrasts** provides the adjusted mean differences and corresponding

p-values for all three pairwise comparisons. We observe the following:

B - A == 0: The difference between Technique B and Technique A is -6.711, with a highly significant p-value of 0.000109. This indicates that Technique A led to significantly higher exam scores than Technique B. The 95% confidence interval (-10.38 to -3.04) confirms this difference, as it does not contain zero.

C - A == 0: The difference between Technique C and Technique A is -8.736, also highly significant ($p < 1e-04$). Technique A resulted in significantly higher exam scores than Technique C. The confidence interval (-12.43 to -5.04) confirms this finding.

C - B == 0: The difference between Technique C and Technique B is -2.025, but the p-value is 0.393089. Since this value is much greater than 0.05, there is **no statistically significant difference** in exam scores between Technique B and Technique C. The corresponding 95% confidence interval (-5.71 to 1.66) contains zero, reinforcing the lack of a significant difference between these two methods.

In conclusion, the ANCOVA and subsequent [post hoc tests](#) demonstrate that **Studying Technique A** is statistically superior, leading to significantly greater adjusted exam scores compared to Techniques B and C. This finding remains robust even after controlling for the baseline academic proficiency measured by the students' current grades.