We use Regression analysis to create models which describe the effect of variation in predictor variables on the response variable. Sometimes, if we have a categorical variable with values like Yes/No or Male/Female etc. The simple regression analysis gives multiple results for each value of the categorical variable. In such scenario, we can study the effect of the categorical variable by using it along with the predictor variable and comparing the regression lines for each level of the categorical variable. Such an analysis is termed as **Analysis of Covariance** also called as **ANCOVA**.

Example

Consider the R built in data set mtcars. In it we observer that the field "am" represents the type of transmission (auto or manual). It is a categorical variable with values 0 and 1. The miles per gallon value(mpg) of a car can also depend on it besides the value of horse power("hp").

We study the effect of the value of "am" on the regression between "mpg" and "hp". It is done by using the **aov()** function followed by the **anova()** function to compare the multiple regressions.

Input Data

Create a data frame containing the fields "mpg", "hp" and "am" from the data set mtcars. Here we take "mpg" as the response variable, "hp" as the predictor variable and "am" as the categorical variable.

```
input <- mtcars[,c("am","mpg","hp")]
print(head(input))
```

When we execute the above code, it produces the following result –

	am	mpg	hp
Mazda RX4	1	21.0	110
Mazda RX4 Wag	1	21.0	110
Datsun 710	1	22.8	93
Hornet 4 Drive	0	21.4	110
Hornet Sportabout	0	18.7	175
Valiant	0	18.1	105

ANCOVA Analysis

We create a regression model taking "hp" as the predictor variable and "mpg" as the response variable taking into account the interaction between "am" and "hp".

Model with interaction between categorical variable and predictor variable

```
# Get the dataset.
input <- mtcars

# Create the regression model.
result <- aov(mpg~hp*am,data = input)
print(summary(result))
```

When we execute the above code, it produces the following result –

```
            Df Sum Sq Mean Sq F value   Pr(>F)
hp           1  678.4   678.4  77.391 1.50e-09 ***
am           1  202.2   202.2  23.072 4.75e-05 ***
hp:am        1    0.0     0.0   0.001    0.981
Residuals   28  245.4     8.8
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
' 1
```

This result shows that both horse power and transmission type has significant effect on miles per gallon as the p value in both cases is less than 0.05. But the interaction between these two variables is not significant as the p-value is more than 0.05.

Model without interaction between categorical variable and predictor variable

```
# Get the dataset.
input <- mtcars

# Create the regression model.
result <- aov(mpg~hp+am,data = input)
print(summary(result))
```

When we execute the above code, it produces the following result –

```
            Df  Sum Sq  Mean Sq   F value   Pr(>F)
hp           1   678.4    678.4    80.15  7.63e-10 ***
am           1   202.2    202.2    23.89  3.46e-05 ***
Residuals   29   245.4      8.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
' 1
```

This result shows that both horse power and transmission type has significant effect on miles per gallon as the p value in both cases is

less than 0.05.

Comparing Two Models

Now we can compare the two models to conclude if the interaction of the variables is truly in-significant. For this we use the **anova()** function.

```
# Get the dataset.
input <- mtcars

# Create the regression models.
result1 <- aov(mpg~hp*am,data = input)
result2 <- aov(mpg~hp+am,data = input)

# Compare the two models.
print(anova(result1,result2))
```

When we execute the above code, it produces the following result –

```
Model 1: mpg ~ hp * am
Model 2: mpg ~ hp + am
  Res.Df    RSS Df  Sum of Sq      F Pr(>F)
1     28 245.43
2     29 245.44 -1 -0.0052515 6e-04 0.9806
```

As the p-value is greater than 0.05 we conclude that the interaction between horse power and transmission type is not significant. So the mileage per gallon will depend in a similar manner on the horse power of the car in both auto and manual transmission mode.