

When we run an ANOVA, we analyze the differences among group means in a sample. In its simplest form, ANOVA provides a statistical test of whether two or more population means are equal, and therefore generalizes the t-test beyond two means.

ANOVA Null and Alternative Hypothesis

The **null hypothesis** in **ANOVA** is that there is no difference between means and the **alternative** is that the means are not all equal.

$$(H_0: \mu_1 = \mu_2 = \dots = \mu_K)$$
$$(H_1: \text{The } \mu_s \text{ Are Not All Equal})$$

This means that when we are dealing with many groups, we cannot compare them pairwise. We can simply answer if the means between groups can be considered as equal or not.

Tukey's HSD

What about if we want to compare all the groups pairwise? In this case, we can apply the [Tukey's HSD](#) which is a single-step multiple comparison procedure and statistical test. It can be used to find means that are significantly different from each other.

Example of ANOVA vs Tukey's HSD

Let's assume that we are dealing with the following 4 groups:

- Group "a": 100 observations from the Normal Distribution with **mean 10** and **standard deviation 5**
- Group "b": 100 observations from the Normal Distribution with **mean 10** and **standard deviation 5**
- Group "c": 100 observations from the Normal Distribution with **mean 11** and **standard deviation 6**
- Group "d": 100 observations from the Normal Distribution with **mean 11** and **standard deviation 6**

Clearly, we were expecting the ANOVA to reject the Null Hypothesis but we would also know that the **Group a and Group b are not statistically different** and the same with the **Group c and Group d**

Let's work in R:

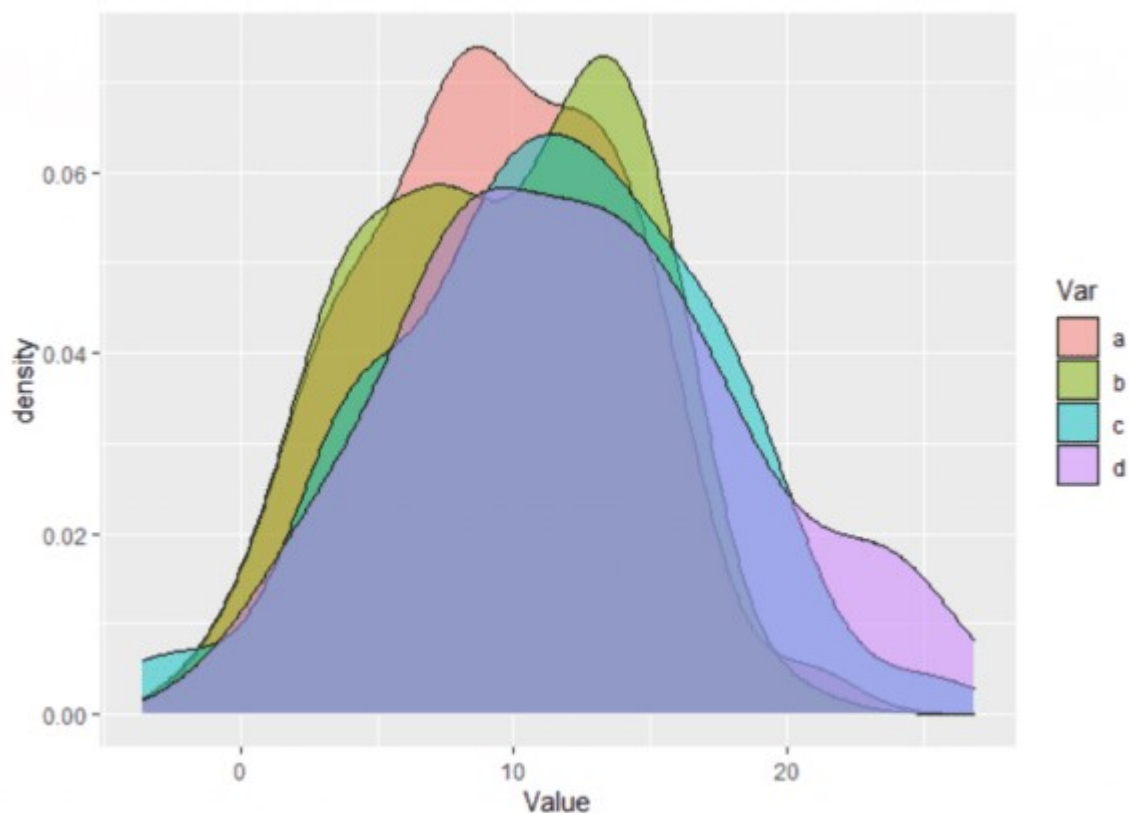
```
library(multcomp)
library(tidyverse)

# Create the four groups
set.seed(10)
df1 <- data.frame(Var="a", Value=rnorm(100,10,5))
df2 <- data.frame(Var="b", Value=rnorm(100,10,5))
df3 <- data.frame(Var="c", Value=rnorm(100,11,6))
df4 <- data.frame(Var="d", Value=rnorm(100,11,6))

# merge them in one data frame
df<-rbind(df1,df2,df3,df4)

# convert Var to a factor
df$Var<-as.factor(df$Var)
```

```
df%>%ggplot(aes(x=Value, fill=Var))+geom_density(alpha=0.5)
```



ANOVA

```
# ANOVA
modell<-lm(Value~Var, data=df)
anova(modell)
```

Output:

Analysis of Variance Table

Response: Value

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Var	3	565.7	188.565	6.351	0.0003257 ***
Residuals	396	11757.5	29.691		

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

Clearly, we reject the null hypothesis since the p-value is **0.0003257**

Tukey's HSD

Let's apply the Tukey HSD test to test all the means.

```
# Tukey multiple comparisons
summary(glht(modell, mcp(Var="Tukey")))
```

Output:

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

```
Fit: lm(formula = Value ~ Var, data = df)
```

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t)
b - a == 0	0.2079	0.7706	0.270	0.99312
c - a == 0	1.8553	0.7706	2.408	0.07727 .
d - a == 0	2.8758	0.7706	3.732	0.00129 **
c - b == 0	1.6473	0.7706	2.138	0.14298
d - b == 0	2.6678	0.7706	3.462	0.00329 **
d - c == 0	1.0205	0.7706	1.324	0.54795

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

As we can see from the output above, the difference between **c vs a** and **c vs b** found not be statistically significant although they are from different distributions. The reason for that is the “issue” with the multiple comparisons. Let's compare them by applying the t-test

t-test a vs c

```
t.test(df%>%filter(Var=="a")%>%pull(), df%>%filter(Var=="c")%>%pull())
```

Output:

Welch Two Sample t-test

```
data: df %>% filter(Var == "a") %>% pull() and df %>% filter(Var == "c") %>% pull()
```

```
t = -2.4743, df = 189.47, p-value = 0.01423
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
-3.3343125 -0.3761991
```

```
sample estimates:
```

```
mean of x mean of y
```

```
9.317255 11.172511
```

t-test b vs c

```
t.test(df%>%filter(Var=="b")%>%pull(), df%>%filter(Var=="c")%>%pull())
```

Output:

Welch Two Sample t-test

```
data: df %>% filter(Var == "b") %>% pull() and df %>% filter(Var == "c") %>% pull()
```

```
pull()
t = -2.1711, df = 191.53, p-value = 0.03115
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.1439117 -0.1507362
sample estimates:
mean of x mean of y
 9.525187 11.172511
```

As we can see from above, the means of the two groups, in both cases, found to be statistically significant, if we ignore the multiple comparisons.

Discussion

When we are dealing with multiple comparisons and we want to apply pairwise comparisons, then Tukey' HSD is a good option. Another approach is to consider the P-Value Adjustments.