ANOVA-One Way Analysis of Variance

One-Way Analysis of Variance (ANOVA) is an extension of hypothesis testing for two population means using the t-distribution. The ANOVA allows us to compare more than two populations' means if the following two conditions are satisfied.

1. The populations are normally distributed.
2. The populations' variances are all equal $\left(\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = ... = \sigma_n^2\right)$.

When you do your project on ANOVA you will have to use Minitab to check condition 1 and the rule of thumb for the equality of variance to check condition 2.

The hypothesis testing for the equality of the means is given below.

Step 1. $H_0 : \mu_1 = \mu_2 = \mu_3 = ... = \mu_n$
$H_a : At\ least\ one\ mean\ is\ different$

Step 2. Test statistic: F (Note: This value comes from the ANOVA table)

Step 3. Reject Ho if $F > F_{v_1, v_2; \alpha}$.
OR If the p-value < α . (Note: we are going to use the p-value)

Step 4. Conclusion.

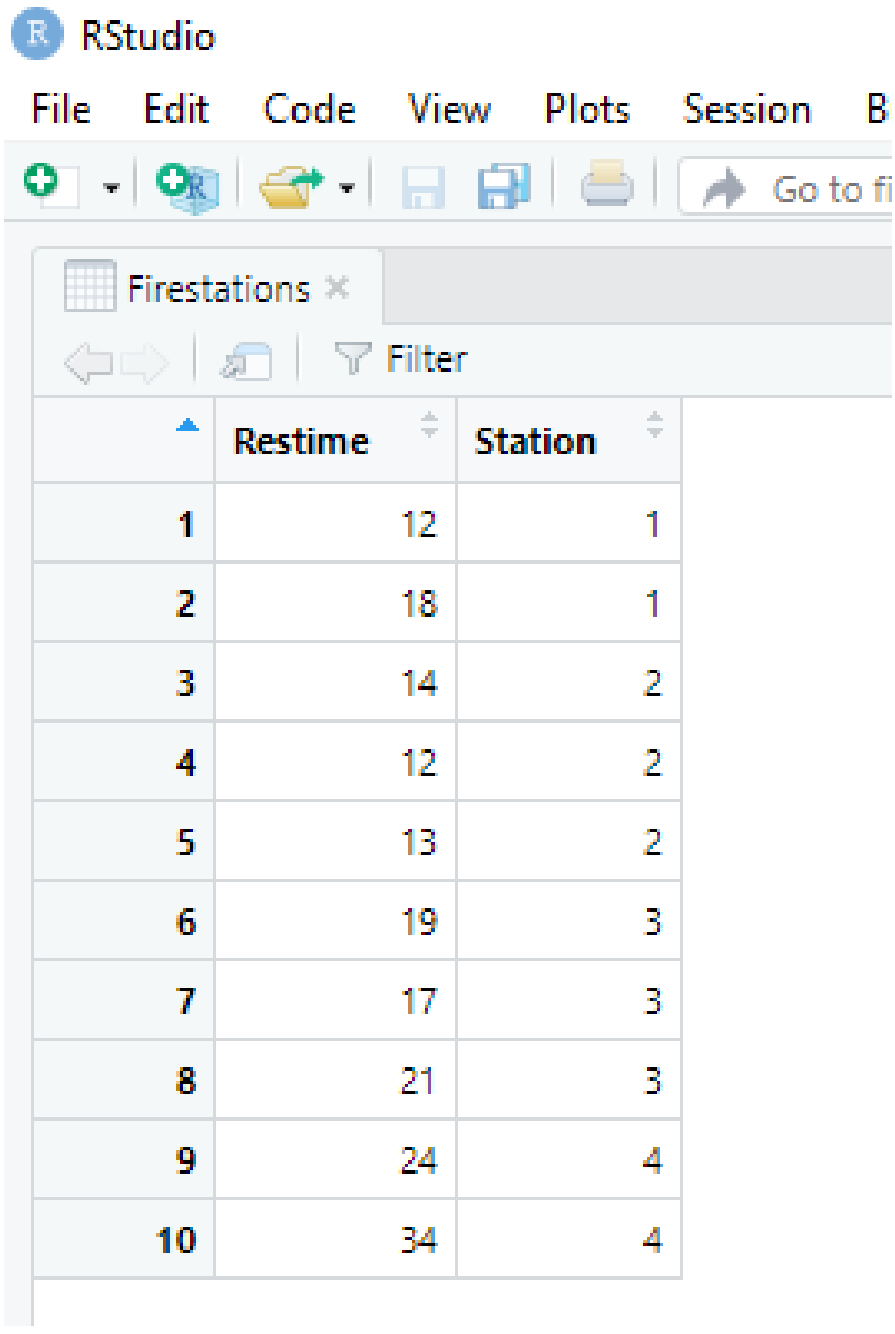
Note: If we fail to reject H_o we conclude that all the means are equal.
If we reject H_o we have to find out which mean or means are different. We have to do a pair-wise comparison of the means. This work will be done in R-Studio.

Example: We would like to compare the average time that it takes a fire station to respond (the time it takes the fire truck to leave the station) to a phone call. There are four different fire stations in town. The following data was provided over a week. Is there a significant difference in the average response time of these four fire stations? Test it at $\alpha = 0.05$.

Station 1	Station 2	Station 3	Station 4
12 min	14 min	19 min	24 min
18 min	12 min	17 min	34 min
	13 min	21 min	

Use R-Studio to do the Hypothesis Testing. The R-Studio commands are provided below.

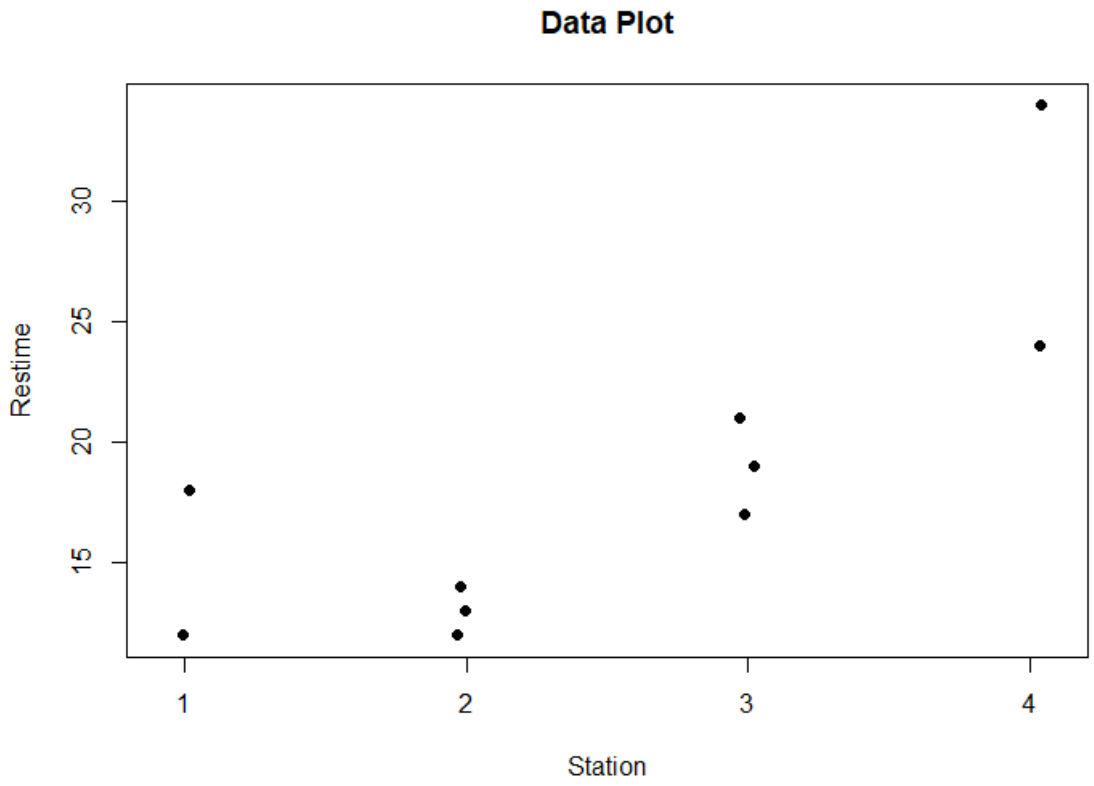
```
Read the Data File Or Import
RData <- read_excel("Regression/Firestations.xlsx")
```



	Restime	Station
1	12	1
2	18	1
3	14	2
4	12	2
5	13	2
6	19	3
7	17	3
8	21	3
9	24	4
10	34	4

This command plots the data.

```
stripchart(Restime~as.factor(Station), main = "Data Plot", vertical = T, pch=19, data=Firestations, xlab="Station", ylab="Restime", method="jitter", jitter=0.04)
```

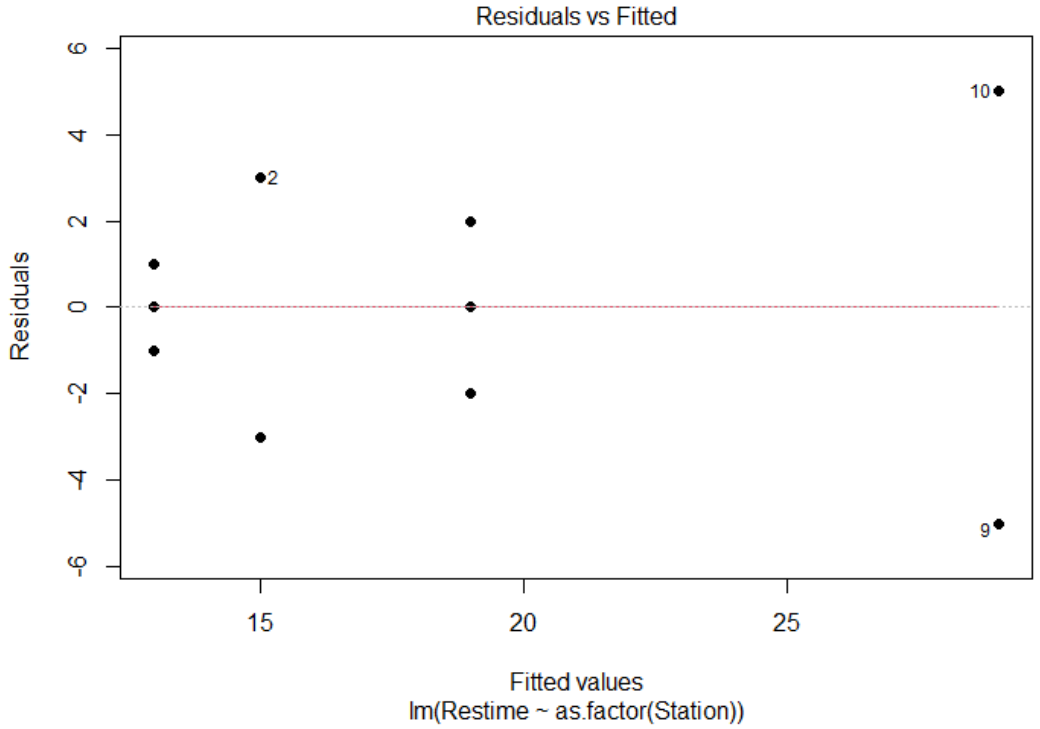



```
analysis <- lm(Restime~as.factor(Station), data=Firestations)
anova(analysis)
```

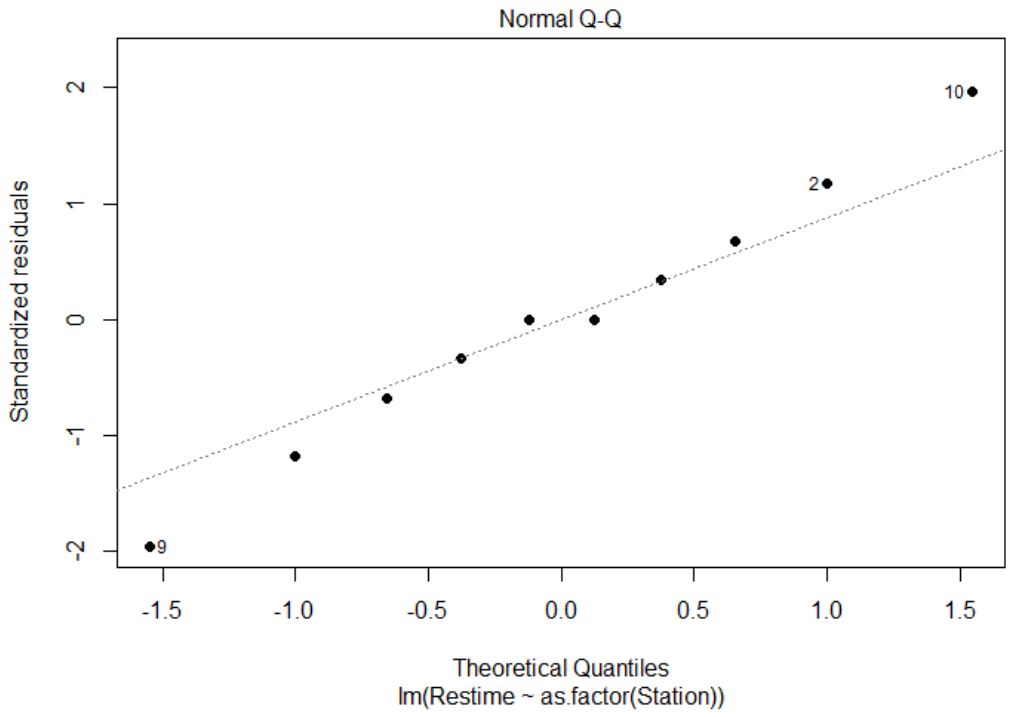
```
> analysis <- lm(Restime~as.factor(Station), data=Firestations)
> anova(analysis)
Analysis of Variance Table

Response: Restime
                   Df Sum Sq Mean Sq F value  Pr(>F)
as.factor(Station)  3  336.4  112.13  8.6256 0.01352 *
Residuals           6   78.0   13.00
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

plot(analysis, which=1, pch=19)



plot(analysis, which=2, pch-19)



```
TukeyHSD(aov(analysis))
```

```
TukeyHSD(aov(analysis))
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = analysis)

$`as.factor(Station)`
      diff       lwr       upr     p adj
2-1     -2 -13.393887  9.393887 0.9259958
3-1      4 -7.393887  15.393887 0.6404247
4-1     14  1.518622  26.481378 0.0310585
3-2      6 -4.191002  16.191002 0.2726170
4-2     16  4.606113  27.393887 0.0111609
4-3     10 -1.393887  21.393887 0.0822023
```

Homework: 12.63, and 12.65 pages 693-694.

Note: Use R-Studio to do Hypothesis testing and draw a conclusion only.