One Way ANOVA Test

T-test only work for one group and two groups of samples.
If we have more than two groups, it seems that we can run multiple t tests to compare the means from different groups and if we found that at least one of the t tests was significant, we can say there is group effect. But the multiple t tests inflate Type1 error for the test of overall group effect (overall α).

If tests are independent, we can calculate exact overall α as:
P(at least one significant result) = $1 - p$(no significant results) $= 1 - (1 - \alpha_1)(1 - \alpha_2)....(1 - \alpha_k)$

For example: 3 groups involve three t-tests. If the tests are independent and α = 0.5 is used for each test, then overall alpha is 1-0.95*0.95*0.95 = 0.143
But tests are often dependent and it is not easy to calculate the exact overall α. Therefore if we have more than two groups, we will go for ANOVA test.

ANOVA Test

Anova compare multiple groups simultaneously which helps to avoid inflated overall alpha. The basic logic of ANOVA is to decompose the total variance of dependent variables to those due to the group effect (between-group variance) and those due to sampling variability (within-group variance).
ANOVA uses a F test for statistical inferences.

$$F = \frac{Between\ group\ variance}{Within\ group\ variance}$$

We use variance because we cannot compute one mean difference with more than two groups. Variance is the average standard error, which can be compared across different sample sizes.

ANOVA table:

Source	SS	df	MS
Between	$SS_{between}$	k-1	$MS_{between}$
Within	SS_{within}	N-k	MS_{within}
Total	SS_{total}	N-1	

k is number of conditions of a factor(levels)

N is total number of samples.

One- Way ANOVA

ANOVA with one categorical independent variable is called One-way ANOVA.

$$F = \frac{SS_{between}/df_{between}}{SS_{within}/df_{within}} = \frac{MS_{between}}{MS_{within}}$$

MS is mean square which is another term for variance. Here F test will have two degrees of freedom.

If null hypothesis is true, then only random variability contributes to any observed differences and therefore ,between group variability and within group variability will be same as both arise from error variance.

Therefore, $F = \frac{MS_{between}}{MS_{within}} = 1$

If null hypothesis is false, $F = \frac{MS_{between}}{MS_{within}} > 1$

Example 1:

Consider, we randomly select students and assign them into three groups with different stress level and observed marks obtained by them. We have to find if stress level affects the scores.

Group1(high stress)	2	3	7	2	6
Group2(moderate stress)	10	8	7	5	10
Group3(low stress)	10	13	14	13	15

Step1: Specify null and alternate hypothesis

Null hypothesis(H0) - Stress level does not affect the scores obtained by the students.

Alternate hypothesis(H1) - Stress level does affect the scores obtained by the students.

Step2: Find total SS and decompose it.

Group1(High stress) X1	Group2(Moderate stress) X2	Group3(Low stress) X3	$(X1 - \overline{X1})^2$	$(X2 - \overline{X2})^2$	$(X3 - \overline{X3})^2$
2	10	10	4	4	9
3	8	13	1	0	0
7	7	14	9	1	1
2	5	13	4	9	0
6	10	15	4	4	4
$\overline{X1}$ = 4	$\overline{X2}$ = 8	$\overline{X3}$ = 13	SS$_1$ = 22	SS$_2$ = 18	SS$_3$ = 14

$\overline{X}_G$ = 8.33 , overall mean

Now, the deviation score SS can be partitioned into:
Deviation of each score from the overall mean = deviation of score from its own group mean + deviation of that group mean from the overall mean.

SS$_{total}$ = SS$_{within}$ + SS$_{between}$

1) SS$_{total}$ (Total variability in the sample)

$$SS_{total} = \Sigma(X - X_G)^2$$
For example: for first score from X1,
X- X$_G$ = 2-8.33 = -6.33, Similary find for all scores and add them .

SS$_{total}$ = 257.33

2) SS$_{within}$ (Individual variability that is due to random error)
$$SS_{within} = SS_1 + SS_2 + SS_3 = \Sigma(X - \overline{X_1})^2 + \Sigma(X - \overline{X_2})^2 + \Sigma(X - \overline{X_3})^2$$

Therefore, for example 1,
SS_{within} = 22+18+14 = 54

3) SS$_{between}$ (individual variability that is affected by the group effect)

$$SS_{between} = n_1(\overline{X_1} - \overline{X_G})^2 + n_2(\overline{X_2} - \overline{X_G})^2 + n_3(\overline{X_3} - \overline{X_G})^2$$
$$SS_{between} = 5 \times (4 - 8.33)^2 + 5 \times (8 - 8.33)^2 + 5 \times (13 - 8.33)^2 = 203.33$$

Therefore,

Source	SS	df	MS
Between	$SS_{between}$ = 203.33	k-1 =2	$MS_{between}$ = 101.67
Within	SS_{within} = 54	N-k =12	MS_{within} = 4.5
Total	SS_{total} = 257.33	N-1 = 14	

k is number of factors,
N is number of samples

Therefore, F(2,12) = 101.67 / 4.5 = 22.59

R code:
Y<- c(2,3,7,2,6,10,8,7,5,10,10,13,14,13,15)

group <- c(rep("high stress",5), rep("moderate stress",5), rep("low stress",5))
data <- data.frame(group,Y)
data$group<- as.factor(group)
model1 <- lm(Y~group , data =data)
anova(model1)

```
Y<- c(2,3,7,2,6,10,8,7,5,10,10,13,14,13,15)
group <- c(rep("high stress",5),rep("moderate stress",5),
           rep("low stress",5))

data <- data.frame(group,Y)
data
data$group<- as.factor(group)

model1 <- lm(Y~group , data =data)
anova(model1)
```

```
>
>
> model1 <- lm(Y~group , data =data)
> anova(model1)
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
group      2 203.33  101.67  22.593 8.539e-05 ***
Residuals 12  54.00    4.50
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Now, to see if there is significant difference between the groups, we have to compare F value with critical value or we can use p value. Here, p value is 8.539e-05 which is very less than 0.5, therefore we can say that there is significant difference between at least two groups.

If p value is greater than 0.5, we fail to reject H0 but if p value is less than 0.5, we can reject null hypothesis and we can say that there is at least one group which has significant difference.
But to see which group has that difference, we have to perform some additional test known as post-hoc tests.

Some Post-hoc tests:
1) Bonferroni Adjustment
This is the simplest and most popular adjustment method. It is based on Bonferroni inequality according to which the familywise Type1 error rate from doing multiple tests will always be less than multiplying α for each test by the number of tests(k).

Overall $\alpha \leq k.\alpha_t \rightarrow Overall \ ^{\alpha}/_k \leq \alpha_t$

So, if we use the desired familywise Type1 error rate divided by k as the alpha foe each t test, then the actual "familywise" Type1 error rate will be less than equal

to the desired level. Or we can multiply the p values by k and compare the adjusted p value to the desired level of the desired "familywise" alpha.

Imp points:
Bonferroni protect us from making at least one type1 error, we sacrifice a great deal of statistical power for each individual test.
Bonferroni adjustments are most common with relatively few tests.

Code:

```
pairwise.t.test(data$Y, data$group, p.adj = "bonferroni",  pool.sd = FALSE)
```

```
pairwise.t.test(data$Y, data$group, p.adj = "bonferroni",
                pool.sd = FALSE)
```

```
>
> pairwise.t.test(data$Y, data$group, p.adj = "bonferroni",
+                 pool.sd = FALSE)

        Pairwise comparisons using t tests with non-pooled SD

data:  data$Y and data$group

                high stress low stress
low stress      0.00057      -
moderate stress 0.06728      0.01305

P value adjustment method: bonferroni
>
```

2) The Holm Method (Holm-Bonferroni Method)
This method is developed to mediate the problem with the classical Bonferroni method. It also controls the maximum familywise error rate at alpha but has higher power than the classical Bonferroni method.
It sorts the p-values from lowest to highest and multiply them by k,k-1,……1 and compare them to alpha.

Code:

```
pairwise.t.test(data$Y, data$group, p.adj = "holm", pool.sd = FALSE)
```

```
pairwise.t.test(data$Y, data$group, p.adj = "holm",
                pool.sd = FALSE)
```

```
>
> pairwise.t.test(data$Y, data$group, p.adj = "holm",
+                 pool.sd = FALSE)

        Pairwise comparisons using t tests with non-pooled SD

data:  data$Y and data$group

                high stress low stress
low stress      0.00057     -
moderate stress 0.02243     0.00870

P value adjustment method: holm
>
```

3) Sidak-Holm Method

This method is extension of Holm method.

It sorts the p-values from lowest to highest and multiply them by

$$1 - (1 - p)^m , \quad 1 - (1 - p)^{m-1}, \quad .... , \quad 1 - (1 - p)^1$$

This method is slightly more powerful than the Holm adjusted values.

Sidak-Holm method is not available in pairwise t -tests. We have to use lsmeans for this method.

Code:

lsmeans(model1, pairwise~ group, adjust="sidak")

```
lsmeans(model1, pairwise~ group, adjust="sidak")
```

```
>
> lsmeans(model1, pairwise~ group, adjust="sidak")
$lsmeans
 group            lsmean    SE df lower.CL upper.CL
 high stress       4 0.949 12    1.93      6.07
 low stress       13 0.949 12   10.93     15.07
 moderate stress   8 0.949 12    5.93     10.07

Confidence level used: 0.95

$contrasts
 contrast                       estimate   SE df t.ratio p.value
 high stress - low stress            -9 1.34 12 -6.708  0.0001
 high stress - moderate stress       -4 1.34 12 -2.981  0.0340
 low stress - moderate stress         5 1.34 12  3.727  0.0086

P value adjustment: sidak method for 3 tests

> |
```

4) **Fisher's Least Significant Difference (Fishers's LSD)**

It is similar to pairwise t tests without any adjustments. The only modification is to replace the pooled variance from only two group with the pooled variance from all groups.

$$t = \frac{\overline{X_i} - \overline{X_j}}{\sqrt{\frac{MS_{within}}{n_i} + \frac{MS_{within}}{n_j}}}$$ df for the test = df for the MS_{within}

Imp points:

It is powerful because it does not adjust p or alpha value.

This method can only protect against inflated type1 error rate when there are 3 or 4 groups. Thus, it is not recommended when there are more than 4 groups.

Code:

pairwise.t.test(data$Y, data$group, p.adj = "none", pool.sd = TRUE)

```
pairwise.t.test(data$Y, data$group, p.adj = "none",
                pool.sd = TRUE)
```

```
>
> pairwise.t.test(data$Y, data$group, p.adj = "none",
+                 pool.sd = TRUE)

        Pairwise comparisons using t tests with pooled SD

data:  data$Y and data$group

                high stress low stress
low stress      2.2e-05     -
moderate stress 0.0115      0.0029

P value adjustment method: none
>
```

5) **Tukey's Honestly Significant Difference (Tukey's HSD)**

Tukey determined the sampling distribution of the largest t-test value when tere were no difference among means (i.e. when H0 is true)
To be statistically significant according to this method, a test must exceed that largest value. By doing so, it ensure that we will have only a 5% chance to find a significant difference when the null hypothesis is true.

Imp points:
It is specifically created for use with doing all of the G(G-1)/2 pairwise comparisons.
This method is not really appropriate for testing more complex contrasts (e.g, compare group 1 to the mean of groups 2 and 3)

Code:

lsmeans(model1, pairwise~group, adjust="tukey")

```
lsmeans(model1, pairwise~group, adjust="tukey")
```

```
> lsmeans(model1, pairwise~group, adjust="tukey")
$lsmeans
 group          lsmean    SE df lower.CL upper.CL
 high stress         4 0.949 12     1.93     6.07
 low stress         13 0.949 12    10.93    15.07
 moderate stress     8 0.949 12     5.93    10.07

Confidence level used: 0.95

$contrasts
 contrast                       estimate   SE df t.ratio p.value
 high stress - low stress             -9 1.34 12 -6.708  0.0001
 high stress - moderate stress        -4 1.34 12 -2.981  0.0287
 low stress - moderate stress          5 1.34 12  3.727  0.0075

P value adjustment: tukey method for comparing a family of 3 estimates

> |
```

* If we have more than one factor (independent variable), we have to look for another method which is known as Factorial ANOVA.

Assumptions of ANOVA:

From the equations of ANOVA, we can say that an ANOVA model can be represented using a linear equation. Therefore, all the assumptions for general linear model will apply.
Three main assumptions for ANOVA are:
1) Errors are independent.
2) Errors are normal.
3) Errors have homogenous variance across all group.

For R code:

Visit Github