Factorial ANOVA

A factorial Anova is any Anova that involves more than one factor (independent variable).

For example, in one way anova we were taking independent variable scores for different groups. But if we have to measure scores for those groups at different situations (i.e. more than one factor), we will use factorial Anova.

In factorial design, we include all combinations of the levels of independent variables. These combinations are called cells.

Pressure level	**Curriculum A**	**Curriculum B**	**Curriculum C**
Low pressure			
High pressure			

This will be our 3X2 factorial design. It has 6 cells and in factorial design, no cell should be empty.

To understand the relationship between independent and dependent variables, we look for different types of effects which are known as Simple effect, Main effect, and Interaction effect.

Simple effects:

This effect concerns about the effect of one independent variable(IV) at a specific level of the other IV.
The effect of curriculum at a specific level of pressure is simple effect. Similarly, effect of pressure at a specific curriculum is also a simple effect.

Pressure level	**Curriculum A**	**Curriculum B**	**Curriculum C**
Low pressure	X	X	X
High pressure	X	X	X

For example: In above table, If the blue highlighted means are different, then there is a simple effect of curriculum for low pressure.

Pressure level	Curriculum A	Curriculum B	Curriculum C
Low pressure	X	X	X
High pressure	X	X	X

Similarly, if the red highlighted means are different, then there is simple effect of pressure on Curriculum A.

Main Effects:

It is the average effect of the factors across all the levels of other IV or average effect of the levels across across all the factors.

Pressure level	Curriculum A	Curriculum B	Curriculum C	
Low pressure	40	40	40	40
High pressure	20	20	20	20
	30	30	30	

Here highlighted cells are average means. For pressure(40,20) are different, therefore there is main effect for pressure. For Curriculum(30,30,30) are same, therefore there is no main effect for curriculum.

Interaction Effects:

This effect occurs when the simple effects of one IV are not constant across all levels of the other IV.

For Example:

Pressure level	Curriculum A	Curriculum B	Curriculum C	
Low pressure	50	30	10	30
High pressure	30	30	30	30
	40	30	20	

Here, there is no simple effect of curriculum for high pressure,
Simple effects of curriculum for low pressure is (50-30, 30-10) = (20,20) which is

constant.
Simple effect of pressure for Curriculum A , B and C are (50-30, 30-30, 10-30) which is not constant.
Therefore, there is an interaction effect.

Also, there will be main effect for Curriculum as average means for all three curriculums are different.

Ordinal and Disordinal Interaction

Ordinal: the effect of A is positive or negative for all values of B.
Disordinal: the effect of A is positive for some values of B and negative for the other values.
More details in mean plot R code.

Calculations:

To conduct two-way Anova, the same general principles from one way ANOVA will apply i.e. total variability is decomposed into stable and random sources.

Here, Sum of squares will contain a component for each effect.

$SS_{Total} = SS_A + SS_B + SS_{AxB} + SS_{within}$
SS_A is the between groups SS for factor A
SS_B is the between groups SS for factor B
SS_{AxB} is the between groups SS for group AxB (interaction effect)
SS_{within} is the random variability, sometimes referred to as SS_{error} or $SS_{residuals}$

Procedure for Two way Anova:
Step1: Calculate SS for each component
Step2: Calculate df for each component
Step3: Calculate MS,
MS = (SS/df) for each component.
Step: There will be F-tests for each component.
F = ratio of $MS_{component}$ to MS_{error}

	SS	df	MS = SS/df	F
A	SS_A	a-1	MS_A	MS_A / MS_{within}
B	SS_B	b-1	MS_B	MS_B / MS_{within}
AB	SS_{AxB}	(a-1)*(b-1)	MS_{AB}	MS_{AB} / MS_{within}
Within/ error	SS_{within}	df_{total} - df_A – df_B – df_{AB}	MS_{within}	
Total				

N is total number of obseravtions
a is number od levels in factor A
b is number of levels in factor B

Example: Consider we randomly assign some students into three curriculums under two conditions, some with low pressure and some with high pressure. Now we have to find effects between independent and dependent variables(scores, curriculum and pressure) and run Anova test to measure effects.

You can download dataset from:
https://github.com/palak-j/Hypothesis-testing-Anova/blob/main/factorial-anova.csv

Firstly, we will load the data and find the cell means
Code:

```
data <- read.csv("D:/r codes/factorial-anova.csv", header=TRUE)

data$Pressure_level = factor(data$Pressure_level)

data$Curriculum = factor(data$Curriculum)

cellmeans = aggregate(data$Scores, by= list(Pressure_level = data$Pressure_level, Curriculum= data$Curriculum), FUN=mean)

xtabs(x~ ., data =cellmeans)
```

```r
data <- read.csv("D:/github/r codes/factorial-anova.csv", header=TRUE)

data$Pressure_level = factor(data$Pressure_level)
data$Curriculum = factor(data$Curriculum)
cellmeans = aggregate(data$Scores,
                      by= list(Pressure_level = data$Pressure_level,
                      Curriculum= data$Curriculum), FUN=mean)
xtabs(x~ ., data =cellmeans)
```

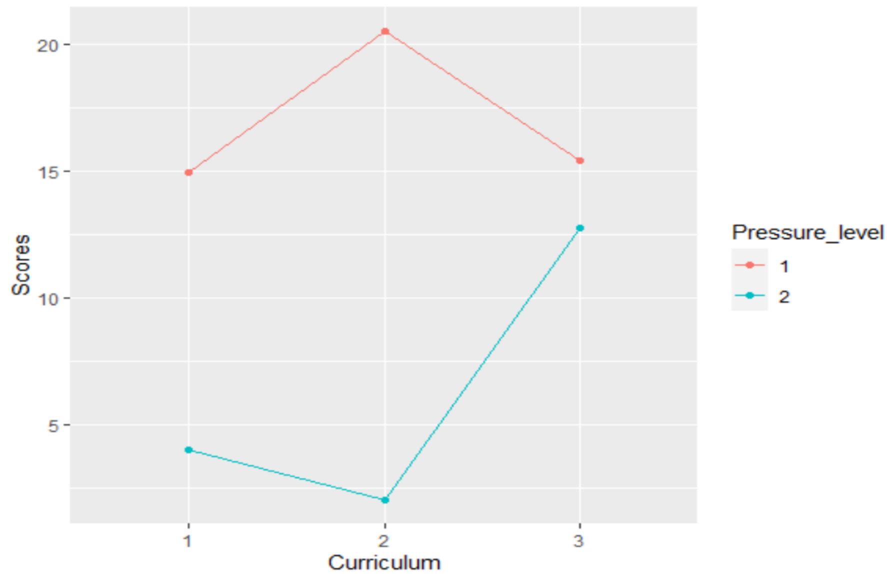
```
> data <- read.csv("D:/github/r codes/factorial-anova.csv", head
er=TRUE)
> data$Pressure_level = factor(data$Pressure_level)
> data$Curriculum = factor(data$Curriculum)
> cellmeans = aggregate(data$Scores,
+                       by= list(Pressure_level = data$Pressure_
level,
+                       Curriculum= data$Curriculum), FUN=mean)
> xtabs(x~ ., data =cellmeans)
              Curriculum
Pressure_level        1        2        3
             1 14.92308 20.53846 15.38462
             2  4.00000  2.00000 12.76923
>
```

Here, we can see means for all combinations of pressure and curriculum(consider pressure level 1 as low and 2 as high).

Now, we will draw mean plot to understand the various effects(simple, main and interaction effects)

Code:

```r
ggplot(data, aes(x = Curriculum, y = Scores, colour = Pressure_level))+
  stat_summary(fun = mean, geom = "point", aes(group = Pressure_level))+
  stat_summary(fun = mean, geom = "line", aes(group=Pressure_level))
```



After analyzing the mean plot and cell means, we can say that there is simple effect, main effect as well as interaction effect.

Now, we will run the Anova test.

```
# Run the Anova test
model <- lm(Scores ~ Pressure_level*Curriculum, data=data)

anova(model)
```

```
> # Run the Anova test
> model <- lm(Scores ~ Pressure_level*Curriculum, data=data)
> anova(model)
Analysis of Variance Table

Response: Scores
                          Df  Sum Sq Mean Sq  F value
Pressure_level             1 2229.35 2229.35 114.5262
Curriculum                 2  281.26  140.63   7.2244
Pressure_level:Curriculum  2  824.54  412.27  21.1791
Residuals                 72 1401.54   19.47
                             Pr(>F)
Pressure_level            < 2.2e-16 ***
Curriculum                 0.001382 **
Pressure_level:Curriculum 5.839e-08 ***
Residuals
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> 
```

From the results, we can see that p value for the model is significantly less than 0.05, therefore we can say there is significant difference in at least one pair of groups.

To understand which pair of groups has mean difference, we will use the appropriate Post-Hoc test. Here, we are using Tukey's HSD.

Code:

lsmeans(model, pairwise ~ Pressure_level, by = "Curriculum", adjust = "tukey")

```r
lsmeans(model, pairwise ~ Pressure_level,
        by = "Curriculum",
        adjust = "tukey")
```

(Top Level) R Script

```
$lsmeans
Curriculum = 1:
 Pressure_level lsmean   SE df lower.CL upper.CL
 1                14.9 1.22 72   12.484    17.36
 2                 4.0 1.22 72    1.561     6.44

Curriculum = 2:
 Pressure_level lsmean   SE df lower.CL upper.CL
 1                20.5 1.22 72   18.099    22.98
 2                 2.0 1.22 72   -0.439     4.44

Curriculum = 3:
 Pressure_level lsmean   SE df lower.CL upper.CL
 1                15.4 1.22 72   12.945    17.82
 2                12.8 1.22 72   10.330    15.21

Confidence level used: 0.95

$contrasts
Curriculum = 1:
 contrast estimate   SE df t.ratio p.value
 1 - 2       10.92 1.73 72   6.312  <.0001

Curriculum = 2:
 contrast estimate   SE df t.ratio p.value
 1 - 2       18.54 1.73 72  10.713  <.0001

Curriculum = 3:
 contrast estimate   SE df t.ratio p.value
 1 - 2        2.62 1.73 72   1.511  0.1351
```

So, now we can clearly see, for curriculum 1 and 2, there is difference in group means because for these two curriculums, p-value is less than 0.5 for low pressure and high pressure.