

Advanced statistical procedures (e.g., ANOVA, REGRESSION etc.) are run using the DATA ANALYSIS ADD-IN of EXCEL.

To add DATA ANALYSIS ADD-IN, do the following:

1. Click the **Microsoft Office Button** and then **Excel Options**.
2. Click **Add-Ins**, and then in the **Manage** box, select **Excel Add-ins**, then **Go**.
3. In the **Add-Ins available** box, select the **Analysis ToolPak** check box, and then click **OK**.

Tip: If **Analysis ToolPak** is not listed in the **Add-Ins available** box, click **Browse** to locate it. If you get prompted that the Analysis ToolPak is not currently installed on your computer, click **Yes** to install it.

4. After you load the Analysis ToolPak, the **Data Analysis** command is available in the **Analysis** group on the **Data** tab.

Example 1 (One-way ANOVA)

Suppose the USGA wants to compare the mean distances obtained when 4 different brands of golf balls are hit with one specific driver. The experiment is carried out with one robot golfer, and following data is obtained:

A	B	C	D
250.8	241.5	247	254.3
250	246.4	280.8	235.9
235.5	282.8	212.6	222.5
255.4	264.7	245.5	275.1
248.7	269.2	228	254
241.8	263.9	232.3	242.4
253.7	248.5	255.7	286
285.2	219.2	252.9	246.4
282.5	218.3	253.1	234.6
235.9	259.6	231.5	256.2
247.4	234.9	260.2	246.7
246.9	253.5	241.3	241.6

Test the null hypothesis that the average distance is same for the 4 brands.

H0: $\mu_A = \mu_B = \mu_C = \mu_D$

H1: H0 is false, i.e., at least two of the means are not equal.

To Run 1-way ANOVA for this example, go to a blank cell and then Data/Data Analysis/ Anova: Single Factor Provide INPUT range for the entire data

select COLUMNS since example has data in columns.

Provide an OUTPUT RANGE if you want the output in the same worksheet, else choose New Worksheet.

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Column 1	12	3033.8	252.8167	248.9033
Column 2	12	3002.5	250.2083	384.2427
Column 3	12	2940.9	245.075	314.1239
Column 4	12	2995.7	249.6417	303.799

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	373.0323	3	124.3441	0.397561	0.755398	2.816466
Within Groups	13761.76	44	312.7672			
Total	14134.79	47				

Since the P-value > .05, we conclude that the mean distances obtained from the 4 brands are equal.

NOTE:

Between Groups = BRAND

Within Groups = Error

Example 2:

Suppose the USGA wants to compare the mean distances obtained when 4 different brands of golf balls are hit with one specific driver. The robot golfer is too expensive, so 3 different golfers are used – one is 19 years old, one 35 years old and one 57 years old: each golfer hits 3 golf balls each from the 4 brands(A, B, C, D). A random sequence is used to hit each of these golf balls.

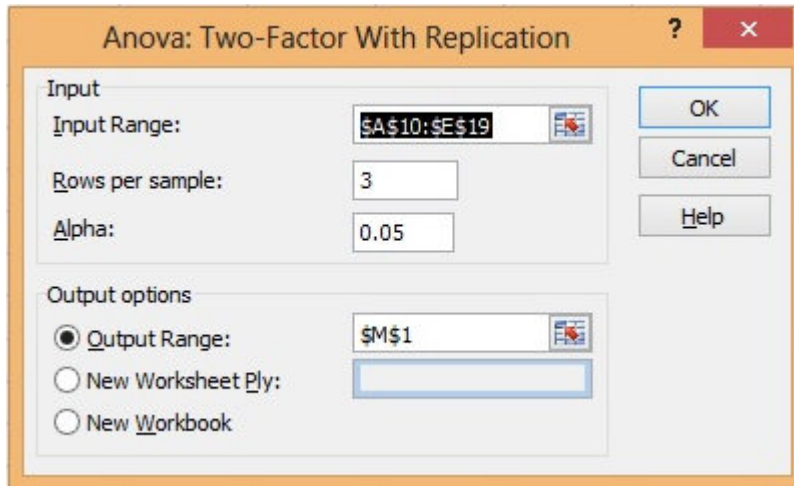
This is an example of a COMPLETELY RANDOMIZED BLOCK DESIGN – with AGE forming the blocks. Data is shown below.

Golfer	A	B	C	D
G1	248.73	222.18	251.76	228.35
G1	258.74	196.71	260.28	274
G1	223.65	238.99	223.3	229.95
G2	240.3	250.1	295.28	253.47
G2	258	245.93	236.26	238.51
G2	272.71	234.33	225.55	271.65
G3	241.07	261.1	263.22	225.15
G3	239.49	248.29	235.57	247.7
G3	249.64	229.46	239	244.8

In Excel:

Data/Data Analysis/Anova:Two-Factor with REplication/

Select data cells (including row and column headings), then select an output range.



The two-way ANOVA results from Excel are shown on the next page.

Anova: Two-Factor With Replication

SUMMARY	A	B	C	D	Total
<i>G1</i>					
Count	3	3	3	3	12
Sum	731.12	657.88	735.34	732.3	2856.64
Average	243.7067	219.2933	245.1133	244.1	238.0533
Variance	326.7524	453.1492	375.0137	671.1475	460.276
<i>G2</i>					
Count	3	3	3	3	12
Sum	771.01	730.36	757.09	763.63	3022.09
Average	257.0033	243.4533	252.3633	254.5433	251.8408
Variance	263.347	66.77363	1410.056	275.4289	394.9951
<i>G3</i>					
Count	3	3	3	3	12
Sum	730.2	738.85	737.79	717.65	2924.49
Average	243.4	246.2833	245.93	239.2167	243.7075
Variance	29.8273	253.2924	227.1493	150.5058	128.8236
<i>Total</i>					
Count	9	9	9	9	
Sum	2232.33	2127.09	2230.22	2213.58	
Average	248.0367	236.3433	247.8022	245.9533	
Variance	200.225	358.3256	514.882	320.2477	

Since the P-value(Sample or ROW) > .05, we conclude that the golfers have no effect on mean distance. P(Columns or BRAND) > .05, all 4 brands yield same mean distances. P(Interaction) > .05, there is no INTERACTION between Golfers and BRAND.

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Sample	1152.863	2	576.4317	1.536317	0.235596	3.402826
Columns	828.4633	3	276.1544	0.736012	0.54082	3.008787
Interaction	991.6913	6	165.2819	0.440512	0.844362	2.508189
Within	9004.887	24	375.2036			
Total	11977.91	35				

Excel uses non-standard statistical terminology: 'Sample' is referred to as 'BLOCK', 'Columns' in this example is 'TREATMENT', and 'Within' is called ERROR in statistics literature.