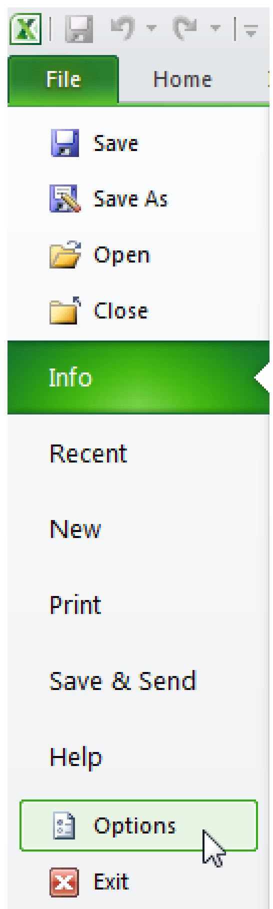
Descriptive statistics using excel

- Data Analysis

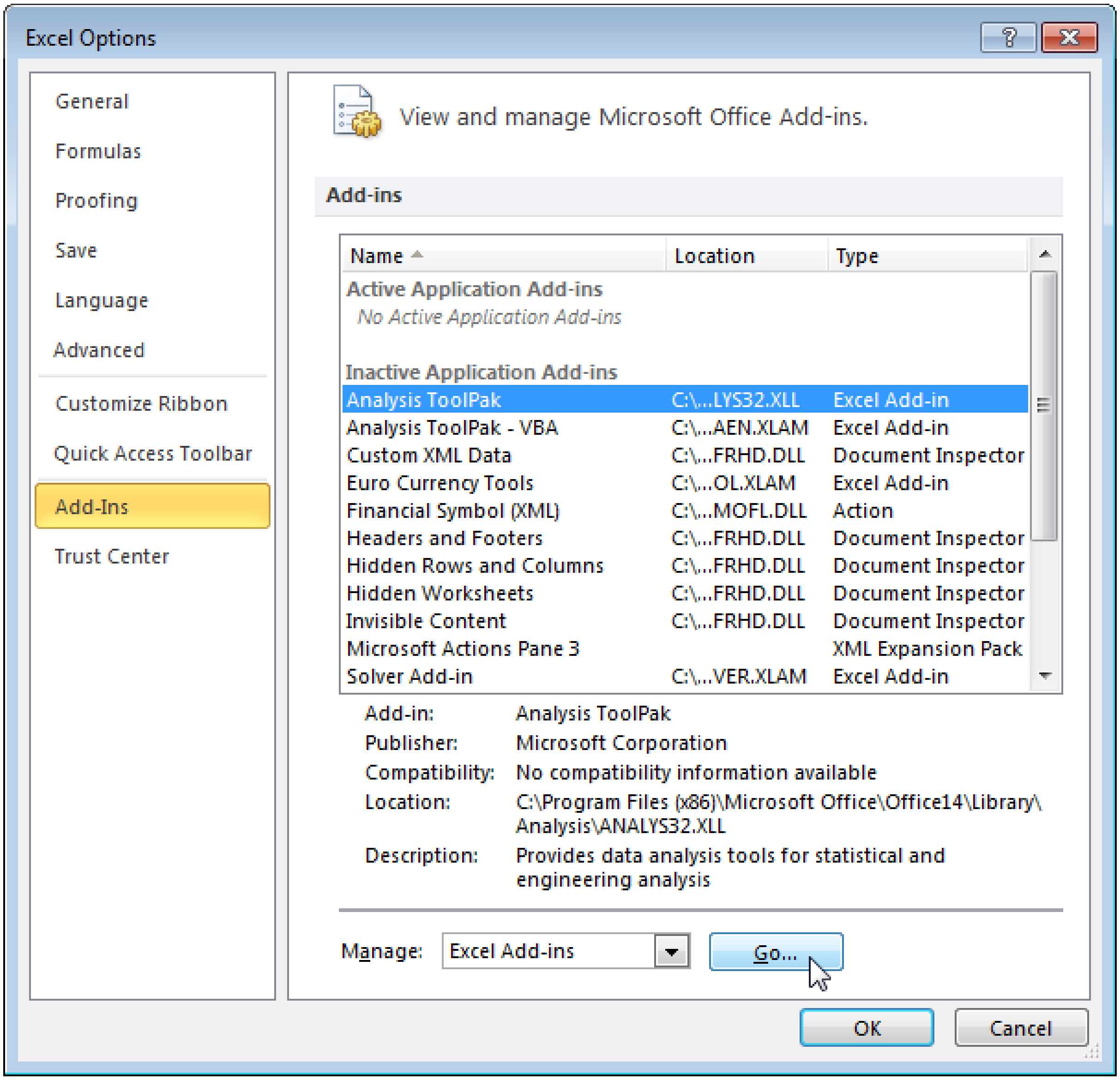
The Analysis ToolPak is an Excel add-in program that provides data analysis tools for financial, statistical and engineering data analysis.

To load the Analysis ToolPak add-in, execute the following steps.

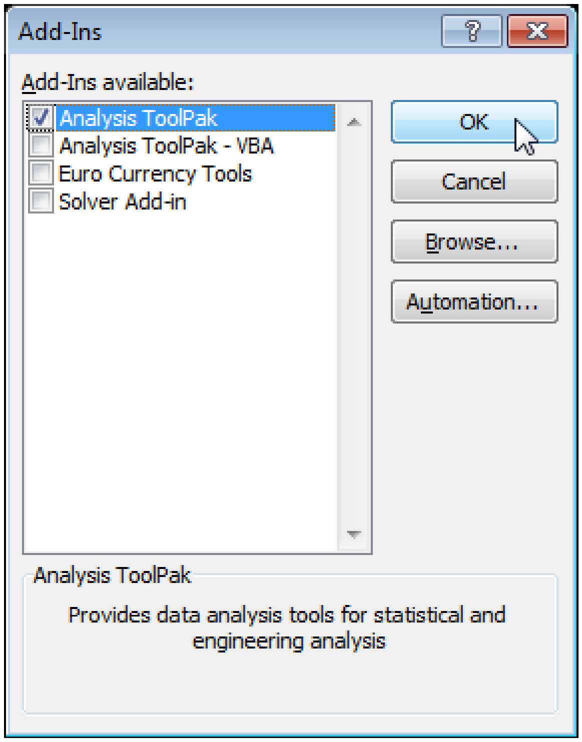
1. Click on Excel Options.



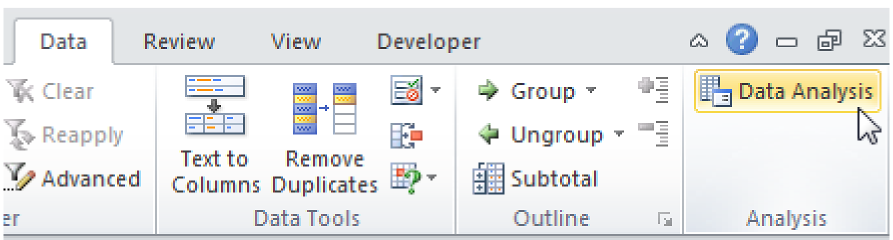
2. Under Add-ins, select Analysis ToolPak and click on the Go button.



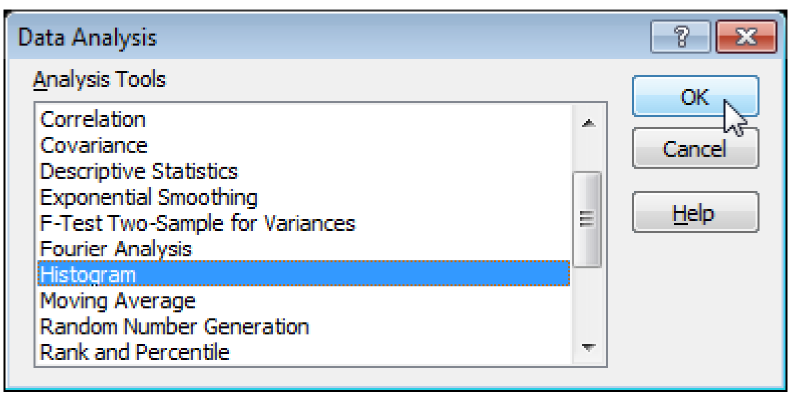
3. Check Analysis ToolPak and click on OK.



4. On the Data tab, you can now click on Data Analysis.



The following dialog box below appears.



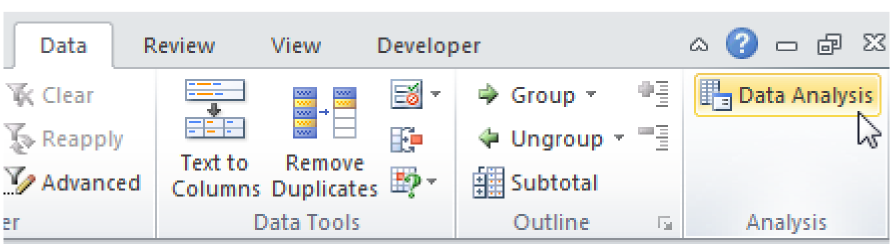
Examples on Data analysis

Descriptive Statistics

You can use the Analysis Toolpak add-in to generate descriptive statistics. For example, you may have the scores of 14 participants for a test.

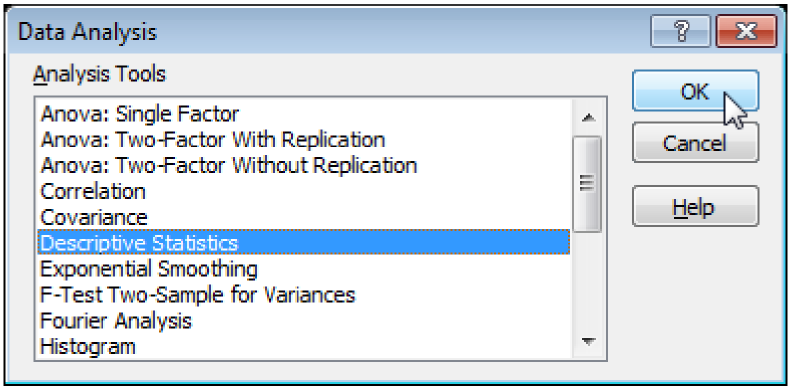
M26		
	A	B
1	Scores	
2	82	
3	93	
4	91	
5	69	
6	96	
7	61	
8	88	
9	58	
10	59	
11	100	
12	93	
13	71	
14	78	
15	98	
16		
17		

To generate descriptive statistics for these scores, execute the following steps.
1. On the Data tab, click Data Analysis.



Note: can't find the Data Analysis button? Click here to load the Analysis ToolPak add-in.

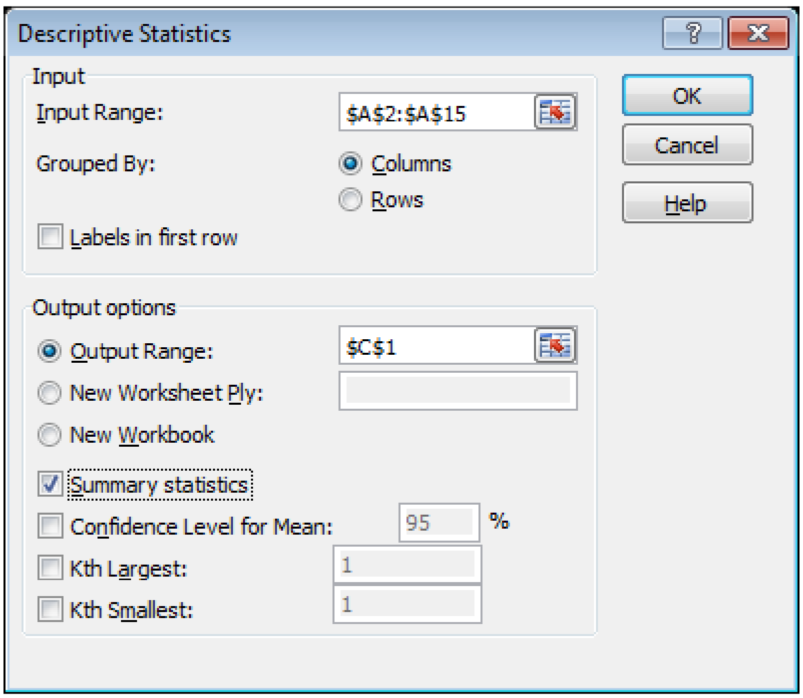
2. Select Descriptive Statistics and click OK.



3. Select the range A2:A15 as the Input Range.

4. Select cell C1 as the Output Range.

5. Make sure Summary statistics is checked.



6. Click OK.

Result:

L28 *fx*

	A	B	C	D	E
1	Scores		*Column1*		
2	82				
3	93		Mean	81.21428571	
4	91		Standard Error	4.045318243	
5	69		Median	85	
6	96		Mode	93	
7	61		Standard Deviation	15.13619489	
8	88		Sample Variance	229.1043956	
9	58		Kurtosis	-1.426053506	
10	59		Skewness	-0.402108004	
11	100		Range	42	
12	93		Minimum	58	
13	71		Maximum	100	
14	78		Sum	1137	
15	98		Count	14	
16					
17					

<u>**Statistical tests using excel**</u>

1- Analysis of variance (ANOVA)

- One Way ANOVA

$\alpha = 0.05.$

<u>**Hypotheses testing:**</u>

$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$

H_1: at least one of the means is different.

<u>**Test statistic:**</u>

$$F = \frac{MSA}{MSE}$$

<u>**Critical region:**</u>

Reject H_0 if $F > F_{1-\alpha,(v_1,v_2)}$

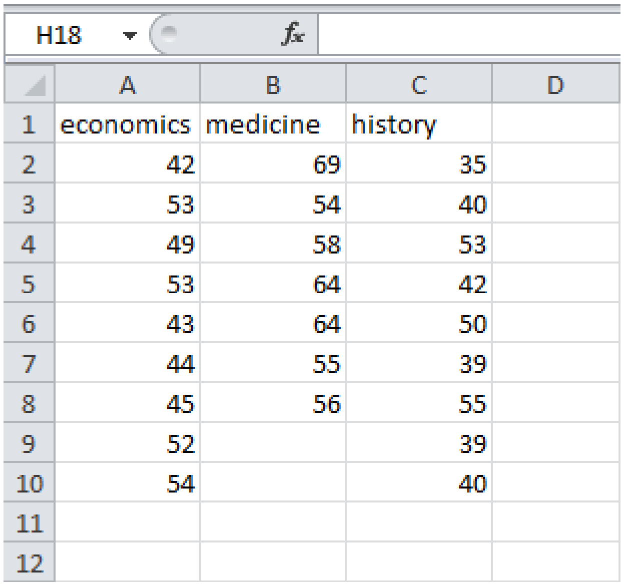
Or

$p - value \leq \alpha$

<u>**Decision:**</u>

This example teaches you how to perform a single factor ANOVA (analysis of variance) in Excel. A single factor or one-way ANOVA is used to test the null hypothesis that the means of several populations are all equal.

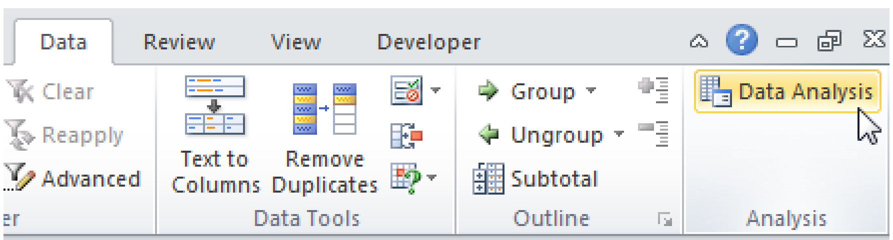
Below you can find the salaries of people who have a degree in economics, medicine or history.



	A	B	C	D
1	economics	medicine	history	
2	42	69	35	
3	53	54	40	
4	49	58	53	
5	53	64	42	
6	43	64	50	
7	44	55	39	
8	45	56	55	
9	52		39	
10	54		40	
11				
12				

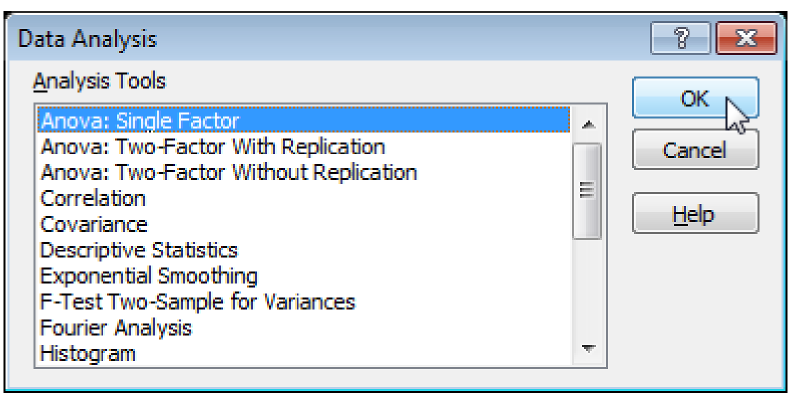
To perform a single factor ANOVA, execute the following steps.

1. On the Data tab, click Data Analysis.



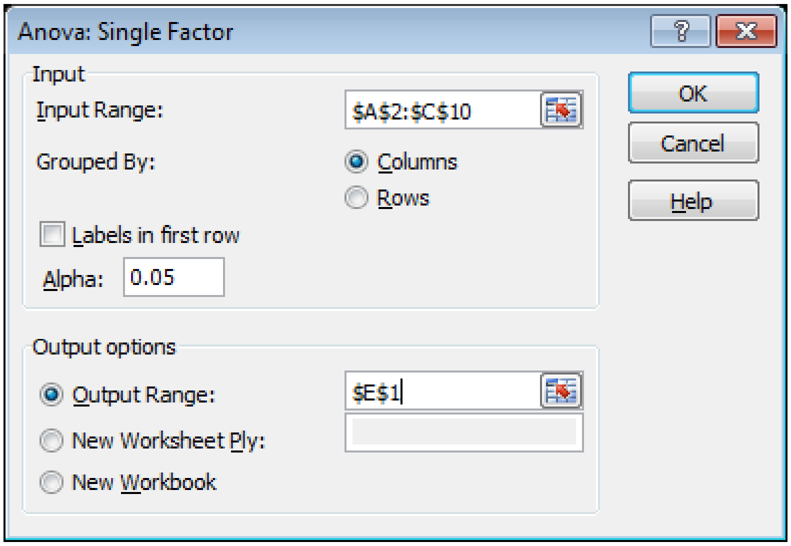
Note: can't find the Data Analysis button? Click here to load the Analysis ToolPak add-in.

2. Select Anova: Single Factor and click OK.



3. Click in the Input Range box and select the range A2:C10.

4. Click in the Output Range box and select cell E1.



5. Click OK.

Result:

E	F	G	H	I	J	K
Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Column 1	9	435	48.33333	23.5		
Column 2	7	420	60	32.33333		
Column 3	9	393	43.66667	50.5		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1085.84	2	542.92	15.19623	7.16E-05	3.443357
Within Groups	786	22	35.72727			
Total	1871.84	24				

$\alpha = 0.05.$

Hypotheses testing:

$H_0: \mu_1 = \mu_2 = \mu_3$

H_1: at least one of the means is different.

Test statistic:

$$F = \frac{MSA}{MSE} = 15.19623$$

Critical region:

$F_{1-\alpha,(v_1,v_2)} = F\ crit = 3.443357$

$p - value = 0.00007163$

Reject H_0 if $F > 3.443$

Or

$p - value \leq 0.05$

Decision:

Therefore, we reject the null hypothesis. The means of the three populations are not all equal. At least one of the means is different. However, the ANOVA does not tell you where the difference lies. You need a t-Test to test each pair of means.

- Two-Factor Analysis of Variance

$\alpha = 0.05.$

The three hypotheses to be tested are as follows:
Hypotheses testing:

1. $H_0: a_1 = a_2 = \cdots = a_a = 0$

   H_1: At least one of the a_i is not equal to zero.

2. $H_0: \beta_1 = \beta_2 = \cdots = \beta_b = 0$

   H_1: At least one of the β_j is not equal to zero.

3. $H_0: (a\beta)_{11} = (a\beta)_{12} = \cdots = (a\beta)_{ab} = 0$

   H_1: At least one of the $(a\beta)_{ij}$ is not equal to zero.

Critical region:

$p - value \leq \alpha$

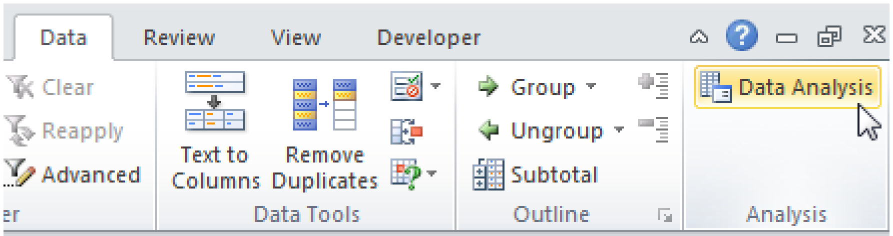
Decision:

- ANOVA two ways with Replication

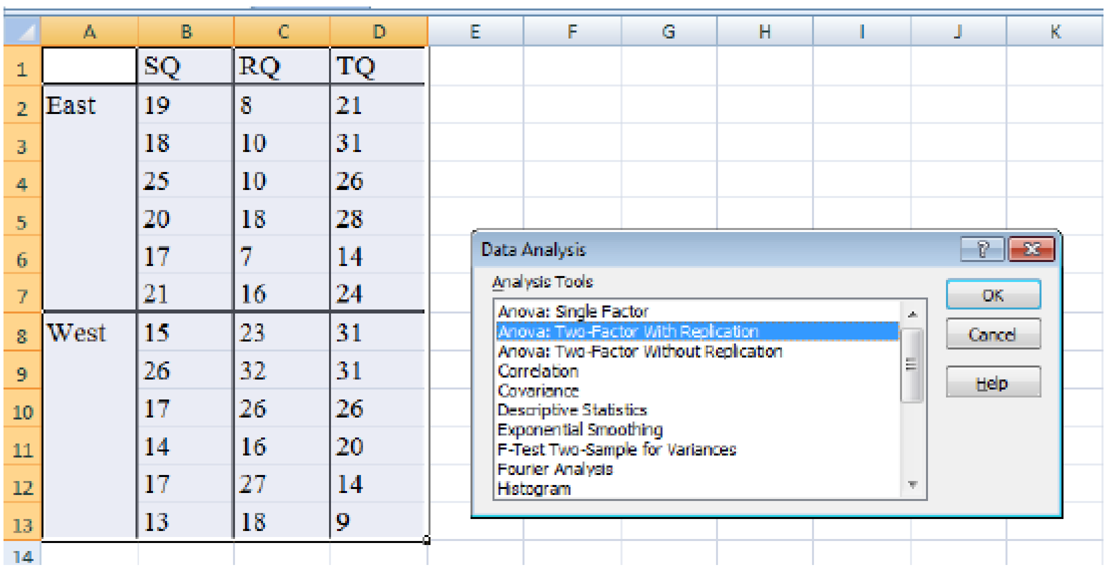
In the following example, the cost of three swimming programs in two regions in the East and the West

	SQ	RQ	TQ
East	19	8	21
	18	10	31
	25	10	26
	20	18	28
	17	7	14
	21	16	24
West	15	23	31
	28	32	31
	17	26	26
	14	16	20
	17	27	14
	13	18	9

1. On the Data tab, click Data Analysis.

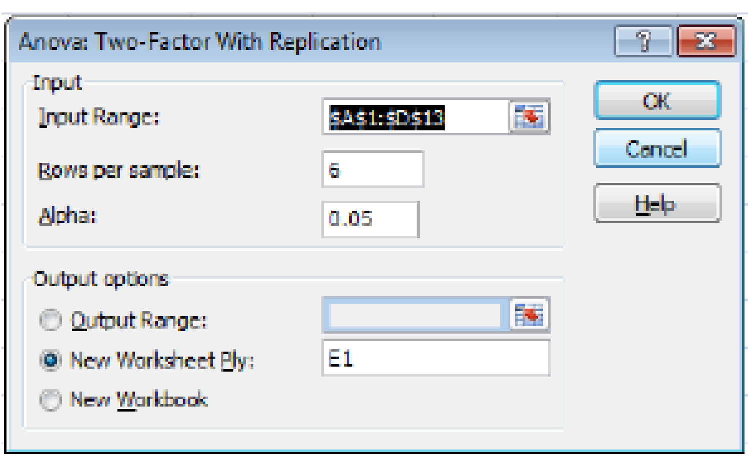


2. Select Anova: Two-Factor With Replication and click OK.



3. Click in the Input Range box and select the range A1:D13.

4. Click in the Output Range box and select cell E2.



5. Click OK.

Result:

Anova: Two-Factor With Replication						
SUMMARY	SQ	RQ	TQ	Total		
East						
Count	6	6	6	18		
Sum	120	69	144	333		
Average	20	11.5	24	18.5		
Variance	8	19.9	35.6	47.44118		
West						
Count	6	6	6	18		
Sum	102	142	131	375		
Average	17	23.66667	21.83333	20.83333		
Variance	22	35.46667	82.96667	49.67647		
Total						
Count	12	12	12			
Sum	222	211	275			
Average	18.5	17.58333	22.91667			
Variance	16.09091	65.53788	55.17424			
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	49	1	49	1.441648	0.239268	4.170877
Columns	195.1667	2	97.58333	2.871036	0.072297	3.31583
Interaction	436.1667	2	218.0833	6.416313	0.004788	3.31583
Within	1019.667	30	33.98889			
Total	1700	35				

$\alpha = 0.05.$

<u>1) Hypotheses testing:</u>

$H_0: a_{East} = a_{West} = 0$

H_1: at least one of the means is different.

<u>Test statistic:</u>

$F = 1.4416$

<u>Critical region:</u>

$F_{1-\alpha,(v_1,v_2)} = F\ crit = 4.1708$

$p - value = 0.2392$

Reject H_0 if $F > 4.1708$ Or $p - value \leq 0.05$

Decision:

Therefore, we not reject the null hypothesis. There are no difference in the means of the cost between the two regions in the east and the west.

2) Hypotheses testing:

H_0: $\beta_{SQ} = \beta_{RQ} = \beta_{TQ} = 0$

H_1: at least one of the means is different.

Test statistic:

$F = 2.871$

Critical region:

$F_{1-\alpha,(v_1,v_2)} = F\ crit = 3.3158$

$p - value = 0.0722$

Reject H_0 if $F > 3.3158$ Or $p - value \leq 0.05$

Decision:

Therefore, we not reject the null hypothesis. There was no significant difference between the mean cost of the three programs.

3) Hypotheses testing:

H_0: $a\beta_{East,SQ} = a\beta_{East,RQ} = a\beta_{East,TQ} = a\beta_{West,SQ} = a\beta_{West,RQ} = a\beta_{West,TQ} = 0$

H_1: at least one of the means is different.

Test statistic:

$F = 6.4163$

Critical region:

$F_{1-\alpha,(v_1,v_2)} = F\ crit = 3.3158$

$p - value = 0.0047$

Reject H_0 if $F > 3.3158$ Or $p - value \leq 0.05$

Decision:

Therefore, we reject the null hypothesis. There is interaction between the two factors (region, type of program).

- ANOVA two ways without Replication

In the following example, the price of three swimming programs in six areas

	SQ	RQ	TQ
1	19	8	21
2	18	10	31
3	25	10	26
4	20	18	28
5	17	7	14
6	21	16	24

$\alpha = 0.05.$

The two hypotheses to be tested are as follows:

Hypotheses testing:

1. $H_0: a_1 = a_2 = \cdots = a_a = 0$

   H_1: At least one of the a_i is not equal to zero.

2. $H_0: \beta_1 = \beta_2 = \cdots = \beta_b = 0$

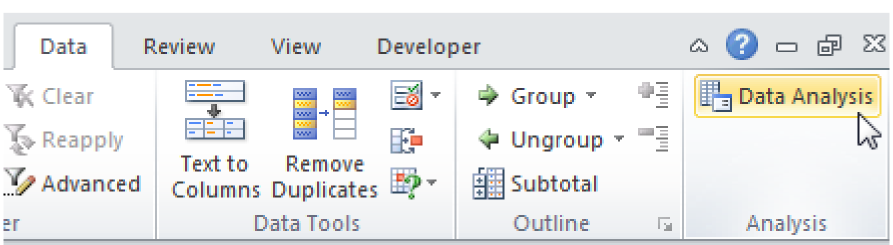
   H_1: At least one of the β_j is not equal to zero.

Critical region:

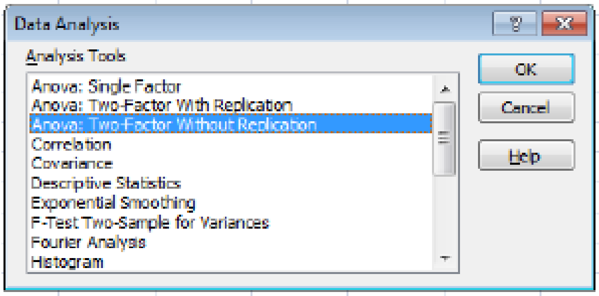
$p - value \leq \alpha$

Decision:

1. On the Data tab, click Data Analysis.

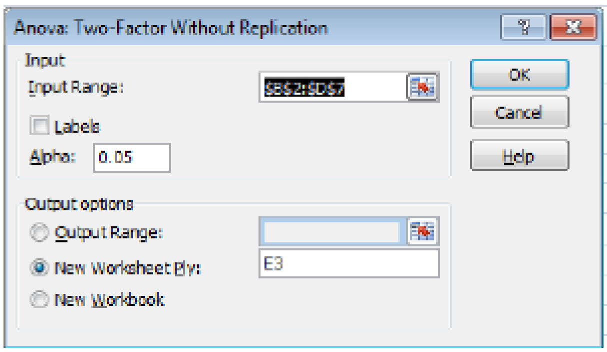


2. Select Anova: Two-Factor With Replication and click OK.



3. Click in the Input Range box and select the range B2:D7.

4. Click in the Output Range box and select cell E3.



5. Click OK.

Result:

Anova: Two-Factor Without Replication						
SUMMARY	Count	Sum	Average	Variance		
Row 1	3	48	16	49		
Row 2	3	59	19.66667	112.3333		
Row 3	3	61	20.33333	80.33333		
Row 4	3	66	22	28		
Row 5	3	38	12.66667	26.33333		
Row 6	3	61	20.33333	16.33333		
Column 1	6	120	20	8		
Column 2	6	69	11.5	19.9		
Column 3	6	144	24	35.6		
ANOVA						
ce of Varic	SS	df	MS	F	P-value	F crit
Rows	181.8333	5	36.36667	2.68059	0.086618	3.325835
Columns	489	2	244.5	18.02211	0.000483	4.102821
Error	135.6667	10	13.56667			
Total	806.5	17				

$\alpha = 0.05.$

1) Hypotheses testing:

$H_0: a_1 = a_2 = \cdots = a_6 = 0$

H_1: at least one of the means is different.

Test statistic:

$F = 2.6805$

Critical region:

$F_{1-\alpha,(v_1,v_2)} = F\ crit = 3.3258$

$p - value = 0.0866$

Reject H_0 if $F > 3.3258$ Or $p - value \leq 0.05$

Decision:

Therefore, we not reject the null hypothesis. There are no difference in the means of the cost between the six areas.

2) Hypotheses testing:

H_0: $\beta_{SQ} = \beta_{RQ} = \beta_{TQ} = 0$

H_1: at least one of the means is different.

Test statistic:

$F = 18.022$

Critical region:

$F_{1-\alpha,(v_1,v_2)} = F\ crit = 4.1028$

$p - value = 0.0004$

Reject H_0 if $F > 4.1028$ Or $p - value \leq 0.05$

Decision:

Therefore, we reject the null hypothesis. There is a significant difference between the mean prices of the three programs.