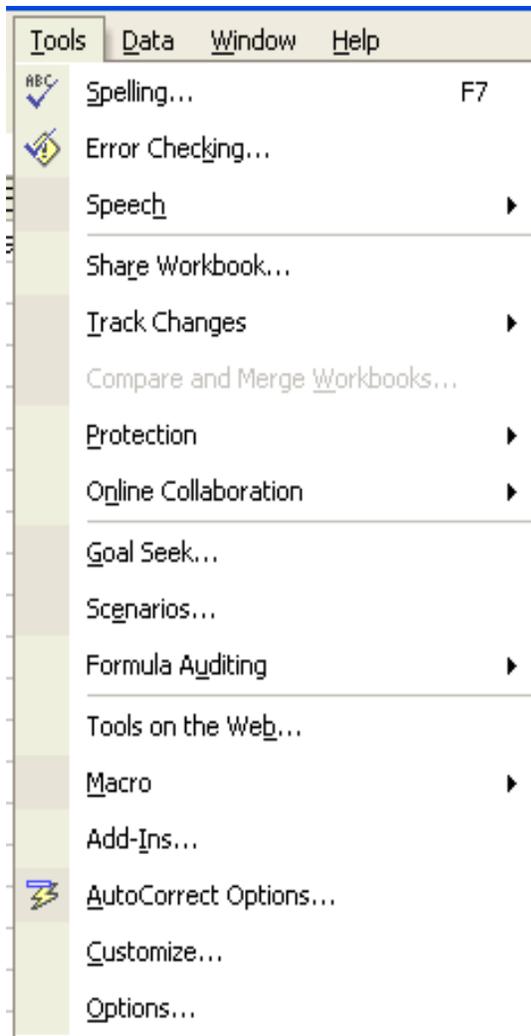


## LECTURE SCHEDULE 10

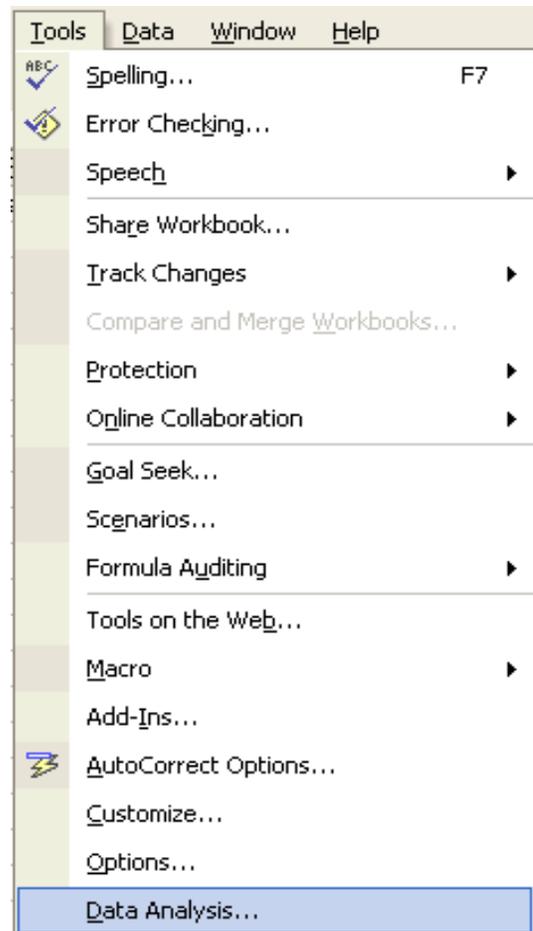
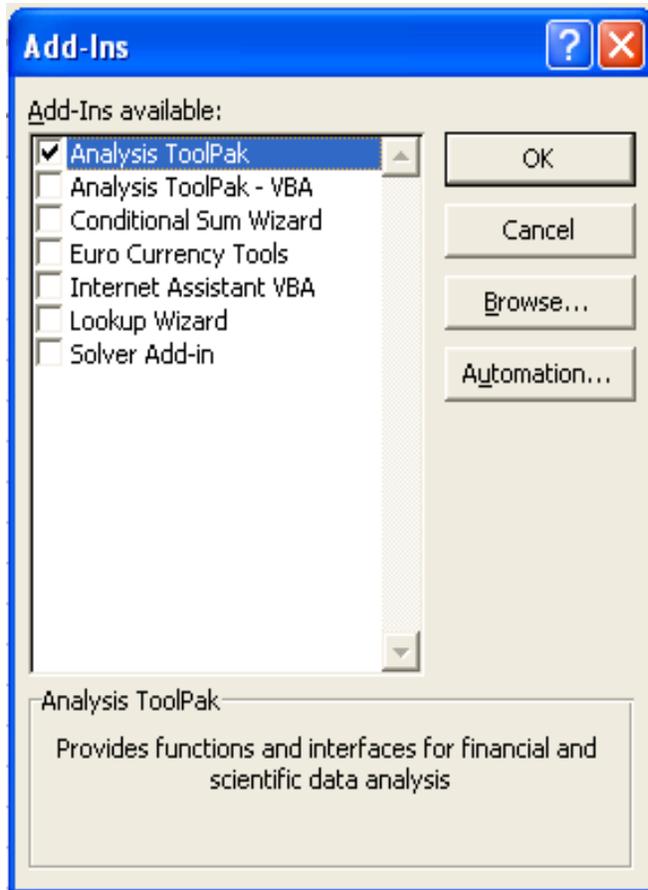
### Data Analysis Tools- Correlation and Regression, t-test for two samples and ANOVA with One-way classification

#### Data Analysis Tools

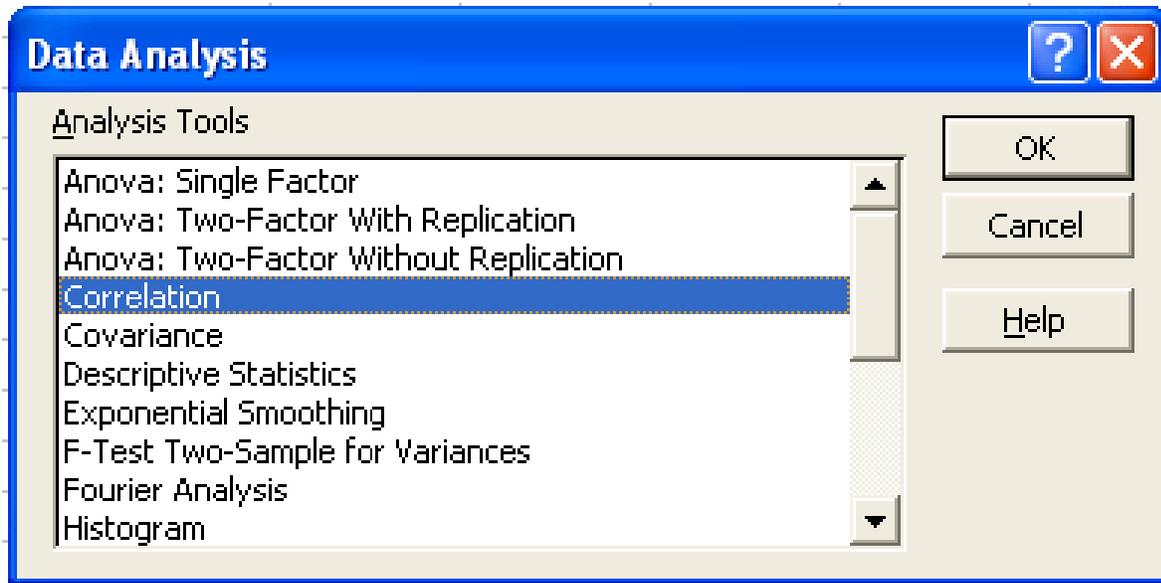
- Most of Excel's statistical procedures are part of the Data Analysis ToolPak. This option is available in TOOLS menu.
- The analysis that can be done are:
  - descriptive statistics,
  - t tests,
  - correlations,
  - one or two way analysis of variance,
  - regression,
  - moving average,
  - Fourier analysis etc.
- To start the data analysis click Tools menu. Search for the Data Analysis tab in the Tools menu.



- If Data Analysis is not found in the menu then click Add-Ins...
- In the Add\_Ins dialog box check Analysis ToolPak. Then Click OK.
- Now choose the Tools menu. You can find the Data Analysis there.



- Click Data Analysis option in the Tools now.
- Data Analysis dialog box will get displayed as shown:

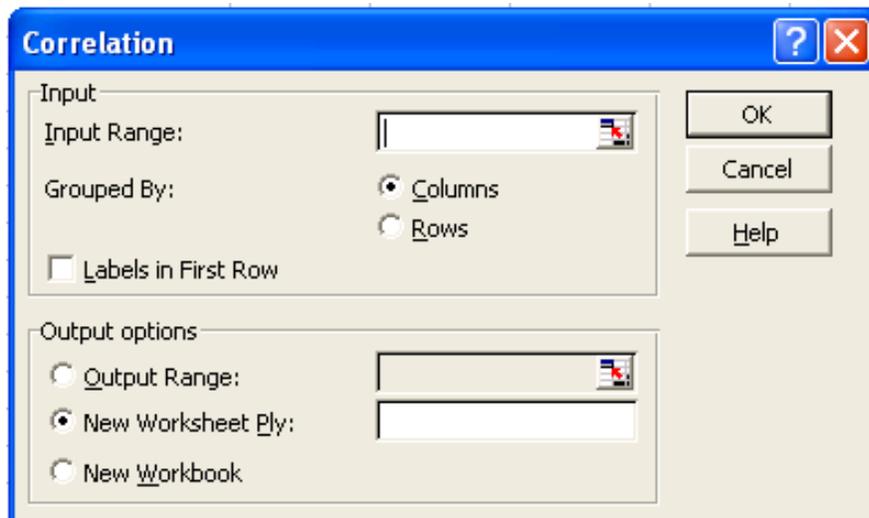


## Correlation

- To do the correlation we consider a problem of finding correlation coefficient of data which relates to the Yield in grams (Y) and the Matured Pods (X) of 10 groundnut plants.
- The data is entered in the Excel sheet as follows:

	A	B	C	D	E	F
1	The followingf data relates to the Yield (in gms)and Matured Pods in numbers of 10 groundnut plants					
2	Yield(Gms)	Matured Pod(in numbers)				
3	14	16				
4	34	40				
5	20	21				
6	16	18				
7	11	14				
8	11	13				
9	20	20				
10	17	35				
11	22	17				
12	17	27				

- Choose Correlation from the Data Analysis window then click OK button.
- Correlation window will get displayed as follows:



- Enter the input range if you know the input range of data or else make use of the address icon to mark the input range in the Excel sheet. In this example the input range is A2:B12.
- Since the data are grouped by column in this example, choose Grouped By Columns option.
- We can add the labels as the first row in the data range. So check the Labels in First Row.
- In the output options choose where you want to display the output. In this example the output range is given as D2.



- Click OK button in the Correlation window.
- The result will be displayed from D2 as shown below:

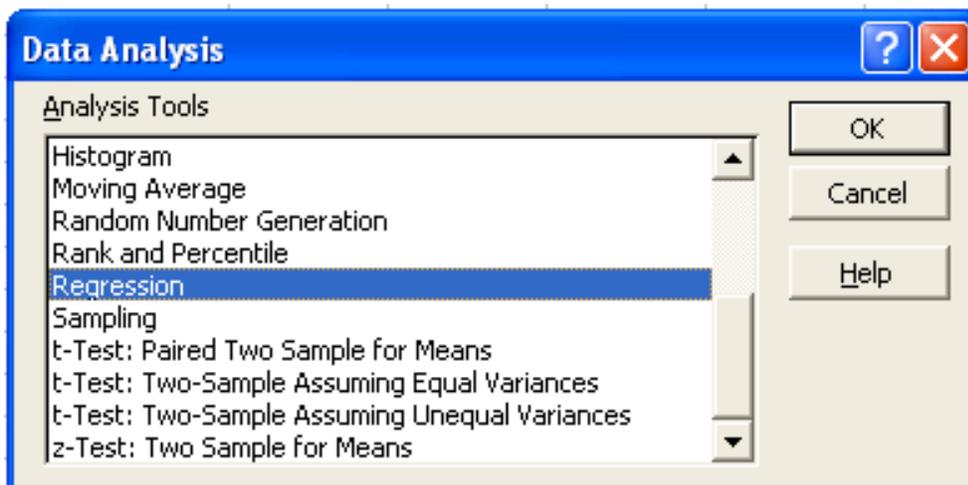
	B	C	D	E	F
1	tes to the Yield (in gms)and Matured Pods in numbers of 10 groundnut plants				
2	Matured Pod(in numbers)			Yield(Gms)	Matured Pod(in numbers)
3	16		Yield(Gms)	1	
4	40		Matured Pod(in numbers)	0.72362809	1
5	21				
6	18				
7	14				
8	13				
9	20				
10	35				
11	17				
12	27				

### **Regression**

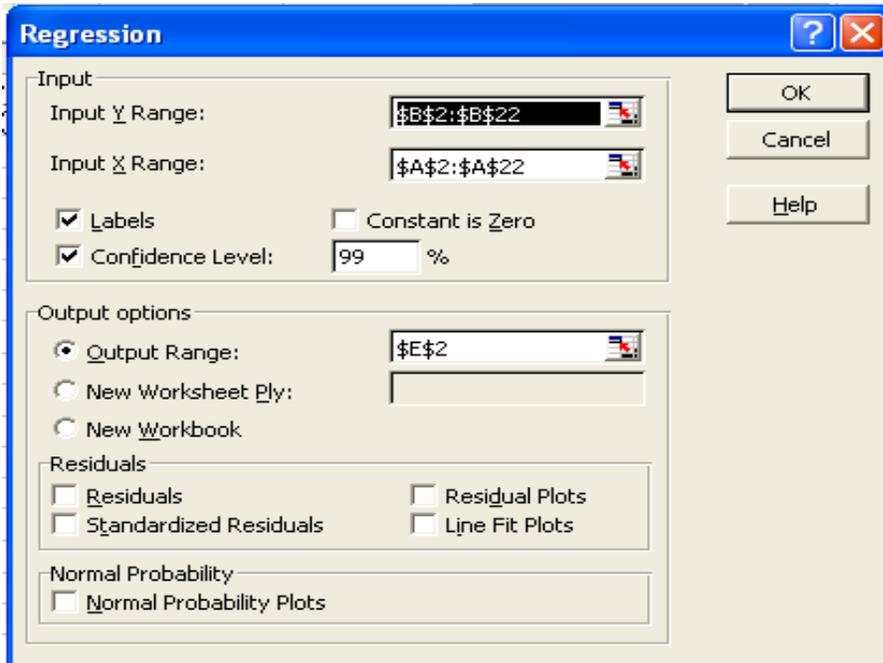
- To do the regression analysis we consider an example with paddy yield in Kg as dependent variable and ear length in cm as the independent variable.
- The data are entered in the Excel sheet as follows:

	A	B	C	D
1	The following data refers to the paddy yield (in Kg) and ear length in (cm)			
2	Paddy Yield(Kg)	Ear Length (cm)		
3	3.5	15		
4	4.15	17		
5	3.8	14		
6	4.2	18		
7	3.95	17		
8	4.17	19		
9	4.35	20		
10	4.32	22		
11	5.01	17		
12	4.87	16		
13	4.05	17		
14	4.28	20		
15	4.6	22		
16	4.1	20		
17	4.72	24		
18	4.25	18		
19	4.1	17		
20	4.65	19		
21	4.89	20		
22	4.99	18		

- The Regression procedure in the Data Analysis tools lets you choose one column as the dependent variable, and a set of contiguous columns for the independents
- It does not tolerate any empty cells anywhere in the input ranges.
- Therefore, if there is any empty cells in the input range, delete the rows contain empty cells.
- Choose Regression from Data Analysis and click OK.



- In the Regression window enter the Y and X ranges if you know the address values or else the click the colored icon in the Y and X range tabs to fill the addresses.
- If the first row of the data entered includes labels then check the Labels box in the Regression window.
- Let the confidence level be set to 99%.
- In the output options choose where you want to display the output of regression analysis. In this example the Output Range is set to E2.



- Click OK. The result will be displayed fro E2 as shown below:

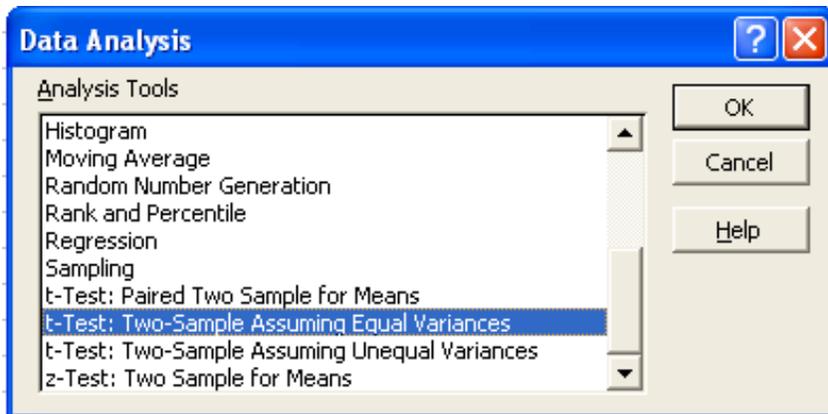
	D	E	F	G	H	I	J	K	L	M
1	in( cm)									
2		SUMMARY OUTPUT								
3										
4		<b>Regression Statistics</b>								
5		Multiple R	0.410574							
6		R Square	0.168571							
7		Adjusted R Square	0.122381							
8		Standard Error	2.304758							
9		Observations	20							
10										
11		<b>ANOVA</b>								
12			<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
13		Regression	1	19.38567	19.38567	3.649474	0.072151			
14		Residual	18	95.61433	5.311907					
15		Total	19	115						
16										
17			<i>Coefficient</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 99.0%</i>	<i>Upper 99.0%</i>
18		Intercept	7.811454	5.618729	1.390253	0.181405	-3.99307	19.61597	-8.36173	23.98464
19		Paddy Yie	2.45855	1.286957	1.91036	0.072151	-0.24525	5.162348	-1.24588	6.16298

### t-test for two samples

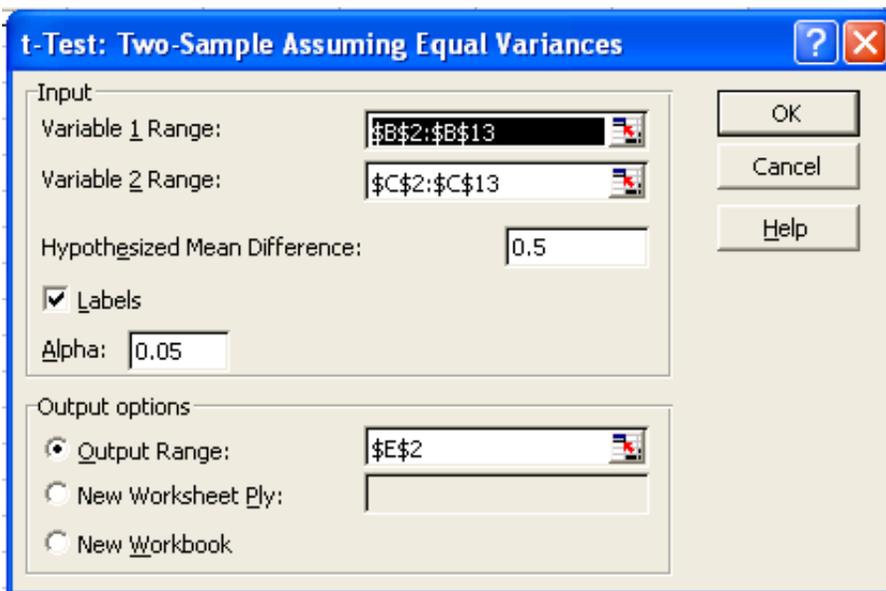
- t-test for two samples assuming equal variances: yield of food in kg/ha in Kharif and Rabi season from 1996-97 to 2006-07.
- The data is entered in Excel sheet as shown below:

	A	B	C	D	E	F	G	H	I	J	K
1	Perform t-test for two samples assuming equal variances: yield of food in kg/ha in Kharif and Rabi season from 1996-97 to 2006-07										
2	Year	Kharif	Rabi								
3	96-97	89	104								
4	97-98	94	109								
5	98-99	94	116								
6	99-2000	78	105								
7	2000-01	93	120								
8	2001-2002	95	119								
9	2002-2003	92	130								
10	2003-2004	106	131								
11	2004-2005	104	134								
12	2005-2006	105	142								
13	2006-2007	110	145								

- Choose t-Test: Two-Samples assuming Equal Variances in the Data Analysis window:



- In the t-Test: Two-Samples assuming Equal Variances window enter the Variable 1 Range and Variable 2 Range.
- Check the Labels option
- Hypothesized mean difference be 0.5
- Let Alpha value be 0.05
- Set output range as E2.



- The result will be displayed from E2 as shown below:

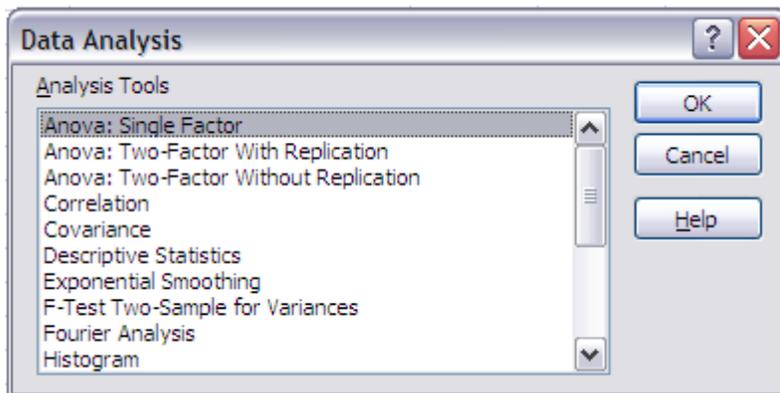
	A	B	C	D	E	F	G	H	I	J	K
1	Perform t-test for two samples assuming equal variances: yield of food in kg/ha in Kharif and Rabi season from 1996-97 to 2006-07										
2	Year	Kharif	Rabi		t-Test: Two-Sample Assuming Equal Variances						
3	96-97	89	104								
4	97-98	94	109			<i>Kharif</i>	<i>Rabi</i>				
5	98-99	94	116		Mean	96.36364	123.1818				
6	99-2000	78	105		Variance	84.65455	203.3636				
7	2000-01	93	120		Observatio	11	11				
8	2001-2002	95	119		Pooled Va	144.0091					
9	2002-2003	92	130		Hypothesis	0.5					
10	2003-2004	106	131		df	20					
11	2004-2005	104	134		t Stat	-5.33873					
12	2005-2006	105	142		P(T<=t) on	1.59E-05					
13	2006-2007	110	145		t Critical of	1.724718					
14					P(T<=t) tw	3.17E-05					
15					t Critical tw	2.085962					
16											

### **ANOVA with One-way Classification**

- Perform ANOVA One Way Classification for yield of food in kg/ha in three seasons from 1996-97 to 2006-07
- The data is entered in Excel sheet as follows:

	A	B	C	D	E	F	G	H
1	Perform ANOVA One Way Classification for yield of food in kg/ha in three seasons from 1996-97 to 2006-07							
2	Year	SEASON1	SEASON2	SEASON3				
3	96-97	89	104	130				
4	97-98	94	109	134				
5	98-99	94	116	135				
6	99-2000	78	105	135				
7	2000-01	93	120	142				
8	2001-2002	95	119	143				
9	2002-2003	92	130	144				
10	2003-2004	106	131	150				
11	2004-2005	104	134	152				
12	2005-2006	105	142	154				
13	2006-2007	110	145	160				

- Choose ANOVA: Single Factor from Data Analysis window.



- In the ANOVA: Single Factor window enter the input range. The input range in the example is B2:D13
- Choose Group by Columns
- Check Labels in the first row
- Output range is set to F2 as follows:

	A	B	C	D	E	F	G	H	I	
1	Perform ANOVA One Way Classification for yield of food in kg/ha in three seasons from 1996-97 to 2006-07									
2	Year	SEASON1	SEASON2	SEASON3						
3	96-97	89	104	130						
4	97-98	94	109	134						
5	98-99	94	116	135						
6	99-2000	78	105	135						
7	2000-01	93	120	142						
8	2001-2002	95	119	143						
9	2002-2003	92	130	144						
10	2003-2004	106	131	150						
11	2004-2005	104	134	152						
12	2005-2006	105	142	154						
13	2006-2007	110	145	160						
14										
15										
16										
17										

**Anova: Single Factor**

Input  
 Input Range:

Grouped By:  Columns  Rows

Labels in first row

Alpha:

Output options  
 Output Range:

New Worksheet Ply:

New Workbook

OK Cancel Help

- The result of the ANOVA with One-way classification is displayed from F2 as shown below:

	A	B	C	D	E	F	G	H	I	J	K	L
1	Perform ANOVA One Way Classification for yield of food in kg/ha in three seasons from 1996-97 to 2006-07											
2	Year	SEASON1	SEASON2	SEASON3		Anova: Single Factor						
3	96-97	89	104	130								
4	97-98	94	109	134		SUMMARY						
5	98-99	94	116	135		<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>		
6	99-2000	78	105	135		SEASON1	11	1060	96.36364	84.65455		
7	2000-01	93	120	142		SEASON2	11	1355	123.1818	203.3636		
8	2001-2002	95	119	143		SEASON3	11	1579	143.5455	91.67273		
9	2002-2003	92	130	144								
10	2003-2004	106	131	150								
11	2004-2005	104	134	152		ANOVA						
12	2005-2006	105	142	154		<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
13	2006-2007	110	145	160		Between Groups	12320.06	2	6160.03	48.67141	3.82E-10	3.31583
14						Within Groups	3796.909	30	126.5636			
15												
16						Total	16116.97	32				