Sub Module 1.5

5. Use of EXCEL for regression analysis

EXCEL is a Microsoft product that comes along with the Office suite of programs. It is essentially a spread sheet program that provides a computing environment with graphic capabilities. The student is encouraged to learn the basics of EXCEL programming so that data analysis, regression analysis and suitable plots may all be done within the EXCEL environment.
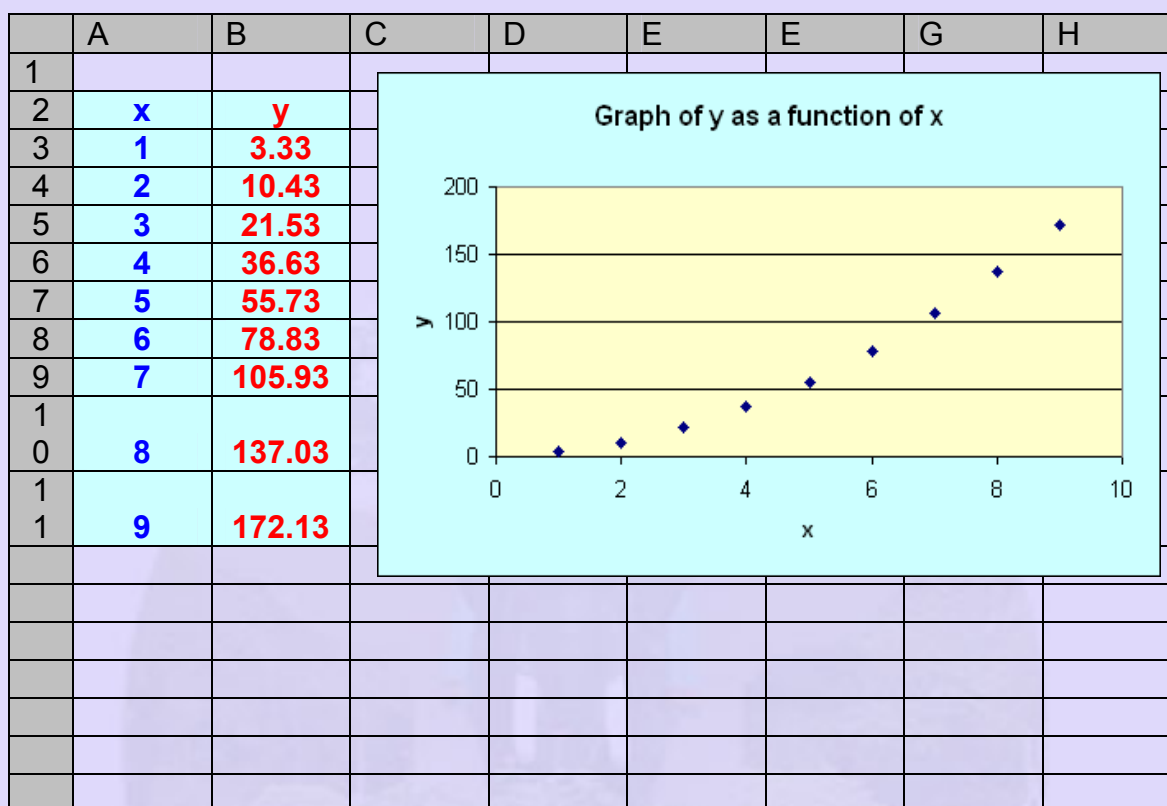
EXCEL work sheet provides a grid with cells in it. The cells form columns and rows as in a matrix. The columns are identified by alphanumeric symbols and the rows by numerals. For example, A1 refers to the cell in the first column and first row. Cell C5 will represent the cell in the 3rd column (column number C) and the 5th row (row number 5). Column identifiers will go from A – Z and then from AA – AZ and so on.. The cell can hold a number, a statement or a formula. A number or a statement is simply written by putting the cursor in the appropriate cell and keying in the number or the statement, as the case may be. A formula, however, is written by preceding the formula by "=" sign. The formula can contain a reference to many built in functions in EXCEL as well as the usual arithmetic operations. The formula can refer to the content of any other cell or cells. The formulas can be calculated repeatedly over a set of rows by simply copying down the formula vertically.

Figure 11 An extract of an EXCEL work sheet shows some of the things one can do!

	A	B	C	D	E	F	G
1							
2	23	88					
3	This is a statement						
4	In cell B4 is the formula "=A2*B2" i.e. product of two numbers	2024					
5	The formula in B4 is acted upon and the result alone appears in the cell B4 as seen above..						
6						x	x^2
7						1	1
8						2	4
9						3	9
10						4	16
11						5	25
12						6	36
13						7	49
14						8	64
15						9	81
16						10	100
17	Sum of G7to G16 is obtained by entering the formula "=SUM(G7:G16) in Cell G17						385
18	SUM() is a built in function in EXCEL						

Data may be keyed into the cells in the form of columns as shown in the work sheet given as Figure 12 below. The plotting is menu driven and the plot may be displayed as a separate plot or within the work sheet. The latter is shown in the case given here. The data range for the plot is specified by simply blocking the Data cells shown by the blue background!

Figure 12 Another extract of an EXCEL work sheet, showing data and the corresponding plot.

	A	B	C	D	E	E	G	H
1								
2	x	y						
3	1	3.33						
4	2	10.43						
5	3	21.53						
6	4	36.63						
7	5	55.73						
8	6	78.83						
9	7	105.93						
10	8	137.03						
11	9	172.13						

Properties of cells, chart (plot is referred to as chart in EXCEL) are changed to suit the requirements with menu driven controls. Student should familiarize oneself by learning these through "HELP" available in EXCEL.

Figure 13 Another extract of an EXCEL work sheet, showing data and the corresponding plot along with the automatically generated fit. The inset in the plot gives the linear relation between y and x. The square of the correlation coefficient is also shown in the inset (symbol R^2).

	A	B	C	D	E	F	G	H
1	The following data is expected to follow a linear law. Obtain such a law							
2	by using the "Trend Line" option in EXCEL.							
3	x	y_d						
4	0.5	0.35						
5	1	1.66						
6	1.5	3.418						
7	2	4.488						
8	2.5	5.306						
9	3	8.584						
10	3.5	9.97						
11	4	12.196						
12	4.5	15.382						
13	5	15.548						
14	5.5	17.274						
15	6	18.704						
16	6.5	20.306						
17	7	21.612						
18	7.5	21.446						
19	8	24.108						

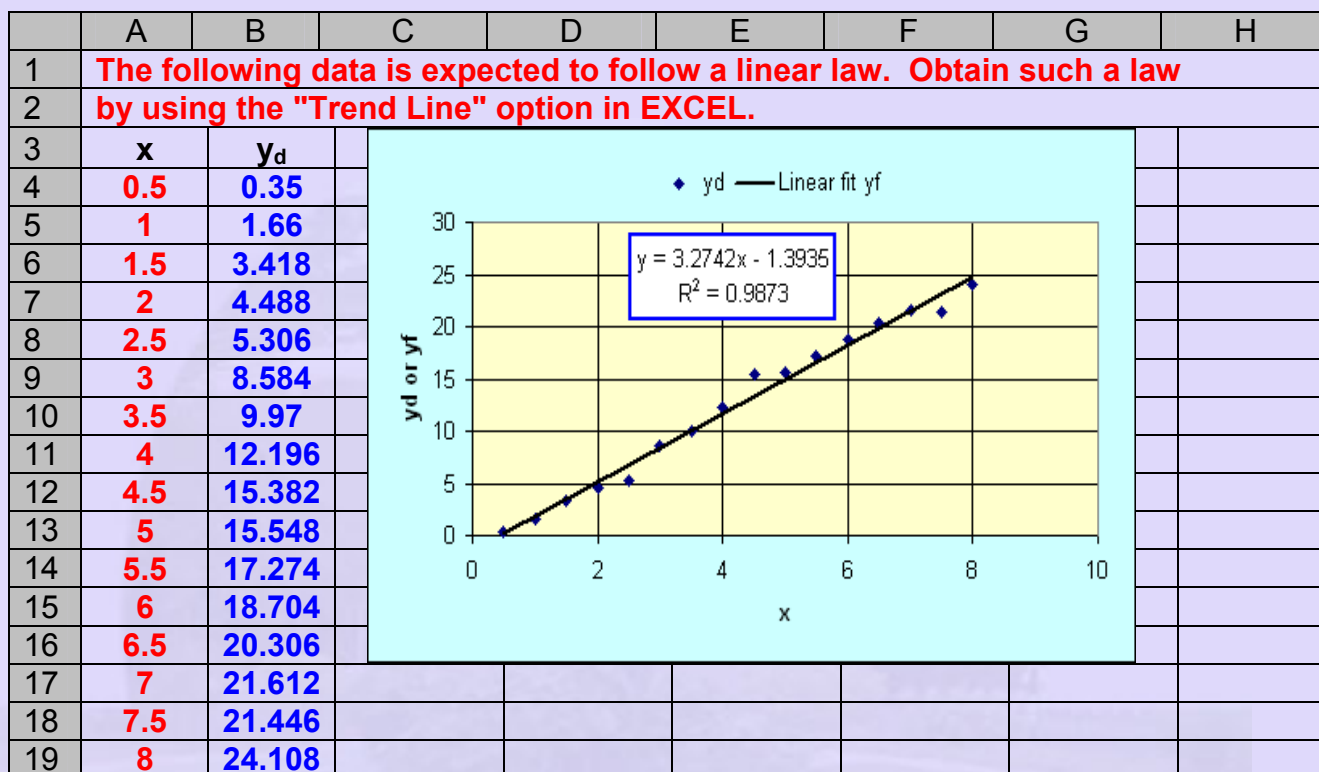The plot inset shows: $y = 3.2742x - 1.3935$, $R^2 = 0.9873$

Figure 13 shows how a trend line can be added to the plot. The inset in the plot shows the relationship that exists between the y and x data values. Correlation coefficient is very high indicating the fit to represent the data extremely well.

Several examples of regression using EXCEL are presented below. The examples are self explanatory and I expect the student to work them out using EXCEL himself/herself.