CHAPTER 9

ADD-INS: ENHANCING EXCEL

This chapter discusses the following topics:

— WHAT CAN AN ADD-IN DO?

— WHY USE AN ADD-IN (AND NOT JUST EXCEL MACROS/PROGRAMS)?

— ADD–INS INSTALLED WITH EXCEL

— OTHER ADD-INS

— THE STATISTICS ADD-IN

— CHOOSING THE ADD-INS

9.1 ADD-INS: INTRODUCTION

An "Add-In" is a software application that adds new functionality to Excel. The Add-In typically seamlessly fits into the Excel interface, providing accessibility to its functionality through

— new menus

— new options in existing menus

— new functions

— new toolbars and specific toolbar icons

9.1.A WHAT CAN AN ADD-IN DO?

Almost anything an imaginative software developer could create. Usually, an Add-In provides functionality that is useful for a particular type of analysis/industry — statistics, finance, real estate, etc.

9.1.B WHY USE AN ADD-IN?

The Add-In could have its base code written in software languages like C, C++, FORTRAN, Pascal, etc. This is important because some algorithms and operations (like simulations) operate best when written in a specific language. Therefore, the developer uses the best language/tool to create the functionality and then packages this inside an Add-In.

9.2 ADD–INS INSTALLED WITH EXCEL

Some Add–Ins are available in the Microsoft Office CD–ROM and are installed (but not activated[10]) along with Excel. I show the use of three Add–ins.

[10] Figure 540 and Figure 542 show how to activate the Add-ins

9.3 OTHER ADD-INS

Many commercially sold Add-Ins can be almost like separate software just needing Excel as the "host." Two examples:

— Crystal Ball™ risk analysis software

— *UNISTAT*™ software for conducting advanced statistics and econometrics from inside Excel

Hundreds of software companies construct Add-Ins. The greatest contribution of this book, if I succeed in doing so, would be the opening of this massive potential functionality to Excel users.

9.4 THE STATISTICS ADD-IN

The Analysis ToolPak Add-In that ships with Excel can conduct several procedures including descriptives, regression, ANOVA, F-test, correlation, T-tests, moving average, and histogram. Let us learn how to use this "Add-In."

9.4.A CHOOSING THE ADD-INS

Choose the menu option TOOLS/ADD-INS. You will see several Add-Ins as shown in Figure 140. (You may not see all the Add–Ins shown in the next two figures.)

Figure 140: Selecting an Add-In

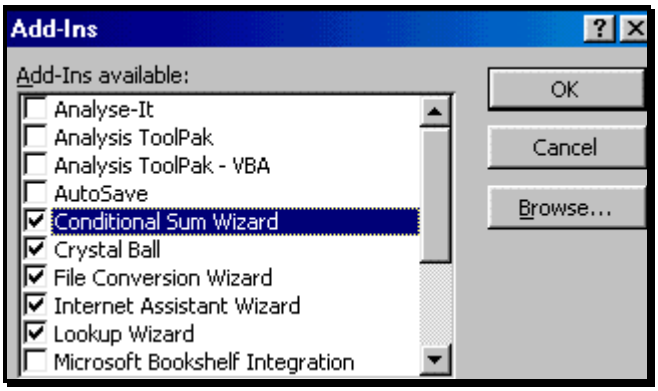
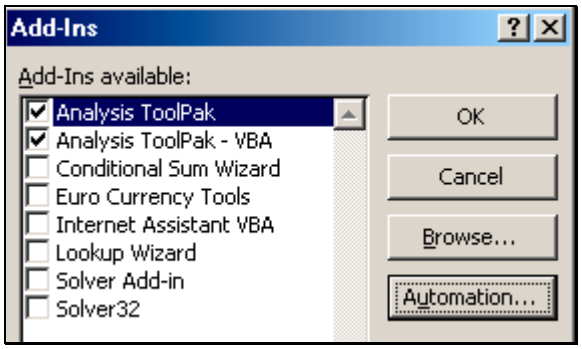


Figure 141: In Excel XP, the Add-Ins dialog provides access to "Automation." This topic is beyond the scope of this book.



You need the "Analysis ToolPak Add-Ins." Select — by clicking on it — the box to the left of these Add-Ins (shown in Figure 142). Execute the dialog by clicking on the button OK and wait for some time while the Add-Ins are "loaded" or "registered" with Excel. An Add-In has to be loaded/registered before it is available for use. The Add-In remains loaded across sessions. It is only "unloaded" when you select the option

TOOLS/ADD-INS and deselect the Add-In[11].

Figure 142: The Add-In pair for data analysis

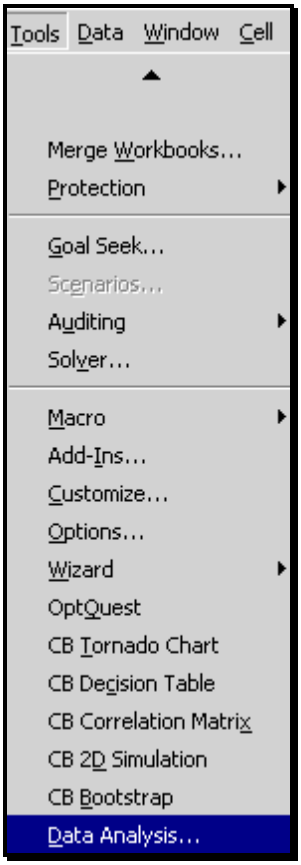
You have activated the "Analysis ToolPak." At the bottom of the menu TOOLS, you will see the option "DATA ANALYSIS the bottom— this option was not there before you accessed the Add-In. (This is illustrated in Figure 143.)

The statistical procedures are accessed through this new option.

> Note:
> Usually Add-Ins expose their functionality by creating new menu options or even new menus. The menu option "Data analysis" provides the statistics functionality available in "Analysis ToolPak" and "Analysis ToolPak VB." The menu options "Optquest" down till CB Bootstrap" are linked to the Add-in "Crystal Ball" (not shipped in the Office CD-ROM).

[11] If too many Add-Ins are loaded, Excel may work too slowly, or even freeze. If you find this problem occurring, then just load the Add-in when you are going to use it and unload it before quitting Excel.

Figure 143: The "Data Analysis" menu option



Page for Notes

CHAPTER 10

STATISTICS TOOLS

This chapter discusses the following topics:

— DESCRIPTIVE STATISTICS

— RANK AND PERCENTILE

— BIVARIATE RELATIONS— CORRELATION, COVARIANCE

A proper analysis of data must begin with an analysis of the statistical attributes of each series in isolation — univariate analysis. From such an analysis, you can learn:

— How the values of a series are distributed — normal, binomial, etc.

— The central tendency of the values of a series (mean, median, and mode)

— Dispersion of the values (standard deviation, variance, range, and quartiles)

— Presence of outliers (extreme values)

The answer to these questions illuminates and motivates further, more complex, analysis. Moreover, failure to conduct univariate analysis may restrict the usefulness of further procedures (like correlation and regression). Reason: even if improper/incomplete univariate analysis may not directly hinder the conducting of more complex procedures, the interpretation of output from the latter will become difficult (because you will not have an adequate understanding of how each series behaves).

Note: I do not go into the details of each statistics procedure. For such details, refer to your statistics textbook or to "SPSS for Beginners" (available at http://www.vjbooks.net and amazon.com).

This chapter requires the Analysis ToolPak Add-Ins; chapter 9 shows how to learn how to launch the Add-Ins.

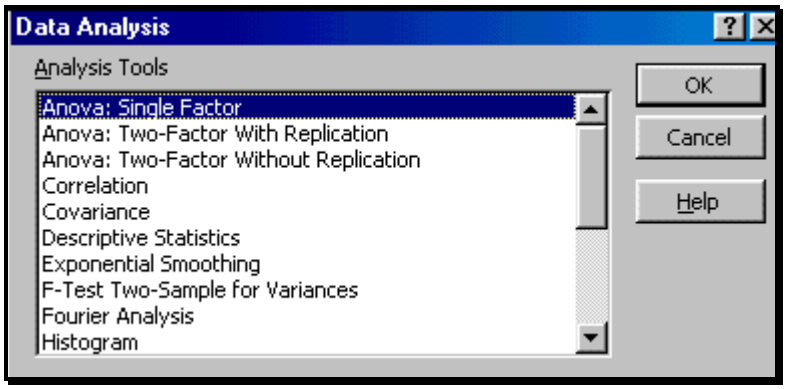
10.1 DESCRIPTIVE STATISTICS

I do not supply the sample data for most of the examples in chapters 36-40. My experience is that many readers glaze over the examples and do not go through the difficult step of drawing inferences from a result if the sample data results are the same as those in the examples in the book.

Choose the menu option TOOLS/DATA ANALYSIS[12]. The dialog shown in Figure 144 opens.

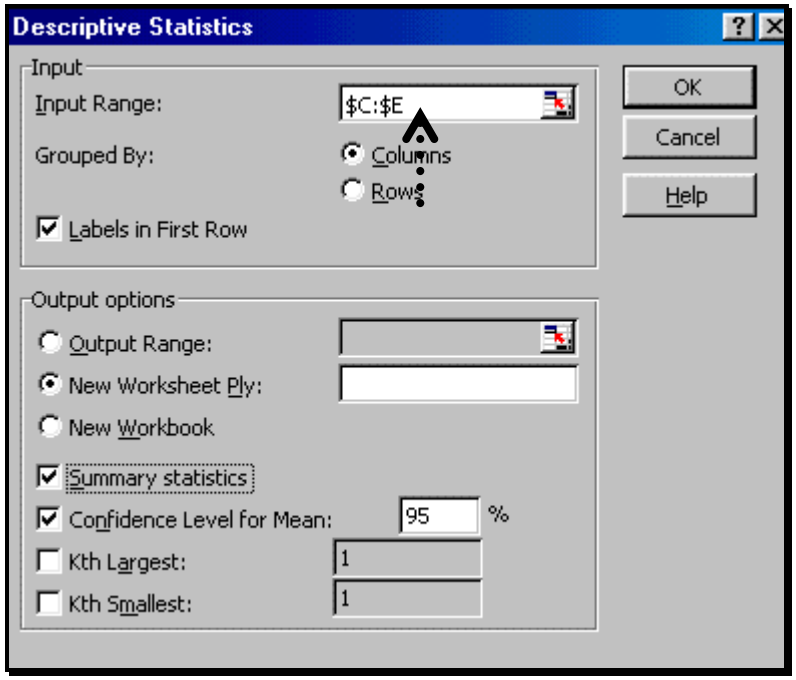
[12] If you do not see this option, then use TOOLS / ADD-INS to activate the Add-In for data analysis. Refer to section 41.4.

Figure 144: The options for the menu TOOLS/DATA ANALYSIS



Choose the statistical procedure "Descriptive Statistics." The dialog for "Descriptive Statistics" opens. Figure 145 shows this dialog (user-input form).

Figure 145: Descriptive Statistics dialog



Input (or, "Source") data

Choose the data series whose descriptives you desire. Click on the edge of the box next to "Input Range" (at the point where the dotted arrow points in Figure 145).

Options

Choose other options shown in Figure 145. Select the option "Labels in first row" because the names of the three series are in the first row of the range you selected (the labels are in cells C1, D1, and E1)— this way Excel picks up the names of the variables and uses these names in the output[13]. Execute the dialog by clicking on the button OK.

Output

Excel produces the descriptive statistics and places the results in a new worksheet. (This is illustrated in Figure 146.)

[13] Note that in the output of this procedure (shown in Figure 546) the first row has the labels for the three variables— 1995, 2000, and 2010.

Figure 146: Output of Descriptive Statistics procedure

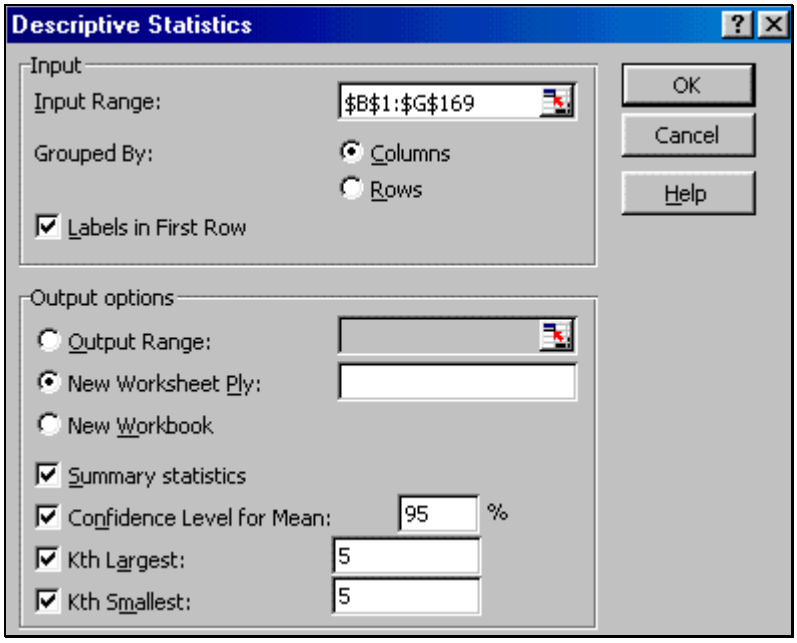
A	B	C	D	E	F
1995		*2000*		*2010*	
Mean	5257914.9	Mean	9440351.4	Mean	7406944.7
Standard Err	2665017.3	Standard Err	5047964.7	Standard Err	3748949.2
Median	492000	Median	589000	Median	799000
Mode	47000	Mode	254000	Mode	51000
Standard De	40853942	Standard De	77383834	Standard De	57470302
Sample Varia	1.669E+15	Sample Varia	5.988E+15	Sample Varia	3.303E+15
Kurtosis	218.59277	Kurtosis	179.18282	Kurtosis	219.84937
Skewness	14.560356	Skewness	13.001367	Skewness	14.619033
Range	617798000	Range	1.109E+09	Range	870291000
Minimum	7000	Minimum	12000	Minimum	25000
Maximum	617805000	Maximum	1.109E+09	Maximum	870316000
Sum	1.236E+09	Sum	2.218E+09	Sum	1.741E+09
Count	235	Count	235	Count	235
Confidence L	5250489.2	Confidence L	9945257.7	Confidence L	7385999.6

This tool generates a report of univariate statistics for data in the input range, providing information about the central tendency and variability of your data

Example 2: Adding additional parameters to the descriptives table

Go to the menu option TOOLS/DATA ANALYSIS. Select the option "Descriptive Statistics." In addition to the statistics requested in the previous example, I request Excel to report on the fifth largest and fifth smallest values for each column/series.

Figure 147: The Descriptives Statistics dialog



Output

The output for the procedure is reproduced in the next table. In one simple step, you have created a table that captures the basic statistical attributes of several data series and the fifth highest and lowest values of each data series.

Table 29: Output of the Descriptive Statistics tool including the Kth largest and smallest values. The names of the three variable are: s1, s2, and x1.

s1		s2		x1	
Mean	7.32	Mean	7.23	Mean	1173.00
Standard Error	0.44	Standard Error	0.49	Standard Error	52.67
Median	5.31	Median	4.81	Median	1173.00
Mode	1.34	Mode	23.00	Mode	#N/A
Standard Deviation	5.72	Standard Deviation	6.33	Standard Deviation	682.73
Sample	32.68	Sample	40.13	Sample	466119.22

s1		s2		x1	
Variance		Variance		Variance	
Kurtosis	–0.22	Kurtosis	0.04	Kurtosis	–1.20
Skewness	0.95	Skewness	1.06	Skewness	0.00
Range	19.66	Range	22.00	Range	2344.00
Minimum	1.34	Minimum	1	Minimum	1
Maximum	21	Maximum	23	Maximum	2345
Sum	1229.79	Sum	1215.395	Sum	197064
Count	168	Count	168	Count	168
Largest (5)	21	Largest (5)	23	Largest (5)	2288.86
Smallest (5)	1.34	Smallest (5)	1	Smallest (5)	57.14
Confidence Level (95.0%)	0.87	Confidence Level (95.0%)	0.96	Confidence Level (95.0%)	103.99

Interpretation of the statistical parameters is discussed in chapter 6, and of Confidence levels is discussed in 7.1.

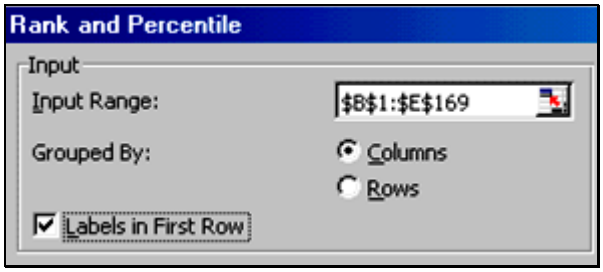
10.2 RANK AND PERCENTILE

This tool produces a table that contains the ordinal and percentage rank of each value in a data set. You can analyze the relative standing of values in a data set. The Percentile values can assist in learning about the spread of the series across its range. For a series provides information on the ranges for the lowest 25%, the next 25%, the next 25%, and the

highest 25%.

Go to[14] the menu option TOOLS/DATA ANALYSIS[15]. Select the option "Rank and Percentile." The dialog is shown in the next figure.

Figure 148: Rank and Percentile tool



The result is reproduced in the next table. Each output table contains four columns:

— The place of the data point in the data series,

— The value of the data (with the label for the series as the label on the output column),

— The rank of the data point within the range, and

[14] I do not supply the sample data for most of the examples in chapter 42 to chapter 46. My experience is that many readers glaze over the examples and do not go through the difficult step of drawing inferences from a result if the sample data results are the same as those in the examples in the book.

[15] If you do not see this option, then use TOOLS / ADD-INS to activate the Add-In for data analysis. Refer to section 41.4.

— The percentage rank of the data point. The columns are sorted in order of ascending rank.

Table 30: Output of the Rank and Percentile tool

Point	*s1*	*Rank*	*Percent*	*Point*	*s2*	*Rank*	*Percent*
24	21.00	1	96.40%	1	23.00	1	96.40%
48	21.00	1	96.40%	25	23.00	1	96.40%
72	21.00	1	96.40%	49	23.00	1	96.40%
96	21.00	1	96.40%	73	23.00	1	96.40%
120	21.00	1	96.40%	97	23.00	1	96.40%
144	21.00	1	96.40%	121	23.00	1	96.40%
168	21.00	1	96.40%	145	23.00	1	96.40%
23	18.63	8	92.20%	2	20.07	8	92.20%
47	18.63	8	92.20%	26	20.07	8	92.20%
71	18.63	8	92.20%	50	20.07	8	92.20%
95	18.63	8	92.20%	74	20.07	8	92.20%
119	18.63	8	92.20%	98	20.07	8	92.20%
143	18.63	8	92.20%	122	20.07	8	92.20%
167	18.63	8	92.20%	146	20.07	8	92.20%
22	16.53	15	88.00%	3	17.51	15	88.00%
46	16.53	15	88.00%	27	17.51	15	88.00%
70	16.53	15	88.00%	51	17.51	15	88.00%
94	16.53	15	88.00%	75	17.51	15	88.00%
118	16.53	15	88.00%	99	17.51	15	88.00%
142	16.53	15	88.00%	123	17.51	15	88.00%
166	16.53	15	88.00%	147	17.51	15	88.00%

Interpreting the output:

The last row's last four columns can be interpreted as—

The 147th data point in the selected range has a value of 17.51, which gives it rank 15 in the selected range, with 88% of the cells in the range having a value less than or equal to this data point.

10.3 BIVARIATE RELATIONS– CORRELATION, COVARIANCE

Correlation analysis

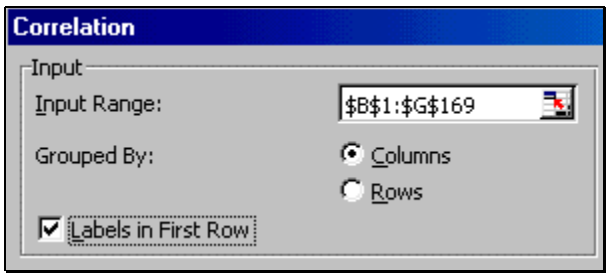
This tool and its formulas measure the relationship between two data sets that are scaled to be independent of the unit of measurement. The correlation coefficient depicts the basic relationship across two variables: "Do two variables have a tendency to increase together or to change in opposite directions and, if so, by how much?" Bivariate correlations measure the correlation coefficients between two variables at a time, ignoring the effect of all other variables.

Go to the menu option TOOLS/DATA ANALYSIS[16]. Select the option "Correlation."

Select the "Input Range" — it must have more than one data series.

[16] If you do not see this option, then use TOOLS / ADD-INS to activate the Add-In for data analysis. Refer to section 41.4.

Figure 149: CORRELATION



The output is reproduced in the next table.

Table 31: Output from Correlation Analysis tool

	s1	*s2*	*x1*	*x2*	*x3*	*x4*
s1	1.00000					
s2	–0.75973	1.00000				
x1	–0.13434	0.13226	1.00000			
x2	0.21423	0.47238	0.01658	1.00000		
x3	0.20122	–0.08459	–0.15748	0.14568	1.00000	
x4	–0.13567	0.12935	0.99998	0.01040	–0.15839	1.00000

Interpreting the output

— A high level of correlation is implied by a correlation coefficient that is greater than 0.5 in absolute terms (that is, greater than 0.5 or less than –0.5).

— A mid level of correlation is implied if the absolute value of the coefficient is greater than 0.2 but less that 0.5.

— A low level of correlation is implied if the absolute value of the coefficient is less than 0.2.