

CHAPTER 1 INTRODUCTION

1.1 Statistics' Computational Intensity

Statistics is a very intense computational discipline. We can get a sense of the computational intensity of statistics by looking at the average statistical formula and the correlation coefficient statistical formula shown below.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Average

$$r = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\left[\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2} \left[\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \right]^{1/2}}$$

Correlation Coefficient

Both of these formulas are relatively complicated and could give the statistician considerable computational trouble. The average formula looks fairly simple, but it can become difficult when n increases. The correlation coefficient formula is obviously more complex, and tedious to perform -- and other computational formulas are even more complicated. Many statistical formulas can take days or weeks by hand or even with a calculator. In the 'old days' before statistical software and computers, statisticians spent a great deal of time doing tedious, repetitive calculations and then tracking down the inevitable errors.

Now consider a situation in which the statistician must calculate the correlation coefficient more than once. Doing it even once by hand is daunting, and having to do it more than once can be overwhelming.

Statistics, of course, involves more than numbers alone. The presentation of information by means of graphs and pictures is an essential part of the discipline. Few people, however, have the time and skill to draw accurate, good-looking graphs.

Fortunately with the advent of personal computers, the availability of computer software has made statistical analysis easier. The computer software can do the following,

1. Do statistical calculations very quickly.
2. Do statistical calculations without mistakes.
3. Do repetitive calculations quickly.
4. Generate statistical graphs quickly and accurately.

In this book we will look at two software packages to do statistical analyses. The two software packages are MINITAB STATISTICAL SOFTWARE Release 12 and Microsoft Excel 97. MINITAB is a software package that is written specifically to do statistical analyses. Microsoft Excel 97 (EXCEL) is a powerful business application that has strong statistical features.

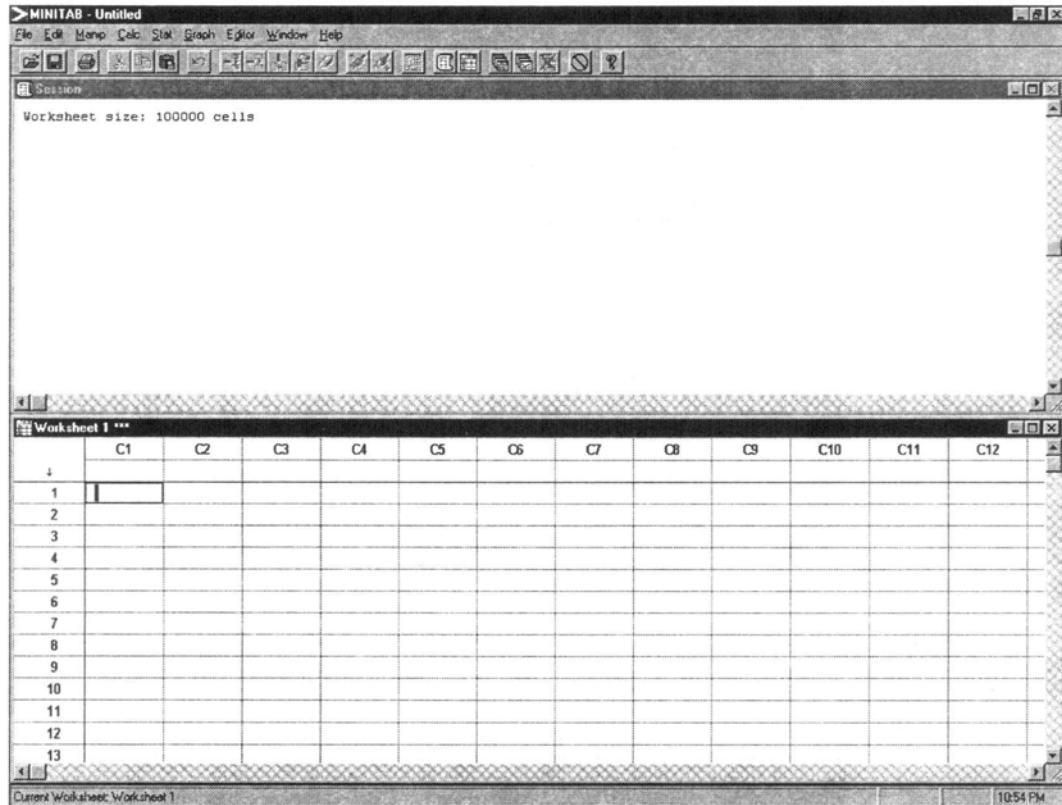
MINITAB's strength is that it is written to do specifically statistical analysis. This is why it is very easy to create a box plot in MINITAB. EXCEL's strength is that it has a very strong programming language called Visual Basic for Applications (VBA). To create a box plot in EXCEL, we would have to write a program using VBA to create a box plot. Even though doing this would be more difficult than using MINITAB, the reward would be worth the effort. We can transfer our knowledge in creating a box plot in EXCEL to other non-statistical EXCEL projects.

An important reason in using EXCEL is it has become a "standard" program in the business world. This makes it very important to study EXCEL.

1.2 Environment of MINITAB

MINITAB consists of two main parts, the *data window* and the *session window*.

The data window is where the statistical data are entered. The session window is where all statistical analysis are shown and where statistical commands called session commands can be issued. When you start MINITAB, the data window and session window are displayed by default. This is shown below.

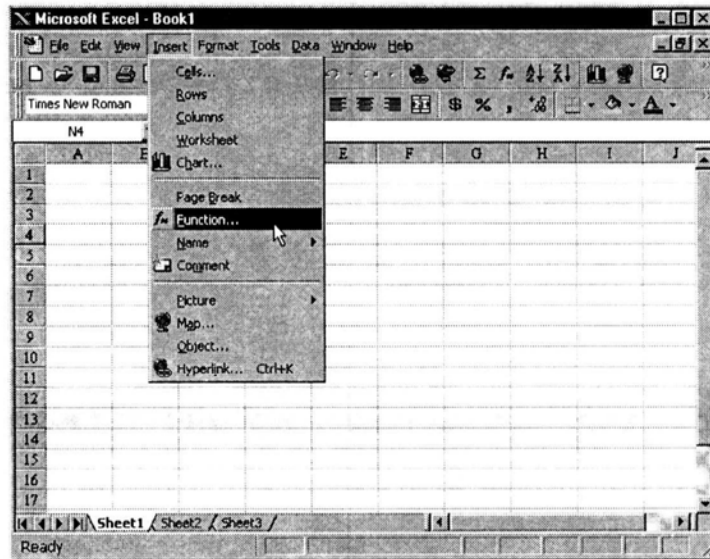


The session window is where statistical commands and reports are displayed. Many statistical calculations require several commands. With MINITAB you can group commands together and give a name to the group of commands. Such a group is called a *macro*. The user can execute the group of commands just by typing the group name. Grouping commands is computer programming. The capability to program is one of MINITAB's strengths.

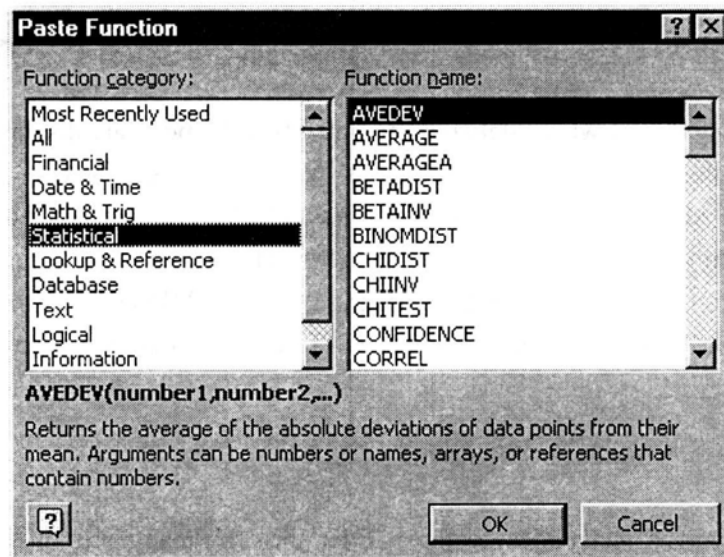
The basic procedures of using MINITAB will be discussed in the next two chapters.

1.3 Statistical Environment of Microsoft Excel 97

EXCEL has many built in statistical features. First, there are many built in statistical functions. To view the statistical functions, go to the *Insert* menu and choose the *Function* menu item as shown below. Going forward we will use the following notation to get to menu items, **Menu → Menu item**. To get to the function menu we will then indicate **Insert → Function**.



The Function menu item will give you the following dialog box to show you all the available built-in functions in EXCEL and to aid you in using the built-in functions in EXCEL.



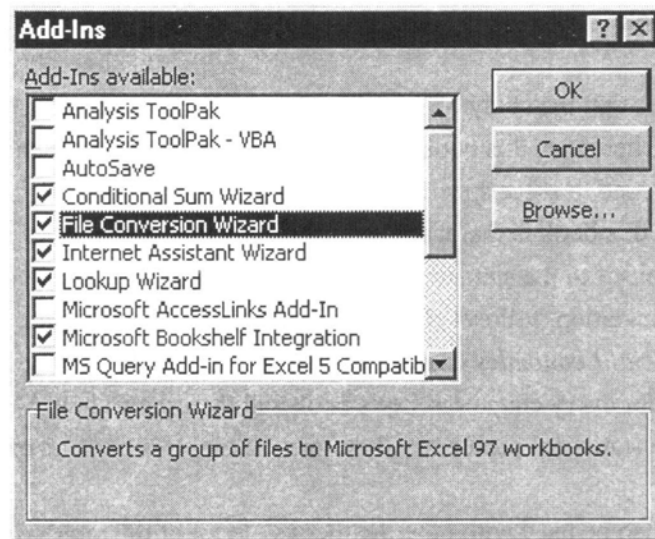
Below is a table showing all the available statistical functions in the above dialog box.

Statistical Functions

AVEDEV	Returns the average of the absolute deviations of data points from their mean
AVERAGE	Returns the average of its arguments
AVERAGEA	Returns the average of its arguments, including numbers, text, and logical values
BETADIST	Returns the cumulative beta probability density function
BETAINV	Returns the inverse of the cumulative beta probability density function
BINOMDIST	Returns the individual term binomial distribution probability
CHIDIST	Returns the one-tailed probability of the chi-squared distribution
CHIINV	Returns the inverse of the one-tailed probability of the chi-squared distribution
CHITEST	Returns the test for independence
CONFIDENCE	Returns the confidence interval for a population mean
CORREL	Returns the correlation coefficient between two data sets
COUNT	Counts how many numbers are in the list of arguments
COUNTA	Counts how many values are in the list of arguments
COVAR	Returns covariance, the average of the products of paired deviations
CRITBINOM	Returns the smallest value for which the cumulative binomial distribution is less than or equal to a criterion value
DEVSQ	Returns the sum of squares of deviations
EXPONDIST	Returns the exponential distribution
FDIST	Returns the F probability distribution
FINV	Returns the inverse of the F probability distribution
FISHER	Returns the Fisher transformation
FISHERINV	Returns the inverse of the Fisher transformation
FORECAST	Returns a value along a linear trend
FREQUENCY	Returns a frequency distribution as a vertical array
FTEST	Returns the result of an F-test
GAMMADIST	Returns the gamma distribution
GAMMAINV	Returns the inverse of the gamma cumulative distribution
GAMMALN	Returns the natural logarithm of the gamma function, $\Gamma(x)$
GEOMEAN	Returns the geometric mean
GROWTH	Returns values along an exponential trend
HARMEAN	Returns the harmonic mean
HYPGEOMDIST	Returns the hypergeometric distribution
INTERCEPT	Returns the intercept of the linear regression line
KURT	Returns the kurtosis of a data set
LARGE	Returns the k-th largest value in a data set
LINEST	Returns the parameters of a linear trend
LOGEST	Returns the parameters of an exponential trend
LOGINV	Returns the inverse of the lognormal distribution
LOGNORMDIST	Returns the cumulative lognormal distribution
MAX	Returns the maximum value in a list of arguments
MAXA	Returns the maximum value in a list of arguments, including numbers, text, and logical values
MEDIAN	Returns the median of the given numbers
MIN	Returns the minimum value in a list of arguments
MINA	Returns the smallest value in a list of arguments, including numbers, text, and logical values
MODE	Returns the most common value in a data set
NEGBINOMDIST	Returns the negative binomial distribution
NORMDIST	Returns the normal cumulative distribution
NORMINV	Returns the inverse of the normal cumulative distribution

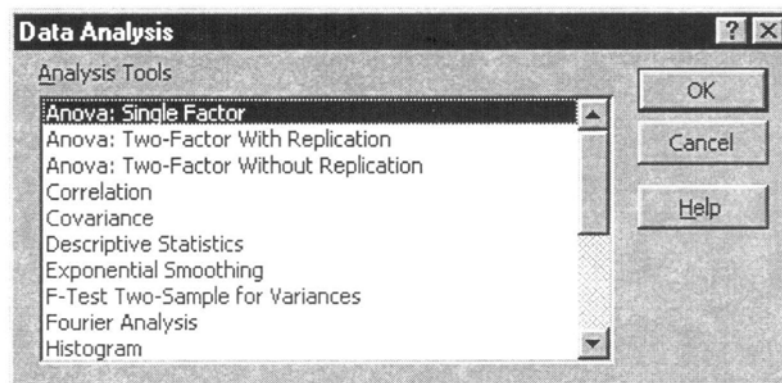
NORMSDIST	Returns the standard normal cumulative distribution
NORMSINV	Returns the inverse of the standard normal cumulative distribution
PEARSON	Returns the Pearson product moment correlation coefficient
PERCENTILE	Returns the k-th percentile of values in a range
PERCENTRANK	Returns the percentage rank of a value in a data set
PERMUT	Returns the number of permutations for a given number of objects
POISSON	Returns the Poisson distribution
PROB	Returns the probability that values in a range are between two limits
QUARTILE	Returns the quartile of a data set
RANK	Returns the rank of a number in a list of numbers
RSQ	Returns the square of the Pearson product moment correlation coefficient
SKEW	Returns the skewness of a distribution
SLOPE	Returns the slope of the linear regression line
SMALL	Returns the k-th smallest value in a data set
STANDARDIZE	Returns a normalized value
STDEV	Estimates standard deviation based on a sample
STDEVA	Estimates standard deviation based on a sample, including numbers, text, and logical values
STDEVP	Calculates standard deviation based on the entire population
STDEVPA	Calculates standard deviation based on the entire population, including numbers, text, and logical values
STEYX	Returns the standard error of the predicted y-value for each x in the regression
TDIST	Returns the Student's t-distribution
TINV	Returns the inverse of the Student's t-distribution
TREND	Returns values along a linear trend
TRIMMEAN	Returns the mean of the interior of a data set
TTEST	Returns the probability associated with a Student's t-test
VAR	Estimates variance based on a sample
VARA	Estimates variance based on a sample, including numbers, text, and logical values
VARP	Calculates variance based on the entire population
VARPA	Calculates variance based on the entire population, including numbers, text, and logical values
WEIBULL	Returns the Weibull distribution
ZTEST	Returns the two-tailed P-value of a z-test

In EXCEL it is possible to extend the functionality of EXCEL. This is done through computer programs called *add-ins*. Excel comes with a statistical add-in. If installed properly you would run the add-in by choosing **Tools** → **Data Analysis**. If it is not there, you would need to add the add-in by choosing **Tools** → **Add-in**. This would give you the following dialog box.



Check the second option, Analysis ToolPak - VBA , to install the statistical add-in.

After doing this you will see a Data Analysis menu item under the Tools menu. The Data Analysis menu item will give you the following dialog box,



The Data Analysis dialog box shows all the statistical analysis available in EXCEL. To do a specific analysis of interest, highlight the statistical analysis of interest and then click the *OK* button.

1.4 Using This Book

This book recognizes that the study of MINITAB and EXCEL cannot be separated from the study of statistics. Therefore, this book first looks at statistics and then shows how MINITAB and EXCEL can be used to solve specific statistical problems.

One strength of this book is that it analyzes the computer output created by MINITAB and EXCEL and then links it to the statistical concepts discussed in the chapter.

The book's organization follows Cheng F. Lee, John Lee, and Alice Lee's *Statistics for Business and Financial Economics* chapter by chapter. It quickly explores and reviews the statistical concepts for every chapter in Lee's book and then shows how MINITAB and EXCEL can be used to solve statistical problems, therefore making this book an excellent companion to Lee's book.

This book can also be used with other textbooks. Most of the topics discussed in other statistics texts are also discussed here. The only difference will be the organization.

What you might discover after studying MINITAB and EXCEL is a greater appreciation for statistics and that statistics is an exciting discipline. MINITAB and EXCEL allows the student of statistics to concentrate more on conceptual understanding, data analysis, real-world applications and less on statistical calculations.