chapter 1-3

Processing Data using MS EXCEL

Introduction

The purpose of this chapter is to highlight those MS EXCEL skills that will be used in the data manipulations used in this text. Those skills include: data entry, entering formulae, statistical functions, statistical analysis using data analysis package, and the Solver function.

In writing this chapter it was assumed that the reader has a basic working knowledge of spreadsheeting with EXCEL. If you require very basic information on how to use EXCEL you can either use the help menu or tutorials provided within EXCEL itself, or there is a very useful website provided by the University of Texas where various Information Technology informational documents can be found, including one on the basics of EXCEL. The EXCEL pdf file can be found at:

http://www.utexas.edu/its/training/handouts/UTOPIA_ExcelGS/

Data Entry

Because of its versatility MS EXCEL is often used for data entry, even though analysis of the data will be carried out with more sophisticated statistical software such as SPSS. It is assumed that you are familiar with entering data into cells and the associated formatting of those cells.

Entering Data

You can enter text, numbers and dates in an Excel worksheet. To enter data, click on the cell to select it and then start typing. The data you type will also appear in the formula bar. When you have finished typing the data for the active cell, the data will only be entered into that cell when you press the Return or Enter key, or click or move away using the arrow keys to another cell.

Adjusting Column Widths

Sometimes the entries in a column are to long for the column width or conversely, the column width is much too long for the length of the entries. In order to correct this you can change column widths by dragging column borders with the mouse. Move the mouse over the right border of a column heading until the mouse icon changes to left and right pointing arrows. When the mouse icon changes shape, click and drag the mouse to left or right, to adjust the column width. You can automatically change the column widths of all columns to optimal width by first selecting the whole sheet by clicking on the top left corner cell which lines up with the letters and numbers (not cell A1). Move the mouse to the right border of any column heading until the mouse icon changes shape to the left and right pointing arrows. When the mouse icon changes double click and the width of all cells will change to be just greater than the largest entry in each column respectively.

As a point of interest, your worksheet is limited to a maximum size of 65,536 rows and 256 columns, which is certainly adequate for most purposes.

Saving Data for Use in Other Applications

A nice feature is that you can save EXCEL files into file formats for many other applications. Although EXCEL will be used for some of the analysis in this text, most of it will be carried out using SPSS statistical software. Although SPSS has the facility for direct entry into its data editor, often researchers prefer to use EXCEL for their initial data entry and clean-up. EXCEL will not save data in SPSS format, but it will save it in various file formats that can be read by SPSS. Note that SPSS will not open modern versions of EXCEL files because of the multiple worksheet workbook format. Files saved in EXCEL 4.x or earlier format can be read using SPSS as early versions of EXCEL only had one worksheet per file.

Another alternative file format that works well is the SYLK (Symbolic Link) format using the extension *.slk. All EXCEL formatting and variable names are saved in SYLK files and its use is preferable over earlier versions of EXCEL because then you are not trying to keep track of files saved in various versions of EXCEL.

Entering Formulae

In data manipulations carried out for this text, fairly complex equations such as the one shown below will be entered into cells in EXCEL.

$$h_t = h_1 - \frac{2(h_1 - h_q)}{e^{[s_0(t-q)]} + e^{[s_1(t-q)]}}$$

The basic mathematical operators used in EXCEL formulae are **+ - / ***, representing addition, subtraction, division and multiplication, respectively. In addition to these there are mathematical functions such as $e^{(n)}$ which is represented by the function EXP(n).

Some basic rules of translating the formula in traditional form to computational form will be reviewed here. Figure 1-3.1 shows part of a simple worksheet with data on three subjects. To enter a formula into a cell, start with the **=** sign. After the **=** sign, type the formula. This will be composed of cell addresses and mathematical operators. For example, to add the values of age for the three subjects the formula entered into B10 would be **=**B6+B7+B8. When you have entered the equation and pressed enter, the result will be displayed in B10, rather than the equation as depicted in the figure. Cell E6 contains the formula for the calculation of Body Mass Index (BMI).

In standard mathematical notation the equation is: $BMI = \frac{Weight}{Height^2}$.

Note that in the equation in cell E6, the ^ sign is used to raise D6 (Height) to the power 2.

	A	B	C	D	E
4					
5	**Subject #**	**Age**	**Weight**	**Height**	**BMI**
6	**141**	23.3	72.3	176.6	**=C6/(D6^2)**
7	**142**	26.5	67.8	172.7	**=C7/(D7^2)**
8	**143**	22.8	86.9	184.9	**=C8/(D8^2)**
9					
10	**Totals**	**=B6+B7+B8**			

Figure 1-3.1: Equation entry in an EXCEL spreadsheet

Copying Formulas

If you now wanted the equation for BMI to be entered into cell E7, you could type it in as = C7/(D7^2). But it would be frustrating if you had to type in the equation every time you wanted it in another cell. Imagine the pain if you had 1000 subjects rather than the 3 here. Fortunately, formulae can be quickly copied to other cells. In our example, if you wanted to copy the BMI equation from E6 to E7 and E8 there are 3 ways to do it. One way you must first select the region E6:E8. Once selected go to the EDIT menu and select the FILL option. This will present you with several options, one of which is DOWN. If you select DOWN you will see the equation fill down into the other two cells. When you are in the EDIT menu you will notice to the right of the word DOWN is Ctrl+D. This is the hot key combination for fill down, meaning that if instead of going to the EDIT menu you just press the Ctrl and D key simultaneously, the formula will be copied down in the same fashion. The easiest way however, to copy down the equation is with the Fill Handle in the lower right corner of the cell. When the cell with the formula in is selected you will see a small square bullet placed over the bottom right hand corner of the selection box. This is the Fill Handle. If you move the mouse over the Fill Handle the mouse icon changes shape to a cross-hair. If you now press and drag over the adjacent cells you want to copy the formula to, the required copying will occur. Sometimes, as illustrated in Figure 1-3.2, you have many subjects or cases represented by many rows. In this example there are 1541 subjects in rows 2 to 1542. Copying down the equation for BMI entered into cell Z2 could be achieved by dragging the Fill Handle down and scrolling down to cell Z1542, but this is tedious with thousands of rows. A quicker way is to select cell Z2 then hold your shift key down while you slide the scroll bar down to cell Z1542. Once there, click on cell Z1542 and the whole region Z2:Z1542 will be selected, then you can press Ctrl+D. However, this still involves a lot of scrolling. This can be avoided by making use of the split window option. Vertically or horizontally you can divide a window into two parts each with its own independent scroll bar, as shown in Figure 1-3.2. Now the selection of large numbers of cells is easy. Using the SHIFT CLICK method, select Z2 press down the SHIFT key and then click on Z1542 which is in view in the lower window. The range Z2 to Z1542 is now selected and you can press Ctrl+D to fill down the formula. Note that you are not limited to filling down but can go up, right and left as well.

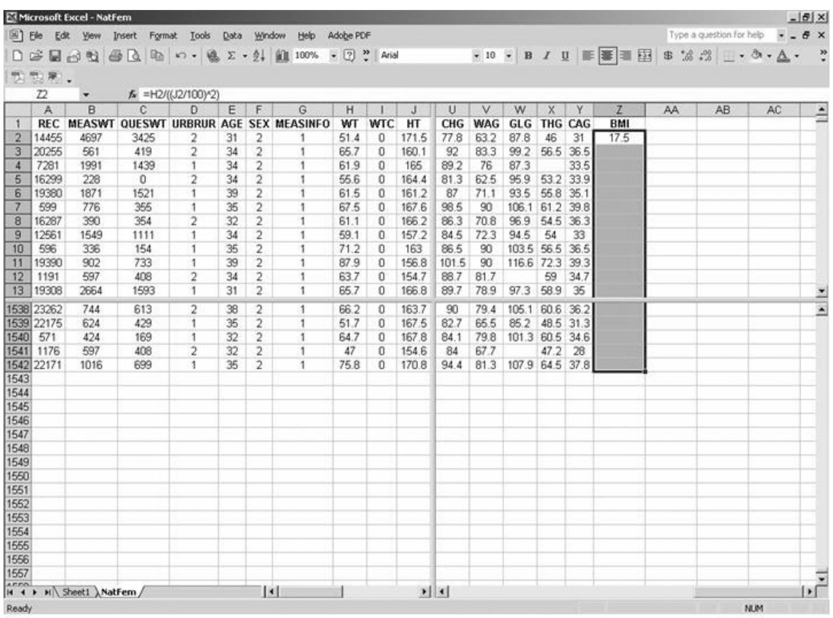


Figure 1-3.2: EXCEL worksheet showing split windows for easy fill down of equations

Relative and Absolute Referencing

When entering a formula it is important to remember that by default so-called relative referencing will be used. The BMI formula shown in Figure 1-3.1, uses relative referencing. Rather than the equation in E6 referring specifically to cells C6 and D6, it can be thought of as referring to the cell two to the left and immediately to the left, respectively. This is very convenient for the copying down of formulae as the cell addresses increment accordingly. However, there are times when you do not want the cell address to increment, but stay refereeing to a specific cell. This can be achieved by changing the referencing to absolute referencing. Figure 1-3.3 shows a slight change to the BMI calculation worksheet, where now the exponent of height is entered into a cell and referred to in the BMI calculation. Note that the cell address for the exponent 2 is not E5 but E5. It is the presence of the $ signs that designates the absolute referencing. Thus, now when the equation in cell E6 is copied down the weight and height cells will increment but the exponent cell will be fixed at E5 by the absolute referencing. This is also useful if you wanted to try different values of the exponent. If you changed the numerical value in E5 the calculations would update reflecting the change in the exponent. We will make use of this capability in several examples later in the text.

	A	B	C	D	E	E
4						**Exponent**
5	**Subject #**	**Age**	**Weight**	**Height**	**BMI**	**2**
6	**141**	23.3	72.3	176.6	**=C6/(D6^E5)**	
7	**142**	26.5	67.8	172.7	**=C7/(D7^E5)**	
8	**143**	22.8	86.9	184.9	**=C8/(D8^E5)**	
9						
10	**Totals**	**=B6+B7+B8**				

Figure 1-3.3: EXCEL worksheet illustrating relative and absolute referencing in the calculation of BMI

Copying and Moving Data

You can copy selected data from one cell to another with the Copy and Paste commands, or with the Drag and Drop procedure. However, if you want to move data rather than copy it, use the Cut and Paste commands, or the Drag and Drop procedure. Since we will be dealing extensively with equations in EXCEL a very useful facility is the PASTE SPECIAL feature. Figure 1-3.4 shows the dialog box for PASTE SPECIAL. Once you have copied a cell or cells containing equations you could PASTE the equations into other cells on the same or different worksheets. If you have relatively referenced cells in the equation then the answers in your new cells will be different than the original because you are now referring to different cells. This may well be what you want if you have set up the new part of the sheet accordingly. However, sometimes you want to take the answers to the calculations and save them in another part of the workbook, while you recalculate the equations with new parameters. If this is the case this can be achieved using the PASTE Special feature under the EDIT menu. Once you have copied the cell or cells with the formulae in, then move to the new location and select PASTE SPECIAL. Up will come a dialog box. There are many options available but the one we want is VALUES. Check values, and when you click OK, the results of the calculations will be pasted into the new cells rather than the formulae. You know have your answers relocated, but bear in mind you have lost any reference to the original equations. Again this is a feature we will find useful in examples later in the text.

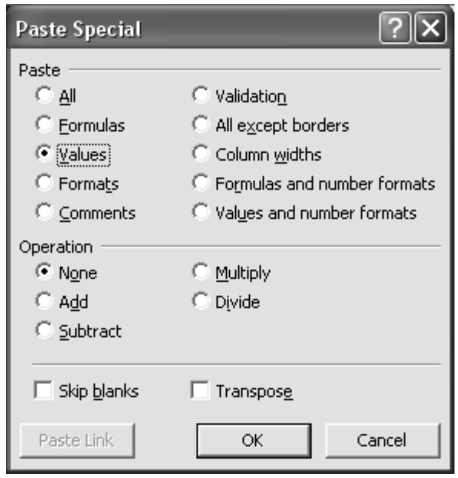


Figure 1-3.4: PASTE SPECIAL dialog box

Statistical Functions

Conveniently EXCEL provides many prewritten functions that save you the trouble of having to insert all the equations necessary for some common statistical procedures. As you saw above, formulas can be very useful on a worksheet. However, what if you want to add up a column of 10 numbers? Do you have to click on ten cells or type ten cell references in a formula? Excel has a more efficient means of dealing with this situation by using functions. The SUM( ) function is probably the most common function in Excel. It adds a range of numbers. To build a SUM( ) function, begin by typing **=SUM.** Next, tell Excel which cells to sum. Using the mouse, press and drag over the range of cells you wish to add. A dotted outline appears around the cells, and Excel displays the cell range in the formula bar. When you have the correct cells selected, release the mouse button, close the parenthesis, and press the Return key. If you do not want to use the mouse, type in the cells you want Excel to sum. For example, to sum cells B6 through B8, type **=SUM(B6:B8).** Excel interprets B6:B8 as the range of cells from B6 to B8.

Excel has many more functions besides the SUM() function described above. For example, you might want to calculate the average of a column of numbers, or count how many entries are in a row, or calculate a standard deviation (Figure 1-3.5).

	A	B	C	D	E
4					
5	**Subject #**	**Age**	**Weight**	**Height**	**BMI**
6	**141**	23.3	72.3	176.6	**=C6/(D6^2)**
7	**142**	26.5	67.8	172.7	
8	**143**	22.8	86.9	184.9	
9					
10	**Totals**	**=SUM(B6:B8)**			
	Mean	**=average(B6:B8)**			
	S.D.	**=STDEV(B6:B8)**			

Figure 1-3.5: EXCEL worksheet illustrating inclusion of statistical functions

Select FUNCTION from the INSERT menu and you will bring up the dialog box shown in Figure 1-3.6. You have a choice of hundreds of functions in categories such as Statistical, Math & Trig, Financial etc. When selected, a line of explanation will be presented about the function, and when OK is clicked, the function will be entered into the cell. The function can now be completed by entering the cell addresses of the parameters required. This is particularly useful if you require the result of one of these functions for direct inclusion in further calculations.

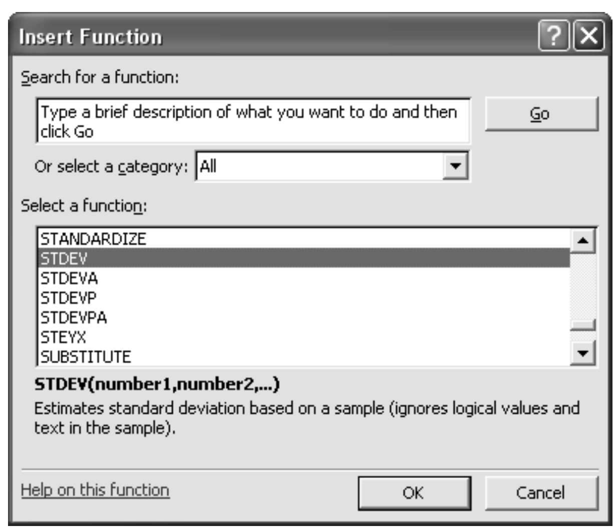


Figure 1-3.6: INSERT FUNCTION dialog box

Statistical Tests using Data Analysis

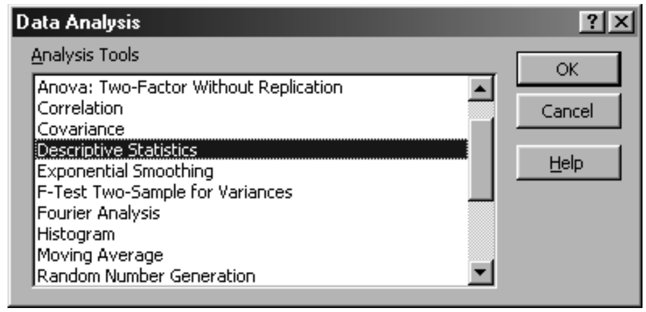


Figure 1-3.7: DATA ANALYSIS dialog box

Although the statistical functions in EXCEL are useful, it is more convenient if all of the results of a statistical analysis are reported in a table. EXCEL has this facility but with a limited statistical package available. EXCEL has many statistical tests that can be carried out. It should be pointed out that EXCEL is not intended to be a sophisticated statistical analysis package. It is a spreadsheet program that offers some statistical analysis features. We will show its facility in several chapters in this text. Sometimes EXCEL might be the only software available to you and the truth is it does provide a lot of utility in this area. Given the option we would like to have a sophisticated package like SPSS, so in many illustrations in this text we will show both applications being used for statistical analysis.

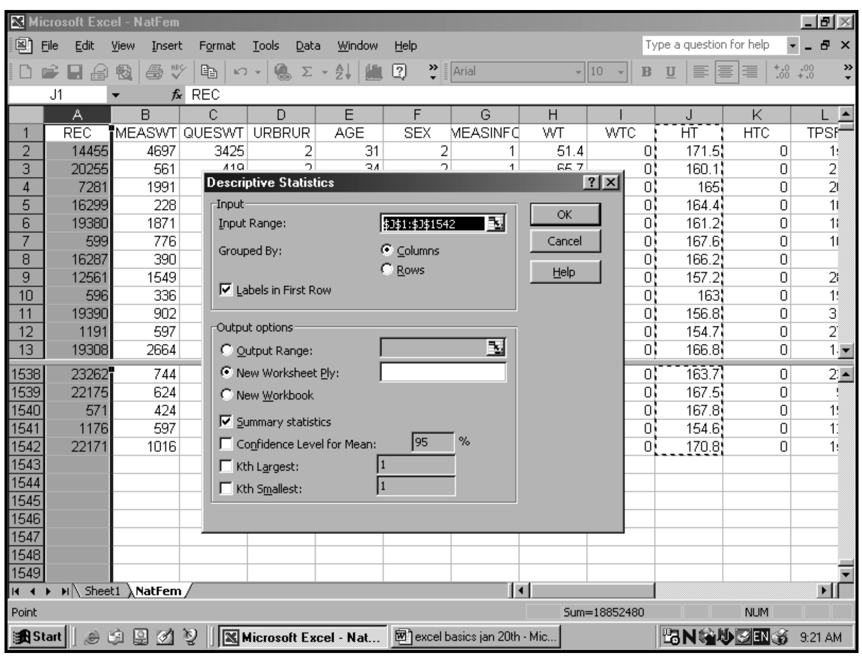
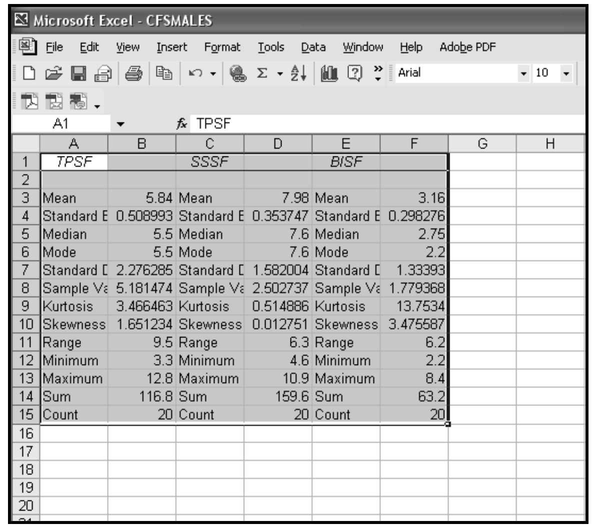


Figure 1-3.8: DESCRIPTIVE STATISTICS dialog box

EXCEL statistical analysis can be found under the TOOLS menu in the DATA ANALYSIS option. When you click on DATA ANALYSIS the dialogue box shown in Figure 1-3.7 will appear. An extensive menu of statistical procedures is available. Click on the procedure you want and you will be presented with the appropriate dialog box. Figure 1-3.8 shows the result of selecting DESCRIPTIVE STATISTICS. The dialog boxes are similar for all the procedures. IINPUT RANGE refers to the cells where the data is for analysis. Note that you must check labels in first row if you wish the output to have variable names included. OUTPUT RANGE is where you want the output table to be printed. You can select a new location on the same sheet or send it to a new worksheet. Dependent upon the particular procedure you may be offered options such as confidence levels which tend to default to 95%.



	A	B	C	D	E	F
1	TPSF		SSSF		BISF	
2						
3	Mean	5.84	Mean	7.98	Mean	3.16
4	Standard E	0.508993	Standard E	0.353747	Standard E	0.298276
5	Median	5.5	Median	7.6	Median	2.75
6	Mode	5.5	Mode	7.6	Mode	2.2
7	Standard D	2.276285	Standard D	1.582004	Standard D	1.33393
8	Sample Va	5.181474	Sample Va	2.502737	Sample Va	1.779368
9	Kurtosis	3.466463	Kurtosis	0.514886	Kurtosis	13.7534
10	Skewness	1.651234	Skewness	0.012751	Skewness	3.475587
11	Range	9.5	Range	6.3	Range	6.2
12	Minimum	3.3	Minimum	4.6	Minimum	2.2
13	Maximum	12.8	Maximum	10.9	Maximum	8.4
14	Sum	116.8	Sum	159.6	Sum	63.2
15	Count	20	Count	20	Count	20

Figure 1-3.9: DESCRIPTIVE STATISTICS dialog box

Figure 1-3.9 shows the output for DESCRIPTIVE STATISTICSS for three variables labeled TPSF, SSSF and BISF. An irritating feature of the outputs from the statistical procedures is that no optimizing of column width occurs. This means that Labels often are not completed within the cell width. This can be cured by manually changing column widths as discussed earlier. In later chapters, as specific statistical procedures are referred to, the details of EXCEL operations will be discussed.

Solver

Solver is a handy function included in EXCEL which we will make use of in two different applications in this text. Solver is one of the what-if analysis tools found in EXCEL. Its purpose is to find an optimal solution for an equation that you have placed in a cell. This cell is called the target cell. The equation in the target cell refers to one or more other cells that have either numerical entries or other equations. These are called the adjustable cells, and Solver adjusts the values in the cells you specify in the Solver set up, to produce the result you specify for the target cell. You can put constraints on the adjustable cells, in order restrict the values Solver can use in the model. Figure 1-3.10 shows application to a height modeling example discussed in chapter 2-7. In this the sum of squares in cell G7 is minimized by

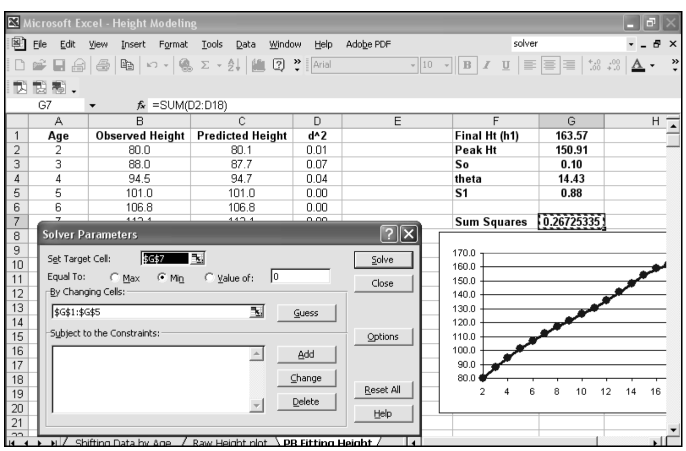


Figure 1-3.10: Solver dialog box for height modeling example

changing the values of model parameters in cells G1 to G5. We will use SOLVER in several applications in this text, and it will be discussed in more detail at that time.