Working with Data in Microsoft Excel 2003

Doc 5.94   Ver 2   *John Matthews*
March 2005   *Central Computing Services*

Abstract

This document provides some examples of handling numeric data using the *Microsoft Excel 2003* spreadsheet package. It assumes that you are familiar with basic techniques covered in CCS document 5.93. For more information on *Microsoft Excel 2003* please refer to 'Software made Simple Guide to *Microsoft Excel 2003*'. Copies of the Guide are available to students and staff of Birkbeck either in paper form from CCS Help Desk, room 151, Malet Street Main Building or from the College Intranet at http://www.bbk.ac.uk/ccs/docs/docs.html#microsoft.
Help Desk is open on weekdays 10.00am - 8.00pm in term time, and in vacations it is open on weekdays 10.00am - 6.00pm.

Prerequisites

These notes assume that you are familiar with entering data and drawing charts, as explained in CCS document 5.93 and 'Software made Simple Guide to *Microsoft Excel 2003*'. You should know how to enter simple formulae and how to use **Autosum** and **Autofill**.

Contents

About this Document

- Words that you type or choose are in **bold type,** e.g. Choose **Programs** from the **Start** menu.
- Keys that you press are shown in small capitals e.g. press ALT.
- Press KEY1 + KEY2 means press both keys together e.g. press CTRL + Z
- Press KEY1, KEY2 means press each key consecutively e.g. press A, B.

➢ General guidelines on how to perform a task are indicated by bulleted lists using the   symbol. Note that these are *not* tasks for you to perform *now* but are advice for you to follow later when you need to work things out for yourself. Exercises for you to practice as you work through the notes are labelled **Exercise**, e.g. **Exercise 1 Importing Database Objects**. The steps that you follow to complete the exercise are marked by round bullets (•).

➢ Hints and tips are indicated by ☞. If you are not very familiar with using computers then you should pay particular attention to the hints and tips. Even if you are fairly familiar with using a computer you should read them, as they can provide explanations of why you do something, or they may suggest an alternative way of doing things.

1   Introduction

You can create tables using either the word-processing package *Microsoft Word for Windows 2003* or the spreadsheet program *Microsoft Excel 2003*. *Word* provides some simple features for sorting and performing calculations with numeric data in tables. However, *Microsoft Excel 2003* provides many more powerful features for handling numeric data. An example is provided in CCS Document 5.93, which explains how to use Autofill to fill in a series and the Function Wizard to calculate a statistical function.

In this document you will see how to

➢ enter formulae
➢ sort data
➢ filter data
➢ Create a pivot table
➢ Manipulate a time series

Although this document provides some examples of data manipulation with *Microsoft Excel 2003* it is by no means a complete guide to all the features of the programme. To cover everything would need a document many times the size of this. However, you do not need be buried under a pile of paper to find out more about *Microsoft Excel 2003*, as help is always available from the Help menu within the programme. If you are unfamiliar with using Help please refer to CCS Document 3.38 'Getting Started on CCS workstations'

2   Working with Formulae

2.1   Creating simple formulae

Document 5.93 explained how to enter a formula and how to use autofill to copy formulae. In general you will have to work out for yourself what formula you need to substitute in order to obtain a particular result. In the following example the formulae have been worked out for you. In this example we have used the variables in our formulae (**HeatRate, NoOfCats** etc.) as the column headers for our spreadsheet. This helps us to understand how the result is calculated. When you perform the calculation you substitute cell addresses into the formulae. *Microsoft Excel 2003* will then perform the calculations using the data values in the cells. If you change one of the data values in a cell the result will change.

Table 1 shows the booking information for a boarding cattery. The data provided includes
➢ The number of cats to be boarded
➢ The number of runs that they will occupy (cats from the same home may share a run)
➢ Whether the cats require heating in the run. This is indicated by the value 0 if no heating is required and the value 1 if heating is required.
➢ The rate per day per run
➢ The rate per day per cat
➢ The rate per day for heating

➢ The rate per day for insurance

The total cost for boarding a cat is calculated using the formula
  BoardingCost = KennelCost+HeatingCost+InsuranceCost
where
  Kennel Cost = (RunRate * NoOfRuns + CatRate * NoOfCats) * NoOfDays
  Heating Cost = Heating * HeatRate * NoOfDays
  Insurance Cost = InsuranceRate * NoOfCats * NoOfDays

➢ Note the use of brackets to enforce an order of precedence for the arithmetic operators in the expression for **KennelCost**
➢ If no heating is required, i.e. **Heating = 0** then the formula for **HeatingCost** returns a zero value. So we can use the same formula to calculate the **BoardingCost** whether or not heating is provided.

Exercise 1 Calculate Boarding Costs

- Log in and start *Microsoft Excel 2003* (click on Start at the bottom left, then Programs and Microsoft Office)
- Enter the data shown in Table 1 into cells **A1:K9** and save the file as **N:\My Documents\cats.xls**
- Enter the column heading **Kennel Cost** in cell **L1**, then enter the formula **=(H2*F2+I2*E2)*D2** in cell **L2**. By referring to the column headings you will see that this corresponds to the formula **Kennel Cost = (RunRate * NoOfRuns + CatRate * NoOfCats) * NoOfDays**
- Enter appropriate headings in cells **M1** and **N1**, then enter the formulae for **Heating Cost** and **Insurance Cost** in cells **M2** and **N2** (hint – use the column headings to work out which cells contain the relevant values to substitute in the formulae)
- Enter an appropriate heading in cell **O1** Select cells **L2:O2** and click on the **Autosum** button on the tool bar to calculate the **BoardingCost**
- Use the **Autofill** handle to copy your formulae into cells **L3:O9**

Table 1 Cats Data

Booking No	OwnerId	Cats In Date	No Of Days	No Of Cats	No Of Runs	Heating	Run Rate	Cat Rate	Heat Rate	Insurance Rate
14	4	31/01/01	3	1	1	1	£2.00	£2.00	£0.50	£0.10
15	7	05/04/01	3	2	1	1	£2.00	£2.00	£0.50	£0.10
16	2	07/06/01	10	4	1	0	£2.00	£2.00	£0.50	£0.10
17	9	03/02/02	10	2	1	1	£2.00	£2.00	£0.50	£0.10
21	2	03/03/02	9	2	1	0	£2.00	£2.00	£0.50	£0.10
22	6	03/03/02	15	1	1	1	£2.00	£2.00	£0.50	£0.10
23	6	03/03/02	3	1	1	1	£2.00	£2.00	£0.50	£0.10
28	2	26/08/02	3	2	1	1	£2.00	£2.00	£0.50	£0.10

Table 2 Calculated Results

Booking No	Owner Id	Cats In Date	No Of Days	No Of Cats	No Of Runs	Heating	Run Rate	Cat Rate	Heat Rate	Insurance Rate	Kennel Cost	Heating Cost	Insurance Cost	Boarding Cost	VAT	Total Cost
14	4	31/01/01	3	1	1	1	£2.00	£2.00	£0.50	£0.10	£12.00	£1.50	£0.30	£13.80	£2.42	£16.22
15	7	05/04/01	3	2	1	1	£2.00	£2.00	£0.50	£0.10	£18.00	£1.50	£0.60	£20.10	£3.52	£23.62
16	2	07/06/01	10	4	1	0	£2.00	£2.00	£0.50	£0.10	£100.00	£0.00	£4.00	£104.00	£18.20	£122.20
17	9	03/02/02	10	2	1	1	£2.00	£2.00	£0.50	£0.10	£60.00	£5.00	£2.00	£67.00	£11.73	£78.73
21	2	03/03/02	9	2	1	0	£2.00	£2.00	£0.50	£0.10	£54.00	£0.00	£1.80	£55.80	£9.77	£65.57
22	6	03/03/02	15	1	1	1	£2.00	£2.00	£0.50	£0.10	£60.00	£7.50	£1.50	£69.00	£12.08	£81.08
23	6	03/03/02	3	1	1	1	£2.00	£2.00	£0.50	£0.10	£12.00	£1.50	£0.30	£13.80	£2.42	£16.22
28	2	26/08/02	3	2	1	1	£2.00	£2.00	£0.50	£0.10	£18.00	£1.50	£0.60	£20.10	£3.52	£23.62

2.2   Absolute and Relative References

We will now calculate the VAT to be added to the boarding cost to give us the total bill. So far you have used *relative cell references*, which are references to cells relative to the position of the formula. However, for some tasks you will need to use *absolute references*, which are cell references that always refer to cells in a specific location. *Absolute references* use the format *CR* where *C* is the column letter and *R* is the row number.

To calculate the VAT we need to multiply the boarding cost by the VAT Rate, using the expression **VAT = VatRate * BoardingCost**. We only need to enter the VAT Rate once, so we will use cell **B11** to store the VAT Rate. If we then calculate VAT in column P, we will not get the correct result. **B11** is a *relative address* and will change to **B12** etc when we autofill the column. We need to specify that the row number is an *absolute address* by putting a dollar sign in front of the row number i.e. **B$11**. Having calculated the VAT we can then calculate the total cost using the formula **Total Cost = BoardingCost + VAT**

Exercise 2 Calculate VAT

- Type a label **VAT Rate** in cell **A11**
- Enter the value **0.175** in cell **B11**
- Enter a column heading **VAT** in cell **P1**
- Enter the formula **=O2*B$11** in cell **P2**
- Enter a column heading **Total Cost** in cell **Q1**
- Calculate the total cost in cell **Q2**
- Use Autofill to enter the Vat and Total Cost values in cells **P3:Q9**
- Save **Cats.xls** again to save the changes.

In the previous example you had to type in the data. However, it is tedious typing in data, so in the following sections we will use alternative techniques (copying, Save As) to get data from sample files.

3   Sorting Data

You can use Excel to sort and manipulate lists of information - rather like a very simple database. In this context it treats column headings (if there are any) as if they were field names.

To sort all the data in a worksheet by the contents of one column:

➢ Click anywhere in the column to be used for sorting
➢ Click on the **Sort** button on the toolbar (see right)
➢ On the dialogue box that pops up, choose to expand the selection – that will keep data in each row together
➢ To sort by more than one column, choose **Sort** on the **Data** menu – that will bring up a dialogue box, allowing you to sort by up to three columns

☞ *For the sorting exercise you will make a copy of some existing Microsoft Excel 2003 data. Note that you can copy data from one spreadsheet to another using the same **Copy/Paste** technique that you will have used in other situations; for example, copying text in a Word for Windows document or cells in a Microsoft Excel 2003 spreadsheet.*

☞ *When you choose **Open** from the Microsoft Excel 2003 **File** menu Excel will look first for a file in the **My Documents** on your personal networked drive, **N:\My Documents**. You can use the drop down menu on the box labelled **Look In** to select files on other drives.*

☞ *If you find double clicking difficult when you are working in Open or Save dialogue boxes then you can single click on the object to select it and then click on the Open button.*
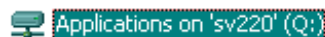
Exercise 3 Open a file on a shared drive.

- Click on the **File** on the menu bar
- Click on **Open** on the file menu. The **open** dialogue box appears
- Click on the downward pointing arrow on the right of the box labelled **Look In.** (Figure 3-1)

Figure 3-1 'Look In' box

- Click on the drive labelled **Q:** (Figure 3-2). The contents of Q: will be displayed

Figure 3-2 Networked Drive Q:

- Double click on the folder labelled **Samples**. Again, the contents of the folder will be displayed
- Double click on the folder labelled **excel**. This time you will see some spreadsheet documents
- Double click on the spreadsheet called **bears.** The document will now open

☞ *The shorthand way of describing the document that you have opened is **Q:\Samples\excel\bears.xls***

☞ *To select a large data range click in the first cell, hold down the shift key, and click on the last cell.*

Exercise 4 Copy shared data.

- If the sheet labelled **BEARS** is not selected then click on the **BEARS** tab to select it
- Select cells **A1: L144**
- Click on the **copy** button on the tool bar
- From the **File** menu choose **New**, select the **workbook icon** in the **general** section of the dialogue box, and click on **OK**
- Click in Cell A1, then click on the paste button
- From the **File** menu choose **Save As**. In the **Save As** dialogue box use the downward pointing arrow on the right of the box labelled **Look In** to locate the **N:** drive
- Double click on the folder labelled **My Documents**
- Type the name **mybears.xls** in the box labelled **File Name**. You have now saved your workbook as **N:\My Documents\mybears.xls**
- Close **Q:\Samples\excel\bears.xls** (Use the **window** menu to switch between files)

Exercise 5 Sort data

- Click in the column labelled **Weight** (*not* on the column header)
- Click on the **Sort Ascending** button
- Save **mybears**.**xls**

You will see that the rows are now sorted in ascending order of weight.

4   Filtering Data

To filter out some of the data - i.e. view only those rows that have specified values in one or more columns:

- ➢ On the **Data** menu, choose **Filter**, and then **AutoFilter**
- ➢ A downward pointing arrow will appear at the top of each column. If you click on one of these, a drop-down list of unique values in the column will be displayed. Click on one of the values, and only those rows that have this value in this column will be displayed. Notice that you are *displaying* the filtered data. You have not removed the rest of the data, only hidden it
- ➢ To view all data, again choose **Data** – **Filter** – **AutoFilter**

If you want Excel to sort or filter dates, you will need to specify that a column has dates in it - i.e. has a *date format.* Otherwise it will be sorted as ordinary text. To specify a format:

- ➢ On the **Format** menu, choose **Cells**, and then the **Number** tab
- ➢ Choose the type of date format that you want, and then click on **OK**

➢ If you want to perform calculations or produce charts from the filtered *sub set* of the data then you will need to copy it to another range of cells, either on the same worksheet, a new worksheet or a new workbook.

For the next exercise you will display the data about the 10 heaviest bears.
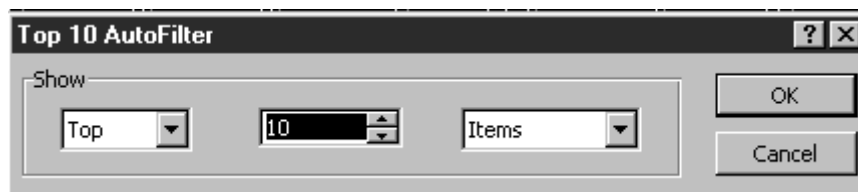
Exercise 6 Filtering Data

• From the **Data** menu choose **Filter**, then choose **Autofilter**
• Click on the downward pointing arrow on the right of the column headed **Weight**. A list of filters drops down (Figure 4-1)

Figure 4-1 AutoFilter Menu
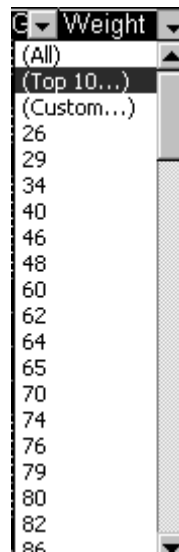
• Click on **Top 10**. A **Top 10 Autofilter** dialogue box appears. (Figure 4-2).

Figure 4-2 Autofilter Dialogue Box

• Click on **OK**

Notice that the *Autofilter Dialogue Box* offers various options for customising the filter.

➢ You can select the **Bottom** instead of **Top** from a drop down menu
➢ You can increase or decrease the number of items using the scroll arrows on the number box
➢ You can choose whether to show **Items** or **Percentage**
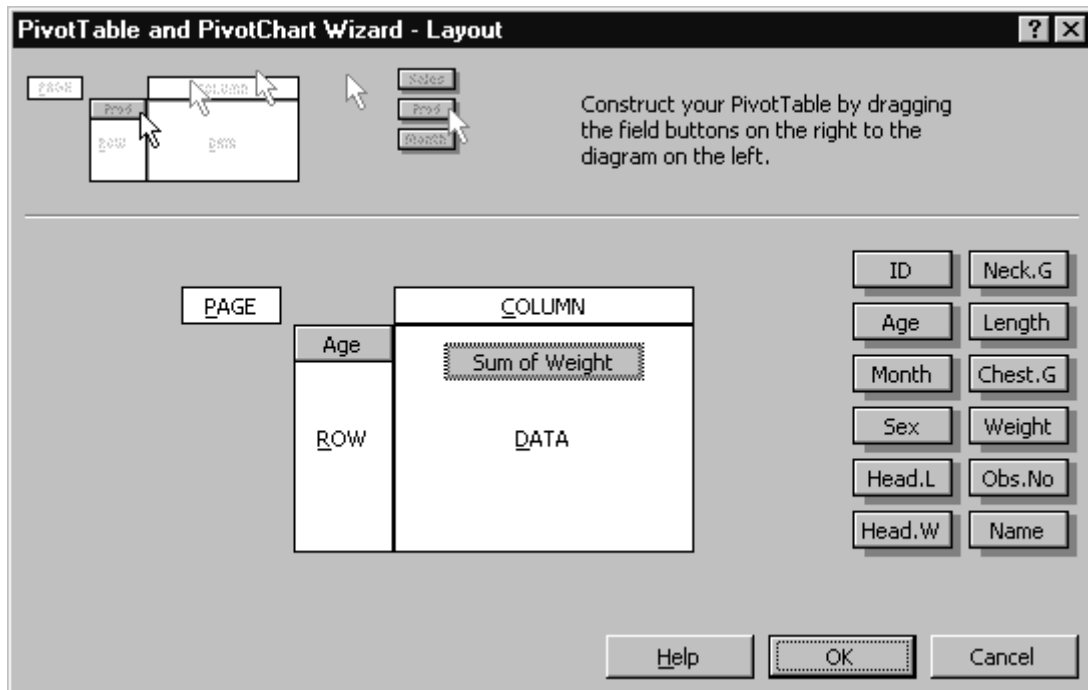
5   Pivot Tables

A PivotTable is an interactive table that quickly summarises, or cross-tabulates, large amounts of data. You may rotate its rows and columns to see different summaries of the source data, filter the data by displaying different pages, or display the details for areas of interest.

☞   *The online-help provides a clear graphical description of the parts of a pivot table and the terminology that is used. If you unfamiliar with the terminology used in the following example look at the Help Contents, analysing data with pivot tables.*

☞   *If you use pivot tables a lot, then choose **Toolbars** from the **View** menu and make sure that **Pivot Table** is checked.*

As an example we will summarise the average weights for bears of different ages. If you have done the filtering exercise, make sure that you are displaying all the data.

Exercise 7 Create a Pivot Table

- Choose **Filter** on the **Data Menu**, then choose **Show All** to redisplay all the data
- Select the range **A1:L144**
- From the **Data** menu choose **PivotTable and PivotChart Report**. As with many *Microsoft Office* tasks a *Wizard* appears to help you
- In the **Pivot Table Wizard Step 1 of 3** box make sure that **Microsoft Office Excel List or Database** is checked in the section headed **Where is the data that you want to analyse?**
Make sure that **PivotTable** is selected in the section headed **What Kind of Report do you want to create?** Click on **Next >**
- In the **Pivot Table Wizard Step 2 of 3** box check that the selected range is displayed as **$A$1:$L$144.** (If you need to select a different range then you can do so at this point by clicking outside the box & dragging to select the required range.). Click on **Next >**
- In the **Pivot Table Wizard Step 3 of 3** box click on the button labelled **Layout**
- Drag the field button labelled **Age** into the **Row** section of the diagram on the left
- Drag the field button labelled **Weight** into the **Data** section of the diagram (Figure 5-1)

Figure 5-1 Choose the parameters

- The default function is **Sum of Weight**. To use another function double click on the button labelled **Sum of Weight.** In the **Pivot Table Field** dialogue box (Figure 5-2) select **Average** from the list of functions and click on **OK**

Figure 5-2 Selecting a pivot table function
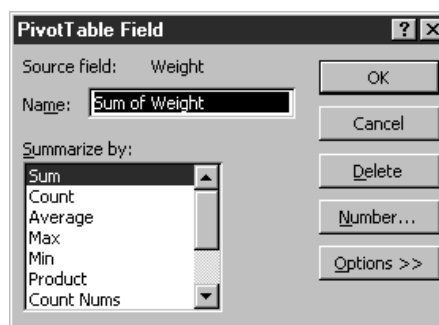
- Your pivot table diagram should now appear as shown in Figure 5-3. Click on **OK**

Figure 5-3 Pivot Table for Age and Average Weights
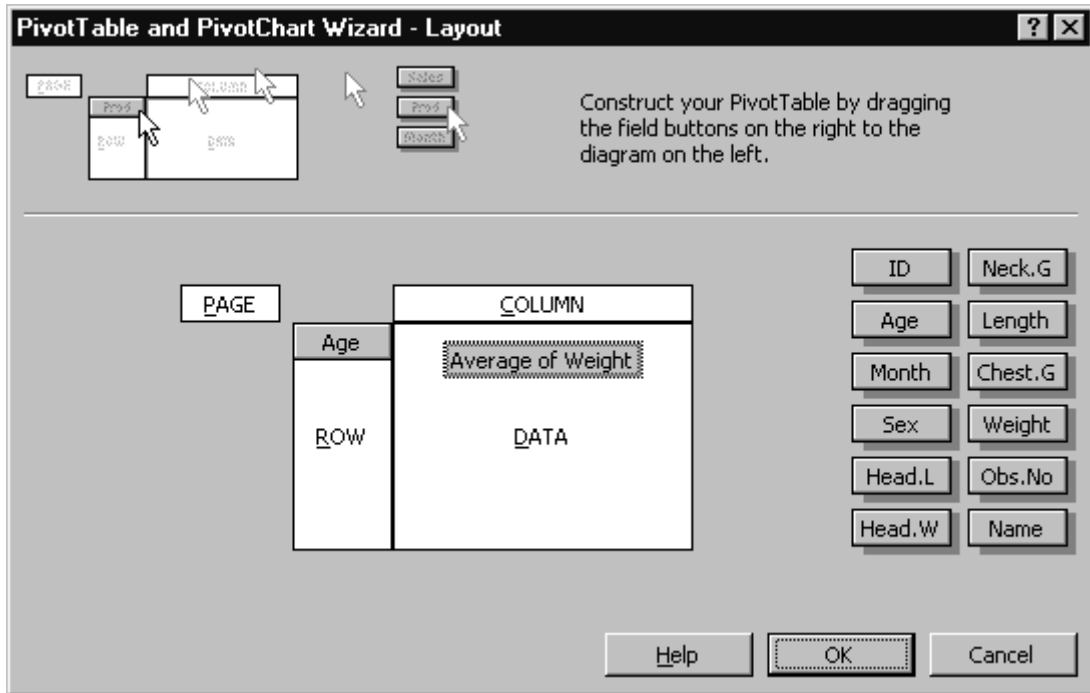
- In the **Pivot Table Wizard Step 3 of 3** box check that **New Worksheet** is selected and click on **Finish**. A new sheet is inserted in your workbook and a table of age and average weight is calculated, as shown in Figure 5-4.
- Save and Close **mybears.xls**

	A	B
1	Average of Weight	
2	Age	Total
3	0	200.2833333
4	8	31.5
5	9	37.33333333
6	10	81.66666667
7	11	70.5
8	16	79.5
9	17	82
10	18	100
11	19	71.66666667
12	20	108.75
13	21	124.8
14	22	176
15	23	143.3333333
16	29	121
17	30	210
18	31	289.5
19	32	173
20	33	154
21	34	207.3333333
22	35	212

Sheet4 / Sheet1 / Sheet2

Figure 5-4 Pivot Table

6   Manipulating and Displaying Time-Series Data

In Exercise 4 we saw how to copy a subset of the data from a shared file into our own file in order to work with it. In the next example we will be working with all the data, not just a subset, so we will need to copy the whole file to our own networked drive.

Exercise 8 Make a copy of the sample data

- Minimise the *Microsoft Excel 2003* window by clicking on the minimise button (see left) in the top right hand corner of the window
- Double click on the *Windows Explorer* icon on the desktop. Be careful to choose *Windows Explorer*, not *Internet Explorer*. *Windows Explorer* is the *Windows* file management utility that allows you to explore the files on your *own* computer. *Internet Explorer* is a *web browser* that allows you to explore the whole Internet.
- In the left hand pane of *Windows Explorer* click on the **My Computer icon**
- In the left hand pane locate the drive labelled **Q:**[1] and click on the **+** symbol [2] next to the drive
- Scroll down the left pane until you see a folder **Samples** beneath the **Q:** drive
- Click on the **+** symbol next to **Samples**
- Click on the **excel** folder. The contents of **Q:\samples\excel** will appear in the right hand pane.
- Click on **ryeprice.xls** in the right-hand pane to select it. It will be highlighted.
- Select **Copy** from the **Edit** menu.
- In the left pane click on the drive where you normally save your files. On CCS student workstations this will be the **N**: drive.
- If the **+** symbol is shown click on it to view the folders on your **N:** drive. You will see that there is a folder labelled **My Documents**.
- Click **My Documents** to display the contents in the right hand pane
- Click in the right hand pane.
- Select **Paste** from the **Edit** menu, to paste the file from the clipboard into the **My Documents** folder
- Go back to the *Microsoft Excel 2003* by clicking on the *Microsoft Excel 2003* button on the task bar, and open **N:\My Documents\ryeprice.doc**
- Save your file as **N:\My Documents\myryeprice.doc**

This gives the price of bread and the wage rate for labourers in Rye, Sussex in the second half of the 16th century. We will try graphing the figures, to see if wages kept up with the increase in bread prices at the end of the 16th

---

[1] If you cannot see the Q: drive then scroll down the left-hand pane until it appears. The drives are sorted in alphabetic order.

[2] If there is a minus sign (-) next to the Q: drive then *Windows NT* has already displayed the contents of Q: and you can proceed to the next step.

century. Note that like many historical data sets, the price series has some missing values.

6.1   More on AutoFill, and Dealing with Missing Values

Let us insert some dates before the first column:

Exercise 9 Autofill a Date Series

- Click on the grey column header at the beginning of the column headed **Weight of 2d Loaf** to select the column.
- On the **Insert** menu, choose **Columns**. A blank column will be inserted before the selected column. The new column will become column A, and the others will be renamed.
- In the first cell **(A1)**, type **Year**
- In **A2**, type **1553**, and **1554** in **A3**
- We can now use **AutoFill** to complete the series. Select **A2:A3**, and then drag the *fill* handle down to **A51**- the last date should be **1602**.

☞   *An alternative method to using the autofill handle is to select **A2:A51**, choose **Fill** on the **Edit** menu, then **Series**, and **OK**.*

☞   *If you had only entered one number and then used AutoFill, the same number would have been entered in each cell. Because there were two numbers, in A2 and A3, Excel calculated the interval between them (**1** in this case), and was then able to complete the series.*

☞   *2d are two old pence in pre-decimal currency. 1p is worth about $2\frac{1}{2}$ d*

The figures for bread prices are not very useful in their present form, because they are expressed in terms of the weight of a 2d loaf. In the next exercise you will convert this to price per lb. of bread.

Where the data is missing, Microsoft Excel 2003 will insert the error message #DIV/0! because you have attempted to divide by zero - although the value is not really zero, it is just not available. This will cause some problems when you format & chart the data.

Firstly, if you use the **Decrease Decimal** button on the tool bar to adjust the number of decimal places in a range that contains an error Microsoft Excel 2003 will beep. One way to overcome this is to first select only those cells that contain valid data using the CTRL key to select non-adjacent cells. A better method is to use the **Format** menu as follows:

➢ Select the whole range.
➢ Choose **cells** from the **Format** menu, click on the **Number** tab.

➢ Choose **Number** from the category box and select the number of decimal places that you want.
➢ Click on **OK**

Secondly, when you try to chart the data, it will interpret #DIV/0! as zero, which is not what you want. Before charting the data you will need to manually delete the #DIV/0!

☞ *To fit "Price of Bread" into the column without increasing the column width use the format painter button to copy the formatting C1 to cell D1. Cell C1 has been formatted to **wrap text** by choosing **Cells (alignment** tab) from the **format** menu.*

Exercise 10 Calculating Missing Values

- In column D, type the heading **Price of Bread**
- In D2, enter the formula **=32/b2**, and press ENTER. Excel will not be able to perform the calculation in this cell, because one of the figures is missing. You can still use *AutoFill* to fill in the rest of the column
- Select the range **D2:D51**
- Choose **cells** from the **Format** menu, click on the **Number** tab
- Choose **Number** from the category box and select the number of decimal places that you want
- Click on **OK**
- Remove each instance of *#DIV/0!* and leave the cell blank

When the data is graphed, Excel should leave gaps where there are missing values.

6.2 Preparing Time-Series Data for Graphical Display

Now let us graph the data in column D, using the labels in column A.

Exercise 11 Charting Non Adjacent columns

- Select cells **A1:A51.**by dragging
- Hold down the control key and select cells **D1:D51** (See right)
- Start the *Chart Wizard*
- When selecting a chart type (*Step 1 of 4* in the *ChartWizard* dialog box), choose **XY (Scatter)**
- In *chart subtype section* choose format **2** (data points connected by smoothed lines) for the scatter plot

	A	B	C	D
		Weight of 2d	Labourer's Daily	Price of
1	Year	Loaf	Wages	Bread
2	1553		9	
3	1554	82	9	0.39
4	1555	66	9	0.48
5	1556	40	9	0.80
6	1557	72	9	0.44
7	1558		9	
8	1559	76	9	0.42
9	1560		9	
10	1561	66	9	0.48
11	1562	66	9	0.48
12	1563		10	
13	1564		10	
14	1565	76	10	0.42
15	1566	44	10	0.73
16	1567	64	10	0.50
17	1568	64	10	0.50
18	1569	53	10	0.60
19	1570	56	10	0.57
20	1571	48	10	0.67
21	1572		10	

- In *Step 4 of 4* choose **As New Sheet**. Your chart will appear on a new worksheet labelled chart 1

If you wish, you can add another graph for the wages data.

It would be useful to put the prices and wages data on the same graph, but they are at present in different units. We can do this if we produce an *index* for each series - i.e. calculate each value as a percentage of the first value in the series. So both series would start with **100** in 1553. Let us put the index for wages in column E, and for prices in column F: We will use a formula that illustrates the difference between a *relative reference* and an *absolute reference.*

6.3   Using an absolute reference in a formula

When you created the formula for the price of bread you divided the cost by the weight to give the price per lb. When you used autofill to copy the formula *Microsoft Excel 2003* automatically updated the row reference of the cell, so you calculated 32/b2, 32/b3, 32/b4 etc. (You can check this by clicking in the cells in column D and viewing the formula in the formula bar.) The cell reference *b2* is a *relative reference;* if you copy it, the references change relative to the original cell. When you are calculating an index you need to divide by constant value. You could just type the value that is in the first cell, i.e. 9. However, if the value in the first cell changed you would have to update all the cells that use that value. It is easier to use an *absolute reference* that always refers to the same cell.

☞ *To switch between sheets in a workbook click on the sheet (or chart) tab at the bottom of the page.*

Exercise 12 Calculate Indexes

- Make sure that **sheet 1** is selected
- In cell **E1** type **Index of Wages**
- In **E2**, enter the formula: **=C2/$C$2** (cell **C2** contains the first value in the series, so all values are expressed as a percentage of this value)
- Use *AutoFill* to fill the rest of the column
- In cell **F1** type **Index of Prices**
- In column **F**, we will have to start with **F3**, because the value in **D2** is missing. Enter the formula: **=D3/$D$3**, and then use *AutoFill* as before
- Select the numeric data in cells **E2:F51**, and use the **%** button on the tool bar to format the data in columns **E & F** as a percentage
- Excel will unfortunately have replaced the missing values with zeroes, so remove the zeroes, leaving the cells blank
- Do another scatter chart for columns **A, E** and **F**. Give your chart the title **Index of Prices and Wages** and put it on a separate sheet

Note that the shape of the graph will be to some extent influenced by the value that you took as a starting point (for the year *1553* in this instance). If

the starting value was very odd, it will not be very meaningful to express the other values as a percentage of it.

It is rather difficult assessing a trend in bread prices because of the extreme variation from year to year, but a method of smoothing irregularities is introduced in the next section.

6.4   Using Statistical Functions in Excel (*Data Analysis* module)

As a final refinement, we could use one of Excel's statistical functions - *Moving Average* - to smooth out irregularities in the Index of bread prices, and make any trend more visible.   To take a three-year moving average, we would take an average of three values at a time, overlapping, so that we get a new series consisting of the averages. For example, the first four values in the Index of bread prices' series are the following:

<div align="center">

100%
124%
205%
114%

</div>

The moving average series would start with 143 (average of 100+124+205), then 148 (average of 124+205+114) etc.

If you wish to see the calculated values moving averages, you first need to check that *Microsoft Excel 2003* statistical functions are available, since they are optional 'Add-Ins':

➢ On the **Tools** menu, check if **Data Analysis** is available (at the bottom of the menu); if not, click on **Add-Ins**
➢ In the *Add-Ins* dialog box, click on **Analysis Toolpak**, and then on **OK**

The item **Data Analysis** should now be available on the **Tools** menu - this gives access to statistical functions. You could now calculate the series (putting the results, say, in column G) and add it to your chart. However, *Microsoft Excel 2003* allows you to fit a trend line directly to the chart itself, so there is no need to go through the calculation first unless you need to see the actual data values.

<div align="center">

Exercise 13 Fit a Trend line to a Chart

</div>

• Select **Chart 2** and *right click* on the **Index of Prices** data series
• Select **Add Trendline** from the pop up menu. The **Add Trendline** dialogue box appears
• Make sure that the **Type** tab is selected
• Select **Moving Average** for the *trend/regression type*
• In the *Based on Series* box make sure **Index of Prices** is selected
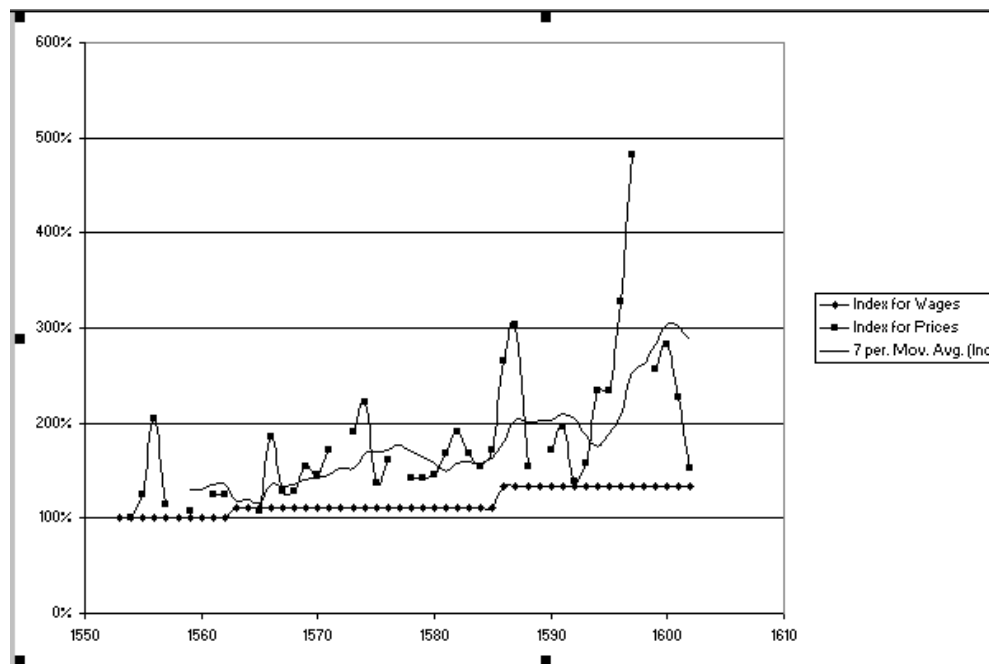• In the *Period* box enter **7**

• Your dialogue box should now appear as in Figure 6-1. Click on **OK**

Figure 6-1 Trendline Dialogue box.
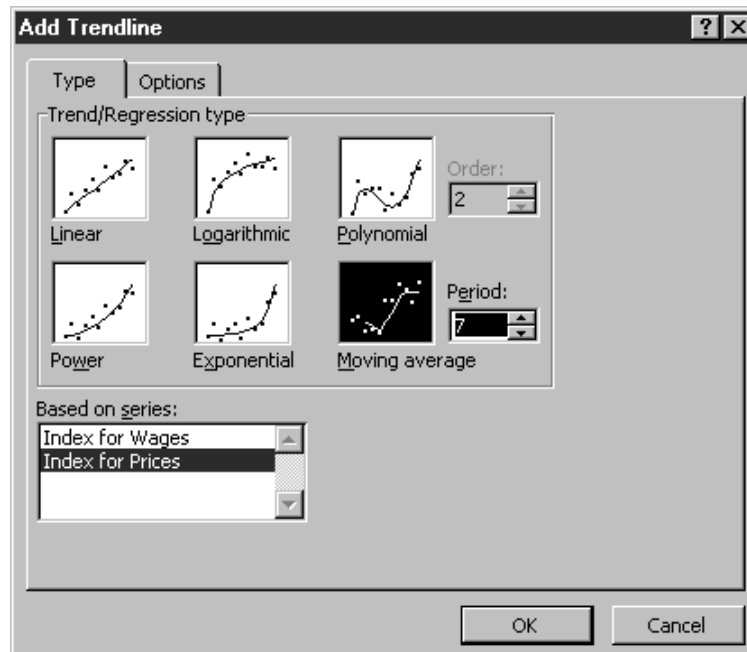
Your chart should now be similar to Figure 6-2

Figure 6-2 Chart with trend line

Do labourers' wages keep up with the increase in bread prices?   To see how different intervals change the chart, you could try 3-year and 5-year moving averages. You need to be careful in choosing the right interval - (1) to balance getting a clearer view of the trend against any resulting loss of information and (2) not to obscure any regular fluctuations that may exist in the data.

To experiment with different intervals for the moving average

➢ Click on the trend line to select it
➢ *Right* click the mouse
➢ From the menu that appears choose **Format Trendline**. (If this option is not available, make sure that you have selected the trend line.)
➢ In the dialogue box that appears select the **Type** tab
➢ Change the *period* value to the one you want
➢ Click on OK