



DARWIN INITIATIVE BASIC TRAINING IN MS EXCEL

March 2007

BASIC TRAINING IN MS EXCEL

OBJECTIVE

This introduction to Microsoft Excel has the overall explicit purpose of bringing everyone up to a certain common user level, whereby, irrespective of your starting point, you will feel confident in carrying out basic spreadsheet design and functions, and explore further possibilities on your own using the help function. In other words, since we are all concerned with collecting data of one sort or another, you will be able to:

- 1) Think logically about the best lay-out (design) of a worksheet;
- 2) Enter data in such a worksheet;
- 3) Manipulate these data using Excel's inbuilt functions;
- 4) Summarise these data efficiently;
- 5) Explore and present these summary data graphically
- 6) Crudely assess statistical trends in these data within the limits of Excel.

The expectation is therefore that you will leave this workshop feeling confident in designing your own spreadsheets, entering and analyzing data, and providing useful summaries of these data. I would hope too that you will feel able to explore the power of Excel on your own using the help facilities as and when your needs dictate. Please note that if you do wish to attend this training workshop, you must bring a laptop with you which has Excel installed, and further you ensure that your copy has the "Analysis Tools" component installed (go to Menu, Tools: Data Analysis). If there is no indication of Data Analysis, you will need to insert the MS Office CD & explicitly request the download of "Analysis ToolPak".

CONTENTS

WHAT IS A SPREADSHEET?

OK – so what is a spreadsheet? Let's start by opening Excel. To do this, either double click on the Excel icon on your desktop, or go to the Start menu, and click on Excel. If neither of these two options is available on your computer, use the Search facility in Windows Explorer and type "excel.exe". The executable file should be sitting in C:\Program Files\Microsoft Office\... Shout if you are having any problems opening Excel.

With Excel open, there are a number of points to note:

1) There is a grid in front of you, consisting of 256 **columns** (labeled A – IV), and 65,536 **rows** (labeled 1 – N). These 256 columns and 65 000 odd rows give nearly 17 million unique **cells**: this may seem a lot, but for many purposes is inadequate, and is one of several reasons why SACWG is placing its data onto a **relational database**. Please note that spreadsheets and workbooks are *not* relational databases, although through specific commands the former are beginning to approach the functionality of the latter.

2) The worksheet grid gives Excel its power: each **cell** is unique, and can be singularly defined by its column and row position e.g. cell A1 refers uniquely to the top left hand cell of the worksheet.

→ *Position your mouse cursor over cell D10.*

→ *What do you notice?*

3) With this column, row identification system, each of the 17 million odd cells in a worksheet can be uniquely referenced. It is this unique referencing system that forms the backbone to Excel's power and functionality. However, it is all very well knowing the ID of a cell, but you need to know which cell you are working on *now*.

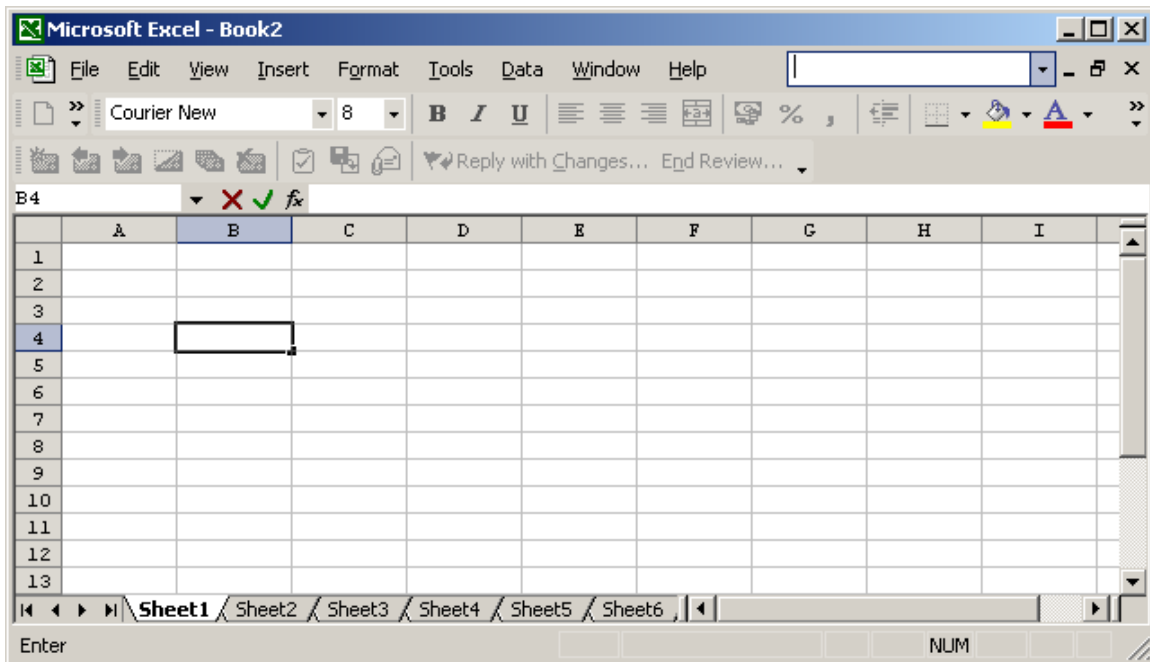
→ *Click on cell B4*

→ *Move to the bottom of the worksheet, and scrolling both across columns and down rows, see if you can reach "the end" i.e. cell IV65536. Do not click on this. If you get bored with the scrolling, stop, and just realize that B4 (your current cell) is not visible.*

→ *To get back to your "current" cell, hit Return. You should now be in cell B5, and it is a simple case to get back to B4. Alternatively, use Edit, Find, and you will be back to B4.*

→ *Make sure that B4 is your current cell. As we discussed above, you can see where you currently are since Excel indicates your current cell position in three ways: i) the current cell ID is given in the name box (top left of sheet); ii) the column B and row 4 are both darkened or coloured (depending on Preferences set); and iii) the border around the cell itself is highlighted.*

At this point, everyone should have a screen looking something like the following:



4) For information only: certain clever-clogs decided they didn't like the standard Column, Row way of identifying cells (and we can agree that BG12 may seem slightly clunky), and instead preferred a solely numeric form known as the **R1C1 reference style**. There are advantages to this system, but *1) we will not use it, and 2) the default is Column, Row ID.*

5) Whilst we are considering cell identification, it is worth pointing out that, for example, cell reference A1 can also be written as \$A1, A\$1, and \$A\$1. Each signifies something very specific in terms of functionality, and these differences are important. We will return to these configurations later. At this point it is sufficient simply to note they all refer to cell A1.

6) The screen in front of you should be familiar, at least in terms of overall "feel". This is because Excel is part of the Microsoft Office suite that includes MS Word, Access etc. We will return to the specifics under the section "The Excel Screen".

7) The last point to note at this stage is that not only can you see a "top" worksheet (labeled in the bottom tab as "Sheet1"), but that there are additional tabbed sheets below. The default number of sheets is three, but this can be set to your own value by going to the menu Tools, Options, General.

OK – so we have now given a very brief overview of what an Excel Worksheet (= spreadsheet) looks like, but we have not yet defined what a spreadsheet is. A simple working definition may simply be “an organized collection of data”. Indeed, an electronic dictionary (*American Heritage Dictionary*) came up with the following two definitions:

1. A piece of paper with rows and columns for recording financial data for use in comparative analysis.
2. *Computer Science* An accounting or bookkeeping program that displays data in rows and columns on a screen.

Common to both these definitions is that a spreadsheet, whether paper or electronic, has data organized in columns and rows. Strictly, tables have headings in both columns and rows, whilst data lists are simply organized according to data column headings.

We can thus expand our above simple working definition into a slightly more restrictive definition, namely

“an organized collection of data, with each entry defined by its specific cell (grid) position, which in turn reflects the type of data entered”

WHY USE SPREADSHEETS?

It is probably a bit early to be answering this question in full, but from what we’ve discussed above a number of key reasons should be apparent:

1. The design of the spreadsheet forces one to think logically: how should the data be collected and recorded, and why collect those specific data.
2. What are the relationships between the differing data being collected
3. Allows cross-referencing e.g. a spreadsheet recording nesting data may include columns for parental ID; a separate spreadsheet may hold data relating to age and other characteristics of each bird
4. Allows data summaries
5. Allows data sharing
6. Allows data back-up
7. Ultimately, allows data analysis.

DESIGNING A SPREADSHEET

This question (how best to design a spreadsheet?) may appear to have an obvious answer: just get on and do it! OK – a typical data sheet that you may wish to create relates to your counts of cranes over time

→ ***Design a spreadsheet with column headings and/or row headings that might allow you to enter count data over time.***

There are clearly a number of alternative designs, and, as importantly, a range of thoughts as to what data should be included in such a spreadsheet. One answer may be to simply set up a number of columns, with headings for Date, Time, Place, Count. But should weather be included? If so, which variables are important? How should one present incomplete counts? Should they be discarded? Instead of a *data list* (columns of data), a *data table* may be more useful, perhaps with date running across the page (columns) and census routes running down the page (rows). None of these variations is necessarily right or wrong: the key is why you want the data entered into a spreadsheet in the first place i.e. What is the *PURPOSE* of the spreadsheet?

→ ***Open sheet1.xls***

This spreadsheet is probably incomprehensible to you, but was a crucial tool for two fieldworkers I had studying the breeding behaviour of Sociable Weavers and White-browed Sparrow Weavers. As you can see, date runs across the page, and nest ID runs down the page. But two things should immediately strike you – i) there are lots of blank spaces; and ii) the spreadsheet runs across a lot of screens. However, this spreadsheet was designed for a very specific purpose: my fieldworkers needed to be able to roll out of bed at four a.m. each morning, and simply look down the relevant date column. This would inform them exactly which nests needed visiting that day, and what activities had to be carried out that nest. But this sheet is of *NO* use if I want to calculate the mean clutch-size or mean lay-date of these birds.

→ ***Design a spreadsheet which would be suitable for recording the relevant breeding parameters for pairs of Blue Cranes in your study area (e.g. clutch-size, lay-date, hatching success, etc.***

As a *general* rule, electronic data should be entered in a form that allows the same data type/class to run down a column (i.e. a column specifically for clutch-size, a column specifically for aspect of nest etc). The reasons for this will become more apparent as we go on, but for now

→ ***Open sheet2.xls.***

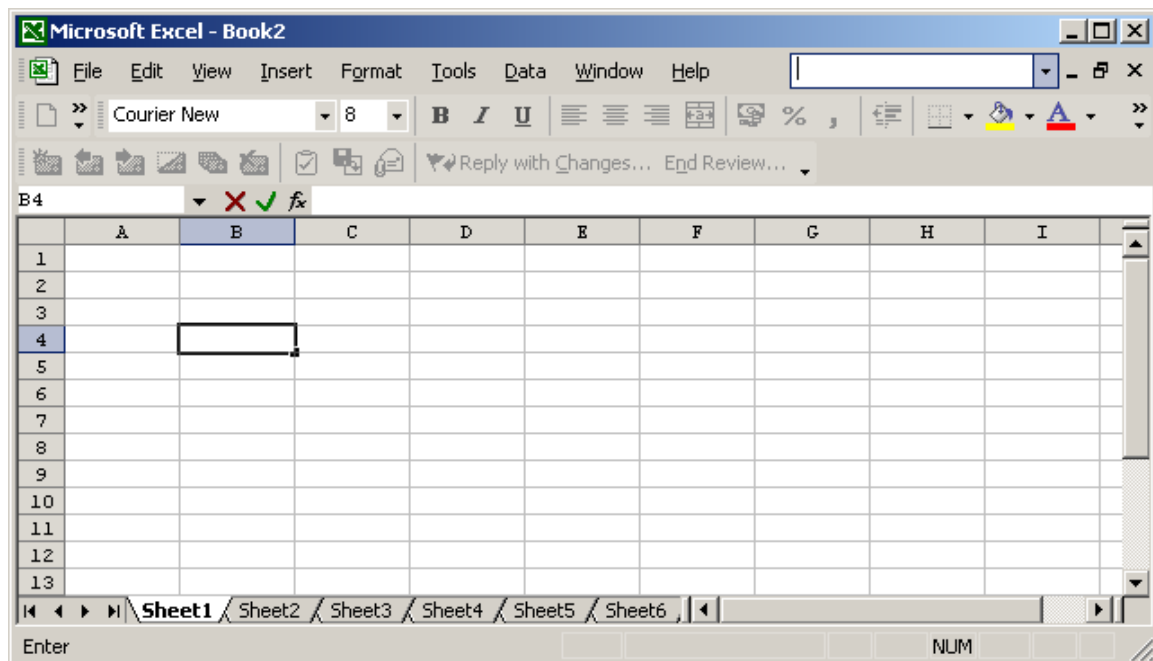
This spreadsheet represents data on the key breeding parameters of a single species collected over time. You can see that with the data stored in this format, that it meets two key criteria of datasheet design: it is *logically structured*, and is *parsimonious* (*efficient*).

It would be very easy to extract summary data, to copy key sections of data, and to store back-ups. I have collected some of these data, and I can assure you the above is true relative to the information collected in the notebooks at the time.

The essential point in spreadsheet design is: *What are these data ultimately going to be used for?* The answer to this question largely determines the structure of the sheet: if it is simply an aid to help you collate your monthly garden list, then the importance of having the correct structure is less than in our situation where Raj and Leon have largely specified a common data sheet format for all your data so that they can then be automatically read into a relational database. But the bottom line is: *think before you create!*

→ *Spend 5 minutes now, designing a spreadsheet which would be suitable for recording the number of cranes that you see on a daily basis. Think about column and row headings as necessary, and save it as “mysheet1.xls”. We will return to it later.*

THE EXCEL SCREEN



Even if you are completely unfamiliar with working in Excel, the screen above should have a feeling of familiarity about it, especially if you use the other Microsoft Office programmes. Working roughly from top to bottom, one has:

1. **The Title Bar:** What file you have open in front of you (here, Book2.xls)
2. **The Menu Bar:** Standard pull-down menus
3. **Toolbar(s):** Default, or designed to one's preference settings: quick button commands
4. **Name Box:** Cell address

5. **Formula Bar:** Inbuilt Excel functions or macro call
6. **Work Area:** Worksheet grids
7. **Tab Area:** options to move between *Worksheets* within *Workbook* (*.xls)
8. **Status Bar:** Reminder of current status re e.g. caps lock, numb lock etc.
9. **Task Pane:** On opening (& at other times) will appear on top of right hand side of worksheet, depending on context (e.g. on opening, will list last *N* files used; on opening help, then new options open here etc).

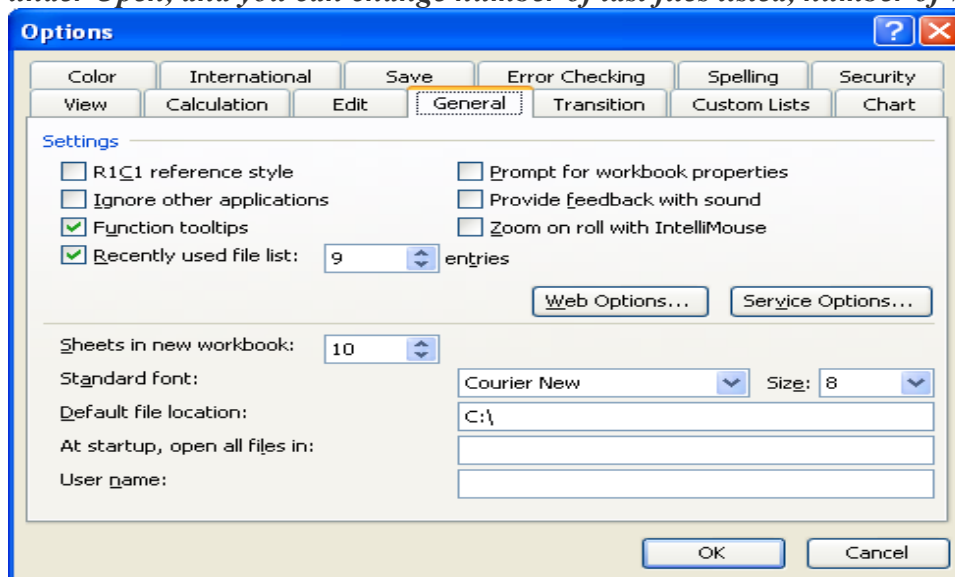
→ *Close any open worksheets (i.e. Close Excel) and re-open. Go to the Menu and use File, Open. Open three files in order: 1.xls; 1.txt and 1.csv. Any comments, observations? Now close all 3 files.*

By and large, the only time to save a worksheet in non-Excel format (i.e. as some kind of *.txt file) is to save space (e.g. say you wish to send a data file to a colleague as an attachment to an e-mail – sending it as a txt file can halve the size of the file). *But note, you will only save the current sheet in this format, NOT the workbook.* Also note some older statistical applications (or versions) required the import of text files – all the common packages now allow *.xls files to be imported.

Before going through the key elements of the Menus, there is one Menu Option to consider first.

→ *Go to the Menu and use Menu, Tools, Customise. Under Toolbars, click (i.e. tick) Standard, Formatting, and Worksheet Menu Bar. You can customize your command buttons at your leisure, but go on to check Options. I would advise against ticking Large Buttons!*

→ *Go to the Menu and use Menu, Tools, Options. Click between the tabs here, but the key one is General, allowing you to change where Excel automatically “starts up” under Open, and you can change number of last files listed, number of worksheets etc.*



May I also make a personal plea: I am frequently having to look at spreadsheets created by students – they delight in psychedelic colours and winking fonts in varying sizes. ***DON'T***. Courier New, Font size 8, at 100% magnification is just fine ...

The File Menu

→ *Does anyone here feel uncomfortable with any of the commands under the File menu, such as New, Open, Save As and Save? If so, shout.*

Otherwise, under the File menu, the only point I wish to discuss concerns setting up sheets for printing.

→ *Open 1.xls. Now go to File, Page Setup. You will see a number of options, but the key one is under Sheet, where you **MUST** get into the habit of ticking gridlines in order for these to print. Do this. Now go to Print Preview. Firstly, click on Margins and move these around. Second, go back to setup, and click gridlines. If you want grids to cover the whole page, drop down to where page ends and add a period (full stop, .) in last cell. Now check that you have required grid in Print Preview. Finally, go to Header and Footer (under Page Setup) and play around with these.*

→ *Any Questions?*

→ *Go back to Menu, & scroll down, checking you are happy with all the menu options.*

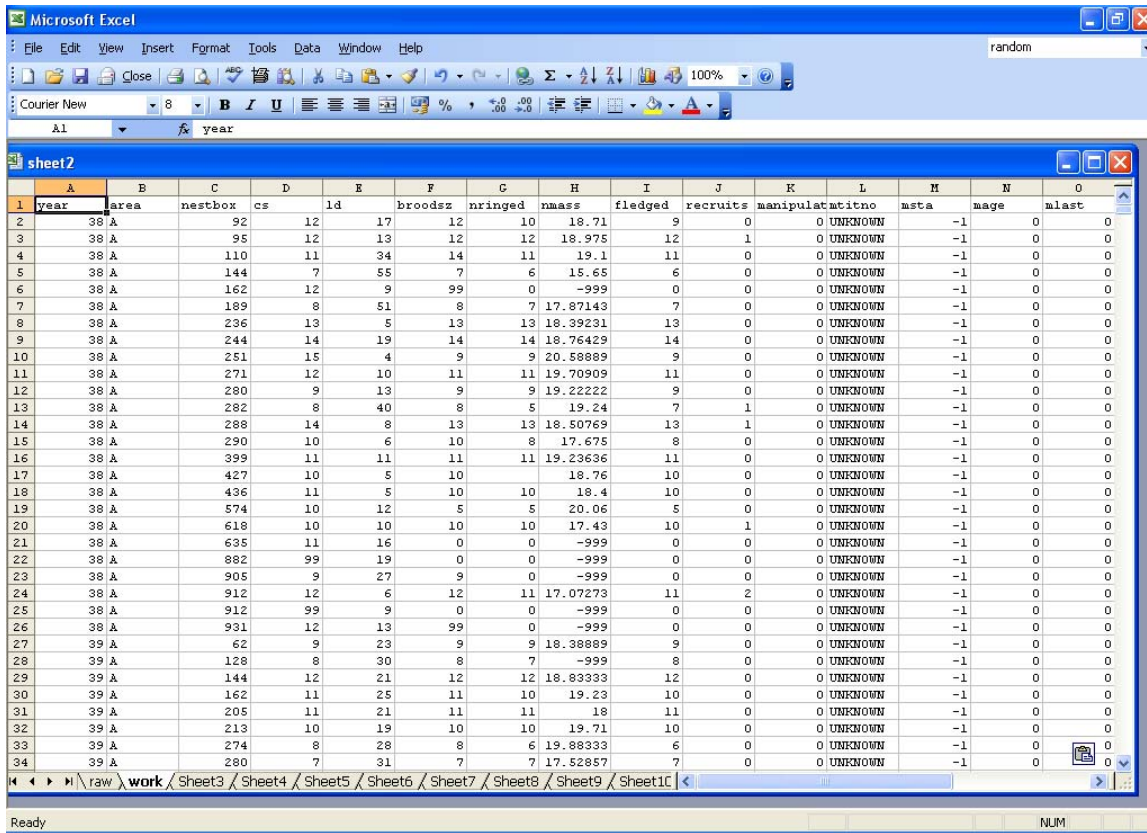
The Edit Menu

This may seem a very straight-forward menu, lifted straight out of MS Word. However, there are little foibles and tricks associated with this menu, so it's worth going through the menu in some detail.

→ *Make sure Sheet2.xls is open. Double click where Sheet1 is: you should now be able to write "raw". Now go back up to A1 and click on the little rectangle in the top left of the worksheet i.e. where the rows and columns meet. The whole sheet should become "blued out" i.e. selected.*

→ *Now using either Ctrl+C or Edit, Copy, copy the selected columns and Paste (Ctrl+V) into Sheet2 by tabbing on Sheet2 to bring it the fore, placing the cursor in cell A1, and then pasting. Call sheet2 "Work". Hit SAVE.*

OK - A FUNDAMENTAL LESSON – NEVER, EVER MANIPULATE YOUR HARD EARNED, RAW DATA. ALWAYS, ALWAYS, ALWAYS COPY IT & PASTE IT ELSEWHERE BEFORE YOU START MANIPULATING IT. ALSO, TRY & GET IN THE HABIT OF ALWAYS HAVING A COMMON REFERENCE COLUMN IN BOTH RAW DATA AND THE DATA TO BE MANIPULATED.



Let's spend a bit of time on copying and pasting. Staying in Sheet2, place the cursor over A1 and click. Note this cell is now highlighted as discussed earlier. Now hit Ctrl+C: the cell outline now has a broken moving line. Move to an empty cell in the worksheet, click and Paste (either Menu, Edit, Paste or Ctrl+V). However, if you use the Menu, you will notice that immediately below Edit, Paste is Paste Special. What is this?



We will come back to Paste Special, but suffice to say here, that when we get on to more advanced usage, especially when employing formulae, knowing whether to simply Paste or select one of the Paste Special options is crucial to retaining the integrity of your worksheet. Another crucial bit of advice for retaining integrity: Excel will generally paste wherever you place the active cell: if there are cells nearby that will get “covered” by the area pasted, then Excel will paste over these already occupied cells and you will “lose” the original data. Just one reason why you should NEVER manipulate your raw data (“Oh I’ll just do a quick look-see, cutting & pasting – I’ll remember” The phone goes, a beer calls ... You come back to your spreadsheet, closing it down & hitting Save when prompted – Whoops – goodbye data).

To copy all the data in the “raw” worksheet into the new “work” worksheet, I asked you to place the cursor over the top left hand corner of the “raw” worksheet, click, and then copy, before pasting into “work”. *Is everybody clear about this action?*: it selects **all** data in the worksheet. There are some very good reasons for doing this (i.e. selecting all data), especially when sorting columns, which we’ll come to later.

→ ***But for now, I want you to select the first five columns of data in “raw”, for the years 40 – 49 inclusive, and paste these into worksheet3.***

Any problems? OK – who bothered 1) to check that the data were in fact sorted by year, and 2) to add column headings so that you know what it is you copied? There are also two ways of doing the selection: either using the Shift+Arrow keys, or selecting with the mouse cursor. Whichever is easier. Note too I asked you to simply select and then paste the cells: I did not say **Cut and Paste** *versus* **Copy and Paste**. Is everybody clear about the differing consequences of these two actions. Please don’t be shy – let me know!

OK – Copy and Insert

→ ***Two separate activities here:***

1) Select cells A1:E100, Copy, and Insert at G1. What happens?

2) Select Column C, Copy, and

a) Paste into Column D, and

b) Insert into Column G. What happens?

Stuck? Let me be more specific:

→ ***Select cells A1:E100. If you have done this selection correctly, there should be “marching ants” circumscribing the outer boundaries of the cells you have selected.***

Any problems with selection?

→ ***Now go to Menu, Edit, Copy, OR go to Copy toolbar button, OR right click and select Copy.***

Any problems with Copy?

→ ***To Paste in Column D, go to D1, place cursor over it & select, and then Menu, Edit, Paste, or select Paste from Toolbar button, or right click and Paste.***

Whoops – you’ve pasted over other existing data. PANIC. Use the Undo button or go to Menu, Edit, Undo.

Nothing ventured, nothing gained, but is one lesson why to leave your RAW data unmanipulated!

→ **OK, still with A1:E100 copied, place cursor on D1 as above, BUT NOW go to Menu, Insert, Copied cells. Note it kindly asks if you want to move cells across or down. Often a clear warning you’ve done something wrong! But this time, insert the cells and see what you’ve done.**

→ **Finally, select all the cells in columns A and B. There are two ways of doing this: a) Place cursor on A1 and holding Shift key down, use down arrow and right arrow to select cells (this may take a bit of practice, but is very quick), or b) use cursor in normal way, or c) absolutely quickest – simply “black out” cells in Columns A & B by placing cursor over these two columns at juncture of column heading and title bar.**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	year	area	nestbox	cs	ld	broodsz	nringed	nmass	fledged	recruits	manipulat	mtino	msta	mage	mlast
2	38 A		92	12	17	12	10	18.71	9	0	0	UNKNOWN	-1	0	0
3	38 A		95	12	13	12	12	18.975	12	1	0	UNKNOWN	-1	0	0
4	38 A		110	11	34	14	11	19.1	11	0	0	UNKNOWN	-1	0	0
5	38 A		144	7	55	7	6	15.65	6	0	0	UNKNOWN	-1	0	0
6	38 A		162	12	9	99	0	-999	0	0	0	UNKNOWN	-1	0	0
7	38 A		189	8	51	8	7	17.87143	7	0	0	UNKNOWN	-1	0	0
8	38 A		236	13	5	13	13	18.39231	13	0	0	UNKNOWN	-1	0	0
9	38 A		244	14	19	14	14	18.76429	14	0	0	UNKNOWN	-1	0	0
10	38 A		251	15	4	9	9	20.58889	9	0	0	UNKNOWN	-1	0	0
11	38 A		271	12	10	11	11	19.70909	11	0	0	UNKNOWN	-1	0	0
12	38 A		280	9	13	9	9	19.22222	9	0	0	UNKNOWN	-1	0	0
13	38 A		282	8	40	8	5	19.24	7	1	0	UNKNOWN	-1	0	0
14	38 A		288	14	8	13	13	18.50769	13	1	0	UNKNOWN	-1	0	0
15	38 A		290	10	6	10	8	17.675	8	0	0	UNKNOWN	-1	0	0
16	38 A		399	11	11	11	11	19.23636	11	0	0	UNKNOWN	-1	0	0
17	38 A		427	10	5	10		18.76	10	0	0	UNKNOWN	-1	0	0
18	38 A		436	11	5	10	10	18.4	10	0	0	UNKNOWN	-1	0	0
19	38 A		574	10	12	5	5	20.06	5	0	0	UNKNOWN	-1	0	0
20	38 A		618	10	10	10	10	17.43	10	1	0	UNKNOWN	-1	0	0
21	38 A		635	11	16	0	0	-999	0	0	0	UNKNOWN	-1	0	0
22	38 A		882	99	19	0	0	-999	0	0	0	UNKNOWN	-1	0	0
23	38 A		905	9	27	9	0	-999	0	0	0	UNKNOWN	-1	0	0
24	38 A		912	12	6	12	11	17.07273	11	2	0	UNKNOWN	-1	0	0
25	38 A		912	99	9	0	0	-999	0	0	0	UNKNOWN	-1	0	0
26	38 A		931	12	13	99	0	-999	0	0	0	UNKNOWN	-1	0	0
27	39 A		62	9	23	9	9	18.38889	9	0	0	UNKNOWN	-1	0	0
28	39 A		128	8	30	8	7	-999	8	0	0	UNKNOWN	-1	0	0
29	39 A		144	12	21	12	12	18.83333	12	0	0	UNKNOWN	-1	0	0
30	39 A		162	11	25	11	10	19.23	10	0	0	UNKNOWN	-1	0	0
31	39 A		205	11	21	11	11	18	11	0	0	UNKNOWN	-1	0	0
32	39 A		213	10	19	10	10	19.71	10	0	0	UNKNOWN	-1	0	0
33	39 A		274	8	28	8	6	19.88333	6	0	0	UNKNOWN	-1	0	0
34	39 A		280	7	31	7	7	17.52857	7	0	0	UNKNOWN	-1	0	0

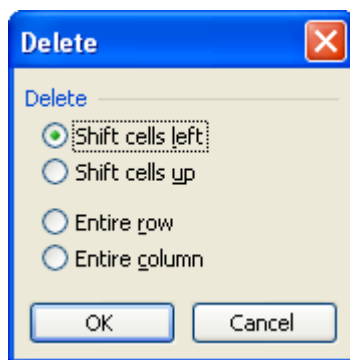
A WORD OF WARNING: whilst the shift arrow keys are quickest way of selecting data, you can run into problems if your sheet has empty cells: the selection process stops as soon as it detects an empty cell. When one is doing a lot of manipulations, it is very easy to select 200 rows of data instead of 2000, simply because cell A201 was blank for some reason. This is one reason why I don’t like blank cells – I use -999 for missing values.

OK – back to example – copy the cells you’ve selected, and then “black out” columns F & G: Go to Menu, Insert, Copied Cells. Any problems??

Deleting and Clearing

Often when working on spreadsheets, one makes mistakes in data entry (and may I recommend that if you work on a **desktop PC**, use **Num Lock** and the Number keyboard – it is much more efficient. I will not comment on pretentious laptops and PDAs (and real fieldworkers use pen and paper ...)). Such mistakes are easily corrected using normal editing procedures. In this section, I want to talk about the removal of a number of cells, for whatever reason – they may be wrongly entered, superfluous, or simply in the wrong place.

→ *Select cells A2:A26. Go to Menu, Edit, Delete. The following option box appears*



If we wanted, for whatever reason, to get rid of all data pertaining to year 38, we would select “Entire row” and go OK.

→ *Do it. What has happened?*

OK – reverse the action (i.e. Undo) and make sure cells A2:A26 are selected. Go to delete again, but this time go with the default “Shift cells sideways”. Comments please!

→ *Yup – a pretty stupid action, since we now have cells displaced and no longer sitting in their proper columns. So “Undo”, ensuring cells still selected, and go back to Menu, Edit, but this time select Clear. A range of options pop up, but generally Clear All suffices. NOTE – you could also select the cells A2:A26 and simply right-click with your mouse, and the option to Clear Contents would be available.*

Often, one copies all ones raw data over into a new worksheet, but the analyses one is interested in may only refer to maybe half a dozen of 20 or more columns. It is generally easier to work on the pared down worksheet: to do this, simply delete the unwanted columns.

→ *Copy all of worksheet “Work” using what you have learnt above, and Paste these data into the next available worksheet in the Workbook (i.e. Sheet3)*

→ *Delete the Columns Area, bs, and everything to the right of the “manipulation” column, with the exception of the last column, “pnum”. This worksheet should now have 10 columns (NOTE BENE (i.e. This is Latin for “wake up”) – if you place your*

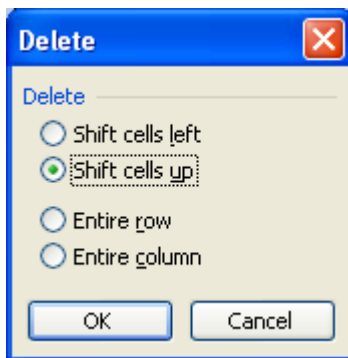
cursor in cell A1 and select by left clicking, and then still with left button pressed down, scroll across the columns, you will notice the number of columns (and rows) being clocked up in the “name box”. The same happens when you convert the cursor to the downward pointing arrow and scroll across the columns. This counting can be useful.

→ *OK, label this sheet “work1”.*

FURTHER MANIPULATION OF DATA SHEETS

We are going to select twenty years worth of data, namely for the years 61 – 80.

→ **Firstly though, using worksheet “work1”, I want you to select the cells in all ten columns for the years 38, 39 and 40. Go to Menu, Edit, Delete.**



→ *Should we go with this and shift the cells up?* In this instance, this action makes sense – none of the cells move into the “wrong columns”

→ *Now delete the remainder of the cells as you see fit. Remember, we are wanting a worksheet that only holds the data for the years 61 – 80. Hint: above we have discussed two ways of selecting data: the first is scrolling down ONE column, but selecting Delete Entire Row, and the other is to place cursor on a cell, and holding Ctrl&Shift, use the Down Arrow. Comments? Help!!?*

→ *What are the dimensions now of “work1” (i.e. how many columns × rows)?*

OK, by using the arrow keys and CtrlShift, we can find the dimensions very easily.

Another activity frequently required when using Excel is the insertion of cells into an already “formatted” worksheet. Generally (but I stress not always) one needs to insert columns or rows, not individual cells. However, whilst “work1” is open,

→ *try inserting a small matrix of cells, namely with dimensions 6 cells across and 3 cells down (ie 18 cells in total). Use A2 as the top left cell.*

→ *What happens? What problems are apparent?*

OK – clearly this is not a useful action to undertake in this specific worksheet, and I would be hard-pressed to think of a reason why one might wish to do this specific action.

→ ***Cancel the above action by using Menu, Edit, Undo or using the Toolbar Undo button. Inserting Rows or Columns is a far more common action when using Excel.***

There are two ways of doing this: insert cursor on cell where one wishes to insert a blank row or column, & then go up to Menu, Edit, Insert (Rows or Columns), or “black out” the row or column where one wishes blank cells to be inserted, and repeat above command sequence.

→ ***Add a new column which will run down as Column C. Undo action (when you’ve got it right).***

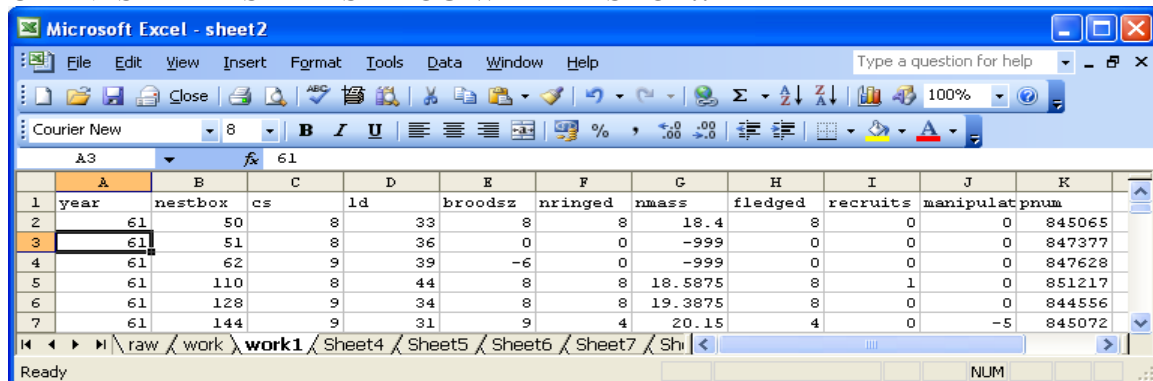
→ ***Add five columns running from C to G as new, blank columns. Undo action.***

Note the last action (adding five columns) is easily done by “blacking out” columns C:G and going to Insert.

One last activity before we get onto the more interesting & useful bits of Excel:

FORMATTING SPREADSHEETS

This is not an activity I want to spend much time on, other than saying that Courier New Fontsize 8 is IDEAL for using in spreadsheets. However, if you wish to use far-out Chick or Pusycat font or whatever – feel free to do so – JUST DON’T EVER USE ON THE CRANES’ DATASHEETS – YOU WILL BE SHOT!!



Firstly, the Font and Fontsize of the current worksheet is displayed in the Standard Toolbar boxes as illustrated above. To change the Font, Fontsize, Colour etc, click on the cell(s) one wishes to change, and then go to Menu, Format, Cells, and play around with the options presented. More usefully, to make universal changes to your worksheets (ie ones that will apply to all new worksheets,

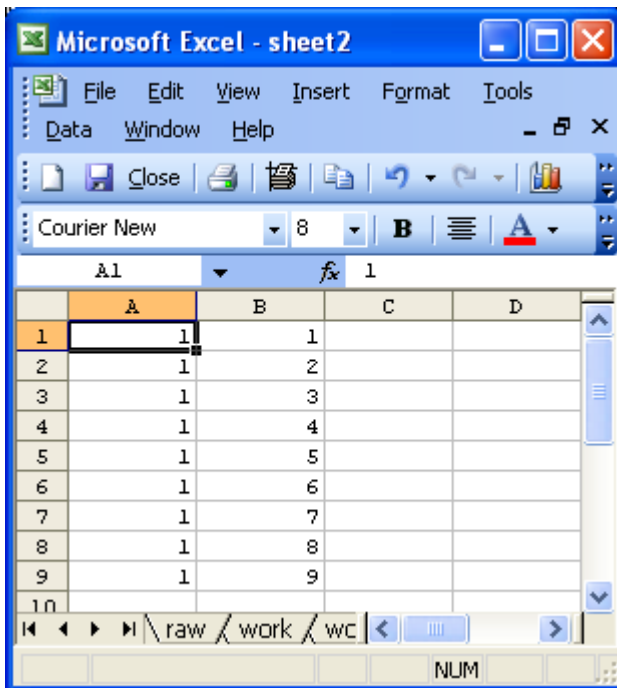
→ ***Go to Tools, Options and Select the General tab. Make the changes as appropriate (i.e, change to Courier New, size 8 ...!)***

MAKING LIFE EASIER: THE FILL FUNCTION

One of the “problems” with electronic data collection, is that initially one assumes that “common” data needs to be repeatedly entered, line after line, whereas in a notebook with pen & paper, one simply writes the date once for the day, and gets on with the observations. In fact, with a bit of fore-thought (see designing a datasheet above) it is possible to automate many repetitive actions.

→ **Open a new worksheet** – either a new workbook (**Menu, Open, New**) or a new worksheet in your current workbook (“sheet2.xls”)

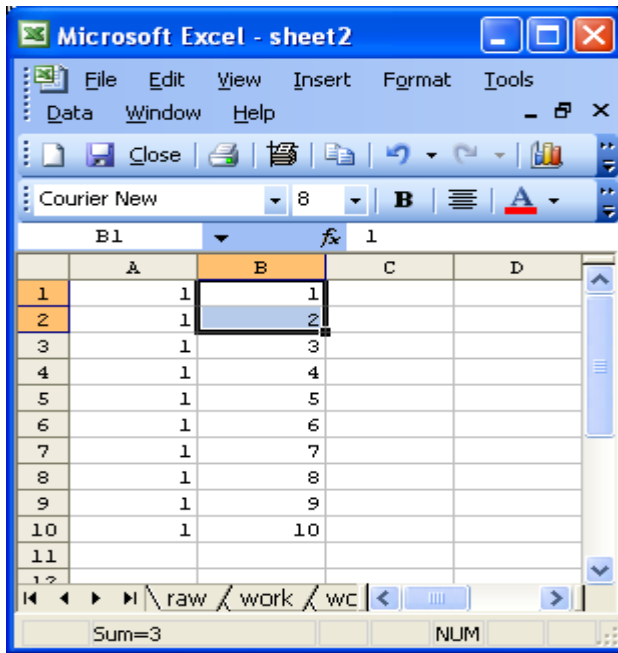
→ **Position your cursor in A1 and make this the current cell. Enter 1 and hit Return. Now select the cell A1 and get a “handle” in the cursor (i.e. the cursor symbol (normally an arrow or “open cross”) turns into a “filled” cross or heavy plus sign) when the cursor “hovers” over the bottom right corner of the active cell (where the cell border turns from a line into a small square).**



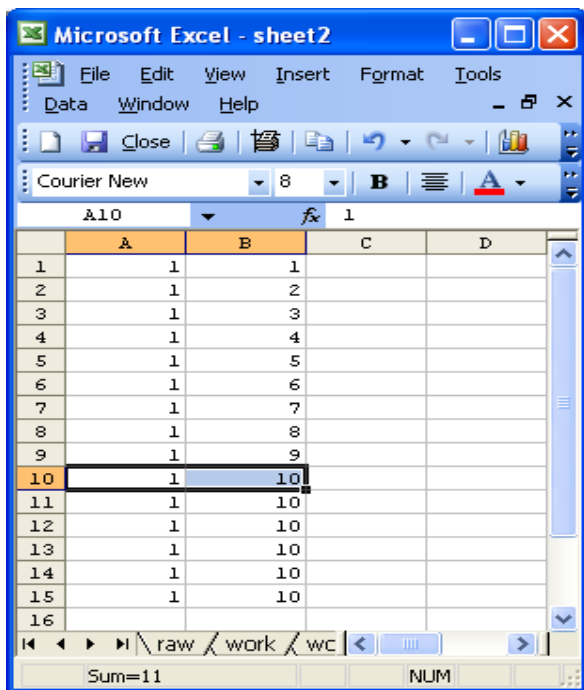
→ **Does everyone have this? OK, left-clicking and holding on your mouse, together with the “solid cross”, drag this down to cell A10. You should get a column of 1’s in Column A.**

Has everybody got this action? – it is important

→ **OK – repeat as above, except this time, type 1, then 2 in cells B1 and B2 respectively. Select both these cells, and drag on handle down to B10: B10 should have “10” entered automatically. Did you notice anything else whilst doing this?**



→ *OK – place the cursor over cells A10:B10. Select these and drag down. What happens now?*



→ *Now try typing “January” in cell C1, and drag down. In D1, type today’s date, and drag down.*

	A	B	C	D	E	F	G	H
1	1	1	January	12.3.07	12/03/2007	12/03/2007		
2	1	2	February	12.3.08	13/03/2007	13/03/2007		
3	1	3	March	12.3.09	14/03/2007	14/03/2007		
4	1	4	April	12.3.10	15/03/2007	15/03/2007		
5	1	5	May	12.3.11	16/03/2007	16/03/2007		
6	1	6	June	12.3.12	17/03/2007	17/03/2007		
7	1	7	July	12.3.13	18/03/2007	18/03/2007		
8	1	8	August	12.3.14	19/03/2007	19/03/2007		
9	1	9	September	12.3.15	20/03/2007	20/03/2007		
10	1	10	October	12.3.16	21/03/2007	21/03/2007		
11	1	10	November	12.3.17	22/03/2007	22/03/2007		
12	1	10	December	12.3.18	23/03/2007	23/03/2007		
13	1	10				24/03/2007		
14	1	10				25/03/2007		
15	1	10				26/03/2007		
16						27/03/2007		
17						28/03/2007		
18						29/03/2007		
19						30/03/2007		
20						31/03/2007		
21						01/04/2007		
22						02/04/2007		
23						03/04/2007		
24						04/04/2007		
25						05/04/2007		
26						06/04/2007		
27						07/04/2007		
28						08/04/2007		

As you can see – Excel needs to “predict” your needs: to fill down simply what is written (ie copy down or across as in Column A), or predict you are wanting a series, and so filling down sequentially as in months of the year (e.g. Column C). Note too the specific format is essential: Excel has not read Column D as a date, but some numbering system, whereas Column E is written in “Date Format”, and Excel has responded in kind. The Final column (F) was written using the Menu, Edit, Fill, Series command, going up to 13 October 2007. Quite how Excel decides whether to fill in by copying down (or across) is a bit of a mystery to me: generally I give Excel a helping hand by entering the first 3 numbers in a series (1.0; 1.1; 1.2 or 5; 10; 15), selecting these three cells, and then dragging down or across until I reach the end-point I want.

→ *Play around with this function until you are happy with how it works.*

However, I can hear some of you muttering that, yes, this Fill Down (or Up or Across) function may be useful, but what happens when I have 10 observations on Day X and 5 observations on Day Y etc. Not a problem. Let's say Day X = 13/03/2007 and Day Y = 14/03/2007.

→ *Go back to the above and copy 13/03/2007, or simply type in 13/03/2007 and copy it. Then select the 10 cells where you wish to paste it, and Paste. Ditto 14/03/2007, except here you select on five cells for Pasting. Below is an example for how your sheet should look just prior to pasting 13/03/2007*

	A	B	C	D	E	F	G	H
1	1	1	January	12.3.07	12/03/2007	12/03/2007		
2	1	2	February	12.3.08	13/03/2007	13/03/2007		
3	1	3	March	12.3.09	14/03/2007	14/03/2007		
4	1	4	April	12.3.10	15/03/2007	15/03/2007		
5	1	5	May	12.3.11	16/03/2007	16/03/2007		
6	1	6	June	12.3.12	17/03/2007	17/03/2007		
7	1	7	July	12.3.13	18/03/2007	18/03/2007		
8	1	8	August	12.3.14	19/03/2007	19/03/2007		
9	1	9	September	12.3.15	20/03/2007	20/03/2007		
10	1	10	October	12.3.16	21/03/2007	21/03/2007		
11	1	10	November	12.3.17	22/03/2007	22/03/2007		
12	1	10	December	12.3.18	23/03/2007	23/03/2007		
13	1	10				24/03/2007		
14	1	10				25/03/2007		
15	1	10				26/03/2007		
16						27/03/2007		
17						28/03/2007		
18						29/03/2007		
19						30/03/2007		
20						31/03/2007		
21						01/04/2007		
22						02/04/2007		
23						03/04/2007		
24						04/04/2007		
25						05/04/2007		
26						06/04/2007		
27						07/04/2007		
28						08/04/2007		

There is one last command I want to deal with in this section, and that is **transpose**.

Transposing data

	A	B	C	D	E	F	G	H	I	J
1	month	day	count	average		month	January	January	January	April
2	January	01/01/2007	321			day	01/01/2007	15/01/2007	30/01/2007	01/04/2007
3	January	15/01/2007	198			count	321	198	302	412
4	January	30/01/2007	302	274		average			274	
5	April	01/04/2007	412							
6	April	15/04/2007	216							
7	April	30/04/2007	299	309						
8	July	01/07/2007	356							
9	July	15/07/2007	195							
10	July	30/07/2007	276	276						
11	October	01/10/2007	352							
12	October	15/10/2007	301							
13	October	30/10/2007	278	310						
14										

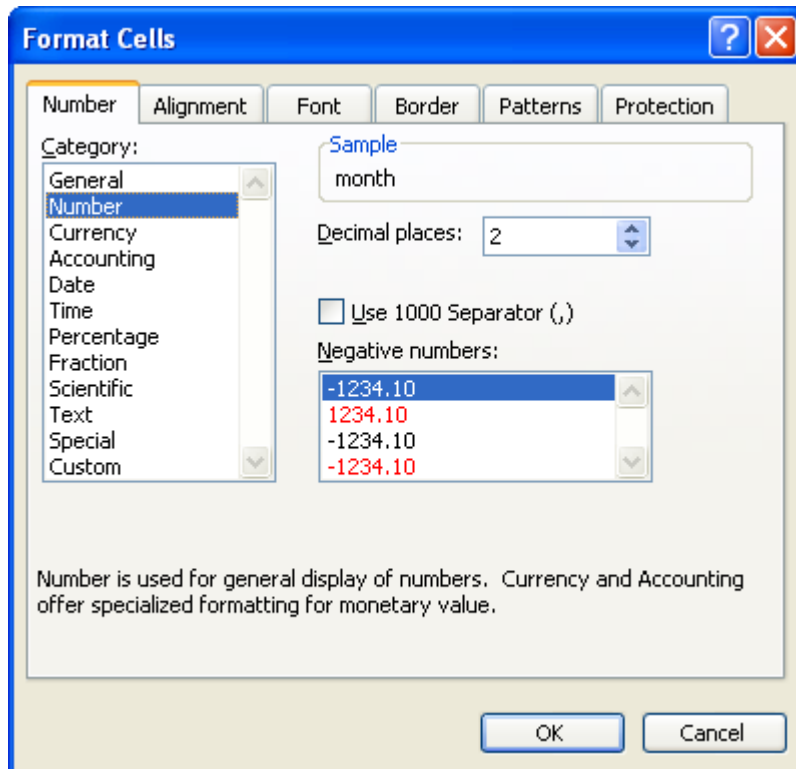
There are occasions when you have collected and entered data in the “correct” format (generally similar data (eg counts) running down the same column), but either in order to summarise data for a report, or for graphical or some other purpose, you need the data to be “spread out” at 900 to how you entered it. In other words, you wish to transpose the data. Look at the lay-out above – a series of counts may quarterly, on three different days each quarter. In your report to Cranes Ltd, you wish to have a table simply showing the quarter concerned and the average count for each quarter. You collected the data in the form shown above in Columns A:C, and you calculated the average counts per quarter in Column D.

→ *Create the sheet above, entering the data and column headings A1:D13. Don’t simply copy each cell (ie write out each cell), but see if you can think of short cuts to make this easier. Once you have the data in the format above for cells A1:D13, copy these cells, place the cursor in F1, and then go to Menu, Edit, Paste Special. In the box, tick on Transpose, and then OK.*



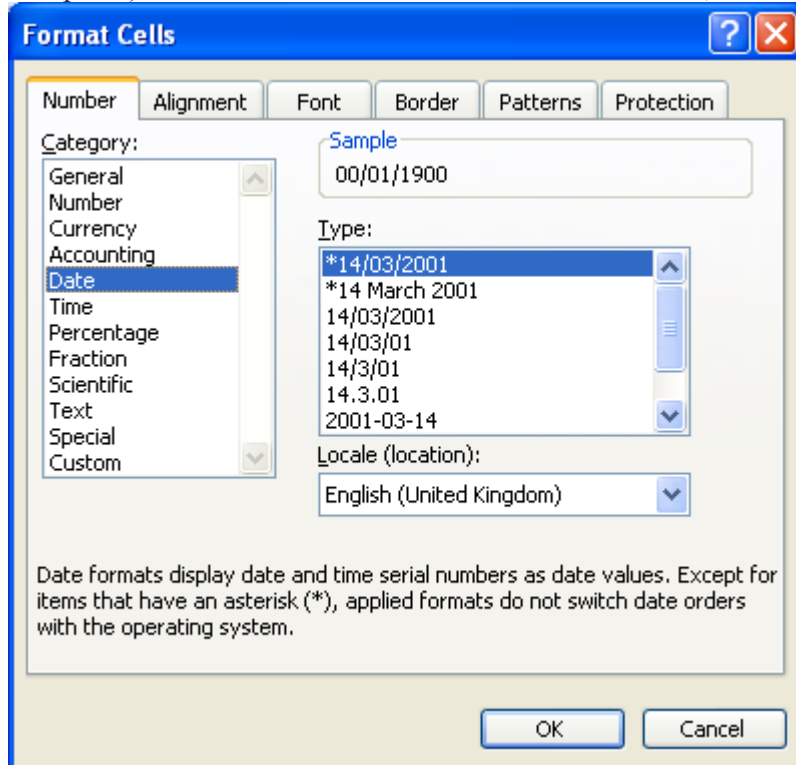
FORMATTING CELLS AND COLUMNS

As in any MS Office suite, one can change formatting, colour, shading etc. However, as I've indicated above, I am unable to read anything that is not in Courier New, 8 point! More seriously, Excel *differs* from Word and the other packages in that it allows the **type of data** in a column to be specified.

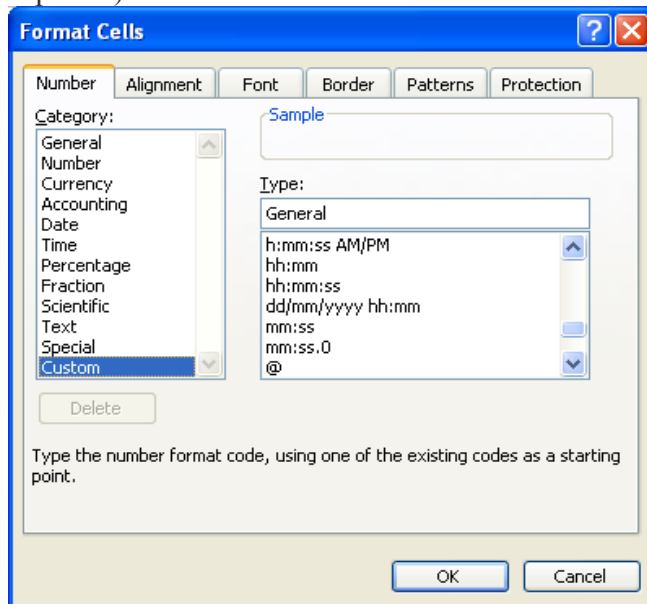


Whether one goes to an individual cell, or, more usefully, a whole column or even worksheet, one can call up the following formatting dialogue box. I will not cover each and every format, but concentrate on the three major ones which will concern us in terms of data collection in the field. The first is the **General** format, and is the one I would use for entering text such as name of farms, provinces, species names etc. The **Number** format is clearly key too. By and large, one does not need to *display* an accuracy greater than two decimal points – often whole numbers make more sense (we would not be interested in a fifth of a crane – round off to whole numbers for count data. Note too, that even if you do set the number of decimal points to 0, and the figure appears as a whole number, Excel is actually *storing* the full string of decimals. Thus π (an irrational number) will be stored as 3.14159265, even though you may specify only 2 decimals (it will thus appear as 3.14), or as 3 if zero decimals are specified. Sometimes, people will send you *.xls files to import, where number are given in what is called **scientific** notation. This is simply a number raised to an exponent e.g. 0.001 may appear as 1.00×10^{-3} . In Excel, this appears as 1.00E-03. Convert this to something readable, namely using the Number format!

The formats which cause me the greatest fun and games (also known as swearing at the computer) are the Date and Time fields. In themselves, these look innocuous enough:

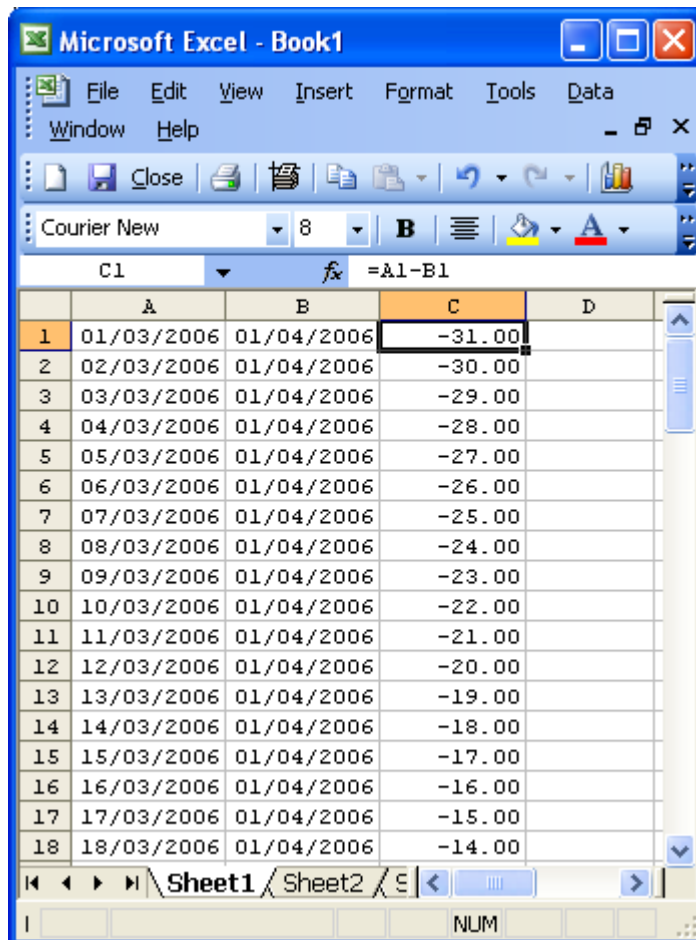


For reasons best known to Excel, I sometimes carefully format my column as a date field, choosing one of the above patterns for entry, only for Excel to reject that what I have entered is a date (or time or whatever – note – under Format cells, Number, one can add date and time as single cell entry (see below) – my advice would be to keep them separate)



The advantage of explicitly specifying date and time fields, is that Excel allows one to carry out basic mathematical operations. In the Wytham tit study for example, lay-date (date of first egg) is calculated relative to 1 April each year. Thus a bird laying its first egg on 15 April is given a lay-date of 15 etc. This is to enable subsequent statistical analysis, such as how clutch-size varies with lay date in this species: in most statistical packages, one needs a numerical value to determine this relationship, not a date value. Variation in reproductive parameters with date is a very common phenomenon, and is one that we actually need to get our heads around with respect to the cranes. But for now, let's just practice what we have preached (or been preached to!).

→ *In a new spreadsheet, enter the 1st March 2006 in whatever format you like that Excel recognizes as a date field. Drag down until 1st June 2006. Create a new, adjoining column, which also has 93 entries, but each needs to read 1st April 2006. (There is a quick tip here: enter and copy 01/04/2006, and then holding Shift+Ctrl down, use the down arrow: this should “block out” the 93 cells you require – now hit Ctrl+V to paste.)*



	A	B	C	D
1	01/03/2006	01/04/2006	-31.00	
2	02/03/2006	01/04/2006	-30.00	
3	03/03/2006	01/04/2006	-29.00	
4	04/03/2006	01/04/2006	-28.00	
5	05/03/2006	01/04/2006	-27.00	
6	06/03/2006	01/04/2006	-26.00	
7	07/03/2006	01/04/2006	-25.00	
8	08/03/2006	01/04/2006	-24.00	
9	09/03/2006	01/04/2006	-23.00	
10	10/03/2006	01/04/2006	-22.00	
11	11/03/2006	01/04/2006	-21.00	
12	12/03/2006	01/04/2006	-20.00	
13	13/03/2006	01/04/2006	-19.00	
14	14/03/2006	01/04/2006	-18.00	
15	15/03/2006	01/04/2006	-17.00	
16	16/03/2006	01/04/2006	-16.00	
17	17/03/2006	01/04/2006	-15.00	
18	18/03/2006	01/04/2006	-14.00	

→ *I now want you to use a very simple formula in Excel. In C1, simply type “=A1-B1” and hit return. This should yield -31. Go back to cell C1 and place the cursor in*

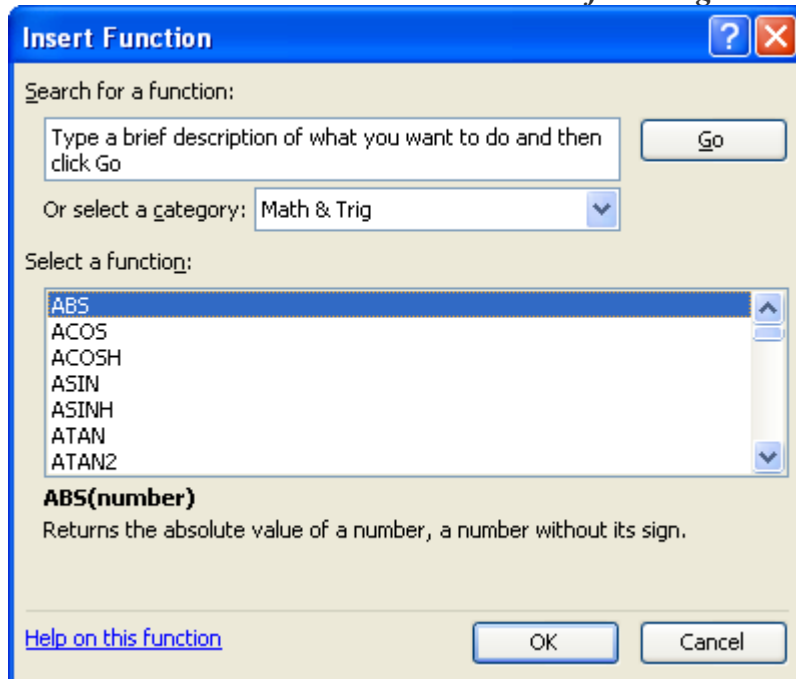
the cell: the formula should now show, PLUS each cell you have specified an action for will be highlighted, in this case A1 and B1. I now want you to go to cell C2. Enter “=” (all formulae and functions need the “equals” sign to operate in Excel) and then place the mouse cursor on A2, click, type in “-“ (minus), and then click on B2. Hit return and you should have an answer of -30. Now COPY C2, and paste in the remainder of the relevant cells (i.e. C3:C93).

SIMPLE FORMULAE IN EXCEL

Remember early on in this manual I said that what gave Excel its unique power was that each of the 17 million odd cells in a worksheet each had their own unique address or ID. This enables the contents of any one cell to “interact” with the contents of another, using a very simple language (essentially “Visual Basic”). So, as above, where we wished to find the difference between two cells, it simply meant writing in a new cell that the new cell would hold the answer (hence the “=” sign) to Cell ID 2 subtracted from Cell ID 1 (e.g written as “= A1-B1). Essentially, what we are doing, is asking Excel to **solve an equation** written using **specified functions**.

Just before the **formula bar** (which expands and shows you the specified function you type) you will see the symbols \times , \checkmark , and f_x . It is the f_x that is the **function** key.

→ *Place the cursor over it & click. The following will come up*



Alternatively, if stuck, go to the following:

List of worksheet functions (by category)

[+ Show All](#)

The following sections list all of the worksheet functions by category.

In this topic

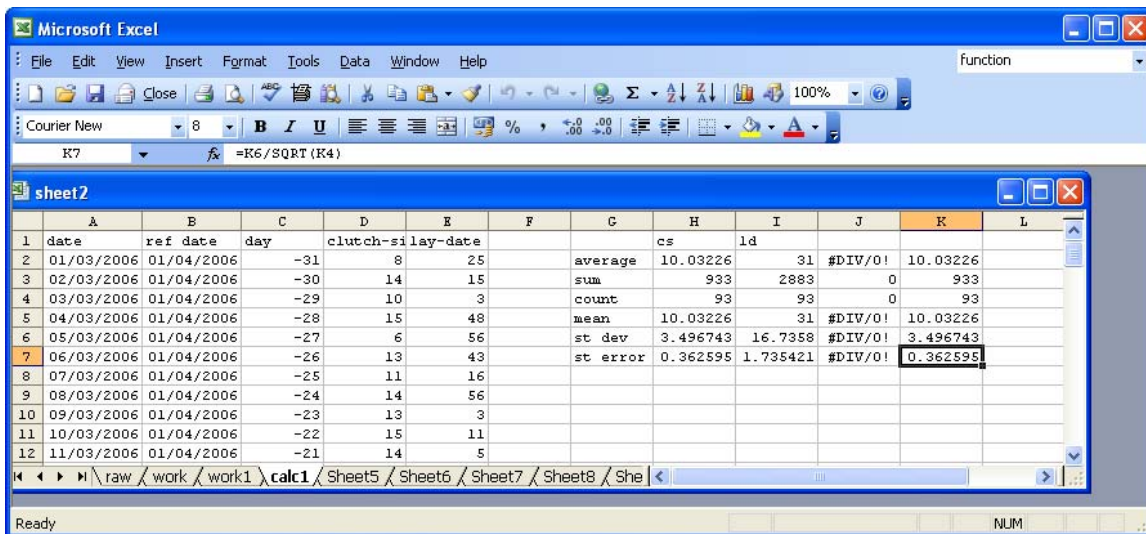
[Database functions](#)[Date and time functions](#)[Engineering functions](#)[Financial functions](#)[Information functions](#)[Logical functions](#)[Lookup and reference functions](#)[Math and trigonometry functions](#)[Statistical functions](#)[Text functions](#)[External functions](#)

Clicking on each of these headings will bring up the relevant function call. Another useful set of help instructions and aide mémoire are given on the next page. For now though, let's just keep to the more useful and hopefully simple functions.

Probably the commonest function we use as field biologists is the mean: what is mean count in a month, what is the mean clutch-size etc. In Excel, this is calculated by the **AVERAGE** function: =average(cell refs). Note cell refs refers simply to the matrix of cells that one is interested in obtaining a function for e.g. A1:A10 or A1:B33.

→ *Open Sheet2.xls and tab open sheet calc1. The first three columns simply retain the exercises with the date function we carried out above. I have then made up two additional rows of data, namely clutch-size and lay-date. I used the RANDBETWEEN function to give me unique values lying between 5 and 15 for clutch-size and 0 and 60 for lay-date. Click on one of these cells in columns D or E and you will see the formula. All I have instructed Excel to do is choose, at random, a number between 5 and 15 (or 0 and 60) in each cell that the formula is copied into.*

→ *Go to cells in the matrix H2:I7. Place the cursor over each of the cells and see if you follow what I have done i.e. why I have chosen the specific function, and what it is returning. Look at Column J: What has happened here? Ditto Column K?*



Earlier I mentioned the different ways of specifying a cell’s identity or address, namely A1, \$A1, A\$1 and \$A\$1. Each of these is referring to the identical cell – what differs is the way that identity or address behaves when it is dragged or copied to a new location. Thus if “=A1” is written in cell B1, this will return the contents of A1. If B1 is then dragged into C1, the formula changes from “=A1” to “=B1” Similarly, if one were to drag B1 into B2, then B2 would contain “=A2”, In other words it automatically updates its address. This can be extremely useful; it can also give you totally the wrong results!

→ *Now place your cursor in the cells in the matrix G11:J11, and drag DOWN four cells (ie to G12:J12). What has happened? Keep playing with these different ways of writing addresses until you are confident of what you are doing and why they give different results. CHECK TOO the difference when these cells are DRAGGED as opposed to COPIED AND PASTED.*

Examples of commonly used formulas

Type of Formula	Example
Conditional	<ul style="list-style-type: none"> Create conditional formulas Check if a number is greater than or less than another number Display or hide zero values Hide error values and error indicators in a cell
Lookup	Look up values in a list of data
Date and Time	Add dates

	<ul style="list-style-type: none">Add timesCalculate the difference between two datesCalculate the difference between two timesCount days before a dateShow dates as the day of the weekInsert the current date and time in a cellInsert Julian dates
Financial	<ul style="list-style-type: none">Calculate a running balanceCalculate a compound annual growth rate (CAGR)
Statistical	<ul style="list-style-type: none">Calculate the average of a group of numbersCalculate the median of a group of numbersCalculate the mode of a group of numbers
Math	<ul style="list-style-type: none">Add numbersSubtract numbersMultiply numbersDivide numbersCalculate percentagesRound a numberRaise a number to a powerCalculate the smallest or largest number in a rangeCalculate the factorial or permutation of a numberCreate a multiplication table
Counting	<ul style="list-style-type: none">Count cells that contain numbersCount nonblank cellsCount how often a value occursCount unique values among duplicatesCount numbers greater than or less than a numberCalculate a running totalCount all of the cells in a rangeCount the number of words in a cell or range
Conversion	<ul style="list-style-type: none">Convert timesConvert dates stored as text to datesConvert numbers stored as text to numbers

Convert measurements
 Convert numbers to different number systems
 Convert Arabic to Roman numerals

Text

Change the case of text
 Check if a cell contains text (case-insensitive)
 Check if a cell contains text (case-sensitive)
 Combine text and numbers
 Combine text with a date or time
 Combine first and last names
 Combine two or more columns by using a function
 Repeat a character in a cell
 Display only the last four digits of identification numbers
 Remove spaces and nonprinting characters from text
 Remove characters from text
 Insert the current Excel file name in a cell
 Split names by using Convert Text to Columns
 Split text among columns by using functions

A final note of caution when using formulae in Excel, and this is to do with the “operational precedence” that Excel employs in calculating formulae.

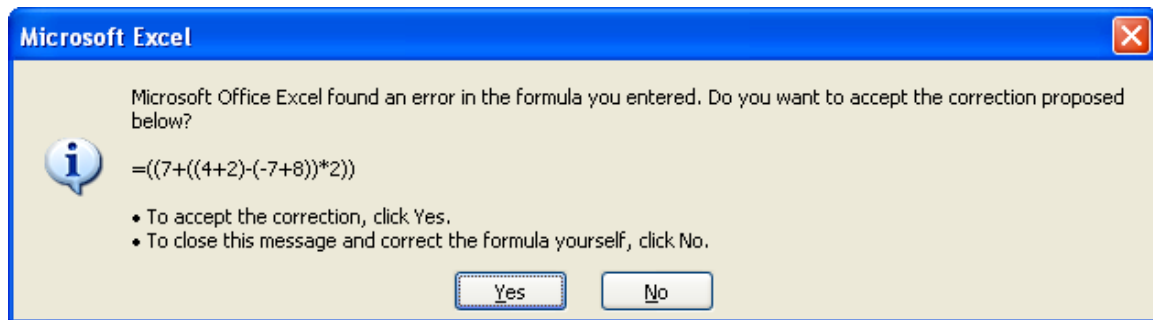
→ ***Find a block of empty cells, and type the following into each cell:***

$=5+6/3$
 $=5+(6/3)$
 $=(5+6)/3$

→ ***Can anybody work out why we are given different answers by Excel?***

Excel works by firstly calculating whatever is within brackets, and then follows a set precedence of multiplying first, then dividing, then adding and finally subtracting. The key to changing this precedence is by using brackets. And note these brackets can be nested (i.e. multiple bracketed expressions), with the “innermost” nesting being calculated first. Thus $((7+((4+2)-(-7+8)))*2)$ is solved sequentially by Excel as Step 1: $(-7+8)=1$; Step 2: $(4+2)-1=5$ (the “1” coming from Step 1); Step 3: $7+5 = 12$ (ie $7 +$ Step 2 Answer); and finally Step 4: $2 \times$ answer from Step 3 = $2 \times 12 = 24$. Note Excel uses the asterisk (*) as the multiplication symbol, and the slash sign (/) as divide. The placement of brackets, **especially when nesting**, is crucial. Excel helpfully clour-codes the brackets in sequence, and I suggest that you make a habit of using your mouse cursor in the formula bar (where it is a narrow vertical line), to check that your brackets are in the correct sequence and each argument is “bracketed off” (i.e. (argument)). Excel will also

helpfully tell you when you have insufficient numbers of brackets, and will actually suggest where to put them, thus



It is **VERY** tempting to simply click yes. **DO NOT do so.** Go back to the formula and manually check it yourself and amend accordingly, otherwise sooner or later there will be tears!

OK – we could spend all week running over the use of formulae etc in Excel. To be honest, unless you use these on a fairly regular basis you will forget them in no time. You can always refer back to these notes, and make use of the Help function in Excel (type in the table headings and these will pop up). Learning this part of Excel is hands-on, and down to you ...

But there is one very useful command that some of you may come to use frequently: the **pivot table**.

PIVOT TABLES

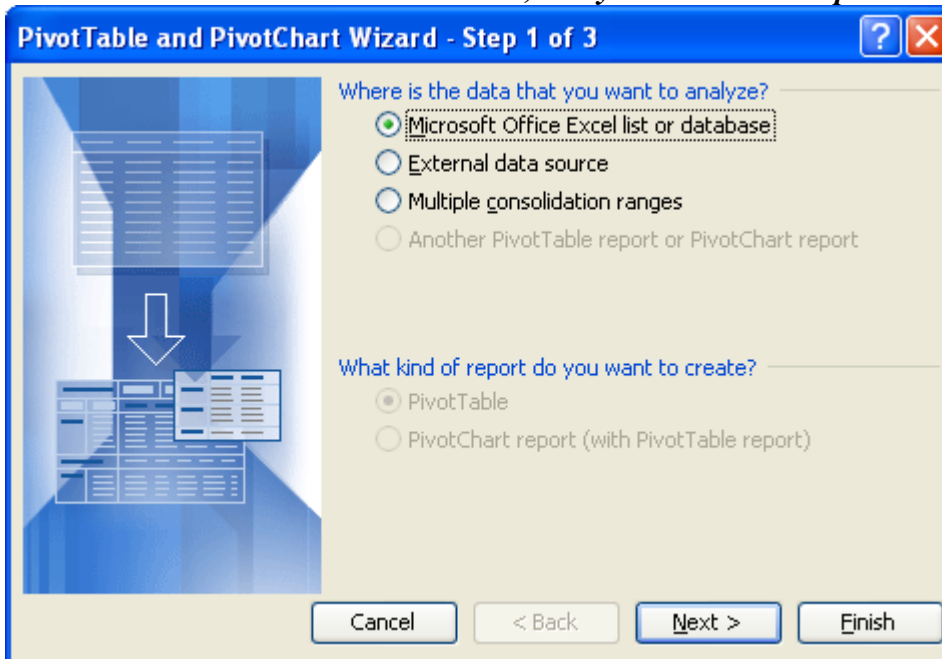
Pivot tables essentially summarise one's data in a form that one specifies. In Sheet2.xls calc1, we have, amongst other data, three columns, namely clutch-size, lay-date and week. Now there are some 93 observations of two continuous variables (clutch-size and lay-date), and a column containing categorical (factor) data, namely week (there are 13 of these categories – call them A, B, C, ... M if you like. Either for your own interest, or because Eskom are asking for a report, or whatever, you wish to see how the two continuous variables (CS and LD) may vary with week.

→ ***Given the information we have gone over above, how might we tackle this problem?***

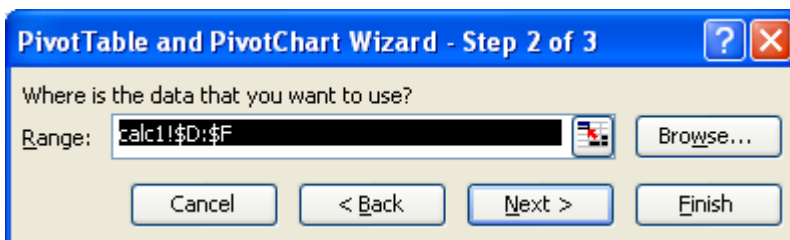
Yes, without too much problem we can sort the data according to week, and then use some of the mathematical and/or statistical functions to summarise the data. But we are on the cusp here of it becoming impractical – look at the sheet “raw” with 11,000+ observations, and imagine the time trying to summarise these data, say by year. **Pivot tables automate this process for you.**

In the sheet “pivot table”, I have already done a pivot table for you of the data in calc1.

→ *Place your cursor anywhere on the table in the sheet “pivot table”, and right click on the mouse. Third down in the listings is “Pivot Table Wizard”. Click on this. Now click TWICE on the BACK button, and you should land up with this screen*

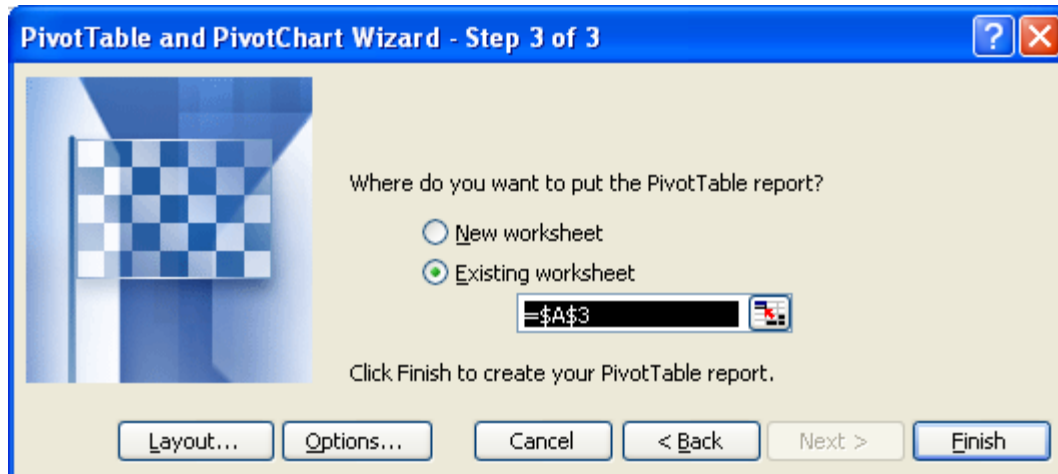


→ *Click on Next button, and we are in Step 2 of the Wizard.*



Here we define the Range of data that we are interested in. There is nothing magical about this: one is simply interested in the matrix of data that contain our 3 columns of CS, LD and week. The Sheet name is included too (calc1!). If you are “stuck” on the wrong sheet, try clicking on Browse, Cancel, and the “marching ants” should show you which data you have selected.

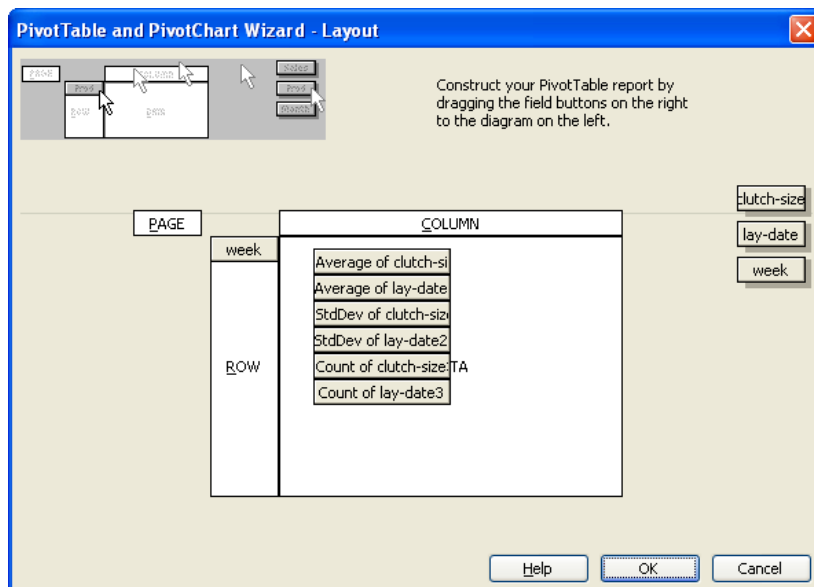
→ *Click on Next button, and we are at the crucial Step 3 of the Wizard. Click on Layout*



Everything you need to do is in this box:

What I have done is placed in the left-hand column (the row headings if you like) the column “week”. I did this simply by clicking on week in the box (see next page) and dragging it into row.

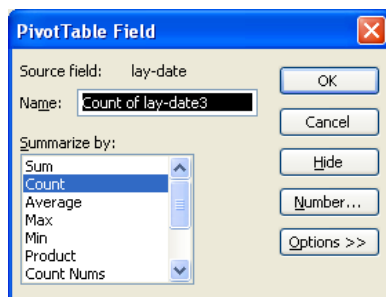
→ *Place your cursor over week, and drag it into the column section. Reverse the action. OK – what this has done is told Excel that we want our data summarized by week, irrespective of what week is. It could be any definition we give to week in the original table that Excel is collecting the data from. Thus we have defined week as running from 1 through 13, but we could have had A through M or donkeys through cats through cranes! So long as there is more than zero “categories” in the row or column data, then you will get summary data for these category(ies). IS EVERYBODY WITH ME?*



OK – so we have the **CATEGORIES** that we want to summarise the data by (ie 13 weeks). Also in the box above are the two continuous variables we want to summarise, namely clutch-size and lay-date. Again, by simply clicking the cursor and dragging the variable into the box, one arrives at the above picture **EXCEPTING** you need to tell Excel how you want that continuous variable summarized. (IS EVERYBODY CLEAR ABOUT DIFFERENCES IN CONTINUOUS AND CATEGORICAL DATA? Please voice your doubts if you have any, as these are **CRITICAL** concepts here).

Can anybody tell me why I have chosen the above summary statistics, namely mean, standard deviation and count?

→ *Place your cursor now over e.g. Count of layd-date3, and DOUBLE CLICK. The following choice box should pop up*



This allows you to choose various summary statistics.

→ *Scroll through these choices and try changing various options. Note too – to get rid of, let's say the variables relating to lay-date, simply grab them and drag them off to one side of the box and “let go”.*

SIMPLE – YES – BUT TWO PROVISOS: 1) Columns chosen for Pivot Table analysis, even if not used in the actual summaries, **MUST** have column headings. 2) Make sure too that the data are “clean” – out of habit (and for good reason) I use -999 as a missing value indicator: if I am asking for a mean (or whatever) where these missing values are included, -999 gets included too! It won't be the first, or the last, time that I have accidentally included such data in my summaries ...

→ *A third consideration is that Pivot tables limit your subsequent manipulation of the data. So the trick is to copy and Paste Special, Values. But one final trick before this is to get the figures into east to manipulate columns. This can be done by getting the mouse cursor to hover over the column marked Data and get it to change into a “four-armed” pointer – move this along into the total box, and Viola!*

Some more practice:

→ *Go the sheet “work1”, and calculate mean, sum and count data for CS, LD, Brood-size and nestling mass for each year, using the Pivot table Wizard. But this time, use a new worksheet for storing your Pivot Table. Label this sheet: graph1*

I strongly suspect that some of you have “nonsensical” data. **Don’t simply accept whatever Excel tells you: LOOK AT YOUR RESULTS AND USE YOUR COMMON SENSE.** For example, who has a mean nestling mass of -582 g? I know great tits are small relative to cranes, but these birds apparently weigh half a kilo less than zero! Why is this? Correct: my indicator of missing values (-999). Why do I bother using -999 instead of leaving the cell blank? It tells me that this is a genuine missing value, and not simply that someone forgot to enter the data, or, alternatively, record the data. But it does require that you “clean” your data ready for analyses. (By the way, another reason to keep and use -999 is that some statistical packages do not like “empty cells”, but allow one to define a value as missing).

→ *How do we go about “cleaning” the data?*

There are two ways: the first is to sort each column sequentially, and check the minimum and maximum values. Why do it this way? Because whilst we might use -999 as missing data, and thus think we only need to remove -999, one may well accidentally have forgotten the minus sign, which will be equally disastrous. Also, one may decide not use all the data range for very good biological reasons: for example in the tits, one tends not to use nests with a CS of less than 3 eggs (sometimes even 5), as these tend to be indicative of something “going wrong”: desertion or whatever, and similarly, clutches over 14 (in great tits but not blues) are thought to be possibly indicative of more than one female laying in the nest.

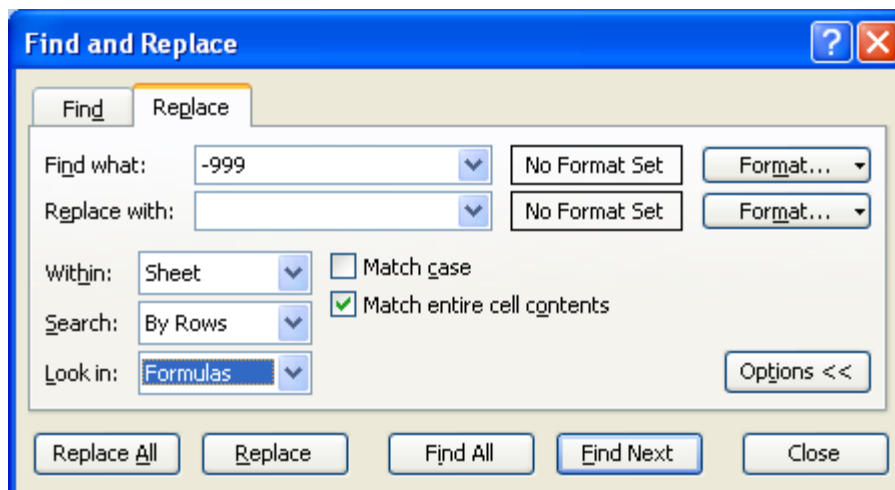
I have also been economical with the truth, just to check who really CHECKS their data as opposed to blindly accepting whatever the machine spews forth: when we are uncertain about records, we precede the **assumed** number by minus. Thus a great tit nest may be given a lay-date of -3 or -10. This is to do with nest checking protocols: generally, >99% of female great tits lay one egg a day, and generally lay more than five eggs. Thus in between lecturing and writing papers etc, one needs to go out and check the nest boxes, of which there are over 1000 in Wytham Woods. On average, if boxes are checked every five to seven days, and birds lay a “normal” clutch, one can “backdate” to get the date of the first egg. Thus I check NB1 on day 5 (= 5 April) and it is empty. I check it again on Day 12 and it contains 2 eggs: I give it a lay-date of 11 (11 April, when first egg laid early in morning – they all lay early in the morning!). But sometimes a nestbox can’t be found, or is accidentally “forgotten”: one then uses the hatch date to calculate the lay-date, but there is much more uncertainty here: thus -10 indicates a **probable** lay-date of 10 April. Note too that LD 1 in this study represents 1 April – soon, with global warming, -10 will represent a lay-date of 22 March ... **Which is why FOR LAY-DATE, missing values are 999, NOT -999 ...**

An absolutely key aspect of sorting in Excel is to ensure that ALL columns are selected prior to sorting, otherwise you sort on, let's say clutch-size, and you DECOUPLE individual clutches from their lay-dates etc etc. DISASTER. **ALWAYS, ALWAYS, ALWAYS click on the top left corner of the active worksheet to ensure that ALL columns are selected.** OK clever-clogs – you decide to simply use the Ctrl+Shift and the arrow keys to whiz around and select all the data so that you don't make this sort of error. Except you forget that last night you were wanting to check some data and added a blank column between CS and LD which is not visible on your screen, and you've forgotten. Net result: DISASTER! I have seen this happen too frequently both to myself and other not to warn you. What I can also be sure of is that at some point you will ignore this lesson ...

One more trick for free: IN YOUR RAW DATA, and in ALL SHEETS YOU COPY, have a **unique reference column** that allows you to go back to your raw data and check a cell's content (there are also powerful lookup functions within Excel that allow you to e.g. add parent ID to your summary data in calc1 etc. But these will ONLY work if you have a common ID in BOTH worksheets (hence "pnum" column in worksheet raw).

→ ***OK – lets clean these data together.***

I said earlier there were two methods for cleaning data: one sorting each column in turn and editing by hand; the other is to use the Find and Replace function.



Obviously, the Find & Replace method is quick and easy IF you know EXACTLY what you are looking for. By its very nature through human data entry, there may well be unpredictable results which you will only find by 1) sorting data, or 2) plotting the data and looking for outliers. So clearly my advice to you is to use the long way round of Sorting data and methodically and physically deleting "unwanted" cells (BUT DON'T DO THIS IN THE RAW DATA (other than to correct mistakes in data entry)).

CHARTS or GRAPHS IN EXCEL

The final section that I want to run over with you in this basic course on using Excel is producing graphs in this package. Now, very generally, there are two main reasons why scientists wish to present data in a graphical format.

→ *Any ideas as to these reasons?*

The first is when all the particular data are plotted, and the person is looking for **patterns** in the data and/or **outliers**. An example of this would be to plot the data on clutch-size in sheet work1.

The second reason is to best **summarise** the data in a visual manner, making any **patterns or trends** obvious to the viewer/reader.

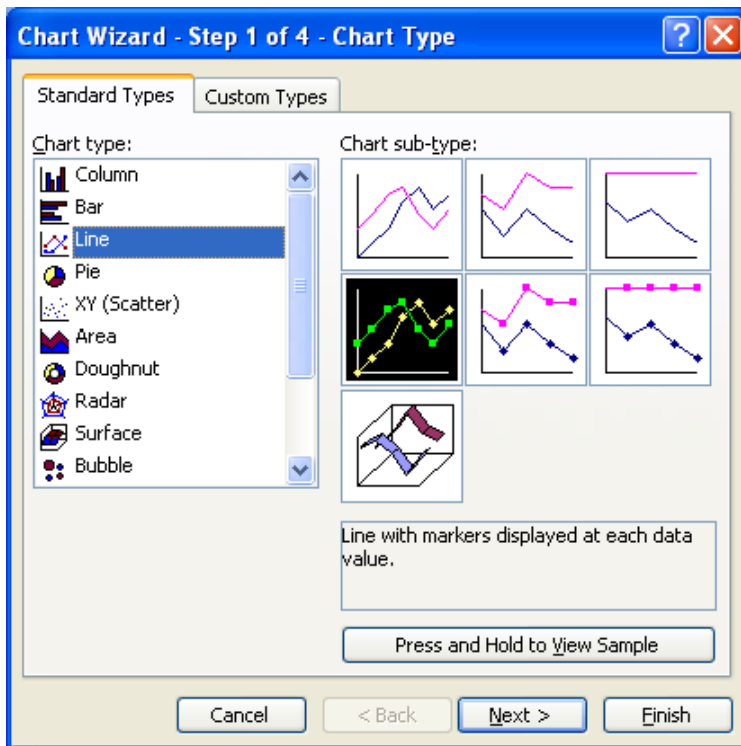
→ *Would plotting all the individual data points as suggested for the first approach above be appropriate here?*

Lets look at the first approach.

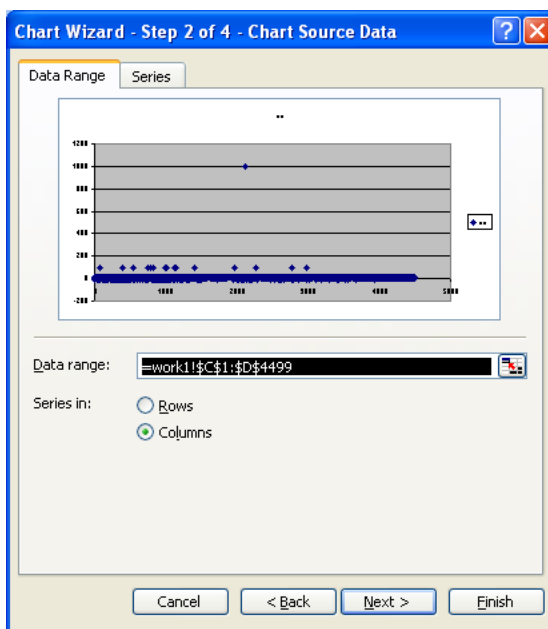
→ *Have work1 open in front of you: Insert a blank column to the left of clutch size. In cell C2, type "1", then in C3, "2" and in C4, type "3". Enter label "Obs" (observation) into C1. Now Select C2:C4 and carefully double-click on the "handle". If you have done this correctly, you will have filled column C sequentially 1, 2, 3 ...4498. Now "black out" ie select Columns C & D (do this by "floating the cursor over the column headers so that the cursor changes into a downward pointing black arrow).*

	A	B	C	D	E	F	G
1	year	nestbox	obs	cs	ld	broodsz	nringed
2	61	1	1	-1	42	0	0
3	61	7	2	-6	45	5	0
4	61	17	3	-6	25	0	0
5	61	31	4	11	23	11	10
6	61	31	5	8	44	7	0
7	61	36	6	9	32	0	0
8	61	40	7	10	29	10	9
9	61	40	8	9	34	-4	0
10	61	50	9	8	33	8	8
11	61	51	10	8	36	0	0
12	61	53	11	6	43	99	6
13	61	55	12	9	30	-5	0
14	61	55	13	-3	32	0	0
15	61	56	14	8	25	8	7
16	61	62	15	9	39	-6	0
17	61	64	16	-8	34	0	0

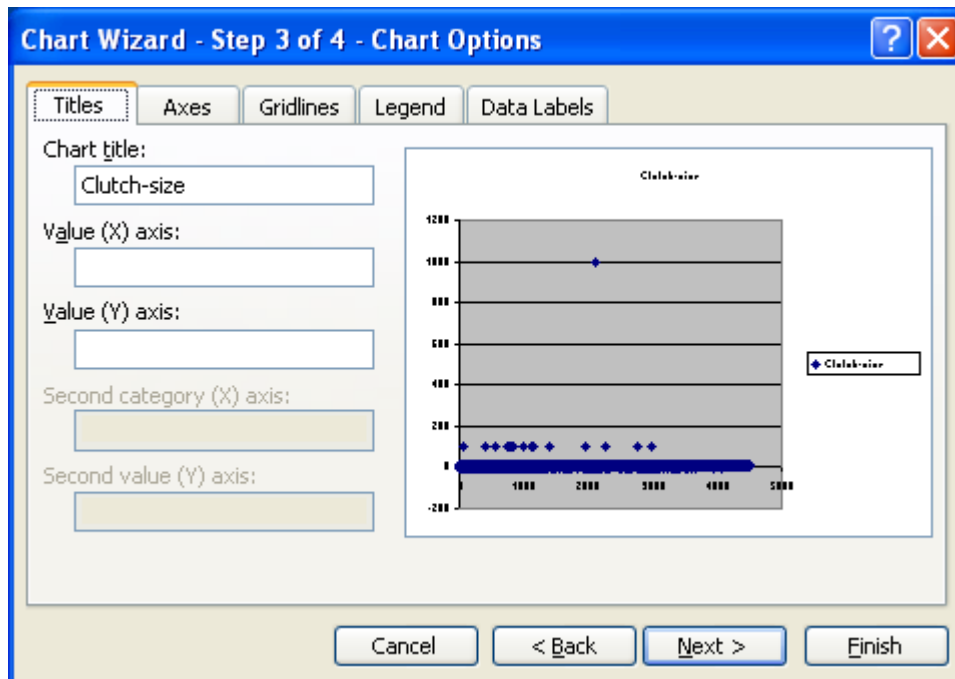
→ *Now go to the little button which looks like a miniature 3D histogram in many colours (next to the zoom button) – this is the Chart Wizard. Click on it and the following appears*



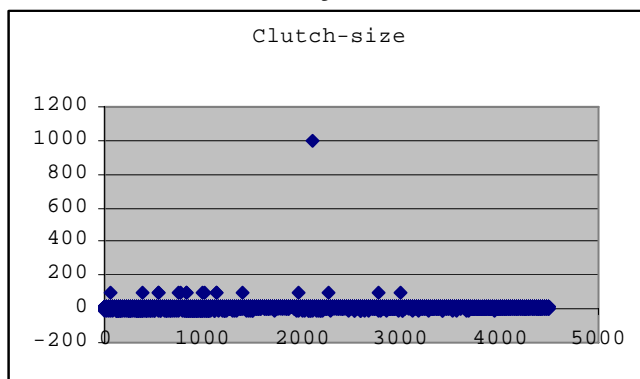
→ *Highlight XY (Scatter) and double-click on top, left-hand graph. Step 2 of Wizard now appears.*



→ *Note the data range: which data are these? By and large I DON'T recommend this way of entering data, but here I am more concerned with the plots. Note too that it is telling you that it is reading the data in columns. The big UNKNOWN though is which is the x-axis and which is the y-axis. Click on the series tab, and the information is given there. Under Name, type in "Clutch-size" and hit Next. Step 3:*



→ *At this stage I am not concerned with the other tabs in Stage 3 of the Wizard. Why might this be? OK – we are using the graph function (Whoops – chart function) to EXPLORE the data. Hit finish, and cut the label:*



This is immediately telling me two things about my data on clutch-size.

→ *What are they?*

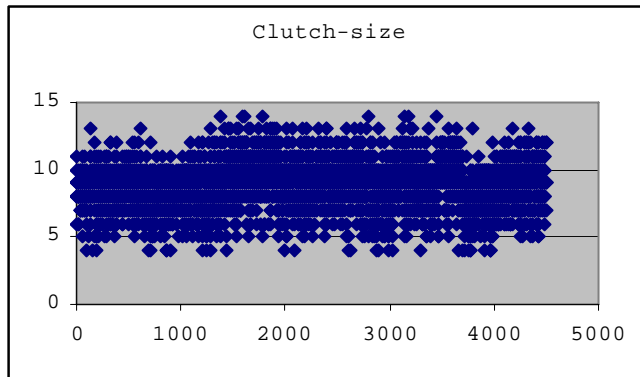
First, there is one or more data points suggesting a clutch-size of around 1000 eggs, and then second (but hard to tell with this scale) that there are a series of further outliers somewhere between 50 and 100.

→ ***How do we tackle this?***

Correct: we go back into work1, Sort, and look at the range in data.

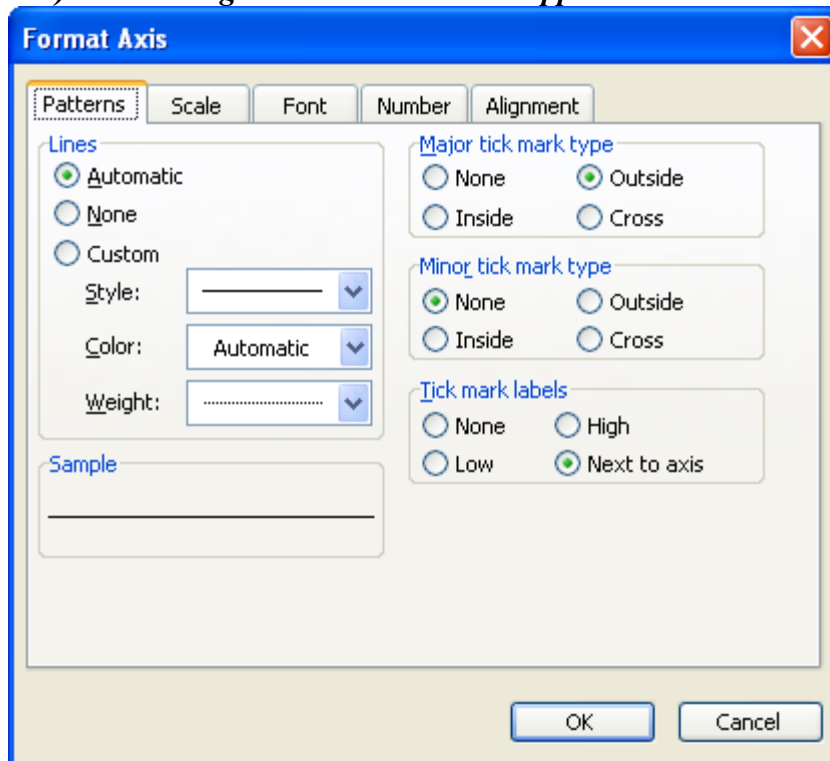
→ ***Sort the data on clutch-size, and Clear CS < 4 and CS > 14.***

→ ***What has happened to the plot?***



Firstly, it has automatically re-scaled the y-axis, and secondly the data now appear “reasonably” distributed i.e. there are no obvious outliers. Just to take you a bit further (although for exploratory purposes I wouldn’t bother with the following)

→ ***Click on the y-axis so that it is shown as selected (two black handles at either end) and then right click: Format Axis appears***



→ ***Click on Scale, and change Minimum to 3 and leave Maximum at 15. Repeat on the X-axis, and set the maximum scale at 4500. This is simply “good” graphical practice in that we are maximizing the scales of the axes to present the maximum amount of information in the most easy to read manner.***

Right, let's move on. As we discussed earlier: plotting serves two functions: one an exploratory one such as we've just done, where the exercise is largely for our own purpose and does not need to be “prettied up”. But the second is to present the data in a way which maximizes both its content, and, let's face it, its **impact**. In other words, the 4500 points plotted above tells us very little other than that the clutch-size in our data ranges between 4 and 14 eggs (because we have defined those limits whilst “cleaning” the data!). However, if we are wanting to write a report or give a talk on clutch-size in tits, then perhaps the “right” scale to be approaching this problem is the variation in mean clutch-size on an annual basis.

→ ***How do we do this?***

Yes – we have already partly tackled this problem with our Pivot tables earlier. In sheet graph2, I have already presented the summary data you will be plotting.

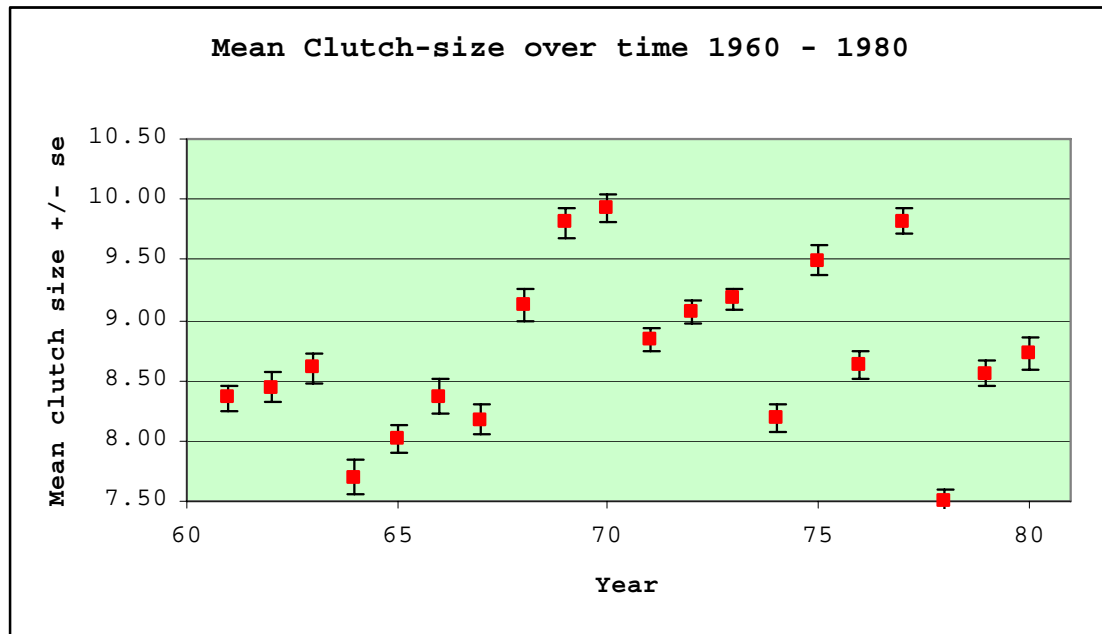
→ ***But I want you to replicate this data lay-out EXACTLY before we proceed. Go to sheet8 and see if you can do it.***

→ ***SHOUT IF YOU REALLY GET STUCK***

→ ***Great – I now need you to do one more lot of data ,manipulations before we are ready to do any plotting (or charting as Microsoft would have us say). Insert a blank columns into C. Remember, the standard error = (the standard deviation)/[square root (number of observations)]. Write a formula in each cell that will calculate the standard error (in Column C).***

→ ***GETTING STUCK? Talk among yourselves.***

→ ***Right, does everyone now have a data set that contains a column (A) for year, mean clutch size (B), and the se (C)? Then plot mean clutch size against year with error bars: I want you to land up with a plot like the one below, or even better, improve on it!***



OK – as I understand it, Excel will be your only graphics package, so let's make sure we all know what we are doing. I now want you to do the following plots:

- *A plot of mean annual clutch-size against mean annual lay-date*
- *I then want you to put a regression line through this plot and show the r^2 values and the line of best fit equation*
- *COPY this graph, paste it, and then plot mean nestling mass against mean clutch-size. What is the relationship? How does mean nestling mass vary with mean lay-date?*
- *Plot a histogram of mean nestling mass with year.*

A FINAL TASK

Kirsten has given me a copy of an old spreadsheet she has found in SACWG's archives.

→ *Open this spreadsheet (crane-old.xls). Look at the column headings (or lack of them), and also notice the different formats in the SAME column (eg for Date). In other words, this worksheet is a mess. I want you to work in pairs, clean the data, and provide a plot suitable for publicity that looks at crane numbers over time.*

Hints – Use Pivot Tables to extract unique dates
Use Vlookup to insert “useful” date formats, meaningful counts etc.

WELL DONE – you are now pretty fluent users of Excel. But the only way you can keep this up is through constant use of what you have learnt – if you don't touch formulae for another year you will have completely forgotten everything, but if you use what you now know on a regular basis, you will soon be able to use the Excel help files to write and solve any equation you wish. Go for it, and thanks. You are a great team.

*Richard Pettifor
March 2007*