A Veterinarian's Guide to Microsoft Excel *

Mark Stevenson †

EpiCentre, IVABS

Massey University, Palmerston North, New Zealand

September 21, 2008

Contents

*Notes to accompany Veterinary Biometrics and Epidemiology, taught as course 227.407 within the BVSc program at Massey University.

†M.Stevenson@massey.ac.nz

1 Using a spreadsheet

Microsoft Excel is a spreadsheet software package that allows the user to store, manipulate, analyse and graph data. Each Excel file consists of a workbook that can hold a maximum of 256 worksheets. In each worksheet there can be up to 256 columns and 65,536 rows (a total of 16,777,216 cells). Within a worksheet columns are identified by letters (`A,B,C, ... AA,AB, ...`) rows are identified by numbers (`1,2,3, ...`). Individual cells are referenced by their column and row identifiers: cell `A1` is the cell in the top left corner of the worksheet. Numbers, text and dates can be entered into these cells.

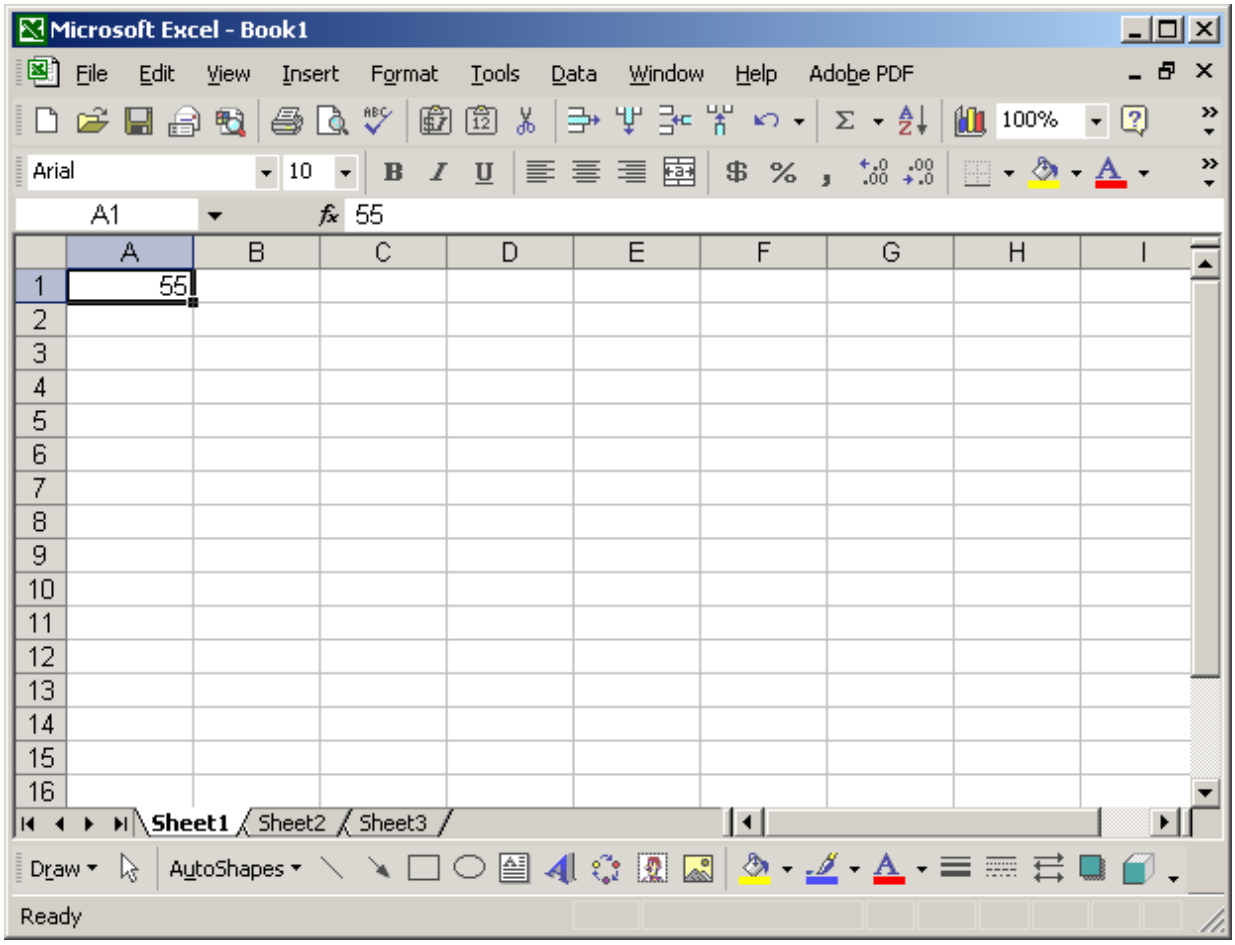


Figure 1: The Excel worksheet. The number '55' has been entered into cell A1.

1.1 Formatting cells

A cell can be formatted to display either text or numbers. To format a cell open the format dialogue box (-FORMAT-CELLS from the menu bar) and select one of the options on the tab titled 'Number.' Another way to access the format dialogue box is to right click on a selected cell and select 'Format Cells' from the drop down list. Using the format dialogue box it is also

possible to change the font of displayed numbers (or text), change the colour of displayed text, add borders around cells and so on.

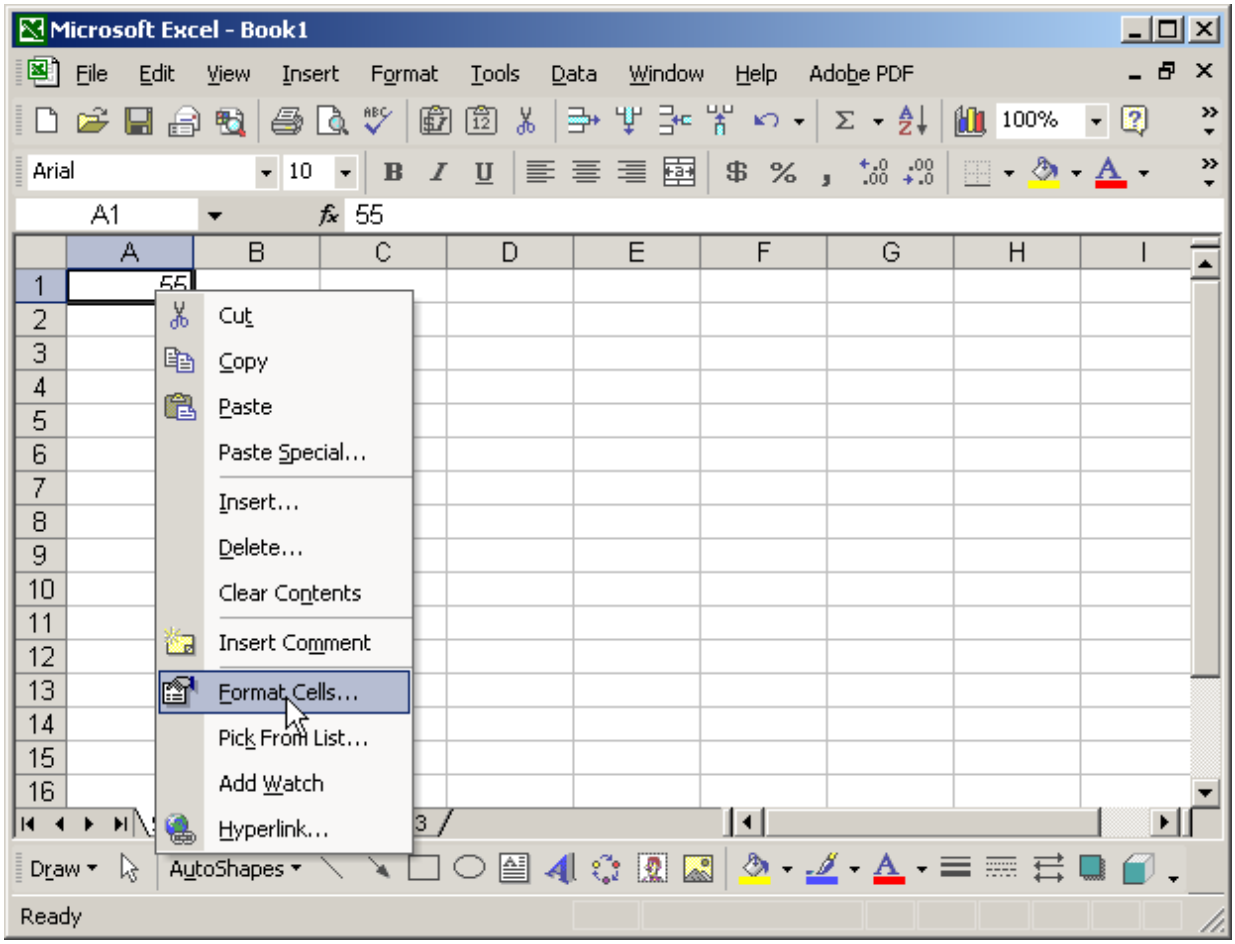


Figure 2: Formatting cells in MS Excel. A cell is selected and the right mouse clicked. Select the Format Cells ... option.

1.2 Sorting data

Once data has been entered into the cells of a worksheet it is possible to sort it (e.g. from lowest to highest, or alphabetical order) using the -DATA-SORT function. When data is sorted text and number formats are sorted separately. If you have numbers and a combination of numbers and text in your list (e.g. 1, 230, 23a, 34b, ...) and you want these values to be sorted together you will need to change the numbers to text before you perform the sort. Once you have done your sort, you might need to change your numbers back to numeric format. Use the paste special (-EDIT-PASTE SPECIAL) function to multiply all numbers by 1.

If you have values in multiple columns and highlight a single column then Excel will sort that single column independent of the others. Be careful! If each row of your worksheet has several items of information related to particular subject (patient, farm etc) sorting in this fashion will result in your data becoming scrambled. To avoid this problem, highlight all columns before selecting -DATA-SORT.

Table 1: Numeric formats in MS Excel.

Format	Description
General	Cells have no specific number format.
Number	Used to display numbers, allows user to specify the displayed number of decimal points.
Dates	Used to format dates and times.
Currency	Used for entering currency values (aligns decimal points and insert a dollar sign).
Percentage	Multiplies the selected number by 100 and displays it with a percent sign.
Scientific	Displays the number in scientific notation format.

1.3 Functions

Excel allows the user to manipulate and analyse data by entering formula into cells. The different types of formula can be grouped into a number of categories including:

- Mathematical
- Logical
- Array
- Statistical
- Look-up
- Rounding and truncating

Formulae in Excel start with an equal sign (`=`) followed by the formula syntax. Formulae can contain either numbers e.g. `=2+3` or a reference to another cell in the spreadsheet e.g. `=A1+A2`. In some cases you will need to stipulate a range of numbers. You can enter either the cell references themselves (e.g. `A1,A2,A3`) or the first and last cell in the range are identified for example `A1` to `A10` is entered as `A1:A10`.

If you type a single quote (') in a cell before anything else Excel will interpret the contents of the cell as text and display the cell contents exactly as it has been entered (without the '). For example, if you type =3+4, 7 will be displayed. However, if you type '=3+4 then =3+4 will be displayed. This feature is useful for making your calculations clear to others using your spreadsheet.

When referring to a cell on its own or as part of an array the reference can be *relative* or *absolute*. An absolute reference is differentiated from a relative reference by the inclusion of `$` in front of the column and/or row reference. For example, the relative reference of the top left cell in the worksheet is `A1` and the absolute reference is `$A$1`. The difference between the two is apparent only when the formula is pasted to another location. An absolute reference always refers to the particular cell no matter where in the worksheet it is pasted. In contrast, when a relative reference is pasted to another cell the cell it refers to is changed based on where the original cell was in relation to the new cell. For example if the reference in cell `B1` was to `A1` and it was copied to `D1` then the reference will now refer to cell `C1`.

To make a relative reference absolute highlight the cell reference and press the F4 key. `$A$1` will be entered, rendering the cell absolute. Click the F4 key again will cycle through `$A1`, `A$1`, and `A1` making the cell column absolute, row absolute, and relative.

Mathematical functions

Mathematical functions can be entered by directly typing them into a cell or by using the function wizard. To open the function wizard click -INSERT-FUNCTION.

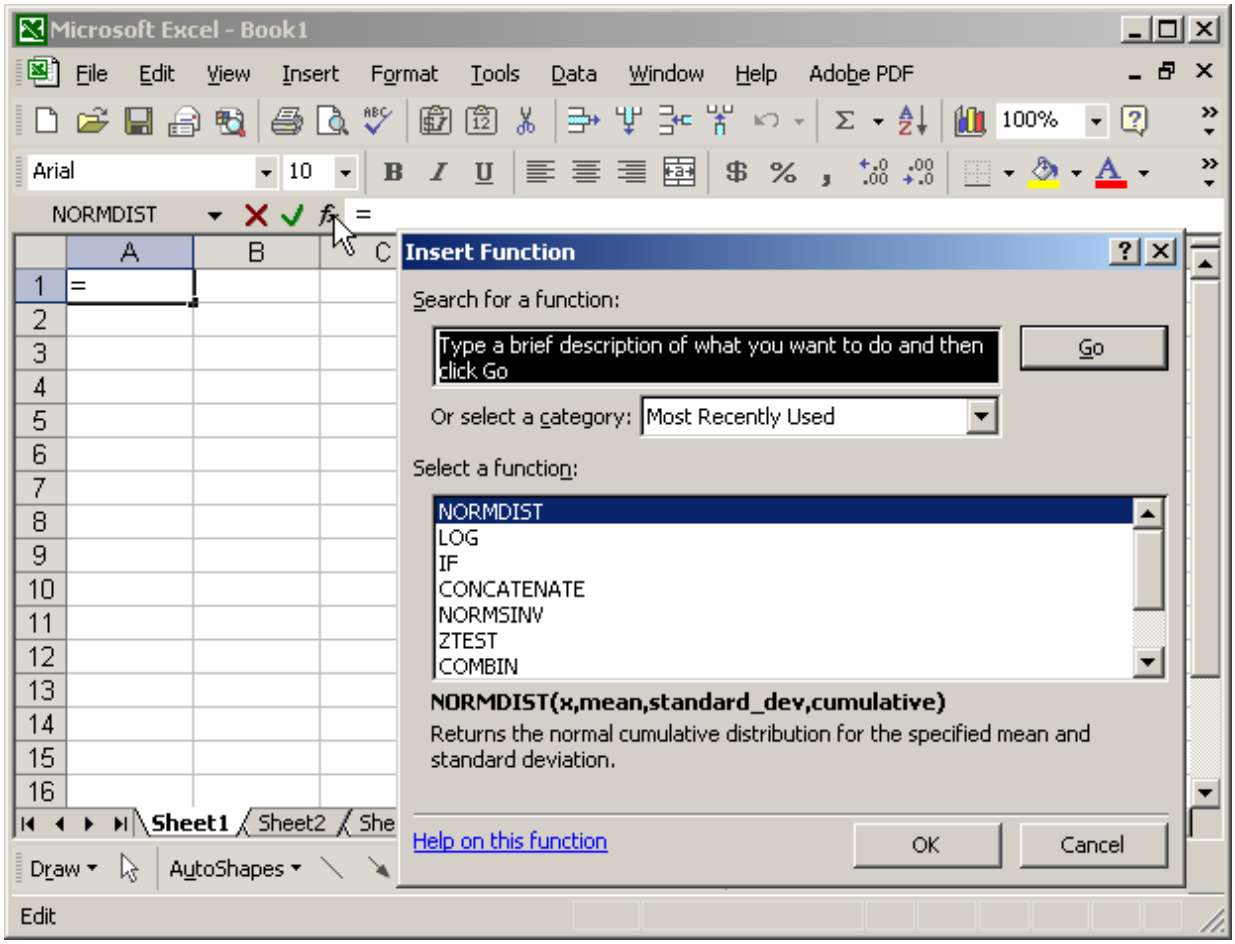


Figure 3: The Excel function wizard.

The keyboard operators for some of the more simple mathematical functions are:

- addition (+)
- subtraction (-)
- multiplication (*)
- division (/)

Table 2: Syntax for commonly used mathematical functions.

Function	Description	Example
SUM	Sums a list of numbers or cell references.	`= SUM(A1, B1:B10)`
MAX	Returns the maximum number in a list of numbers.	`= MAX(A1, A20)`
SQRT	Returns the square root.	`= SQRT(9)`
EXP	Returns e raised to given number.	`= EXP(A1)`
LN	Returns natural of log of a given number.	`= LN(A1)`
LOG10	Returns base-10 log of a number.	`= LOG(10)`
LOG	Returns base-n log of a number.	`= LOG(2,5)`
POWER	Returns the result of one number raised to another.	`= POWER(A1,A2)`

After entering the name of a function into a cell Excel can be made to prompt you for the required values by entering CTL-A or CTL-SHIFT-A.

If several operators (+, - *, /) occur in the same formula the program will calculate the higher order precedence first. That is when entering a formula Excel will calculate parentheses, any to the power of another number, multiplication/division and then addition/subtraction. For example, when calculating `=3+4*(3-1)2` Excel will calculate 3 - 1 to give 2, it will then square 2 to get 4 and multiply 4 by 4 to give 16 and finally add 3 to give 19. If events are of the same precedence (e.g. multiplication and division or addition and subtraction) then Excel will work from left to right. If you have any doubt as to how Excel will deal with a formula, enter brackets to ensure the formula behaves as you intend it to.

Absolute and relative cell references can be used to complete running totals. For example, Figure 6 shows a spreadsheet of month and sales and the running total. The running total (column C) was calculated by entering `=SUM($B$2:B2)` and copying the formula into the other cells in column C. When the formula is pasted the start of the array (cell `$B$2`) stays fixed because it is an absolute reference but the end of the array is free to move depending on what row the formula is pasted into.

When using functions the formula not the cell value is recorded. To copy the actual results to a new location select -EDIT-PASTE SPECIAL and then highlight 'value' only.

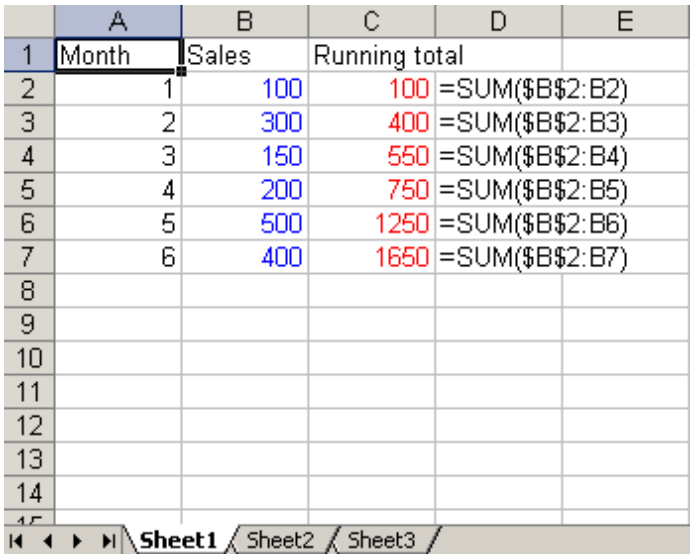
Logical functions

Logical functions are used to compare two numerical values or strings and return a TRUE or FALSE value. Logic operators can also be used to return one value if certain criteria are met and another if not.

The IF function can be used to specify the value to be returned if the condition is TRUE and FALSE. The general form of the IF function is:

```
= IF(logical condition, value if true, value if false)
```

For example `=IF(weight<990,NA(),1)` returns an empty cell if weight is greater than 990 and a 1 if it is less than 900. Up to seven IF statements can be nested within a single formula:



	A	B	C	D	E
1	Month	Sales	Running total		
2	1	100	100	=SUM($B$2:B2)	
3	2	300	400	=SUM($B$2:B3)	
4	3	150	550	=SUM($B$2:B4)	
5	4	200	750	=SUM($B$2:B5)	
6	5	500	1250	=SUM($B$2:B6)	
7	6	400	1650	=SUM($B$2:B7)	
8					
9					
10					
11					
12					
13					
14					

Sheet1 Sheet2 Sheet3

Figure 4: Use of sum and absolute and relative cell references to calculate a running total of sales by month.

Table 3: Syntax for commonly used logical functions.

Function	Description	Example
AND	TRUE when both conditions are met and FALSE otherwise.	`= AND(A2>=1,A2<=10)`
OR	TRUE if either condition met and FALSE otherwise.	`= OR(A2<1,A2>10)`
=	TRUE if values are the same and FALSE otherwise.	`= (A1=A2)`
NOT	TRUE if values are different and FALSE otherwise.	`= NOT(A1=A2)`

`=IF(weight<300,5,IF(weight>800,15,10))` returns 5 if weight is less than 300, 10 if weight is between 300 and 800, and 15 if weight is greater than 800.

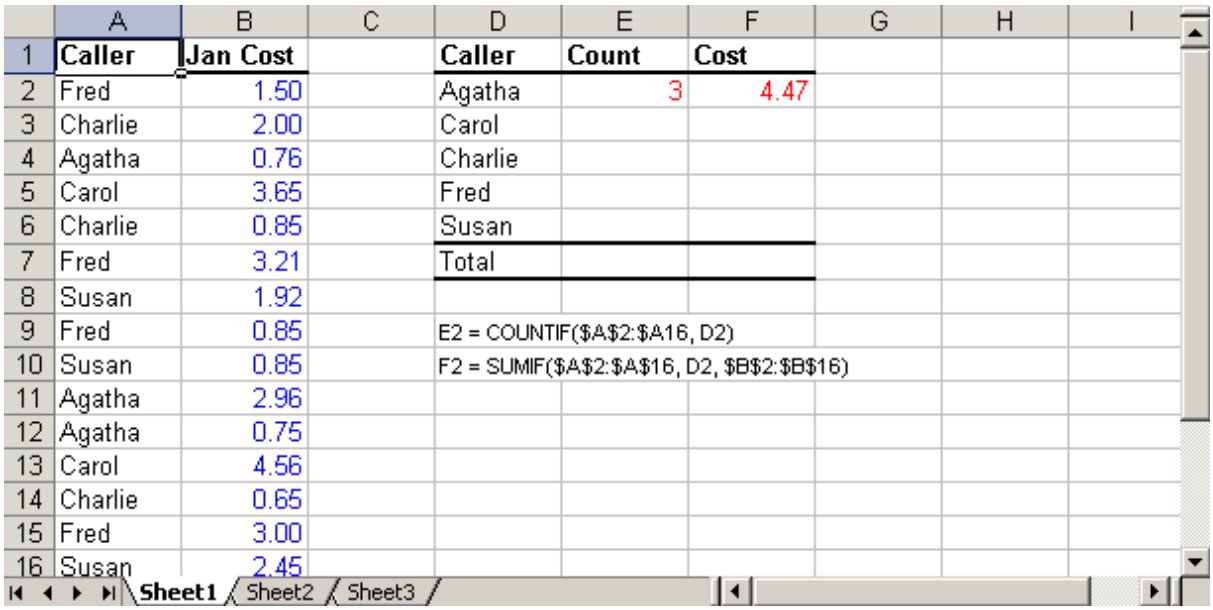
It is possible to count and sum values that meet certain criteria, for example sales for each agent, costs from a supplier. This is done using the `COUNTIF` and `SUMIF` functions. The arguments for these functions are:

```
= COUNTIF(range, criteria)
= SUMIF(range, criteria, sum range)
```

Array functions

Array formulae operate on more than one cell at a time and return more than one value. Array formulae can be recognised by different enclosing brackets {} that are entered by Excel after the formula is entered. To ensure results of an array formula are written to an array (rather than a single cell) you must press SHIFT + CTRL + ENTER when entering the formula. If you just press ENTER then the value will be returned to a single cell.

Figure 6 contains data arranged in three columns (columns A, B, and C). We wish to create a frequency histogram of these data. First of all we create a list of 'bin' values (cells `E2:E11`).



	A	B	C	D	E	F	G	H	I
1	**Caller**	**Jan Cost**		**Caller**	**Count**	**Cost**			
2	Fred	1.50		Agatha	3	4.47			
3	Charlie	2.00		Carol					
4	Agatha	0.76		Charlie					
5	Carol	3.65		Fred					
6	Charlie	0.85		Susan					
7	Fred	3.21		Total					
8	Susan	1.92							
9	Fred	0.85		E2 = COUNTIF($A$2:$A16, D2)					
10	Susan	0.85		F2 = SUMIF($A$2:$A$16, D2, $B$2:$B$16)					
11	Agatha	2.96							
12	Agatha	0.75							
13	Carol	4.56							
14	Charlie	0.65							
15	Fred	3.00							
16	Susan	2.45							

Sheet1 Sheet2 Sheet3

Figure 5: Using the COUNTIF and SUMIF functions in MS Excel.

We now want to determine the frequency of each bin value - to do this we use the `FREQUENCY` function. Select the cells `F2:F11`, type `=FREQUENCY(A2:C21,E2:E11)`, click CTRL + SHIFT + ENTER.

Text functions

Text functions change numbers to a text or string. You can specify what format the text will be displayed in by using some of text functions.

To convert a number stored as text to a number put the number 1 in an empty cell and copy the cell. Then highlight the cells you wish to change and select -EDIT-PASTE SPECIAL from the menu bar. In the past special dialog box and tick multiple in the operator section.

Statistical functions

It is often necessary to summaries data using statistics such as the average, standard error, minimum and maximum. In Excel it is possible to calculate a number of values for a data set by typing in the statistical function you require.

Excel also has an Analysis ToolPak add-in which contains a number of statistical functions including:

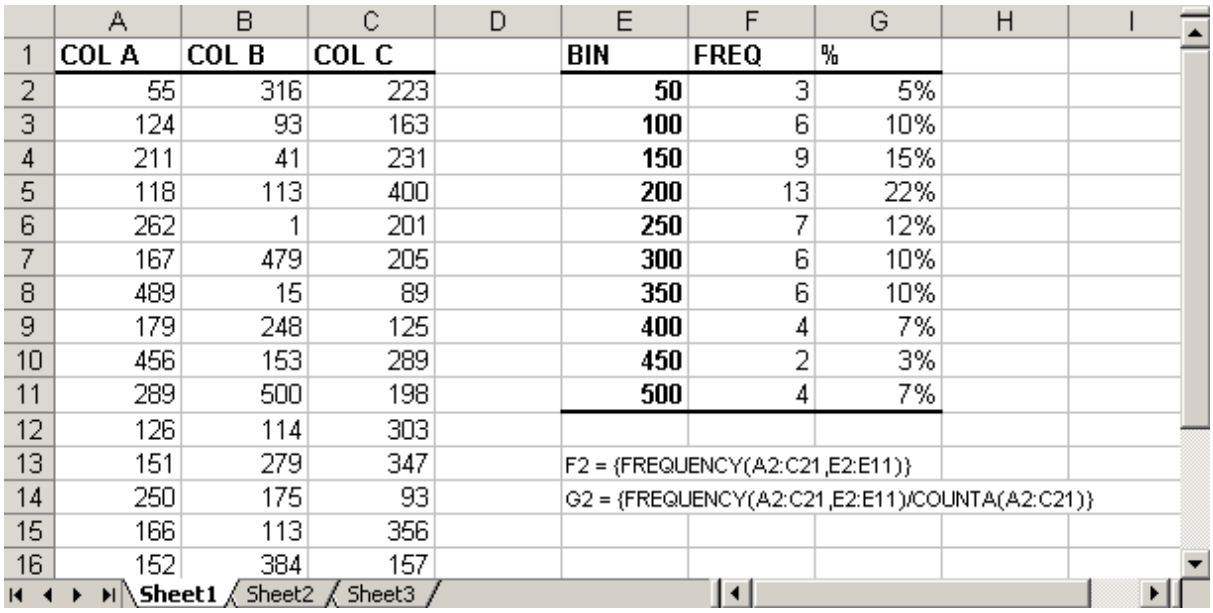
- ANOVA
- Correlation
- Descriptive statistics
- t-tests

Table 4: Syntax for commonly used text functions.

Syntax	Description
DOLLAR()	Returns a number as text in currency format.
FIXED()	Returns as a text rounded to the specified number of digits.
T()	Return the value as text.
VALUE()	Returns the number value of a string.
CONCENTENTATE	Combines two or more strings together.
FIND()	Locates a substring (case sensitive).
LEFT()	Extracts characters from the left.
LEN()	Returns the number of characters in a string.
LOWER()	Converts a string to lower case.
MID()	Extracts a substring.
PROPER()	Capitalizes the first letter of each word in the string.
REPLACE()	Replaces a substring.
RIGHT()	Extracts characters from the right of a string.
SEARCH()	Locates a substring (not case sensitive).
TRIM()	Removes leading and trailing spaces.
UPPER()	Converts a string to upper case.

Table 5: Syntax for commonly used statistical functions.

Function	Description	Example
AVERAGE	Returns the mean of a series of numbers.	`= AVERAGE(B1:B10)`
CORREL	Returns the correlation between two values.	`= CORRELMAX(A1:A20, B1:B2))`
COUNTBLANK	Counts the number of cells that have no value.	`= COUNTBLANK(A1:E20)`
LARGE	Returns the kth largest value in the data set.	`= LARGE(A1:E20, 3)`
MEDIAN	Returns the median.	`= MEDIAN(A1:A20)`
MODE	Returns the mode.	`= MODE(A1:A20)`
STD	Returns the standard deviation of a range of numbers.	`= STD(A1:A20)`



	A	B	C	D	E	F	G	H	I
1	COL A	COL B	COL C		BIN	FREQ	%		
2	55	316	223		50	3	5%		
3	124	93	163		100	6	10%		
4	211	41	231		150	9	15%		
5	118	113	400		200	13	22%		
6	262	1	201		250	7	12%		
7	167	479	205		300	6	10%		
8	489	15	89		350	6	10%		
9	179	248	125		400	4	7%		
10	456	153	289		450	2	3%		
11	289	500	198		500	4	7%		
12	126	114	303						
13	151	279	347		F2 = {FREQUENCY(A2:C21,E2:E11)}				
14	250	175	93		G2 = {FREQUENCY(A2:C21,E2:E11)/COUNTA(A2:C21)}				
15	166	113	356						
16	152	384	157						

Sheet1 Sheet2 Sheet3

Figure 6: Using ARRAY functions in MS Excel.

- Moving averages
- Histograms

To access the ToolPak see -TOOLS-DATA ANALYSIS from the menu bar. If the feature is not available then click -TOOLS-ADD INS and select 'Analysis ToolPak.' If the option to select Analysis ToolPak is not visible you will need the Microsoft Office installation media to install the Analysis module into your version of Microsoft Office. The descriptive statistics option in the Analysis ToolPak is useful for calculating a range of summary statistics.

Probability functions

Syntax for the various functions related to probability distributions are shown in Table 6.

To return the z value for a given probability value:

```
= NORMSINV(0.975)
= 1.96
```

To return the probability of a given value of z:

```
= NORMSDIST(1.96)
= 0.975
```

A study has found that the bodyweight of newborn Friesian calves is normally distributed with a mean of 45 kg and SD of 5 kg. What is the probability that a calf will be greater than 55 kg at birth?

```
= NORMDIST(55,45,5,TRUE)
= 0.98
```

The probability that a calf will be up to and including 55 kg at birth is 0.98. The probability that a calf will be greater than 55 kg is (1 - 0.98) = 0.02. There is a 2% chance that a calf will be greater than 55 kg at birth.

Table 6: Syntax for commonly used probability distributions.

Function	Description	Example
BETADIST	Cumulative beta pdf.	`= BETADIST(x, alpha, beta, A, B)`
BETAINV	Inverse of the cumulative beta pdf.	`= BETAINV(x, alpha, beta, A, B)`
BINOMDIST	Individual term binomial distribution probability.	`= BINOMDIST(no_s, trials, prob_s, cumul)`
CHIDIST	One-tailed probability of the chi-squared distribution.	`= CHIDIST(x, dof)`
CHIINV	Inverse of the one-tailed probability of the chi-squared distribution.	`= CHIINV(x, dof)`
CHITEST	Value from the chi-squared distribution for the statistic and df.	`= CHITEST(act_range, exp_range)`
EXPONDIST	Exponential distribution.	`= EXPONDIST(x, lambda, cumul)`
FDIST	F probability distribution.	`= FDIST(x, dof1, dof2)`
FINV	Inverse of the F probability distribution.	`= FINV(prob, dof1, dof2)`
GAMMADIST	Gamma distribution.	`= GAMMADIST(x, alpha, beta, cumul)`
GAMMAINV	Inverse of the gamma cumulative distribution.	`= GAMMAINV(prob, alpha, beta)`
LOGNORMDIST	Cumulative lognormal distribution.	`= LOGNORMDIST(x, mean, sd)`
NEGBINOMDIST	Negative binomial distribution.	`= NEGBINOMDIST(no_f, no_s, prob_s)`
NORMDIST	Normal distribution for the specified mean and SD.	`= NORMDIST(x, mean, sd, cumul)`
NORMINV	Inverse of the normal distribution for the specified mean and SD.	`= NORMINV(prob, mean, sd)`
NORMSDIST	Standard normal cumulative distribution function.	`= NORMSDIST(z)`
NORMSINV	Inverse of the standard normal cumulative distribution.	`= NORMSINV(prob)`
POISSON	Poisson distribution.	`= POISSON(x, mean, cumul)`
TDIST	Percentage points for the Student's t-distribution.	`= TDIST(x, dof, tails)`
TINV	t-value of the t distribution as a function of the probability df.	`= TINV(prob, dof)`

Table 7: Lookup functions in MS Excel.

Function name	Purpose
VLOOKUP	Returns value in a cell based on the value in another and looking to and array.
CHOOSE	Returns a specific value from a list of values (up to 29) supplied as arguments.
HLOOKUP	Same as VLOOKUP but in the horizontal direction.
INDEX	Returns a value or a reference to a value from within a table or range.
MATCH	Locates the position within a vector.

Lookup functions

Lookup functions return a value either from a one-row or one-column range or from an array. There are several lookup functions available within Excel, as listed below. These notes will outline the use of the `VLOOKUP` function. The arguments for the `VLOOKUP` function are:

```
= VLOOKUP(lookup_value, table_array, col_index_num, range_lookup)
```

Figure 7 shows a spreadsheet where columns A, B, C and D list details of patients recorded at a hopsital clinic. Patient identifiers have been entered into column F and in column G the `VLOOKUP` function has been used to: (1) read the `ID` value recorded in column F, (2) find the row with a matching `ID` variable in the array `A1:D10`, and (3) return the appropriate value for `SEX`.

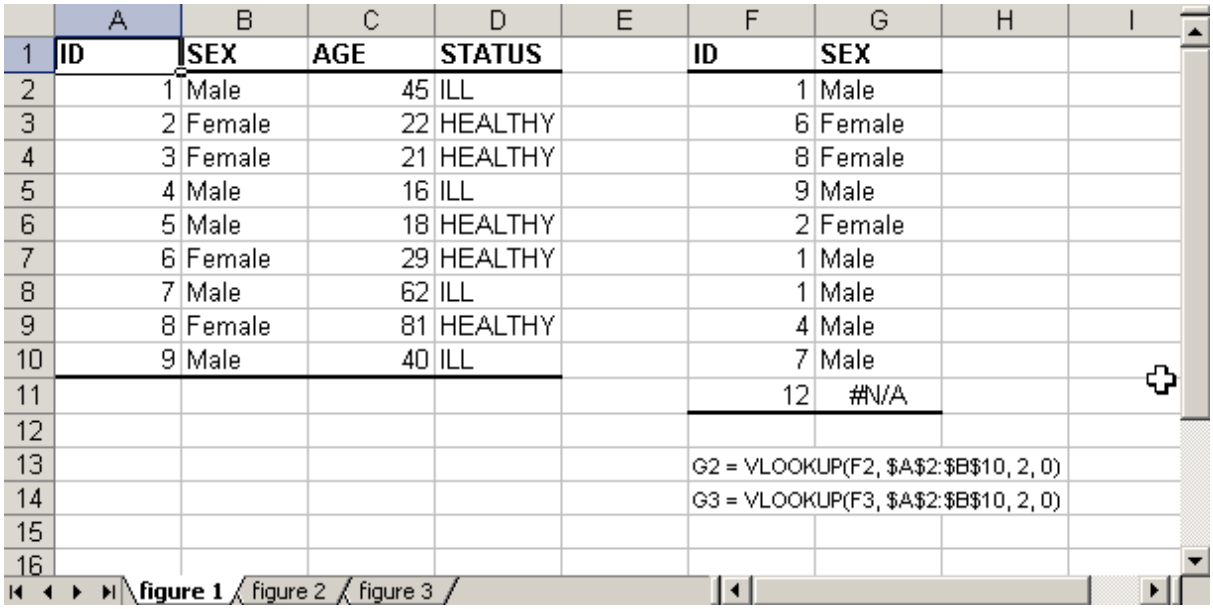


Figure 7: Using the VLOOKUP function in MS Excel.

In the above example, note the use of 0 for the `range_lookup` argument. `range_lookup` is a logical value that specifies whether you want `VLOOKUP` to find an exact match or an approximate match. If TRUE or omitted, an approximate match is returned. In other words, if an exact

Table 8: Functions that change the way numbers are stored in a MS Excel spreadsheet.

Function	Description	Example
ABS	Returns the absolute value.	`= ABS(-12.55)`
FLOOR	Rounds the number towards the nearest multiple.	`= FLOOR(1.255, 0.5)`
INT	Returns the nearest integer.	`= INT(2.4)`
ROUND	Rounds a number to the required digits.	`= ROUND(1.378, 2)`
ROUNDDOWN	Rounds the number down the required number of digits.	`= ROUNDDOWN(1.378, 2)`
ROUNDUP	Rounds the number up the required number of digits.	`= ROUNDUP(1.378, 2)`

match is not found, the next largest value that is less than lookup_value is returned. If FALSE, `VLOOKUP` will find an exact match. If one is not found, the error value `#N/A` is returned.

Rounding and truncating functions

Excel can change the way in which a number is displayed (leaving the entered number unchanged). It is possible to change the actual stored number by using rounding and truncating functions.

1.4 Graphs

To enter a graph into an Excel worksheet click -INSERT-CHART from the menu bar to open the chart wizard. Select the type of graph you require and follow the prompts to produce a chart. Once a chart is completed you can right click on a feature within the chart to alter its features.

Error bars

To add error bars right click on the data points, select Format Data Series from the drop down menu, then select the Y Error Bars tab. Error bars can be either a fixed value, a percentage, a number of standard deviations or custom defined. The custom feature is extremely useful if each point has a different standard error, you can enter the standard errors for each point in a column beside the value and then enter the cell range into custom + and/or -. It is also possible to enter different values for the + and - error bars.

Trend lines

Trend lines can also be added to data points by right clicking on the data series and selecting the Add Trendline option. This will open the Trendline dialogue box shown in Figure 10. Select the type of trend line you require and use the options to display the formula on the graph.

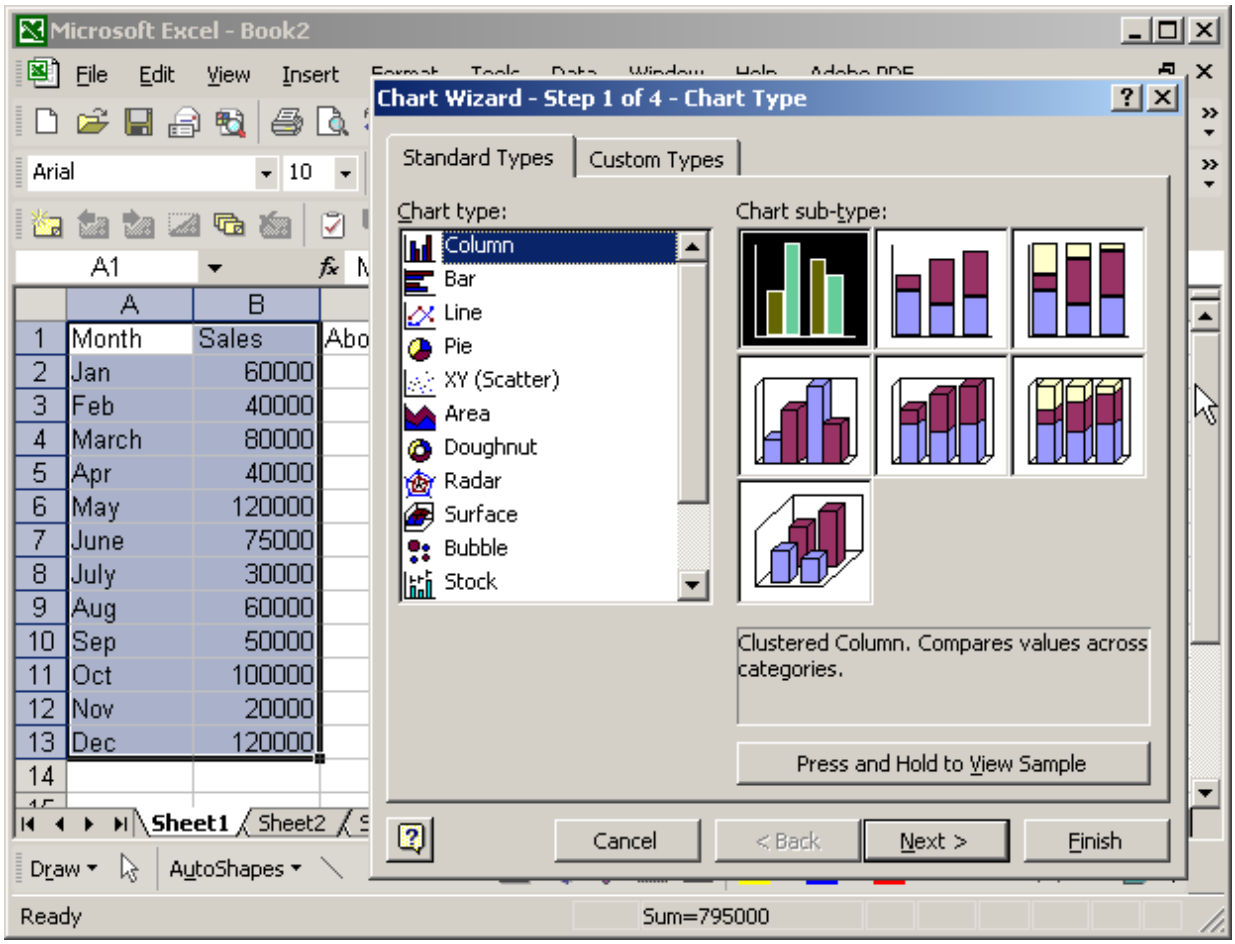


Figure 8: The MS Excel chart wizard.

Shewhart charts

Often we need to monitor variables and how they change over time. A typical example in manufacturing would involve monitoring the number of defective units per batch of goods processed. If an increase in the number of defective units is detected then steps can be taken to investigate (and rectify) the problem. Graphs are a useful way to visualise this process. Here, date (and/or time) is plotted on the horizontal axis and a value for the monitored parameter is plotted on the vertical axis.

A problem that often arises is: two 'abnormal' values have been recorded over a short period of time. Is this sufficient to justify an investigation, or are these values just part of 'random fluctuation'? Shewhart charts (also known as control charts) can be useful in this situation by displaying the level of normal, random variation in a data and by revealing the observations that indicate real change. The steps to create a Shewhart chart are as follows:

- Create a time series chart.
- Add a centre line (usually the mean) for central reference.

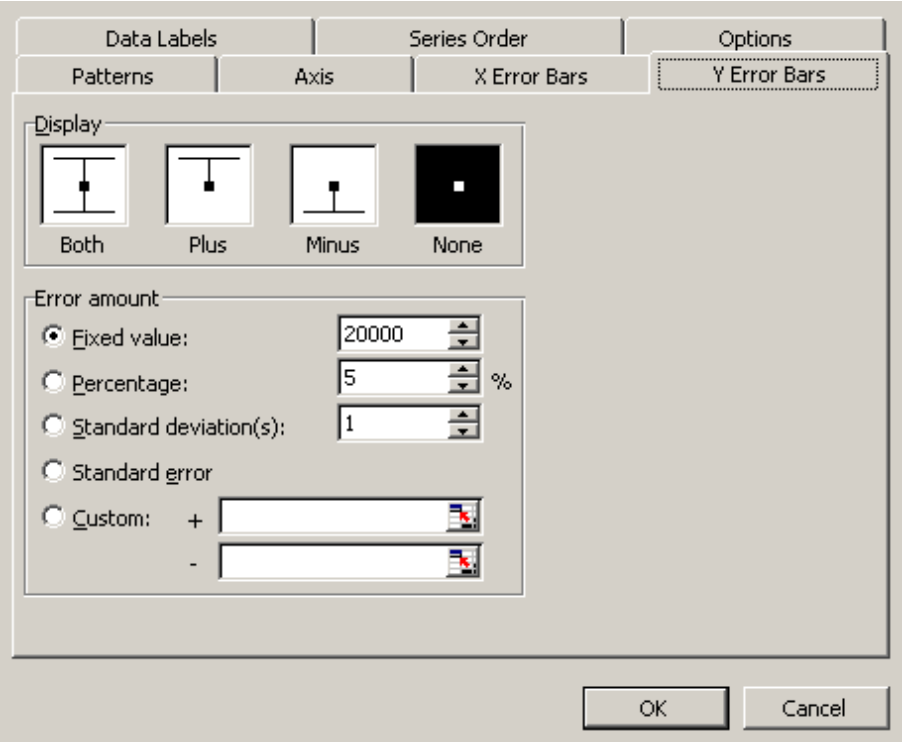


Figure 9: Options for error bar charts.

- Add control limits, computed from the data and based on the common cause variation only, equidistant on either side of the centre line.
- Apply tests to distinguish between data points resulting from special causes and data points resulting from common causes only.

When a single data point is observed outside of one of the control limits, the probability is that this point is not a real change is around 3 out of 1000. Additional tests can be applied to indicate, with a high probability, that a real change has occurred. Using several tests simultaneously increases the sensitivity of inferences but also increases the possibility of a false positive. Wheeler (1993) uses three simple tests:

- Test 1: a single point outside of control limits.
- Test 2: three out of four consecutive points closer to the control limit than to the centre line.
- Test 3: eight or more successive points on one side of the centre line.

Figure 11 shows weekly corrected feed conversion ratios (cFCR) measured for batches of broilers on a poultry farm. cFCR values range between 1.5 and 2.0 kg/kg. Date and cFCR have been recorded in columns A and B. In column C the `IF` function has been used to return the value recorded in column B if the average of the previous three weeks cFCR is greater than 1.85 kg/kg and `#N/A` otherwise.

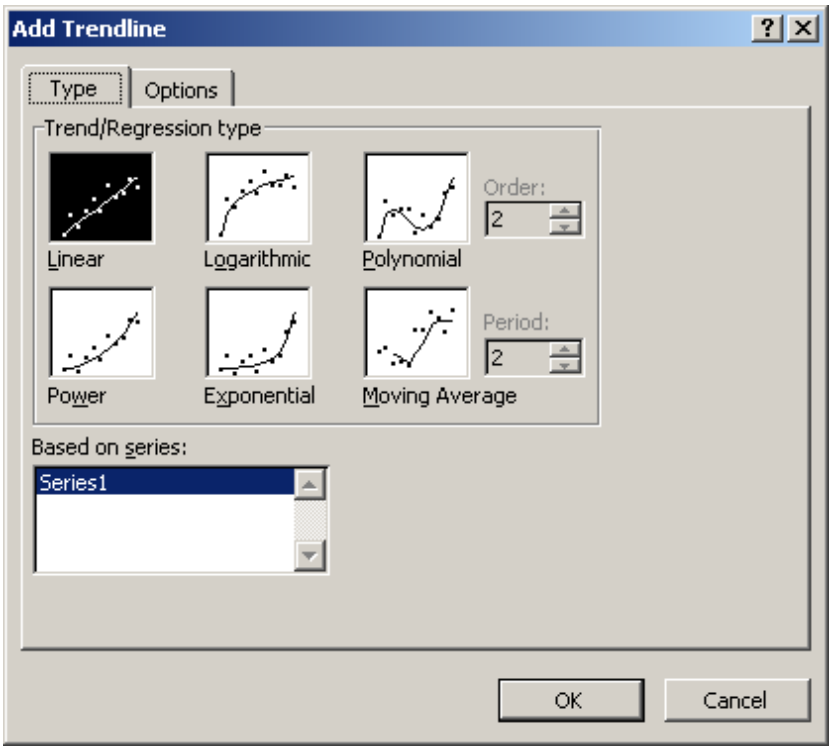


Figure 10: Options for adding trendlines to scatterplot graphs.

Frequency histograms

Histograms are used to summarise discrete or continuous data that are measured on an interval scale. A histogram divides up the range of possible values in a data set into classes or groups. For each group, a rectangle is constructed with a base length equal to the range of values in that specific group, and an area proportional to the number of observations falling into that group. This means that the rectangles will be drawn of non-uniform height. A histogram has an appearance similar to a vertical bar graph, but when the variables are continuous, there are no gaps between the bars. When the variables are discrete, however, gaps should be left between the bars. The data shown in column A in Figure 12 is to be plotted as a frequency histogram. The data bins to be used are shown in column C.

With the Analysis ToolPak loaded, click -TOOLS-DATA ANALYSIS... and select the Histogram option. Define the input range (cells `A2:A100` in this example), the bin range (cells `C2:C21E1`) and click OK to produce a graph similar to Figure 13. Note that this graph has been manipulated to eliminate spaces between the bars of the histogram.

The histogram function in Excel works with data formated as dates, as well as numbers. This facility is useful for plotting epidemic curves.

1.5 Simple random sampling

You are presented with a herd of 100 cows with ear tags 1 to 100. You would like to take a simple random sample of 20 from this group. Fill column A of your spreadsheet with values 1

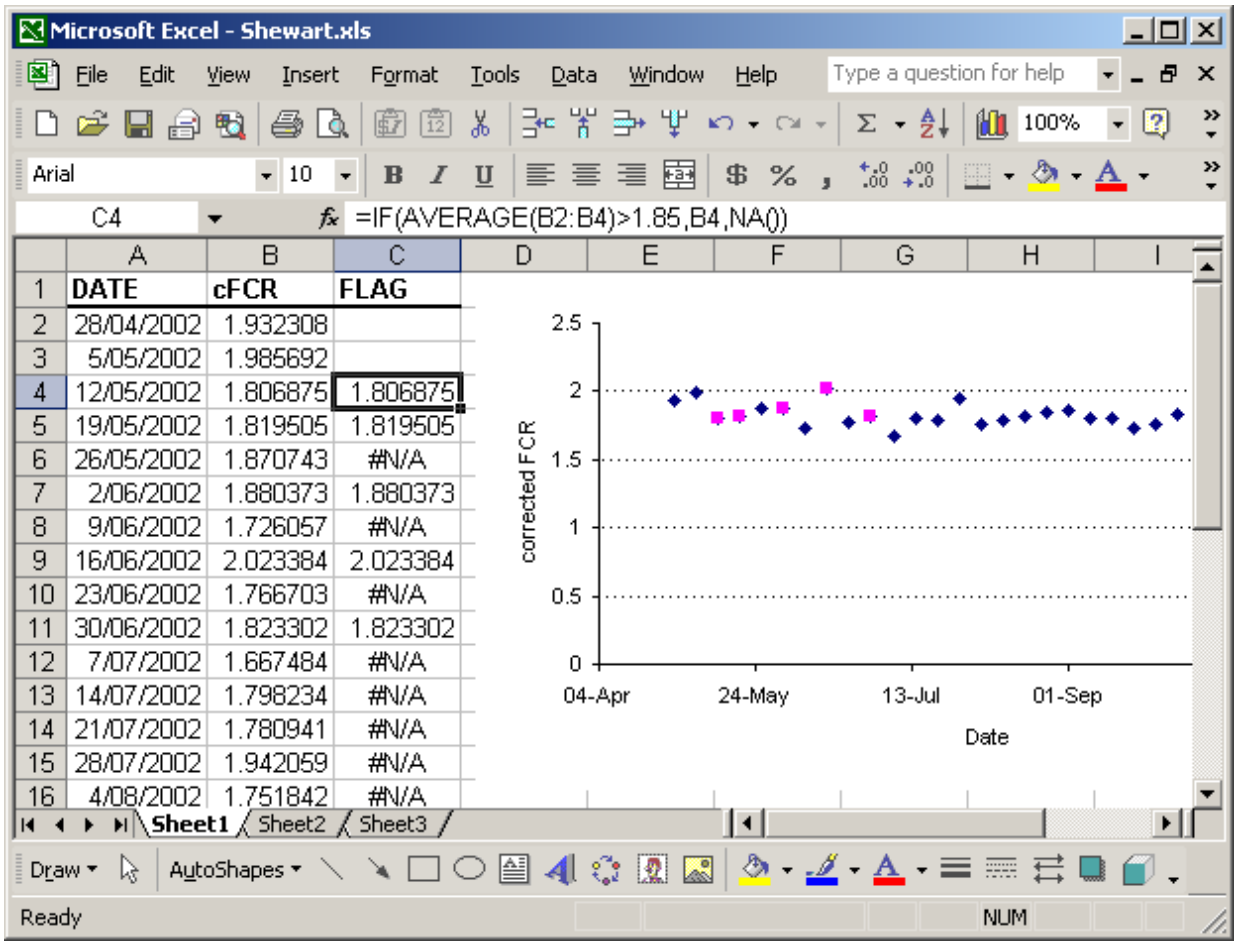


Figure 11: A time series chart. Data points where the average of the readings taken over the previous three days is greater than 1.85 kg/kg are shown in pink.

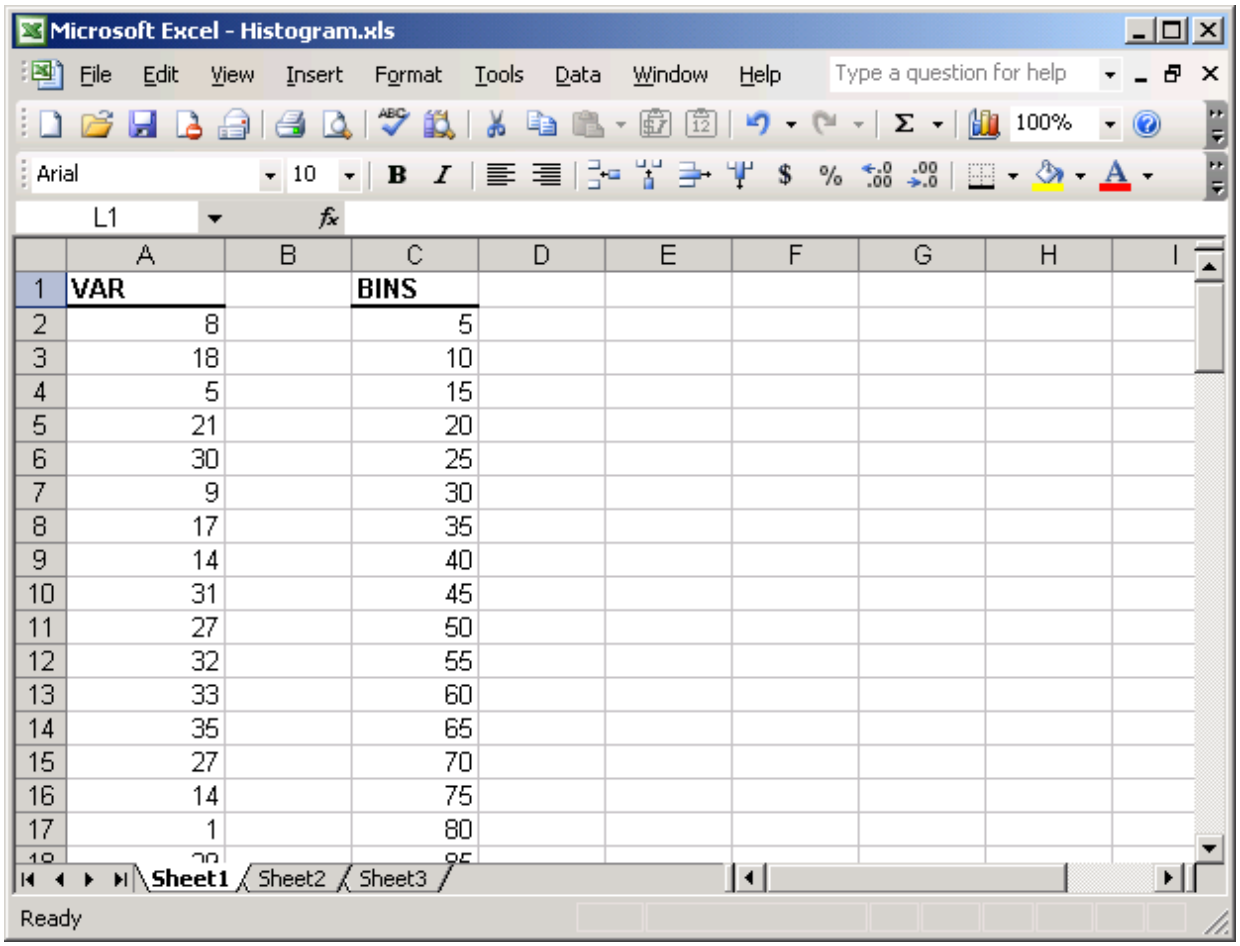
to 100. In cell `A1` enter 1. Select column A by clicking the column heading then select -EDIT-FILL-SERIES. Enter values in the dialogue box as shown in Figure 14 and click OK. In practice it would be usual to copy and paste the identifiers of animals eligible for sampling from a herd health software package into your spreadsheet.

Column A should now contain the numbers 1 to 100. We now fill rows 1 to 20 of column C with a random sample selected from column A. Select -TOOLS-DATA ANALYSIS-SAMPLING. Enter values in the sampling dialogue box as shown in Figure 15 and click OK.

Column C will now contain a random sample of 20 values from Column A.

1.6 Pivot tables

In epidemiology, pivot tables provide a quick way to retrieve event counts that can be used to populate a 2×2 table. Imagine you are interested in working out the association between cows that have a veterinary assisted dystocia and a diagnosis of metritis during the first 50 days of the subsequent lactation. The first 17 records of your data set are shown in Figure 16.



	A	B	C	D	E	F	G	H	I
1	VAR		BINS						
2	8		5						
3	18		10						
4	5		15						
5	21		20						
6	30		25						
7	9		30						
8	17		35						
9	14		40						
10	31		45						
11	27		50						
12	32		55						
13	33		60						
14	35		65						
15	27		70						
16	14		75						
17	1		80						

Figure 12: Frequency histograms in MS Excel. The data to be plotted are shown in column A. The bins to be used for the histogram are shown in column C.

From the menu bar see -DATA-PIVOT TABLE AND PIVOT CHART REPORT. Select PivotTable as the type of report you want to create, then click Next. Define the source data and again, click Next. Select the option to create the report in a New worksheet, and then, before you click finish, click Layout. Drag the box titled 'METRITIS' (the outcome) into the pivot table column. Drag the box titled 'DYSTOCIA' (the exposure) into the pivot table row. Finally, drag the box titled 'COW' into the centre of the pivot table, and then double click on it. Select to summarise cow as a count (rather than the default, sum). The form should look like Figure 17.

Click OK, then Finish to run the report. The completed report should look similar to Figure 18. Due to the way DYSTOCIA and METRITIS have been coded, the default table is inappropriately formated for estimation of risk and odds ratios. A correctly formated table, referencing each cell of the Pivot Table, is shown in cells `A9:G13`. A veterinary assisted dystocia increased the risk of metritis in the first 50 days of the subsequent lactation by a factor of 4.5.

1.7 Shortcuts

Table 9: Keyboard shortcuts for selecting data.

Key combination	Description
CTRL+SHIFT+*	Selects the current range around the active cell, the largest rectangle of data surrounded by white space.
SHIFT+↑↓	Select cells around and active cell.
END+SHIFT+↑ ↓	Select cells around and active cell to the first non blank cell in the direction selected.
CTRL+A	Selects all the cells in the spreadsheet.
CTRL+SPACEBAR	Selects an entire column.
CTRL+SHIFT	Selects an entire row.
CTRL+SHIFT+END	Extend the selection to the last cell in the worksheet.
CTRL+SHIFT+HOME	Extend the selection to the last non blank cell in the same column or row as the active cell.

Table 10: Keyboard shortcuts for moving around a workbook.

Key combination	Description
PAGE UP	Move up on screen.
PAGE DOWN	Move down on screen.
ALT+PAGE UP	Move one screen to the left.
ALT+PAGE DOWN	Move one screen to the right.
CTRL+HOME	Move to the beginning of the worksheet.
CTRL+END	Move to the end of the worksheet.
CTRL+PAGE DOWN	Move to the next sheet in the workbook.
CTRL+PAGE UP	Move to the previous sheet in the workbook.
CTRL+F6	Move to the next workbook or window.
CTRL+TAB	Move to the next workbook or window.

Table 11: Miscellaneous shortcuts.

Key combination	Description
F4	Toggles between absolute and relative references.
SHIFT+F9	Recalculates the active worksheet.
CTRL+ALT+F9	Recalculates all open worksheets.
CTRL+C	Copies the highlighted section.
CTRL+X	Cuts the highlighted section.
CTRL+V	Pastes the selection on the Windows clipboard.
CTRL+;	Enter the current date.
CTRL+SHIFT+:	Enter the current time.
CTRL+ENTER	Fill the selected cells with the current entry.
SHIFT+F2	Create a cell comment.
ALT+back space	Undo the last command.
ALT+ENTER	Repeat the last command.
CTRL+F	Choose a font.

Table 12: Error messages.

Error	Cause
#DIV/0	Attempting to divide by zero or by an empty cell.
#NAME	Formula contains an undefined variable or function name.
#N/A	No value is available; this can be entered as =NA().
#NULL	The result has no value.
#NUM	Numeric overflow for example when asking for the square root of a negative value.
#REF	Invalid cell reference.
VALUE	Invalid argument (for example asking for the sum when the cell contains text).
######	The column is to narrow to view the result or the formula returns an invalid date e.g. prior to 1900 or negative time

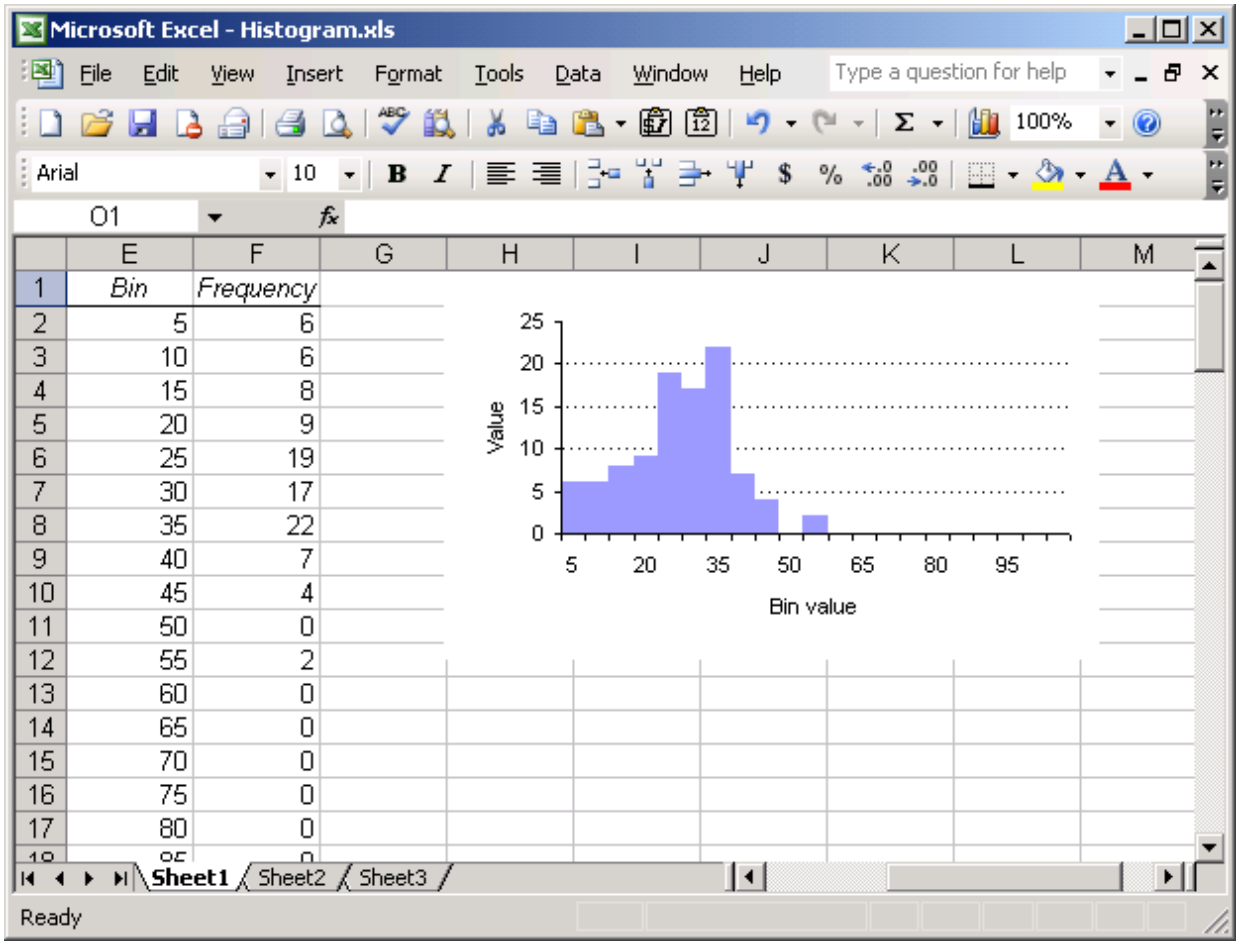


Figure 13: Frequency histograms in MS Excel. Bin and frequency values are shown in cells E1 to F22. A frequency histogram, based on these data, is shown in cells H1:L12.

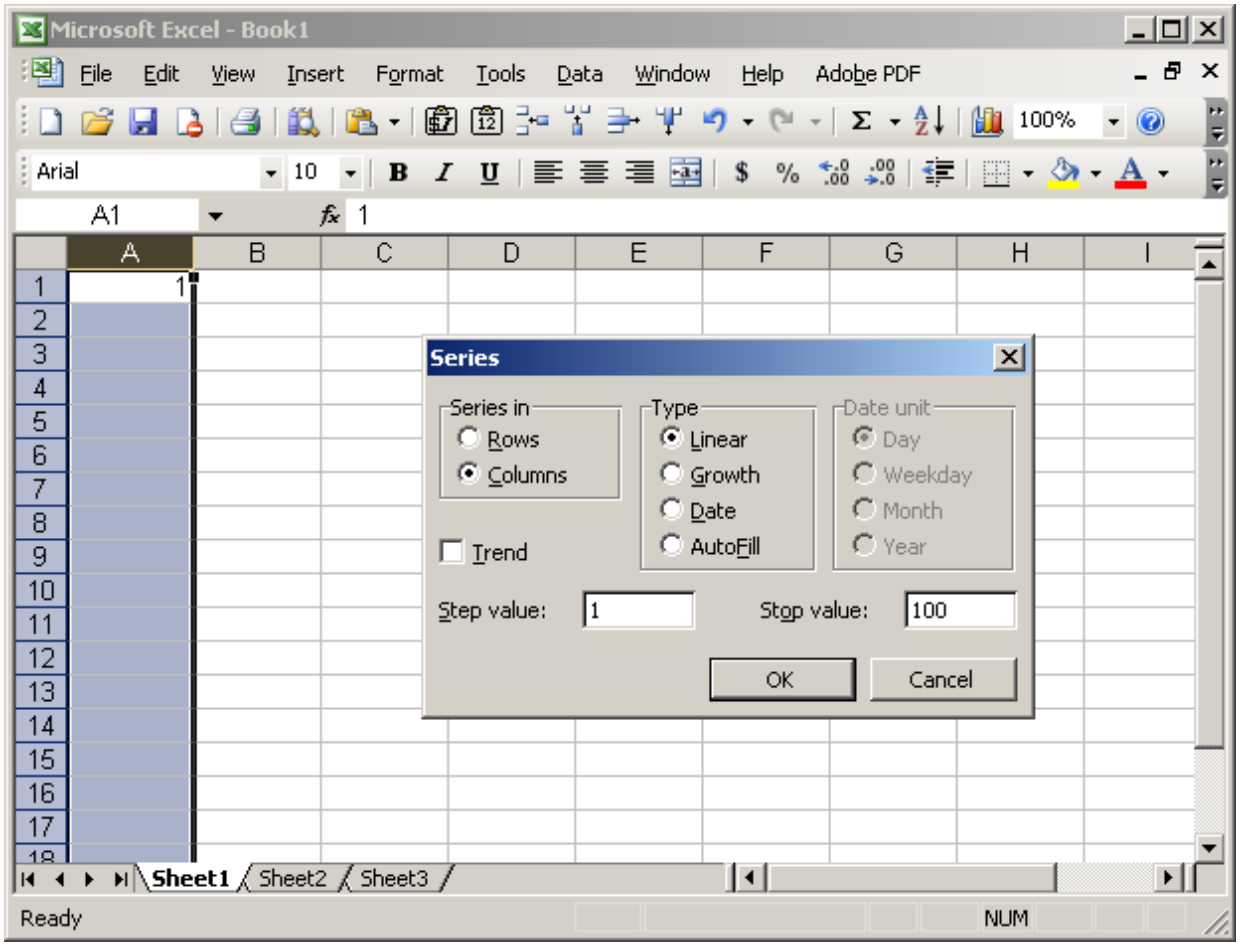


Figure 14: Filling column A with a linear series of numbers from 1 to 100.

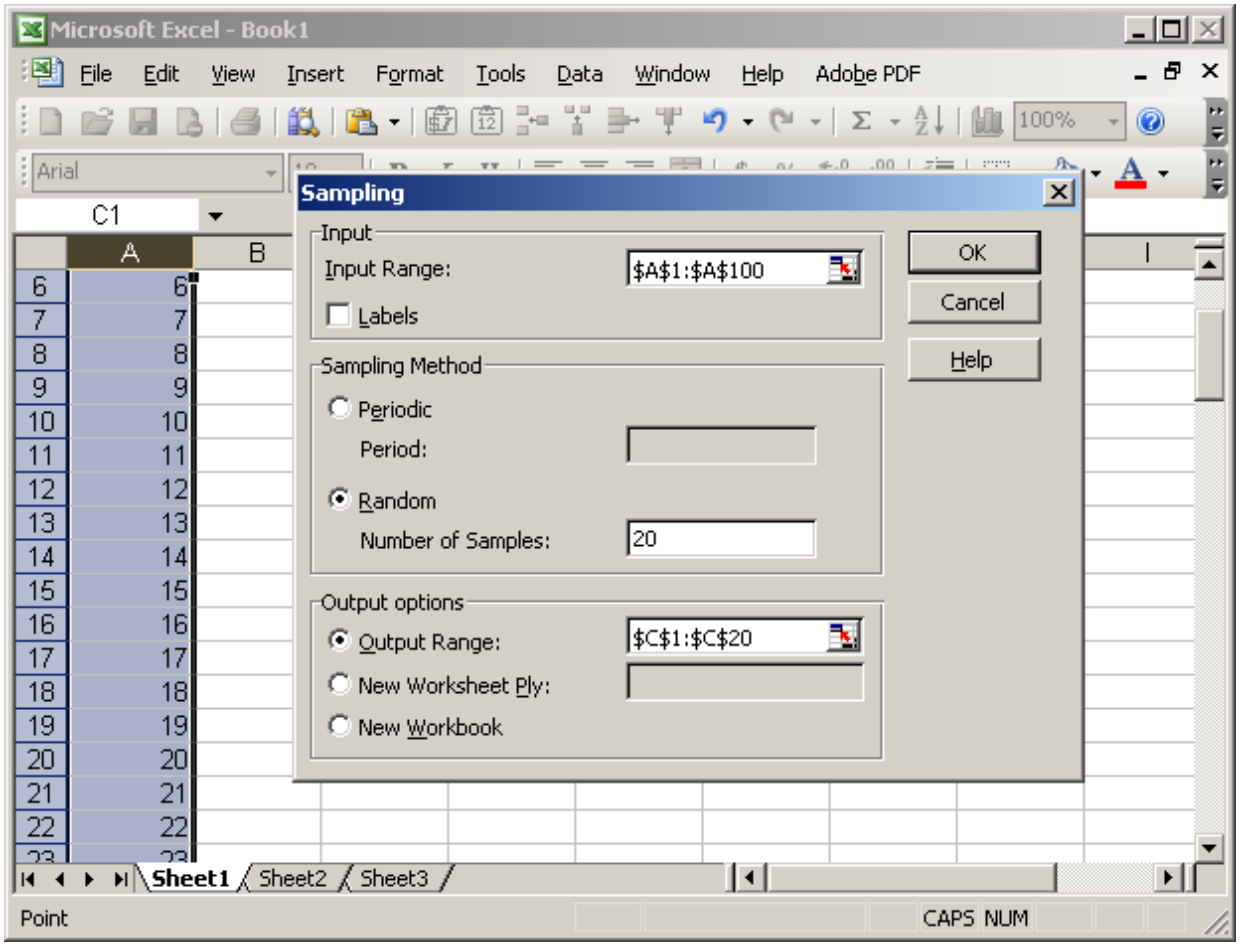
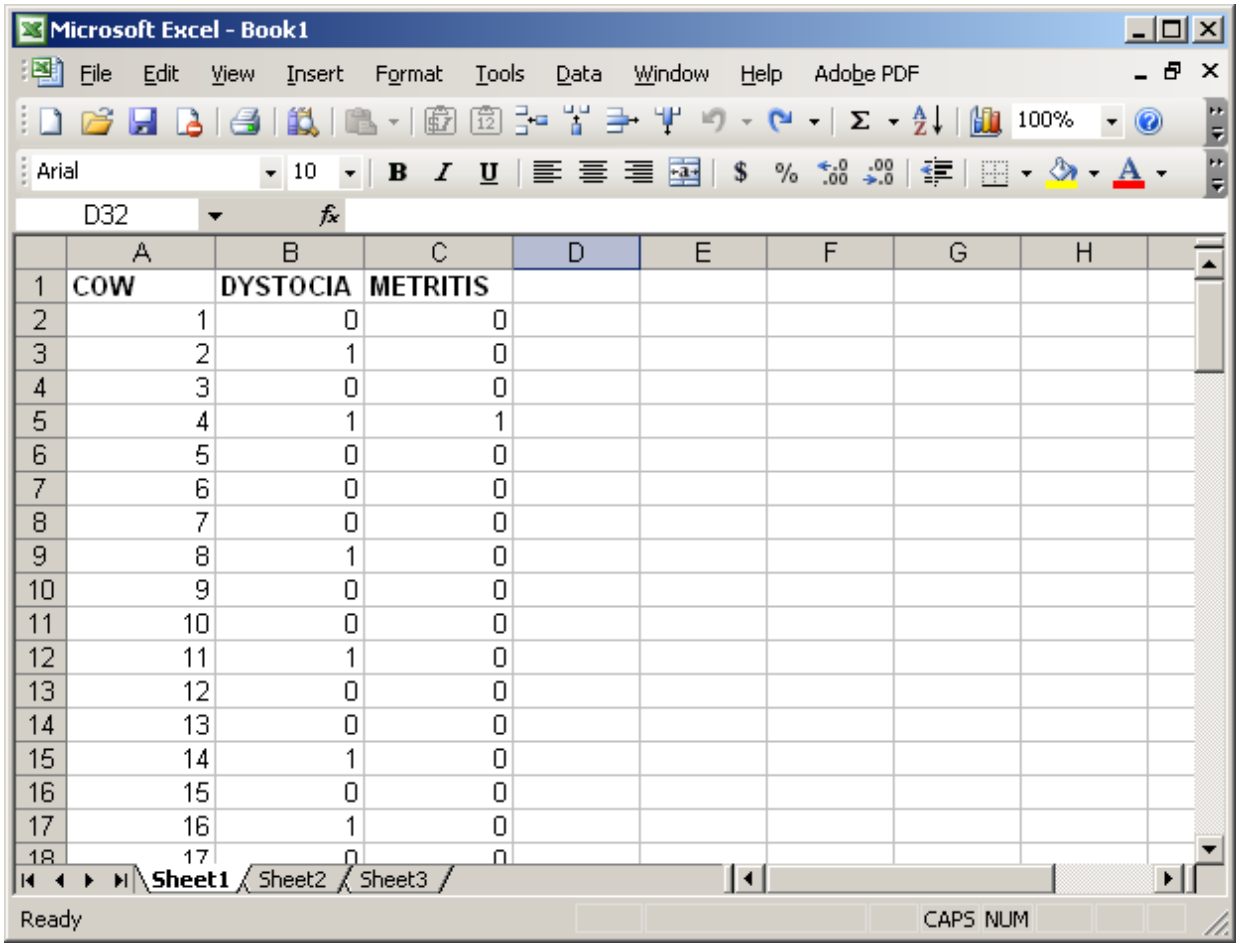


Figure 15: Values entered in the sampling dialogue box, specifying that 20 samples will be taken from cells $A1{:}A100$.



	A	B	C
1	COW	DYSTOCIA	METRITIS
2	1	0	0
3	2	1	0
4	3	0	0
5	4	1	1
6	5	0	0
7	6	0	0
8	7	0	0
9	8	1	0
10	9	0	0
11	10	0	0
12	11	1	0
13	12	0	0
14	13	0	0
15	14	1	0
16	15	0	0
17	16	1	0

Figure 16: Quantifying the association between dystocia and metritis during the first 50 days of the subsequent lactation. Arrangement of individual cow data in MS Excel. There are three columns: column A cow identifier, column B dystocia flag (0 = dystocia absent 1 = dystocia present), column C metritis flag (0 = metritis absent, 1 = metritis present).

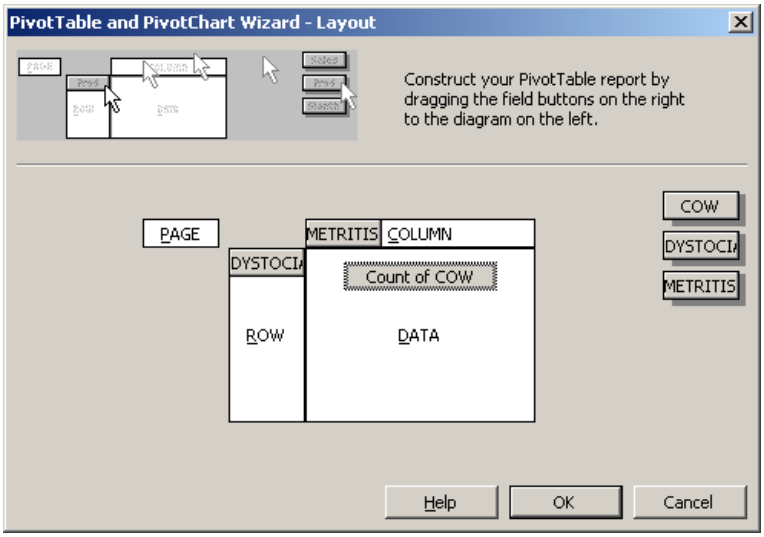


Figure 17: Layout of the Pivot Table report: METRITIS (the outcome) status as columns, DYSTOCIA (the exposure) status as rows. Counts of cows in each cell of the table will be returned using the 'Count of Cow' option.

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help Adobe PDF

Arial 10

B10 =GETPIVOTDATA("COW",$A$3,"DYSTOCIA",1,"METRITIS",1)

PivotTable

	A	B	C	D	E	F	G	H
1								
2								
3	Count of COW	METRITIS						
4	DYSTOCIA	0	1	Grand Total				
5	0	82	8	90				
6	1	6	4	10				
7	Grand Total	88	12	100				
8								
9		Metritis +	Metritis -	Total				
10	Dystocia +	4	6	10		R_E	0.40	
11	Dystocia -	8	82	90		R_O	0.09	
12		12	88	100				
13						RR	4.5	
14								
15								
16								
17								

Sheet6 Sheet1 Sheet2 Sheet3

Ready NUM

Figure 18: Completed Pivot Table report.