



MS Excel Short Course in Data Handling & Statistical Function

*A basic course for new users of MS Excel with
a special emphasis on data handling, data
organizational and statistical functions.*

Developed by
the International Centre for Allied Health Evidence (iCAHE)
School of Health Science, Division of Health Science, University of South
Australia

Intention of the Course

This booklet provides you with a structured, enjoyable and practical approach to learn to use MS Excel. These skills will enable you to undertake research and data handling activities within your clinical practice.

This booklet is designed for all skill levels, from absolute beginner to more advanced knowledge.



University of
South Australia

International Centre for
Allied Health Evidence

iCAHE



Content of the booklet

In the first session of this course, you will learn:

- ◆ About the properties of MS Excel
- ◆ Why use MS Excel
- ◆ The difference between a spread sheet, a database and a word document
- ◆ How to get into Excel
- ◆ How to open a new file
- ◆ How to enter data
- ◆ How to save a file
- ◆ How to open an existing file
- ◆ How to copy and paste data

In the second session of this course, you will learn:

- ◆ Shortcuts to assist data entry
 - ◆ Fill
 - ◆ Data filter
- ◆ How to move data around
- ◆ How to sort data
- ◆ How to delete and clear data
- ◆ How to format data
 - ◆ Different colours
 - ◆ Different fonts
 - ◆ Different data types
 - ◆ Different borders
- ◆ How to interpret a prepared data sheet

In the third session of this course, you will learn:

- ◆ How to change a column of data (multiplying, adding, subtracting etc.)
- ◆ How to link two columns of data
- ◆ How to apply formulas
 - ◆ Means, standard deviations, maximum, minimum, percentiles
 - ◆ Change data
- ◆ How to apply statistics tests
 - ◆ T-test, ANOVA, linear regression

In the fourth session of this course, you will learn:

- ◆ How to make graphs
- ◆ How to change graphs
- ◆ How to convert data from Word to Excel and back
- ◆ How to manage audit data

Self-check to complete at the end of the booklet*How well do you understand these aspects of MS Excel?**My understanding is:*

Core competencies	Good	OK	Poor
About the properties of MS Excel			
Why use MS Excel			
The difference between a spread sheet, a database and a word document			
How to get into Excel			
How to open a new file			
How to enter data			
How to save a file			
How to open an existing file			
How to copy and paste data			
Paste Special			
Shortcuts to assist data entry			
♦ Fill			
♦ Drag Cells			
♦ Data filter			
How to move data around			
How to sort data			
How to delete and clear data			
How to format data			
♦ Different colours			
♦ Different fonts			
♦ Different data types			
♦ Different borders			
How to interpret a prepared data sheet			
How to change a column of data			
How to link two columns of data			
How to apply formulas			
♦ Means, standard deviations etc			
How to apply statistics tests			
♦ T-test, ANOVA, linear regression, chi square			
How to make graphs			
How to change graphs			
How to convert data from Word to Excel and back			
How to manage audit data			



Important Information

This booklet is based on MSOffice 2010.

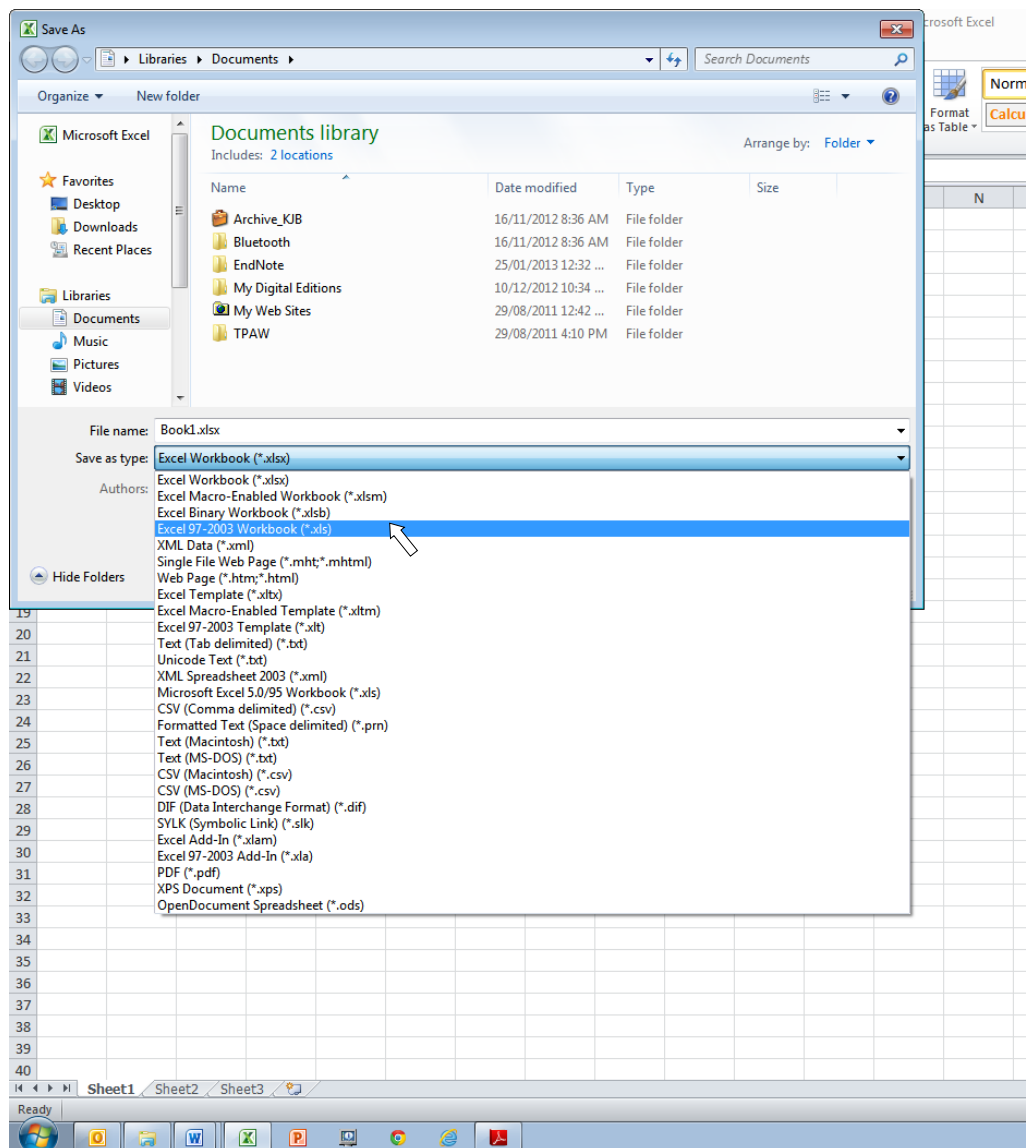
If you have an older version of MSOffice on your laptop and/or computer, please make sure to save all your files in the 97-2003 formats.

93-2003 formats are saved as *.doc or *.xls or *.ppt

2010 files are saved as *.docx, *.xlsx or *.pptx

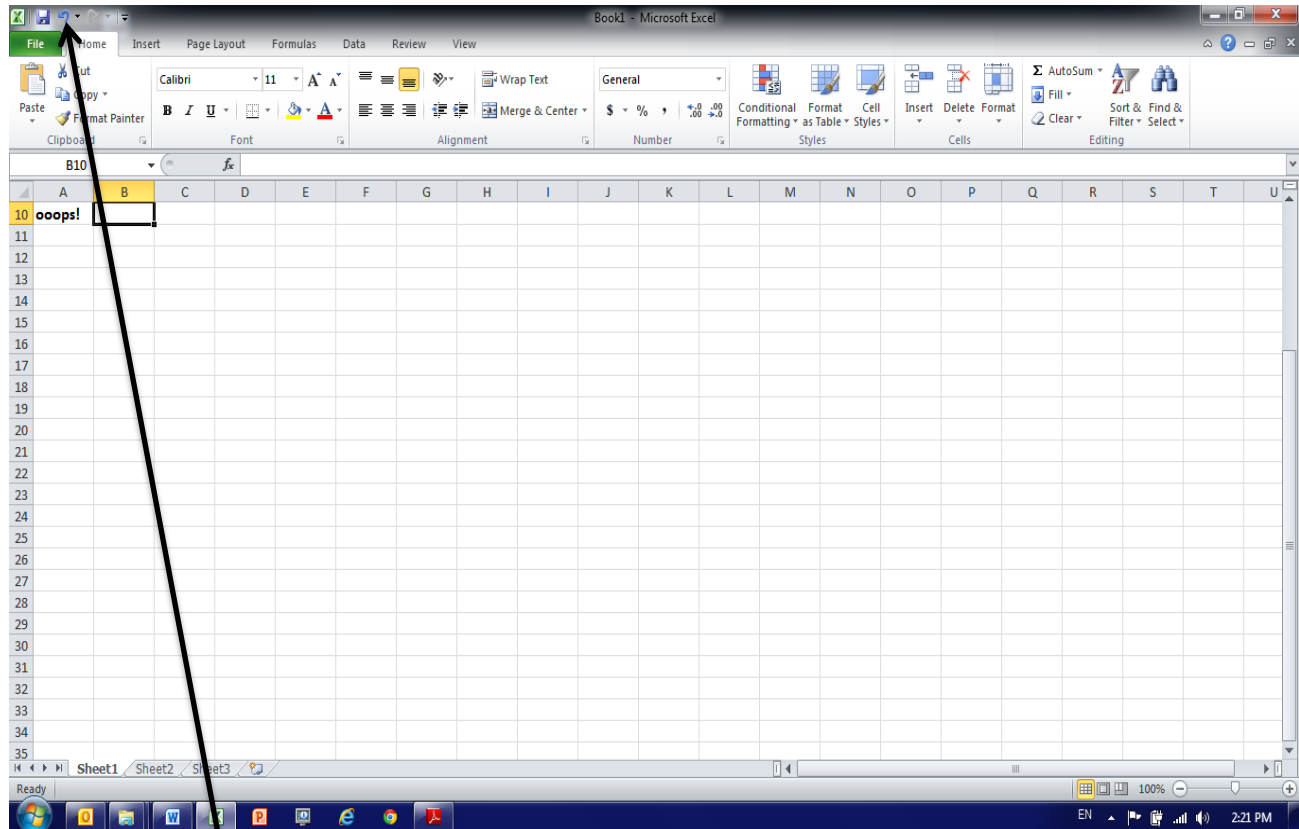
MSOffice 2010 has the option of saving files compatible to an older version in all of their individual programs like Word, Excel, and PowerPoint etc.

This is how you can do it: -



About the properties of MS Excel

Excel is a spreadsheet – what you see on the screen is a series of cells, into which data can be entered. The spreadsheet offers a simple way of collecting and examining data. Moving from one cell to the next is by cursors, or your mouse. To embed a value in a cell, or to complete a calculation, hit the ENTER key.

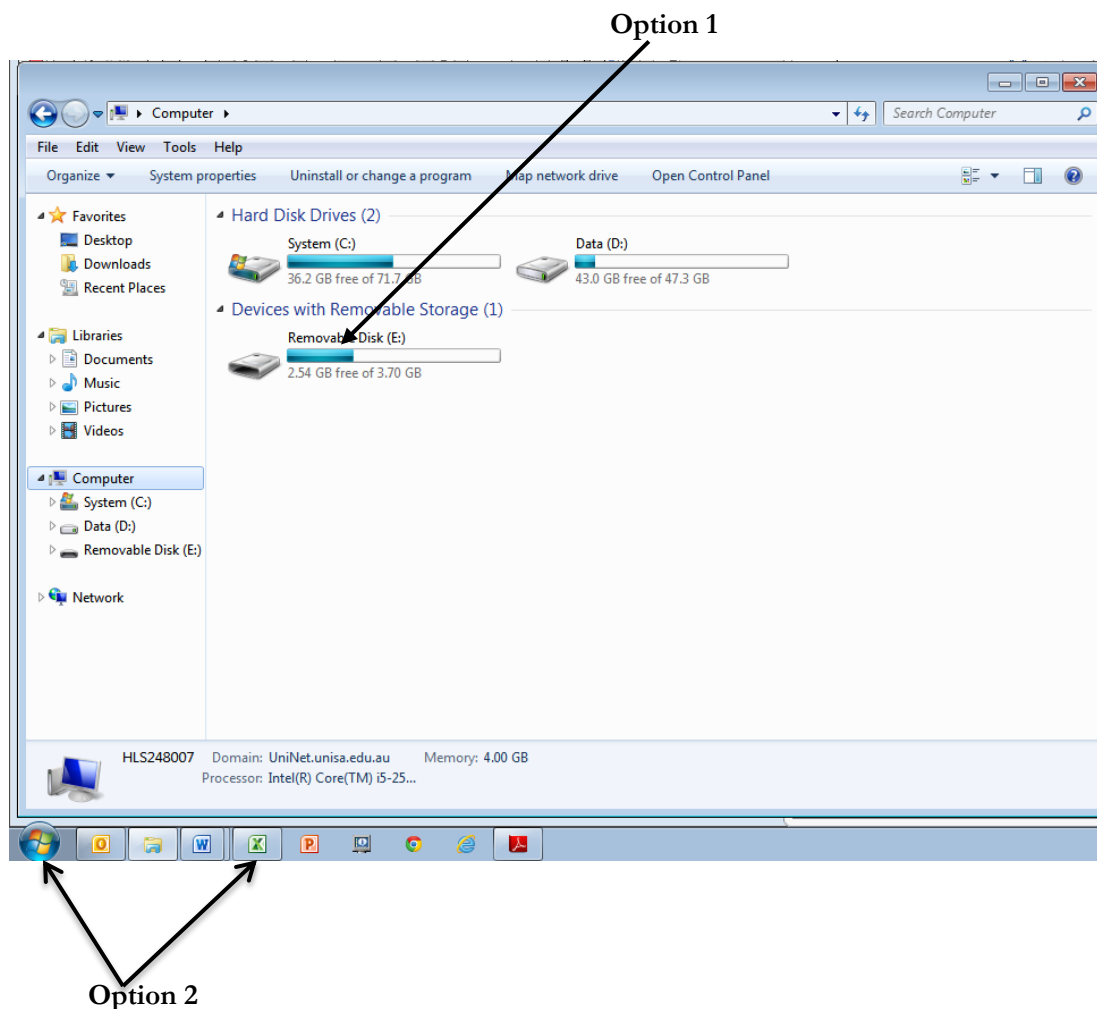


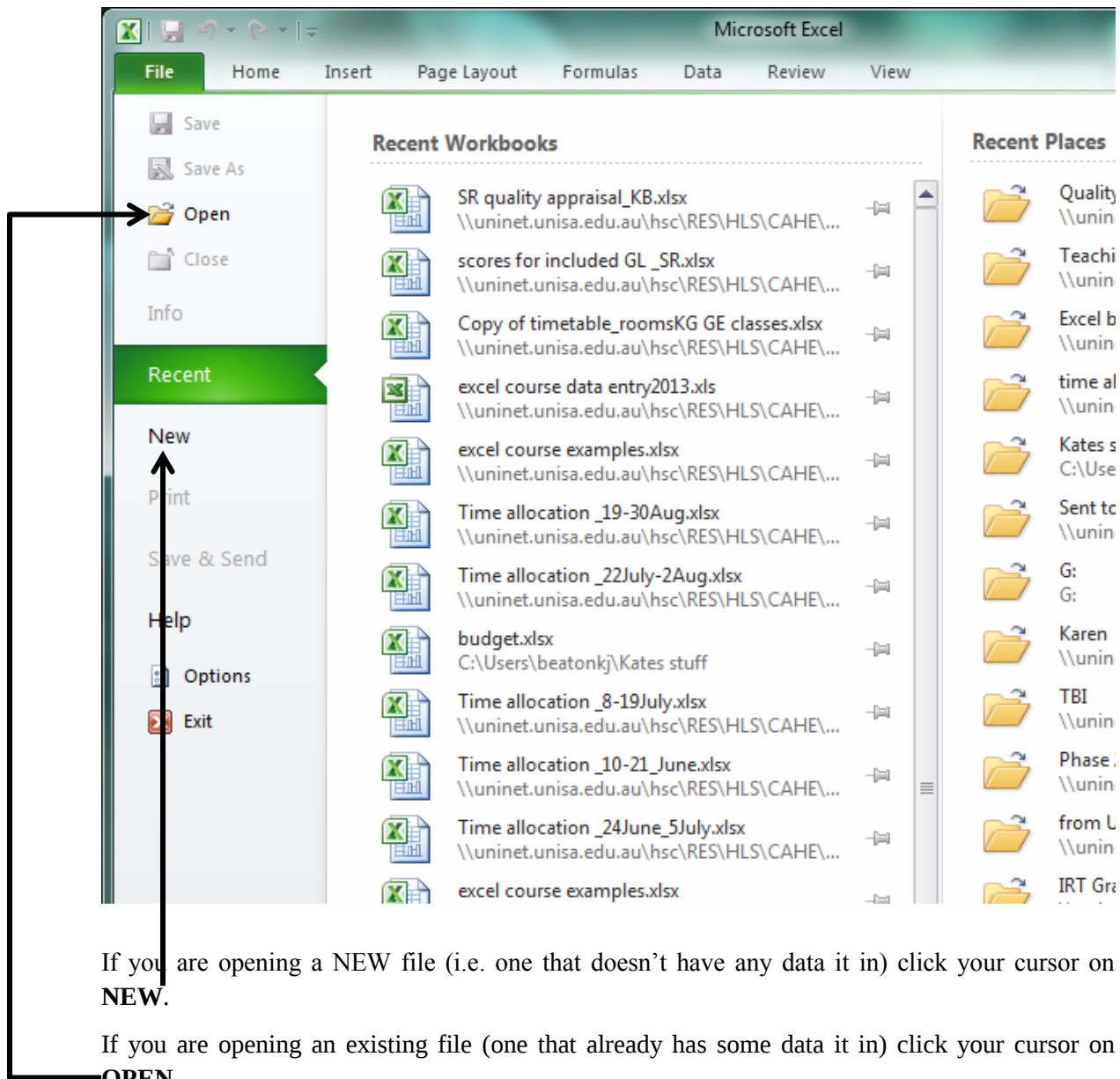
REMEMBER: if you make a mistake, **DON'T PANIC!** Immediately go to the undo button [the little back arrow] and click – that will take you back 1 step. So if you do something that you aren't happy with, think before you attempt to do anything else to retrieve a situation!

How to open a file

To open a file in Excel, you have a number of options.

1. If you are working from a USB, you can double click the USB icon under Computer in Windows, and you will see all of the existing files on the USB. Double click on the file that you want, to open it.
2. You can open up MSExcel from the green MSExcel icon, or from the program files, and then open up your file by going through the File tab. Again, double click on the file you want to open.





If you are opening a NEW file (i.e. one that doesn't have any data in it) click your cursor on **NEW**.

If you are opening an existing file (one that already has some data in it) click your cursor on **OPEN**.

You can always tell which files have been generated and saved in MS Excel because they have the file extension of .xls or .xlsx.

Similarly a Word file has an extension .doc or .docx and a MS Powerpoint file has an extension of .ppt or .pptx.

Exercise 1

You have been given one MS Excel file (Excel Course Data Entry.xlsx) with three named sheets in it, as a practice piece. Practice opening and closing the file and looking at each of the sheets.

How to enter data

Page 5 shows you an example of what a NEW Excel file looks like (a clean spreadsheet into which you can enter data). You can put names in for each column which make sense to you, colour cells to code or organize data, add, remove or change borders and you can also nominate the type of data that you expect to find in each column. Like this...

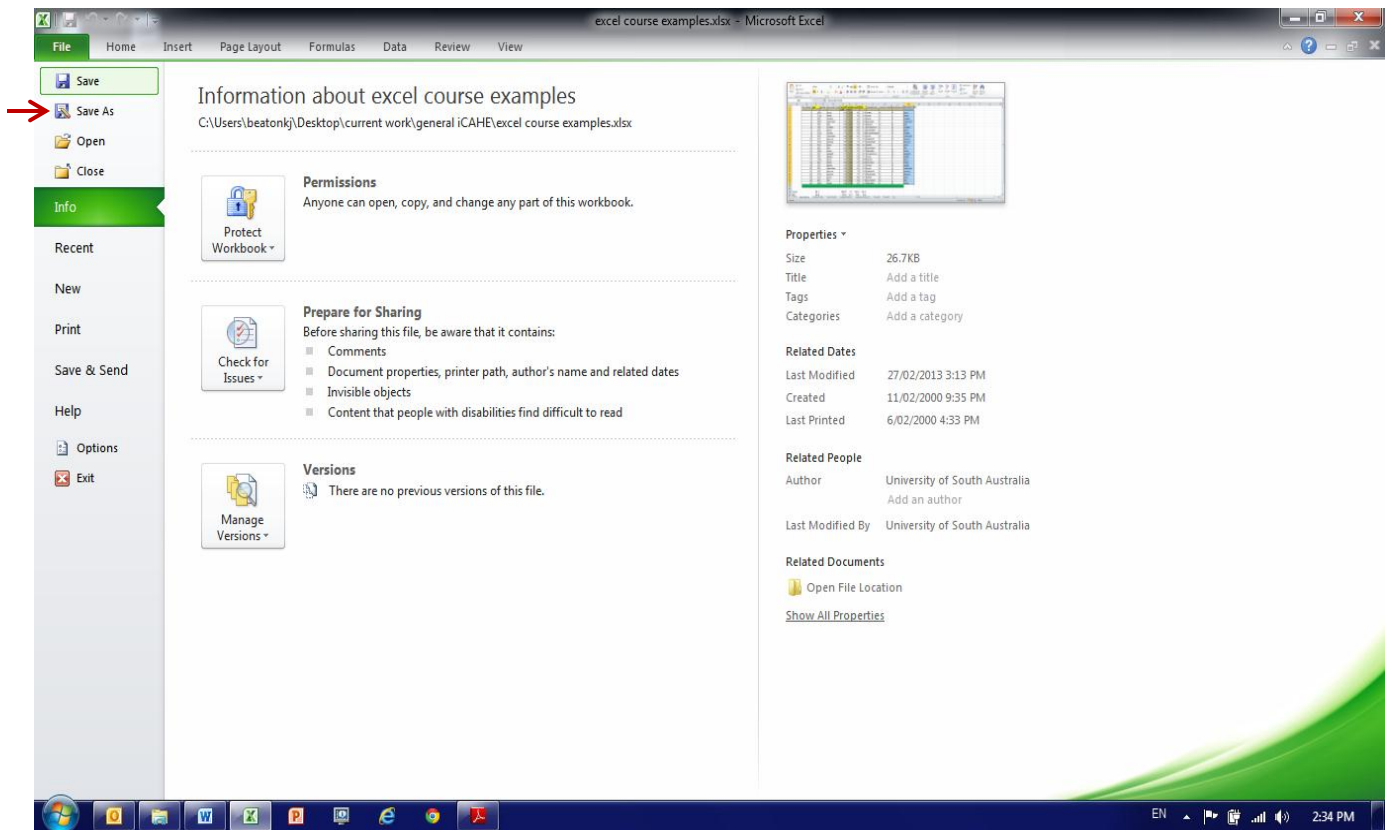
MAIN HEADING	ID number	Today's Date	Date of Birth	Dominant Hand	Country of Birth	Mother's Country of Birth	Father's Country of Birth	Bag type	Bag brand
SUB HEADING									
ANSWER DETAILS								Question 1	Question 2
								1_2_3	
									10 characters max
4 digits		11 digits	11 digits	R_L	8 characters max	8 characters max	8 characters max		
QUESTION CODE		date03	dob03	hand03	born03	p1brn03	p2brn03	bagt03	brand03
EXAMPLE ANSWERS	1001	04-Mar-02	27-Jun-86	R	Australia	Australia	England	1	hot chilli
	1003	04-Mar-02	16-Nov-85	R	Australia	Australia	England	1	blackwolf
	1004	05-Mar-02	25-Nov-85	R	Australia	Australia	Australia	1	go pak

How to save a file

To save a file, always go 'SAVE AS' instead of save, because that allows you to double check your file name before saving. This stops you saving over the top of another file. When you go 'SAVE AS' check the **location** of where you are saving the file to, as well as its **name**. Sometimes a lot of work gets lost because the file was saved to the wrong directory! The first example Excel page on the next page of this workbook shows you where to find the 'Save' and 'Save As' commands.

Exercise 2

Now, open your practice MS Excel file (Excel course data entry.xls). Go to the Sheet named 'Descriptive'. Enter data for another five subjects into this file (make up anything you want), and save your file using a new file name 'My File.xlsx'. Close the file. Now open it up again and check that your new data is there, on the same sheet!!! The second example on the next page is of the data you should expect to find on the sheet named 'Descriptive' when you first open it up.



Exercise 3.

Open a new Excel file, and label the first four columns **ID, gender, age, height**, and then enter fictitious data for five people. Give the file a name and save it to your USB. Now save the file into the computer hard drive. Close and open both files to see that you have saved and labeled them correctly.

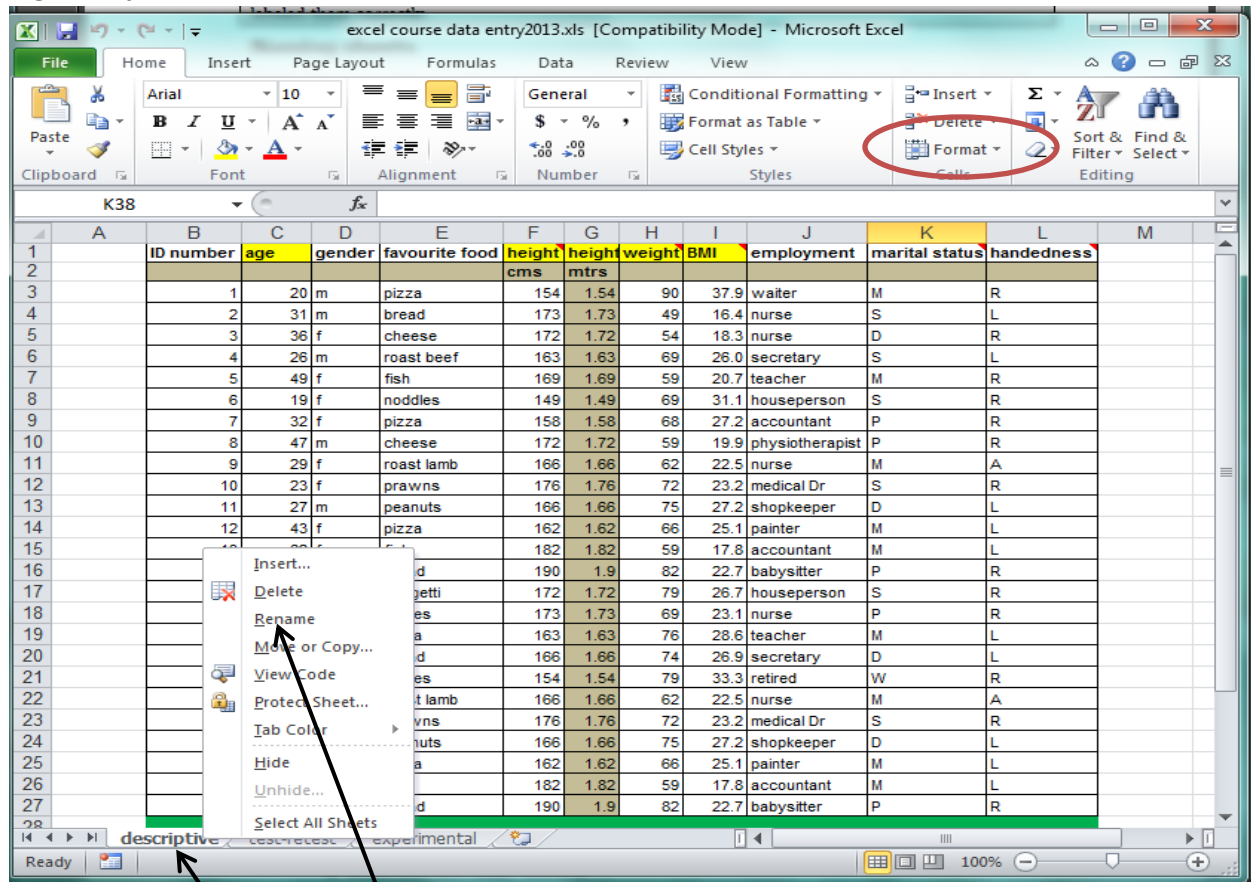
Naming sheets

Excel files are workbooks which contain a lot of sheets (or pages). Imagine your Excel file like an exercise book, a really big one that you can keep adding to! You can put your calculations on different sheets and give each sheet a unique name so you know where everything is stored. Your practice Excel file is an example of an Excel file in which the first three sheets have been renamed to reflect the type of data found in each sheet.

You rename a sheet by either

- going to the Format button on the home ribbon (circled in red on the next page), and selecting rename sheet (all using the left mouse button) OR
- putting the cursor on the sheet label (at the bottom) that you want to rename, and using the right mouse button, click the name, choose 'rename' from the pop up menu and enter in another name. OR
- using the left mouse button, double click on the name you wish to change and type in the new name

Try renaming the sheets in your course file 'Excel course data entry.xlsx', using names that are logical to you.



Right click and select 'Rename' from the pop up menu, OR double left click.

How to copy and paste data

Using your left mouse button, highlight the data in one column, go to the Menu and click on COPY (or CUT) and then click your cursor into the first cell of the new column into which you want to paste your data.

Then put your cursor on the PASTE icon and Click!!! The data then appears in the new column. It is wise if you are uncertain to always use COPY rather than CUT, and then when you are sure that your data is securely placed in the new location, you can always delete the old column!

MS Excel Short Course in Statistics

excel course examples.xlsx - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M
		id	height	height	weight	BMI	employment	marital status	handedness	favourite food			
		cms	mtrs										
1		1	20	m	pizza	154	1.54	90	37.9	waiter	M	R	pizza
2		2	31	m	bread	173	1.73	49	16.4	nurse	S	L	bread
3		3	36	f	cheese	172	1.72	54	18.3	nurse	D	R	cheese
4		4	26	m	roast beef	163	1.63	69	26.0	secretary	S	L	roast beef
5		5	49	f	fish	169	1.69	59	20.7	teacher	M	R	fish
6		6	19	f	noodles	149	1.49	69	31.1	houseperson	S	R	noodles
7		7	32	f	pizza	156	1.56	68	27.2	accountant	P	R	pizza
8		8	47	m	cheese	172	1.72	59	19.9	physiotherapist	P	R	cheese
9		9	29	f	roast lamb	166	1.66	62	22.5	nurse	M	A	roast lamb
10		10	23	f	prawns	176	1.76	72	23.2	medical Dr	S	R	prawns
11		11	27	m	peanuts	166	1.66	75	27.2	shopkeeper	D	L	peanuts
12		12	43	f	pizza	162	1.62	66	25.1	painter	M	L	pizza
13		13	62	f	fish	182	1.82	59	17.8	accountant	M	L	fish
14		14	49	m	bread	190	1.9	82	22.7	babysitter	P	R	bread
15		15	28	f	spaghetti	172	1.72	79	26.7	houseperson	S	R	spaghetti
16		16	36	f	apples	173	1.73	69	23.1	nurse	P	R	apples
17		17	18	m	pizza	163	1.63	76	28.6	teacher	M	L	pizza
18		18	62	f	bread	166	1.66	74	26.9	secretary	D	L	bread
19		19	32	f	apples	154	1.54	79	33.3	retired	W	R	apples
20		20	29	f	roast lamb	166	1.66	62	22.5	nurse	M	A	roast lamb
21		21	23	f	prawns	176	1.76	72	23.2	medical Dr	S	R	prawns
22		22	27	m	peanuts	166	1.66	75	27.2	shopkeeper	D	L	peanuts
23		23	43	f	pizza	162	1.62	66	25.1	painter	M	L	pizza
24		24	62	f	fish	182	1.82	59	17.8	accountant	M	L	fish
25		25	49	m	bread	190	1.9	82	22.7	babysitter	P	R	bread
30	mean		36.1			168.9	1.7	69.0	24.5				
31	SD		13.6			10.3	0.1	9.8	5.0				

excel course examples.xlsx - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
		id	height	height	weight	BMI	employment	marital status	handedness	favourite food					
		cms	mtrs												
1		1	20	m	pizza	154	1.54	90	37.9	waiter	M	R	pizza		
2		2	31	m	bread	173	1.73	49	16.4	nurse	S	L	bread		
3		3	36	f	cheese	172	1.72	54	18.3	nurse	D	R	cheese		
4		4	26	m	roast beef	163	1.63	69	26.0	secretary	S	L	roast beef		
5		5	49	f	fish	169	1.69	59	20.7	teacher	M	R	fish		
6		6	19	f	noodles	149	1.49	69	31.1	houseperson	S	R	noodles		
7		7	32	f	pizza	156	1.56	68	27.2	accountant	P	R	pizza		
8		8	47	m	cheese	172	1.72	59	19.9	physiotherapist	P	R	cheese		
9		9	29	f	roast lamb	166	1.66	62	22.5	nurse	M	A	roast lamb		
10		10	23	f	prawns	176	1.76	72	23.2	medical Dr	S	R	prawns		
11		11	27	m	peanuts	166	1.66	75	27.2	shopkeeper	D	L	peanuts		
12		12	43	f	pizza	162	1.62	66	25.1	painter	M	L	pizza		
13		13	62	f	fish	182	1.82	59	17.8	accountant	M	L	fish		
14		14	49	m	bread	190	1.9	82	22.7	babysitter	P	R	bread		
15		15	28	f	spaghetti	172	1.72	79	26.7	houseperson	S	R	spaghetti		
16		16	36	f	apples	173	1.73	69	23.1	nurse	P	R	apples		
17		17	18	m	pizza	163	1.63	76	28.6	teacher	M	L	pizza		
18		18	62	f	bread	166	1.66	74	26.9	secretary	D	L	bread		
19		19	32	f	apples	154	1.54	79	33.3	retired	W	R	apples		
20		20	29	f	roast lamb	166	1.66	62	22.5	nurse	M	A	roast lamb		
21		21	23	f	prawns	176	1.76	72	23.2	medical Dr	S	R	prawns		
22		22	27	m	peanuts	166	1.66	75	27.2	shopkeeper	D	L	peanuts		
23		23	43	f	pizza	162	1.62	66	25.1	painter	M	L	pizza		
24		24	62	f	fish	182	1.82	59	17.8	accountant	M	L	fish		
25		25	49	m	bread	190	1.9	82	22.7	babysitter	P	R	bread		
30	mean		36.1			168.9	1.7	69.0	24.5						
31	SD		13.6			10.3	0.1	9.8	5.0						

excel course examples.xlsx - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
		id	number	age	gender	favourite food	height	height	weight	BMI	employment	marital status	handedness	favourite food	weight						
							cms	mtrs													
1		1	20	m	pizza	154	1.54	90	37.9	waiter	M	R	pizza	90							
2		2	31	m	bread	173	1.73	49	16.4	nurse	S	L	bread	49							
3		3	36	f	cheese	172	1.72	54	18.3	nurse	D	R	cheese	54							
4		4	26	m	roast beef	163	1.63	69	26.0	secretary	S	L	roast beef	69							
5		5	49	f	fish	169	1.69	59	20.7	teacher	M	R	fish	59							
6		6	19	f	noodles	149	1.49	69	31.1	houseperson	S	R	noodles	69							
7		7	32	f	pizza	156	1.56	68	27.2	accountant	P	R	pizza	68							
8		8	47	m	cheese	172	1.72	59	19.9	physiotherapist	P	R	cheese	59							
9		9	29	f	roast lamb	166	1.66	62	22.5	nurse	M	A	roast lamb	62							
10		10	23	f	prawns	176	1.76	72	23.2	medical Dr	S	R	prawns	72							
11		11	27	m	peanuts	166	1.66	75	27.2	shopkeeper	D	L	peanuts	75							
12		12	43	f	pizza	162	1.62	66	25.1	painter	M	L	pizza	66							
13		13	62	f	fish	182	1.82	59	17.8	accountant	M	L	fish	59							
14		14	49	m	bread	190	1.9	82	22.7	babysitter	P	R	bread	82							
15		15	28	f	spaghetti	172	1.72	79	26.7	houseperson	S	R	spaghetti	79							
16		16	36	f	apples	173	1.73	69	23.1	nurse	P	R	apples	69							
17		17	18	m	pizza	163	1.63	76	28.6	teacher	M	L	pizza	76							
18		18	62	f	bread	166	1.66	74	26.9	secretary	D	L	bread	74							
19		19	32	f	apples	154	1.54	79	33.3	retired	W	R	apples	79							
20		20	29	f	roast lamb	166	1.66	62	22.5	nurse	M	A	roast lamb	62							
21		21	23	f	prawns	176	1.76	72	23.2	medical Dr	S	R	prawns	72							
22		22	27	m	peanuts	166	1.66	75	27.2	shopkeeper	D	L	peanuts	75							
23		23	43	f	pizza	162	1.62	66	25.1	painter	M	L	pizza	66							
24		24	62	f	fish	182	1.82	59	17.8	accountant	M	L	fish	59							
25		25	49	m	bread	190	1.9	82	22.7	babysitter	P	R	bread	82							
30	mean		36.1				168.9	1.7	69.0	24.5											
31	SD		13.6				10.3	0.1	9.8	5.0											

Exercise 4

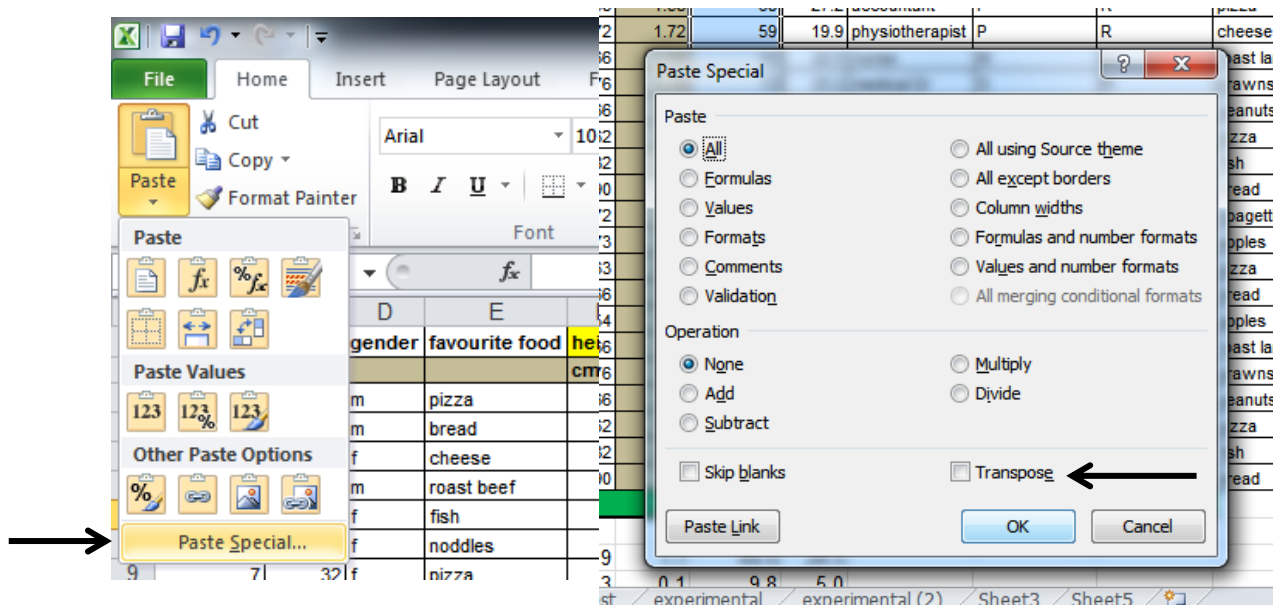
Copy three columns of data from the 'Descriptive' sheet in your new MS Excel file (my File.xlsx) and paste them in other places on the sheet. Practice deleting the original column of data, and then copying the information back. Don't worry if you cannot retrieve it because you have a backup of the original file (Excel Course Data entry.xlsx)!!!

Paste Special

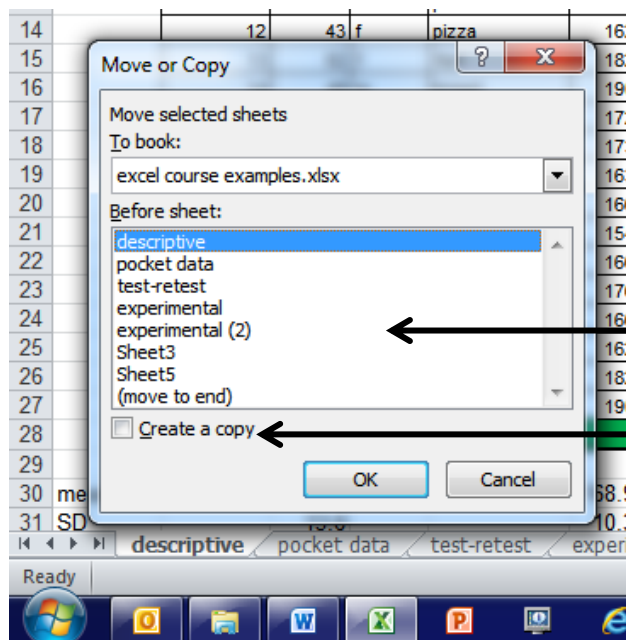
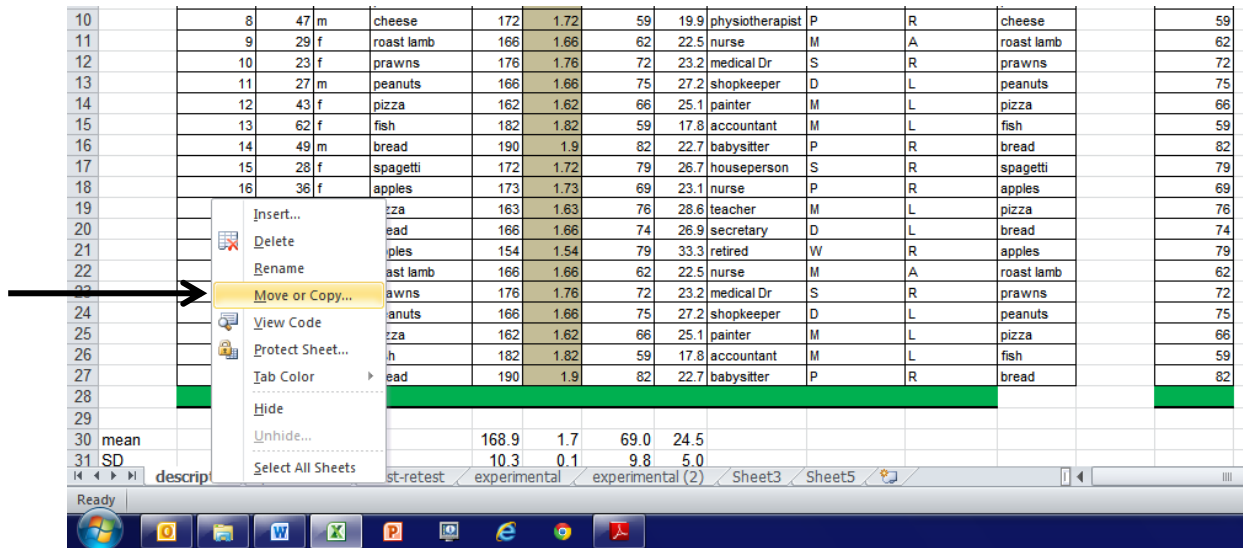
PASTE SPECIAL (found in the drop down menu under the paste button on the home ribbon) is the function that you use if you want to paste information into a cell in other than number form. It is particularly useful if you are transferring formulae around your excel sheet, or if you want to transpose numbers (from vertical to horizontal).

Exercise 5

Highlight data in one of the columns in the sheet 'Descriptive' and copy it. Go to PASTE SPECIAL, select 'Transpose'. Click your cursor into a cell on a new sheet in the same workbook, and click on PASTE. Notice what happens to the data.

**How to copy a sheet**

If you are using data for statistical calculations, it is really important that you keep your raw data intact (that is, don't do any calculations on it!!). **This is so that you always have a correct set of data in case you make a mistake!!** Make sure that you have a spare sheet in your workbook (you know it is spare because it hasn't been renamed, and is blank). Right click on the name tab of the sheet you wish to copy and select Move or Copy from the pop up menu.



Select the page you want your copy to appear

To create a **copy** you **must** select this box!

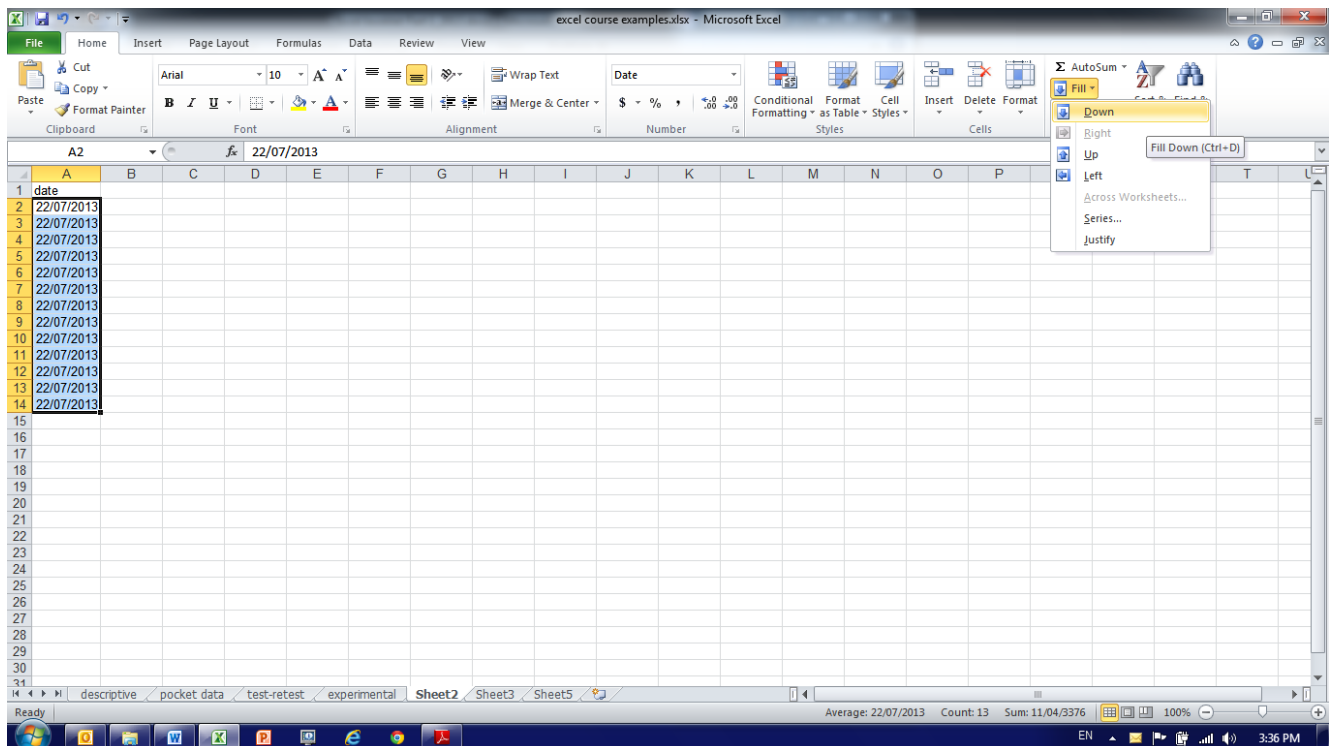
Select 'create a copy' if you click on this, it will copy all the information in '**descriptive**' (or whatever sheet you are currently viewing) to your new location. A new sheet will appear in the workbook called '**descriptive (2)**'. Now rename the sheet that you copied (in this instance **descriptive**) as 'Raw Data' and the newly copied sheet as 'worksheet' or something else that will remind you of what you are doing in this sheet. Now it doesn't matter if you make a mistake in the copied sheet, because you can always retrieve information from your raw data sheet. **Remember once you have a Raw Data sheet, never do any calculations on it, in case you need to retrieve anything!!**

Exercise 6

Practice copying, moving and deleting sheets in you're my File.xlsx workbook, and renaming them, so that you really understand these functions.

The Fill Function

The Fill function allows you to enter one value and then 'Fill' down by holding the left cursor down and highlighting the cells that you want to fill. You can put the same value in all these highlighted cells (for instance if you want to enter data that was all collected on the same day you would enter the date and then **FILL down** and the same date would appear in as many cells as you had highlighted) or you can use the **Series** function in FILL to change the values (usually in increments). So you could enter value 1, go to FILL, select **series**, indicate that you want to stop at 10, highlight the next 9 cells, and MS Excel would automatically fill those cells 2-10.



Drag cells or formulae

One of the most useful features of Excel is the **DRAG** feature. It is particularly useful for filling cells with values or to calculate formulae without having to retype information. The FILL function is a type of **DRAG** feature. To be really useful the **DRAG** feature can be used to copy formulae from one set of data to another with minimum typing. To learn to **DRAG** you need to learn a little bit about turning Excel from a spread sheet to a calculator.

Set's set up the following example. In your MS Excel My File.xlsx, the page tabs at the bottom and click on the new page tab. A new numbered worksheet will be inserted.

27		25	49	m	bread	190	1.9	82	22.7	babysitter	P	R
28												
29												
30	mean		36.1			168.9	1.7	69.0	24.5			
31	SD		13.6			10.3	0.1	9.8	5.0			

Enter the data as outlined on the page below. Imagine that Column 1 is the code for 20 different items, and Column B is the price per item. The manager has decided that everything will be increased by \$0.10. You can recalculate the price of the items by typing = into cell C2, clicking into cell B2 (or typing B2 in Cell C2) and then typing + 0.10.

C2				fx =B2+0.1			
	A	B	C	D			
1	item number	price	price +0.10				
2	1	12.65	12.75				
3	2	13.95					
4	3	18.2					
5	4	6.5					
6	5	7.44					
7	6	15.86					
8	7	23.9					
9	8	13.98					
10	9	17.96					
11	10	2.54					
12	11	11					
13	12	15.2					
14	13	46.52					
15	14	18.65					
16	15	13.12					
17	16	0.55					
18	17	14.5					
19	18	12.3					
20	19	5.21					
21	20	19.85					
22							
23							

You hit the ENTER key, and the new price for the first item will appear. Take your cursor to the bottom right corner of this cell (C2) and watch it change from an open white cross to a hard black cross. Holding the left mouse button down and moving this hard black cross down the column, will fill the cells below with the new formula. Watch the new prices appear, using the formula that you put into cell C2 (which references to each subsequent B column reference).

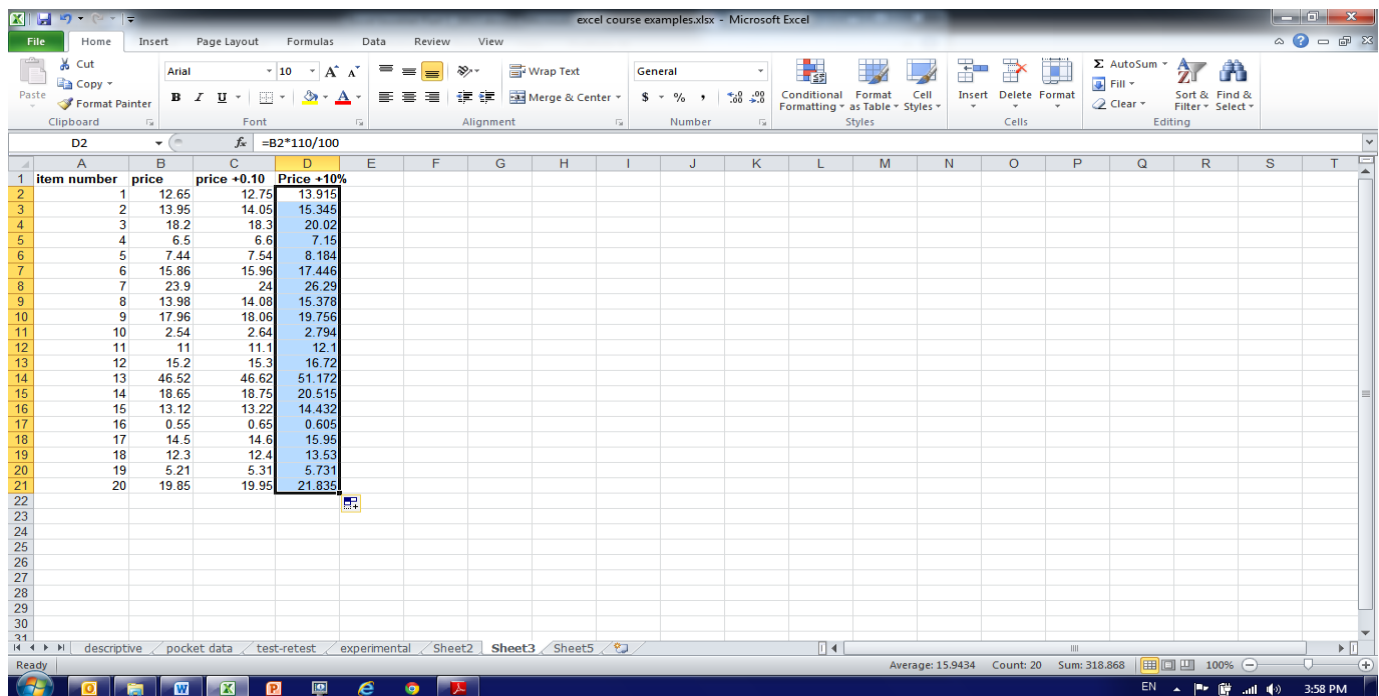
You can do exactly the same feature using the PASTE SPECIAL function. Instead of dragging the formula, you could copy what is in Cell C2 (a formula), and PASTE SPECIAL it into Cell D2 (or somewhere else) by clicking into the PASTE FORMULAS function. Dragging is a lot

quicker! The advantage of the PASTE SPECIAL function is that if you want to copy the actual value that you have created from the formula (say, for further analysis), you can use the PASTE VALUE feature, and it gets rid of the formula and leaves you only with the value.

Words of warning:

When you are creating formulae, don't enter the actual value that is in cell B2 (12.55) after the = sign in C2, always enter the cell reference (ie B2). If you enter the value for cell B2 and add 10c, and then drag down, all the rest of the cells in column C will have the same value (12.65)

Now, let's try increasing the price of each item by 10%. We will need to use a different formula in cell D2. Type =, click into cell B2, then type *110/100 (which will increase the value in B2 by 110 and divide it by 100 (a quick way of adding 10%)). If you hit the ENTER key, and then drag this formula down you can see there is a different set of values appearing.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	item number	price	price +0.10	Price +10%																
2	1	12.65	12.75	13.915																
3	2	13.95	14.05	15.345																
4	3	18.2	18.3	20.02																
5	4	6.5	6.6	7.15																
6	5	7.44	7.54	8.184																
7	6	15.86	15.96	17.446																
8	7	23.9	24	26.29																
9	8	13.98	14.08	15.378																
10	9	17.96	18.06	19.756																
11	10	2.54	2.64	2.794																
12	11	11	11.1	12.1																
13	12	15.2	15.3	16.72																
14	13	46.52	46.62	51.172																
15	14	18.65	18.75	20.515																
16	15	13.12	13.22	14.432																
17	16	0.55	0.65	0.605																
18	17	14.5	14.6	15.95																
19	18	12.3	12.4	13.53																
20	19	5.21	5.31	5.731																
21	20	19.85	19.95	21.835																

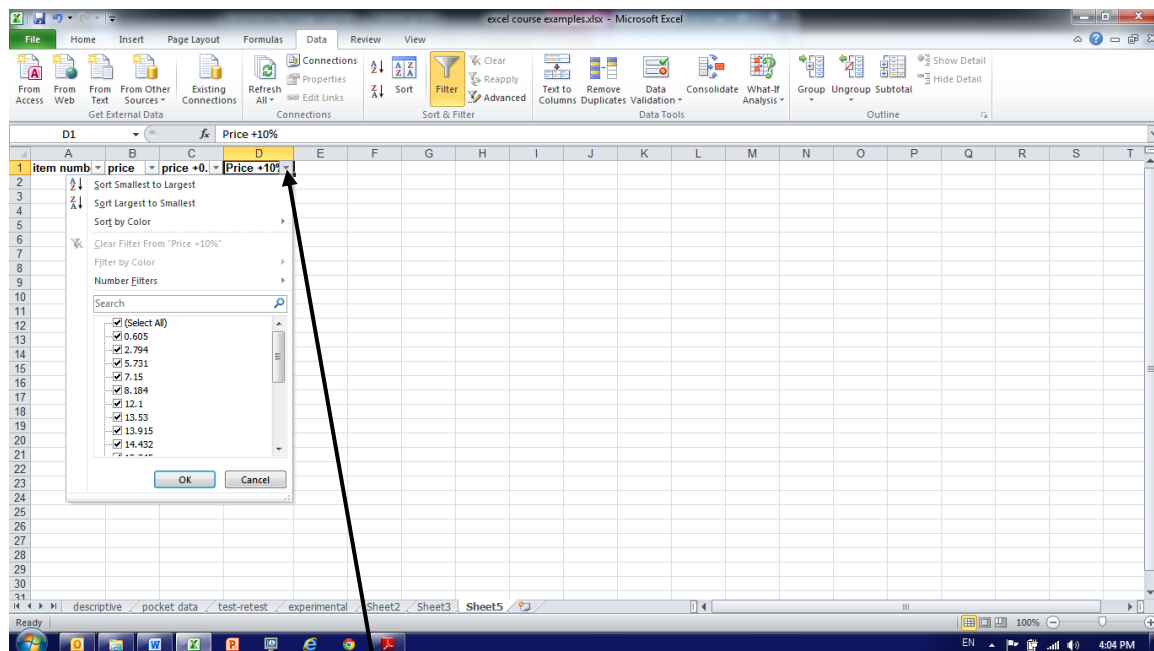
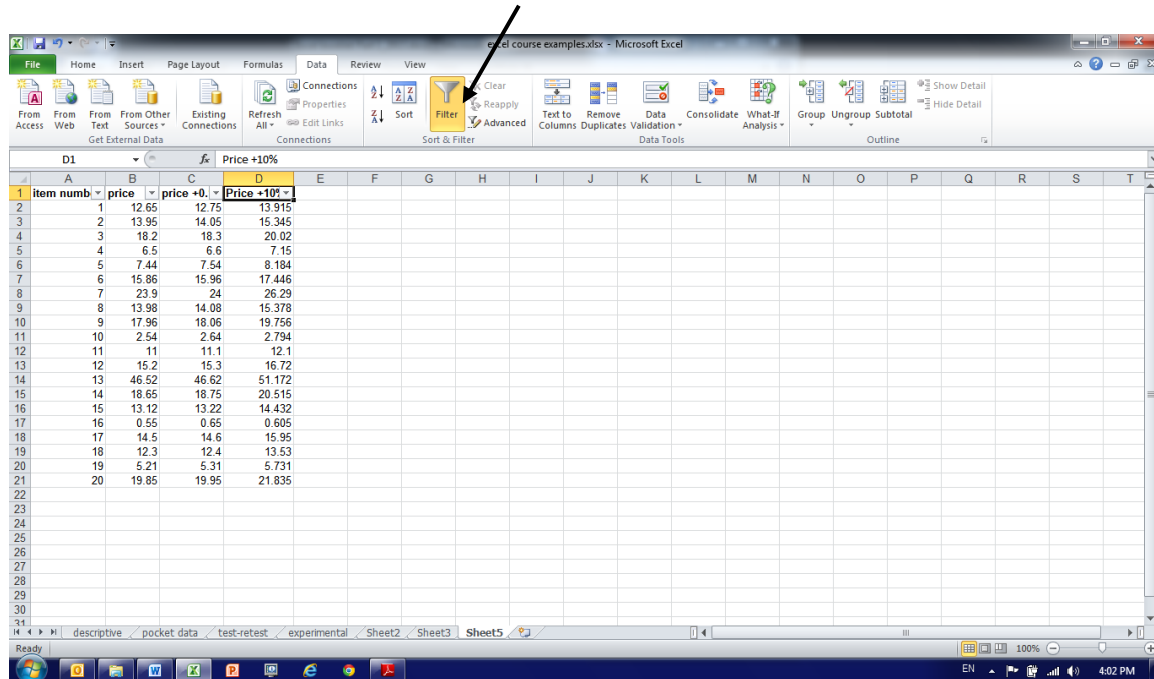
Remember the DRAG function when we come to applying statistical tests!!

Exercise 7.

Open a new MS Excel file and using the FILL and DRAG functions, enter 10 rows of data, assigning an ID number of 1, today's date and a score of 10 to the first row, and then automatically complete the remaining 9 rows of data (without any further manual data entry). Increase the ID number by one for each row, retain the date, and increase the score by 2 for each subsequent row. Play around with this data, once you have completed this task. Save the file onto your USB using another file name.

Data filter

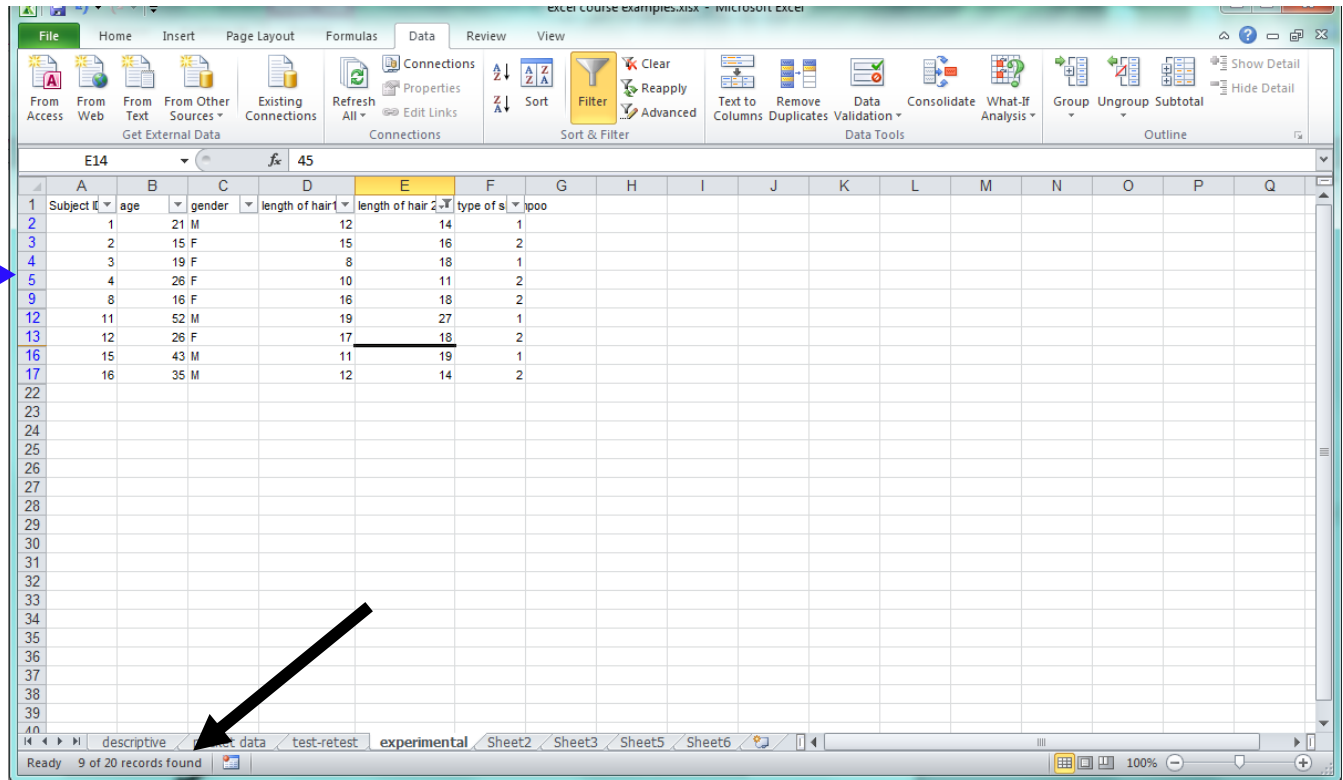
The data filter is another Excel wonder! Click your cursor into any one of the heading cells (the names that you give to your columns of data). Point your cursor at DATA in the Menu, and click on FILTER. This brings up a little button on the side of each of the header cells.



If you click onto any of the buttons, you can see all of the values that have been entered into this column (not how many of them at the moment, simply all the possible choices). So now you can start to see the value of this function. If for instance when entering data, if you are putting

information on gender into this column, you would only realistically expect to find M or F. Any other value is probably an error, so you can highlight the error, which will bring up the row (or rows) which have this value in it, and you can change it to the correct value (first checking your raw data of course!)

So what do you do now?



If you click on the value that you want to check, this filters out all of the occurrences of the value that you have selected in the column. You will notice that when Excel is in Filter mode, all of the filtered values appear in blue in the row header. **This is a very good reminder that you are in filter mode.**

You will notice that down in the bottom left corner of the program page you are told how many of the values you selected are to be found in the filtered column of data – a very useful feature when you have lots of data!!!

A word of warning:

When you are in filter mode, never attempt to conduct any calculations on your filtered data. Always copy the filtered data into another sheet for analysis purposes. If you forget to do this, you will find that while MS Excel tells you it is in FILTER MODE, it actually will do all calculation on all of the underlying data. **So BEWARE!!!!**

How to sort data

You can sort data by selecting the whole page (use the little arrow in the top left of the row header), then go to the Data ribbon and going to the SORT function, you can put some order into your data if it is jumbled up (like our Excel course data set Descriptive Sheet!).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
		ID number	age	gender	favourite food	height cms	height mtrs	weight	BMI	employment	marital status	handedness	favourite food								
1		1	20	m	pizza	154	1.54	90	37.9	waiter	M	R	pizza								
2		2	31	m	bread	173	1.73	49	16.4	nurse	S	L	bread								
3		3	36	f	cheese	172	1.72	54	18.3	nurse	D	R	cheese								
4		4	26	m	roast beef	163	1.63	69	26.0	secretary	S	L	roast beef								
5		5	49	f	fish	169	1.69	59	20.7	teacher	M	R	fish								
6		6	19	f	noddles	149	1.49	69	31.1	houseperson	S	R	noddles								
7		7	32	f	pizza	158	1.58	68	27.2	accountant	P	R	pizza								
8		8	47	m	cheese	172	1.72	59	19.9	physiotherapist	P	R	cheese								
9		9	29	f	roast lamb	166	1.66	62	22.5	nurse	M	A	roast lamb								
10		10	23	f	prawns	176	1.76	72	23.2	medical Dr	S	R	prawns								
11		11	27	m	peanuts	166	1.66	75	27.2	shopkeeper	D	L	peanuts								
12		12	43	f	pizza	162	1.62	66	25.1	painter	M	L	pizza								
13		13	62	f	fish	182	1.82	59	17.8	accountant	M	L	fish								
14		14	49	m	bread	190	1.9	82	22.7	babysitter	P	R	bread								
15		15	28	f	spagetti	172	1.72	79	26.7	houseperson	S	R	spagetti								
16		16	36	f	apples	173	1.73	69	23.1	nurse	P	R	apples								
17		17	18	m	pizza	163	1.63	76	28.6	teacher	M	L	pizza								
18		18	62	f	bread	166	1.66	74	26.9	secretary	D	L	bread								
19		19	32	f	apples	154	1.54	79	33.3	retired	W	R	apples								
20		20	29	f	roast lamb	166	1.66	62	22.5	nurse	M	A	roast lamb								
21		21	23	f	prawns	176	1.76	72	23.2	medical Dr	S	R	prawns								
22		22	27	m	peanuts	166	1.66	75	27.2	shopkeeper	D	L	peanuts								
23		23	43	f	pizza	162	1.62	66	25.1	painter	M	L	pizza								
24		24	62	f	fish	182	1.82	59	17.8	accountant	M	L	fish								
25		25	49	m	bread	190	1.9	82	22.7	babysitter	P	R	bread								
26																					
27																					
28																					
29																					
30	mean		36.1			168.9	1.7	69.0	24.5												
31	SD		13.6			10.3	0.1	9.8	5.0												

Word of warning:

You MUST select the whole page when you are sorting data that has more than one column per row. If you only select the column you want to sort it will change your dataset and unlink the data that should be connected (ie: ID number 25 may end up with ID number 11's data for the column you have sorted, you can see how this could be a problem!)

ID number	age	gender	favourite food	height cms	height mtrs	weight	BMI	employment	marital status	handedness	favourite food
1	20	m	pizza	154	1.54	90	37.9	waiter	M	R	pizza
2	31	m	bread	173	1.73	49	16.4	nurse	S	L	bread
3	36	f	cheese	172	1.72	54	18.3	nurse	D	R	cheese
4	26	m	roast beef	163	1.63	69	26.0	secretary	S	L	roast beef
5	49	f	fish	169	1.69	59	20.7	teacher	M	R	fish
6	19	f	noddles	149	1.49	69	31.1	houseperson	S	R	noddles
7	32	f	pizza	158	1.58	68	27.2	accountant	P	R	pizza
8	47	m	cheese	172	1.72	59	19.9	physiotherapist	P	R	cheese
9	29	f	roast lamb	166	1.66	62	22.5	nurse	M	A	roast lamb
10	23	f	prawns	176	1.76	72	23.2	medical Dr	S	R	prawns

If you choose ascending order for the column of data for 'favorite foods', and sort, this is what you will get!! You can see that the whole data set has been rearranged and sorted alphabetically by 'favorite foods'. To be certain that your data is the same (just rearranged) check that you have the same other values linked with the favorite foods by checking the earlier dataset. Practice changing the order of (sorting) some of the other variables in this dataset.

Remember how you applied the data filter to find out how many times one type of data entry occurred in a column of data? The SORT function is similar, as you can see how many there are of different types of data in a column. If there are lots of data then you can still use the FILTER function to automatically count how many times a particular value has been entered.

Practice playing around with the SORT and FILTER functions in this file!! Don't worry if you make mistakes because we won't be saving the file!

Clearing data (or deleting it)

We did a little of this previously. Highlight the column of data that you want to clear (or delete). Go to home ribbon in the Menu Bar; select the CLEAR button and then Clear All from the drop down menu. All the values will be cleared in that column but the column space will remain intact. If you choose the DELETE button on the Home ribbon, not only will the values disappear

but the column space will also disappear. Remember that if you delete a column of data in error, there is always the UNDO button (the little back arrow) on the top of the page. If you aren't sure that you really won't need this data again, copy your data onto a new sheet and then delete the column of data in the copy – that way you always have a backup!! **(Remember to always keep the raw data sheet “clean” (unused) so that you could go back to this if you needed it!)**

How to link two columns of data

How to apply formulas

You have already attempted some of this task when you learnt about the DRAG feature. Let's look at linking two columns of data first. First consider your data set with the item codes and prices. Add another column that includes the number of times each item has been purchased. To calculate the total cost of this item, you would put an = sign into the first cell of the column into which you want the new information to go, and then put in a formula that multiplies the dollar cost by the number of times the item had been purchased. Remember to use the cell references, not the actual value!

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	item number	price	Number of purchases	total cost												
2	1	12.65	5	63.25												
3	2	13.95	6													
4	3	18.2	2													
5	4	6.5	1													
6	5	7.44	45													
7	6	15.86	12													
8	7	23.9	13													
9	8	13.98	2													
10	9	17.96	5													
11	10	2.54	1													
12	11	11	47													
13	12	15.2	8													
14	13	46.52	9													
15	14	18.65	1													
16	15	13.12	12													
17	16	0.55	13													
18	17	14.5	10													
19	18	12.3	5													
20	19	5.21	25													
21	20	19.85	4													
22																
23																
24																
25																
26																
27																
28																
29																
30																

As you have done previously, hit the ENTER key and the value for the first item appears. Then drag this formula down for all of the other items and the total cost of each item appears.

The columns of data that you want to link by formula don't have to be next to one another – they can be anywhere on the data sheet – they can even be on other sheets of data in the one workbook (file). As long as your formula and the cell references are correct you can use the DRAG function and the formula entry function to link any number of columns of data.

Now, another wonderful Excel feature! If the price of any item changes, instead of recalculating everything, simply change the value in the appropriate cell, and watch Excel recalculate everything automatically!

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	item number	price	Number of purchase	total cost												
2	1	12.65	5	63.25												
3	2	13.95	6	83.7												
4	3	18.2	2	36.4												
5	4	6.5	1	6.5												
6	5	7.44	45	334.8												
7	6	15.86	12	190.32												
8	7	23.9	13	310.7												
9	8	13.98	2	27.96												
10	9	17.96	5	89.8												
11	10	2.54	1	2.54												
12	11	11	47	517												
13	12	15.2	8	121.6												
14	13	46.52	9	418.68												
15	14	18.65	1	18.65												
16	15	13.12	12	157.44												
17	16	0.55	13	7.15												
18	17	14.5	10	145												
19	18	12.3	5	61.5												
20	19	5.21	25	130.25												
21	20	19.85	4	79.4												
22																
23																
24																
25																
26																
27																
28																
29																
30																

Exercise 8.

Using the price file, practice applying formulae and dragging. You can choose what formulae you want to apply, and to what columns of data, but be ready to explain what you are doing to the tutor or the class!!

Descriptive stats and statistical tests

Calculating AVERAGES (Means)

On the sheet called TEST-RETEST, imagine that it has been generated on a timed up and go test which has been measured by 20 members of a student group in an exercise on reliability of measurement technique. Each of the students has recorded three measurements of the test, in an attempt to demonstrate that they are reliable. This is the way that the data might look. To start with you want to calculate the AVERAGE value for the group for Test 1, 2 and 3. Leave a row under your raw data, and insert a column to the left of your first column of data, so that it gives you a new column into which to add some references for yourself. **Now keep your eye on these boxes!!!!**

	A	B	C	D	E	F	G	H	I	J	K	L
		ID	test 1	test2	test3	test 1	test2	test3				
1												
2												
3			1	12	13	13	a	b				
4			2	15	14	16	b	b				
5			3	17	17	18	a	b				
6			4	15	16	16	c	a	b			
7			5	20	21	20	a	b	c			
8			6	21	22	22	b	b	a			
9			7	25	25	26	c	c	b			
10			8	26	27	26	c	b	c			
11			9	19	20	20	a	b	b			
12			10	20	21	22	a	a	a			
13			11	25	27	26	b	b	a			
14			12	27	28	28	b	b	b			
15			13	28	26	27	a	a	a			
16			14	21	19	21	a	a	b			
17			15	19	17	18	c	c	c			
18			16	18	17	18	c	c	b			
19			17	20	19	20	b	b	a			
20			18	21	20	22	a	b	b			
21			19	30	29	27	b	a	b			
22			20	42	41	39	c	b	a			
23												
24												
25												
26												

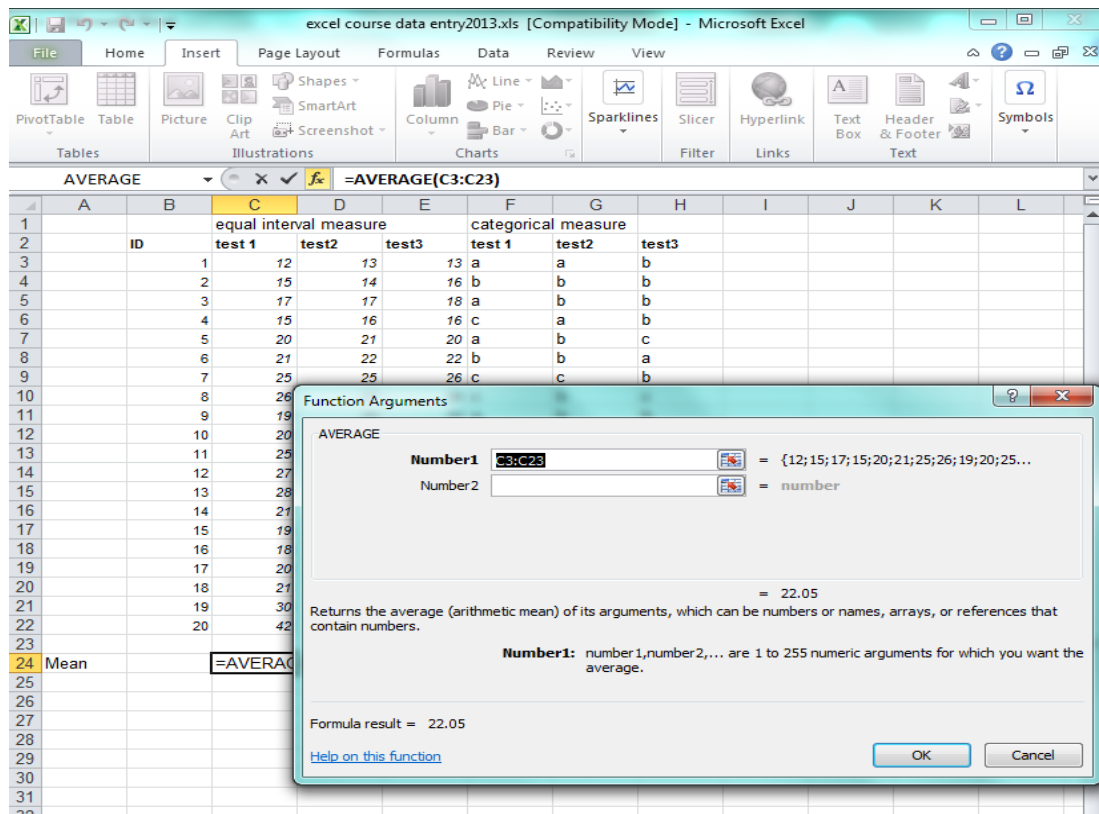
In your new column, type mean (or AVERAGE) for your own reference (leaving a blank row first!) Under the Test 1 data, type =. You will notice that the left hand box that you have your eye on changes, and lists for you (if you click on the button on the right) all the recently used statistical tests.

You select the statistical function that you want. Each of the function names are shortened, so you should expect that you will know what you are looking for! But if in doubt, you can click on a function and it will tell you what it does. The Right hand box (the Comments Box), is where

your formula appears. If you aren't sure what formula (or value) is in a cell, click into the cell, and look at what appears in the comments box.

The screenshot shows a Microsoft Excel window titled "excel course data entry2013.xls [Compatibility Mode] - Microsoft Excel". The ribbon includes File, Home, Insert, Page Layout, Formulas, Data, Review, and View. The Formulas tab is active, showing a dropdown menu for the AVERAGE function. The spreadsheet data is as follows:

	C	D	E	F	G	H	I	J	K	L
	equal interval measure			categorical measure						
	test 1	test2	test3	test 1	test2	test3				
12	10	20	21	22	a	a	a			
13	11	25	27	26	b	b	a			
14	12	27	28	28	b	b	b			
15	13	28	26	27	a	a	a			
16	14	21	19	21	a	a	b			
17	15	19	17	18	c	c	c			
18	16	18	17	18	c	c	b			
19	17	20	19	20	b	b	a			
20	18	21	20	22	a	b	b			
21	19	30	29	27	b	a	b			
22	20	42	41	39	c	b	a			
23										
24	Mean	=								
25										
26										
27										



To calculate the average (or the mean) for TEST 1, now select (or type) AVERAGE after the = sign under the data for Test 1 (leaving your row). This is so that you don't mix up your statistics with your raw data. Then the program asks you to define the range of variables from which you want to calculate the average (mean). You can enter these by clicking into the first cell of data, holding your Left mouse button down, and dragging down to the last cell. You can also type the cell references directly into the reference cell, but this can be slow and prone to error. Hit the RETURN key, and there is your Average value for that column of data!!

Now, using that DRAG option (move the cursor down to the bottom right corner of the cell into which you typed the AVERAGE formula and watch it change in nature), move the cursor across to the next two columns to calculate the group means for Test 2 and Test 3. This DRAG option will only work if you have the same number of cells with data in them as the column in which you first calculated the formula. Don't worry, if you have missing data in a cell within the cell references in the formula, MS Excel treats it as missing data (no value) and calculates over the top of it!!

Now you can apply exactly the same approach to calculate the Standard Deviation (STDEV).

excel course data entry2013.xls [Compatibility Mode] - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View

PivotTable Table Picture Clip Art SmartArt Screenshot Charts Sparklines Slicer Hyperlink

H34 fx

	A	B	C	D	E	F	G	H	I
1			equal interval measure			categorical measure			
2		ID	test 1	test2	test3	test 1	test2	test3	
3		1	12	13	13	a	a	b	
4		2	15	14	16	b	b	b	
5		3	17	17	18	a	b	b	
6		4	15	16	16	c	a	b	
7		5	20	21	20	a	b	c	
8		6	21	22	22	b	b	a	
9		7	25	25	26	c	c	b	
10		8	26	27	26	c	b	c	
11		9	19	20	20	a	b	b	
12		10	20	21	22	a	a	a	
13		11	25	27	26	b	b	a	
14		12	27	28	28	b	b	b	
15		13	28	26	27	a	a	a	
16		14	21	19	21	a	a	b	
17		15	19	17	18	c	c	c	
18		16	18	17	18	c	c	b	
19		17	20	19	20	b	b	a	
20		18	21	20	22	a	b	b	
21		19	30	29	27	b	a	b	
22		20	42	41	39	c	b	a	
23									
24	Mean		22.05	21.95	22.25				
25	Stdev		6.613185	6.508901	5.784507				
26									

Other descriptive data functions in MS Excel that you can obtain in exactly the same way are MEDIAN, MODE, PERCENTILE, MIN(imum), MAX(imum). AVERAGES (or means), Standard Deviations, Medians, Percentiles can only be calculated for equal interval data (number data). You cannot calculate these statistical tests on categorical data (such as a, b, c etc).

Exercise 9

Check that you have calculated the same value for the Average and Standard Deviation as the person sitting next to you. Now calculate all the descriptive data techniques listed above for the equal interval data Test-retest' sheet.

Exercise 10.

Go to the Sheet 'Experimental' in Excel Course Data Entry.xlsx and calculate all the descriptive data techniques listed above for the equal interval data in age, length of hair 1, and length of hair 2.

Exercise for the people who want more challenge

As a way of reflecting on what you have learnt so far, open a new MS Excel file, and identify in the first row, the header names of the likely expenses that you will have whilst living in Adelaide. Label column 1 'DATE', and then list at least another 5 columns of your likely types of expenses (Transport, food, dinning etc)

Now make up some costs for each of the next seven days (one day per row). Play around with the numbers, working out (using your formulae skills), how much living in Adelaide could cost you each day, and across a week. Calculate average costs across a week for your expenses, and see what happens when you increase or decrease your expenses by a percentage.

If you are working to a budget, type your budget for a week into another column, and practice taking away from this amount the money you are spending each day. Don't worry if you blow your budget – this is pretend!! But this is a very good way of keeping track of what you are spending, and how you could change your spending patterns.

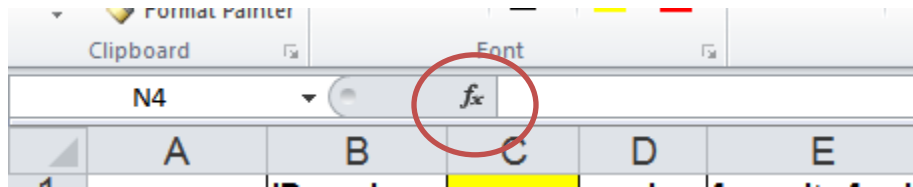
Statistical calculations

Remember the TEST-RETEST sheet in your Excel file? We now need to test whether there are significant differences between the three sets of measurements to see whether the students were reliable

We do this in 3 ways

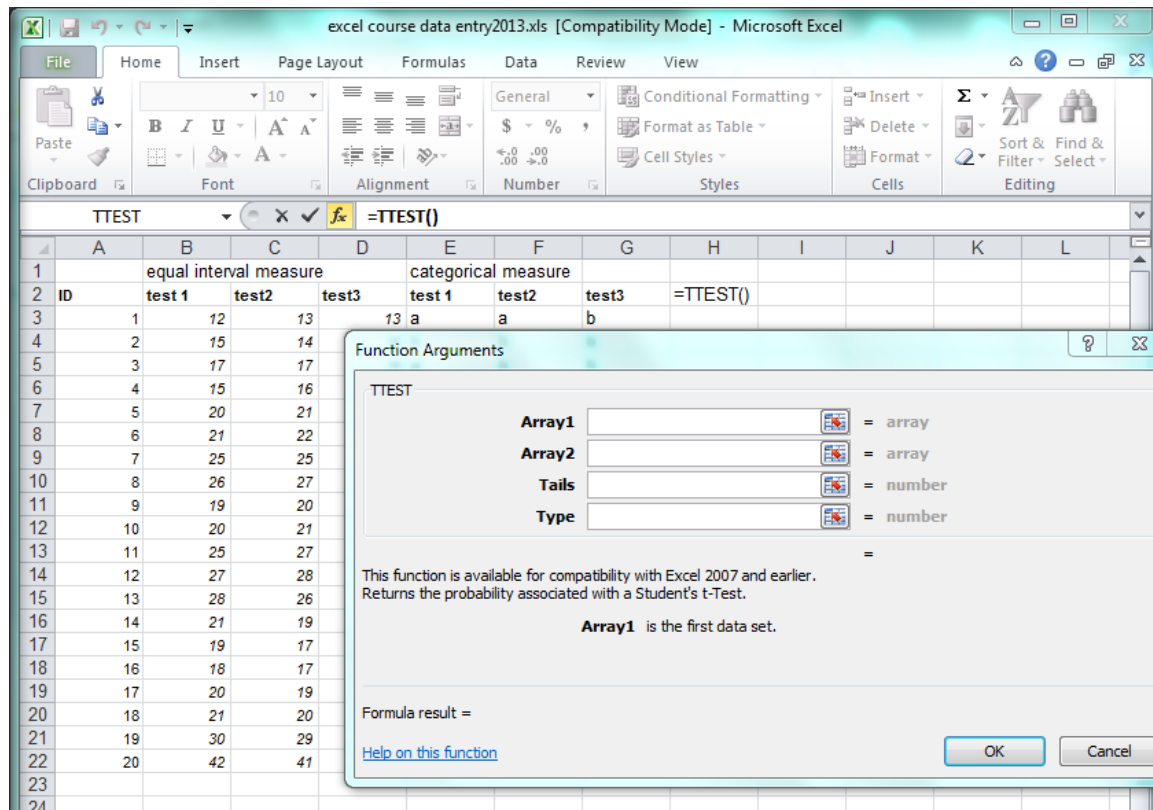
1. We conduct a paired t-test on the data.

Choose a cell into which you want your t-test result to go (usually under the descriptive statistics), type **=T.test** into the Comments Box, and a small box will appear under it outlining the format of the equation you need to enter. Alternatively you can click on the function button and select TTEST from the menu and a pop up box will appear.



For T-tests (in our example there is an initial and then 2 repeat measures on same group of subjects), you pair sets of these measures (set 1 compared with 2, set 2 compared with 3, set 1 compared with 3), and test significant differences (seeking to accept the null hypothesis and demonstrate that the measures are no different from each other). First is **array 1** put the cell references for the raw data for test 1 in first, second is **array 2**, put the cell references for raw data for test 2 (making sure that you don't include the AVERAGE value) in here (make sure you separate each data set with a comma, as seen in the little box), the 3rd is the number of **tails** (1 or 2) (put in 2 for 2 tailed t-test, always do this!!!) and the 4th is type, type 1 is for paired t-tests (important to remember when you are doing reliability tests). The answer is the *p value* associated with the T-test. If it is less than 0.05 then the two sets of measurements are significantly different.

Sometimes the p value is so small that Excel expresses it as an exponential value (E5.0006) for instance, which indicates that there are 5 0's after the decimal point (a sure indication that the p value is very small indeed and therefore definitely significant!!!! In that the p value is definitely less than 0.05)



2. We can also test for significance using Analysis of Variance (ANOVA).

To test significance on this same set of data using ANOVA, use single factor ANOVA (the initial data and the 2 repeats all considered together), and test significant differences (seeking to accept the null hypothesis and demonstrate that the measures are no different from each other).

You will need to go to the Data ribbon, and select Data Analysis. Sometimes the computers that you are using will need to be primed first, as you may not find Data Analysis in the list. If this is the case, go to the FILE TAB, choose OPTIONS, ADD INS, and click GO next to the manage add ins drop down box. Tick both box for Analysis Toolpak. Wait until the Data Analysis tools have been retrieved and activated, and then you will find Data Analysis on that computer every time you go back to it.

For a simple replication test on this type of data you need to test the Single factor ANOVA model. Click on this option and follow the prompts. If you want to calculate an Intraclass Correlation to test for reliability, you will need to calculate an ANOVA first.

excel course data entry2013.xls [Compatibility Mode] - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View

Get External Data Refresh All Edit Links Connections Sort & Filter Filter Clear Reapply Advanced Text to Columns Remove Duplicates Data Tools Group Ungroup Subtotal Outline Analysis Data Analysis

H2

	A	B	C	D	E	F	G	H	I	J	K	L
1		equal interval measure			categorical measure							
2	ID	test 1	test2	test3	test 1	test2	test3					
3	1	12	13	13	a	a	b					
4	2	15	14	16	b	b	b					
5	3	17	17	18	a	b	b					
6	4	15	16	16	c	a	b					
7	5	20	21	20	a	b	c					
8	6	21	22	22	b							
9	7	25	25	26	c							
10	8	26	27	26	c							
11	9	19	20	20	a							
12	10	20	21	22	a							
13	11	25	27	26	b							
14	12	27	28	28	b							
15	13	28	26	27	a							
16	14	21	19	21	a							
17	15	19	17	18	c							
18	16	18	17	18	c							
19	17	20	19	20	b							
20	18	21	20	22	a	b	b					
21	19	30	29	27	b	a	b					
22	20	42	41	39	c	b	a					

Data Analysis

Analysis Tools

- Anova: Single Factor
- Anova: Two-Factor With Replication
- Anova: Two-Factor Without Replication
- Correlation
- Covariance
- Descriptive Statistics
- Exponential Smoothing
- F-Test Two-Sample for Variances
- Fourier Analysis
- Histogram

OK Cancel Help

excel course data entry2013.xls [Compatibility Mode] - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View

Get External Data Refresh All Edit Links Connections Sort & Filter Filter Clear Reapply Advanced Text to Columns Remove Duplicates Data Tools Group Ungroup Subtotal Outline Analysis Data Analysis

B3

	A	B	C	D	E	F	G	H	I	J	K	L
1		equal interval measure			categorical measure							
2	ID	test 1	test2	test3	test 1	test2	test3					
3	1	12	13	13	a	a	b					
4	2	15	14	16	b	b	b					
5	3	17	17	18	a	b	b					
6	4	15	16	16	c							
7	5	20	21	20	a							
8	6	21	22	22	b							
9	7	25	25	26	c							
10	8	26	27	26	c							
11	9	19	20	20	a							
12	10	20	21	22	a							
13	11	25	27	26	b							
14	12	27	28	28	b							
15	13	28	26	27	a							
16	14	21	19	21	a							
17	15	19	17	18	c							
18	16	18	17	18	c							
19	17	20	19	20	b							
20	18	21	20	22	a	b	b					
21	19	30	29	27	b	a	b					
22	20	42	41	39	c	b	a					

Anova: Single Factor

Input

Input Range:

Grouped By: ☒ Columns ☐ Rows

☐ Labels in first row

Alpha:

Output options

☒ Output Range:

☐ New Worksheet Ply:

☐ New Workbook

OK Cancel Help

excel course data entry2013.xls [Compatibility Mode]

File Home Insert Page Layout Formulas Data Review View

From Access From Web From Text From Other Sources Existing Connections Refresh All Connections Sort & Filter Filter Clear Reapply Advanced Text to Columns Remove Duplicates Data Validation Consolidate What-If Analysis

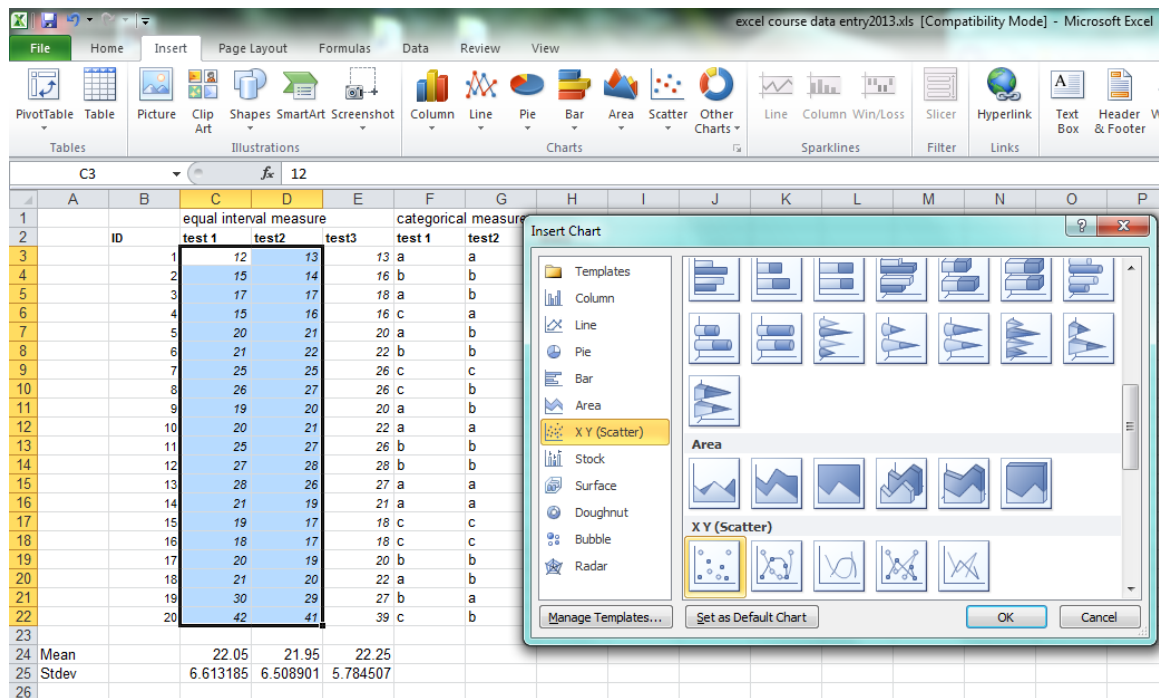
H2 Anova: Single Factor

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		equal interval measure			categorical measure									
2	ID	test 1	test2	test3	test 1	test2	test3	Anova: Single Factor						
3	1	12	13	13	a	a	b	SUMMARY						
4	2	15	14	16	b	b	b	Groups Count Sum Average Variance						
5	3	17	17	18	a	b	b	Column 1 20 441 22.05 43.73421						
6	4	15	16	16	c	a	b	Column 2 20 439 21.95 42.36579						
7	5	20	21	20	a	b	c	Column 3 20 445 22.25 33.46053						
8	6	21	22	22	b	b	a	ANOVA						
9	7	25	25	26	c	c	b	Source of Variance SS df MS F P-value F crit						
10	8	26	27	26	c	b	c	Between Groups 0.933333 2 0.466667 0.01171 0.988361 3.158843						
11	9	19	20	20	a	b	b	Within Groups 2271.65 57 39.85351						
12	10	20	21	22	a	a	a	Total 2272.583 59						
13	11	25	27	26	b	b	a							
14	12	27	28	28	b	b	b							
15	13	28	26	27	a	a	a							
16	14	21	19	21	a	a	b							
17	15	19	17	18	c	c	c							
18	16	18	17	18	c	c	b							
19	17	20	19	20	b	b	a							
20	18	21	20	22	a	b	b							
21	19	30	29	27	b	a	b							
22	20	42	41	39	c	b	a							

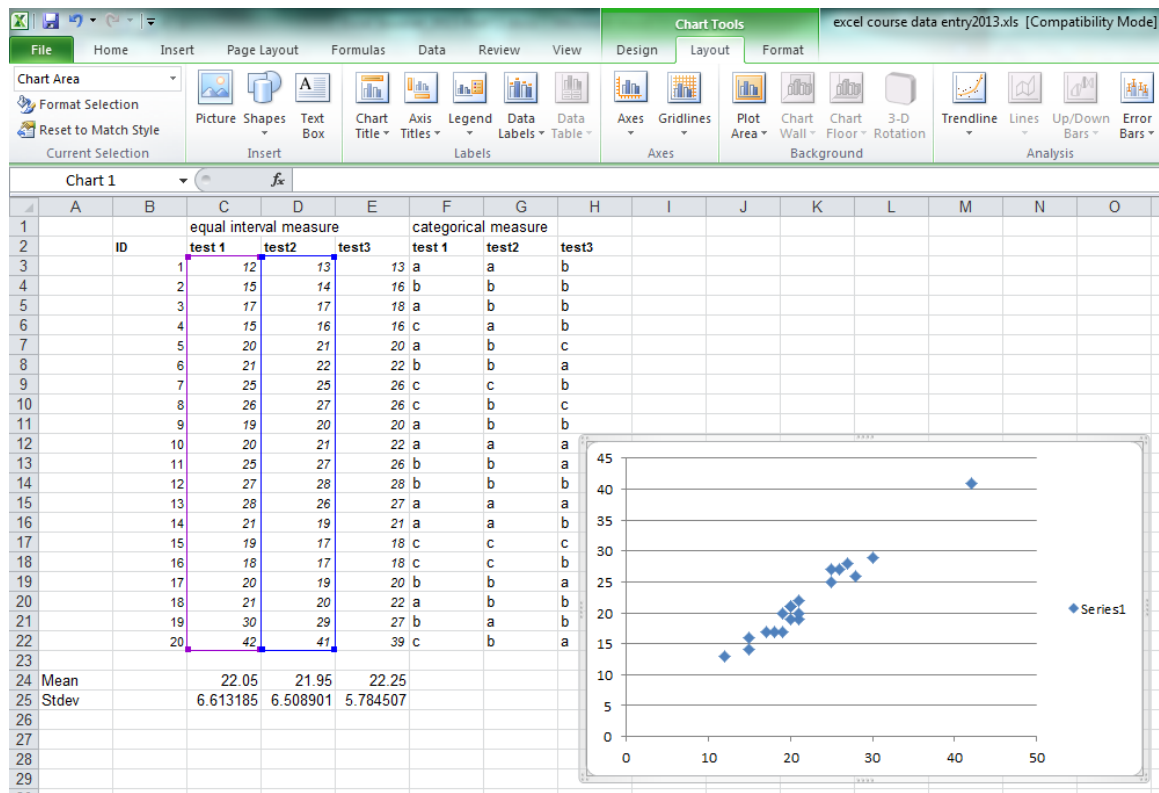
This is the output from the ANOVA model, which appears after you have completed all the steps as outlined in the DATA ANALYSIS process. You are not surprised to find that the *p* value is not significant, as it was not significant when you tested differences in the data using the T-test statistic. If you wanted to calculate an ICC, you would use these values from the ANOVA output in your formula (the within subject mean square, and the within-group mean square).

We now need to test for the correlations in the data, to see how valid and stable the measures were that the students took. This is usually undertaken by graphing the data, and then by a statistical test (Pearson *r*) for correlation. Using two of the columns of equal interval measures (tests 1 and 2), click on the drop down arrow (bottom right) in 'Charts' on the insert ribbon and follow the steps.

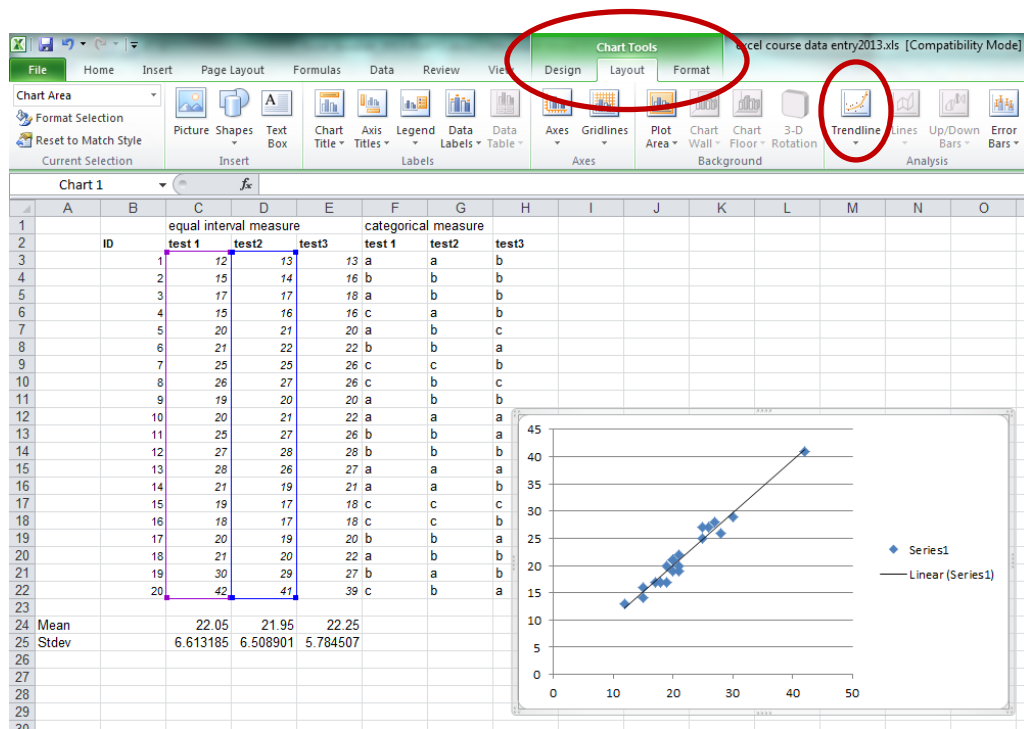
You want to select the scatter plot graph, which arranges the Test 1 data on one axis (usually the *x* or horizontal axis) against the Test 2 data (usually on the *y* or vertical axis). To be assured that the data is reliable it needs to be arranged closely around a line of 45 degrees (the line of best fit). You can superimpose this line once the graph has been made, or you can 'eyeball' it!!



If you follow the instructions for graph making, you will produce a graph such as this. This shows that when linked together the repeated test data is scattered variably around a straight line.



To impose the line of best fit on this graph, after the graph has been constructed and is sitting on the Excel sheet, click on the graph to highlight it, go to Chart Tools, select the Layout tab, and select the Trendline button. You have a choice then of all sorts of trend lines, what you want to test correlation in a linear sense is a Linear Trendline.



Correlation Coefficients

To put a value on the correlation outlined in this graphical display, use Pearson r statistics. When you calculate the Pearson r value, you go back to the two columns of data (Tests 1 and 2) and select a cell into which you want the calculation to appear. Remember how you calculated the AVERAGE and the STANDARD DEVIATION (STDEV)?

Type = PEARSON (you might need to go looking for this statistical function in the other statistical tests as it may not have been a regularly used feature). The program asks you to define the range of data (again Test 1 and Test 2) and then it calculates the statistic for you. The closer the r value is to 1 the better it is!

See the example on the next page. The Pearson r value for the test-retest is 0.98 (or 98%). This indicates that 98% of the values in the initial test are explained by the values in the second test. So there is 2% variability overall, which is influencing the variability of the measures. If you were the statistician on the students' team, you would be happy with this result. Ideally there should be a close correlation between data from the two tests, because the test isn't changing – so any error must be coming from the students! We will talk more about this in class!

The screenshot shows an Excel worksheet titled 'excel course data entry2013.xls [Compatibility Mode] - Microsoft Excel'. The worksheet has columns A through M and rows 1 through 31. The data is organized into two main sections: 'equal interval measure' (columns C, D, E) and 'categorical measure' (columns F, G, H). The 'equal interval measure' section has columns for 'test 1', 'test 2', and 'test 3'. The 'categorical measure' section has columns for 'test 1', 'test 2', and 'test 3'. The data is as follows:

	ID	test 1	test 2	test 3	test 1	test 2	test 3
1							
2							
3	1	12	13	13	a	a	b
4	2	15	14	16	b	b	b
5	3	17	17	18	a	b	b
6	4	15	16	16	c	a	b
7	5	20	21	20	a	b	c
8	6	21	22	22	b	b	a
9	7	25	25	26	c	c	b
10	8	26	27	26	c	b	c
11	9	19	20	20	a	b	b
12	10	20	21	22	a	a	a
13	11	25	27	26	b	b	a
14	12	27	28	28	b	b	b
15	13	28	26	27			
16	14	21	19	21			
17	15	19	17	18			
18	16	18	17	18			
19	17	20	19	20			
20	18	21	20	22			
21	19	30	29	27			
22	20	42	41	39			
23							
24	Mean	22.05	21.95	22.25			
25	Stdev	6.613185	6.508901	5.784507			
26							
27							
28							
29	pearsons r						
30							
31							

The 'Function Arguments' dialog box for the PEARSON function is open, showing Array1 as C3:C22 and Array2 as D3:D22. The formula result is 0.98190775.

Exercise 11.

Practice constructing these graphs and doing statistical tests for the paired data in the Test-Retest sheet.

Now how to we treat categorical data?

Let's imagine that the students have viewed a video of someone undertaking a specified test. It can be correctly or incorrectly performed (Yes = correct, No = incorrect). You want to see how good the agreement was between the students in whether the test was correctly performed or not. The students have viewed the video twice and have provided a response on two separate occasions. This is how your data might look. Can you enter this data (outlined in the next page) onto a new worksheet in your MS Excel workbook?

	A	B	C	D	E	F	G	H	I	J	K
1	Items being measured	Measure	Test 1	Test 2							
2	Brace test 1		1 Y	Y							
3	Brace test 1		2 Y	N							
4	Brace test 1		3 N	N							
5	Brace test 1		4 Y	Y							
6	Brace test 1		5 N	Y							
7	Brace test 1		6 N	N							
8	Brace test 1		7 Y	N							
9	Brace test 1		8 Y	Y							
10	Brace test 1		9 N	Y							
11	Brace test 1		10 N	N							
12											
13											
14											
15											
16											
17											
18											

To make sense of this data, and to summarise it, use the Filter Feature in the Data option on the Menu Bar. Filter the data into a 2x2 table so that you can see how many students said the test was successful on both occasions of testing, how many said that the test was unsuccessful on the two occasions of testing, and how many students disagreed on the two occasions of testing (YN, or NY).

	Yes1	No1	
Yes2			Total
No2			

You could calculate % agreement (the YES1_Yes2 and the No1_No2) scores as a percentage of the Total, however this doesn't make a lot of sense because you want all the students to either agree that the test was completed successfully or that it was not completed successfully.

If however you wanted to test the differences between the values provided in Test1 and the values provided in Test2 (the distributions of data as outlined below), you would use a chi square

statistic, which looks at the difference between the expected values and the actual values. Type =CHITEST from the statistical functions list. You will get a box like this.....

The screenshot shows an Excel spreadsheet with the following data:

Items being measured	Measure	Test 1	Test 2
Brace test 1	1	Y	Y
Brace test 1	2	Y	N
Brace test 1	3	N	N
Brace test 1	4	Y	Y
Brace test 1	5	N	Y
Brace test 1	6	N	N
Brace test 1	7	Y	N
Brace test 1	8	Y	Y
Brace test 1	9	N	Y
Brace test 1	10	N	N
TEST 1	5	5	
TEST 2	5	5	

The 'Function Arguments' dialog box for CHITEST is shown with the following details:

- Actual_range:** B14:C14 = {5,5}
- Expected_range:** B15:C15 = {5,5}
- Formula result:** = 1

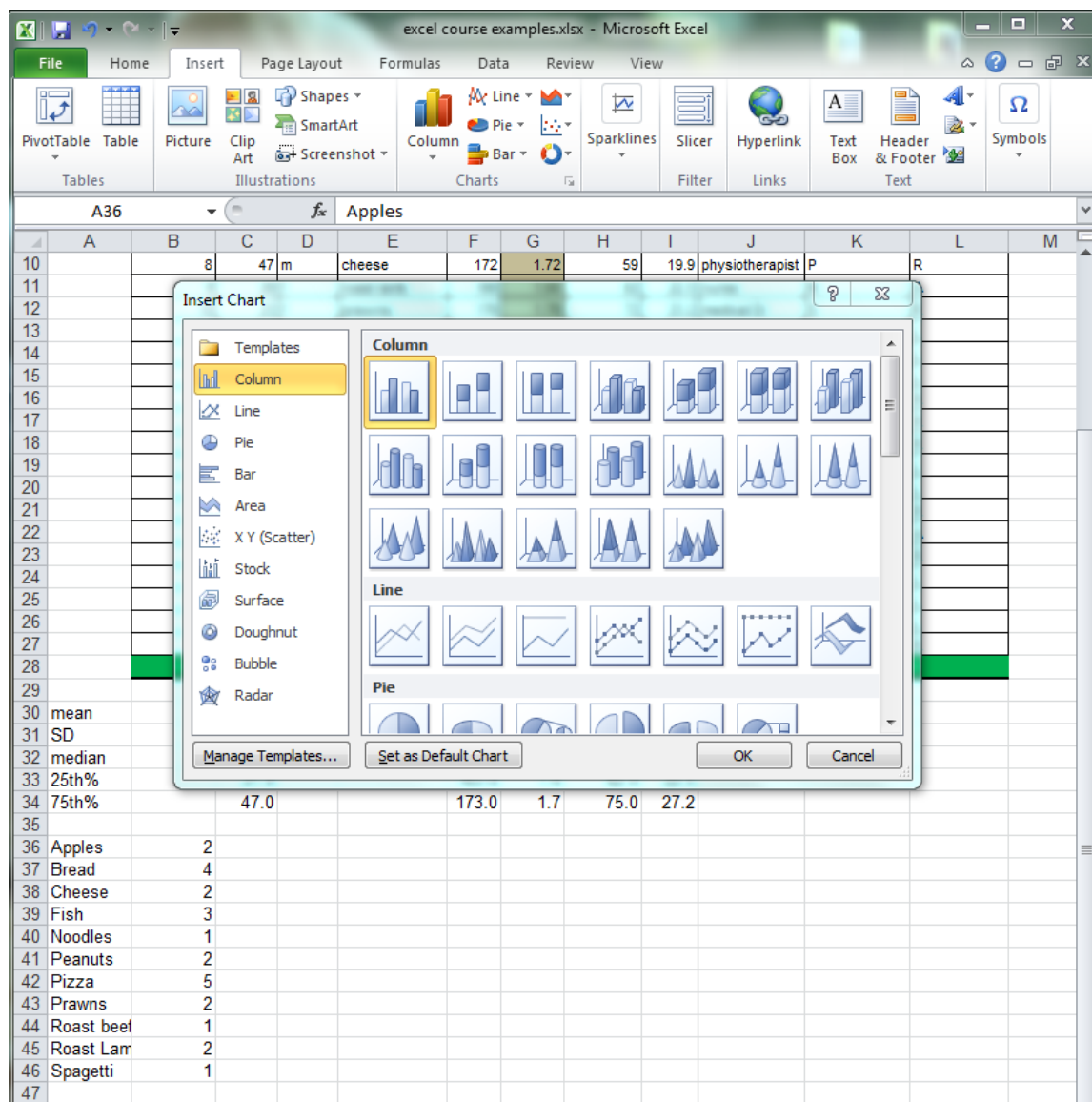
The actual data is the data from Test 1 and the expected data is from Test 2. The *p* value from the chi square test is 1, which indicates that there were no differences between tests 1 and 2. However, this doesn't really help us because we don't know whether the test itself was conducted correctly or not! The *p* value that demonstrates reliability is always greater than 0.05. If it is less, then you have significant differences. You can calculate chi squared statistics from any number of categories in MS Excel.

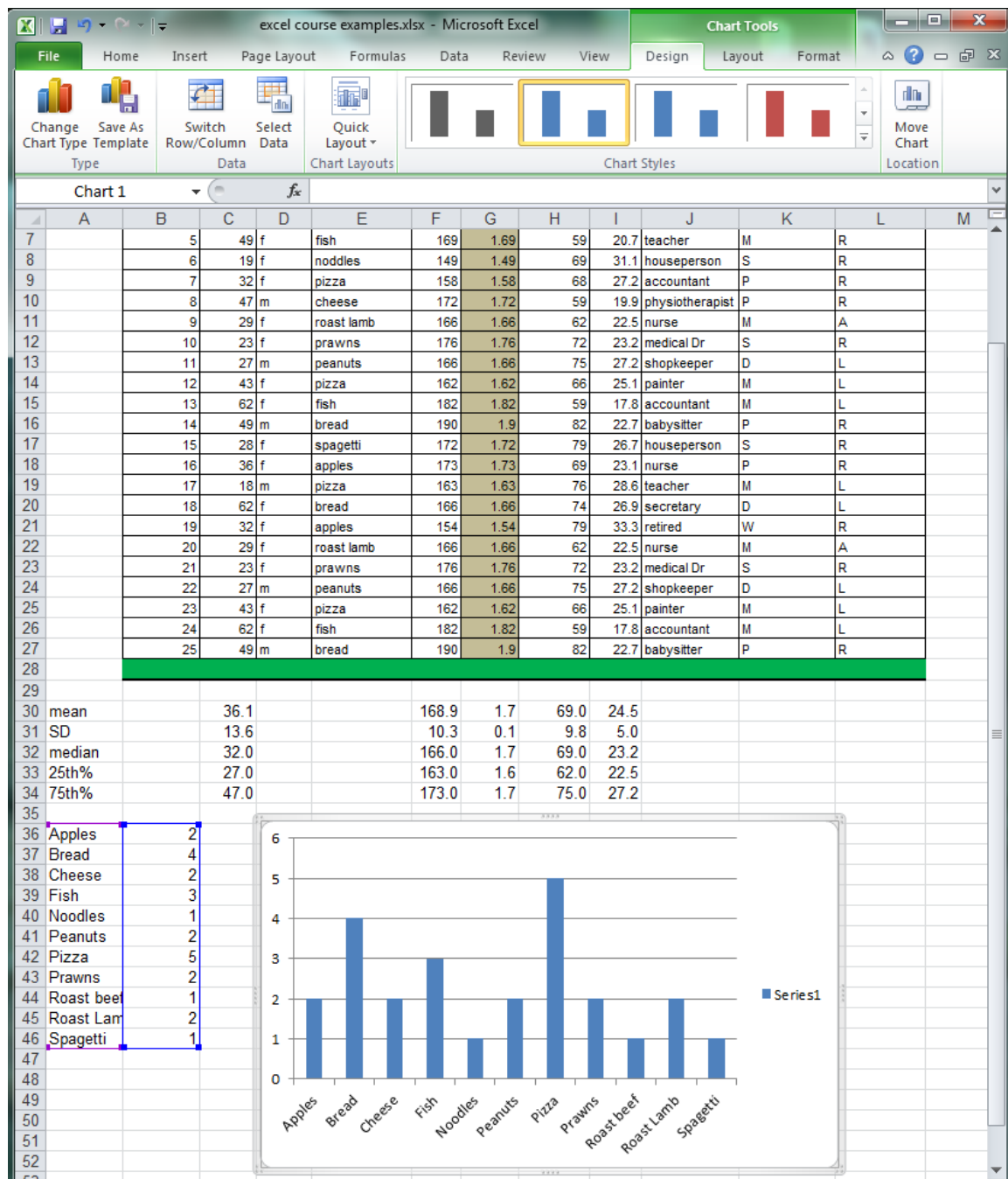
Exercise 12.

Go to the Test-retest sheet and use the repeated categorical data. Work out how to apply the statistics from this section to the three sets of categories.

Making graphs

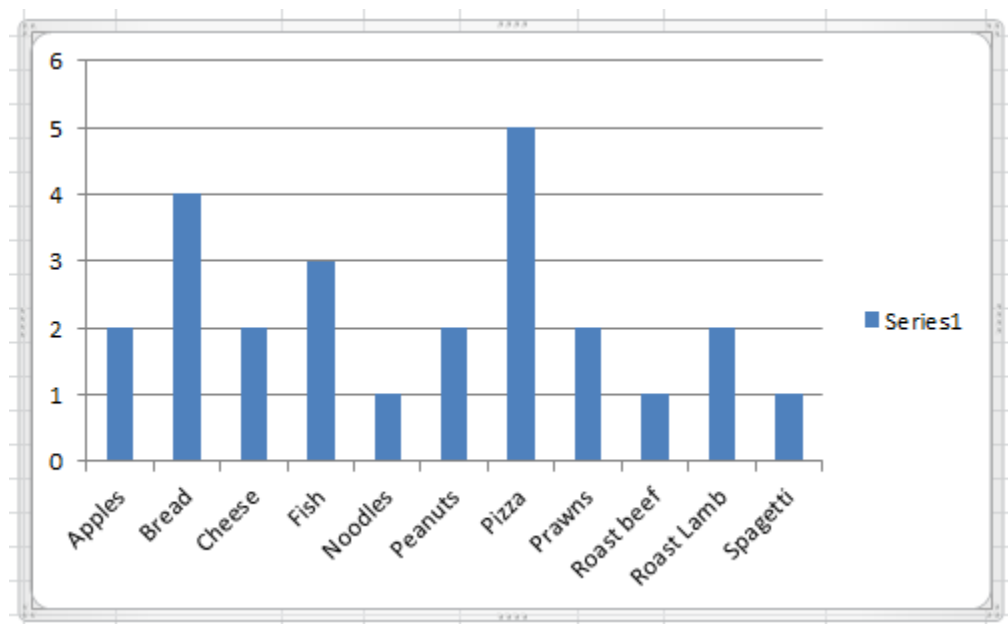
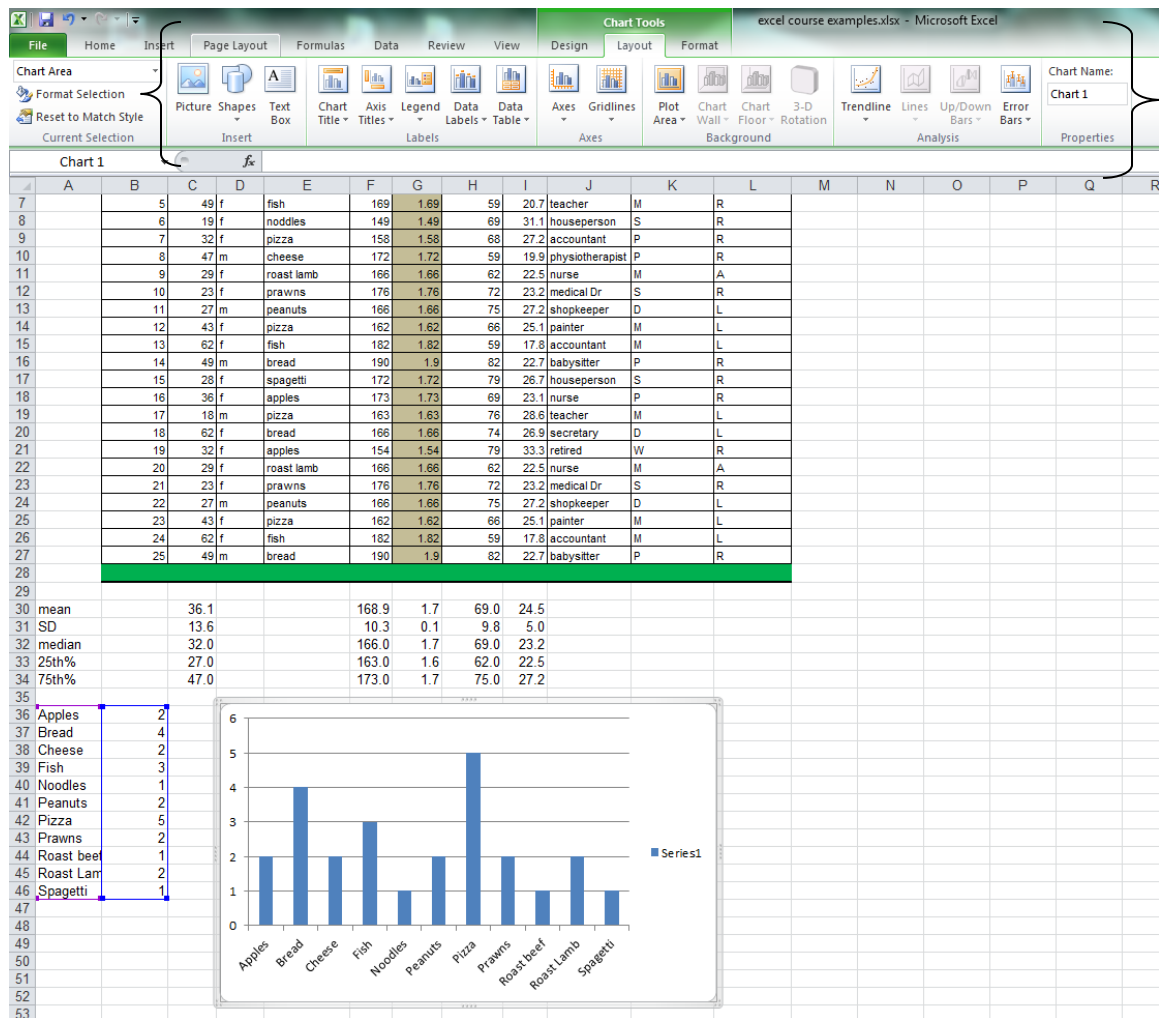
One of the wonderful features of Excel is its great graph making facilities. You have already made a scatter plot. Let's look at the Descriptive sheet (the favourite foods information) in the Excel file. To make sense of the information on the favourite food column, first you have to summarise it by using the Data filter function, and then once you have summarised the information in a couple of new columns, click on Charts. Summarising the category information will allow you to put it into a summary graph like a column graph, or a bar graph!





At this stage in the graph making, you can decide how you want your axes to be labelled (if at all), whether you want a legend, whether you want those gridlines – all by clicking onto the Chart Tools ribbons and turning things off and on, or putting in the text for your labels.

MS Excel Short Course in Statistics



This is a very basic graph! Once it is made, you can double click onto any element of the graph or in the Design, Layout or Format ribbons after it has been made, and change it. You can change its colour (bars, background etc), its axes (font, scale) and its labels (font, size, orientation). Remember, graphing needs to tell a story – if your graph doesn't tell an interesting story then it doesn't add anything to your statistics!

Exercise 13.

Using the data in any of the sheets in the Excel file, create different types of graphs. It doesn't matter what you make, as long as you have a reason for doing what you are doing! Try to make graphs that have some meaning.

Exercise 14

The dataset outlined in the sheet 'Experimental' was generated from a fictitious experimental study. A summary of the research is.....

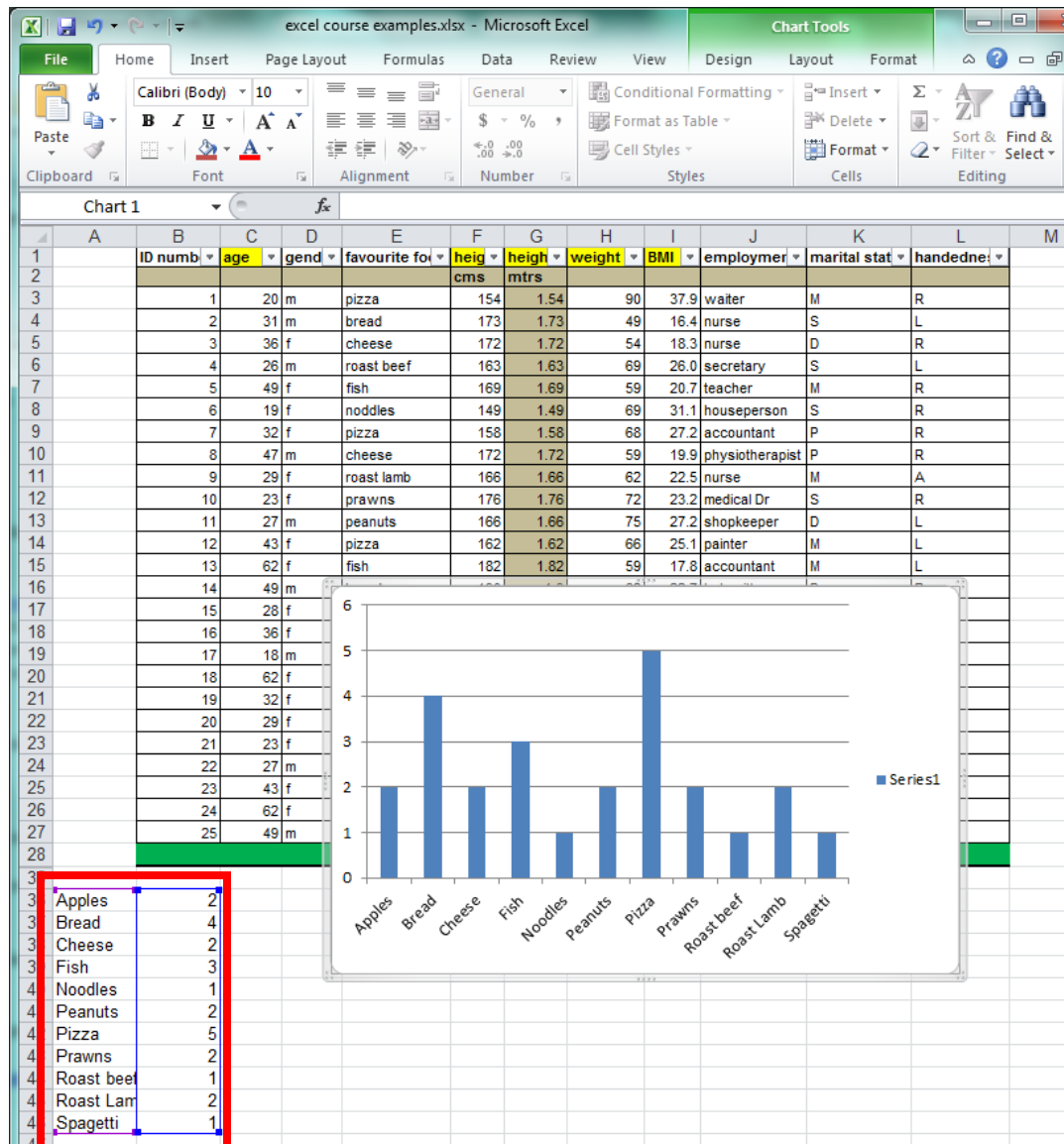
A chemical company tested two types of shampoo on the growth of hair (shampoo type 1 and 2). It took measurements of hair length before and after application of the shampoo. The participants were randomly allocated either shampoo 1 or 2, and used it the same number of times, in the same way over the same time period. The length of their hair was measured again after the test period. The shampoo company wanted to test whether the length of hair grown over the specified time period differed depending on the type of shampoo used.

Do the following tests on this dataset, following all the instructions in the booklet

- Divide the data into men and women using the data filter function.
- Sort the male data into ascending age
- Calculate the mean age and Standard Deviation for females
- Calculate the individual difference between the two length of hair measurements, and then calculate the mean difference, and the standard deviation
- Divide the data into the types of shampoo.
- Divide this into men and women
- Now calculate the mean difference in length of hair growth, for men and women, who used the two types of shampoo.

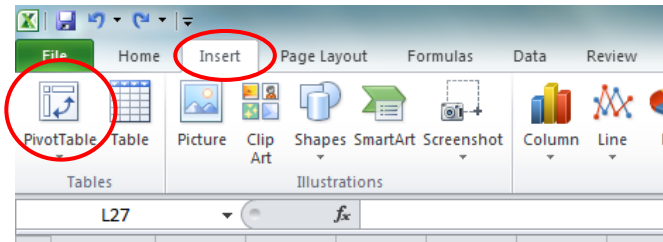
Making PivotTables

Another feature of MS Office is the function called PivotTable. A PivotTable report is useful to summarise, analyse, explore, and present summary of your data. PivotTable enables you to easily see comparisons, patterns and trends in data. For example, in the data set below, the summary table (highlighted in red) could be automatically created.



How to do it:

Select the cell or range of cells that contain the data, make sure that you include the column headings in the selection. On the **Insert** tab, in the **Tables** group, click **PivotTable**



In the **Create PivotTable** dialog box, make sure **Select a table or range** is selected and then select where to place the PivotTable, either a new worksheet or the existing worksheet. Click Ok. The PivotTable Field List box will appear on right hand side of worksheet. Place the fields you want summarized in the layout fields, by clicking and dragging the section to the field required. Nonnumeric fields are added to the Row Labels area, numeric fields are added to the Values area, and date and time hierarchies are added to the column labels area. For example, if you wanted to summarise the number of people that selected the same favourite food, you would put Fav Food in the Row Labels area and Gender or Employment or Marital Status (a field that is nonnumeric) in Values area.

Go to the next page for the example

The screenshot shows Microsoft Excel with a PivotTable and the PivotTable Field List task pane. The PivotTable is located in the range M6:Q19. The PivotTable Field List task pane is on the right, showing the following fields:

- Choose fields to add to report:
 - ☐ ID number
 - ☐ age
 - ☐ gender
 - ☒ favourite food
 - ☐ height
 - ☐ height2
 - ☐ weight
 - ☐ BMI
 - ☒ employment
 - ☒ marital status
 - ☒ handedness
- Drag fields between areas below:
 - Report Filter: (empty)
 - Column Labels: (empty)
 - Row Labels: favourite food
 - Σ Values: Count of emplo..., Count of marit..., Count of hand...

The PivotTable data is as follows:

Row Labels	Count of employment	Count of marital status	Count of handedness
apples	2	2	2
bread	4	4	4
cheese	2	2	2
fish	3	3	3
noddles	1	1	1
peanuts	2	2	2
pizza	5	5	5
prawns	2	2	2
roast beef	1	1	1
roast lamb	2	2	2
spagetti	1	1	1
(blank)			
Grand Total	25	25	25

By placing different fields into different areas, summary of various data are possible.

You can also apply further filters to your data, by adding **Slicers**. These allow you to filter based on certain fields in an interactive way. For example, if you wanted to know how many females said pizza was their favourite food. To add Slicers, go to the **Options** Tab in the **PivotTable Tools** ribbon, click on **Insert Slicer**. In the Insert Slicers dialog box **check the tick box** of the field you wish to filter based on, e.g. Gender, then **OK**. A dialog box will open and you can select the particular value you would like to filter by, eg Male or Female.

The screenshot shows the Microsoft Excel interface with the PivotTable Tools ribbon active. The ribbon includes the Options and Design tabs. The Options tab is selected, showing various actions like Expand Entire Field, Collapse Entire Field, Group Selection, Ungroup, Group Field, Sort, and Insert Slicer. The Design tab shows options like Refresh, Change Data Source, Clear, Select, Move PivotTable, Summarize Values By, Show Values As, Fields, Items, & Sets, and PivotChart.

The data table is as follows:

	D	E	F	G	H	I
Gender	Fav Food	Height	Weight	Employment	Marital Status	Handedness
1	pizza	154	55	waiter	M	R
2	bread	173	78	nurse	S	L
3	cheese	174	73	nurse	D	R
4	roast beef	176	79	secretary	S	L

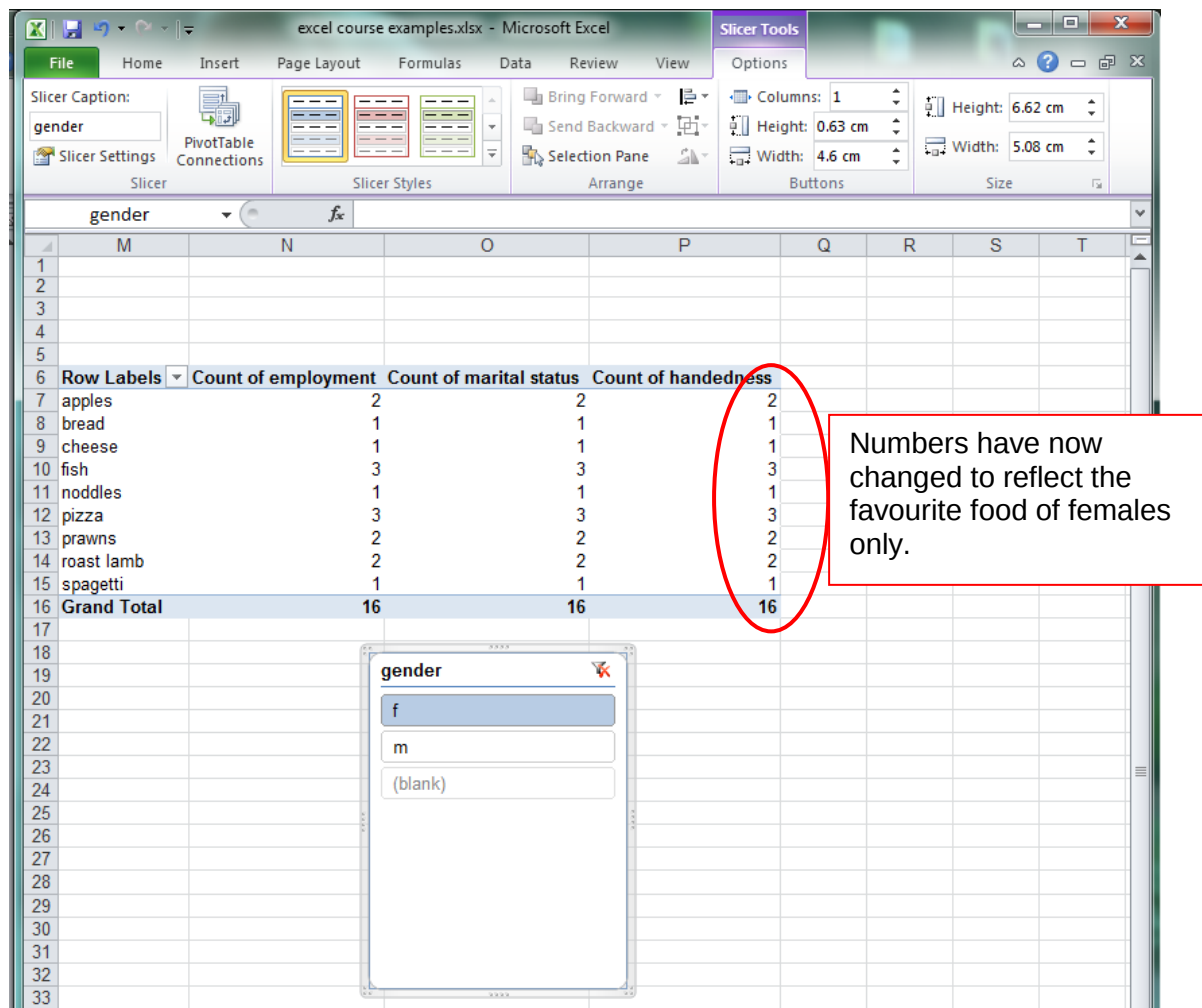
The PivotTable is located in the range L1:M4, with the formula bar showing the formula =COUNTIF(D2:I4, "M"). The PivotTable is titled "Count of Marital Status" and shows the following data:

	L	M
Count of Marital Status		
M	2	
S	1	
D	1	
L	2	

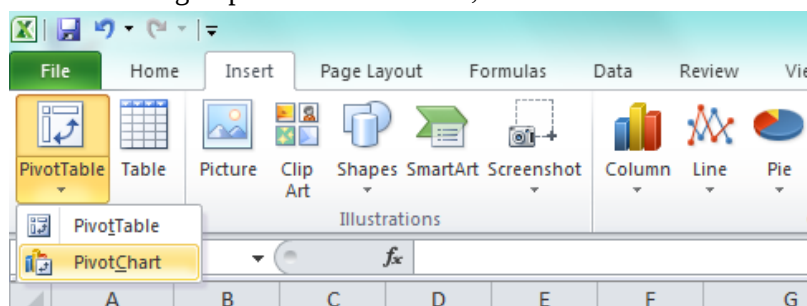
The Insert Slicer dialog box is open, showing a list of fields to choose from. The fields are:

- ID Number
- Age
- Gender
- Fav Food
- Height
- Weight
- Employment
- Marital Status
- Handedness

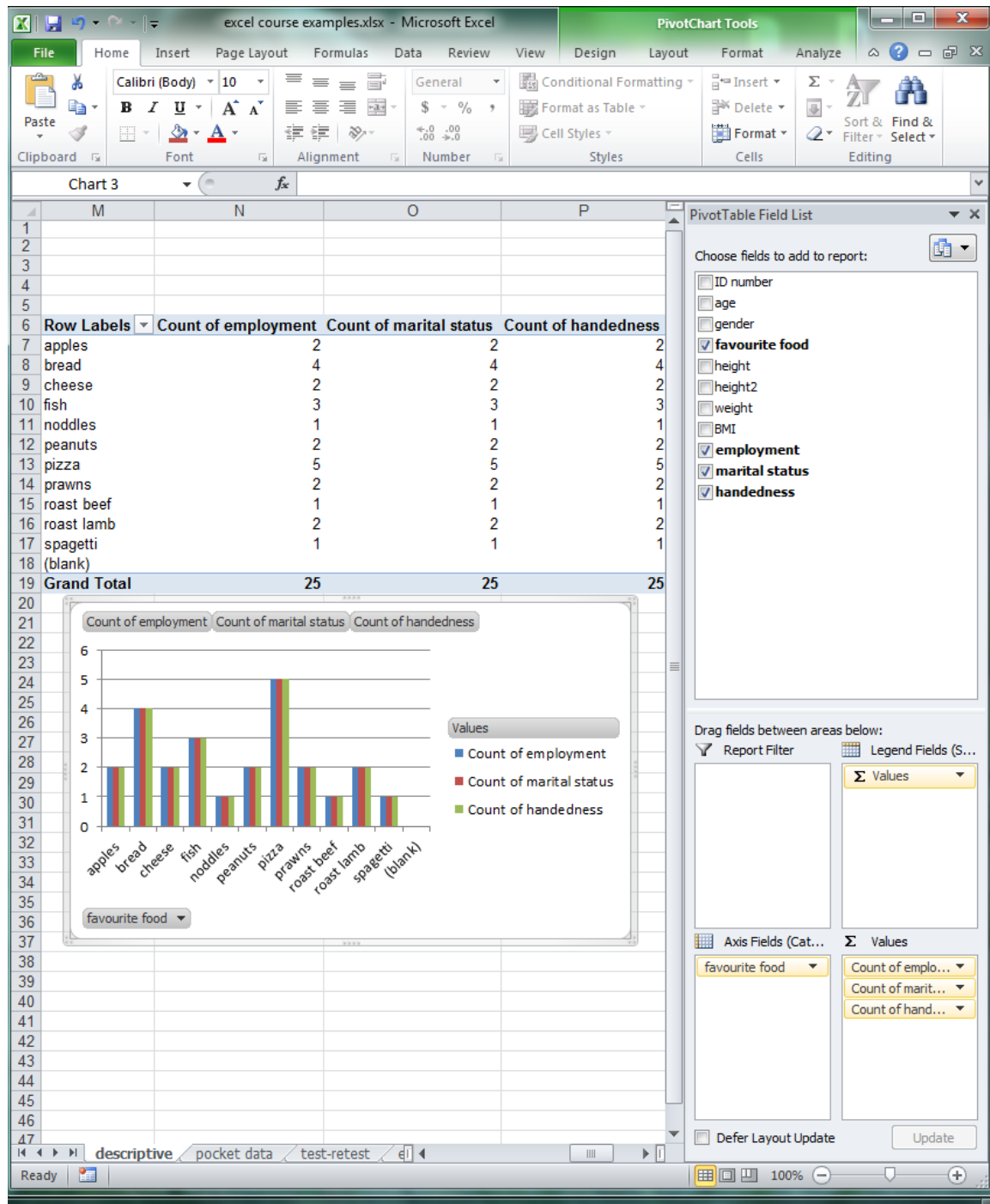
The "Gender" field is selected in the list.



PivotTable can automatically generate a graph of the summarized data by using the PivotChart Tool. Instead of selecting PivotTable when creating a table at the start, click the drop down menu of the **Tables** group in the **Insert** menu, select **PivotChart**.



By selecting PivotChart and then doing the same steps as above you would create both a summary of the data and a graph representing the number of people who chose a certain favourite food.



Transferring data around the MS Office Suite

MS Office uses interactive programs that have similar Menu bars, and which will accept information from all other programs. It is a good idea to become familiar with moving between the programs in the MSOffice suite, as it saves you time and effort, and allows you to make very professional presentations. This final section of the course gives you the opportunity to explore copying information from one program to another. It also teaches you about the interactivity of

the programs, whereby you can change a table or graph generated in Excel, in one of the other programs.

Transferring data to and from MS Excel and MS Word

Open up a new MS Word file. Use the same approach as outlined in the beginning of this workbook for MS Excel. Copy several columns from one of the MS Excel datasheets, go across to the Word document, and paste them in. Now, do the same with one of the graphs that you have made. See how the tables and graph transfer so easily. Whilst you have the MS Excel file open behind the MS Word file, you can make changes to the graph in MSWord by clicking onto the graph, changing the data table which appears behind it, and then clicking out of the graph.

Exercise 15a

Using your findings from Exercise 14, write 100 words or so on what you found from the experiment. Illustrate your words with a table, and a graph or two of the findings.

Transferring data to and from MS Excel and MS PowerPoint

Open up a new MS PowerPoint file. Use the same approach as outlined in the beginning of this workbook for MS Excel. You will see a number of choices of slides. Choose one which you can paste a graph into. Copy one of the graphs that you have made in Excel, and paste it in. See how the graph transferred so easily. Whilst you have the MS Excel file open behind the MS PowerPoint file, you can make changes to the graph in MS PowerPoint by clicking onto the graph, changing the data table which appears behind it, and then clicking out of the graph. Play around with MS PowerPoint, putting in headings and other text. There is a lot that you can do in MS PowerPoint, including using different fonts and font sizes, different backgrounds, importing pictures and logos etc. Think about the style of visual presentation that works best for you – colors, fonts, backgrounds etc. When your lecturers present to you, think about what is the most effective teaching approach using visual aids.

Exercise 15b

Using your findings from Exercise 14, construct 4-5 slides of what you found from the experiment. Imagine that you are going to use these slides to illustrate a short presentation on the experiment and the results.

Moving from Tables to Text and Back

Open the MS Word file 'References for playing around with.doc'. Using the Tab key put a tab between the separate sections of the reference (authors, date, title, journal, volume, and page numbers). Do this only for the first few references as it soon gets boring. Now highlight the references that you have 'tabbed', copy them, and put them on another new file. While they are still highlighted, go to the Insert ribbon, click on the arrow under Table, and select Convert Text to Table, and see if you can put your references into a multiple column table form in Word. You should have as many columns in your table as you have tabs. If your tables haven't worked as expected, make sure that you have put in the Tabs as you were instructed. Now copy this table across into MS Excel. To do this, open a new sheet, click into the first cell of the sheet, and PASTE what you have copied. You will notice that the word data is now arranged in columns

that you can SORT, FILTER etc. You can also sort in MSWord. Think about the usefulness of being able to move this sort of data around between MSWord and MS Excel. Could you use this technique to keep track of your references this year, or your weekly timetable?

Exercise 16

Go to the checklist at the front of the workbook and rate yourself. If you rate OK or Good for every one of the competencies, you have successfully completed this course. You might need to practice the skills to keep yourself fit!!! If you don't rate yourself as highly, this book provides you with all the information you need to repeat MS Excel steps. Using MS Excel has lots of benefits. It is simply a tool that makes light work of statistical and descriptive calculations. It can help you to store your personal data; it is great for making graphs and keeping lists (like references). It is great for quick calculations that you would normally make mistakes when using pen and paper, or a calculator.

Audits

MS Excel is a great tool for undertaking audits. For instance you could take a pre-established and piloted MS Excel sheet on a laptop to the records office (or wherever you are accessing the patient records) and you could enter your data without having to use an interim paper-based step. You can also enter more columns of data as you need them (**make sure you only put one piece of information in each cell**), and the MS Excel sheets can hold thousands of rows of patient data. Keep your raw data collection to a minimum (only collect what you need). Remember that MS Excel formulae features readily assist you to calculate derived data such as Length of Stay (LOS) (subtracting the admissions date from the discharge data), or the delay between admission and therapy assessment dates etc.

Exercise 17

Look at the sample audit data (Audit) in the practice MS Excel file. This is 'made up' data which could well be from an audit that aimed to describe features of a hospital admission for patients admitted to two hospital wards (2A, 2B) for hip and knee replacements. Now add another 10 rows of information for this (pretend) patient population. Be outrageous in your numbers so that you have something to play with. **NB** Notice that there is no need to enter information on LOS, or delay in therapy, as this can be derived from raw data using the appropriate formulae. Practice populating new columns of derived data. The formulas for LOS and therapy delay are provided in the first cell for your reference.

What other 'new' data could be derived from these existing data items? Practice establishing and then calculating other derived data.

NOW.....go to the Sheet 'Audit Benchmarks'. These are 'hypothetical' benchmarks for average LOS for the different diagnoses and age groups (and the allowable variability to extend the LOS for a particular reason), and also the average expected delay between admission and therapy assessment, and between therapy assessment and commencement of therapy.

Considering only the data in the Audit sheet, what descriptive audit (research) questions could be asked? Are there any cross-correlations which could be undertaken between the data items?

Using the Audit Benchmark data, superimpose one of the benchmarks on your audit data. How does your sample comply with the benchmark (% compliance?) Could this audit have collected other information which might have explained variations in compliance?

Go to the checklist at the front of the booklet, and see how many of the items in it you feel you are now competent in!!!