statstutor

Statistical Methods
4. Introduction to Excel and Pivot Tables

Based on materials provided by Coventry University and Loughborough University under a National HE STEM Programme Practice Transfer Adopters grant

National HE STEM Programme

www.statstutor.ac.uk

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Summary

❑ Opening Excel

❑ File handling

❑ Excel interface

❑ Pivot tables for percentage frequencies of category values of a variable

❑ Built-in statistical functions of a single series

❑ Pivot tables for statistical functions of category values of a variable

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Opening Excel

To open Excel without a data file:

1. Click on the Start button

2. Select "All Programs"

3. Select "Microsoft Office"

4. Select "Microsoft Excel 2010"

You can also select Excel from the recently used programs list, if it is visible in the list

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield
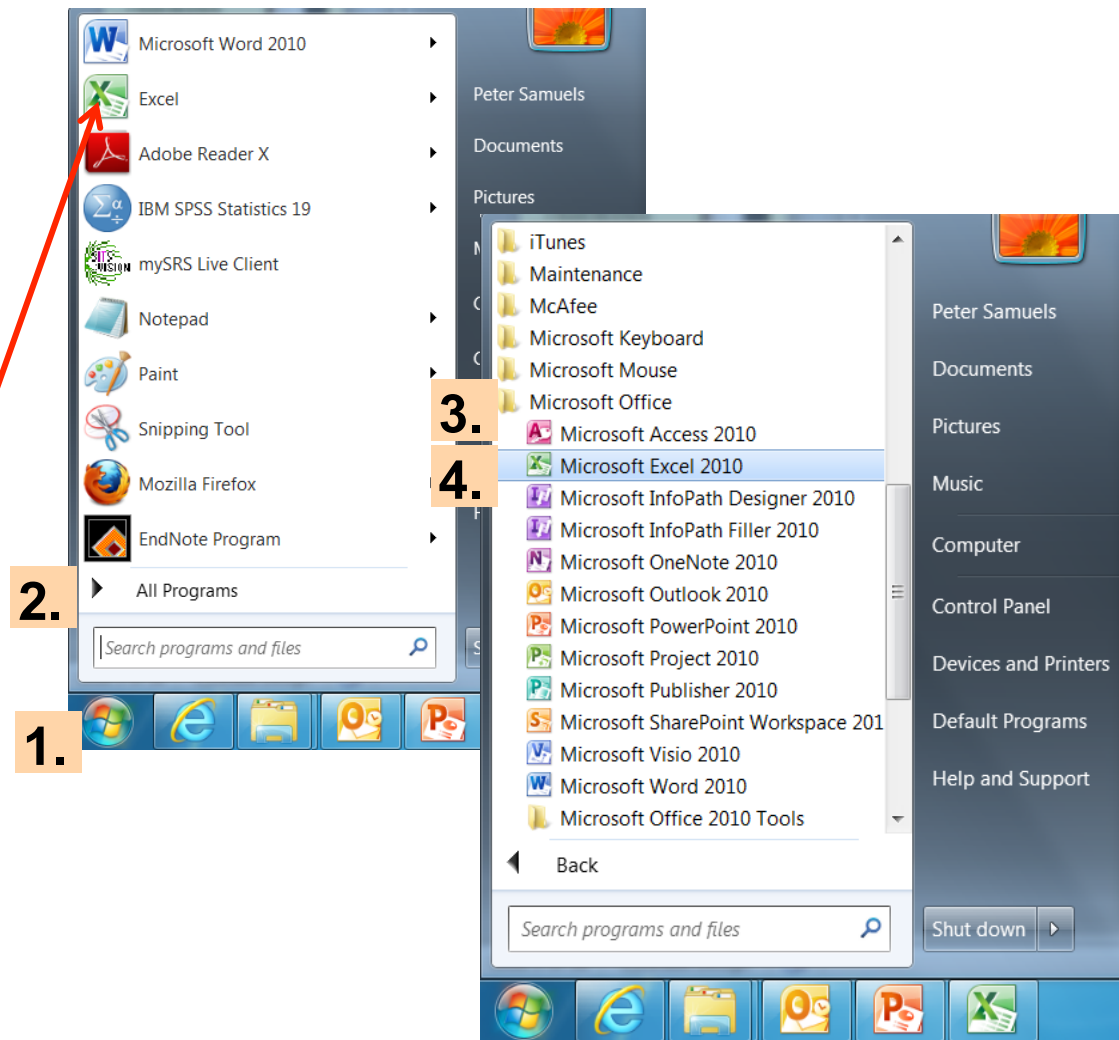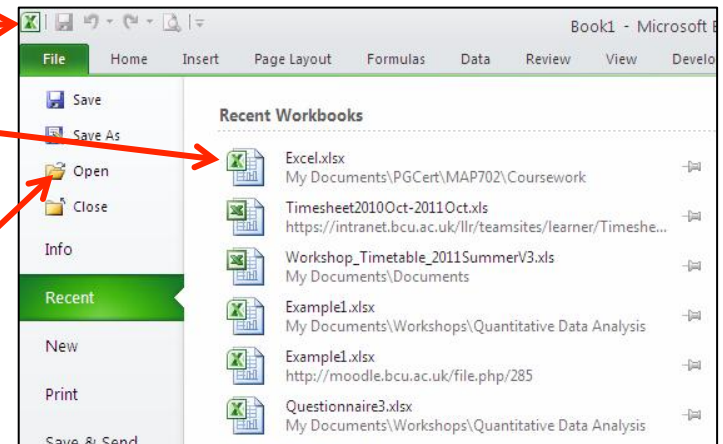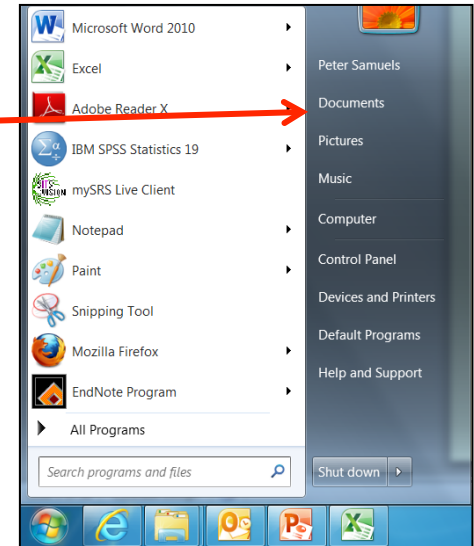
How to open an Excel data file

❑ Use a file manager such as Documents:

   ➢ Move to the folder containing the file

   ➢ Double click on the file

or

❑ Open Excel and select the File menu:

   ➢ Select the file name from the list on the right hand side, **or**

   ➢ Select Open and locate the file

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

File handling 1: How to copy a file onto your computer

- ❑ Open two file managers
- ❑ Navigate to the folder containing your file on the memory stick (or other storage media) in one and the Documents folder (or a subfolder) in the other
- ❑ Click and drag a file from the memory stick (or other storage media) folder to the Documents folder (or a subfolder)
- ❑ Take care not to drag into a subfolder by mistake
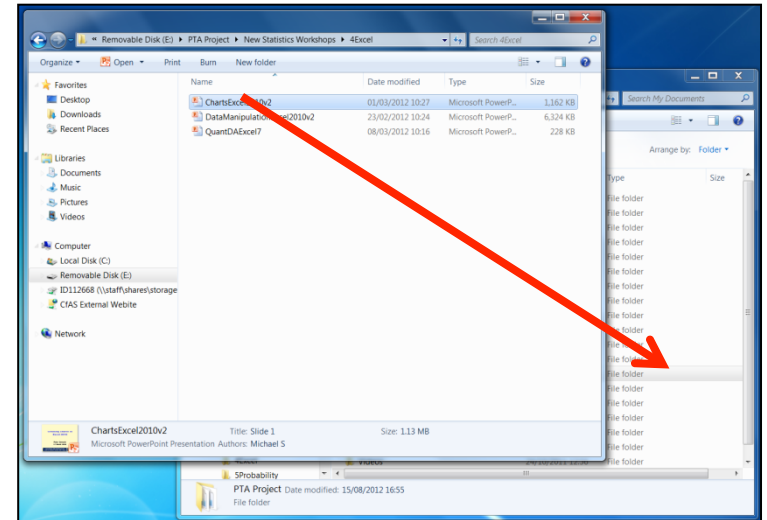
If you are requested for permission to overwrite a file with the same name, you normally need to overwrite with a **newer** one

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

File handling 2: Creating a subfolder

1. Open the "Documents" application from the Start button

2. Select: "New folder"
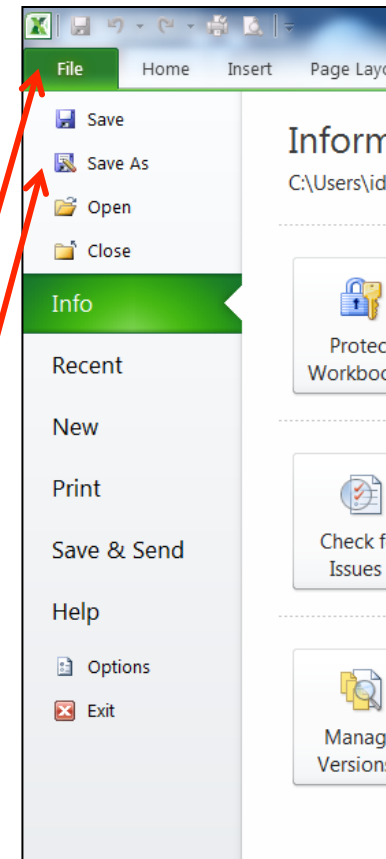
3. Enter the name of the new folder as "4ExcelPivotTables"

You can create one folder for each workshop in this course

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield
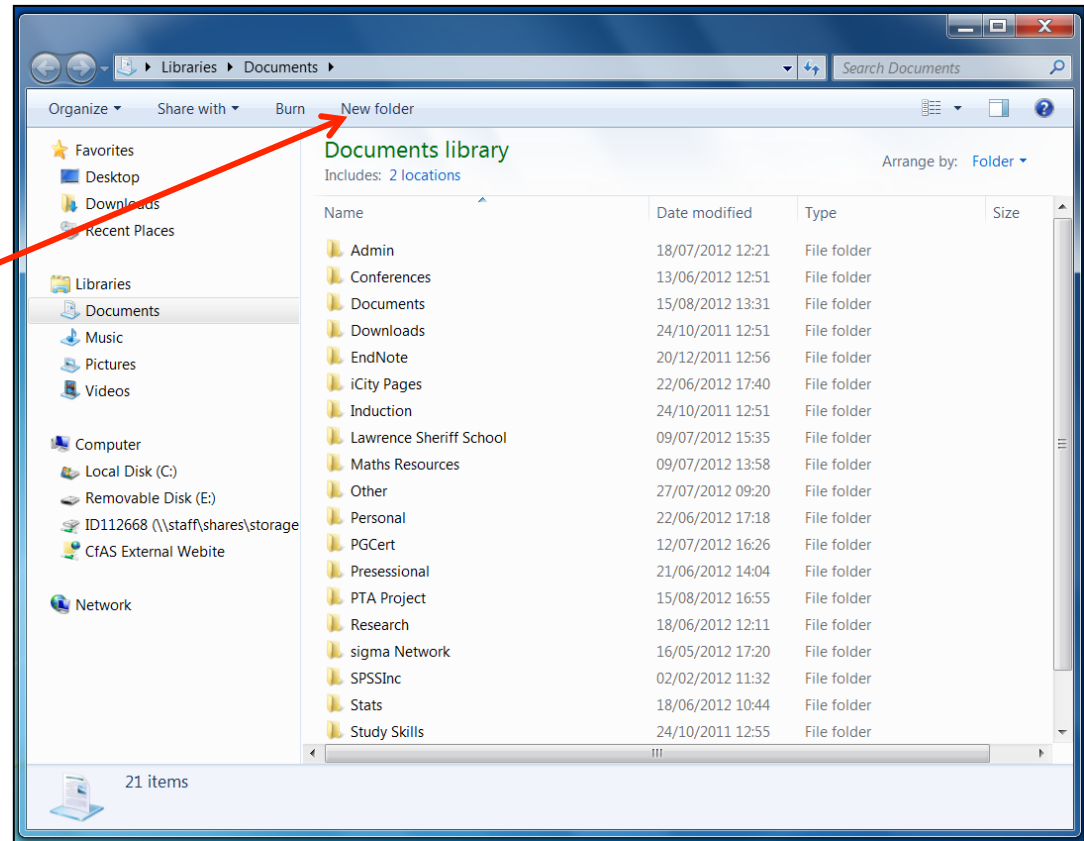
File handling 3: Creating a new version of a file

1. Open the folder "4ExcelPivotTables" in a file manager

2. Make a copy of the file SurveyData.xlsx (associated with this presentation) in this folder by selecting the file and using the Save as option

3. Open this copy of the file by double clicking on it in the Documents application

4. Rename it as SurveyData2.xlsx using File – Save As

Always rename your version at the **start** of a new session

Peter Samuels
Birmingham City University

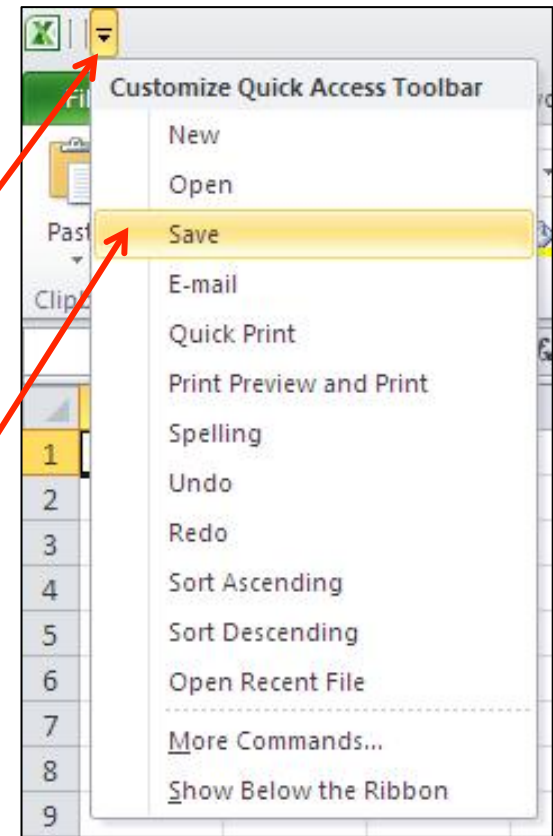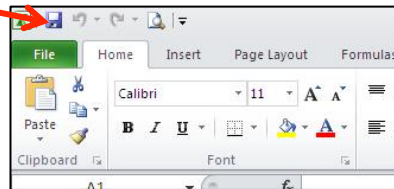Reviewer: Ellen Marshall
University of Sheffield

File handling 4: Saving your work regularly

Good practice to save your work regularly (e.g. every 5 to 10 minutes)

Create a Save shortcut button (floppy disk picture) on the Quick Access Toolbar by:
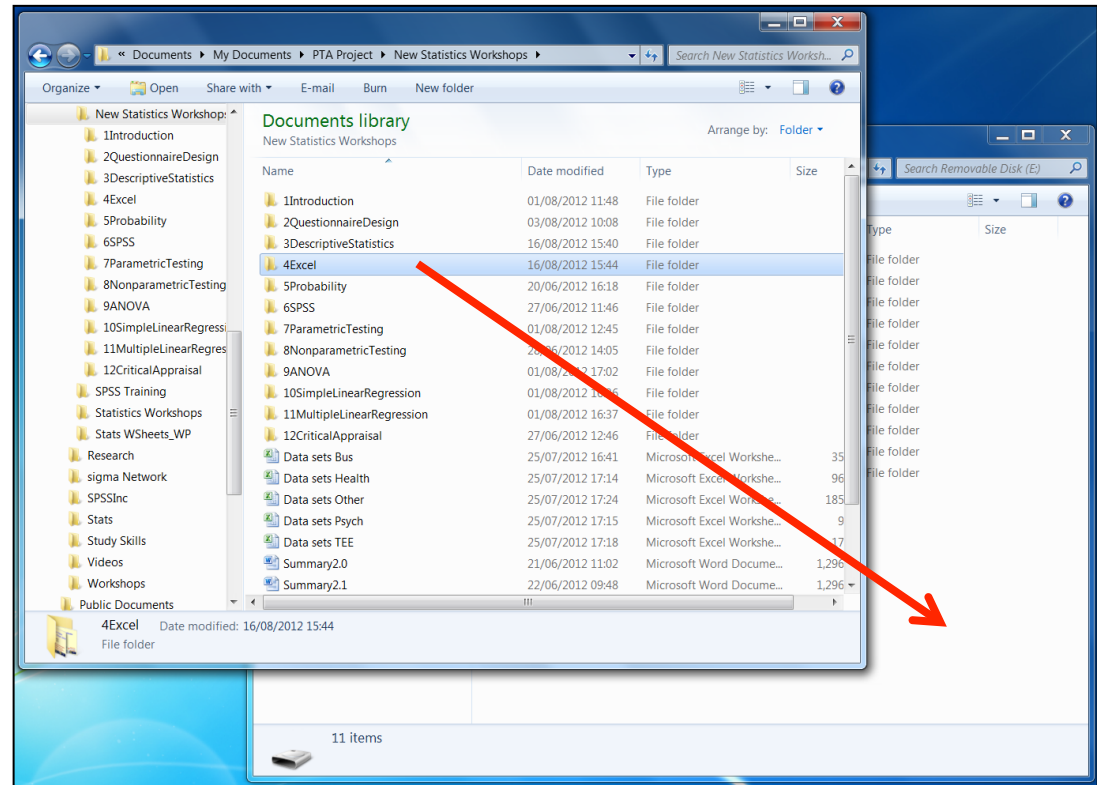
❑ Selecting the Quick Access Toolbar

❑ Selecting Save

❑ Selecting the Save button which has been created

www.**stats**tutor.ac.uk

Peter Samuels
Birmingham City University

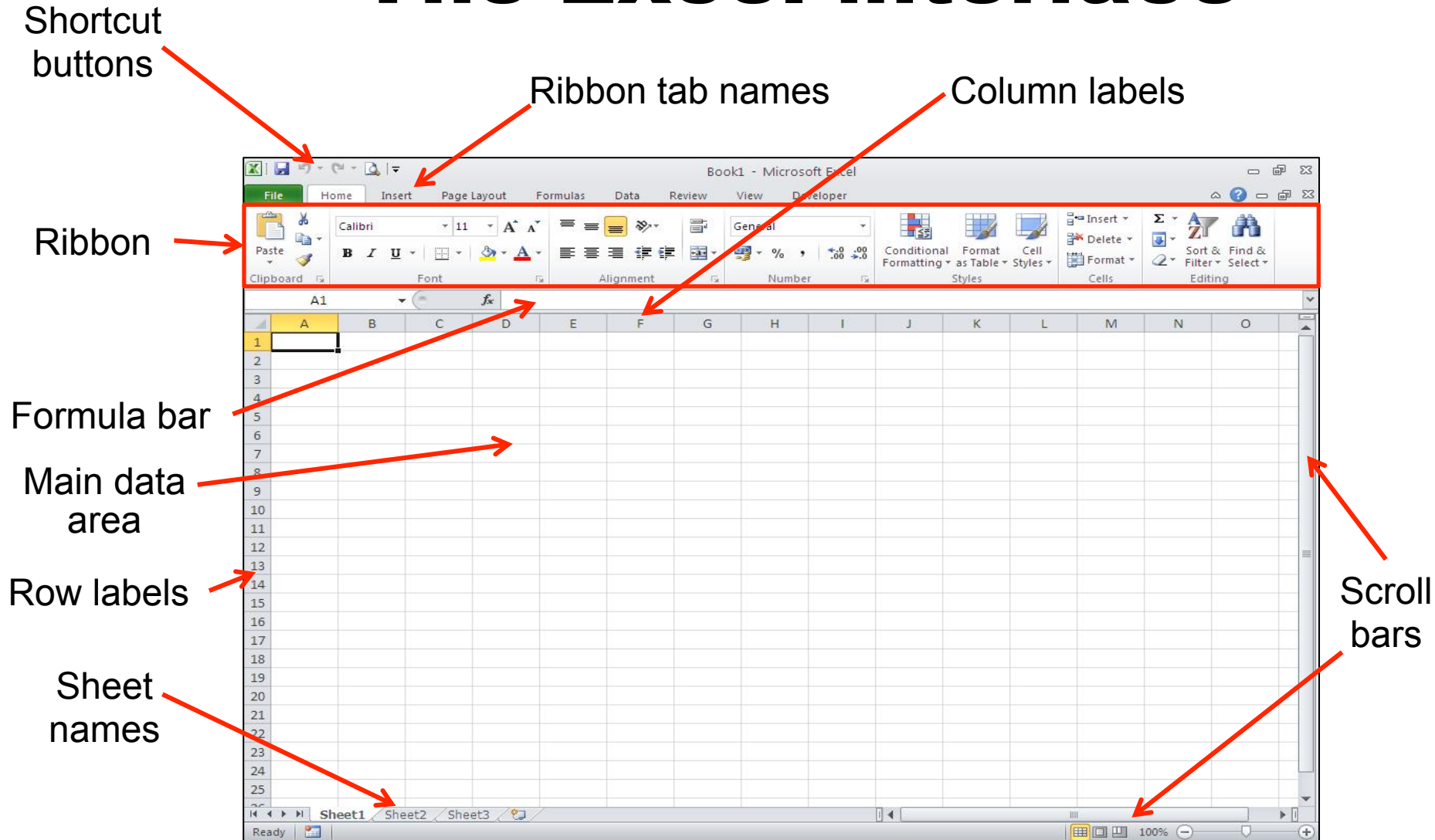Reviewer: Ellen Marshall
University of Sheffield

File handling 5: Backing up files

It is always good practice to keep your files on 2 or 3 different media.

Copy the folder 4ExcelPivotTables from your computer onto a memory stick using two file managers

The Excel interface

Shortcut buttons

Ribbon tab names

Column labels

Ribbon

Formula bar

Main data area

Row labels

Scroll bars

Sheet names

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Pivot tables: aims

❑ To demonstrate the usefulness of pivot tables in analysing data

❑ To help you feel comfortable with the creation and manipulation of pivot tables

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Initial definition: datasets

A **dataset** is a block of cells containing headings and data, surrounded by empty cells (or the edges of the worksheet)

So this is a dataset

and so is this

but this isn't

Useful keyboard shortcuts

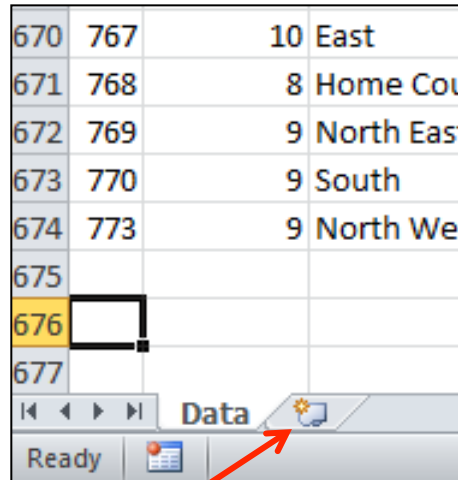| Shortcut	Effect
Shift + an arrow key	Selects cells
Ctrl + Home	Moves to the first cell of the sheet (A1) ('+' means press together)
Ctrl + an arrow key	Moves to the edge of a dataset
Ctrl + C	Copies selected text to the clipboard
Ctrl + V	Pastes from the clipboard

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Creating a pivot table step by step
1. Open your Excel file

Open the file SurveyData2 you created earlier

	A	B	C	D	E	F	G	H
1	ID	YearGroup	Location	Gender	Age	Height	FootLengthCm	ElbowToW
2	1	9	South West	Male	13	155	25.5	
3	2	8	South East	Male	13	176	27	
4	3	9	South	Male	14	172	25	
5	4	7	North West	Male	11	152	26	
6	5	9	South West	Male	14	165	27	
7	6	9	North West	Male	14	165	27	
8	7	8	South	Male	13	164	26	
9	8	7	East Midlands	Male	12	156	25	
10	9	0	Hong Kong	Male	14	174	25	
11	10	8	South West	Male	12	157	24	
12	11	9	Home Counties	Male	13	150	22	

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

2. Create a new sheet
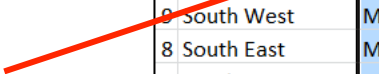
Click here, at the bottom of your sheet

And here's your new sheet

www.**stats**tutor.ac.uk

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield
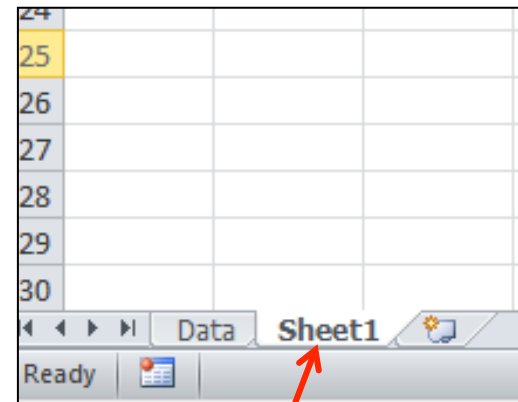
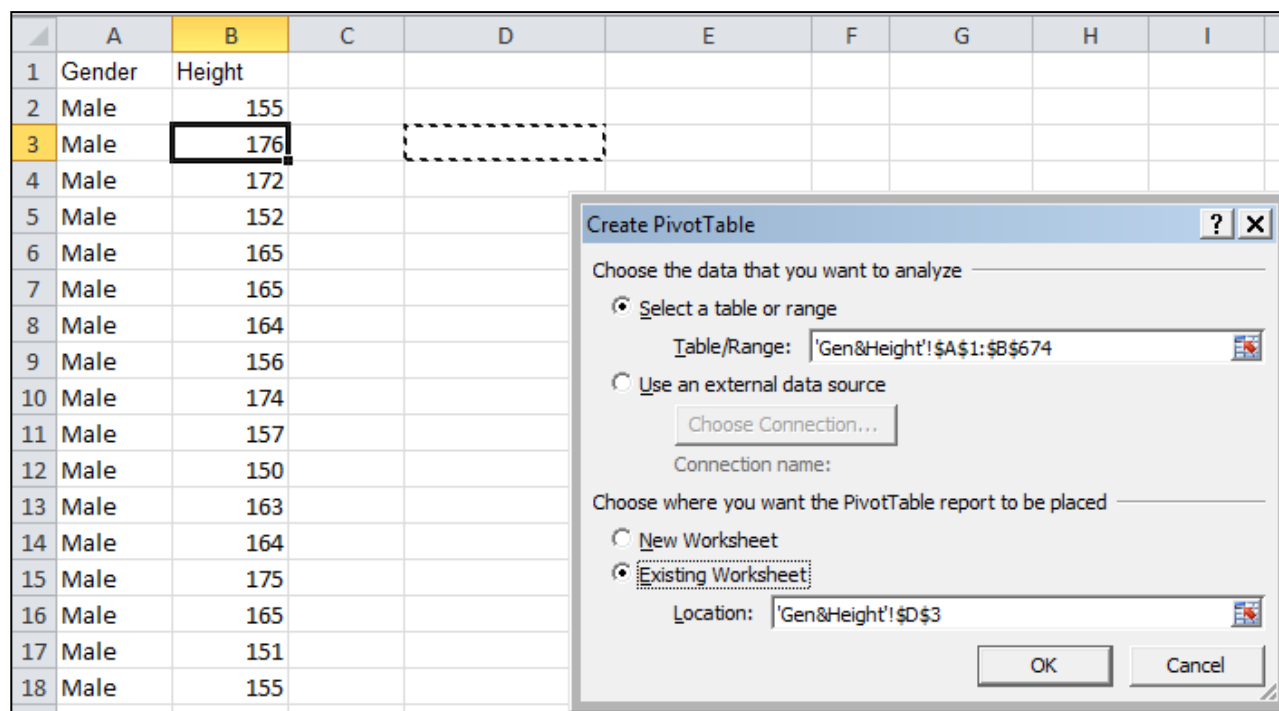# 3. Copy and paste the *Gender* and *Height* data from the *Data* worksheet onto *Sheet1*

❑ Select the column using the column label

❑ Use **Ctrl + C** to Copy

❑ Select a blank column on a new sheet

❑ Use **Ctrl + V** to Paste

	C	D	E	F	
p	Location	Gender	Age	Height	Foc
9	South West	Male	13	155	
8	South East	Male	13	176	
9	South	Male	14	172	
7	North West	Male	11	152	
9	South West	Male	14	165	
9	North West	Male	14	165	
8	South	Male	13	164	
7	East Midlands	Male	12	156	
0	Hong Kong	Male	14	174	
8	South West	Male	12	157	
9	Home Counties	Male	13	150	
0	East Midlands	Male	14	163	
9	Home Counties	Male	14	164	
9	North West	Male	13	175	
9	North East	Male	13	165	
8	West Midlands	Male	12	151	

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

4. Open the PivotTable dialog

❑ Click inside your new dataset on Sheet1

❑ Use **Insert > PivotTable** to bring up the PivotTable dialog

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

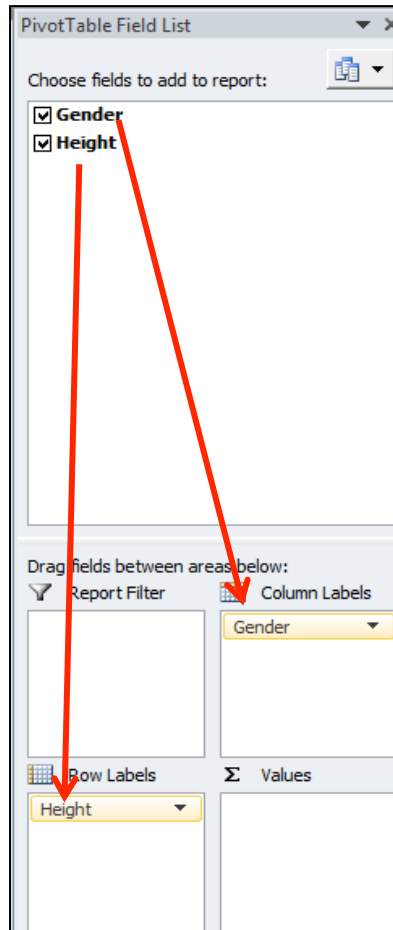# 5. Decide where to put your pivot table

❑ Ensure the Existing worksheet radio button is selected
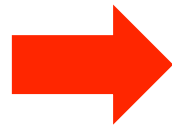
❑ Select the cell D3 then select OK
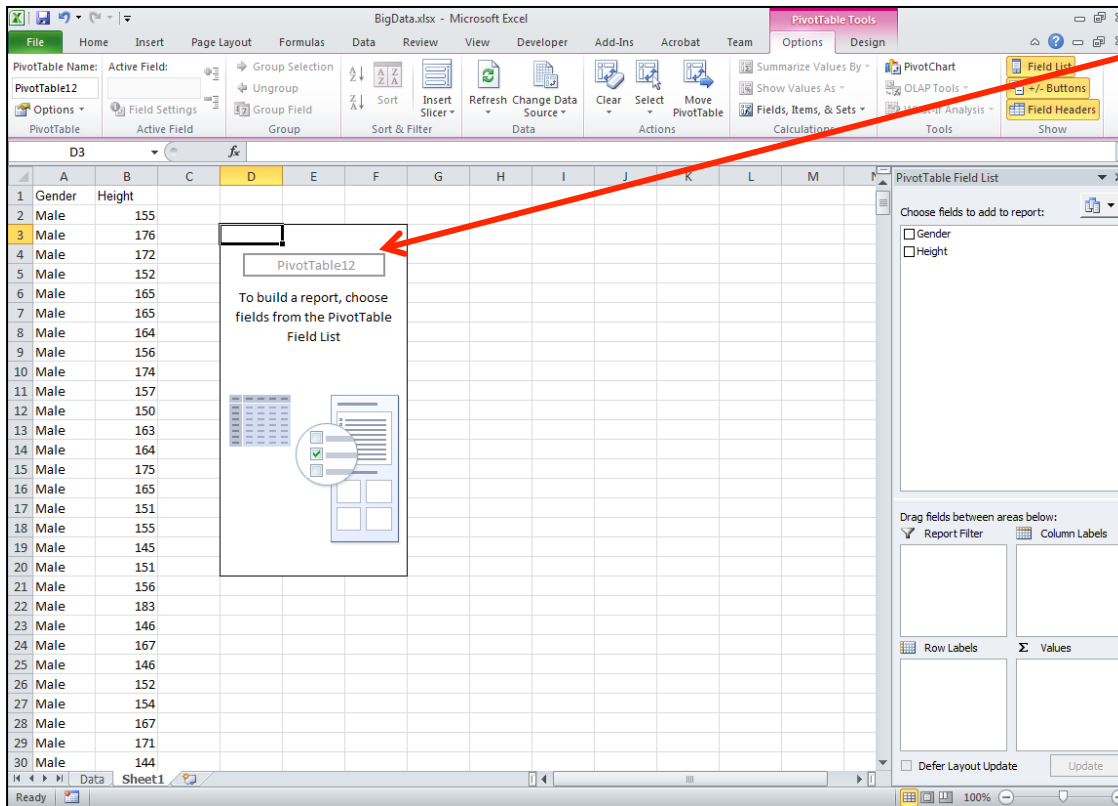
❑ This creates a pivot table on the same sheet as the data

❑ If you have a lot of columns of data you might prefer to create the pivot-table on a new sheet

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield
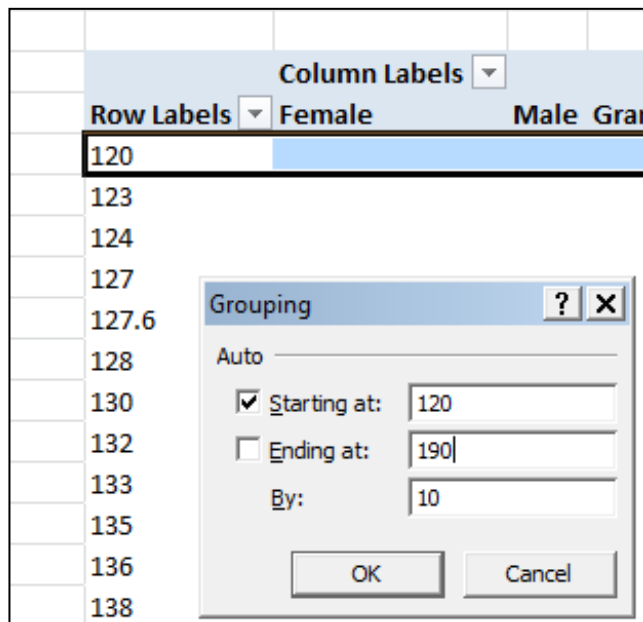
6. Create the row and column labels

In the PivotTables Field List dialogue box on the right hand side, drag and drop *Height* to the **Row Labels** area and *Gender* to the **Column Labels** area

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

7. Group the row labels

❑ Right-click in one of the numbers in the Row Labels area and choose Group
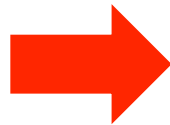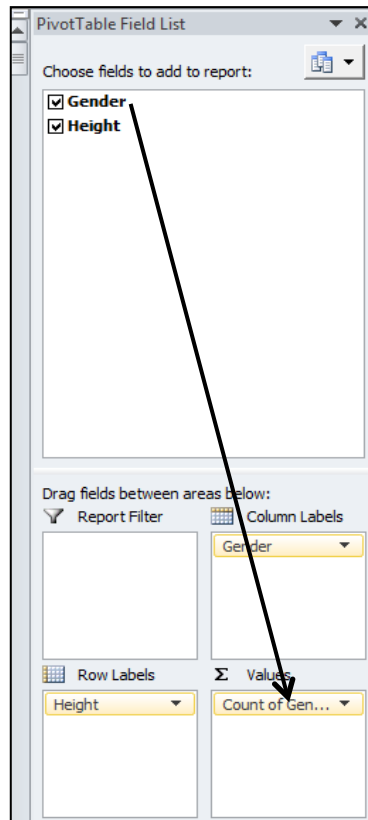
❑ Select 190 for Ending At then click on OK

Note: A Group in Excel means:

   lower limit ≤ x < upper limit

So the Group 120-130 contains values from 120 up to (but not including) 130

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

8. Create the values

Drag Gender to the **Values** area

Change the values to count:
- ❑ Right-click on a cell in the Values area
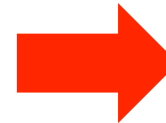- ❑ Choose **Summarize Values By > Count**

9. Change the values format

❑ Right-click on a cell in the Values area

❑ Choose "**Show Values As**"

❑ Change the settings to "% of Column Total"

You can also Right-click and choose "**Number Format**" to change the number of decimal places, etc.

10. Add row and column labels

- ❑ Click inside your pivot-table
- ❑ Choose **Design > Report Layout > Show in Tabular Form**

Count of Gender	Column Labels ▼		
Row Labels ▼	Female	Male	Grand Total
120-130	0.00%	2.20%	1.34%
130-140	2.27%	1.71%	1.93%
140-150	8.71%	9.54%	9.21%
150-160	35.98%	31.05%	32.99%
160-170	42.05%	29.83%	34.62%
170-180	8.33%	19.56%	15.16%
180-190	2.65%	6.11%	4.75%
Grand Total	100.00%	100.00%	100.00%

Count of Gender	Gender ▼		
Height ▼	Female	Male	Grand Total
120-130	0.00%	2.20%	1.34%
130-140	2.27%	1.71%	1.93%
140-150	8.71%	9.54%	9.21%
150-160	35.98%	31.05%	32.99%
160-170	42.05%	29.83%	34.62%
170-180	8.33%	19.56%	15.16%
180-190	2.65%	6.11%	4.75%
Grand Total	100.00%	100.00%	100.00%

www.statstutor.ac.uk

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

11. Rename your sheet and save the file

Double-click on the sheet tab and enter the name *GenderHeight*
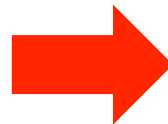
Save the file

www.**stats**tutor.ac.uk

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Activity

On a new sheet, create a pivot table for percentage frequencies of *ArmSpan* against *Gender* where the percentages are calculated as proportions of the column total and the *ArmSpan* values are split into categories of 10cm.

Count of Gender	Gender		
ArmSpan	**Female**	**Male**	**Grand Total**
90-100	3.41%	3.67%	3.57%
100-110	0.38%	3.91%	2.53%
110-120	1.52%	1.47%	1.49%
120-130	3.03%	3.91%	3.57%
130-140	4.55%	5.62%	5.20%
140-150	15.53%	13.94%	14.56%
150-160	31.82%	23.23%	26.60%
160-170	28.79%	22.49%	24.96%
170-180	9.85%	14.18%	12.48%
180-190	1.14%	4.40%	3.12%
190-200	0.00%	1.47%	0.89%
200-210	0.00%	1.22%	0.74%
210-220	0.00%	0.49%	0.30%
Grand Total	**100.00%**	**100.00%**	**100.00%**

Statistical functions: aims

To show you how to:

❑ Use basic statistical functions in Excel

❑ Use Pivot tables to extract the same data

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

# Create a dataset on a new sheet

- ❑ Open the file SurveyData2.xlsx
- ❑ Save it as SurveyData3.xlsx
- ❑ Create another new sheet (it will automatically be called *Sheet2*)
- ❑ Copy and paste the data and headings for *Gender* and *Age* to *Sheet2* as a **dataset**
- ❑ Rename your sheet *GenderAge* and resave your file

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

AVERAGE function

On the sheet *GenderAge*:

1. Type the word "Mean" in cell D2
2. Select cell E2 then type "=average(" then select the cell B2
3. Select **Shift + End + DownArrow** or enter ":" and the address of last data cell in the column: "B674"
4. Type ")" and press the Return key
5. "13.04903" should be then displayed in cell E2. This is the mean *Age* for the whole dataset.
6. Select cell E2 then click on this button repeatedly to reduce the number of decimal places to 2
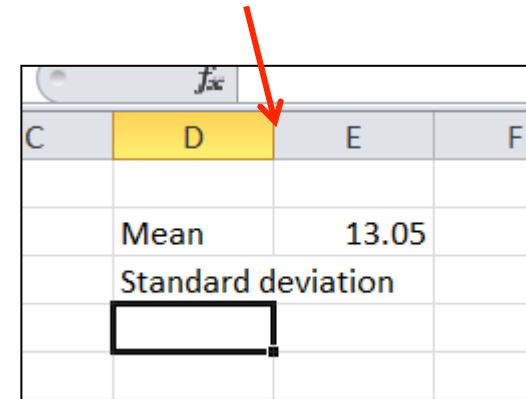
Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield
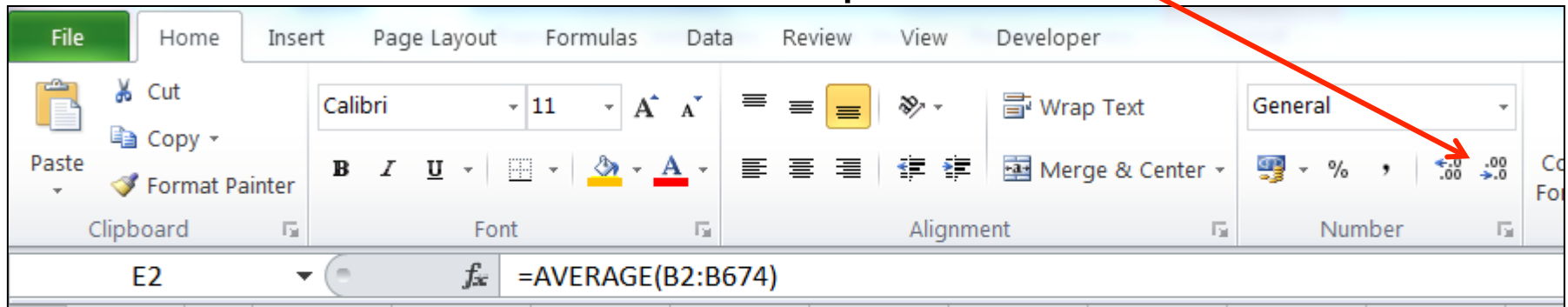
STDEV.S function

On the same sheet:

1. Type "Standard deviation" in cell D3

2. Hover the pointer between the D and E column tabs and double click to auto-increase the width of column D

3. Select the cell E3 and type "=stdev.s("

4. Select the cell B2

5. Select **Shift + End + DownArrow** or enter ":B674"

6. Type ")" and press the Return key

7. "1.070006" should be displayed in cell E3

8. As before, reduce the number of decimal places to 2

This is the standard deviation of the Ages in the dataset assuming that they are a sample from a larger population.

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Activity

On a separate sheet, calculate the mean and standard deviation of *ArmSpan*. Display them to 2 decimal places.

Mean	153.18	
Standard deviation	21.17	

Calculating means and standard deviations using a pivot table

❑ Suppose we want to find the Means and Standard Deviations for the *Ages* of the *Male* and *Female* respondents respectively

❑ We could use formulae (the AVERAGE and STDEV.S functions, in this case), but how would we separate the *Male* and *Female* data?

❑ We could sort the data then select the relevant rows, but a Pivot table is simpler…

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

1. Create a pivot table

❏ Click inside the dataset on the *GenderAge* sheet

❏ Use **Insert > PivotTable** to bring up the PivotTable dialog

❏ Create a Pivot table on the same sheet (select Existing worksheet and the cell D5)

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

2. Calculate the mean *Age* for each value of *Gender*

❑ Drag *Gender* to the **Row Labels** area in the PivotTables field list

❑ Drag *Age* to the **Values** area (it might default to Sum, but we're going to change that…)

❑ Select **Design > Report Layout > Show in Tabular Form**
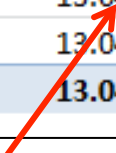
You should get something like this.
It might count *Age* instead.

Gender	Sum of Age
Female	3445
Male	5337
Grand Total	**8782**

- ❑ **Right-click** a data cell in the Age column of the Pivot table

- ❑ Choose "**Summarize Values By…**" and click on "**Average**"

- ❑ You should have something like this:

- ❑ Too much detail!

Gender	Average of Age
Female	13.04924242
Male	13.04889976
Grand Total	13.04903418

- ❑ **Right-click** on one of the "Average of Age" values of the Pivot table and choose "**Number Format…**"

- ❑ Then choose "**Number**"

- ❑ Reduce the number of decimal places to 2

- ❑ You should then have something like this:

Gender	Average of Age
Female	13.05
Male	13.05
Grand Total	13.05

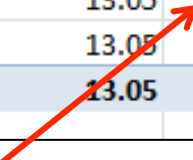# 3. Calculate the standard deviation of *Age* for each value of *Gender*

❑ Drag **Age** to the Values area of PivotTables field list again (again, it might default to Sum)

❑ You should have something like this:

Gender ▼	Average of Age	Sum of Age
Female	13.05	3445
Male	13.05	5337
Grand Total	13.05	8782

❑ **Right-click** inside the "Sum of Age" values area of the Pivot table

❑ Choose "**Summarize Values By…**" and click on "**More Options**"

❑ Scroll down the list and select "**StdDev**"

❑ You should have something like this:

Gender ▼	Average of Age	StdDev of Age
Female	13.05	1.164036259
Male	13.05	1.006135595
Grand Total	13.05	1.070006309

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

- ❑ Too much detail again!
- ❑ **Right-click** in the "**StdDev of Age**" values area of the Pivot table
- ❑ Choose "**Number Format**" and "**Number**"
- ❑ Make sure you have 2 decimal places
- ❑ You should then have something like this:

Gender ▼	Average of Age	StdDev of Age
Female	13.05	1.16
Male	13.05	1.01
Grand Total	**13.05**	**1.07**

- ❑ You can see that the average ages of males and females are the same (to 2 decimal places), but the ages of the female respondents vary a little more than those of the males
- ❑ Save the file

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Activity

On a new sheet, find the means and standard deviations for the number of *TeethFillings* of *Male* and *Female* respondents (to 2 decimal places)

Gender ▼	Average of TeethFillings	StdDev of TeethFillings
Female	1.62	2.54
Male	1.59	3.16
Grand Total	**1.60**	**2.93**

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Recap

We have looked at:

❑ Opening Excel

❑ File handling

❑ Excel interface

❑ Pivot tables for percentage frequencies of category values of a variable

❑ Built-in statistical functions of a single series

❑ Pivot tables for statistical functions of category values of a variable

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield