

# **Session 7.2: Trainee Notes**

## **Introduction to Statistics**

### **Contents**

1. Data types
2. Tables and distributions
3. Entering and importing/exporting data
4. Pivot tables in Excel
5. Data analysis in Excel

### **Learning Objective**

After completing this session, you will be able to:

- Get data in and out of Excel
- Manipulate data in Excel

# 1. Data types

The choice of statistical method is partly determined by the type of data that are to be analysed.

Data can usually be placed into one of two broad categories:

- Quantitative: numerical values. Quantitative variables can be discrete or continuous
- Qualitative: data made up of categories. Qualitative variables can be ordinal or nominal

## Discrete data

- Restricted to a number of distinct integer (whole number) values.
- These are also known as count data since all discrete data can be counted.
- There are situations when we treat discrete data as continuous. If the number of counts is very large, data are generally assumed to be continuous,
  - For example the number of red blood cells.

## Continuous data

- Variables that are measured (measuring is not the same as counting)
- These can take any value within the limits of the precision of the measuring apparatus.
- Examples include weight, height and growth rate.

## Nominal data

- Consist of names or labels
- There is no ordering of the values, even though they may be labelled as numbers.
- For example, there are four blood types A, B, AB, O. There is no natural order to these values.

## Ordinal data

- Contain extra information about the rank (order) of values so that we can make statements such as a is less b.
- Example: questionnaire responses - strongly disagree, disagree, neither, agree or strongly agree.

## 2. Tables and distributions

When you first examine any data you should take a careful look at the data. If this is a large data set then it is impossible to check data individually.

Tables are an excellent way of checking and presenting discrete quantitative data or qualitative data if you have only a few different values.

(Excel) **Insert > Pivot table**

You can present several variables against each other and include totals and percentages.

### Exercise 7.1

Open the 'SAR' worksheet from Copy of Datasets.xls

Prepare a frequency table for sex using **pivot tables**:

1. Drag age into row values
2. Drag Customer ID into values
3. Change metric in values to sum by left clicking on Customer ID and selecting Value field settings

Prepare a frequency table for age using **pivot tables**

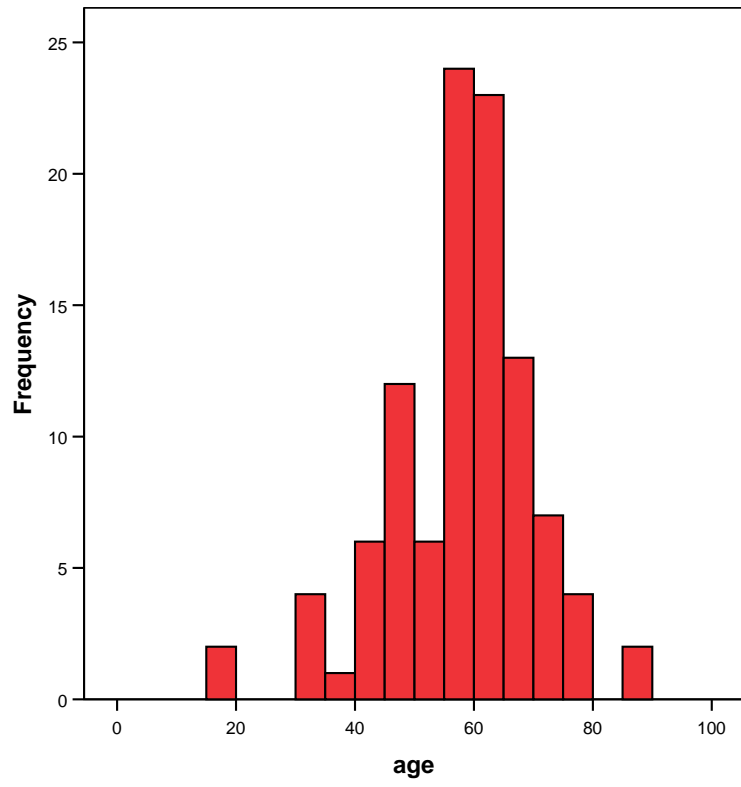
Prepare a frequency table for sex against *agegrp* by putting *sex* in *column values* and *agegrp* in *row values*

Add row percentages to this frequency table

1. Drag Customer ID into values
2. Change metric in values to sum by left clicking on Customer ID and selecting Value field settings > Summarize by
3. Change values to % of row by left clicking on Customer ID and selecting Value field settings > Show values as

Prepare a frequency table with totals and column percentages.

What is the difference between the percentages?



### 3. Entering and Importing/Exporting data

#### Entering new data

To open a new Data editor window

**File → New → Data**

Set up your variables and you can input your data by hand.

Don't forget  
to save as  
you go  
along!

#### *Code your data numerically if possible*

You may need to code string variables e.g. Male = 1, Female = 2 (see Section 4, Variable View, values for labelling your numeric codes).

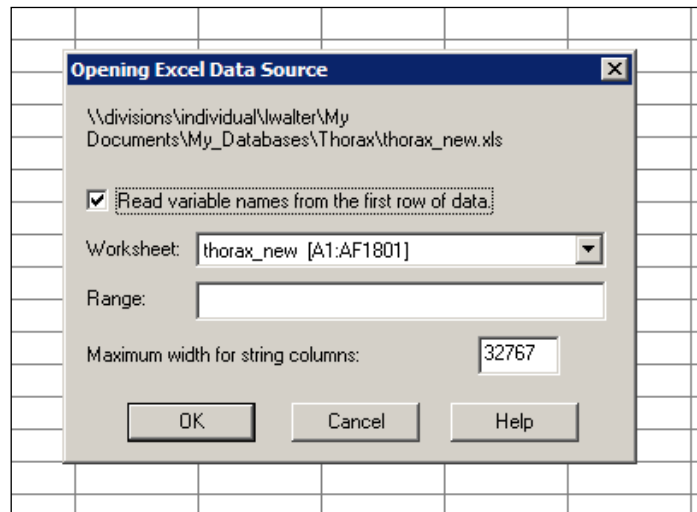
Some string variables, e.g. postcodes, may be necessary – change the type to *String*

#### Importing data from Excel

Close the Excel file you want to import.

**File → Open → Data**

Select Files of Type: Excel (\*.xls) and find your file.



Pick the worksheet you require and check whether the first row is variable names

### Warnings when importing from Excel

- It is only meaningful to import data in a list format
- Excel workbooks must remain closed
- Very long column names will be truncated to 64 characters, spaces omitted
  - Complete original name, with spaces, will be read into variable label
- Formulae in cells will be lost
- Errors may be caused by field names containing special characters, like /

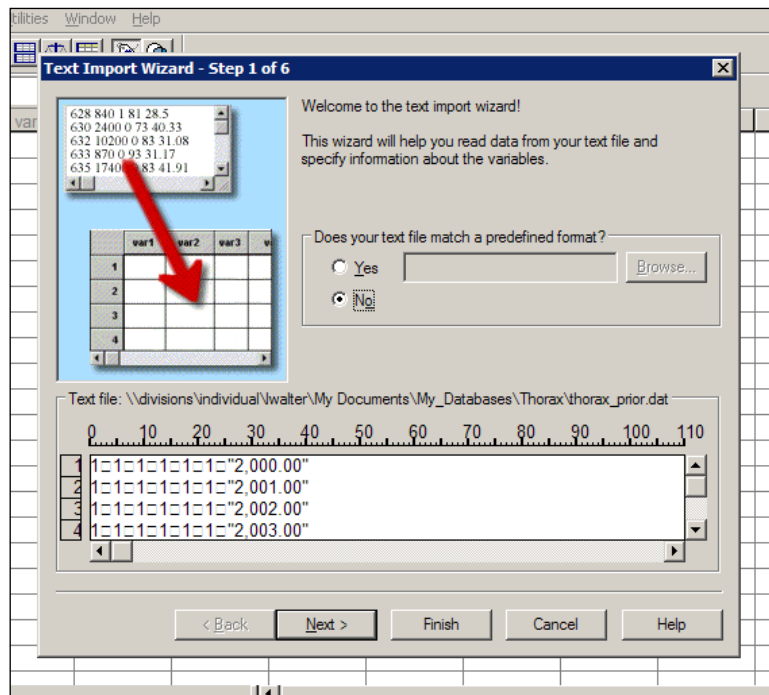
### Importing data from ASCII files

Close the file you want to import.

#### **File → Open → Data**

Select Files of Type: Text (\*.txt) or Data (\*.dat) and find your file.

This opens the text wizard. This takes you through the stages of importing text.



In step 4, choose a delimiter: tab, comma, space, semicolon, other

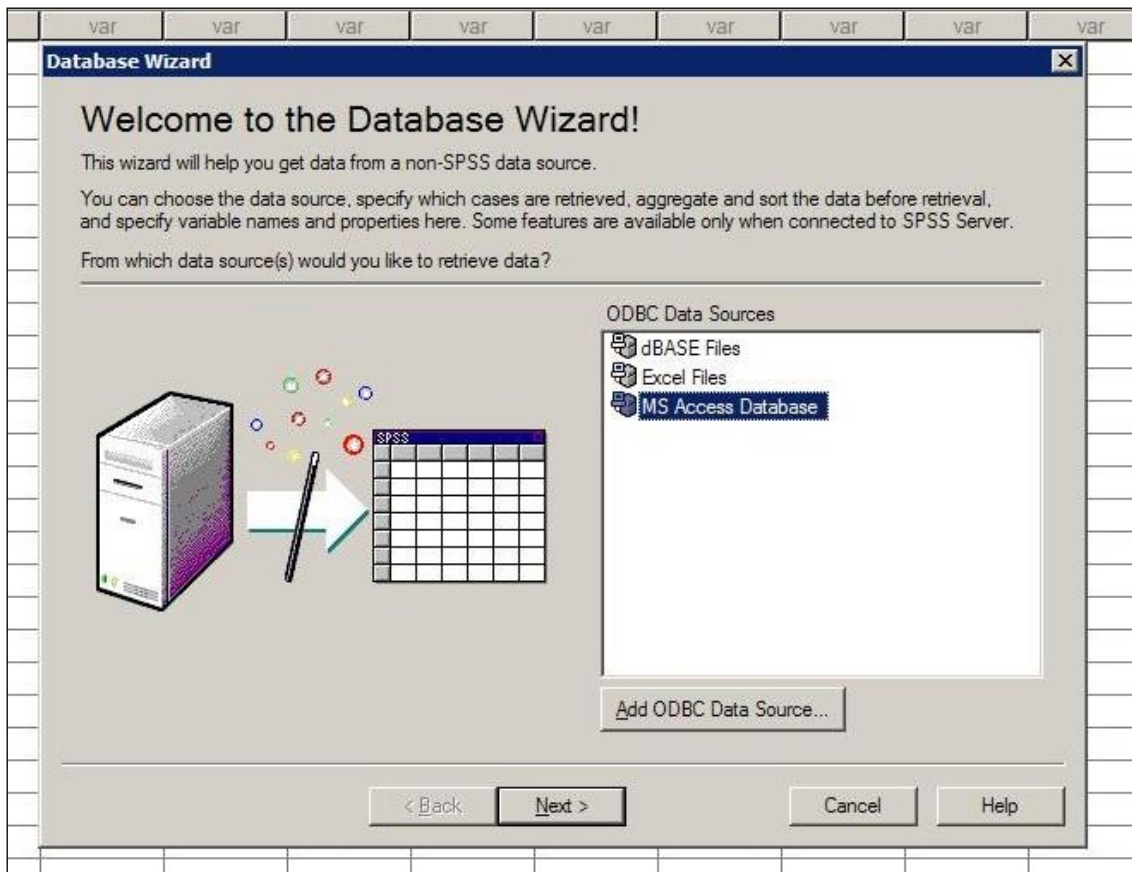
In step 5, choose variables to import and their formats

## Importing data from Access files

Close the database you want to import.

### **File → Open Database → New query**

This opens the database wizard. Select MS Access as the ODBC data source and locate your file.



Select the fields you wish to work with. To select an item, click on it in the 'available list' and, with the mouse button pressed, drag it over to the 'Retrieve Fields' list

Dragging a table selects all of its fields.

### *Features of the ODBC link*

- You can select all columns by dragging a table
- Or select a subset of rows based on conditional expressions [in step 4] such as:
  - specified values of a field or fields
  - results from functions and logical operators

## 4. Pivot tables in Excel

A Pivot Table is an interactive way to quickly summarize large amounts of data. A Pivot Table is especially designed for subtotalling and aggregating numeric data and summarizing data by categories.

### Insert→ Pivot table

Select the range of data (including the titles) and place it in a new worksheet – (NB you need a count variable).

Drag and drop the count variable into values.

Drag and drop the other variable to create the frequency table or cross tabulation you want.

### Exercise 7.2

**Open** the Excel file '*Copy of Datasets.xls*'

**Create** a pivot table for '*cohort data 2*' showing a frequency table of '*Instrg* – number of instructors'


Number of instructors ever had

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	can't remember	1	.0	.0	.0
	none	1	.0	.0	.1
	1	1247	55.3	59.2	59.3
	2	551	24.4	26.1	85.4
	3	204	9.0	9.7	95.1
	4	63	2.8	3.0	98.1
	5	21	.9	1.0	99.1
	6-10	10	.4	.5	99.5
	11-20	9	.4	.4	100.0
	20+	1	.0	.0	100.0
	Total	2108	93.4	100.0	
Missing	System	149	6.6		
Total		2257	100.0		

## 10. Data analysis tool in Excel

Excel has the capability of doing some statistical analysis via an Excel add-in program. For example, t-tests, histograms, ANOVA etc.

If the program has been loaded the **data analysis** command is available under the **data** tab. If not then it can be easily installed:

1. Click the **Office Button** , and then click **Excel Options**.
2. Click **Add-Ins**.
3. In the **Manage box**, click **Excel Add-ins**, and then click **Go**.
4. In the Add-Ins available box, select the **Analysis ToolPak** and the **Analysis ToolPak - VBA** check box, and then click **OK**.

### Exercise 7.3

Install the analysis toolpak if the data analysis command is not available under the data tab

Open the Excel file '*Copy of Datasets*'

Create a histogram for 'age'

**data → data analysis → histogram**