

**B126\_1**

**Data analysis: visualisations in Excel**

**About this free course**

This free course is an adapted extract from the Open University course B126 Business data analytics and decision making - [www.open.ac.uk/courses/modules/b126](https://www.open.ac.uk/courses/modules/b126https://www.open.ac.uk/courses/modules/b126https://www.open.ac.uk/courses/modules/b126?LKCAMPAIGN=ebook_&amp;MEDIA=ou).

This version of the content may include video, images and interactive content that may not be optimised for your device.

You can experience this free course as it was originally designed on OpenLearn, the home of free learning from The Open University –

There you’ll also be able to track your progress via your activity record, which you can use to demonstrate your learning.

First published 2023.

Unless otherwise stated, copyright © 2023 The Open University, all rights reserved.

**Intellectual property**

Unless otherwise stated, this resource is released under the terms of the Creative Commons Licence v4.0 <http://creativecommons.org/licenses/by-nc-sa/4.0/deed.en>. Within that The Open University interprets this licence in the following way: [www.open.edu/openlearn/about-openlearn/frequently-asked-questions-on-openlearn](http://www.open.edu/openlearn/about-openlearn/frequently-asked-questions-on-openlearn). Copyright and rights falling outside the terms of the Creative Commons Licence are retained or controlled by The Open University. Please read the full text before using any of the content.

We believe the primary barrier to accessing high-quality educational experiences is cost, which is why we aim to publish as much free content as possible under an open licence. If it proves difficult to release content under our preferred Creative Commons licence (e.g. because we can’t afford or gain the clearances or find suitable alternatives), we will still release the materials for free under a personal end-user licence.

This is because the learning experience will always be the same high quality offering and that should always be seen as positive – even if at times the licensing is different to Creative Commons.

When using the content you must attribute us (The Open University) (the OU) and any identified author in accordance with the terms of the Creative Commons Licence.

The Acknowledgements section is used to list, amongst other things, third party (Proprietary), licensed content which is not subject to Creative Commons licensing. Proprietary content must be used (retained) intact and in context to the content at all times.

The Acknowledgements section is also used to bring to your attention any other Special Restrictions which may apply to the content. For example there may be times when the Creative Commons Non-Commercial Sharealike licence does not apply to any of the content even if owned by us (The Open University). In these instances, unless stated otherwise, the content may be used for personal and non-commercial use.

We have also identified as Proprietary other material included in the content which is not subject to Creative Commons Licence. These are OU logos, trading names and may extend to certain photographic and video images and sound recordings and any other material as may be brought to your attention.

Unauthorised use of any of the content may constitute a breach of the terms and conditions and/or intellectual property laws.

We reserve the right to alter, amend or bring to an end any terms and conditions provided here without notice.

All rights falling outside the terms of the Creative Commons licence are retained or controlled by The Open University.

Head of Intellectual Property, The Open University

# Contents

* [Introduction](#Session1)
* [Learning outcomes](#Session2)
* [1 Excel spreadsheets](#Session3)
  + [1.1 Using Excel](#Session3_Section1)
  + [1.2 Opening an Excel file](#Session3_Section2)
  + [1.3 Adding the Data Analysis ToolPak in Excel](#Session3_Section3)
  + [1.4 Decimal points and dates](#Session3_Section4)
  + [1.5 Using shortcut keys in Excel](#Session3_Section5)
  + [1.6 Use of Excel spreadsheets](#Session3_Section6)
* [2 Univariate data visualisation](#Session4)
  + [2.1 Frequency tables](#Session4_Section1)
  + [2.2 Types of frequency distribution](#Session4_Section2)
  + [2.3 Histograms: a graphical visualisation of frequency tables](#Session4_Section3)
  + [2.4 Frequency density](#Session4_Section4)
* [3 Bivariate data](#Session5) 
  + [3.1 Contingency tables](#Session5_Section1)
  + [3.2 Scatter diagrams](#Session5_Section2)
* [Conclusion](#Session6)
* [Acknowledgements](#Session7)
* [Solutions](#Solutions1)
* [Descriptions](#Descriptions1)

## Introduction

The objective of this course is to explore the ways to visualise data sets, such as univariate and bivariate data, in Excel and to familiarise yourself with the functions used in Excel to explore the relationship between variables. Univariate data here refers to data that consists of one variable, and bivariate data refers to data that consists of two variables.

Start of Figure



**Figure 1** Making sense of visualising and exploring data

[View description - Figure 1 Making sense of visualising and exploring data](" \l "Session1_Description1)

[View alternative description - Figure 1 Making sense of visualising and exploring data](" \l "Session1_Alternative1)

End of Figure

This course requires Microsoft Excel in some activities. Therefore, this course should be completed on a desktop or laptop rather than a mobile device. If you do not have access to Microsoft Excel, there are various other free options – such as Google Sheets, Apple Numbers or LibreOffice.

This OpenLearn course is an adapted extract from the Open University course [B126 Business data analytics and decision making](https://www.open.ac.uk/courses/modules/b126).

## Learning outcomes

After studying this course, you should be able to:

* explore the functionalities of Excel that are used for problem solving in a business context
* demonstrate the numeracy skills required for gathering and organising data for decision making related to a specific problem
* use graphical techniques (histograms and scatter diagrams) to provide a visual summary of available data
* recognise data presentation and communication techniques used in a range of traditional and electronic media
* describe the relationship between two variables (independent and dependent variables).

## 1 Excel spreadsheets

Before making any decision for a business, it is usually a good idea to get a clear picture of the data, which can provide you with an overview of the relevant information. It is good practice, therefore, to organise and present data in such ways that make it useful for decision making and problem solving. Microsoft Excel is used widely for data analysis in a professional context. Researchers and analysts alike use this tool for various applications in the real world, such as in business, medicine, academia, tax and auditing, marketing, accounting and finance. Moreover, it is flexible enough to be used with many types of data, irrespective of whether it is qualitative or quantitative data.

In this course, you will make extensive use of Excel spreadsheets. In this section you will familiarise yourself with the basics of using Excel. This will enhance your analytical skills as well as your employability skills.

This section briefly explains the various features and functions of Excel that are used by researchers and data analysts to explore, organise and analyse data.

Start of Box

**Excel for OU students**

If you are currently studying with the OU as a fee-paying student, you have free access to Microsoft Office 365. This includes the spreadsheet software Excel. For this you need to go to the OU Computing Guide and scroll down to ‘Microsoft Office 365’. Click on the link. If you have not done so, then you should follow the instructions to sign up to get access for your free version of the software. If you have already installed Microsoft Excel on your laptop, then you may prefer to use your own version. Although earlier versions of Excel are not significantly different, the layout of some tabs and menus may vary slightly.

End of Box

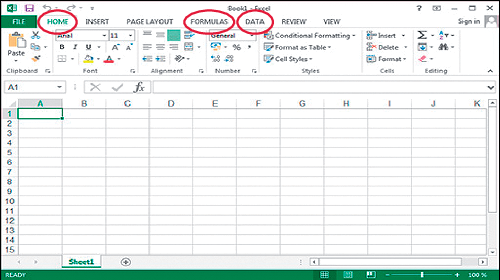
## 1.1 Using Excel

You can run Excel by double-clicking on the Excel icon on your desktop or laptop (or select **Excel 2013** or your version from your list of programs).

Excel will open with a clean, new worksheet called **Book 1** that contains only one worksheet (**Sheet1**). You can add more sheets by clicking on the plus sign at the bottom of the worksheet or spreadsheet in case one sheet becomes too small or too cluttered. This will help you to organise and manage your large sets of data, or a variety of data.

Add a new worksheet by clicking on the plus sign. The new sheet **Sheet2** looks exactly like **Sheet1**. Each individual cell is labelled according to the **A1** reference style. This means that columns are labelled with letters from A to XFD and rows with number 1 to 1048576. You will be using this style, as opposed to the **R1C1** reference style, which labels both rows and columns with numbers. However, it is sometimes useful to think of a value or formula as being in a cell with a specific row number and column number.

Start of Figure



**Figure 2** An empty worksheet in Excel

[View description - Figure 2 An empty worksheet in Excel](" \l "Session3_Description1)

[View alternative description - Figure 2 An empty worksheet in Excel](" \l "Session3_Alternative1)

End of Figure

The latest versions of Excel, including Excel 2013, are organised using a **ribbon** interface, which consists of a collection of icons for each tab. The screenshot above in Figure 2 shows the **Home** ribbon with several icons for general editing such as font size, text alignment or cell styles. In this section, you will use the **Insert**, **Formulas** and **Data** tabs in Excel. You may click on each tab and familiarise yourself with the large main icons.

You should not be worried if you are not very familiar with Excel, or do not know how to use the various icons, tabs and functions yet. The aim of this course is to gradually build up your familiarity with the software.

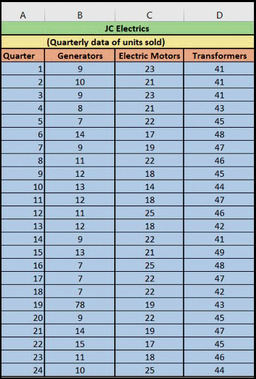
## 1.2 Opening an Excel file

Excel stores the data in files that contain one or multiple worksheets. When you open a worksheet by double-clicking the Excel icon, it will usually display the data contained in the last active sheet. For this course, you will be using a pre-populated worksheet for JC Electrics, a company manufacturing heavy machinery, such as generators, transformers and electric motors. Download the file at the link below. You can then save it to your computer so you can return to it at different points in the course. The file will also be linked at future points in the course.

* Download the file [**JC Electrics**](http://www.open.edu/openlearn/ocw/mod/oucontent/olinkremote.php?website=B126_1&targetdoc=JC%20Electrics%20spreadsheet%20-%20OpenLearn).
* Click on **File**, then **Open** and navigate to the folder in which you saved this file and open the file by double-clicking it.

By opening the worksheet, you should see quarterly data of units sold by JC Electrics in four columns. The screenshot below shows a spreadsheet with these columns labelled as: ‘Quarter’, ‘Generators’, ‘Transformers’ and ‘Electric Motors’.

Start of Figure



**Figure 3** Data in Excel

[View description - Figure 3 Data in Excel](" \l "Session3_Description2)

[View alternative description - Figure 3 Data in Excel](" \l "Session3_Alternative2)

End of Figure

## 1.3 Adding the Data Analysis ToolPak in Excel

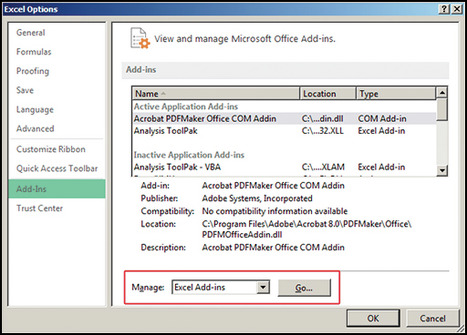
In this section, you are going to learn how to add the Data Analysis ToolPak in Excel and what its purpose is. If you are using another spreadsheet software, you may skip this section. Alternatively, similar functionality is usually available within the software, and all tasks within the Excel ToolPak can be carried out using other functions in Excel as well.

The Data Analysis ToolPak is a Microsoft Excel add-in function that provides you with a set of statistical tools for analysing data efficiently and effectively. This add-in comes with Excel during installation; however, if it has not already been loaded or activated on the machine you are using, you need to load or activate it before you can use it.

To load the Data Analysis ToolPak:

* Open the Excel worksheet.
* Click on the **File** tab, then click on **Options**, and select Add-Ins.
* In the **Manage** box, select Excel **Add-ins** as shown in the screenshot below and click **Go**.
* In the **Add-Ins** dialog box, select the **Analysis Tool Pak** check box, and then click **OK**.

Start of Figure



**Figure 4** Activating the Data Analysis ToolPak

[View description - Figure 4 Activating the Data Analysis ToolPak](" \l "Session3_Description3)

[View alternative description - Figure 4 Activating the Data Analysis ToolPak](" \l "Session3_Alternative3)

End of Figure

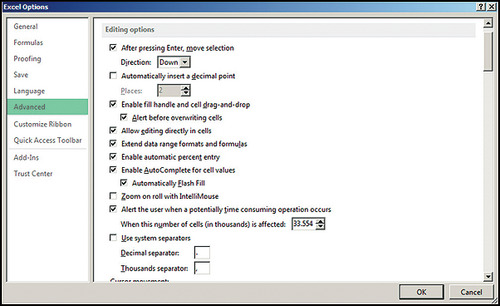
The Data Analysis icon should now be available on the **Data** ribbon in the **Analysis** group.

## 1.4 Decimal points and dates

If you are using a version of Excel other than the English (UK) one, you may have noticed the different date format and decimal point. In addition, one of the Excel’s habits of automatically adjusting the format of cell contents can sometimes produce unwanted or incorrect results. Therefore, it is useful to learn how things can be corrected and adjusted in Excel.

In your Excel spreadsheet, click on **File,** select **Options** and then select the **Advanced tab**; the results of this are shown in Figure 5.

Start of Figure



**Figure 5** Advanced options in Excel

[View description - Figure 5 Advanced options in Excel](" \l "Session3_Description4)

[View alternative description - Figure 5 Advanced options in Excel](" \l "Session3_Alternative4)

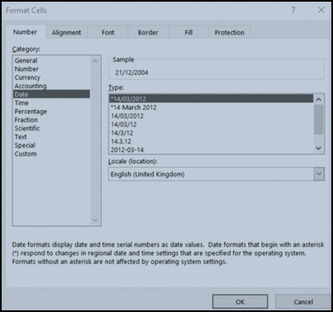
End of Figure

If you untick the box ‘Use system separators’, you can enter alternative symbols to use instead.

You can adjust the date format by selecting a cell that contains a date, then clicking on **Format** in the **Home** ribbon under **Cells** and **Format Cells …** or by right-clicking on the cells and selecting **Format Cells …** from the context menu.

Figure 6 below shows you the ‘Format’ dialog for an English version of Excel installed on a computer. You can produce the desired date format; for example, **DD/MM/YYYY** by selecting the option and then clicking **OK**.

Start of Figure



**Figure 6** Changing the date format

[View description - Figure 6 Changing the date format](" \l "Session3_Description5)

[View alternative description - Figure 6 Changing the date format](" \l "Session3_Alternative5)

End of Figure

## 1.5 Using shortcut keys in Excel

One of the most beneficial things in Excel is being able to control the user interface without using the mouse. This will substantially speed up your Excel projects, especially when working under pressure in a professional environment.

Some of the most useful shortcuts are listed in the tables below. Try some of these out while reading the list.

Start of Table

Table 1   Navigating inside and between worksheets

|  |  |
| --- | --- |
| Arrow keys | Move around the spreadsheet. |
| Page Down/Page Up | Move screen down or up. |
| Ctrl + Arrow keys | Move to the edge of a region. This is useful for navigating large blocks of data, particularly with the Ctrl + Shift + Arrow selection functionality. |
| Home | Move to the beginning of a row. |
| Ctrl + Home | Move to the beginning of a worksheet. This is most useful if you have multiple worksheets and want to prepare a nice-looking workbook, by cycling through all worksheets pressing Ctrl + Page Down and Ctrl + Home for each sheet, which quickly puts the cursor in the upper-left corner. |
| Ctrl + F | Display the ‘Find’ dialog box. |
| Ctrl + H | Display the ‘Find and Replace’ dialog box. |
| Ctrl + Tab | Set focus on next workbook if multiple workbooks are open. |

End of Table

Start of Table

Table 2   Data selection

|  |  |
| --- | --- |
| Shift + Space | Select the entire row at the cursor position. |
| Ctrl + Space | Select the entire column at the cursor position. |
| Ctrl + A | Select the entire worksheet or the data-containing area. Press Ctrl + A a second time to select the entire worksheet. |
| Ctrl + Shift + Page Up | Select the current and previous sheet in a workbook. This is useful if you have similar worksheets and want to edit cells in all of them at the same time. |
| Shift + Arrow key | Extend the selection by one cell. This is one of the most useful shortcuts. |
| Ctrl + Shift + Arrow key | Extend the selection to the last cell with content in row or column. You can do this with the Page Up/Down keys. |
| Shift + F8 | Add another range to the selected range of cells. |
| Esc | Cancel selection. |

End of Table

Start of Table

Table 3    Editing

|  |  |
| --- | --- |
| Ctrl + C | Copy active selection. |
| Ctrl + X | Cut active selection. Think carefully about whether you want to copy or cut a selection before pasting in each situation, because cell references in copied selections will point to other cells and not the original references when pasted. |
| Ctrl + V | Paste active selection. |
| Ctrl + Z | Undo last action. |
| Ctrl + Y | Redo last action. |
| F2 | Edit current cell. |
| F4 | Repeat last formatting action. |
| Alt + Enter | Start a new line in the same cell when entering text. |
| Ctrl + D | Copy above cell down. |
| Ctrl + ‘+’ | Insert row/column. |
| Ctrl + ‘-’ | Delete row/column. |
| Ctrl + 1 | Show the ‘Format cells’ dialog. |
| Shift + F11 | Insert new worksheet. |

End of Table

Start of Table

Table 4   Formulas and special functions

|  |  |
| --- | --- |
| Ctrl + Shift + Enter | Enter an array formula. Must have a range selected first. (This is shown here only for reference and will be explained later.) |
| Shift + F3 | Display the ‘Insert Function’ dialog. |
| F4 | When editing a cell reference (e.g. ‘H5’), pressing F4 makes this reference absolute (e.g., ‘$H$5’). Pressing F4 repeatedly makes only row or column absolute. |
| F9 | Force re-calculation of worksheets. It can be used to calculate part of a formula, when selecting part of formula and pressing F9, this evaluates the selected part. |
| Shift + F9 | Calculate the active worksheet. |
| F12 | Display the ‘Save As’ dialog. |
| Ctrl + S | Save the current workbook. Extremely useful for the occasional power outage or computer crash. |
| Ctrl + F1 | Minimise or show the ribbon. |

End of Table

## 1.6 Use of Excel spreadsheets

Excel spreadsheets are used to manage, organise and present data in a systematic way. They are particularly useful when there is a specific relationship between results in different cells. They can be used in many ways in different fields; for example, in finance, they are mostly used to present and analyse data such as accounting transactions (e.g. sales, payments, cost, forecasting and budgeting). Spreadsheets are also used to design templates of financial statements and survey results.

Some other uses are listed below:

* In schools and universities, spreadsheets are often used to manage student data in areas such as their grade performance, attendance or personal biography.
* In hospitals, spreadsheets are used to manage patient data, like their personal information, details of their illness or details of the medicines they use.
* Data is often exported from more complicated computer systems, such as manufacturing, financial or marketing systems, to allow managers or analysts to manipulate the data once it has been created, and to carry out forecasts, simulations and ‘what-if’ exercises.

Excel also has many formatting options (borders, colour highlighting), to allow you to draw attention to aspects of your figures. One of the more advanced features is conditional formatting, in which Excel automatically assigns distinct colours to cells according to their value (e.g. red for negative, green for positive and appropriate shades in between).

Start of Activity

**Activity 1  Test your Excel knowledge**

Allow approximately 15 minutes to complete this activity

Start of Question

Choose an answer to each of the questions.You can check your answer to each question as you go.

Which of the following functions counts all cells?

End of Question

a)   SUMIF

b)   COUNTIF

c)   AVERAGE

d)   COUNT

[View answer - Part](" \l "Session3_Interaction1)

Start of Question

What is the result of the following formula?

4 + 3\*10

To enter the formula, click on an empty cell and type ‘=4+3\*10’.

End of Question

a)   24

b)   34

c)   30

[View answer - Part](" \l "Session3_Interaction2)

Start of Question

What is the shortcut key to copy a cell in an Excel spreadsheet?

End of Question

a)   Ctrl + C

b)   Ctrl + F

c)   Shift + F3

[View answer - Part](" \l "Session3_Interaction3)

Start of Question

What is the shortcut key to save the work in an Excel spreadsheet?

End of Question

a)   Ctrl + V

b)   Ctrl + C

c)   Ctrl + S

[View answer - Part](" \l "Session3_Interaction4)

Start of Question

What is the shortcut key in the keyboard to edit the formula or text?

End of Question

a)   F9

b)   F2

c)   Ctrl + F

[View answer - Part](" \l "Session3_Interaction5)

Start of Question

What is the shortcut key to cancel the selection in the sheet or cell?

End of Question

a)   Ctrl + Alt + Delete

b)   Esc

c)   F12

[View answer - Part](" \l "Session3_Interaction6)

Start of Question

What is the shortcut key to insert a new worksheet in an Excel file?

End of Question

a)   Ctrl + Z

b)   Ctrl + V

c)   Alt + Enter

d)   Shift + F11

[View answer - Part](" \l "Session3_Interaction7)

Start of Question

What is the shortcut key to display the insert function dialog?

End of Question

a)   Shift + F3

b)   Ctrl + Z

c)   Shift + F2

[View answer - Part](" \l "Session3_Interaction8)

End of Activity

In the next section, you will learn how to present and summarise a univariate dataset in a table and graphical form.

## 2 Univariate data visualisation

In practice, there are two main ways to visualise and summarise data in Excel. These are:

* tabular form
* graphical form.

While presenting and summarising data in Excel, it is important to know the features of data. If your data is **univariate** – that is, the data consists of many observations for only one variable – then you can either use a frequency table or a histogram to summarise the data and get an idea of its features. However, if your data is **bivariate** – that is, the data consists of two variables (an independent variable and a dependent variable) – and you need to know the relationship between these two variables, then you can use either a contingency table or scatter diagram to summarise the data and get an idea of its structure. You will learn about bivariate data visualisation later in the course.

The next section will briefly explain frequency tables.

## 2.1 Frequency tables

Before learning how to make frequency tables and histograms in Excel, you first need to know what a frequency distribution is, and why we need histograms.

The frequency of a value is the number of times that value appears in a data set. A frequency distribution table displays the pattern of frequencies of a variable in a tabular form. It gives the information of how many times each value of a variable occurs in a data set. A frequency distribution table is an effective way to summarise and organise the collected raw data so that all its features are summarised in a table form. The first step that a researcher or analyst must do with collected raw data is to organise and present the data in such a way that makes it meaningful and easy to digest.

Frequency distribution tables are also called frequency tables, and in practice both terms are used interchangeably. In short, a frequency table gives you a snapshot of how your data is distributed and spread out.

A frequency distribution table has two columns: Column A and Column B. Column A presents the outcome of the values and Column B presents the frequency of the outcomes. You can understand this better with the example below.

Start of Example

Anna is an analyst, and she works Monday and Tuesday in a hospital. On Wednesday and Thursday, she works in a small accounting firm. On Friday, Saturday and Sunday she works in a bank.

Now you can display this data through a frequency distribution table, as shown in Figure 7.

Start of Figure



**Figure 7** Displaying the data in the frequency distribution table

[View description - Figure 7 Displaying the data in the frequency distribution table](" \l "Session4_Description1)

[View alternative description - Figure 7 Displaying the data in the frequency distribution table](" \l "Session4_Alternative1)

End of Figure

This table gives you a clear idea of how many days Anna works in each different organisation.

End of Example

In the next section, you will learn about various types of frequency distribution table.

## 2.2 Types of frequency distribution

There are four types of frequency distribution table:

* ungrouped frequency distribution
* grouped frequency distribution
* relative frequency distribution
* cumulative frequency distribution.

Before describing each type of frequency distribution table, you need to know the difference between **ungrouped data** and **grouped data**.

In simple terms, **ungrouped data** is raw data that has not been categorised. For example, a manager in a firm knows that 100 employees work in their firm; this is raw data because it does not tell you how many employees work in each department (e.g. production and sales). However, if you have raw data that is categorised, it is defined as **grouped data**. For example, if this manager knows that 50 employees work in production and 50 employees work in sales, it means that the data is organised in such a way that it provides a clear indication of how many employees work in each department.

### 2.2.1 Concepts involved in frequency tables

The following terms are frequently used in frequency distribution:

**Class interval or class limit**: the lowest and the highest value defined for a class or group are called class limits. The lowest value is called the lower-class limit and the highest value is called the upper-class limit of that class. In the example in Table 5, the lower-class limits are 7, 9, 11, 13, 15, and the upper limits are 8, 10, 12, 14, 16. The terms class and class interval are often used interchangeably, although the class interval is a symbol for the class.

**Class boundaries**: a class boundary is the number that is used to separate the two different classes. It is the midpoint between the upper limit of a class and the lower limit of the next class. Each class has both an upper and a lower limit boundary. The lower boundary of a class is calculated by subtracting half of the value of the interval from the lower-class limit, while the upper boundary of a class is calculated by adding half of the value of the interval to the upper-class limit.

Referring to the example from JC Electrics, its class boundaries are given in Table 5.

Start of Table

Table 5   Class intervals and boundaries for JC Electrics

|  |  |
| --- | --- |
| **Class intervals** | **Class boundaries** |
| 7–8 | 6.5–8.5 |
| 9–10 | 8.5–10.5 |
| 11–12 | 10.5–12.5 |
| 13–14 | 12.5–14.5 |
| 15–16 | 14.5–16.5 |

End of Table

Referring to Table 5, you can say that the lower limit of the first-class interval is 6.5, as all values between 6.5 and 7.5 are recorded as 7. Meanwhile, the upper-class limit of 8 is 8.5, as all values between 7.5 and 8.5 are recorded as 8. The real class limit of a class is called a class boundary. A class boundary is obtained by adding two successive class limits and dividing the sum by 2. The value so obtained is taken as the upper-class boundary for the previous class, and lower-class boundary for the next class.

**Midpoint or class mark**: this is the average of a class interval, and is obtained by dividing the sum of upper- and lower-class limits by 2. Thus, the class mark of the interval 7–8 is 7.5, as (7+8)/2=7.5.

**The size or the width of a class interval**: the size, or width, of a class interval is the difference between the lower- and upper-class boundaries and is also referred to as the class width, class size, or class length. If all class intervals of a frequency distribution have equal widths, this common width is denoted by c.

**Range**: this is the difference between the maximum value and the minimum value of the data set. For example, in the JC Electrics data set the maximum number of Electric Motors sold has a value of 25, while the minimum is 14. Hence, to calculate the range, you must calculate 25–14=9.

### 2.2.2 Ungrouped frequency distribution tables

When you are summarising small amounts of data, then it is better to organise and represent it in an ungrouped frequency distribution table. This is a type of distribution that shows how many times each individual value occurs in a data set; they are usually used to calculate the accurate frequency of individual data values.

For example, say you are interested to know how many plants people have in their homes. A survey gives the following figures as number of plants that 18 different people have in their homes:

Number of plants = 1, 5, 2, 2, 3, 3, 5, 5, 1, 1, 1, 3, 4, 4, 2, 3, 3, 3

To answer your question, first you need to see the frequency of each value in the data. Value 1 occur 4 times, so you can describe it as 4 people having 1 plant. Then you do the same for the rest of the values, so: 3 people have 2 plants, 6 people have 3 plants, 2 people have 4 plants, and 3 people have 5 plants.

In the following activity, you will learn how to make ungrouped frequency tables in Excel.

Start of Activity

**Activity 2  How to make an ungrouped frequency table in Excel**

Allow approximately 30 minutes to complete this activity

Start of Question

In this activity, you will learn how to make an ungrouped frequency table in Excel. Once you have produced the ungrouped frequency table in Excel, you may need to compare it with the final output by clicking ‘Reveal discussion’. This will help you to see whether you have produced the accurate ungrouped frequency table or not.

Watch Video 1, which gives on how to create a frequency table, or follow the instructions below.

Start of Media Content

Video content is not available in this format.

**Video 1**

[View transcript - Video 1](" \l "Session4_Transcript1)

Start of Figure



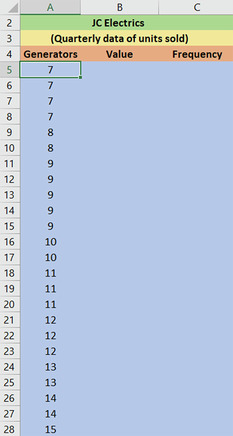
[View alternative description - Uncaptioned Figure](" \l "Session4_Alternative2)

End of Figure

End of Media Content

* Open the Excel file [**JC Electrics**](http://www.open.edu/openlearn/ocw/mod/oucontent/olinkremote.php?website=B126_1&targetdoc=JC%20Electrics%20spreadsheet%20-%20OpenLearn). This file contains the quarterly data of number of generators sold. Make sure that the data is arranged in columns.
* Copy the data containing the number of generators sold to **Column A** of a new worksheet.
* Label **Column B** ‘Value’ and label **Column C** ‘Frequency’.
* Find the minimum and maximum value in the data. In this example: **=MAX(A5:A28)**, which is 15, and **=MIN(A5:A28)**, which is 7
* Calculate the range: **(MAX – MIN)**, so 15 minus 7 equals 8.
* Arrange the values in Column A in ascending order. Select the values **(A5:A28)** in Column A, click **Data** in the toolbar and then click **Sort**, select **Continue with current selection** and **press Enter**. Figure 8 shows how your information should be displayed.

Start of Figure



**Figure 8** Arranging the data from ascending to descending order

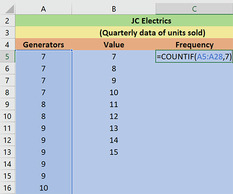
[View description - Figure 8 Arranging the data from ascending to descending order](" \l "Session4_Description2)

[View alternative description - Figure 8 Arranging the data from ascending to descending order](" \l "Session4_Alternative4)

End of Figure

* To count the number of quarters in which 7 units were sold, you need to calculate the frequency in Column C. Type **=COUNTIF (Range, value)**. For example, **=COUNTIF (A5:A28,7)**

Start of Figure



**Figure 9** Calculating the frequency of ungroup data

[View description - Figure 9 Calculating the frequency of ungroup data](" \l "Session4_Description3)

[View alternative description - Figure 9 Calculating the frequency of ungroup data](" \l "Session4_Alternative5)

End of Figure

* You should now save your file as you will return to this ungrouped frequency table in a later activity.

End of Question

[View discussion - Activity 2  How to make an ungrouped frequency table in Excel](" \l "Session4_Discussion1)

End of Activity

As is mentioned above, ungrouped frequency tables are useful when you have a small set of data and you want to easily observe the frequency of each value in the data set. However, if you have a large data set then a grouped frequency distribution table is the best option; you will learn about these in the next section.

### 2.2.3 Grouped frequency distribution tables

When you are summarising large amounts of raw data, it is useful to represent the data in groups. The groups are commonly known as classes or class intervals. You might then want to determine the number of values belonging to each class or class interval; this is called class frequency. A tabular arrangement of data by class together with the corresponding class frequencies is called a grouped frequency distribution table. This is a more efficient way to find the trends within the data, but there is a possibility that the grouping process may sacrifice much of the original detail of the data.

In the following activity, you will learn how to make a grouped frequency table in Excel.

Start of Activity

**Activity 3  How to make a grouped frequency distribution table in Excel**

Allow approximately 30 minutes to complete this activity

Start of Question

In this activity, you need to produce a grouped frequency table in Excel either by watching the screencast in Video 2 or by following the instructions given below. Once you have produced the grouped frequency table in Excel, you can check your answer by clicking ‘Reveal discussion’.

Start of Media Content

Video content is not available in this format.

**Video 2**

[View transcript - Video 2](" \l "Session4_Transcript2)

Start of Figure



[View alternative description - Uncaptioned Figure](" \l "Session4_Alternative7)

End of Figure

End of Media Content

* Open the Excel file [**JC Electrics**](http://www.open.edu/openlearn/ocw/mod/oucontent/olinkremote.php?website=B126_1&targetdoc=JC%20Electrics%20spreadsheet%20-%20OpenLearn). This file contains quarterly data of the number of generators sold. Make sure that the data is arranged in columns.
* Find the range which is the difference between the maximum and minimum value in the data set. You can do this either by entering the formula **=MAX (A2:A25)-MIN (A2:A25)**, or by simply using the results you have calculated in Column H as, **=H10-H11** (see Figure 11).

Start of Figure



**Figure 11** Calculating the range

[View description - Figure 11 Calculating the range](" \l "Session4_Description5)

[View alternative description - Figure 11 Calculating the range](" \l "Session4_Alternative8)

End of Figure

* Decide the class interval width. There are no firm rules on how to choose the width. However, the following formula is the most common method to calculate the width:

Start of $1

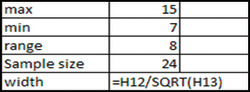
width equals range over square root of sample size

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative9)

End of $1

* You can round this value to a whole number or a number that is convenient to add (such as multiple of 10). For example, the width calculated in the given data set is 1.6, so will be taken as **2** (see Figure 12).

Start of Figure



**Figure 12** Calculating the width

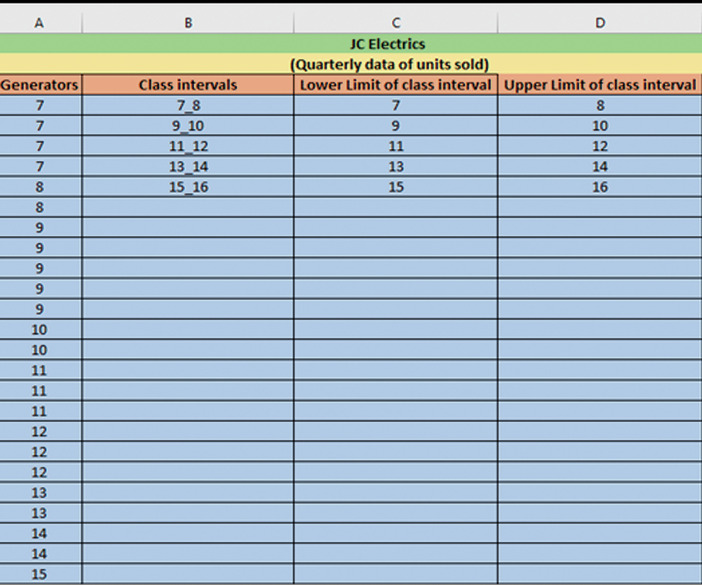
[View description - Figure 12 Calculating the width](" \l "Session4_Description6)

[View alternative description - Figure 12 Calculating the width](" \l "Session4_Alternative10)

End of Figure

* Decide the number of groups or class intervals into which data is to be distributed. Each class interval is defined by a lower limit and an upper limit. The lower limit of first-class interval is the lowest value in the data set. Add the class interval width to find the upper limit of the first interval and the lower limit of the next class interval. Keep adding the interval width to calculate more class intervals until you exceed the highest value. For example, in the given data set, you determined the class intervals width equals 2, so you should make the class intervals as 7–8, 9–10, 11–12, 13–14, 15–16.
* This means that the first-class interval has lower limit of 7 and upper limit of 8. See Figure 13 below.

Start of Figure



**Figure 13** Maing the class intervals

[View description - Figure 13 Maing the class intervals](" \l "Session4_Description7)

[View alternative description - Figure 13 Maing the class intervals](" \l "Session4_Alternative11)

End of Figure

* The next step is to calculate the frequency. Select the range **E2:E6** and enter FREQUENCY function as shown in the Figure 14 in the discussion.
* Press **CTRL + SHIFT + ENTER** to submit the FREQUENCY formula above as an array formula. If it is entered correctly, you would see a formula wrapped in curly braces {}.
* You should now save your file as you will return to this grouped frequency table in a later activity.

End of Question

[View discussion - Activity 3  How to make a grouped frequency distribution table in Excel](" \l "Session4_Discussion2)

End of Activity

As mentioned above, a grouped frequency table is the best option to visualise the frequency of values in a large data set. However, if you are interested to know the proportion of a particular value in relation to the total number of values in the data set, then a relative frequency distribution table is the better option. In the next section, you will learn how to produce a relative frequency table in Excel.

### 2.2.4 Relative frequency distribution tables

Relative frequency distribution is another type of frequency distribution. This type of distribution tells you the proportion of each value or class interval of a variable. In other words, relative frequency distribution describes the number of times a particular value occurs in relation to the total number of values. You can use this type of frequency distribution for any type of variable when you are more interested in comparing frequencies than the actual number of observations.

For example, Team A has won 6 football games from a total of 12 football games played. The frequency of winning is 6 and the relative frequency of winning is 50% (i.e. 60/12=0.5).

You will learn how to make a relative frequency table in Excel in the following activity.

Start of Activity

**Activity 4  How to make a relative frequency table in Excel**

Allow approximately 30 minutes to complete this activity

Start of Question

In this activity, you will build a relative frequency table using the ungrouped frequency distribution table from Activity 2. Once you have made the relative frequency distribution table in Excel, check your answer by clicking ‘Reveal discussion’ below.

The ungrouped frequency distribution table consists of three columns. Column A is labelled ‘Generators’, Column B is labelled ‘Value’ and Column C is labelled ‘Frequency’. Add a fourth column to the table for the relative frequencies.

To calculate the relative frequencies, you need to divide each frequency by the sample size (frequency / sample size). You can calculate the sample size by taking the sum of all the frequencies in Column C, which is 24.

End of Question

[View discussion - Activity 4  How to make a relative frequency table in Excel](" \l "Session4_Discussion3)

End of Activity

In the next section, you will learn how to make cumulative frequency distribution tables in Excel.

### 2.2.5 Cumulative frequency distribution tables

Cumulative frequency distribution is the fourth type of frequency distribution table. It is the sum of the frequencies less than or equal to each value or class interval of a variable. This type of frequency distribution can be used for **ordinal** or **quantitative variables**, especially when you want to understand how often observations fall below certain values.

For example, Company A sells 250 books in the first week, 150 books in the second week and 400 books in the third week. The cumulative number of books sold in the second week by Company A is 400 books (250 books in the first week + 150 books in the second week). The cumulative number in the third week is 800 books (250 books in the first week + 150 books in the second week + 400 books in the third week).

In the following activity, you will learn how to build a cumulative frequency distribution table in Excel.

Start of Activity

**Activity 5  How to make a cumulative frequency distribution table in Excel**

Allow approximately 30 minutes to complete this activity

Start of Question

In this activity, you will build a cumulative frequency distribution table using the grouped frequency distribution table in Activity 3. Once you have built the cumulative frequency distribution table, you can check your answer by clicking ‘Reveal discussion’ below.

Borrow the grouped frequency distribution table from Activity 3. This table consists of five columns. Column A is labelled Generators, Column B is labelled Class intervals, Column C is labelled Lower limit of class interval, Column D is labelled Upper limit of class interval and Column E is labelled Frequency.

Add another column, Column F, to the table for the cumulative frequency. The cumulative frequency is calculated by adding each frequency from a frequency distribution table to the sum of its predecessors. The last value will always be equal to the total for all observations, since all frequencies will already have been added to the previous total.

End of Question

[View discussion - Activity 5  How to make a cumulative frequency distribution table in Excel](" \l "Session4_Discussion4)

End of Activity

In the next section, you will learn how to visualise these tables by drawing histograms in Excel.

## 2.3 Histograms: a graphical visualisation of frequency tables

A histogram is a popular visualisation tool to summarise the distribution of continuous data. In a histogram, the variable is divided into intervals called ‘bins’. You then count the number of observations in each bin and plot the resulting table in a bar chart. The horizontal x-axis displays the ‘bins’ and the vertical y-axis displays the number of observations in each bin. Histograms can help you to see whether the data is clustered around certain values or whether there are many small or many large values. A typical histogram in Excel looks like the bar chart below. Note that the values on the x-axis show the upper limit of the interval. In a proper histogram, there are no spaces or gaps between the bars.

Start of Figure

A histogram in Excel

**Figure 17** A typical histogram in Excel

[View description - Figure 17 A typical histogram in Excel](" \l "Session4_Description11)

[View alternative description - Figure 17 A typical histogram in Excel](" \l "Session4_Alternative15)

End of Figure

In the following activity, you will learn how to plot a histogram in Excel.

Start of Activity

**Activity 6 Using Excel to draw a histogram**

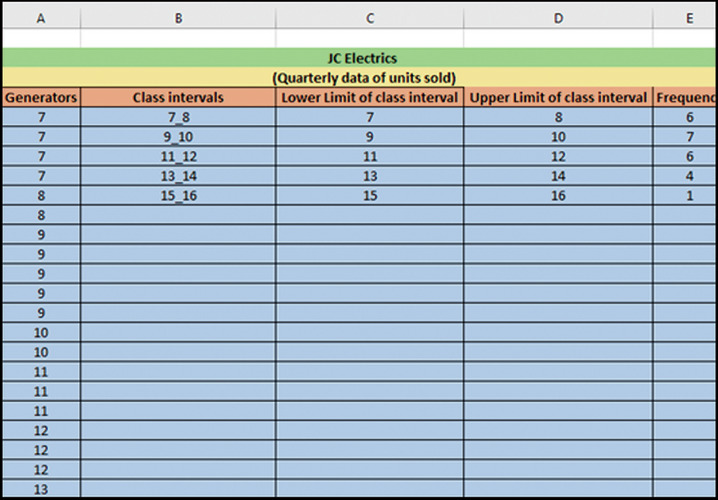
Allow approximately 35 minutes to complete this activity

Start of Question

In this activity, you will learn how to produce a histogram in Excel by following the instructions that are given below. Once you have produced the histogram in Excel, you can check your answer by clicking ‘Reveal discussion’ below.

* Open the Excel file called [**JC Electrics**](http://www.open.edu/openlearn/ocw/mod/oucontent/olinkremote.php?website=B126_1&targetdoc=JC%20Electrics%20spreadsheet%20-%20OpenLearn), which contains the quarterly data of number of generators sold. The third column C contains information about the number of generators sold in each quarter of the year.
* Find the minimum and maximum value in the data set. You can obtain them through the min (range) and max (range) functions in Excel. Type **=MAX(A5:A28)** into cell L10 and **=MIN(A5:A28)** into cell L11. This will give the minimum and maximum values of the data set, which are 7 and 15.
* Next, you need to specify a range of intervals (often called ‘bins’) for which to count the number of observations that fall into each bin. The maximum value is 15 and the minimum value is 7, so you can make the class intervals 7–8, 9–10, 11–12, 13–14, 14–15, 15–16 etc. This means that the first class has the lower value 7 and the maximum value 8 and so on. See Columns C and D in the worksheet in Figure 18.

Start of Figure



**Figure 18** A frequency distribution table in Excel

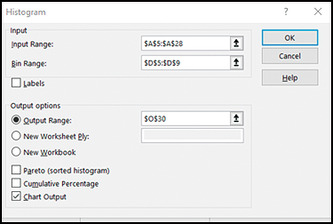
[View description - Figure 18 A frequency distribution table in Excel](" \l "Session4_Description12)

[View alternative description - Figure 18 A frequency distribution table in Excel](" \l "Session4_Alternative16)

End of Figure

* There are many ways to calculate the width of the bin in Excel. One of the easiest ways to calculate it is as the width of the bin or class intervals (sample size / range), which is 3 (i.e. 24/8=3). In this example, the bin width is 2.
* Click on ‘Data Analysis’ in the ‘Data’ ribbon. This will bring up a list of some of the statistical analyses that you can perform in Excel.
* Select ‘Histogram’ and click ‘OK’.
* Specify the input range as A5:A28 and the bin range as D5:D9
* Tick the box ‘Chart Output’ and specify the output location as H5, as shown in Figure 19 below.

Start of Figure



**Figure 19** Histogram dialog box in Excel

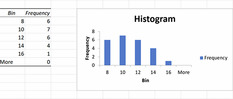
[View description - Figure 19 Histogram dialog box in Excel](" \l "Session4_Description13)

[View alternative description - Figure 19 Histogram dialog box in Excel](" \l "Session4_Alternative17)

End of Figure

Click ‘OK’. Excel will put the histogram next to your frequency table.

Start of Figure



**Figure 20** Histogram of number of units sold (Generators)

[View description - Figure 20 Histogram of number of units sold (Generators)](" \l "Session4_Description14)

[View alternative description - Figure 20 Histogram of number of units sold (Generators)](" \l "Session4_Alternative18)

End of Figure

* To remove the space between the bars, right click a bar, click **Format Data Series**, and change the Gap Width to 0%.
* To add borders, right click a bar, click **Format Data Series**, click the **Fill & Line** icon, click **Border**, and select a colour.
* Now click ‘Reveal discussion’ to compare what you have made against the answer.

End of Question

[View discussion - Activity 6 Using Excel to draw a histogram](" \l "Session4_Discussion5)

End of Activity

## 2.4 Frequency density

Frequency density is defined as the frequency per unit of the data in each class. Frequency density is calculated by dividing the frequency by the class width (the class width is the difference between the upper limit of the class interval and the lower limit of the class interval). Frequency density allows for a meaningful comparison of different classes where the class width may not be equal.

Start of $1

Frequency density equals frequency over class width

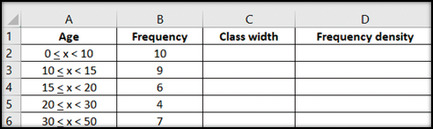
[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative20)

End of $1

The frequency density **gives the ratio of the frequency of a class to its width**. Frequency density is used to plot a frequency density histogram; here, you plot frequency density instead of frequency on the y-axis. Frequency density gives you the total area of bars and tells you about the frequency in the histogram (rather than the height).

You can calculate frequency density when you have a set of grouped data that consists of unequal widths of class intervals. For example, see the following Excel worksheet in Figure 22, which shows information about the ages of a group of people playing football.

Start of Figure



**Figure 22** Information about the age of people playing football in an Excel file

[View description - Figure 22 Information about the age of people playing football in an Excel file](" \l "Session4_Description16)

[View alternative description - Figure 22 Information about the age of people playing football in an Excel file](" \l "Session4_Alternative21)

End of Figure

To calculate the frequency densities:

* In Column C, find the class width of the class intervals by finding the difference of upper and lower bounds/limits. (For example, 10 minus zero equals 10 comma 15 minus 10 equals five, and so on.)

Start of Figure



**Figure 23** Calculating the class width

[View description - Figure 23 Calculating the class width](" \l "Session4_Description17)

[View alternative description - Figure 23 Calculating the class width](" \l "Session4_Alternative28)

End of Figure

* Then, in Column D, divide the frequency of each class interval by its width.

Start of Figure



**Figure 24** Calculating the frequency density

[View description - Figure 24 Calculating the frequency density](" \l "Session4_Description18)

[View alternative description - Figure 24 Calculating the frequency density](" \l "Session4_Alternative34)

End of Figure

In the activity below, you will test your knowledge of the difference between a frequency density histogram and a frequency histogram.

Start of Activity

**Activity 7 The difference between frequency histograms and frequency density histograms**

Allow approximately 15 minutes to complete this activity

Start of Question

Watch the video below and note down in the box the difference between frequency histograms and frequency density histograms.

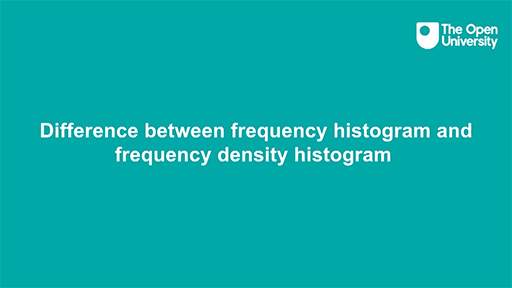
Start of Media Content

Video content is not available in this format.

**Video 3**

[View transcript - Video 3](" \l "Session4_Transcript3)

Start of Figure



[View alternative description - Uncaptioned Figure](" \l "Session4_Alternative40)

End of Figure

End of Media Content

End of Question

*Provide your answer...*

End of Activity

## 3 Bivariate data

Bivariate data refers to an instance in which two separate variables are examined and compared. For example, a performance manager may be interested to know how well employees perform their work, that is, to measure the efficiency of the employees. In this example, the performance manager may examine two variables: the number of tasks they complete and the quality of the tasks.

Bivariate data is collected to explore the relationship between two variables and then use this relationship to inform future decisions. One of the main aims of the researcher is to find out whether changes in one variable may be caused by changes in another variable. This type of research involves two basic types of variables: independent variables and dependent variables.

**Independent variables**: an independent variable is **one that stands alone and is not changed by the other variable you are trying to measure**. The researcher changes the independent variable to see the effect it will have on the dependent variable.

**Dependent variables**: a dependent variable is **the one that changes because of independent variable manipulation**. It is the outcome you are interested in measuring, and it ‘depends’ on your independent variable. In statistics, dependent variables are also called response variables (as they respond to a change in another variable).

For example, say a researcher is interested to know whether mature students’ performance in a maths class changes based on the time of the class. To answer this question, the researcher measures mature students’ performance in a morning class and an evening class. The study finds that mature students perform better in the evening class than in the morning class.

What are the independent and dependent variables in the example above? The independent variable is the time of the class, and the dependent variable is mature students’ performance in maths, as it might change in relation to the independent variable.

In the next activity, you will expand your knowledge of bivariate data.

Start of Activity

**Activity 8 Bivariate data**

Allow approximately 15 minutes to complete this activity

Start of Question

Watch the following video and make notes on bivariate data in the free response box below.

Start of Media Content

Video content is not available in this format.

**Video 4**

[View transcript - Video 4](" \l "Session5_Transcript1)

Start of Figure



[View alternative description - Uncaptioned Figure](" \l "Session5_Alternative1)

End of Figure

End of Media Content

End of Question

*Provide your answer...*

End of Activity

Bivariate data can be visualised using contingency tables and scatter diagrams. In the next section you will learn about contingency tables.

## 3.1 Contingency tables

A contingency table is called a cross table or two-way table and counts observations for each unique combination of values in two variables. It is a table of data in which the row entries tabulate the data according to one variable and the column entries tabulate the data according to another variable.

The tool best suited for a data set will depend on the variable’s scale of measurement. The most important distinction here is whether a variable is discrete or continuous. If it is discrete, a convenient method to summarise bivariate data is a contingency table. If it is continuous, a scatter diagram can be more useful in visualising the data set.

Contingency tables are used in statistics to understand the relationship between categorical variables. For example, say you want to summarise the following sample of firms in Table 7 regarding their industry sector and size.

Start of Table

Table 7 Sample of firms

|  |  |  |
| --- | --- | --- |
| **Firm** | **Sector** | **Employees** |
| 1 | Technology | <50 |
| 2 | Food | 50+ |
| 3 | Technology | <50 |
| 4 | Food | <50 |
| 5 | Food | <50 |
| 6 | Food | 50+ |
| 7 | Technology | <50 |
| 8 | Technology | <50 |
| 9 | Technology | 50+ |
| 10 | Food | 50+ |

End of Table

See the following cross table (Table 8), which summarises the information of the sample data. It counts the number of firms for each combination of sector and number of employees.

Start of Table

Table 8 Cross table for sample firms

|  |  |  |  |
| --- | --- | --- | --- |
| **Sector** | **<50 Employees** | **50+ Employees** | **Total employees** |
| Technology | 4 | 0 | 4 |
| Food | 2 | 4 | 6 |
| Total | 6 | 4 | 10 |

End of Table

The cross table shows that there are two firms in the food manufacturing sector that have less than 50 employees. However, the cell 50+ shows that there are four firms in the food manufacturing sector that have more than 50 employees. The sum of the total food manufacturing firm is six which is the 60% of the grand total.

Contingency tables vary in size and type because the size of the contingency table depends on the sample size and number of observations.

There is no formula to draw a contingency table in Excel. However, analysts use a PivotTable to build contingency tables. A PivotTable is considered a powerful statistical tool to summarise bivariate and multivariate data sets in an Excel spreadsheet or database table and obtain the desired report. This tool does not actually change the spreadsheet or database itself; it simply pivots or turns the data to view it from different perspectives. Researchers and analysts use PivotTables especially when they have large amounts of data that would be time consuming to calculate by hand. A PivotTable can perform a few data processing functions such as identifying sums, averages, ranges or outliers. It then arranges this information in a simple and meaningful way that draws attention to key values. If you would like to experiment with PivotTables, go to the ‘Insert’ ribbon in Excel and select ‘PivotTable’.

## 3.2 Scatter diagrams

A scatter diagram is another way to visualise a quantitative bivariate data set. This is a two-dimensional diagram or graph with one variable on the x-axis (the independent variable) and the other variable on the y-axis (the dependent variable). You can then plot the corresponding point on the diagram.

In the next activity, you will produce a scatter diagram in Excel either by following the video or the instructions provided.

Start of Activity

**Activity 9 Drawing scatter diagrams**

Allow approximately 35 minutes to complete this activity

Start of Question

The screencast in Video 5 gives you instructions on how to draw scatter plots in Excel.

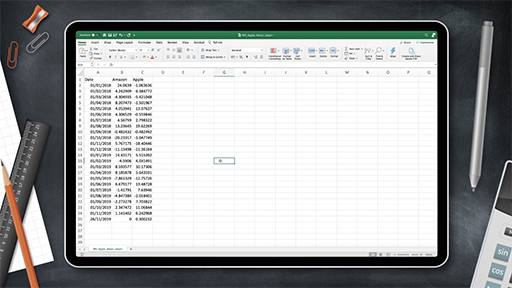
Start of Media Content

Video content is not available in this format.

**Video 5**

[View transcript - Video 5](" \l "Session5_Transcript2)

Start of Figure



[View alternative description - Uncaptioned Figure](" \l "Session5_Alternative2)

End of Figure

End of Media Content

Look at the following example, which shows a data set relating to the temperature on several days in June, and the number of Pepsi drinks sold in a small shop.

Start of Table

Table 9   Temperature and number of drinks sold

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Temperature (X) | 12 | 14 | 15 | 17 | 22 | 13 | 20 | 23 |
| Pepsi (Y) | 12 | 16 | 16 | 19 | 32 | 10 | 24 | 40 |

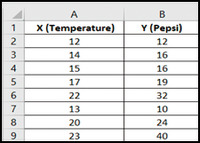
End of Table

Here, temperature is the independent variable and number of drinks is the dependent variable, as the sale of Pepsi is affected by changes in the temperature. Hence, you will plot temperature on the x-axis and drinks on the y-axis.

To find out if there is any relationship between variable X (Temperature) and variable Y (Pepsi), execute the following steps in Excel.

* Select the range A1:B9

Start of Figure



**Figure 25** Spreadsheet of Pepsi sold and Temperature

[View description - Figure 25 Spreadsheet of Pepsi sold and Temperature](" \l "Session5_Description1)

[View alternative description - Figure 25 Spreadsheet of Pepsi sold and Temperature](" \l "Session5_Alternative3)

End of Figure

* On the **Insert** tab, in the **Charts** group, click the **Scatter** symbol.

Start of Figure



**Figure 26** How to select the Scatter symbol

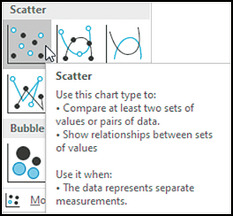
[View description - Figure 26 How to select the Scatter symbol](" \l "Session5_Description2)

[View alternative description - Figure 26 How to select the Scatter symbol](" \l "Session5_Alternative4)

End of Figure

* This will open a drop-down menu showing various types of scatter plots. The standard type is the one with unconnected dots in the top left. Click the **Scatter** symbol to insert this chart.

Start of Figure



**Figure 27** How to select the Scatter chart type

[View description - Figure 27 How to select the Scatter chart type](" \l "Session5_Description3)

[View alternative description - Figure 27 How to select the Scatter chart type](" \l "Session5_Alternative5)

End of Figure

End of Question

[View discussion - Activity 9 Drawing scatter diagrams](" \l "Session5_Discussion1)

End of Activity

## Conclusion

In this course, you have started to familiarise yourself with the spreadsheet software Excel, which is widely used in workplaces, and useful in many different fields and contexts: for example, in business, medicine, marketing, tax and auditing, accounting and finance.

You have also studied the basics of data analysis. The focus here was on the several ways to visualise and summarise data using tools available in Microsoft Excel, such as frequency tables, histograms, and scatter diagrams or plots. The main objective of data analysis and statistical modelling is to help make more evidence-based decisions. The various data visualisation tools studied in this course are only the first step toward starting the decision-making process using data.

The next step could be to study descriptive statistics, which gets you closer to a comprehensive analysis of the data. You could then become more confident when examining and summarising data and using Excel tools such as measures of location and measures of dispersion to numerically analyse data.

A second OpenLearn course on data analysis, [Data analysis: hypothesis testing](https://www.open.edu/openlearn/science-maths-technology/data-analysis-hypothesis-testing/content-section-0), is now also available should you wish to take your studies further.

This OpenLearn course is an adapted extract from the Open University course [B126 Business data analytics and decision making.](https://www.open.ac.uk/courses/modules/b126)

## Acknowledgements

This free course was written by Henry Lahr.

Except for third party materials and otherwise stated (see [terms and conditions](http://www.open.ac.uk/conditions)), this content is made available under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 Licence](http://creativecommons.org/licenses/by-nc-sa/4.0/deed.en).

The material acknowledged below is Proprietary and used under licence (not subject to Creative Commons Licence). Grateful acknowledgement is made to the following sources for permission to reproduce material in this free course:

Course image: ismagilov / iStock / Getty Images Plus

Figure 1: jadamprostore / iStock / Getty Images Plus

Figures 2–28: © Excel

Every effort has been made to contact copyright owners. If any have been inadvertently overlooked, the publishers will be pleased to make the necessary arrangements at the first opportunity.

**Don't miss out**

If reading this text has inspired you to learn more, you may be interested in joining the millions of people who discover our free learning resources and qualifications by visiting The Open University – [www.open.edu/openlearn/free-courses](http://www.open.edu/openlearn/free-courses?LKCAMPAIGN=ebook_&MEDIA=ol).

## Solutions

## Activity 1  Test your Excel knowledge

### Part

#### Answer

**Right:**

d)   COUNT

**Wrong:**

a)   SUMIF

b)   COUNTIF

c)   AVERAGE

[Back to - Part](" \l "Session3_Part1)

### Part

#### Answer

**Right:**

b)   34

**Wrong:**

a)   24

c)   30

[Back to - Part](" \l "Session3_Part2)

### Part

#### Answer

**Right:**

a)   Ctrl + C

**Wrong:**

b)   Ctrl + F

c)   Shift + F3

[Back to - Part](" \l "Session3_Part3)

### Part

#### Answer

**Right:**

c)   Ctrl + S

**Wrong:**

a)   Ctrl + V

b)   Ctrl + C

[Back to - Part](" \l "Session3_Part4)

### Part

#### Answer

**Right:**

b)   F2

**Wrong:**

a)   F9

c)   Ctrl + F

[Back to - Part](" \l "Session3_Part5)

### Part

#### Answer

**Right:**

b)   Esc

**Wrong:**

a)   Ctrl + Alt + Delete

c)   F12

[Back to - Part](" \l "Session3_Part6)

### Part

#### Answer

**Right:**

d)   Shift + F11

**Wrong:**

a)   Ctrl + Z

b)   Ctrl + V

c)   Alt + Enter

[Back to - Part](" \l "Session3_Part7)

### Part

#### Answer

**Right:**

a)   Shift + F3

**Wrong:**

b)   Ctrl + Z

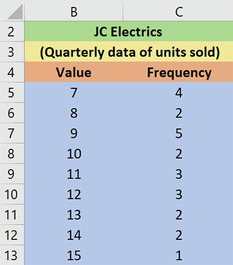
c)   Shift + F2

[Back to - Part](" \l "Session3_Part8)

## Activity 2  How to make an ungrouped frequency table in Excel

#### Discussion

Start of Figure



**Figure 10** Ungrouped frequency table

[View description - Figure 10 Ungrouped frequency table](" \l "Session4_Description4)

[View alternative description - Figure 10 Ungrouped frequency table](" \l "Session4_Alternative6)

End of Figure

Figure 10 above shows the completed frequency table. The same data is shown in Table 6 below.

Start of Table

Table 6 Ungrouped frequency table

|  |  |
| --- | --- |
| **Value** | **Frequency** |
| 7 | 4 |
| 8 | 2 |
| 9 | 5 |
| 10 | 2 |
| 11 | 3 |
| 12 | 3 |
| 13 | 2 |
| 14 | 2 |
| 15 | 1 |

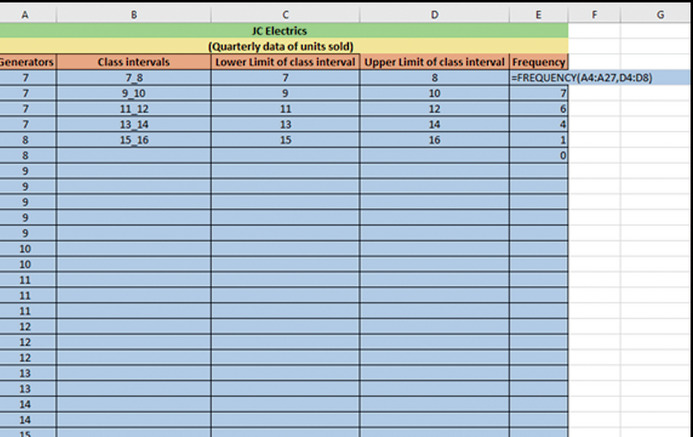
End of Table

[Back to - Activity 2  How to make an ungrouped frequency table in Excel](" \l "Session4_Activity1)

## Activity 3  How to make a grouped frequency distribution table in Excel

#### Discussion

Start of Figure



**Figure 14** Calculating the frequency by using an array formula

[View description - Figure 14 Calculating the frequency by using an array formula](" \l "Session4_Description8)

[View alternative description - Figure 14 Calculating the frequency by using an array formula](" \l "Session4_Alternative12)

End of Figure

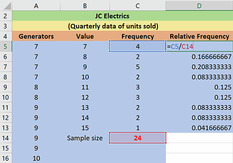
Figure 14 shows the results of using the array formula to calculate frequency. It is important to be aware that any error entered may result in an incorrect grouped frequency table and provide false information about the business.

[Back to - Activity 3  How to make a grouped frequency distribution table in Excel](" \l "Session4_Activity2)

## Activity 4  How to make a relative frequency table in Excel

#### Discussion

Start of Figure



**Figure 15**  Calculating the relative frequency of ungrouped data

[View description - Figure 15  Calculating the relative frequency of ungrouped data](" \l "Session4_Description9)

[View alternative description - Figure 15  Calculating the relative frequency of ungrouped data](" \l "Session4_Alternative13)

End of Figure

Figure 15 represents the quarterly data of number of units of generators sold. Column A (‘Generators’) contains values between 7 to 15. Column B is labelled ‘Value’. Column C is labelled ‘Frequency’.

To calculate the relative frequency in Column D, you need to divide each frequency by sample size. The sample size of 24 can be found by summing the ‘Frequency’ column.

[Back to - Activity 4  How to make a relative frequency table in Excel](" \l "Session4_Activity3)

## Activity 5  How to make a cumulative frequency distribution table in Excel

#### Discussion

Start of Figure



**Figure 16** Calculating the cumulative frequency

[View description - Figure 16 Calculating the cumulative frequency](" \l "Session4_Description10)

[View alternative description - Figure 16 Calculating the cumulative frequency](" \l "Session4_Alternative14)

End of Figure

Figure 16 presents the data of the number of generators sold. Column A (‘Generators’) shows the values between 7 to 15 in ascending order. Column B shows the class intervals. Column C shows the values of the lower limit of each class intervals. Column D shows the values of upper limit of each class intervals. Column E shows the frequency.

To calculate the cumulative frequency in Column F, add each frequency to the frequencies in the previous rows. If you do it correctly, the value in the last row will be equal to the sample size.

[Back to - Activity 5  How to make a cumulative frequency distribution table in Excel](" \l "Session4_Activity4)

## Activity 6 Using Excel to draw a histogram

#### Discussion

Start of Figure

A histogram in Excel

**Figure 21**  Histogram showing units sold of generators

[View description - Figure 21  Histogram showing units sold of generators](" \l "Session4_Description15)

[View alternative description - Figure 21  Histogram showing units sold of generators](" \l "Session4_Alternative19)

End of Figure

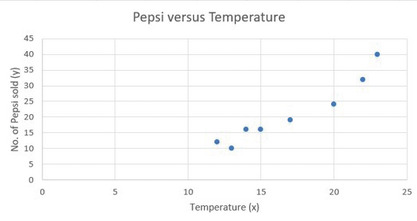
Working through the steps given above should return the completed histogram shown in Figure 21.

[Back to - Activity 6 Using Excel to draw a histogram](" \l "Session4_Activity5)

## Activity 9 Drawing scatter diagrams

#### Discussion

Start of Figure



**Figure 28** Temperature versus Pepsi sales

[View description - Figure 28 Temperature versus Pepsi sales](" \l "Session5_Description4)

[View alternative description - Figure 28 Temperature versus Pepsi sales](" \l "Session5_Alternative6)

End of Figure

The final image shows the positive relationship between the number of Pepsis sold and the temperature. This shows that as the temperature increases, sales of Pepsi also increase.

[Back to - Activity 9 Drawing scatter diagrams](" \l "Session5_Activity2)

## Descriptions

### Figure 1 Making sense of visualising and exploring data

A graphical illustration of different types of charts (e.g. pie chart, line graph).

[Back to - Figure 1 Making sense of visualising and exploring data](" \l "Session1_Figure1)

### Figure 1 Making sense of visualising and exploring data

A colourful graphical illustration of different types of charts (e.g. pie chart, line graph). In the foreground is a histogram with a magnifying glass hovering over a section of the graph.

[Back to - Figure 1 Making sense of visualising and exploring data](#Session1_Figure1)

### Figure 2 An empty worksheet in Excel

A blank Excel worksheet with the different editing tools at the top

[Back to - Figure 2 An empty worksheet in Excel](" \l "Session3_Figure1)

### Figure 2 An empty worksheet in Excel

A blank Excel worksheet with the different editing tools at the top; it shows columns labelled A to K and rows numbered 1 to 15. The tabs running along the very top of the sheet are: File, Home, Insert, Page layout, Formulas, Data, Review and View. Home, Formulas and Data are circled in red.

[Back to - Figure 2 An empty worksheet in Excel](#Session3_Figure1)

### Figure 3 Data in Excel

A spreadsheet that displays the quarterly data of units sold for JC Electrics

[Back to - Figure 3 Data in Excel](" \l "Session3_Figure2)

### Figure 3 Data in Excel

A screenshot of a spreadsheet that displays the quarterly data of units sold for JC Electrics, which manufactures three products: generators, electric motors and transformers. This spreadsheet consists of four columns and each column is labelled as: Quarter, Generators, Electric Motors, and Transformer.

[Back to - Figure 3 Data in Excel](#Session3_Figure2)

### Figure 4 Activating the Data Analysis ToolPak

‘Add Ins’ dialogue box

[Back to - Figure 4 Activating the Data Analysis ToolPak](" \l "Session3_Figure3)

### Figure 4 Activating the Data Analysis ToolPak

This figure shows the ‘Add Ins’ dialogue box to activate the ‘DataAnalysis Toolpack’ in Excel.

[Back to - Figure 4 Activating the Data Analysis ToolPak](#Session3_Figure3)

### Figure 5 Advanced options in Excel

Screenshot showing how to select the advanced option in Excel

[Back to - Figure 5 Advanced options in Excel](" \l "Session3_Figure4)

### Figure 5 Advanced options in Excel

A screenshot showing how to select the advanced option in Excel to format the date.

[Back to - Figure 5 Advanced options in Excel](#Session3_Figure4)

### Figure 6 Changing the date format

Format Cells window in Excel

[Back to - Figure 6 Changing the date format](" \l "Session3_Figure5)

### Figure 6 Changing the date format

Figure 6 shows the Format Cells window in Excel. The Number tab is selected; under the Category menu on the left, Date is highlighted, and on the right is the Type menu showing various different date formatting options. At the bottom of the window are the OK and Cancel buttons.

[Back to - Figure 6 Changing the date format](#Session3_Figure5)

### Figure 7 Displaying the data in the frequency distribution table

Table displaying the data in the frequency distribution table

[Back to - Figure 7 Displaying the data in the frequency distribution table](" \l "Session4_Figure1)

### Figure 7 Displaying the data in the frequency distribution table

This figure displays a table and describes clearly how many days in a week Anna, who is an analyst, works in these three different organisations: a hospital, an accounting firm and a bank. This figure consists of two columns. The first column is labelled as Column A, which describes the types of organisations (hospital, accounting firm and bank) in which Anna works each week. The second column is labelled as Column B, which describes the number of days Anna works in the above mentioned organisations: 2 in hospital, 2 in accounting firm, 3 in bank).

[Back to - Figure 7 Displaying the data in the frequency distribution table](#Session4_Figure1)

### Uncaptioned Figure

[Back to - Uncaptioned Figure](" \l "Session4_Figure2)

### Figure 8 Arranging the data from ascending to descending order

JC Electrics Excel sheet

[Back to - Figure 8 Arranging the data from ascending to descending order](" \l "Session4_Figure3)

### Figure 8 Arranging the data from ascending to descending order

This figure shows the JC Electrics Excel sheet Generators column values on the left, which are: 7, 7, 7, 7, 8, 8, 9, 9, 9, 9, 9, 10, 10, 11, 11, 11, 12, 12, 12, 13, 13, 14, 14, 15. Next to this column are two empty columns labelled Value and Frequency. All three columns are under the headings ‘JC Electrics’ followed by (Quarterly data of units sold).

[Back to - Figure 8 Arranging the data from ascending to descending order](#Session4_Figure3)

### Figure 9 Calculating the frequency of ungroup data

JC Electrics Excel sheet

[Back to - Figure 9 Calculating the frequency of ungroup data](" \l "Session4_Figure4)

### Figure 9 Calculating the frequency of ungroup data

This figure shows the JC Electrics Excel sheet Generators column values on the left, which are: 7, 7, 7, 7, 8, 8, 9, 9, 9, 9, 9, 10. Next to this column are two further columns, with the first labelled as Value and the second labelled as Frequency. The Value column contains the values 7, 8, 9, 10, 11, 12, 13, 14, 15. The Frequency column in the row Generator:7, Value 7 shows the formula: =COUNTIF(A5:A28,7).

[Back to - Figure 9 Calculating the frequency of ungroup data](#Session4_Figure4)

### Figure 10 Ungrouped frequency table

Excel spreadsheet for JC Electrics

[Back to - Figure 10 Ungrouped frequency table](" \l "Session4_Figure5)

### Figure 10 Ungrouped frequency table

This figure shows an Excel spreadhseet for JC Electrics (Quarterly data of units sold). The first column, labelled Values, lists the values 7, 8, 9, 10, 11, 12, 13, 14, 15 The second column, labelled Frequency, contains the values 4, 2, 5, 2, 3, 3, 2, 2, 1.

[Back to - Figure 10 Ungrouped frequency table](#Session4_Figure5)

### Uncaptioned Figure

[Back to - Uncaptioned Figure](" \l "Session4_Figure6)

### Figure 11 Calculating the range

JC Electrics Excel sheet

[Back to - Figure 11 Calculating the range](" \l "Session4_Figure7)

### Figure 11 Calculating the range

This figure shows the JC Electrics Excel sheet Generators column values on the left, which are: 7, 7, 7, 7, 8, 8, 9, 9, 9, 9, 9, 10, 10, 11, 11, 11, 12, 12, 12, 13, 13, 14, 14, 15. Next to this column are two empty columns, with the first also labelled as Generators and the second labelled as Frequency. All three columns are under the headings ‘JC Electrics’ followed by (Quarterly data of units sold). To the right of this table are three cells labelled, from top to bottom, max, min and range. The max cell shows a value of 15, the min cell shows a value of 7, and the range cell contains the formula =H10-H11.

[Back to - Figure 11 Calculating the range](#Session4_Figure7)

### Uncaptioned Equation

width equals range over square root of sample size

[Back to - Uncaptioned Equation](" \l "Session4_Equation1)

### Figure 12 Calculating the width

Table arranged in rows showing max, min, range, sample size and width

[Back to - Figure 12 Calculating the width](" \l "Session4_Figure8)

### Figure 12 Calculating the width

This figure shows a table arranged in rows as follows: max: 15; min: 7; range: 8; sample size: 24; width: =H12/SQRT(H13).

[Back to - Figure 12 Calculating the width](#Session4_Figure8)

### Figure 13 Maing the class intervals

Excel spreadsheet for JC Electrics

[Back to - Figure 13 Maing the class intervals](" \l "Session4_Figure9)

### Figure 13 Maing the class intervals

This figure shows the JC Electrics Quarterly data of units sold, with columns labelled from left to right as: Generators, Class intervals, Lower limit of class interval, Upper limit of class interval. From top to bottom, the Generators column contains the values 7, 7, 7, 7, 8, 8, 9, 9, 9, 9, 9, 10, 10, 11, 11, 11, 12, 12, 12, 13, 13, 14, 14, 15. The Class intervals column contains the values 7\_8, 9\_10, 11\_12, 13\_14, 15\_16. The Lower limit of class interval contains the values 7, 9, 11, 13, 15; and the Upper limit of class interval contains the values 8, 10, 12, 14, 16.

[Back to - Figure 13 Maing the class intervals](#Session4_Figure9)

### Figure 14 Calculating the frequency by using an array formula

Excel spreadsheet for JC Electrics

[Back to - Figure 14 Calculating the frequency by using an array formula](" \l "Session4_Figure10)

### Figure 14 Calculating the frequency by using an array formula

This figure shows the JC Electrics Quarterly data of units sold, with columns labelled from left to right as: Generators, Class intervals, Lower limit of class interval, Upper limit of class interval and Frequency. From top to bottom, the Generators column contains the values 7, 7, 7, 7, 8, 8, 9, 9, 9, 9, 9, 10, 10, 11, 11, 11, 12, 12, 12, 13, 13, 14, 14, 15. The Class intervals column contains the values 7\_8, 9\_10, 11\_12, 13\_14, 15\_16. The Lower limit of class interval contains the values 7, 9, 11, 13, 15; and the Upper limit of class interval contains the values 8, 10, 12, 14, 16. Finally, the top cell of the Frequency column contains the formula =FREQUENCY(A4:A27,D4:D8), with th rest of the column containing the values 7, 6, 4, 1, 0.

[Back to - Figure 14 Calculating the frequency by using an array formula](#Session4_Figure10)

### Figure 15  Calculating the relative frequency of ungrouped data

Excel spreadsheet for JC Electrics

[Back to - Figure 15  Calculating the relative frequency of ungrouped data](" \l "Session4_Figure11)

### Figure 15  Calculating the relative frequency of ungrouped data

This figure shows the JC Electrics Quarterly data of units sold, with columns labelled from left to right as: Generators, Value, Frequency and Relative frequency. From top to bottom, the first Generators column contains the values 7, 7, 7, 7, 8, 8, 9, 9, 9, 9, 9, 10. The Value column contains the values 7, 8, 9, 10, 11, 12, 13, 14, 15 and Sample size. The Frequency column contains the values 4, 2, 5, 2, 3, 3, 2, 2, 1, and in red, 24. The top cell of the Relative frequency column contains the formula =C5/C14, and underneath are the values 0.166666667, 0.208333333, 0.083333333, 0.125, 0.125, 0.083333333, 0.041666667.

[Back to - Figure 15  Calculating the relative frequency of ungrouped data](#Session4_Figure11)

### Figure 16 Calculating the cumulative frequency

Excel spreadsheet for JC Electrics

[Back to - Figure 16 Calculating the cumulative frequency](" \l "Session4_Figure12)

### Figure 16 Calculating the cumulative frequency

This figure shows the JC Electrics Quarterly data of units sold, with columns labelled from left to right as: Generators, Class intervals, Lower limit of class interval, Upper limit of class interval, Frequency and Cumulative frequency. From top to bottom, the Generators column contains the values 7, 7, 7, 7, 8, 8, 9, 9, 9, 9, 9, 10, 10, 11, 11, 11, 12, 12, 12, 13, 13, 14, 14, 15. The Class intervals column contains the values 7\_8, 9\_10, 11\_12, 13\_14, 15\_16. The Lower limit of class interval contains the values 7, 9, 11, 13, 15. The Upper limit of class interval contains the values 8, 10, 12, 14, 16. The Frequency column contains the values 6, 7, 6, 4, 1; and the Cumulative frequency column contains the values 6, 13, 19, 23, 24.

[Back to - Figure 16 Calculating the cumulative frequency](#Session4_Figure12)

### Figure 17 A typical histogram in Excel

A histogram in Excel

[Back to - Figure 17 A typical histogram in Excel](" \l "Session4_Figure13)

### Figure 17 A typical histogram in Excel

This figure shows a histogram in Excel. The horizontal x-axis the numbers from minus 3 to 5 in increments of 0.5, with a final bar labelled ‘>5’. The vertical y-axis shows the ‘frequency’ and ranges from 0 to 140 in increments of 20.

[Back to - Figure 17 A typical histogram in Excel](#Session4_Figure13)

### Figure 18 A frequency distribution table in Excel

Excel spreadsheet for JC Electrics

[Back to - Figure 18 A frequency distribution table in Excel](" \l "Session4_Figure14)

### Figure 18 A frequency distribution table in Excel

This figure shows the JC Electrics Quarterly data of units sold, with columns labelled from left to right as: Generators, Class intervals, Lower limit of class interval, Upper limit of class interval and Frequency. From top to bottom, the Generators column contains the values 7, 7, 7, 7, 8, 8, 9, 9, 9, 9, 9, 10, 10, 11, 11, 11, 12, 12, 12, 13, 13, 14, 14, 15. The Class intervals column contains the values 7\_8, 9\_10, 11\_12, 13\_14, 15\_16. The Lower limit of class interval contains the values 7, 9, 11, 13, 15. The Upper limit of class interval contains the values 8, 10, 12, 14, 16. The Frequency column contains the values 6, 7, 6, 4, 1.

[Back to - Figure 18 A frequency distribution table in Excel](#Session4_Figure14)

### Figure 19 Histogram dialog box in Excel

Histogram dialog box in Excel

[Back to - Figure 19 Histogram dialog box in Excel](" \l "Session4_Figure15)

### Figure 19 Histogram dialog box in Excel

This figure shows the histogram dialog box in Excel. On the left is a heading that reads Input, under which are Input range and Bin range, with boxes to the right to add the required information. Under the range tags is an empty tick box named Labels. This is followed by the header Output options, with radio buttons underneath called Output range (which is selected), New worksheet ply, and New workbook. Output range and New worksheet ply have boxes to the right to input values. Under the radio buttons are three tick boxes labelled, from top to bottom, Pareto (sorted histogram), Cumulative percentage, and Chart output. Chart output is selected. To the top right of the window are buttons labelled OK, Cancel, and Help.

[Back to - Figure 19 Histogram dialog box in Excel](#Session4_Figure15)

### Figure 20 Histogram of number of units sold (Generators)

Excel spreadsheet showing a table and histogram for JC Electrics

[Back to - Figure 20 Histogram of number of units sold (Generators)](" \l "Session4_Figure16)

### Figure 20 Histogram of number of units sold (Generators)

This figure shows a table to the left and a histogram to the right. The table has two columns headed Bin and Frequency. The values in Bin are 8, 10, 12, 14, 16, More; the values in Frequency are 6, 7, 6, 4, 1, 0. The histogram shows five blue bars. The x-axis is labelled Bin and runs from 8 to More in increments of 2. The y-axis is labelled Frequency and runs from 0 to 8 in increments of 2. The bars represent the values from the table, showing 8 on the x-axis with 6 on the y-axis, followed by 10 (x) with 7 (y); 12 (x) with 6 (y), 14 (x) with 4 (y) and 16 (x) with 1 (y). There is no bar in the More (x) with 0 (y) space.

[Back to - Figure 20 Histogram of number of units sold (Generators)](#Session4_Figure16)

### Figure 21  Histogram showing units sold of generators

A histogram in Excel

[Back to - Figure 21  Histogram showing units sold of generators](" \l "Session4_Figure17)

### Figure 21  Histogram showing units sold of generators

This figure shows a histogram with five blue bars. The x-axis is labelled Bin and runs from 8 to More in increments of 2. The y-axis is labelled Frequency and runs from 0 to 8 in increments of 2. The bars show 8 on the x-axis with 6 on the y-axis, followed by 10 (x) with 7 (y); 12 (x) with 6 (y), 14 (x) with 4 (y) and 16 (x) with 1 (y). There is no bar in the More (x) with 0 (y) space.

[Back to - Figure 21  Histogram showing units sold of generators](#Session4_Figure17)

### Uncaptioned Equation

Frequency density equals frequency over class width

[Back to - Uncaptioned Equation](" \l "Session4_Equation2)

### Figure 22 Information about the age of people playing football in an Excel file

A histogram in Excel

[Back to - Figure 22 Information about the age of people playing football in an Excel file](" \l "Session4_Figure18)

### Figure 22 Information about the age of people playing football in an Excel file

This figure shows a spreadsheet which consists of four columns labelled Age, Frequency, Class width and Frequency density. The Age column contains the values multirelation zero less than or equals multiplication less than 10, multirelation 10 less than or equals multiplication less than 15, multirelation 15 less than or equals multiplication less than 20, multirelation 20 less than or equals multiplication less than 30, multirelation 30 less than or equals multiplication less than 50. The Frequency column contains the values 10, 9, 6, 4, 7. The remaining two columns are empty.

[Back to - Figure 22 Information about the age of people playing football in an Excel file](#Session4_Figure18)

### Figure 23 Calculating the class width

Excel spreadsheet

[Back to - Figure 23 Calculating the class width](" \l "Session4_Figure19)

### Figure 23 Calculating the class width

This figure shows a spreadsheet which consists of four columns labelled Age, Frequency, Class width and Frequency density. The Age column contains the values multirelation zero less than or equals multiplication less than 10, multirelation 10 less than or equals multiplication less than 15, multirelation 15 less than or equals multiplication less than 20, multirelation 20 less than or equals multiplication less than 30, multirelation 30 less than or equals multiplication less than 50. The Frequency column contains the values 10, 9, 6, 4, 7. The Class width column contains the values 10, 5, 5, 10, 20. The Frequency density column is empty.

[Back to - Figure 23 Calculating the class width](#Session4_Figure19)

### Figure 24 Calculating the frequency density

Excel spreadsheet

[Back to - Figure 24 Calculating the frequency density](" \l "Session4_Figure20)

### Figure 24 Calculating the frequency density

This figure shows a spreadsheet which consists of four columns labelled Age, Frequency, Class width and Frequency density. The Age column contains the values multirelation zero less than or equals multiplication less than 10, multirelation 10 less than or equals multiplication less than 15, multirelation 15 less than or equals multiplication less than 20, multirelation 20 less than or equals multiplication less than 30, multirelation 30 less than or equals multiplication less than 50. The Frequency column contains the values 10, 9, 6, 4, 7. The Class width column contains the values 10, 5, 5, 10, 20. The Frequency density column contains the values 1, 1.8, 1.2, 0.4, 0.35.

[Back to - Figure 24 Calculating the frequency density](#Session4_Figure20)

### Uncaptioned Figure

[Back to - Uncaptioned Figure](" \l "Session4_Figure21)

### Uncaptioned Figure

[Back to - Uncaptioned Figure](" \l "Session5_Figure1)

### Uncaptioned Figure

[Back to - Uncaptioned Figure](" \l "Session5_Figure2)

### Figure 25 Spreadsheet of Pepsi sold and Temperature

Spreadsheet of Pepsi sold and Temperature

[Back to - Figure 25 Spreadsheet of Pepsi sold and Temperature](" \l "Session5_Figure3)

### Figure 25 Spreadsheet of Pepsi sold and Temperature

A spreadsheet which consists of two columns. Column A is labelled as Temperature and Column B is labelled as Pepsi. From top to bottom, the Temperature column contains the values 12, 14, 15, 17, 22, 13, 20, 23. The Pepsi column contains the values 12, 16, 16, 19, 32, 10, 24, 40.

[Back to - Figure 25 Spreadsheet of Pepsi sold and Temperature](#Session5_Figure3)

### Figure 26 How to select the Scatter symbol

The Recommended charts toolbar in Excel

[Back to - Figure 26 How to select the Scatter symbol](" \l "Session5_Figure4)

### Figure 26 How to select the Scatter symbol

This figure shows a section of the Recommended charts toolbar in Excel. Over the label ‘Charts’ are different chart icons for histograms, boxplots, bar charts, pie charts and so on. the cursor hovers over the icon representing scatter plots. Next to the collection of icons is a globe icon labelled ‘Maps’ with a dropdown arrow, and a pivot chart icon labelled ‘PivotChart’, also with a dropdown arrow.

[Back to - Figure 26 How to select the Scatter symbol](#Session5_Figure4)

### Figure 27 How to select the Scatter chart type

Section of the Scatter type selection menu in Excel

[Back to - Figure 27 How to select the Scatter chart type](" \l "Session5_Figure5)

### Figure 27 How to select the Scatter chart type

This figure shows a section of the Scatter type selection menu in Excel with icons representing different styles of scatter charts arranged around the top and left edge. In the centre is text which reads: ‘Scatter. Use this chart type to: Compare at least two sets of values or pairs of data; show relationships between sets of values. Use it when: The data represents separate measurements.’

[Back to - Figure 27 How to select the Scatter chart type](#Session5_Figure5)

### Figure 28 Temperature versus Pepsi sales

Scatter diagram showing the relationship between temperature and Pepsi sales

[Back to - Figure 28 Temperature versus Pepsi sales](" \l "Session5_Figure6)

### Figure 28 Temperature versus Pepsi sales

Scatter diagram showing the relationship between Temperature (x-axis) and Pepsi sales (y-axis), with the heading ‘Pepsi versus Temperature’. The Temperature axis ranges from 0 to 25, and the Pepsi sales axis ranges from 0 to 45. The points plotted show that as temperature increases, so do the sales of Pepsi.

[Back to - Figure 28 Temperature versus Pepsi sales](#Session5_Figure6)

# Video 1

## Transcript

NARRATOR

In this video, I will be showing you how to make an ungrouped frequency distribution table in Excel. Start by opening the Excel file JC\_electrics. You can see here that there are four sheets within this Excel file.

For this activity, we are going to use the sheet titled ungrouped frequency table. Go ahead and select it by clicking the tab. Here you will find the quarterly data of the number of generators sold. Make sure that the data is arranged in columns. Label column B with the variable name Criteria and column C as Frequency.

Now, first, we will find the minimum and maximum values in the data. Select a cell outside of the range and type =MAX. And select the range, which is A5 to A28, then press Enter. This will give you the maximum value, which is 15.

To find the minimum value in column G, type =MIN. And select the range, which is A5 to A28. Press Enter. This will give you the minimum value, which in this case is seven.

To calculate the range in column G, type equals. And then select the maximum value, in this case, 15 minus. And then select the minimum value, which is seven. Then press Enter. This calculates the range of eight for this data set.

We will now arrange the values in column A labeled Generators from ascending to descending order. To do this, select the values A5 to A28 in column A. Then click Data in the toolbar. Select Sort. Then select Continue with selection and press Sort.

In column B, which we label Criteria, we will type in every value from our range starting from the minimum value of seven to the maximum value of 15. We will use this column to calculate the frequency.

Let's now assume we want to calculate in how many quarters the company sold a maximum of seven units. To do this, we need to calculate the frequency. Column A contains sales with values between 7 to 15, which we have arranged in descending order. Go to the first row of column C under the label Frequency and type =COUNTIF.

Select the range A5 to A28 in column A. Then select the first criteria, which is seven from column B. This will produce a frequency of four. This means that seven units were sold in four quarters. Repeat this step for each criteria until you reach the number of quarters in which 15 units are sold. Note how in this formula only the criteria value changes each time.

And here we have an ungrouped frequency table.

[Back to - Video 1](" \l "Session4_MediaContent1)

# Video 2

## Transcript

NARRATOR

Hello. In this video, I will be showing you how to make a grouped frequency distribution table in Excel. We will do this by following these steps. Firstly, we will calculate the range, then we will determine the class interval width. Before moving on to determining the number of groups of distribution, finally, we will end this activity by creating the class intervals.

Start by opening the Excel file JC Electrics and then click on the tab named grouped frequency table to access the correct sheet. This sheet contains the quarterly data of the number of generators sold. Make sure that the data is arranged in columns. Column A named Generators contains values between 7 to 15, and all the values are arranged in ascending order.

Now we will calculate the range, which is the difference between the maximum and minimum value in the data set. The maximum value in the given data set is 15 and the minimum value is 7. In Excel in column H, type equal max. Then select the data in column A and press Enter.

This will give you the maximum value which is 15. To calculate the minimum value in column H again, type equal min. Then select the data in column A and press Enter. This will give you the minimum value which is 7.

Let's now have a look at how we calculate the range. The formula for range is the difference between the maximum and minimum value. You can enter this formula anywhere on the sheet outside the table.

Type Equal then select the cell that contains the maximum value. Here it is H10. Type minus and select the cell which contains the minimum value. Here it is H11. Then press Enter. This works out as 15 minus 7 equals 8.

Now let's look at how we determine the class interval width. There are no firm rules on this. However, this formula is the most common method to calculate the width. Width is equal to range divided by the square root of sample size.

To calculate the sample size in column H, you need to type Equal then Count and select the data range A4 to A27. This will give you the sample size which is 24. To calculate the width of the class interval with excel, we use the following formula. Type equals and select the cell where you calculated the range. Here it is H12.

Type Divide and then SQRT for square root followed by the cell for sample size. Here it is H13. This will divide the range which is 8 by the square root of 24, which in this example is the sample size. The width calculated in the data set is 1.6. You can round this value to a whole number or a number that is convenient to add such as a multiple of 10. For our example, we will round the number up to 2.

Next up, we will determine the number of groups or class intervals into which data is to be distributed. Each class interval or group consists of a lower limit and an upper limit. The lower limit is represented by the lowest value in the group, and the upper limit is the highest value in that group.

To determine the lower and upper limit of each class interval or group in the data set, you need to find out the lowest value in the data set. In this data set, it is 7. This is the lower limit of the first group or class interval. You then add the width of the class interval to this lower limit. As you calculated earlier, the width is 2. So this works out as 7 plus 2 equals 9.

This value 9 is the lower limit of the next class interval/group. Keep adding the interval width to the previous lower limit of the class interval to calculate more class intervals until you exceed the highest value. Now you can make the class intervals or groups.

In this data set, the intervals are 7 to 8, 9 to 10, 11 to 12, 13 to 14, 15 to 16. As you can see here, the first class intervals lower limit equals 7 and upper limit equals 8. Let's review.

Column A named generators shows the values between 7 to 15 in ascending order. Column B is labeled as class interval. Each class interval consists of a lower limit and an upper limit. These are presented in column C and column D, respectively.

To find the class interval, add the width which in this case is 2 to the lower limit of the first class interval, which is 7. This will give you the lower limit of the second-class interval. The next step is to calculate the frequency.

In column E, enter the frequency function type equals freq open bracket A4 to A27 comma open bracket D4 to D8. Close bracket and press Enter.

[Back to - Video 2](" \l "Session4_MediaContent2)

# Video 3

## Transcript

NARRATOR

Difference between histograms. In this video, we will discuss the following topics-- what we mean by frequency density, how we calculate frequency density and construct the frequency density histogram, what is the difference between the frequency histogram and the frequency density histogram.

Let's first have a look at what frequency density is. Frequency density is defined as the frequency per unit of the data in each class. Frequency density gives us the ratio of the frequency of a class to its width. Frequency density is used to plot the histogram. And it allows for a meaningful comparison of different classes where the class width may not be equal.

So how is frequency density calculated? Frequency density is calculated by dividing the frequency with the class width. We calculate frequency density when we have a set of group data that consists of an unequal width of class intervals.

Let's have a look at this worksheet for an example. This Excel sheet consists of information relating to the age of people playing football. The spreadsheet consists of four columns. Column A shows the age range of the people playing football in distinct groups. Column B shows the frequency. Column C is labeled as class width. And column D is labeled as frequency density.

To calculate the frequency density of each class interval or group, first, you should determine the class width of each group. In column C, we can find the class width of the class intervals or groups by determining the difference of upper and lower bounds, also called limits. For example, in column C on the first row, type equals 10 minus 0 equals 10, equals 15 minus 10 equals 5, and so on.

Then in column D, you will calculate the frequency density by using the formula for frequency density, which is the frequency of each class interval or group divided by its width. For example, on the first row of column D, we divide the frequency of the first class interval or group, which is 10 by its class width, which is also 10, so equals 10 divided by 10 equals 1, which is the frequency density of the first class interval or group.

You can calculate the frequency density of the rest of the groups in the same way. And we get nine divided by five equals 1.8. 18 divided by 35 equals 0.5. 7 divided by 10 equals 0.7. And 5 divided by 10 equals 0.5.

Once you calculate the frequency density of each group, you will be able to construct the frequency density histogram. As we said earlier, frequency density is used to plot the histogram and allows for a meaningful comparison of different classes where the class width may not be equal. Here you can see that we plot the frequency density on the vertical y-axis and the class interval or group, which in this case is the age range of the people playing football, on the horizontal x-axis.

The height of each bar shows the frequency density of each group. You will also see that the bar of range 15 to 50 is wider than the bar of range 0 to 10. In the frequency density histogram, the frequency of each group is equal to the area of each class. And it can be calculated by multiplying the class width by the frequency density.

So the frequency is equal to the area of the class, which is equal to the class width multiplied by the frequency density.

Let's now have a look at what the difference between the frequency histogram and frequency density histogram is.

Here, we have a frequency histogram and frequency density histogram and their corresponding frequency tables. In each case, we are measuring quantitative data that is continuous. This here is the frequency histogram. And you will notice that the bars or groups consist of a range of quantitative data values that are equally spaced out, whereas in the case of the frequency density histogram, which is this histogram here, the bars or groups of each range of quantitative data values are not equally spaced out.

This is not always the case as you can have a frequency density histogram which has equally spaced out bars or columns. In both the cases of the frequency histogram and the frequency density histogram, the bars are connected with each other. And there are no spaces between the bars.

Another distinction between the frequency histogram and the frequency density histogram is that in the case of the frequency histogram, the height of the bars corresponds to the frequencies, whereas in the case of the frequency density histogram, the height of the bars corresponds to the frequency density. The order or sequence of the bars cannot be changed again. For example, once you determine the class intervals or groups, then you cannot change their orders.

[Back to - Video 3](" \l "Session4_MediaContent3)

# Video 4

## Transcript

NARRATOR

Hello. In this video, we will discuss what we mean by bivariate data. Well, just by the word bivariate data, we know that we are dealing with some kind of data. But what kind of data is the question. To better understand the word bivariate, we need to break it down in two. The prefix bi means two or twice, and variant means a quantity that can take the value of any member of a particular set, a variable.

So bivariate data are data that have two variables with different values. With bivariate data, we are comparing two different sets of variables. The one variable is the dependent variable, and the other variable is the independent variable. Let's look at an example.

As you can see, this table has three columns. The columns are labeled as days, temperature, and the number of Pepsi cans sold. So this table is showing the data relating to the number of Pepsi cans sold in the different days of a week. The two sets of variant data we have here are the temperature, and the number of Pepsi cans sold.

The temperature on this table is the independent variable, and the number of Pepsi cans is the dependent variable, because the number of Pepsi cans sold depends on the increase or decrease in temperature. As you can see in this table, on the day the temperature was 12 degrees celcius, the number of Pepsi cans sold was 12. When the temperature increased from 12 to 18 degrees celcius, the sale of Pepsi cans also increased.

Now, let's have a look at how these two different sets of data can be compared. We can plot this data into a graph, and the scatter plot is one of the best graphs for representing bivariate data. On this scatter plot here, we have the x-axis, which is the horizontal axis, representing the temperature, which is our independent variable. And on the y-axis, which is the vertical axis, we have the dependent variable, which is the number of Pepsi cans sold.

So from the graph, you can see that with the increase in temperature, we also have an increase in the number of Pepsi cans sold. In other words, this graph shows a positive correlation between the two variables because with the increase in temperature, the sale of Pepsi cans has also increased.

There can be cases where there is no positive correlation, as with the example here, where the increase in temperature has not seen an increase in the number of packets of biscuits sold. This is what we call a negative correlation.

In this video, you learned what bivariate data are, and that they are best graphed in a scatter plot to show either a positive or negative correlation.

[Back to - Video 4](" \l "Session5_MediaContent1)

# Video 5

## Transcript

NARRATOR

This video shows how to draw scatter plots in Microsoft Excel. We'll use the data file W4\_Apple\_Amzn\_returns.xlxs to draw a scatter plot of monthly stock returns for Apple and Amazon. The data file has three columns. The first column, Date, represents calendar months. The second column, Amazon, shows monthly stock returns for Amazon. And the third column, Apple, shows monthly stock returns for Apple.

We see in the month that ended on the 1st of January, 2018, Amazon had positive stock returns of about 24%, while Apple stock went down by about 1%. The data has 25 rows, including one row for the labels, which means that our data has 24 observations of monthly stock returns for the two firms.

Let's highlight our data columns Amazon and Apple by clicking on the columns B and C. Go to Insert. Look under Charts and choose Scatter, which is under the little graph that looks like a few unconnected dots scattered around. Excel will now make a scatter plot from the highlighted data.

A scatter plot helps you visualise how two sets of numerical values are related. Each dot represents two separate measurements for a single instance. So each dot represents one month. And Amazon stock returns for that month are plotted on the x-axis, while Apple returns are plotted on the y-axis.

Once we have a scatter plot, we can add a title. We want to describe both of our measurements in the title, so we'll call this Monthly Returns, Amazon versus Apple. Next, we need to add labels for the x and y-axis. Click on the Chart area.

Go to Add Chart element. Choose Axis Titles. And click on Primary Horizontal to add the x-axis label. And choose Primary Vertical for the y-axis label. It's important that the axis labels show the units of measurement. Both the x-axis and y-axis show percentage returns, so we label them accordingly.

It can be useful to change the position of the horizontal and vertical axes. To change the position of the vertical axis, right-click on the horizontal axis and choose Format Axis. We can move the vertical axis by changing the value of the horizontal axis at the point which it crosses the vertical axis.

If we want to move the vertical axis all the way to the left, we tick Axis value and type in the lowest value for the horizontal axis. The lowest value can be seen here under Bounds. It's minus 30. So we type minus 30 here and click Enter. The y-axis has now moved all the way to the left of the graph and crosses the x-axis at minus 30.

We can move the position of the horizontal axis in the same way. Right-click on the vertical axis and choose Format Axis. To move the horizontal axis down, tick Axis value and type in the lowest value for the vertical axis. The lower vertical bound for our plot is minus 25, so type in minus 25 and click Enter. The x-axis is now moved all the way to the bottom. And the graph or scatter plot is now less cluttered.

[Back to - Video 5](" \l "Session5_MediaContent2)