Your logo here

COURSE MANUAL

C7: Quantitative Techniques

Module 2

[Add institute name here]
[Add School/Department name here]

Copyright

Commonwealth of Learning
1055 West Hastings Street
Suite 1200
Vancouver, BC V6E 2E9
CANADA

Email: info@col.org

[Add institute name here]
[Add School/Department name here]

[Add address line 1]
[Add address line 2]
[Add address line 3]
[Add country]

Fax: +[Add country code] [Add area code] [Add telephone #]
Email: [Add e-mail address]
Website: www.[Add website address]

Acknowledgements

The Commonwealth of Learning (COL) wishes to thank those below for their contribution to the development of this course:

Course author	Barbara Swart, PhD Professor, Quantitative Management University of South Africa, South Africa
Subject matter experts	Samuel Kwame Amponsah, PhD Kwame Nkrumah University of Science and Technology, Ghana
	Kim Loy Chee, PhD Wawasan Open University, Malaysia
	S.A.D. Senanayake Open University of Sri Lanka, Sri Lanka
Educational designers	Symbiont Ltd. Otaki, New Zealand
Course editor	Symbiont Ltd. Otaki, New Zealand

COL would also like to thank the many other people who have contributed to the writing of this course.

Contents

Module 2 1

Unit 5 2

Unit 6 15

Unit 7 33

Module 2

Statistical tools and probability theory

Introduction

This module introduces the basic ideas of probability and statistics. These ideas provide the basis for interpreting many properties of the data sets you will encounter in management situations.

Statistics help us to analyse practical situations. Although there are many statistical tools available for interpreting data, care must be taken when basing decisions on statistical results. We all know the quote, attributed to Benjamin Disraeli: “There are lies, damned lies, and statistics.”

Probability theory is a challenging subject, dealing with notions of chance and likelihood. These are important ideas in a world where uncertainty and volatility affect decision making.

Upon completion of this module you will be able to:

Outcomes

- **display** data in the form of graphs and charts;
- **identify** appropriate statistical tools in analysing data (mean, median, and so on);
- **interpret** results correctly to inform decisions;
- **understand** probabilities of events;
- **handle** probability distributions and understand their role in modelling; and
- **apply** expected values, variance and standard deviation to risk and volatility.

Unit 5

Collecting and presenting data

Introduction

The basis for decision making lies in data and information. However, these two concepts are not the same: *data* has to be analysed and interpreted before it can be presented as *information* and used to make management decisions.

Data can be collected and analysed in a variety of ways for use in decision making:

- Data that occurs in pairs of numbers may be represented by scatter graphs or line graphs.
- Data in categories may be represented by column charts or pie charts (with the measured values on the vertical axis in column charts or displayed as sections of a disc in pie charts).
- Where measured values are grouped into classes and the number of observations for each class is of interest, data may be displayed in frequency tables and histograms (with the number of observations or frequency on the vertical axis, while the horizontal axis displays the grouped data, which may be either discrete or continuous).

Upon completion of this unit you will be able to:

Outcomes

- **classify** different types of data;
- **represent** data in tables;
- **display** data in charts and graphs; and
- **draw** and **display** frequency tables.

Data collection and tables for data

Data can be collected from published sources, specially designed questionnaires, surveys or complicated research projects.

Nominal data or **categorical data** is organised into categories. For example, 60 per cent of respondents feel positive about the economy, 26 per cent feel negative and 14 per cent are neutral.

Ordinal data has a scale or ranking connected to the observations. For example, people may rank their feelings about a product on a scale from 1 to 10 – from very dissatisfied to highly satisfied.

Cardinal data has some measurable property such as length, weight, price, and so on. The values of the measurements may form discrete points in the real number system or a continuous interval or section of the real number line. Therefore **discrete data** may consist of natural numbers 1, 2, 3 and so on to describe the number of children in a family, whereas **continuous data** may consist of the numbers between 20 and 300 to describe the mass of adults (in, say, kilograms).

Data can consist of individual observations or be grouped into classes of values with the number of observations in each class.

Samples

The collection of all possible entities from which data could be collected is called the **population**. This could refer to all adults in a country, all animals in a specific area, all shares traded on a stock exchange, and so on. It is usually not feasible to collect data from each entity, so we take samples. A **sample** is a smaller subset of the entire population.

The choice of a sample is an extremely important part of the process of collecting and interpreting data. It must represent the population in an unbiased way and be large enough to be properly representative. If you are investigating the effects of an economic downturn on people in a city, you cannot interview only 10 people or only people from the wealthy suburbs.

A **random sample** means that each entity in the population has an equal chance of being selected. Computers can generate random numbers (we will discuss this shortly), and these can be used to pick units for the sample from a population.

Here's an example:

A random sample of 100 students from a population of 500 students needs to be chosen. Assume the students' registration numbers have five digits.

- Use a computer program to randomly generate 100 five-digit numbers.
- Choose the students with the matching registration numbers for the random sample.

Systematic samples are useful when you want to inspect items on a production line. A manager may decide, for example, to inspect every 10th item coming off the production line.

Quota samples can be used to make the sample representative. If you want to interview people aged over 40 and the population consists of 55 per cent females and 45 per cent males, then a quota sample of 1,000 people can be made to consist of 550 females and 450 males.

This can be refined even further using stratified sampling, where you ensure that different strata or groups are included.

Charts, graphs and histograms

The graphic representation of data is an excellent tool for analysing information because it can make patterns and the relationships between quantities clearer.

Module 1 explained the use of tables and graphs to represent linear and certain non-linear (quadratic) relations. This can be extended to other types of graphs and charts. The type of data often determines the type of graph or chart you will use. If the data is in the form of pairs of measurements, you would use a column chart or a scatter or line graph. If the data is in the form of categories or time periods, with measurements or percentages attached for each category, you could use bar charts, histograms or pie charts.

Here's an example:

The table in Figure 1 gives the sales figures for the number of pairs of sunglasses sold in a shop in an island resort.

Month and sales figure	Month and sales figures
January: 8	July: 32
February: 6	August: 44
March: 9	September: 36
April: 15	October: 21
May: 18	November: 12
June: 26	December: 10

Figure 1

If you number the months from 1 to 12 on the horizontal axis and indicate sales on the vertical axis, you get the scatter graph or chart shown in Figure 2.

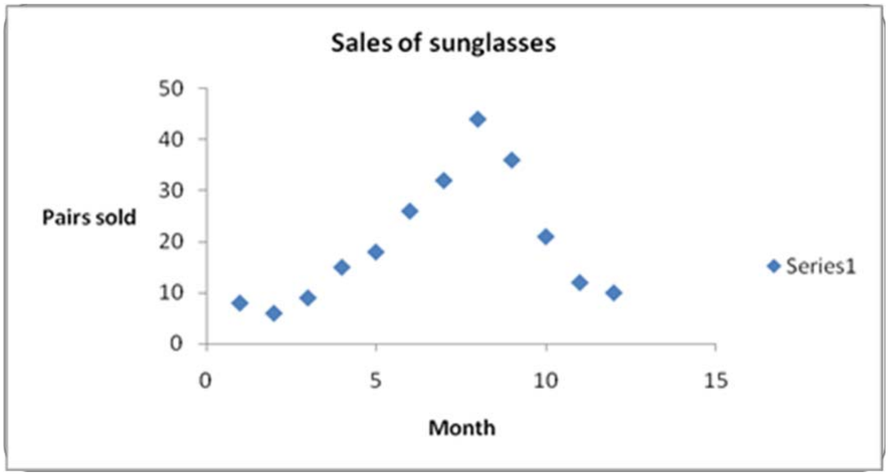


Figure 2

"Sunglasses sold" is considered the dependent quantity.

You can see that sales increased over the months leading up to August and then decreased again. If you have data for a few years, you may be able to pick up a trend or seasonal pattern in the sales.

Note: A change in scale can make a huge difference and distort the analysis, as Figure 3 shows.

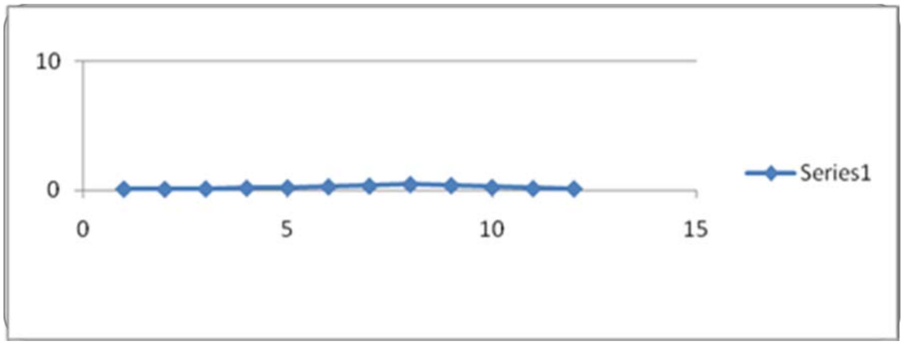


Figure 3

The scale on the vertical axis has been compressed to show units of 1,000. This makes it hard to distinguish between the quantities and the variation in the sales figures cannot easily be seen. Always choose an appropriate scale.

The data from Figure 1 can be presented as a piecewise linear graph. A piecewise linear graph consists of a number of straight lines. The lines can be joined together or there can be jumps (discontinuities):

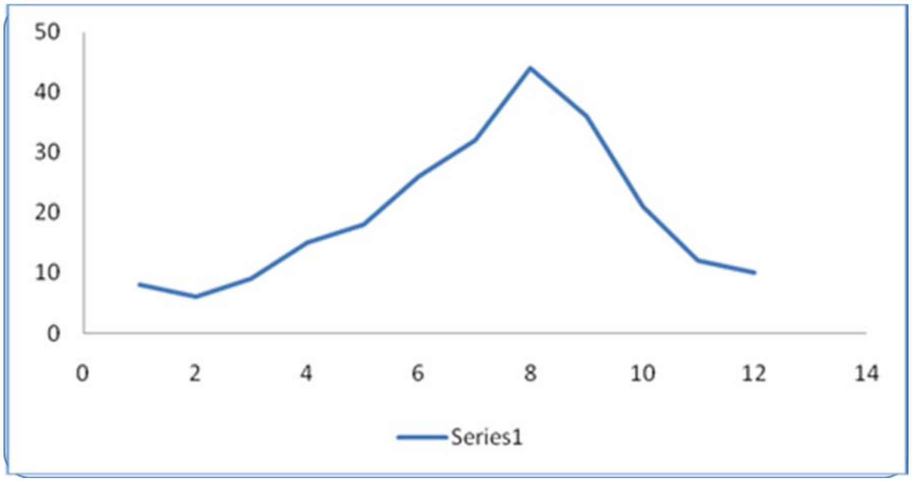


Figure 4

Or the data may be set out as a column chart:

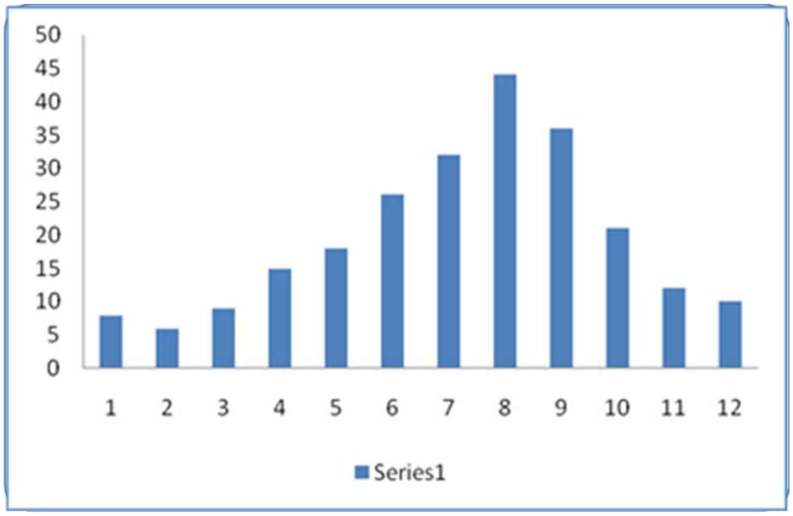


Figure 5

These charts can be drawn using computer software such as Excel or Open Office, and this will be discussed further in Module 3.

Each representation has advantages and disadvantages.

Sometimes data is in the form of pairs of values, with no clear indication of which variable should be **dependent** (represented on the vertical axis) or **independent** (represented on the horizontal axis). For example, the data may be numbers of sunglasses and numbers of caps sold and you want to see if there is any relation between the two. You may have either on the vertical axis, as the two quantities do not have a dependent/independent relationship. That is, there is no **causal relationship**.

Pie charts

Pie charts are more effective than column charts when there is a small number of categories.

Here's an example:

The popularity of four car colours was tested. A survey of 40 people showed that 18 preferred black cars, 12 preferred silver, six preferred blue and four preferred red cars.

The associated pie chart looks like this:

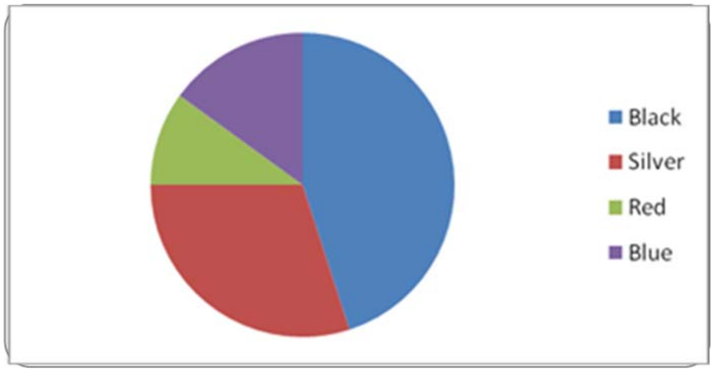


Figure 6

The full pie represents 100% of the people. Black was chosen by 18 out of 40 people, which gives a percentage of 45% (just less than half or 50%). Remember that $\frac{18}{40}$ = 0.45 = 45%. The full pie also represents a circular sweep of 360 degrees. Now 45% of 360˚ = 162˚, so the colour black forms a slice with central angle 162˚. Similarly, silver was chosen by

$\frac{12}{40}$ = 30% of respondents and represents a slice of size 0.30 × 360˚ = 108˚.

Activity 2.1

Activity

Interpret graphs

What will you do?

1. Calculate the size of slice each colour makes in the pie chart in Figure 6 and show that the slices add up to 360˚.
2. Represent the car colour data as a column chart.

Frequency distributions: Tables and histograms

Sometimes data takes the form of **observations** made for classes of values. You may not have individual measurements but be more interested in how many **entities** (observations) have a specific property. For example, weekly sales figures taken over a year may be divided into classes such as:

- 10 to 19 units sold,
- 20 to 29 units sold,
- 30 to 39 units sold, and
- 40 to 49 units sold.

For each class, you count the number of weeks that sales fell into a particular class. There were eight weeks in which the weekly sales figures were between 10 and 19 units sold and 12 weeks in which the weekly sales figures were between 20 and 29 units sold. (Note that the number of weeks adds up to 52.)

Sales class (units sold per week)	Number of weeks (observation)
10–19	8
20–29	12
30–39	28
40–49	4

Figure 7

For this type of situation, you draw up a frequency table showing the frequency distribution, as in Figure 7. (Frequencies mean number of observations and are always whole numbers.) The graphic representation is called a **histogram**, with the classes represented on the horizontal axis and the frequency (number of entities with each class property) on the vertical axis. The classes can represent discrete data such as number of units sold, or continuous data such as the mass of items. ("Discrete" in this context means whole numbers and "continuous" means any real value in an interval of values.)

Here's an example:

A study was conducted by the municipal manager to measure the daily consumption of water by 125 households in a village. The water consumption is denoted by five possible intervals of values for litres (x); the number of households consuming a certain range of litres is denoted by $f(x)$.

A frequency table shows the number of households using 0–20 litres per day, 20–40 litres per day, 40–60 litres per day, 60–80 litres per day and 80–100 litres per day. There is a limit of 100 litres per day.

Litres per day (x)	Number of households (frequency f(x))
0–20	10
20–40	36
40–60	45
60–80	28
80–100	6

Figure 8

The table in Figure 8 is called the frequency distribution for the water consumption. We will agree that 20–40 means greater than or equal to 20 but less than 40. There is a continuous range of values for water consumption; a histogram is used to display continuous data. The horizontal axis will have a continuous scale for the litres of water consumed.

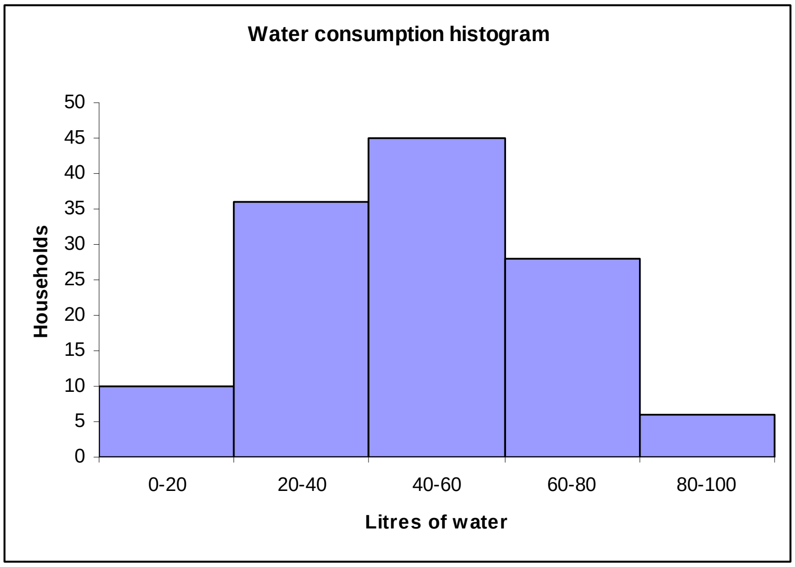


Figure 9

Note: In histograms, the width (range of values) and height (frequency or number of observations) of each rectangular bar is important. Each area (width × height) has meaning.

In this case, each area gives the maximum total number of litres used on a weekly basis by that group of households. It is best to have the horizontal classes of the same width.

In the frequency table shown in Figure 10, you can add a column for cumulative frequency and a column for relative frequency. Cumulative frequency $F(x)$ adds frequencies up to a point x, while relative frequency is the ratio of frequency of a particular class to the total number of observations. In the case of water consumption, you get:

Litres per day (x)	Frequency f(x)	Cumulative frequency F(x)	Relative cumulative frequency
0–20	10	10	0.08
20–40	36	46	0.288
40–60	45	91	0.36
60–80	28	119	0.224
80–100	6	125	0.048

Figure 10

Figure 10 shows that 46 households use between 0 and 40 litres per day, and 119 households use less than 80 litres per day.

Note: The final value for *F* must equal the total number of observations and the relative frequencies must add up to 1.

Relative frequencies can be expressed as percentages: 36 per cent of households use between 40 and 60 litres per day. As percentages, the relative frequencies must add up to 100 per cent. In Unit 6 you will interpret the relative frequencies as the probabilities that a household uses a certain range of litres.

Cumulative frequencies can be represented by a graph called the **ogive**. It has a characteristic S shape. Relative cumulative frequencies can be used to scale the graph from 0 to 1.

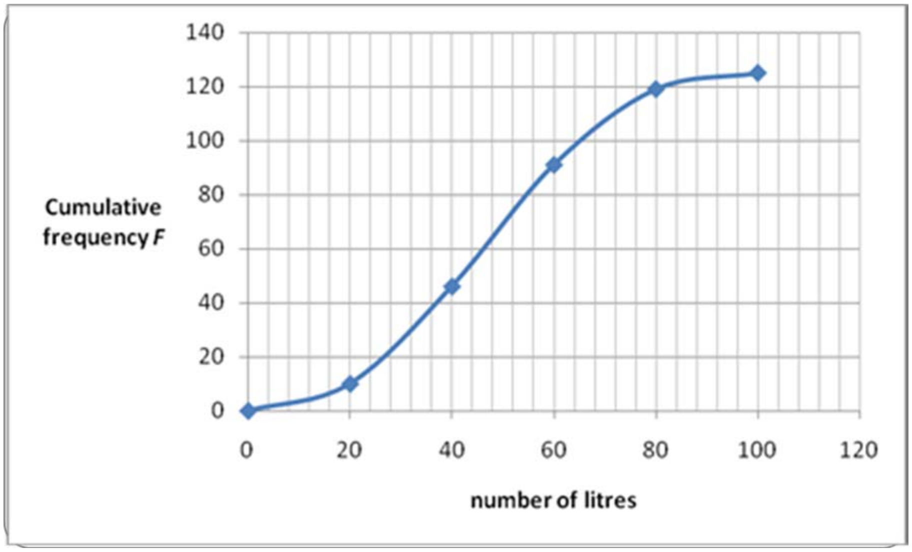


Figure 11

Histograms and ogives are useful tools for analysis, as you will see in the next unit.

Activity 2.2

Activity

Make graphs and tables

What will you do?

1. You want to conduct a survey to find out which new flavour of coffee will sell well in your country. The available flavours are chocolate, hazelnut, cinnamon and rum. Describe how you would go about gathering data.
2. Your supplier in another country has this data for her own country:
   - 36% of people preferred hazelnut,
   - 28% preferred chocolate,
   - 22% preferred rum
   - 8% preferred cinnamon, and
   - the remainder didn't like any of the flavours.

   Present the supplier's findings graphically in two ways. What conclusion can you draw?
3. Draw a frequency table, histogram and ogive for the following data. Of 2,000 adults surveyed:
   - 8% had a mass between 120 kg and 140 kg,
   - 20% had a mass between 100 kg and 120 kg,
   - 36% had a mass between 80 kg and 100 kg,
   - 28% had a mass between 60 kg and 80 kg, and
   - the remainder had a mass between 40 kg and 60 kg.

Activity 2.3

Activity

Understand the terminology

What will you do?

Use this terminology table to record any terms or words you are uncertain about.

This activity is an opportunity to consolidate your understanding of new terminology and concepts you encountered in Unit 5. Fill in the terms you have learned and write your own descriptions of them.

Terminology

Term		**Description**
	:	
	:	
	:	
	:	
	:	
	:	
	:	
	:	
	:	
	:	

Remember these key points

- Data can be collected and analysed for decision making.
- Data that occurs in pairs of numbers can be represented by scatter graphs or line graphs.
- Data organised in categories can be represented by column charts or pie charts. The measured values are on the vertical axis in column charts or displayed as sections of a disc in pie charts.
- Where measured values are grouped into classes and the number of observations for each class is of interest, data can be displayed in frequency tables and histograms. The number of observations or frequency is on the vertical axis. The horizontal axis displays the grouped data, which may be either discrete or continuous.

Unit summary

Summary

You have successfully completed this unit if you can:

- **present** data in tables;
- **draw** graphs, charts, histograms and ogives;
- **distinguish** between discrete and continuous data, and between ungrouped and grouped data; and
- **calculate** cumulative frequencies.

Unit 6

Measures for analysing data

Introduction

Although graphs and tables are important ways of looking at data, more precise quantitative measures are needed for analysing it. Sometimes the volume of data is so big that charts may be difficult to draw. Tracking daily share price movements over years and displaying the data as a scatter graph may indicate trends, but provide no concrete information.

In this unit, you will look at measures of location to identify central values of data and measures of spread, to see how data deviates from the central values. Measures of location are **mean** or **average**, **median** and **mode**; measures of spread are **range**, **standard deviation** and **variance**. You will look at discrete and continuous data, including grouped data (discrete or continuous) with frequency distributions.

Upon completion of this unit you will be able to:

Outcomes

- **calculate** the mean value of discrete data;
- **calculate** the median and mode of discrete data;
- **determine** the mean and median for grouped data with frequency distributions;
- **apply** different measures of spread or dispersion of data;
- **compare** different data sets using co-variance or correlation; and
- **critique** some uses of statistics.

Mean, median and mode

Discrete data

Average (arithmetic mean)

The average or mean value of a set of data is a well-known concept. By average, we usually understand the arithmetic mean. There are also weighted averages and geometric means.

Here's an example:

If the average age of a group of people is 26 years, we understand that there will be people younger than 26 and people older than 26 in the group. In fact, there may be no-one aged exactly 26! If there were four people aged 24 and four people aged 28, then the average age would be 26:

$$\frac{1}{8}[24+24+24+24+28+28+28+28] = 26$$

If the ages of the people were 25, 25, 25, 25, 26, 26, 27 and 27, then the average age would also be 26.

Let's revise:

The general formula for averages can be formulated algebraically: Suppose there are N discrete data values denoted by x_1, x_2– x_3… x_N. The shorthand notation for the data is x_i, $i = 1, 2, 3\ldots N$. The subscripts distinguish the different data values from each other. N can be any number and the values x_i can be any real numbers.

The shorthand notation for the sum of all data values x_i, $i = 1, 2, 3\ldots$ N is: $x_1 + x_2 + x_3 + \ldots + x_N = \sum_{i=1}^{N} x_i$ where ∑ is the Greek letter S and stands for "sum". The average or mean of all the x-values is denoted by:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

This gives some indication of the centre of the data, but although it is much-used, an average on its own can lead to misleading conclusions.

Here's another example:

Say a country has a population of 1 million working people and it is said that the average monthly income is GBP 5,000. Does this tell us much? Not necessarily. It could be that 10,000 people each earn GBP 500,000 per month and all the rest earn nothing, which would be a very unfair distribution of wealth. Or it could be that each person earns GBP 5,000 per month, which would be a more equitable distribution. The average in both cases is GBP 5,000.

Median

The median is the middle value when all the data is arranged from smallest to largest value. When there is an odd number of data points the median will be the actual middle value, and when there is an even number of data points the median will be the average of the two middle values. This can give a better indication of typical values in a data set.

Here's an example:

In the income distribution example, the situation where 10,000 people each earned GBP 500,000 per month and all the rest earned nothing would have a median of 0. (Clearly, the two middle values of a string of 990,000 zeroes followed by 10,000 values of 500,000 will be two zeroes.) Saying that the median income is zero is a better measure here than saying the average income is GBP 5,000.

Mode

The mode is the data value that occurs the most. The easiest way to find the mode is for situations where the discrete data is organised in frequency tables. Data can have several modes, as it is possible for certain data values to occur with the same frequency. The mode is not commonly used to analyse data.

Here's an example:

Students are asked about the number of hours they watch television each day. The data from the survey is displayed in a frequency table.

Number of hours watching TV (*x*)	Number of students *f*(*x*)
1	120
2	254
3	200
4	68
5	12

Figure 12

The mode for this example is two hours.

The other measures may be calculated by noting that there are 654 students in the survey and altogether they watch 1,560 hours per day. The median is the average of the two middle values – that is, the average number of hours for students numbered 327 and 328. These two values are both 2 hours, so the median is two hours.

The average is $\frac{1{,}560}{654} = 2.385$ hours.

The number of observations or frequency is on the vertical axis. The horizontal axis displays the discrete grouped data x_i.

Activity 2.4

Activity

Find the mean, median and mode

What will you do?

1. Find the mean, median and mode of this data set:
   12, 3, 25, 6, 22, 34, 12, 41, 33, 14, 18, 5, 22, 18
2. The frequency table is given for discrete data describing numbers of faulty items produced in a factory in a week, together with the number of flaws found in each item. No item had more than four flaws.

Number of flaws per faulty item (x_i)	Number of faulty items $f(x_i)$
1	62
2	23
3	12
4	4

Figure 13

a) Calculate the mean, median and mode for the distribution of flaws.

b) Derive formulas for the measures for frequency distributions of discrete data.

c) Draw a column chart for the data.

Here's our feedback

1. Mean: $\frac{1}{14}$ [265] = 18.93 (to two decimals) — that is, mean = 19 to the nearest whole number.

   Median: Arrange numbers in ascending order:
   3, 5, 6, 12, 12, 14, 18, 18, 22, 22, 25, 33, 34, 41

   The median is the average of the middle two numbers since there's an even number of data points. Median = ½[18 + 18] = 18.

   Modes are 12, 18 and 22.

2. a) There are four groups of data: $x_1 = 1$, $x_2 = 2$, $x_3 = 3$, $x_4 = 4$. The corresponding frequencies are $f(x_i)$.

   Total number of faulty items produced:
   $= \sum f(x_i) = 62 + 23 + 12 + 4 =101$

   Total number of flaws:
   $= \sum(f(x_i) \times x_i) = (62\times1) + (23\times2) + (12\times3) + (4\times4) = 160$

   Mean number of flaws per faulty item = $\frac{160}{101}$ = 1.58.

   This is almost two flaws per faulty item.

   Median: If you arrange numbers of flaws in ascending order, you get 62 occurrences of one flaw, 23 occurrences of two flaws, 12 occurrences of three flaws and four occurrences of four flaws. This gives a row of 160 numbers with middle numbers at positions 80 and 81. These positions fall into the group with two flaws, because from numbers 63 to 85 you are in the group with two flaws. The median is ½[2 + 2] = 2.

Mode = 1 because most faulty items have one flaw.

b) Mean = $\frac{1}{N}\sum_{i=1}^{N} x_i f(x_i)$ where $N = \sum f(x_i)$

Median: Arrange data values x_i with each x_i repeated $f(x_i)$ times. Find the middle value.

Mode: Choose the value(s) x_i with maximum value for $f(x_i)$.

c)

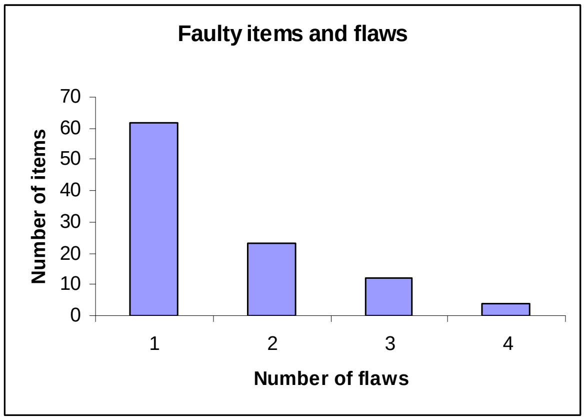


Figure 14

Grouped or continuous data

Remember, for continuous data x you do not have individual observations but a continuum of classes of values for x. You then note the number or frequency of observations in each class.

Average (arithmetic mean)

For the purpose of calculating the mean, we will associate with each observation in a class the value at the midpoint of the class. We add a column with the midpoints of classes, denoted x_i, to the frequency table. Frequency $f(x_i)$ then denotes the number of observations with value (x_i), and $\sum f(x_i)$ is the total number of observations.

Here's an example:

Return to the problem of the frequency distribution for continuous data for the problem of daily consumption of water.

Litres consumed	Midpoint x_i	Number of households Frequency $f(x_i)$
0–20	10	10
20–40	30	36

Litres consumed	Midpoint x_i	Number of households Frequency $f(x_i)$
40–60	50	45
60–80	70	28
80–100	90	6

Figure 15

Therefore it can be assumed that each of the 10 households in the class [0–20] uses 10 litres per day, that each of the 36 households in the class [200–40] uses 30 litres per day, and so on.

In class [00–20] a total of 10 × 10 = 100 litres are used per day, while in class [200–40] a total of 36 × 30 = 1,080 litres are used per day.

If x_i is the midpoint of each class, then:

$$\text{Mean} = \frac{1}{N}\left[\sum x_i f(x_i)\right] \text{ where } N = \sum f(x_i)$$

Each product $x_i f(x_i)$ denotes the area of the rectangle in the histogram with height $f(x_i)$ and the total area of the histogram is $\sum x_i f(x_i)$

For the same example:

$$\text{Mean consumption} = \frac{1}{125}\left[10\times10 + 30\times36 + 50\times45 + 70\times28 + 90\times6\right] =$$

47.44 litres per household per day.

Median

Use the cumulative frequency table to help you find the median. The method is as follows:

There are $\sum f(x_i)$ observations. Find the position of the middle observation. Determine the class in which it falls and its position in the class. Then:

Median = lower limit of class + fraction of width of class to position of observation.

For our example:

Litres consumed (x)	Number of households $f(x)$	Cumulative Frequency $F(x)$
0–20	10	10
20–40	36	46

Litres consumed (x)	Number of households $f(x)$	Cumulative Frequency $F(x)$
40–60	45	91
60–80	28	119
80–100	6	125

Figure 16

There are 125 observations. The middle value is observation number 63. This falls in interval 40–60 because there are 46 observations before class 40–60 and 91 at the end of class 40–60. Since 63 – 46 = 17, the middle observation will take position 17 of the 45 values in class 40–60.

Median = $40 + (\frac{17}{45} \times 20) = 47.56$. Therefore half of the households use less than 47.56 litres and the rest use more than 47.56 litres.

The median and mean are almost equal. This is because the distribution is nearly symmetrical, as we can see from the spread of data in Figures 9 and 16.

Note: An easy but approximate way of finding the median is by using the ogive. Find the halfway position on the vertical axis (cumulative frequency) and then the corresponding value of x on the horizontal axis. Do this using the ogive in Figure 11.

Activity 2.5

Activity

Practice some more

What will you do?

1. Find the mean, median and mode of this set of data:
   5, 2, 6, 4, 10, 7, 9, 11, 4, 7, 8, 3

   Represent the data in a scatter graph.

2. Study the frequency table for continuous data:

Class of values (*x*)	Frequency *f*(*x*)
2.0–4	4
4.0–6	7
6.0–8	15
8.0–10	10
10.0–12	2

Figure 17

a) Expand the table to include cumulative and relative frequencies.

b) Find the mean and median for the distribution.

c) Draw a histogram and ogive.

d) Use the ogive to confirm the value of the median.

Skewness and spread of data

Skewness

The position of the median with respect to the mean gives a measure of the **skewness** of a distribution. Skewness shows how far the distribution of data deviates from symmetry. If the mean and median are the same, the histogram will be symmetrical and the skewness will be zero.

The skewness is proportional to the difference: (mean – median).

In the water consumption example, the difference is:
(47.44 – 47.56) = –0.12

The median is to the right of the mean. This means that more than half the values lie to the right of the mean value. The distribution is not symmetrical, but is negatively skewed.

Spread of data

Now consider range, quartiles and the most-used measures: variance and standard deviation.

Range

Range is one of the measures of the spread of data around a centre. It cannot be used for grouped data.

Range = maximum data value – minimum data value.

Here's an example:

For question 1 in Activity 2.5, range = 41 – 3 = 38.

One problem with using the range as a measure is that a single value lying far away from the rest of the values would result in a very big range as the measure of spread. This may not be a reasonable reflection of the spread of data.

Quartiles

Quartiles can be used for grouped data. Let Q_1 denote the value below which 25% of observations are found, Q_2 the value below which 50% of observations are found (Q_2 is the median) and Q_3 the value below which 75% of observations are found.

Then the interval between Q_1 and Q_3 contains 50% of observations. The bigger the value of $(Q_3 – Q_1)$, the wider is the spread of data around the median Q_2. This spread around the median is described by quartile deviation:

Quartile deviation = $\dfrac{Q_3 - Q_1}{2}$

Variance

Variance measures the spread of data values x_i ($i = 1, 2, 3... N$) about the mean $\bar{x}$ and can be used for ungrouped or grouped data. It is defined as:

variance $(x) = s^2 = \dfrac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}$ for ungrouped discrete data values x_i.

variance $(x) = s^2 = \dfrac{\sum_{i=1}^{N}(x_i - \bar{x})^2 f(x_i)}{\sum_{i=1}^{N} f(x_i)}$ for grouped data values with midpoint values x_i of classes and frequencies $f(x_i)$.

Variance is additive for uncorrelated data sets: This means if you have the variance of data set A and the variance for uncorrelated data set B, then the variance for the combined data set is the sum of the individual variances.

Here's an example:

If the variance in the mass of a person is 20 kg, then the variance in mass for six people is 6 × 20 = 120 kg. This is because the measurements of the mass of people are uncorrelated.

Standard deviation

Standard deviation (std dev) also measures the spread of data about the mean. It is defined as the square root of variance: std dev (x) = $\sqrt{s^2}$ and is denoted by the symbol s.

If in finance you collect data on the daily or monthly returns on share prices, the standard deviation of the returns is known as the volatility of the share price. This is a very important measure in the risk management of share portfolios. The units of the measure are percentages because returns are measured as percentages.

Reminder: If a share price changes from 20 to 16 over a single day, the daily return is $\frac{16-20}{20}$ = –0.20 = –20%.

Sample versus population data

If you want to find the variance for a large population for which you have collected limited data, you must adapt your formula. If the sample size is N, use the formula:

variance (x) = $s^2 = \dfrac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}$ for the variance of the population. The effect of dividing by $N - 1$ instead of by N is to make the variance larger. This is important because, by taking a sample you may have omitted large values in the population that would have given a larger variance.

Co-efficient of variation

This measure is used to compare the spread of two data sets. Unless they have the same mean, you cannot easily compare standard deviations of two sets to come to a conclusion. The co-efficient of variation does this:

Co-efficient of variation = $\dfrac{std\ dev}{mean} = \dfrac{s}{\bar{x}}$

Co-variance of two ordered sets of data

The cases where measurements are linked to units of time are particularly interesting. For each unit of time t we have a data value $x(t)$, and the order of the measurements is important. The data would be presented as a scatter graph, with time on the horizontal axis. We consider two sets of measured values for a situation such as share prices, sales figures and so on, over a time period.

Here's an example:

The share prices of two shares, 1 and 2, are tracked over 10 days. The data is presented in Figure 18. Share price S_1 is indicated by Δ and share price S_2 is indicated by ♦.

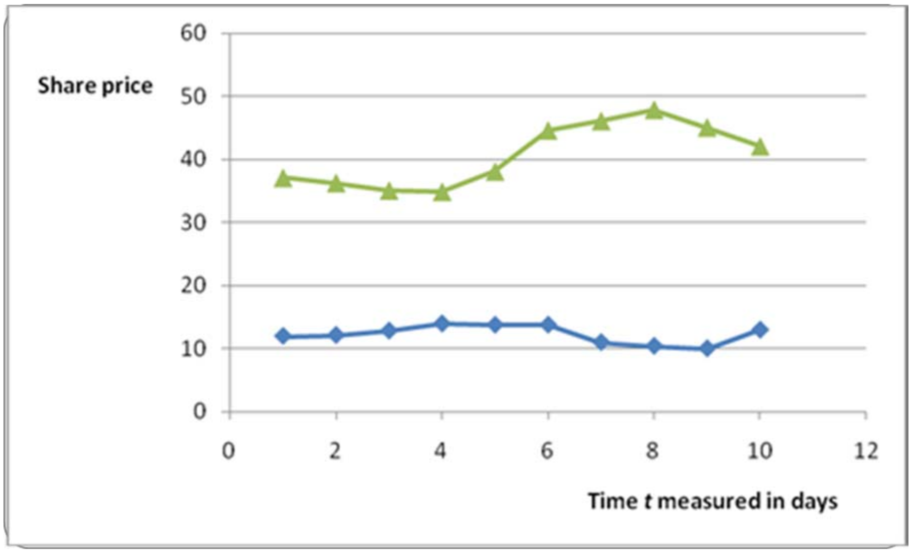


Figure 18

In these cases, it is clear that the data is ordered from day 1 to day 10, and this order is important.

Co-variance and correlation co-efficient

If you have two ordered sets of data for a specific situation, you can see whether they are correlated in any way. For example, sales figures for cold drinks and for sunglasses are collected each month for a year; or the daily prices of two shares are tracked over a number of days, as shown in Figure 18. The question is whether there is any correlation in the movements of the sales figures for cold drinks and sunglasses, or in the movements of the two share prices. This can be measured in terms of the co-variance between the two sets of data: let the two sets of values be x_i and y_i respectively, where $i = 1, 2, 3\ldots N$.

$$\text{Covariance } (x, y) = \operatorname{cov}(x, y) = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{N}$$

If the co-variance is positive, then the two sets of variables tend to move up or down together. If the co-variance is negative then the two sets of variables tend to move in opposite directions. As one rises above its mean, the other tends to drop below its mean. If the co-variance is zero, then the two sets of variables are uncorrelated.

The co-variance can also be normalised to give the correlation co-efficient $\rho\,(x, y)$:

$\rho\,(x,y) = \dfrac{\operatorname{cov}(x, y)}{s(x)s(y)}$ where $s(x)$ and $s(y)$ are the standard deviations of the x- and y-values, respectively. The correlation co-efficient is always between -1 and $+1$, with -1 denoting perfect negative correlation and $+1$ denoting perfect positive correlation. A value of zero indicates no correlation.

Activity 2.6

Activity 2.6

Consolidate these skills

What will you do?

1. Calculate the skewness, range, variance and standard deviation of this data: 12, 10, 14, 12, 15, 8, 11, 16, 21, 14, 43

   Discuss the advantages and disadvantages of the measures.

2. Determine the variance and standard deviation of the frequency distribution for continuous data *x* in the table:

Class (*x*)	Frequency *f*(*x*)
2.0–4	4
4.0–6	7
6.0–8	15
8.0–10	10
10.0–12	2

Figure 19

3. The data for Figure 18 is given in this table:

Day	Share price S_1	Share price S_2
1	12.00	37.00
2	12.15	36.20
3	12.85	35.00
4	14.00	34.80
5	13.80	38.00
6	13.80	44.50
7	11.00	46.00
8	10.40	47.80
9	10.00	45.00
10	13.00	42.00

Figure 20

a) Calculate the mean share price for each of S_1 and S_2 over the 10 days.

b) Determine the standard deviation and co-efficients of variation of the share price in each case.

c) Calculate the co-variance and correlation co-efficient for S_1 and S_2.

d) Interpret and discuss your results.

Here’s our feedback

1. Skewness is proportional to the difference (mean – median). Mean = 16 and median = 14. Skewness is positive: More than half the values lie to the left of the mean value. The distribution is not symmetrical.

   Range = 43 − 8 = 35. Variance = $s^2 = \frac{1}{11}\sum(x_i - 16)^2$ = 83.64 and standard deviation = s = 9.145. A standard deviation of about 9 units (around the mean value of 16) is a better reflection of the spread of data than the range value of 35. The value 43 is clearly an outlier explain ‘outlier’? and possibly an incorrect measurement.

2. Take the midpoints of the classes to be 3, 5, 7, 9 and 11. The mean is $\frac{1}{38}[\ \sum x_i f(x_i)\ ]$ = 6.95 (rounded off)

   $$\text{Variance }(x) = s^2 = \frac{\sum_{i=1}^{N}(x_i - \overline{x})^2 f(x_i)}{\sum_{i=1}^{N} f(x_i)}$$

   $$= \frac{1}{38}[62.41 + 26.62 + 0.038 + 42.03 + 32.81] = 4.31$$

   Standard deviation = s = 2.08 (rounded off)

3. a) $\overline{S_1}$ = 12.3; $\overline{S_2}$ = 40.63

   b) Std dev(S_1) = 1.45, std dev(S_2) = 4.96, co-efficient of variation for S_1 : 0.118, co-efficient of variation for S_2 : 0.122

   c) Co-variance(S_1, S_2) = −3.91, correlation co-efficient = −0.6

d. Prices for share 1 are less spread out than prices for share 2. Share 1 is therefore probably less risky than share 2. The negative correlation shows that as one share price tends to move up, the other share price tends to move down.

Case study:
Investment decisions

You are analysing the daily price movements of two shares, with a view to making a recommendation about buying one, or both, of the shares.

These numbers are the share prices (in dollars) of shares S1 and S2 collected daily at the close of the trading day for 21 days in February 2009.

Day	1	2	3	4	5	6
S_1	12,00	12,15	12,85	14,00	13,80	13,80
S_2	37,00	36,20	35,00	34,80	38,00	44,50

Day	7	8	9	10	11
S_1	11,00	10,40	10,00	13,00	15,50
S_2	46,00	47,80	45,00	42,00	39,80

Day	12	13	14	15	16
S_1	16,50	16,00	15,00	14,50	14,20
S_2	35,00	48,00	52,40	54,00	54,20

Day	17	18	19	20	21
S_1	13,00	12,20	13,00	14,00	14,20
S_2	56,00	58,20	52,40	51,00	50,00

The daily rate of return for S_1 at the end of day two is calculated as:

$$\frac{12{,}15 - 12{,}00}{12{,}00} = 0{,}0125 = 1{,}25\%$$

Similarly, the daily rate of return for S_2 for day 10 is:

$$\frac{42{,}00 - 45{,}00}{45{,}00} = -\,0{,}067 = -\,6{,}7\%$$

1. Draw a table to show the daily prices and the 20 daily rates of return for each share.
2. Use graphs to present your data in an attractive way.
3. Find the mean daily rates of return over the month for each share.
4. Calculate the one-day standard deviation of rates of return for each share over the month. Use this to estimate the monthly volatility of each share.
5. Calculate the correlation between the two share price processes, as well as the correlation between the returns for the month.
6. Do some research on the topic of "diversification in investments" and write a short report on this. Show how diversification is linked with correlation of share prices.
7. Write a short paragraph comparing the two shares in terms of expected returns and volatility. What recommendations will you make for investment in these shares?

Activity 2.7

Activity

Understand the terminology

What will you do?

Use this terminology table to record any terms or words you are uncertain about.

This activity is an opportunity to consolidate your understanding of new terminology and concepts you encountered in Unit 6. Fill in the terms you have learned and write your own descriptions of them.

Terminology

Term Description

:

:

:

:

:

:

:

:

Remember these key points

- Location of data is measured by mean, median and mode.
- The average or mean of all the x-values is denoted by:
  $$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$
- The median is the middle value when all the data is arranged from smallest to largest value. The mode is the value occurring most often.
- For grouped data: Mean = $\frac{1}{N}\sum_{i=1}^{N} x_i f(x_i)$ where $N = \sum_{i=1}^{N} f(x_i)$
- If data is continuous, then the x_i are the midpoints of the classes.
- Median = the lower limit of the class in which the median will fall + fraction of the width of class to the position of middle value.
- The ogive curve can be drawn from the cumulative frequency table and used to find the median. Spread of data is measured by range, quartiles, variance and standard deviation.
- Variance (x) = $s^2 = \frac{1}{N}\sum (x_i - \bar{x})^2$ for ungrouped data values x_i.
- Variance (x) = $s^2 = \frac{\sum (x_i - \bar{x})^2 f(x_i)}{\sum f(x_i)}$ for grouped data with midpoint values of classes x_i and frequencies $f(x_i)$.
- Data can be ordered by linking it to time. The movement of two sets of data x and y over time can be compared by calculating the co-variance $\sigma(x, y)$or correlation co-efficient $\rho(x, y)$:
  $$\sigma(x,y) = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{N}$$
  $$\rho(x,y) = \frac{\sigma(x,y)}{s(x)s(y)}$$
- Statistics cannot be used without being aware of the limitations of, and dangers in, interpreting the values of statistical measures.

Unit summary

Summary

You have successfully completed this unit if you can:

- **understand** the location and spread of data;
- **calculate** mean, median and mode for ungrouped discrete data;
- **determine** the mean and median for grouped data;
- **calculate** variance and standard deviation for data;
- **measure** co-variance and correlation co-efficients for two sets of data; and
- **critique** the use of statistics in interpreting data.

Unit 7

Basic probability theory

Introduction

Up to now, data and variables have been considered from the point of view of quantities that are known with certainty. If you have a list of last year's sales figures, you are dealing with known numbers; if you have a formula relating profit *P* to quantity sold *Q*, you are dealing with deterministic variables: If you know *Q,* you can determine *P* immediately and exactly.

In this unit, you will look at variables from the point of view of uncertainty. Such variables are called **random** or **stochastic** variables. Looking ahead, we do not know what next year's sales figures for automobiles will be or the future profit of a company or the number of unemployed people in a country in five years' time – but we are not entirely ignorant either.

The topic of this unit is to formalise the knowledge that does exist about random variables in terms of probabilities. Probability theory was developed in the 17th century by Blaise Pascal and Pierre de Fermat and today is an important part of decision making.

Upon completion of this unit you will be able to:

Outcomes

- **describe** outcomes and events for experiments;
- **calculate** probabilities for outcomes and events;
- **understand** conditional probability and Bayes' formula; and
- **identify** mutually exclusive and independent events.

Outcomes and probabilities

An action such as tossing a coin or dice, or projecting sales figures can be called an experiment. An experiment produces a number of possible outcomes that can be specified beforehand. The set of all possible outcomes is called the **sample space** and denoted by Ω.

What we cannot specify beforehand is which one of the outcomes will actually happen. But we can specify beforehand (with some work) what the probability *P* of each outcome is. The probability of something happening can be seen as the likelihood (sometimes called chance) or relative frequency of it happening, based on historical data.

Here are some examples:

1. The umpire at a cricket match tosses a coin to decide which captain may choose whether to bat or bowl first. West Indies captain Chris Gayle chooses heads. The outcome of this “tossing a coin experiment” is uncertain. But we know the outcome will be either heads or tails, and if the coin is fair (balanced) we say there is a 50 per cent likelihood (chance) of Gayle winning the toss. The probability of getting heads is *P*(heads) = ½ or 50 per cent and the probability of getting tails is *P*(tails) = ½ or 50 per cent.
2. Next year’s profit for your company is uncertain. However, based on past data and making some assumptions about the world economy, you may be able to predict a 70 per cent probability that the profit will be between GBP 1.2 million and GBP 2.1 million.

A probability is a number between 0 and 1. A certainty has a probability of 1 and an impossible outcome or event has a probability of 0. Probabilities can also be expressed as percentages between 0 per cent and 100 per cent.

The set of all outcomes with their associated probabilities, {Ω, *P*}, is called the probability space. Each outcome can be seen as an observation. A set of selected outcomes can be seen as a group or class of observations and forms an event.

Calculating probabilities

The probability of a specific outcome is:

$$P(\text{outcome}) = \frac{\textit{number of observations of this outcome}}{\textit{total number of outcomes}}$$

For an event:

$$P(\text{event}) = \frac{\textit{number of outcomes for event}}{\textit{total number of outcomes}}$$

In the language of Unit 6:

$$P(\text{outcome or event } x_i) = \frac{f(x_i)}{\sum f(x_i)}$$

The probability of something happening = *P*(Ω) = 1.

Every event *E* has a complement E^c, which is the event containing all outcomes of Ω that *are not in E*. Remember that events are sets or collections of outcomes. We must then have:
$E \cup E^c = \Omega$. The symbol ∪ denotes the union of sets.
Since either *E* or E^c must occur, we have $P(E) + P(E^c) = 1$.

Here’s an example:

A fair dice is thrown (this is the experiment). The number showing on the top face is the outcome. The set of possible outcomes is Ω = {1, 2, 3, 4, 5, 6}. Note that curly brackets { } are used to denote sets of objects. The total number of outcomes is six. The probability of throwing a 1 (getting outcome 1) is denoted by *P*(1).

$P(1) = 1/6$, $P(2) = 1/6$, $P(3) = 1/6$, and so on. The sum of all the probabilities is 1.

We can define events such as the event of throwing an even number = {2, 4, 6}, the event of throwing a number greater than two = {3, 4, 5, 6} and so on.

P(event "throw an even number") = 3/6 = 1/2, because there are three possible even numbers and six possible outcomes.

P(event "throw an odd number ") = 3/6 = 1/2, because there are three possible odd numbers and six possible outcomes.

P(event "throw a number greater than 2") = 4/6.

P(event "throw a number greater than 0") = 6/6 = 1, because there are six possible numbers greater than zero on the dice and six possible outcomes. You are certain to throw a number greater than 0: *P*("certainty") = 1.

P(event "throw a number greater than 6") = 0/6 = 0, because there are no numbers greater than six on the dice and six possible outcomes. You are certain to never throw a number greater than six: it is impossible and *P*(event "impossible") = 0.

Dependent, independent and mutually exclusive events

And/or calculations

Sometimes you need the probability that *either* of two events *E* or *F* will occur — that is, the probability that an outcome will be realised in either of them. You write this as: $P(E \text{ or } F) = P(E \cup F)$. (Mathematically, "or" means the union of sets, which is denoted by the symbol "$\cup$".)

Sometimes you need the probability that *both* events *E* and *F* will occur — that is, the probability that an outcome will be realised that is in both of them. You write this as: $P(E \text{ and } F) = P(E \cap F)$. (Mathematically, "and" means the intersection of sets, which is denoted by the symbol "$\cap$". The intersection of two sets is those outcomes that lie in both sets.)

How do you calculate these probabilities?

Mutually exclusive events

Mutually exclusive events do not have any outcomes (observations) in common. They are called disjoint and their intersection is empty. An event *E* and its complement E^c are mutually exclusive. The event "throw an odd number" and the event "throw an even number" are mutually exclusive. The event "throw a number greater than two" and the event "throw an even number" are not mutually exclusive, since they both contain outcomes four and six.

If *E* and *F* are mutually exclusive or disjoint events, then:
$P(E \text{ or } F\ ) = P(E \cup F) = P(E) + P(F)$

If *E* and *F* are not mutually exclusive, then:
$P(E \cup F) = P(E) + P(F) - P(E \cap F)$

You have to deduct the outcomes they have in common, otherwise you will be adding some probabilities twice.

Independent and dependent events

These are more tricky concepts. The outcomes in the sample space are assumed to be independent of each other. The presence of any one outcome should not affect the presence of another outcome in Ω.

Events can be either dependent or independent. Independent is not the same as mutually exclusive. In particular, complementary events (which are, by definition, mutually exclusive) are always dependent. For example, the events "throw an even number" and "throw an odd number" are mutually exclusive but not independent. If I throw an even number, it precludes me from throwing an odd number and so there is a dependency. The events E_1 = {"throw an even number"} and E_2 = {"throw a number greater than two"} are independent of each other but not mutually exclusive.

If E and F are independent events, then:
$P(E \text{ and } F) = P(E \cap F) = P(E) \times P(F)$

For independent events E_1 and E_2 above, you find:
$P(E_1 \text{ and } E_2) = P(E_1 \cap E_2) = P(E_1) \times P(E_2) = ½ \times ⅔ = ⅓$

This is correct because: $E_1 \cap E_2 = \{4, 6\}$ so that $P(E_1 \cap E_2) = 2/6$

On the other hand, $P(E_1 \text{ or } E_2) = P(E_1 \cup E_2) \neq P(E_1) + P(E_2)$ because E_1 and E_2 are not mutually exclusive, but:
$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) = ½ + ⅔ - ⅓ = ⅔$

If E and F are not independent, then the above formula does not necessarily hold and you will have to calculate $P(E \cap F)$ from first principles.

Note: $P(E \cap F)$ is often written as $P(EF)$.

Probability trees

Simple experiments that are repeated over time can be expressed graphically in the form of a tree. At each time step, the tree branches out into the various possible outcomes. This graphical representation can be very useful when calculating probabilities or trying to visualise a sequence of decisions. Decision trees are discussed in Module 4.

Here's an example:

A coin is tossed three times. Represent this as a tree:

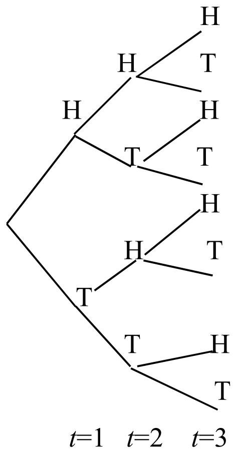


Figure 21

Each toss is independent. At each point (node), the tree branches into heads (*H*) or tails (*T*). Note that the experiment has eight outcomes: there are eight possible sequences of heads and tails: *HHH, HHT, HTH. HTT*. and so on. Each outcome (sequence) has a probability of 1/8 of occurring.

Each separate up or down branch carries a probability of ½. Therefore the probability of *TTT* can also be calculated as $P(TTT) = ½ × ½ × ½ = 1/8 = 0.125$. The occurrence of each *T* is independent of the others.

Activity 2.8

Activity

Calculate probability

What will you do?

1. What is the probability of getting an odd number smaller than five when throwing a dice?
2. Let E be an event with $P(E) = k$, $k \neq 0$. Prove that complementary events E and E^c must be dependent.
3. A fair coin is tossed 10 times. What is the probability of getting tails every time?
4. What is the probability of picking a jack of hearts from a pack of cards? What is the probability of picking any jack from a pack of cards?

Here's our feedback

1. The sample space is $\{1, 2, 3, 4, 5, 6\}$. The event "get an odd number smaller than five" is the set $\{1, 3\}$. The probability is $P(\{1, 3\}) = 2/6 = 1/3$ or 33.33%. Or: Let E be the event "get an odd number" and let F be the event "get a number smaller than five". The required event is $E \cap F$. Since they are independent:
   $P(E \cap F) = P(E\ ) \times P(F) = ½ \times 4/6 = 1/3$

2. Suppose they are independent. Then the formula
   $P(E \cap E^c) = P(E) \times P(E^c)$ must hold true. But $E \cap E^c$ is an empty set (impossible event) because complementary events have nothing in common. So (i): $P(E \cap E^c) = 0$

   You know $P(E) = k$, $0 < k < 1$ and $P(E^c) = 1 - P(E) = 1 - k$. So (ii):
   $P(E \cap E^c) = P(E) \times P(E^c) = P(E)(1 - P(E)) = k(1 - k) \neq 0$

   You get both (i) $P(E \cap E^c) = 0$ and (ii) 0 $P(E \cap E^c) \neq 0$, which is impossible. Therefore your supposition that E and E^c are independent is wrong. You have proved that E and E^c must be dependent.

3. All tosses are independent of each other. $P(\text{tails}) = ½$.

   $P(\text{tails 10 times in a row}) = ½ \times ½ \times ½\ldots \times ½$ (10 times)
   $= (½)^{10}$
   $= 0.000977$

4. There are 52 cards in a pack. There is one jack of hearts. The probability of picking it is $\frac{1}{52}$. There are 4 jacks of different suits.

   The probability of picking any one is $\frac{4}{52} = \frac{1}{13}$.

Activity 2.9

Activity

Calculate more probabilities

What will you do?

1. Event E has probability 0.65 and event F has probability 0.3.

   a) If the sample space Ω has 200 outcomes, how many observations are there in each event?

   b) What is $P(E^c)$? Calculate $P(E \cup E^c)$ and discuss your answer. Calculate $P(E \cap E^c)$ and discuss your answer.

   c) If E and F are independent, determine $P(EF)$.

2. You have 120 staff members: 86 female and 34 male. There are five per cent of staff members who are disabled. What is the probability that a staff member is female and disabled?

3. What is the total number of outcomes in having to choose any three letters from the English alphabet? What is the total number of outcomes in choosing three letters from the English alphabet if you may only choose each letter once?

4. A lottery requires you to pick five different numbers, each from the set of numbers from 1 to 20. What is the probability of picking the correct number? How does the probability change when you can choose from numbers 1 to 50?

5. The share price today is USD 20. At the end of the week, it will go either up by 15 per cent with probability 0.3 or down by 10 per cent with probability 0.7, respectively. At the end of the following week, the then price will go either up or down by 10 per cent with probabilities 0.5 and 0.5, respectively. Draw a probability tree. What is the probability that the share price will be USD 19.80 at the end of the second week?

Conditional probability and Bayes' formula

Two or more events on a sample space can be dependent. This means knowledge of the occurrence of one event affects the probability of the other event. Look, for example, at the data in the following table:

	Full-time	Part-time
Male students	50	24
Female students	62	77

The table shows there are 74 male and 139 female students in a group of 213, some of whom are full-time (112) and some part-time students (101).

Four events can be considered: M (male), F (female), FT (full-time), PT (part-time). The events are not all independent: If you know a student is male, his probability of being a full-time student is affected by this.

a) $P(FT)$ = number of full-time students / total number = (112/213) = 0.53 (rounded)

b) But $P(FT$, *given that the student is male*$)$ = (50/74) = 0.68

This probability is denoted by $P(FT \mid M)$. It is called a conditional probability and read as: the probability that event *FT* occurs, *given that* event *M* occurs.

c) P(F) = (139 / 213) = 0.65

d) But P(F | PT) = (77 / 101) = 0.76

Calculation: How do you calculate conditional probabilities in general? The formula, for general events *E* and *F*, is:

$$P(E \mid F) = \frac{P(EF)}{P(F)}$$

Explanation: Both events *E* and *F* must occur (that is, the intersection *EF*), but you must work relative to event *F* because that is the given event.

So, in case *d*): $P(F \mid PT) = \dfrac{P(FandPT)}{P(PT)} = 77 / 101$

Note: Clearly, if *E* and *F* are independent then $P(EF) = P(E)\,P(F)$, and so $P(E \mid F) = \dfrac{P(E)P(F)}{P(F)} = P(E)$. This makes sense. If *E* and *F* are independent, then the occurrence of *F* does not affect the occurrence of *E*.

Here's an example:

Two dice are thrown and the outcomes denoted by pairs of numbers: (2, 4) means that the first dice showed two and the second showed four. As an exercise, write down all 36 possible outcomes. Let *E* denote the event that the first dice shows three. Let *F* denote the event that the second dice shows a six.

$E = \{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}$.

$F = \{(1, 6), (2, 6), (3, 6), (4, 6), (5, 6), (6, 6)\}$.

Then Then $P(E) = \dfrac{6}{36} = \dfrac{1}{6}$ because there are 36 outcomes in total and *E* has six outcomes.

And the conditional probability that the first dice shows three, *given that* the second dice shows six, namely $P(E \mid F)$, is also $P(E)$ $(= \dfrac{1}{6})$

You should be able to see that also $P(E)$ $(=\dfrac{1}{6})$ because if you work relative to event *F*, you have six possible outcomes and outcome (3, 6) is one of the six.

When you make decisions based on probabilities, it is important to know which events are conditional upon others.

Bayes' formula

Suppose F_1, F_2… F_N are mutually exclusive events that combine to form sample space Ω. Suppose E is another event in Ω (overlapping with and therefore conditional on some of the F_i). We assume we know the probabilities $P(F_i)$ and $P(E \mid F_i)$ for each $i = 1, 2\dots N$.

Then $P(E) = \sum_{i=1}^{N} P(E|F_i)P(F_i)$

Bayes' formula states that for any specific k:

$$P(F_k \mid E) = \frac{P(E|F_k)P(F_k)}{\sum_{i=1}^{N} P(E|F_i)P(F_i)}$$

Here's an example:

There are three drawers in your cabinet. A file is equally likely to be in one of the drawers. You are in a hurry to find the file and you quickly look in each drawer. Let the probability be 0.2 that you will find the file in drawer 1 if you quickly look in drawer 1, given that the file is, in fact, in drawer 1. Suppose you quickly look through drawer 1, but do not see the file. What is the probability that the file is actually in drawer 1?

The mutually exclusive events are that the file is in drawer 1, or in drawer 2, or in drawer 3. Let F_i be the event that the file is in drawer i for $i = 1, 2, 3$. Then $P(F_i) = ⅓$ for every i.

Let E be the event that you quickly look but do not see the file in drawer 1. We want to find $P(F_1 \mid E)$.

$$P(F_1 \mid E) = \frac{P(E|F_1)P(F_1)}{P(E|F_1)P(F_1) + P(E|F_2)P(F_2) + P(E|F_3)P(F_3)}$$

$$= \frac{0.8 \times (1/3)}{0.8 \times (1/30 + 1 \times (1/3) + 1 \times (1/3)}$$

$$= 0.286$$

Therefore there is almost a 29 per cent chance that the file is in drawer 1, given that you did not see it there.

Activity 2.10

Activity

Consolidate probability

What will you do?

1. Suppose events E and F are independent, with $P(E) = 0.3$ and $P(F) = 0.55$. Calculate:

   a) $P(E \cup F)$

   b) $P(E \mid F)$

2. You have two managers in your division. What is the conditional probability that both are female (f), given that at least one of them is female?

Hint: The sample space is $\Omega = \{(f, f), (f, m), (m, f), (m, m)\}$. Let E be the event that both are female and F the event that at least one of them is female.

3. Your staff are given a competency test in the form of a multiple-choice test. Each question has four answers to choose from. You want to evaluate their performance on a specific question. A person either knows the answer to this question or guesses. Assume that the probability of a person knowing the answer is 0.5 and the probability of them guessing is 0.5. Assume also that the probability that a person who guesses gets the answer correct is ¼.

   What is the conditional probability that a person really knew the answer to the question, given that she or he answered it correctly?

Hint: Use Bayes' formula, with E being the event that they actually knew the answer and F the event that they gave the correct answer.

Activity 2.11

Activity

Understand the terminology

What will you do?

Use this terminology table to record any terms or words you are uncertain about.

This activity is an opportunity to consolidate your understanding of new terminology and concepts you encountered in Unit 7. Fill in the terms you have learned and then write your own descriptions of them.

Terminology

Term		**Description**
	:	
	:	
	:	
	:	
	:	
	:	
	:	
	:	
	:	
	:	
	:	

Remember these key points

Summary of definitions in probability

Experiment	An action that produces observable results. For example, tossing a coin or throwing a dice.
Outcomes	The results of the experiment. For example: The two possible outcomes for tossing a coin are heads H and tails T. The six outcomes for throwing a dice are 1, 2, 3, 4, 5 and 6.
Sample space	The set or list of all possible outcomes: $\Omega_{coin} = \{ H, T\}$ is the sample space for coin tossing. $\Omega_{dice} = \{1, 2, 3, 4, 5, 6\}$ is the sample space for the dice-throwing experiment.
Events	An event is a subset of the sample space. For example: Getting heads is the event. $E = \{H\}$ Throwing an even number on dice is the event E, with three outcomes. $E = \{2, 4, 6\}$ Throwing a seven on dice is the empty (impossible) event. $E = \{ \}$
Mutually exclusive events	Mutually exclusive events do not have any outcomes (observations) in common. An event E and its complement E^c are mutually exclusive. If E and F are mutually exclusive events, then: $P(E \text{ and } F) = P(E \cap F) = 0$ and $P(E \text{ or } F) = P(E \cup F) = P(E) + \mathrm{P}(F)$ Otherwise: $P(E \text{ or } F) = P(E \cup F) = P(E) + \mathrm{P}(F) - P(E \cap F)$
Dependence and independence	Independent events do not affect each other. If E and F are independent events, then: $P(E \text{ and } F) = P(E \cap F)$ $= P(EF)$ $= P(E) \times \mathrm{P}(F) = P(E)P(F)$
Probabilities of outcomes	$P(\text{outcome}) = \dfrac{\textit{number of observations of outcome}}{\textit{total number of outcomes}}$ Coin tossing: $P(H) = P(T) = ½ = 0.5$ Dice: $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$ In each case, all outcomes have the same probability

	and are called *equally likely.* In each case, the sum of probabilities of all outcomes = 1.
Probabilities of events	$P(\text{event}) = \dfrac{\textit{number of outcomes for event}}{\textit{total number of outcomes}}$
Probability of sample space	P(even dice number event) = 3/6 = ½ P(odd dice number event) = 3/6 = ½
Probability of impossible events	$P(\Omega)$ = 1 for every sample space Ω P(7 on dice) = 0

- Conditional probabilities are calculated when one event can be affected by the occurrence of another event. The probability of E given the occurrence of F is $P(E|F) = \dfrac{P(EF)}{P(F)}$
- In this case, Bayes' formula is often useful.
- Bayes' formula states that for any specific k:

$$P(F_k \mid E) = \frac{P(E|F_k)P(F_k)}{\sum_{i=1}^{N} P(E|F_i)P(F_i)}$$

Unit summary

Summary

You have successfully completed this unit if you can:

- **construct** outcomes and events for experiments;
- **calculate** probabilities for outcomes and events;
- **explain** conditional probability and Bayes’ formula; and
- **work** with mutually exclusive and independent events.

Unit 8

Probability distributions and applications

Introduction

Now you will look at probabilities from the perspective of random variables. A **random variable** has an uncertain value corresponding to the outcome in an experiment. If you toss a fair coin three times and let X denote the number of tails that occurred, then X is variable and uncertain. X is defined on sample space Ω where:
$\Omega = \{ HHH, HHT, HTT, HTH, THH, TTH, THT, TTT\}$

X can take on values 0, 1, 2 or 3. In particular, $X(HHH) = 0$, $X(HHT) = 1$, and so on.

The profit of a firm is a random variable. The number of rainy days in a year is a random variable. Random variables may be **discrete** (number of days, number of heads or tails, and so on) or **continuous** (amount of electricity consumed, rate of return, and so on).

In all these cases, our ignorance is not complete. We have theoretical and empirical knowledge (for example, data studies) that could be used to determine the distribution of the random variable. This is similar to what was covered in Units 5 and 6, but at a higher level of sophistication. Although you may not know which value random variable X will assume, you may know the probability that it will assume a certain value.

In Units 5 and 6, frequency tables were set up, sometimes adding a relative and cumulative frequency column, and column charts (for discrete data) or histograms (for continuous data) were drawn. The classes for grouped data are the events, the relative frequencies are just the probabilities of being in a class (event), and the histograms show the form of the distribution.

Here, random variable X is considered as a function from the space Ω into the set of real numbers. The aim is to construct distribution functions that describe random variables and enable you to calculate mean values and the spread of values of random variables.

Upon completion of this unit you will be able to:

Outcomes

- **identify** discrete random variables and their probability mass functions;
- **apply** binomial distribution and Poisson distribution;
- **identify** continuous random variables;
- **understand** the normal and uniform distributions and their density functions; and
- **apply** the normal distribution in market research by **calculating** confidence intervals and **applying** hypothesis tests.

Discrete random variables

In this case, the variable X can take on various possible discrete numerical values denoted by $x_1, x_2\ldots x_N$. You construct the probability distribution of a discrete random variable X as follows:

1. Find all the possible different values of X on Ω .That is, find all values x_i.
2. Calculate the probabilities and define the **probability mass function** $\boldsymbol{p}$ for X: $p(x_i) = P(X = x_i)$.
3. Define the **cumulative distribution function** $\boldsymbol{F}$ for X:

   $F(x_i) = P(X \le x_i)$
4. Draw graphs for p and F.
5. Find the mean and variance of the distribution of X.

Note: Mean = $\overline{X} = \sum x_i\, p(x_i)$ summed over all possible values x_i.

Variance = $\mathrm{var}(X) = \sigma^2(X) = \overline{X^2} - (\overline{X})^2$ where $\overline{X^2} = \sum x_i^2\, p(x_i)$.

Here's an example:

Apply these five steps to random variable X where X describes the number of heads when a coin is tossed three times.

1. $\Omega = \{HHH, HHT, HTT, HTH, THH, TTH, THT, TTT\}$.The corresponding numbers of heads are $\{3, 2, 1, 2, 2, 1, 1, 0\}$. Therefore the possible values x_i of X are $x_1 = 0, x_2 = 1, x_3 = 2, x_4 = 3$.

   There are eight possible outcomes. (Note that $8 = 2^3$.)
2. Probability $P(X = 0) = p(x_1)$ is the probability that there is no head in the sequence of tosses. Since there is only one outcome satisfying this, namely (TTT), the probability is $p(x_1) = \frac{1}{8}$. In the same way we determine the probabilities $p(x_2)$, $p(x_3)$, and $p(x_4)$. The probability mass function p is defined by:

   $p(x_1) = \frac{1}{8}$; $p(x_2) = \frac{3}{8}$; $p(x_3) = \frac{3}{8}$; $p(x_4) = \frac{1}{8}$

   Note that the probabilities add up to 1.
3. The cumulative distribution function F for X is:

   $F(x_1) = P(X \le x_1) = \frac{1}{8}$; $F(x_2) = P(X \le x_2) = \frac{4}{8}$;

   $F(x_3) = P(X \le x_3) = \frac{7}{8}$; $F(x_4) = P(X \le x_4) = \frac{8}{8} = 1$

4. Draw a histogram for *p* and a histogram-type ogive for *F*.

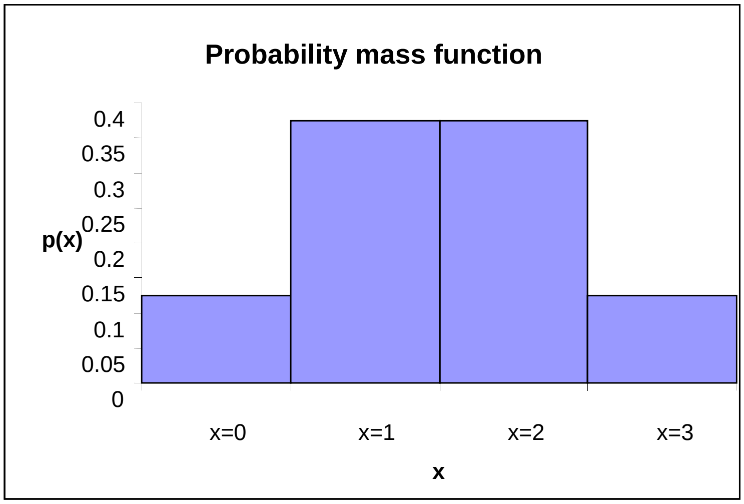


Figure 22

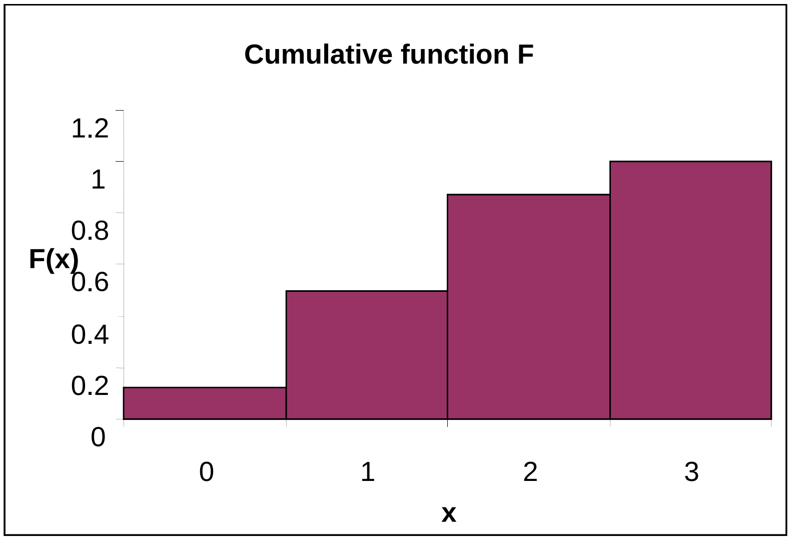


Figure 23

5. Find the mean and variance of the distribution of *X*.

$$\overline{X} = \sum x_i\, p(x_i) = 0 \times 0.125 + 1 \times 0.375 + 2 \times 0.375 + 3 \times 0.125 = 1.5$$

$$\overline{X^2} = \sum x_i^2\, p(x_i) = 0.375 + 4 \times 0.375 + 9 \times 0.125 = 3$$

$$\text{var}(X) = \overline{X^2} - (\overline{X})^2 = 3 - (1.5)^2 = 0.75$$

From this example, it can be seen that this discrete random variable *X* has a symmetrical probability distribution. The skewness is zero.

The most commonly used discrete random variables are the **binomial** and **Poisson** distributions.

Binomial random variable distribution

In this distribution, the basic experiment has only two outcomes, success or failure. Each success has a probability p and so each failure has probability $1 - p$. The binomial random variable counts the number of successes a in a fixed number N of independent runs of the experiment. This could refer to faulty items in a production process or even the number of heads in a sequence of coin tosses.

Here's an example:

In the coin tossing example, heads could be seen as a success and tails as failure with $p = ½ = 1 - p$. Random variable X is thus a binomial random variable with $N = 3$, $\overline{X} = 1.5 = 3 \times ½ = Np$; and $\mathrm{var}(X) = 0.75 = 3 \times ½ \times ½ = Np(1 - p)$.

In general: For sequences of N runs, there are 2^N outcomes in Ω. The probability of a successes is $p^a(1 - p)^{N-a}$. The reason is that if there are a successes each with probability p, then there must be $(N - a)$ failures each with probability $(1 - p)$.

How many ways are there of getting a successes from N runs? From number theory we learn that the number is denoted by symbol $\binom{N}{a}$ which is read as "N combination a" and means:

$$\frac{N(N-1)(N-2)\ldots\times 2\times 1}{[a(a-1)(a-2)\ldots\times 2\times 1].[(N-a)(N-(a-1))\ldots\times 2\times 1]}.$$

(For example, the number of ways of getting 2 heads from runs of 3 tosses is:

$\binom{3}{2}$. The value of this is $\dfrac{3(2)(1)}{[2(1)]\times[1]} = 3$.)

We can now construct the general mass function p for the binomial variable X:

$p(a) = P(X = a) = \binom{N}{a} p^{a} (1 - p)^{N-a}$

There are also formulas for the mean, variance and standard deviation of binomial random variable X:

Mean (expected value) = $E[X] = \overline{X} = Np$.

Variance of $X = \mathrm{var}(X) = \sigma^2(X) = Np(1 - p)$ and

Standard deviation of $X = \mathrm{std.dev}(X) = \sigma(X) = \sqrt{Np(1-p)}$.

Poisson distribution

The Poisson distribution also describes the random number of successes in a large number of trials. The difference is that the situation is one where the number of trials is quite large and the number of successes (each with probability p) is small. Note that "success" can mean faults, accidents, mistakes, and so on.

Here are some examples:

- The number of faults per kilometre section in an oil pipeline of 1,000 km.
- The number of accidents in a factory each month over 5 years.

Properties

For Poisson distributed random variable *X*:

Mean of *X* (expected value) = $\mu = Np$ and variance of $X = \sigma^2(X) = Np$.

Standard deviation of $X = \sigma(X) = \sqrt{Np}$

The probability mass function is:

$$p(a) = P(X = a) = \frac{e^{-\mu}\mu^{a}}{a(a-1)(a-2)\ldots 1}$$

Remember: *e* refers to the special exponential function. (See Module 1, Unit 3).

Activity 2.12

Activity

Calculate distributions, probabilities, means and standard deviations

What will you do?

1. Calculate $\begin{pmatrix}5\\3\end{pmatrix}, \begin{pmatrix}8\\1\end{pmatrix}, \begin{pmatrix}10\\0\end{pmatrix}$

2. Laptop components are packed into boxes at a factory. Each box contains 25 items. It is said that 15 per cent of all items produced are defective.

   a) What distribution would you use to describe the number of defective items per box?

   b) Give the probability mass function.

   c) Calculate the mean number of defective items and the standard deviation of the number of defective items.

   d) What is the probability of there being no defective items per box?

3. A major oil pipeline in Nigeria has random faults along its 2,400 km length. There are, on average, four faults per 10 km of pipeline. Let *X* denote the number of faults per 10 km.

   a) What distribution would you use to describe the number of faults per 10 km? Justify your answer.

   b) Give the probability mass function of *X*.

   c) Calculate the standard deviation of the number of faults per 10 km.

   d) What is the probability of there being one fault per 10 km stretch?

Here's our feedback

1. $\begin{pmatrix}5\\3\end{pmatrix} = \dfrac{5\times4\times3\times2\times1}{(3\times2\times1)(2\times1)} = \dfrac{120}{12} = 10.$

   $\begin{pmatrix}8\\1\end{pmatrix} = \dfrac{8\times7\times6\times5\times4\times3\times2\times1}{1\times(7\times6\times5\times4\times3\times2\times1)} = \dfrac{8}{1} = 8.$

   $\begin{pmatrix}10\\0\end{pmatrix} = 1$ (by definition).

2. a) The binomial distribution. Variable X is the number of defective items (successes) per box.

   b) $p(a) = P(X = a) = \begin{pmatrix}25\\a\end{pmatrix} 0.15^{a}(0.85)^{25-a}$

   c) $\overline{X} = 25(0.15) = 3.75, \quad \sigma_X = \sqrt{25(0.15)(0.85)} = 0.6375$

   d) $p(0) = P(X = 0) = \begin{pmatrix}25\\0\end{pmatrix} 0.15^{0}(0.85)^{25} = 1\times1\times(0.85)^{25} = 0.0172$

3. a) The Poisson distribution with mean $\mu = 4/10 = 0.4$. Justification: N is large ($N = 240$).

   b) $p(a) = P(X = a) = \dfrac{e^{-\mu}\mu^{a}}{a(a-1)(a-2)...1} = \dfrac{e^{-0.4}(0.4)^{a}}{a(a-1)(a-2)...1}$

   c) Standard deviation of $X = \sigma_X = \sqrt{Np} = \sqrt{\mu} = 0.632$

   d) $p(1) = P(X = 1) = e^{-0.4}(0.4)^{1} = 0.268$

Continuous random variables

When the values that a random variable X can take on form a continuum on the real number line, continuous probability distributions are used. In this case, we cannot calculate $p(a) = P(X = a)$ for a specific number a. Instead, we calculate the probability that X takes on values in an interval of real numbers – for example, $P(a \le X \le b)$ or $P(X \le c)$. The situation is similar to that of grouped continuous data in Unit 6.

Remember that the real number system or real line can be expressed by $(-\infty, \infty)$. So $P(X \le c)$ means $P(-\infty < X \le c)$, or the probability that X takes on values less than or equal to number c. The expression $P(a \le X \le b)$ means the probability that random variable X takes on values between (and including) a and b.

The distribution is described by a probability density function $f(x)$ and cumulative distribution function $F(x)$, with x along the x-axis.

Properties

- $f(x) = \dfrac{dF}{dx}(x)$ and $F(x) = \int f(x)dx$
- $P(X \le c) = F(c) = \int_{-\infty}^{c} f(x)dx$
- $P(a \le X \le b) = F(b) - F(a) = \int_{a}^{b} f(x)dx$

Note: The distribution of *X* is usually displayed as a graph of the density function *f*(*x*).

It is a good idea to do some revision of integration (covered in Module 1, Unit 4).

The equation $P(a \le X \le b) = F(b) - F(a) = \int_a^b f(x)dx$ means: the area under the graph of *f*(*x*) between values *a* and *b* gives the probability that *X* takes on values between *a* and *b*.

Similarly the equation $P(X \le b) = F(b) = \int_{-\infty}^{b} f(x)dx$ means: the area under the graph of *f*(*x*) from −∞ to *b* gives the probability that *X* takes on values less than *b*.

The total area under the graph is $\int_{-\infty}^{\infty} f(x)dx\ (= F(\infty))$ and gives the probability that *X* takes on any real number values. This is of course 1.

The mean and variance are calculated as:

Expected value (mean) of $X = \mathrm{E}[X] = \overline{X} = \int x.f(x)dx$

Variance of $X = \sigma_X^2 = \overline{X^2} - (\overline{X})^2$ where $\overline{X^2} = \int x^2 f(x)dx$

The uniform distribution

If random variable *X* is uniformly distributed on interval (0, 1), then the probability density function *f*(*x*) is: $\begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{for all other values of } x \end{cases}$

The graph looks like this:

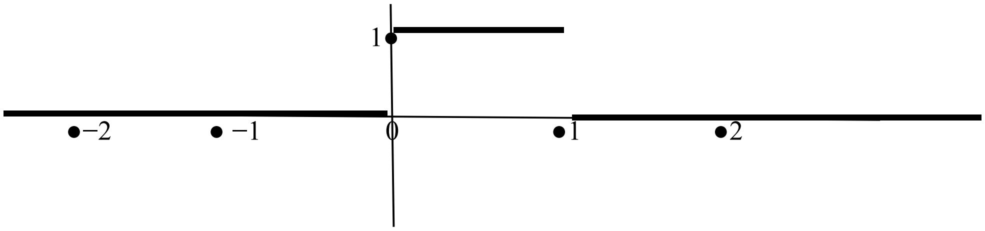


Figure 24

For numbers a and b with $0 \le a < b \le 1$, the probability $P(a \le X \le b)$ is calculated as:

$$P(a \le X \le b) = \int_a^b f(x)dx = \int_a^b 1dx = x \text{ (at } x = b) - x \text{ (at } x = a) = b - a$$

For numbers a and b with $a \le 0$ and $b \ge 1$ the probability $P(a \le X \le b)$ is calculated as: $\int_0^1 1dx = x(\text{at } x = 1) - x(\text{at } x = 0) = 1 - 0 = 1$

If $b > a > 1$, then $P(a \le X \le b) = \int_0^1 0dx = 0$

This can be generalised to a random variable X uniformly distributed on interval (α, β), with probability density function

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} \text{ if } \alpha < x < \beta \\ 0 \text{ for all other values of } x \end{cases}$$

The total area under the graph of the density function f is:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

$$\mathrm{E}[X] = \int_\alpha^\beta xf(x)dx = \frac{1}{\beta - \alpha}\int_\alpha^\beta xdx = \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} = \frac{\beta + \alpha}{2}$$

$$\text{Variance of } X = \sigma_X^2 = \int_\alpha^\beta x^2 f(x)dx - (\frac{\beta + \alpha}{2})^2 = \frac{(\beta - \alpha)^2}{12}$$

The uniform random distribution can be used to generate random numbers for simulations.

The normal distribution

One of the most well-known distributions is the **normal** or **Gaussian** distribution. The parameters of importance are the mean value μ and the standard deviation σ. The normal distribution is usually displayed as a continuous bell-shaped curve given by the density function $f(x)$ where:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

The graph of the density function f of the normal distribution is symmetrical about the mean and extends from $-\infty$ to ∞. The tails on either side become thinner as they tend towards the horizontal axis.

The total area under the graph equals 1: $\int_{-\infty}^{+\infty} f(x)dx = 1$.

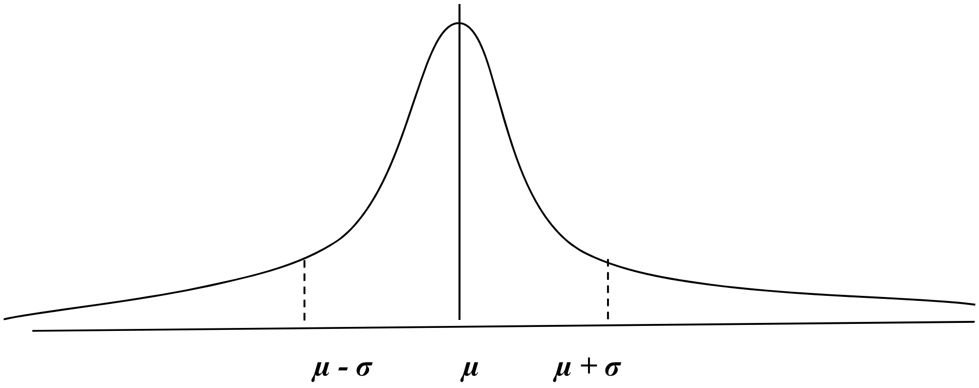


Figure 25

Important facts about the normal distribution:

- Mean = median = mode = μ
- $P(X \le \mu) = \int_{-\infty}^{\mu} f(x)dx = P(X \ge \mu) = 0.5$ (half the total area)
- $P(X \le -a) = 1 - P(X \ge a)$ for any number a
- $P(\mu-\sigma \le X \le \mu+\sigma) = \int_{\mu-\sigma}^{\mu+\sigma} f(x)dx = 0.68$
- $P(\mu-2\sigma \le X \le \mu+2\sigma) = 0.95$ (approximately)
- $P(\mu-3\sigma \le X \le \mu+3\sigma) = 0.99$ (approximately)
- The standard normal distribution is given by $Z = \dfrac{X-\mu}{\sigma}$

  Variable Z then has mean 0 and standard deviation 1. We write this as $Z \sim \text{n}(0, 1)$. There are tables for calculating probabilities such as $P(0 < Z < s)$ for any positive number s. (See the Appendix.)
- $P(-1.96 \le Z \le 1.96) = 0.95$
- $P(-2.33 \le Z \le 2.33) = 0.99$
- The binomial distribution can be approximated by the normal distribution.

Kurtosis

Kurtosis characterises the relative peakedness or flatness of a unimodal distribution in comparison with the normal distribution. As you know, a unimodal distribution is a probability distribution which has a single mode. If a distribution has a broader peak than the normal distribution, the tails are thinner. If a distribution has a thinner peak than the normal distribution, the tails are fatter. The reason for this is that the total area under the density function must always be equal to one.

Let X be a random variable with values x_i, (i = 1, ...,N); mean value $\overline{X}$ and variance σ^2. The measure for kurtosis is: $\sum \frac{(x_i - \overline{X})^4}{N\sigma^4} - 3$

The normal distribution has zero kurtosis.

Negative kurtosis means thinner or shorter tails and lower probability of extreme values.

Positive kurtosis means fatter or longer tails and higher probability of extreme values.

Fat-tailed distributions have become quite important and are discussed in Module 5.

The normal distribution has important applications in finance and economics. We often assume that the returns on share prices are normally distributed.

In operations management, it is often assumed that process variations are normally distributed, as is employee performance in human resource management.

Always remember that using certain probability distributions to describe and analyse data is an approximation for a possibly complex situation. Using the normal distribution to describe returns on assets may lead to predictions and decisions that are not always realistic. If the tails of the distribution of random variable X are fatter that predicted by the normal distribution, the probabilities that X takes on small values will be increased. (You will look at this again in Module 5, Unit 20.)

Activity 2.13

Activity

Practice probability

What will you do?

1. A continuous random variable X has density function f defined by:

   $f(x) = \frac{1}{2}x$, for $0 \leq x \leq 2$ and $f(x) = 0$ for all other x

   a) Determine the probability that X takes on a value between 0 and 1.

   b) What is the probability that X is smaller than 0.5?

   c) Calculate the mean value and standard deviation of X.

2. If X has normal distribution with $\mu = 10$ and $\sigma = 5$, define the associated standard normal variable Z. Now determine $P(X > 18)$.

Here's our feedback

1. a) $P(0 < X < 1) = \frac{1}{2}\int_{0}^{1} xdx$

$$= \frac{1}{2}[\frac{1}{2}x^2 \,(\text{where } x = 1) - \frac{1}{2}x^2 \,(\text{where } x = 0)]$$

$$= \frac{1}{4}$$

b) $P(X < 0.5) = \frac{1}{2}\int_{-\infty}^{0.5} xdx = \frac{1}{2}\int_{0}^{0.5} xdx = \frac{1}{2} . \frac{1}{2}x^2 \,(\text{where } x = 0.5)$

$$= \frac{1}{16}$$

c) $\overline{X} = \frac{1}{2}\int_{0}^{2} (x \times x)dx = \frac{1}{6}x^3 \,(\text{where } x = 2) - \frac{1}{6}x^3 \,(\text{where } x = 0)$

$$= \frac{8}{6}$$

$$\overline{X^2} = \frac{1}{2}\int_{0}^{2} (x^2 \times x)dx = \frac{1}{8}x^4 \,(\text{where } x = 2) - \frac{1}{8}x^4 \,(\text{where } x = 0)$$

$$= 2$$

Variance of $X = \sigma_X^2 = \overline{X^2} - (\overline{X})^2 = 2 - (\frac{8}{6})^2 = 0.22$ (rounded off)

Standard deviation of $X = \sqrt{0.22} = 0.469$

2. $Z = \frac{X - 10}{5}$

Then $P(X > 18) = P(\frac{X - 10}{5} > \frac{18 - 10}{5}) = P(Z > 1.6)$

First, note that $P(Z > 1.6) = 0.5 - P(0 < Z < 1.6)$. Now use the table in the Appendix to see that $P(0 < Z < 1.6) = 0.4452$.

Therefore $P(Z > 1.6) = 0.0548$ and $P(X > 18) = 0.0548$. There is a 5.48 per cent probability that X will be greater than 18.

Activity 2.14

Activity

Revisit probability

What will you do?

1. A machine produces items with a 10 per cent probability of being defective. What is the probability that in a sample of three items, at most one will be defective?
2. The number of accidents occurring on a road each week has a Poisson distribution with mean 6. Determine the probability that there will be 10 accidents the following week.
3. A continuous random variable X has probability density function $f(x) = 0.5$ for $0 < x < 2$, and $f(x) = 0$ for all other x.

   Calculate $\overline{X}$ and $\sigma^2(X)$.
4. Returns on assets are assumed to be normally distributed, with mean value 20 per cent and standard deviation $\sigma = 40$ per cent.

   What is the probability that the return will be:

   a) greater than 20 per cent?

   b) between −20 per cent and 60 per cent?

Application of the normal distribution to market research

In this section, you will apply your knowledge of statistical tools and probability theory to an important business application, market research.

Organisations often undertake market research to find out what people think of their product or service. The information that is gathered is then analysed and used by the manager to make strategic decisions.

The important thing to remember is that market research reflects data from a sample and not the entire population. As manager, you have to assess the reliability of the information. This can be done by **statistical inference** (inference means deduction or conclusion).

Sample and population

Some aspects of working with a sample versus a population have already been discussed. To summarise: population, or rather, statistical population means the entire collection (set) of items or people that are relevant to the research, and a sample is a part (subset) of the population. You use a sample to gather information because it is usually too expensive or even impossible to use the whole population.

Here are some examples:

- A sample of 100 customers from a population of 1,000 customers visiting a shop every month, to test their satisfaction with a product.
- A sample of 50 items from a population (batch) of 500 items manufactured every day, to test for defects.

- A sample of 200 patients waiting at a clinic from a population of 3,500 people using the clinic, to find out about service at the clinic.

The process of applying the findings from the sample to the statistical population as a whole (drawing conclusions) is called statistical inference. In this process, we must remember these points:

1. The sample must be representative of the population. A cross-section of items or people should be included. A sample of items to be tested should be drawn a number of times during the production phase, not just at the end of the process. A sample of shoppers should include people from different income groups, different ages, and males and females.
2. The data collected must be reliable or you must be aware that the results may not be accurate. People don't always answer questions truthfully.

Sampling distributions

The following example illustrates the concept of sampling distribution.

A government wants to put various programmes in place to reduce poverty. It needs data to determine the average (mean) annual household income for the entire population. This unknown quantity is denoted by μ. It also needs the standard deviation σ of incomes to understand the spread of incomes below and above the mean.

First, it surveys 1,000 households in the population (Survey 1) to find out the average annual income per household in the sample. Suppose the survey shows an average income of USD 700. The data can be displayed as: $\overline{x}_1 = 700$

If the survey is repeated for another, different, sample of 1,000 people (Survey 2), the results will, of course, be different and there will be a new average income $\overline{x}_2$. Say $\overline{x}_2 = 836$.

Theoretically, the survey can be repeated with new samples and a large number n of sample means collected: $\overline{x}_1$, $\overline{x}_2$, $\overline{x}_3$... $\overline{x}_n$. This gives a distribution of sample means, called the sampling distribution of the mean. It has its own mean value $\overline{X} = \frac{1}{n}\sum_{i=1}^{n}\overline{x_i}$ and standard deviation S.

How do these values relate to the unknown population mean income μ and population standard deviation σ that you are trying to find?

The central limit theorem gives the answer.

It says that the sample means are normally distributed for n large ($n \to \infty$) and that, in this case, the mean of the distribution of means approaches the population mean μ. Taking a large number n of samples is really equivalent to taking one very large sample. So if your sample is large, its mean $\overline{x}$ is close to the population mean μ. In this case, the sampling distribution has a small standard deviation S and the relation between S and σ is: $S\sqrt{n} = \sigma$.

Let’s summarise:

Suppose a population has mean value μ and standard deviation σ. If the population is normally distributed, then the sampling distribution of the mean is also normally distributed. More importantly, if the sample size n is large (more than 30 at least), then the sampling distribution of the mean is normally distributed, regardless of the population distribution. Best of all, the mean of the sampling distribution equals the population mean μ and the standard deviation S of the distribution of the mean is $S = (\sigma/\sqrt{n})$. So the bigger the sample size, the more closely its mean will approximate the population mean.

In this example, if you took many more samples n or one very large sample and find the average $\bar{x}$ = \$745, then you can approximate the population average as 745 USD. If your sample has standard deviation s = 84, then you approximate population standard deviation σ by s and standard deviation S of the sampling distribution by $s/\sqrt{n}$.

(Mathematically, the central limit theorem says:
If you have any sequence $X_1, X_2, X_3\ldots$ of independent, identically distributed random variables, each with mean μ and variance σ^2, then $\dfrac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$ tends to the Standard Normal distribution as n becomes very large.)

Confidence intervals

How good an approximation of the population mean is *our particular sample* mean? Remember, our sample is only one observation of the sampling distribution. In practice, we cannot really take $n \to \infty$.

We use probability theory to calculate how far apart our sample mean and the population mean could lie in the sampling distribution. Consider a 95 per cent confidence interval around μ. This means considering the interval around μ so that the probability that sample mean $\bar{x}$ lies in that interval is 95 per cent. That is, we are 95 per cent sure that the sample mean lies in this interval around μ. This is the interval $(\mu - 1.96\, s/\sqrt{n}, \mu + 1.96\, s/\sqrt{n})$. This also means that:

- the maximum distance between $\bar{x}$ and μ is $1.96\, s/\sqrt{n}$
- we are 95% sure that unknown population mean μ lies in interval $(\bar{x} - 1.96\, s/\sqrt{n},\ \bar{x} + 1.96\, s/\sqrt{n})$, which is what we really are interested in.

Here's an example:

In the population income study, we have
$\bar{x}$ = 745, s = 84 and sample size n = 1,000.

Calculate $1.96\, s/\sqrt{n} = 1.96\,(84/31.623) = 5.206$. Therefore we are 95% sure that the sample mean and population mean are not further apart than about USD 5.21.

The 95% confidence interval is: $(\bar{x} - 1.96\, s/\sqrt{n},\ \bar{x} + 1.96\, s/\sqrt{n})$
= (739.79, 750.21)

This is a very valuable conclusion: The population mean will lie in this interval around $\bar{x}$ with 95 per cent certainty.

Figure of sampling distribution and 95 per cent confidence interval

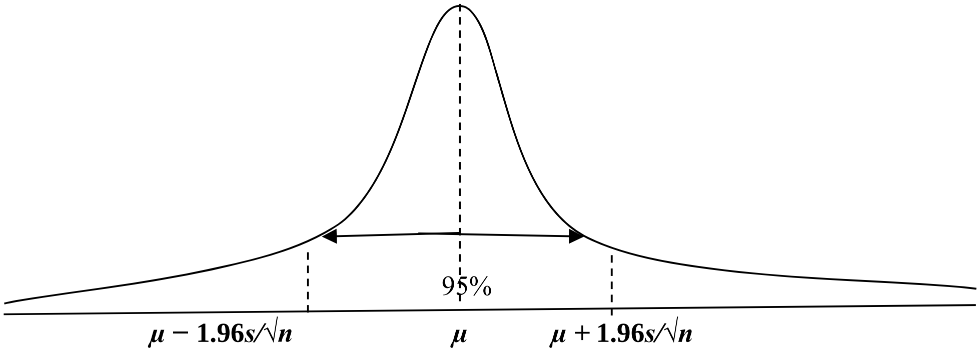


Figure 26

The government can now design its poverty alleviation programme with confidence in the data.

Note it!

Formula for limits of confidence interval

1. Consider a sample with mean sample standard deviation s and sample size n.

   A 95% confidence interval around sample mean $\bar{x}$ is:
   $(\bar{x} - 1.96\ s/\sqrt{n},\ \ \bar{x} + 1.96\ s/\sqrt{n})$

   This means we are 95% sure that the sample mean and population mean are not further apart than about 1.96 $s/\sqrt{n}$.

2. When the number of interest is a percentage (or proportion) л, the standard deviation of the sampling distribution is $\sqrt{\dfrac{\pi(1-\pi)}{n}}$ instead of the value $s/\sqrt{n}$

Activity 2.15

Activity

Calculate confidence interval and average

What will you do?

Samples are taken annually to calculate the average number of hours worked per month by agricultural labourers, and their average monthly wages. Standard deviations (std dev) are also calculated.

The data for years 2005–2008 are shown in this table:

	2005	2006	2007	2008
Average monthly hours worked (and std dev)	178 (10)	162 (14)	148 (7.2)	140 (12.1)
Average monthly wages in pounds (and std dev)	370.00 (10.56)	450.20 (5.53)	500.00 (12.54)	480.60 (14.32)
Sample size *n* (number of workers)	200	200	138	112

1. Calculate the 95 per cent confidence interval for each year. Discuss the meaning of the values.
2. Determine the average wages per hour for each year.

Hypothesis testing

A **hypothesis** is a statement or a belief about a population that still has to be tested against data obtained from a sample. Let’s explore a brief discussion of hypothesis testing.

Here's an example:

The government now wants to establish whether it is likely that private business will support the poverty alleviation programme. It hopes more than 50 per cent of all businesses will support the programme, otherwise the project will fail.

A sample of $n = 400$ businesses has shown that 55 per cent of sampled businesses are willing to take part in the programme, but the **actual** population percentage may, of course, be lower than 55 per cent or even lower than 50 per cent.

The question now is whether it is likely that less than 50 per cent of the total population of businesses will support the programme. The government would like to know that it can reject this possibility, because in this case the programme will not be sustainable. This problem can be investigated with hypothesis testing.

The basic statement to be tested is called the **null hypothesis**.

Note it!

Steps in hypothesis testing

1. Formulate a **null hypothesis H_0 and alternative hypothesis H_1**.
2. Set a significance level α. This is the probability of rejecting the null hypothesis if it is, in fact, true. If H_0 is true, we may, based on the sample data, reject H_0. In this case, we will have made an error (Type I). If H_0 is false, we may, based on the sample data, not reject it. In this case, we will have made a Type II error. We would like to minimise the probability of making errors. However, the two probabilities of errors are linked: the lower the probability of Type I errors, the higher the probability of Type II errors. Significance levels are either $\alpha = 5$ per cent or $\alpha = 1$ per cent.
3. Calculate the critical and test statistical Z values, Z_α and Z_{calc}. **(**Z refers to the standard normal distribution discussed in Unit 7.)

   Z_{calc} is the standard normal distribution value calculated for the sample This means $Z_{calc} = \dfrac{\overline{x} - \mu}{s / \sqrt{n}}$ remembering that $\overline{x}$ is the sample mean, μ is the population mean and the standard deviation of the sampling distribution is $s/\sqrt{n}$. (Review Unit 8.1.)

   $Z_\alpha = 2.33$ for $\alpha = 0.01$ and $Z_\alpha = 1.96$ for $\alpha = 0.05$ (from the table for the standard normal distribution in the Appendix).
4. Decide which hypothesis is accepted.

   The rule is that if $|Z_{calc}| > |Z_\alpha|$ then reject H_0.
   If $|Z_{calc}| < |Z_\alpha|$ then we cannot reject H_0. Remember that $|Z_\alpha|$ stands for absolute value and is always positive.

Activity 2.16

Activity

Carry out the hypothesis test

What will you do?

Carry out the hypothesis test for the example of government, private enterprise and the poverty alleviation programme.

Step 1: The null hypothesis H_0 is that less than 50 per cent of all businesses will support the programme. Denote the population percentage by л. H_0: л < 50%

The alternative hypothesis H_1 is that 50 per cent or more of businesses will support the programme: H_1: л ≥ 50%

You must choose between these two hypotheses. You either reject H_0 and accept H_1 or you do not reject H_0. It is important to note that not rejecting H_0 does not mean you accept H_0. There is merely not enough evidence to reject it.

Step 2: The actual, but unknown, situation is that H_0 is either true or false. If H_0 is true, you may, based on the sample data, reject H_0. In this case, you will have made an error (Type I). In your example, you want the probability of a Type I error to be as small as possible and choose $\alpha = 0.01$.

Step 3: Determine the critical Z values. $Z_\alpha = 2.33$ for $\alpha = 0.01$, and Z_{calc} = [*x* − л]/std dev is calculated from the sampling distribution. The population mean is л = 50 per cent. The observed value was 56 per cent. The standard deviation $s/\sqrt{n}$ must be adapted for percentages. It becomes: $\sqrt{\frac{\pi(100-\pi)}{n}} = \sqrt{\frac{50(100-50)}{400}}$ =2.5 and thus Z_{calc} = $\frac{56-50}{2.5}$ =2.4

The meaning of this value is that the sample result of 56 per cent is 2.4 standard deviations away from the hypothesised population mean of 50 per cent. This is outside the region determined by 2.33 standard deviations.

Step 4: The decision can be made. In this case, you can reject hypothesis H_0 because 2.4 > 2.33, and you accept the alternative hypothesis H_1.

Based on this test, the government can accept that 50 per cent or more of businesses will support the poverty project.

Note: Remember, that calculated values come from data collected from a *sample*. If this sample is not reliable or representative of the population, you may draw the wrong conclusions.

As a manager, you should not base important decisions on the results of a simple hypothesis test. You should analyse the situation carefully, for example by considering *p* values or looking at *t* tests or $\chi 2$ tests. (These are not discussed here; it is up to you to investigate when needed.)

Activity 2.17

Activity

Understand the terminology

What will you do?

Use this terminology table to record any terms or words you are uncertain about.

This activity is an opportunity to consolidate your understanding of new terminology and concepts you encountered in Unit 8. Fill in the terms you have learned and then write your own descriptions of them.

Terminology

Term **Description**

:

:

:

:

:

:

:

:

:

:

:

Case study: Market research

Attempt this case study in market research and hypothesis testing before doing the assessment for this module.

Companies are awarded a special rating for their social commitment, which gives them certain tax benefits. The rating system is based on assessments by the companies themselves as to how much they spend on social projects, improving the living conditions of their workers, combating pollution, and so on. They send an annual report to the Internal Revenue Service (IRS).

The IRS awards the ratings, but it cannot check each company's self-assessment every year. It was initially thought that at least 90 per cent of the self-assessments would be correct and fair. However, the possibility exists of mistakes or outright fraud.

After two years, the IRS has thoroughly checked a sample of the reports of 2,100 companies. It found that 12.01 per cent of reports were not quite correct.

Discuss and test the initial view (hypothesis) of the IRS that at least 90 per cent of reports would be correct.

Remember these key points

- Random or stochastic variables take on values in a random, unpredictable way. The values can be discrete real numbers or form a continuous section of the real line.
- The probability distribution of a random variable X gives information on the mean and standard deviation of values. It takes the form of a probability mass function p for discrete variables and a probability density function f for continuous variables.
- For discrete variables:

  Mean = $\overline{X} = \sum x_i\, p(x_i)$ summed over all possible values x_i

  Variance = var(X) = $\sigma^2(X) = \overline{X^2} - (\overline{X})^2$ where $\overline{X^2} = \sum x_i^2\, p(x_i)$
- The main discrete distributions are the binomial distribution and Poisson distribution.
- For continuous variables:

  Mean value of $X = \mu = \overline{X} = \int x.f(x)dx$

  Variance of $X = \sigma_X^2 = \overline{X^2} - (\overline{X})^2$ where $\overline{X^2} = \int x^2 f(x)dx$
- The main distributions are the normal (Gaussian) distribution and the uniform distribution. The probability density for a normally distributed random variable X is $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$
- If X is normal, then the standard normal variable is $Z = \dfrac{X-\mu}{\sigma}$
- A normal random variable has zero kurtosis. Fat-tailed distributions have positive kurtosis and a higher probability of taking on extreme values.
- The normal distribution has an important application in market research via the central limit theorem. Confidence intervals and critical Z-values help managers determine the reliability of results based on samples from populations.

Unit summary

Summary

You have successfully completed this unit if you can:

- **construct** probability mass functions for discrete random variables;
- **apply** binomial distribution and Poisson distribution;
- **construct** probability density functions for continuous random variables;
- **apply** the normal distribution in market research;
- **interpret** results correctly to inform decisions; and
- **handle** probability distributions and understand their role in modelling.

Assessment

Assessment — Module 2

1. Find the mean, median and mode of this set of data. Represent the data in a scatter graph:
   15, 6, 14, 7, 12, 11, 7, 10, 8, 13
2. Study the frequency table for the consumption of diesel by a group of trucks. (For example, there are six trucks using between 25 and 30 litres of diesel per 100 km.)

Class: Litres per 100 km	Frequency: Number of trucks
20–25	2
25–30	6
30–35	12
35–40	8
40–45	5

   a) Expand the table to include cumulative and relative frequencies.

   b) Find the mean and median for the distribution.

   c) Draw a histogram and ogive.

   d) Use the ogive to confirm the value of the median.

   e) Is the distribution positively or negatively skewed?

   f) Calculate the standard deviation for the data.

3. Sales figures S_1 and S_2 (income in a local currency) for a week for two branches of a coffee shop are:

Day	Sales for shop 1	Sales for shop 2
1	1,200	1,300
2	2,300	2,450
3	2,400	3,500
4	1,800	2,000
5	1,100	1,500
6	2,000	2,200
7	2,200	1,800

   a) Determine the mean and standard deviation of sales for each shop.

   b) Calculate the co-variance and correlation co-efficient for sales figures S_1 and S_2.

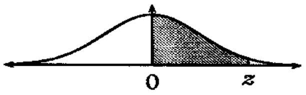
   c) Present the data in column or scatter graphs.

   d) Interpret and discuss your results for 3a) – 3c).

4. What is the probability of getting a sum greater than 10 when throwing two dice?
5. Let E and F be mutually exclusive and independent events. What can you say about the probabilities $P(E \cup F)$ and $P(E \cap F)$?
6. A fair coin is tossed three times. What is the probability of throwing two tails and one head? Use a probability tree for your calculations.
7. The conditional probability $P(E|F)$ is 0.4 and the probability $P(E \cap F)$ is 0.2. What is the value of $P(F)$?
8. Random variable X is binomially distributed. There are 20 trials and the probability of success is 0.15. What is the probability of 10 successes?
9. Random variable Y has the Poisson distribution with mean value 4.2.
   a) What is the variance of the distribution?
   b) What is the probability of 14 successes?
10. Random variable X is normally distributed, with mean value 30 and a variance of 14.
    a) What is the probability that X will take on values greater than 44? (Hint: No tables are necessary.)
    b) What is the probability that X will take on values less than 20? (Hint: Transform X to a standard normal variable Z and use tables.)
11. A company wants to find out how long it takes a worker to assemble a certain component of a machine. A sample of 40 workers shows an average time of 76.4 seconds, and a standard deviation of 17.2 seconds. Find the 95 per cent confidence interval for the population mean.
12. A sample of 500 patients is given a new treatment for AIDS. Testing the patients after one year shows that 67 per cent of these patients have responded very well, with increased CD4- or T-cell counts. Determine a 95 per cent confidence interval for the percentage of patients in the population who will have increased CD4 counts after the treatment.

Appendix

The table entries give the probabilities $P(0 \leq Z \leq z)$, where $Z \sim n(0; 1)$, denoted by the shaded area in the following figure.



Second decimal place of z

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,00000	0,00399	0,00798	0,01197	0,01595	0,01994	0,02392	0,02790	0,03188	0,03586
0,1	0,03983	0,04380	0,04776	0,05172	0,05567	0,05962	0,06356	0,06749	0,07142	0,07535
0,2	0,07926	0,08317	0,08706	0,09095	0,09483	0,09871	0,10257	0,10642	0,11026	0,11409
0,3	0,11791	0,12172	0,12552	0,12930	0,13307	0,13683	0,14058	0,14431	0,14803	0,15173
0,4	0,15542	0,15910	0,16276	0,16640	0,17003	0,17364	0,17724	0,18082	0,18439	0,18793
0,5	0,19146	0,19497	0,19847	0,20194	0,20540	0,20884	0,21226	0,21566	0,21904	0,22240
0,6	0,22575	0,22907	0,23237	0,23565	0,23891	0,24215	0,24537	0,24857	0,25175	0,25490
0,7	0,25804	0,26115	0,26424	0,26730	0,27035	0,27337	0,27637	0,27935	0,28230	0,28524
0,8	0,28814	0,29103	0,29389	0,29673	0,29955	0,30234	0,30511	0,30785	0,31057	0,31327
0,9	0,31594	0,31859	0,32121	0,32381	0,32639	0,32894	0,33147	0,33398	0,33646	0,33891
1,0	0,34134	0,34375	0,34614	0,34849	0,35083	0,35314	0,35543	0,35769	0,35993	0,36214
1,1	0,36433	0,36650	0,36864	0,37076	0,37286	0,37493	0,37698	0,37900	0,38100	0,38298
1,2	0,38493	0,38686	0,38877	0,39065	0,39251	0,39435	0,39617	0,39796	0,39973	0,40147
1,3	0,40320	0,40490	0,40658	0,40824	0,40988	0,41149	0,41308	0,41466	0,41621	0,41774
1,4	0,41924	0,42073	0,42220	0,42364	0,42507	0,42647	0,42785	0,42922	0,43056	0,43189
1,5	0,43319	0,43448	0,43574	0,43699	0,43822	0,43943	0,44062	0,44179	0,44295	0,44408
1,6	0,44520	0,44630	0,44738	0,44845	0,44950	0,45053	0,45154	0,45254	0,45352	0,45449
1,7	0,45543	0,45637	0,45728	0,45818	0,45907	0,45994	0,46080	0,46164	0,46246	0,46327
1,8	0,46407	0,46485	0,46562	0,46638	0,46712	0,46784	0,46856	0,46926	0,46995	0,47062
1,9	0,47128	0,47193	0,47257	0,47320	0,47381	0,47441	0,47500	0,47558	0,47615	0,47670
2,0	0,47725	0,47778	0,47831	0,47882	0,47932	0,47982	0,48030	0,48077	0,48124	0,48169
2,1	0,48214	0,48257	0,48300	0,48341	0,48382	0,48422	0,48461	0,48500	0,48537	0,48574
2,2	0,48610	0,48645	0,48679	0,48713	0,48745	0,48778	0,48809	0,48840	0,48870	0,48899
2,3	0,48928	0,48956	0,48983	0,49010	0,49036	0,49061	0,49086	0,49111	0,49134	0,49158
2,4	0,49180	0,49202	0,49224	0,49245	0,49266	0,49286	0,49305	0,49324	0,49343	0,49361
2,5	0,49379	0,49396	0,49413	0,49430	0,49446	0,49461	0,49477	0,49492	0,49506	0,49520
2,6	0,49534	0,49547	0,49560	0,49573	0,49585	0,49598	0,49609	0,49621	0,49632	0,49643
2,7	0,49653	0,49664	0,49674	0,49683	0,49693	0,49702	0,49711	0,49720	0,49728	0,49736
2,8	0,49744	0,49752	0,49760	0,49767	0,49774	0,49781	0,49788	0,49795	0,49801	0,49807
2,9	0,49813	0,49819	0,49825	0,49831	0,49836	0,49841	0,49846	0,49851	0,49856	0,49861
3,0	0,49865	0,49869	0,49874	0,49878	0,49882	0,49886	0,49889	0,49893	0,49896	0,49900
3,1	0,49903	0,49906	0,49910	0,49913	0,49916	0,49918	0,49921	0,49924	0,49926	0,49929
3,2	0,49931	0,49934	0,49936	0,49938	0,49940	0,49942	0,49944	0,49946	0,49948	0,49950
3,3	0,49952	0,49953	0,49955	0,49957	0,49958	0,49960	0,49961	0,49962	0,49964	0,49965
3,4	0,49966	0,49968	0,49969	0,49970	0,49971	0,49972	0,49973	0,49974	0,49975	0,49976
3,5	0,49977	0,49978	0,49978	0,49979	0,49980	0,49981	0,49981	0,49982	0,49983	0,49983
3,6	0,49984	0,49985	0,49985	0,49986	0,49986	0,49987	0,49987	0,49988	0,49988	0,49989
3,7	0,49989	0,49990	0,49990	0,49990	0,49991	0,49991	0,49992	0,49992	0,49992	0,49992
3,8	0,49993	0,49993	0,49993	0,49994	0,49994	0,49994	0,49994	0,49995	0,49995	0,49995
3,9	0,49995	0,49995	0,49996	0,49996	0,49996	0,49996	0,49996	0,49996	0,49997	0,49997
4,0	0,49997	0,49997	0,49997	0,49997	0,49997	0,49997	0,49998	0,49998	0,49998	0,49998

References

References

Anderson, D. R., Sweeney, D.J. & Williams, T.A. (2007). *Statistics for business and economics.* Dundee, Scotland: Thomson Publishers.

Bradley, T. & Patton, P. (2002). *Essential mathematics for economics and business.* New York: John Wiley.

Portal: Statistics. (n.d.). In *Wikipedia*. Retrieved May 2, 2011, from http://en.wikipedia.org/wiki/Portal:Statistics

Ross, S. M. (2006). *Introduction to probability models.* San Diego, CA: Harcourt/Academic Press.

Swift, L. (1997). *Mathematics and statistics for business management and finance.* London: Macmillan.

University of California, Berkeley. (n.d.). Stat labs: Mathematical statistics through applications [website]. Retrieved from http://www.stat.berkeley.edu/users/statlabs/

Venture Data. (2009). ResearchInfo [website]. Retrieved May 2, 2011, from http://www.researchinfo.com/docs/library/

Waters, D. (1997). *Quantitative methods for business.* Boston, MA: Addison-Wesley.

Wisniewski, M. (2006). *Quantitative methods for decision makers.* Upper Saddle River, NJ: Prentice Hall.

Wisniewski, M. (2010). *Quantitative methods for decision makers* [webpage]. Retrieved from http://www.pearsoned.co.uk/wisniewski/

Further reading

Reading

Where else can I look?

These texts and websites may be consulted for additional information.

Anderson, D. R., Sweeney, D.J. & Williams, T.A. (2007). *Statistics for business and economics.* Dundee, Scotland: Thomson Publishers.

Portal: Statistics. (n.d.). In *Wikipedia*. Retrieved May 2, 2011, from http://en.wikipedia.org/wiki/Portal:Statistics

Ross, S. M. (2006). *Introduction to probability models.* San Diego, CA: Harcourt/Academic Press.

Swift, L. (1997). *Mathematics and statistics for business management and finance.* London: Macmillan.

University of California, Berkeley. (n.d.). Stat labs: Mathematical statistics through applications [website]. Retrieved from http://www.stat.berkeley.edu/users/statlabs/

Venture Data. (2009). ResearchInfo [website]. Retrieved May 2, 2011, from http://www.researchinfo.com/docs/library/

Waters, D. (1997). *Quantitative methods for business.* Boston, MA: Addison-Wesley.

Wisniewski, M. (2006). *Quantitative methods for decision makers.* Upper Saddle River, NJ: Prentice Hall.

Wisniewski, M. (2010). *Quantitative methods for decision makers* [webpage]. Retrieved from http://www.pearsoned.co.uk/wisniewski/