

**M140\_1**

**Prices, location and spread**

**About this free course**

This free course is an adapted extract from the Open University course M140 Introducing statistics [www.open.ac.uk/courses/modules/m140](http://www.open.ac.uk/courses/qualifications/details/m140).

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

You can experience this free course as it was originally designed on OpenLearn, the home of free learning from The Open University – [www.open.edu/openlearn/science-maths-technology/prices-location-and-spread/content-section-0](http://www.open.edu/openlearn/science-maths-technology/prices-location-and-spread/content-section-0?utm_source=openlearn&utm_campaign=ol&utm_medium=ebook).

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

Unless otherwise stated, copyright © 2016 The Open University, all rights reserved.

**Intellectual property**

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

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

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

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

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

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

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

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

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

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

Head of Intellectual Property, The Open University

978-1-4730-2079-5.kdl 978-1-4730-2078-8.epub

1.1

# Contents

* [Introduction](#Session1)
* [Learning outcomes](#Session2)
* [1 Measuring location](#Session3)
  + [1.1 Data on prices](#Session3_Section1)
  + [1.2 The median](#Session3_Section2)
  + [1.3 The arithmetic mean](#Session3_Section3)
  + [1.4 The mean and median compared](#Session3_Section4)
  + [Exercises on Section 1](#Session3_Section5)
* [2 Weighted means](#Session4)
  + [2.1 The mean of a combined batch](#Session4_Section1)
  + [2.2 Further uses of weighted means](#Session4_Section2)
  + [2.3 More than two numbers](#Session4_Section3)
  + [Exercises on Section 2](#Session4_Section4)
* [3 Measuring spread](#Session5)
  + [3.1 The range](#Session5_Section1)
  + [3.2 Quartiles and the interquartile range](#Session5_Section2)
  + [3.3 The five-figure summary and boxplots](#Session5_Section3)
  + [Exercises on Section 3](#Session5_Section4)
* [4 A simple chained price index](#Session6)
  + [4.1 A two-commodity price index](#Session6_Section1)
  + [Exercise on Section 4](#Session6_Section2)
* [5 The UK government price indices](#Session7)
  + [5.1 What are the CPI and RPI?](#Session7_Section1)
  + [5.2 Calculating the price indices](#Session7_Section2)
  + [5.3 Using the price indices](#Session7_Section3)
  + [Exercises on Section 5](#Session7_Section4)
* [Conclusion](#Session8)
* [Text](#Session9)
* [Solutions](#Solutions1)
* [Descriptions](#Descriptions1)

## Introduction

This free course, Prices, location and spread, examines aspects of the question:

Start of Quote

Are people getting better or worse off?

End of Quote

The course concentrates on the statistical aspects of the question, focusing on statistics about prices. However, it is not the case that statistics can provide all the answers – or even the best answer – to the question of whether people are getting better or worse off. There are many non-statistical issues which are relevant and it is important to put the statistical approach in its correct perspective.

In the question, people does not refer specifically to you, the readers of the course, but to the whole of society in the UK. That is quite a big batch (more than 62 million in 2010, according to an estimate from the UK’s Office for National Statistics), consisting of men, women and children, living alone, in large or small households, or in institutions; some of them working, others unemployed, some retired and others still at school.

It is not possible, using statistical techniques, to provide a complete answer to this one question covering such a big theme, particularly an answer which is valid for all these people and their varied economic and social circumstances; data and techniques both have to be used with common sense. Instead, the aim of this text is more modest: to explore small batches of data relevant to the question (and relating to some individuals and groups in society), using basic analytical and graphical techniques. The sections in this course cover the following:

* Section 1 looks at ways to measure the overall ‘location’ (the typical value) of a batch of data. Two very important measures looked at are the ‘median’ and the ‘arithmetic mean’. The section also looks at patterns in data using diagrammatic methods.
* Section 2 shows how to calculate the ‘weighted mean’, which is a quantity related to the arithmetic mean. You will learn about some circumstances where it makes sense to calculate a weighted mean.
* Section 3 shows how to calculate one particular measure of spread for a batch: the ‘interquartile range’. It also shows some diagrammatic methods for representing the spread and shape of the distribution of values in a batch.
* Section 4 introduces the notion of a ‘price index’ for indicating changes in the price of a single item and for two or more different items.
* Section 5 looks at the UK’s Retail Prices Index (RPI) and Consumer Prices Index (CPI), which measure changes in prices over time. (This section is longer than all the other sections, so you should plan your study time accordingly.)

This OpenLearn course is an adapted extract from the Open University course [M140 Introducing statistics](http://www3.open.ac.uk/study/undergraduate/course/m140.htm).

## Learning outcomes

After studying this course, you should be able to:

* find the mean and median of a batch of data
* find the weighted mean of two or more numbers
* find the lower and upper quartiles, interquartile range and five-figure summary of a batch of data
* calculate a simple chained price index
* use the Retail Prices Index and the Consumer Prices Index to measure price changes.

## 1 Measuring location

Measuring location has two components:

* gathering data about the quantity of interest
* determining a value to represent the location of the data.

The task of gathering appropriate data is somewhat problem-specific – general strategies are available, but exact details usually need to be decided for each problem. To determine the price of an electric kettle, for example, we would have to decide the size and type of kettle we’re interested in, where and when its purchased, and so forth. In contrast, choosing a value to summarise the location of a set of data is more straightforward. In this section, we will focus on the two most common measures of location: the median and the mean. The data gathered about the quantity of interest does not affect the way we calculate these location measures.

## 1.1 Data on prices

In order to measure how prices change, we need data on prices and some way of measuring their overall location. Price data take many forms.

In examining the overall location, prices of all goods are relevant, but some are more important than others. Ballpoint pens are relatively unimportant in most people’s shopping baskets, coffee prices are unimportant for tea drinkers, and chicken prices are of little concern to vegetarians. The first batch of price data we will look at is coffee prices.

Start of Box

**Example 1 Price data for jars of coffee**

Table 1 shows prices of a 100 g jar of a well-known brand of instant coffee obtained in 15 different shops in Milton Keynes on the same day in February 2012.

Start of Table

Table 1 Coffee prices (in pence)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 299 | 315 | 268 | 269 | 295 |
| 295 | 369 | 275 | 268 | 295 |
| 279 | 268 | 268 | 295 | 305 |

End of Table

There are several points to note concerning these prices.

* They relate to a particular brand of coffee. You might expect the price to vary between brands.
* They relate to a standard 100 g jar. You might expect the price per gram of this brand of coffee to vary depending upon the size of the jar – larger jars are often cheaper (per gram).
* They relate to a particular locality. You might expect the price to vary depending upon where you buy the coffee (e.g. central London, a suburb, a provincial town, a country village or a Hebridean island).
* They relate to a particular day. You might expect the price to vary from time to time depending upon changes in the cost of raw coffee beans, costs of production and distribution, and the availability of special offers.

Nevertheless, although we have data for a fixed brand of coffee, size of jar, locality and date of purchase, this batch of prices still varies from the lower extreme of 268p to the upper extreme of 369p. (In symbols: uppercase E subscript uppercase L end = 268 and uppercase E subscript uppercase U end = 369.) One of the most likely reasons for this is that the prices were collected from different kinds of shops (e.g. supermarket, petrol station, ethnic grocery and corner shop).

For all these reasons, it is impossible to state exactly what the price of this brand of instant coffee is. Yet its price is, in its own small way, relevant to the question: Are people getting better or worse off? That is, if you drink this particular coffee, then changes in its price in your locality will affect your cost of living. Similarly, your costs and economic well-being will also be affected by what happens to the prices of all the other things you need or like to consume.

On the other hand, someone who never buys instant coffee will be unaffected by any change in its price; they will be much more interested in what happens to the prices of alternative products such as ground coffee, tea, milk or fruit juice. The problem of measuring the effect of price changes on individuals with different consumption patterns will be considered in Section 5.

End of Box

## 1.2 The median

Despite the variability in the data, Table 1 does provide some idea of the price you would expect to pay for a 100 g jar of that particular instant coffee in the Milton Keynes area on that particular day. The information provided by the batch can be seen more clearly when drawn as a stemplot, shown in Figure 1 of Example 2.

Start of Box

**Example 2 Picturing the coffee data**

Start of Figure

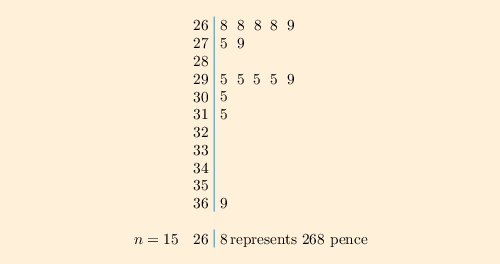


Figure 1 Stemplot of coffee prices from Table 1

[View description - Figure 1 Stemplot of coffee prices from Table 1](" \l "Session3_Description1)

End of Figure

This stemplot shows at a glance that if you shop around, you might well find this brand of coffee on sale at less than 270p. (Indeed some stores seem to have been ‘price matching’ at the lowest price of 268p.) On the other hand, if you are not too careful about making price comparisons then you might pay considerably more than 300p (£3). However, you are most likely to find a shop with the coffee priced between about 270p and 300p. Although there is no one price for this coffee, it seems reasonable to say that the overall location of the price is a bit less than 300p.

The **median** of the batch is a useful measure of the overall location of the values in a batch. It is defined as the middle value of a batch of figures when the values are placed in order. Let us examine in more detail what that means.

The stemplot in Figure 1 shows the prices arranged in order of size. We can label each of these 15 prices with a symbol indicating where it comes in the ordered batch. A convenient way of showing this is to write each value as the symbol x plus a subscript number in brackets, where the subscript number shows the position of that value within the ordered batch. Figure 2 shows the 15 prices written out in ascending order using this subscript notation.

Start of Figure

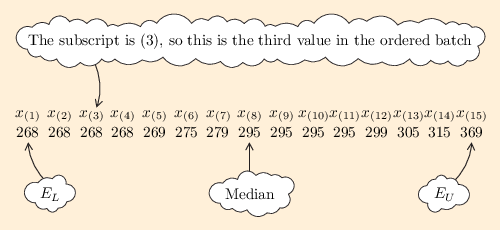


Figure 2 Subscript notation for ordered data

[View description - Figure 2 Subscript notation for ordered data](" \l "Session3_Description2)

End of Figure

The lower extreme, uppercase E subscript uppercase L end, is labelled x subscript open bracket 1 close bracket end and the upper extreme, uppercase E subscript uppercase U end, is labelled x subscript open bracket 15 close bracket end. The middle value is the value labelled x subscript open bracket 8 close bracket end since there are as many values, namely 7, above the value of x subscript open bracket 8 close bracket end as there are below it. (This is not strictly true here, since the values of x subscript open bracket 9 close bracket end, x subscript open bracket 10 close bracket end and x subscript open bracket 11 close bracket end happen also to be actually equal to the median.)

This is illustrated in Figure 3 by a V-shaped formation. The median is the middle value, so it lies at the bottom of the V. (This way of picturing a batch will be developed further in Subsection 3.2.)

Start of Figure

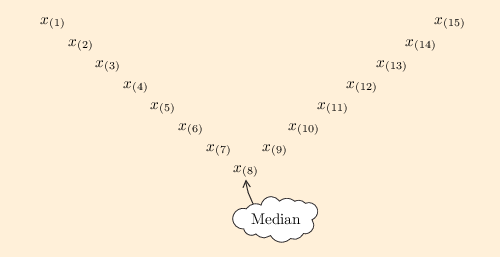


Figure 3 Median of 15 values

[View description - Figure 3 Median of 15 values](" \l "Session3_Description3)

End of Figure

If you wanted to make a more explicit statement, then you could write: The median price of this batch of 15 prices is 295p.

End of Box

If we picture any batch of data as a V-shape like Figure 3, the median of the batch will always lie at the bottom of the V. In the ordered batch, it is more places away from the extremes than any other value.

In general, the median is the value of the middle item when all the items of the batch are arranged in order. For a batch size n, the position of the middle value is fraction 1 over 2 end open bracket n+1 close bracket. For example, when n = 15, this gives a position of fraction 1 over 2 end open bracket 15+1 close bracket =8, indicating that x subscript open bracket 8 close bracket end is the median value. When n is an even number, the middle position is not a whole number and the median is the average of the two numbers either side of it. For example, when n = 12, the median position is 6 fraction 1 over 2 end, indicating that the median value is taken as halfway between x subscript open bracket 6 close bracket end and x subscript open bracket 7 close bracket end.

Example 3 uses prices of a digital camera to illustrate how the median is found for an even number of values.

Start of Box

**Example 3 Digital cameras**

Table 2 shows prices for a particular model of digital camera as given on a price comparison website in March 2012.

Start of Table

Table 2 Prices for a digital camera (to the nearest £)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 60 | 70 | 53 | 81 | 74 |
| 85 | 90 | 79 | 65 | 70 |

End of Table

If we put these prices in order and arrange them in a V-shape, they look like Figure 4.

Start of Figure

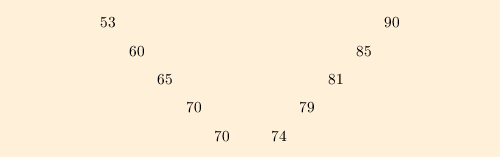


Figure 4 Prices of 10 digital cameras

[View description - Figure 4 Prices of 10 digital cameras](" \l "Session3_Description4)

End of Figure

Because 10 is an even number, there is no single middle value in this batch: the position of the middle item is fraction 1 over 2 end open bracket 10+1 close bracket = 5 fraction 1 over 2 end. The two values closest to the middle are those shown at the bottom of the V: x subscript open bracket 5 close bracket end = 70 and x subscript open bracket 6 close bracket end =74. Their average is 72, so we say that the median price of this batch of camera prices is £72.

End of Box

The following activity asks you to find the median for an even number of values, using a stemplot of prices for small flat-screen televisions.

Start of Activity

**Activity 1 Small flat-screen televisions**

Start of Question

Figure 5 is a stemplot of data on the prices of small flat-screen televisions. (The prices have been rounded to the nearest £10. Originally all but one ended in 9.99, so in this case it makes reasonable sense to ignore the rounding and treat the data as if the prices were exact multiples of £10.) Find the median of these data.

Start of Figure

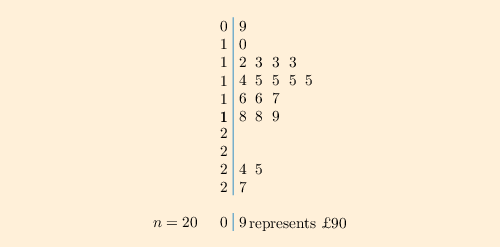


Figure 5 Prices of all flat-screen televisions with a screen size of 24 inches or less on a major UK retailer’s website on a day in February 2012

[View description - Figure 5 Prices of all flat-screen televisions with a screen size of 24 inches or ...](" \l "Session3_Description5)

End of Figure

End of Question

[View discussion - Activity 1 Small flat-screen televisions](" \l "Session3_Discussion1)

End of Activity

This subsection can now be finished by using some of the methods we have met to examine a batch of data consisting of two parts, or sub-batches.

Start of Activity

**Activity 2 The price of gas in UK cities**

Start of Question

Table 3 presents the average price of gas, in pence per kilowatt hour (kWh), in 2010, for typical consumers on credit tariffs in 14 cities in the UK. These cities have been divided into two sub-batches: as seven northern cities and seven southern cities. (Legally, at the time of writing, Ipswich is a town, not a city, but we shall ignore that distinction here.)

Start of Table

Table 3 Average gas prices in 14 cities

|  |  |  |  |
| --- | --- | --- | --- |
| **Northern cities** | **Average gas price (pence per kWh)** | **Southern cities** | **Average gas price (pence per kWh)** |
| Aberdeen | 3.740 | Birmingham | 3.805 |
| Edinburgh | 3.740 | Canterbury | 3.796 |
| Leeds | 3.776 | Cardiff | 3.743 |
| Liverpool | 3.801 | Ipswich | 3.760 |
| Manchester | 3.801 | London | 3.818 |
| Newcastle-upon-Tyne | 3.804 | Plymouth | 3.784 |
| Nottingham | 3.767 | Southampton | 3.795 |

End of Table

End of Question

Start of Question

(a)   Draw a stemplot of all 14 prices shown in the table.

End of Question

[View discussion - Part](" \l "Session3_Discussion2)

Start of Question

(b)   Draw separate stemplots for the seven prices for northern cities and the seven prices for southern cities.

End of Question

[View discussion - Part](" \l "Session3_Discussion3)

Start of Question

(c)   For each of these three batches (northern cities, southern cities and all cities) find the median and the range. Then use these figures to find the general level and the range of gas prices for typical consumers in the country as a whole, and to compare the north and south of the country.

End of Question

[View discussion - Part](" \l "Session3_Discussion4)

End of Activity

Activity 2 illustrates two general properties of sub-batches:

* The range of the complete batch is greater than or equal to the ranges of all the sub-batches.
* The median of the complete batch is greater than or equal to the smallest median of a sub-batch and less than or equal to the largest median of a sub-batch.

## 1.3 The arithmetic mean

Another important measure of location is the arithmetic mean. (Pronounced arithmetic.)

Start of Box

**Arithmetic mean**

The arithmetic mean is the sum of all the values in the batch divided by the size of the batch. More briefly,

Start of $1

mean = fraction sum over size end .

[View alternative description - Uncaptioned Equation](" \l "Session3_Alternative41)

End of $1

End of Box

There are other kinds of mean, such as the geometric mean and the harmonic mean, but in this course we shall be using only the arithmetic mean; the word mean will therefore normally be used for arithmetic mean.

Start of Box

**Example 4 An arithmetic mean**

Suppose we have a batch consisting of five values: 4, 8, 4, 2, 9. In this simple example, the mean is

Start of $1

fraction sum over size end = fraction 4 + 8 + 4 + 2 + 9 over 5 end = fraction 27 over 5 end = 5.4.

[View alternative description - Uncaptioned Equation](" \l "Session3_Alternative42)

End of $1

End of Box

Note that in calculating the mean, the order in which the values are summed is irrelevant.

For a larger batch size, you may find it helpful to set out your calculations systematically in a table. However, in practice the raw data are usually fed directly into a computer or calculator. In general, it is a good idea to check your calculations by reworking them. If possible, use a different method in the reworking; for example, you could sum the numbers in the opposite order.

The formula ‘mean = sum / size’ can be expressed more concisely as follows. Referring to the values in the batch by x, the ‘sum’ can be written as sum x. Here sum is the Greek (capital) letter Sigma, the Greek version of S, and is used in statistics to denote ‘the sum of’. Also, the symbol overline x is often used to denote the mean – and as you have already seen in stemplots, n can be used to denote the batch size. (Some calculators use keys marked sum x and overline x to produce the sum and the mean of a batch directly.)

Using this notation,

Start of $1

mean = fraction sum over size end

[View alternative description - Uncaptioned Equation](" \l "Session3_Alternative51)

End of $1

can be written as

Start of $1

overline x = fraction sum x over n end .

[View alternative description - Uncaptioned Equation](" \l "Session3_Alternative52)

End of $1

In this course we shall normally round the mean to one more figure than the original data.

Start of Activity

**Activity 3 Small televisions: the mean**

Start of Question

The prices of 20 small televisions were given in [Activity 1](#open-u2act1-0) (Subsection 1.2). Find the mean of these prices. Round your answer appropriately (if necessary), given that the original data were rounded to the nearest £10.

End of Question

[View discussion - Activity 3 Small televisions: the mean](" \l "Session3_Discussion5)

End of Activity

## 1.4 The mean and median compared

Both the mean and median of a batch are useful indicators of the location of the values in the batch. They are, however, calculated in very different ways. To find the median you must first order the batch of data, and if you are not using a computer, you will often do the sorting by means of a stemplot. On the other hand, the major step in finding the mean consists of summing the values in the batch, and for this they do not need to be ordered.

For large batches, at least when you are not using a computer, it is often much quicker to sum the values in the batch than it is to order them. However, for small batches, like some of those you will be analysing in this course without a computer, it can be just as fast to calculate the median as it is to calculate the mean. Moreover, placing the batch values in order is not done solely to help calculate the median – there are many other uses. Drawing a stemplot to order the values also enables us to examine the general shape of the batch. In Section 3 you will read about some other uses of the stemplot.

Comparisons based on the method of calculation can be of great practical interest, but the rest of this subsection will consider more fundamental differences between the mean and the median – differences which should influence you when you are deciding which measure to use in summarising the general location of the values in a batch.

Many of the problems with the mean, as well as some advantages, lie in the fact that the precise value of every item in the batch enters into its calculation. In calculating the median, most of the data values come into the calculation only in terms of whether they are in the 50% above the median value or the 50% below it. If one of them changes slightly, but without moving into the other half of the batch, the median will not change. In particular, if the extreme values in the batch are made smaller or larger, this will have no effect on the value of the median – the median is resistant to outliers. In contrast, changes to the extremes could have an appreciable effect on the value of the mean, as the following examples show.

Start of Box

**Example 5 Changing the extreme coffee prices**

For the batch of coffee prices in [Figure 1](#open-u2fig1-1) (Subsection 1.2), the sum of the values is 4363p, so the mean is

Start of $1

fraction 4363 p over 15 end simeq 290.9 p .

[View alternative description - Uncaptioned Equation](" \l "Session3_Alternative58)

End of $1

Suppose the highest and lowest coffee prices are reduced so that

Start of $1

x subscript open bracket 1 close bracket end = 240 and x subscript open bracket 15 close bracket end = 340.

[View alternative description - Uncaptioned Equation](" \l "Session3_Alternative59)

End of $1

The median of this altered batch is the same as before, 295p. However, the sum of the values is now 4306p and so the mean is

Start of $1

fraction 4306 p over 15 end simeq 287.1 p .

[View alternative description - Uncaptioned Equation](" \l "Session3_Alternative60)

End of $1

End of Box

Start of Box

**Example 6 Changing the small television prices**

Suppose the highest two television prices in [Activity 1](#open-u2act1-0) (Subsection 1.2) are altered to £350 and £400. The median, at £150, remains the same as that of the original batch, whereas the new mean is

Start of $1

fraction pounds 3470 over 20 end = pounds 173.5 simeq pounds 174

[View alternative description - Uncaptioned Equation](" \l "Session3_Alternative61)

End of $1

compared with the original mean of £162.

Now, even with the very high prices of £350 and £400 for two televisions, the overall location of the main body of the data is still much the same as for the original batch of data. For the original batch the mean, £162, was a reasonably good measure of this. However, for the new batch the mean, £174, is much too high to be a representative measure since, as we can see from the stemplot in Activity 1, most of the values are below £174.

End of Box

Example 6 is the subject of the following screencast. [Note that the reference to ‘Unit 2’ should be ‘this course’. Unit 2 is a reference to the Open University course from which this material is adapted.]

Start of Media Content

Video content is not available in this format.

Screencast 1 Effects on the median and mean when data points change

[View transcript - Screencast 1 Effects on the median and mean when data points change](" \l "Session3_Transcript1)

Start of Figure



End of Figure

End of Media Content

Start of Box

A measure which is insensitive to changes in the values near the extremes is called a **resistant measure**.

The median is a **resistant** measure whereas the mean is **sensitive**.

End of Box

In the following activities, you can investigate some other ways in which the median is more resistant than the mean.

Start of Activity

**Activity 4 Changing the gas prices**

Start of Question

In [Activity 2](#open-u2act1-1) (Subsection 1.2) you may have noticed that Cardiff and Ipswich had rather low gas prices compared to the other southern cities. Here you are going to examine the effect of deleting them from the batch of southern cities. Complete the following table and comment on your results.

Start of Table

|  |  |  |
| --- | --- | --- |
| **Batch** | **Mean** | **Median** |
| Seven southern cities |  |  |
| Five southern cities (excluding Cardiff and Ipswich) |  |  |

End of Table

End of Question

[View discussion - Activity 4 Changing the gas prices](" \l "Session3_Discussion6)

End of Activity

Start of Activity

**Activity 5 A misprint in the gas prices**

Start of Question

Suppose the value for London had been misprinted as 8.318 instead of 3.818 (quite an easy mistake to make!). How would this affect your results for the batch of five southern cities (again omitting Cardiff and Ipswich)?

Start of Table

|  |  |  |
| --- | --- | --- |
| **Batch** | **Mean** | **Median** |
| Five cities (correct data) |  |  |
| Five cities (with misprint) |  |  |

End of Table

End of Question

[View discussion - Activity 5 A misprint in the gas prices](" \l "Session3_Discussion7)

End of Activity

Suppose you wanted to use these values – the correct ones, of course – to estimate the average price of gas over the whole country. The simple arithmetic mean of the 14 values given in [Table 3](#open-u2table1-3) (Subsection 1.2) would not allow for the fact that much more gas is consumed in London, at a relatively high price, than in other cities. To take account of this you would need to calculate what is known as a weighted arithmetic mean. Weighted means are the subject of Section 2.

## Exercises on Section 1

The following exercises provide extra practice on the topics covered in Section 1.

Start of Exercise

**Exercise 1 Finding medians**

Start of Question

For each of the following batches of data, find the median of the batch. (We shall also use these batches of data in some of the exercises in Section 3.)

End of Question

Start of Question

(a)   Percentage scores in arithmetic obtained by 33 school students.

Start of Figure

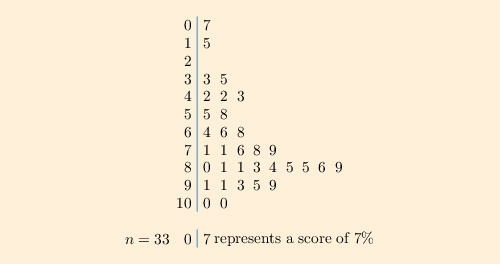


Figure 8

[View description - Figure 8](" \l "Session3_Description8)

End of Figure

End of Question

[View discussion - Part](" \l "Session3_Discussion8)

Start of Question

(b)   Prices of 26 digital televisions with 22- to 26-inch LED screens, quoted online by a large department store in February 2012. The prices have been rounded to the nearest pound (£).

Start of Table

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 170 | 180 | 190 | 200 | 220 | 229 | 230 | 230 | 230 |
| 230 | 250 | 269 | 269 | 270 | 279 | 299 | 300 | 300 |
| 315 | 320 | 349 | 350 | 400 | 429 | 649 | 699 |  |

End of Table

End of Question

[View discussion - Part](" \l "Session3_Discussion9)

End of Exercise

Start of Exercise

**Exercise 2 Finding means**

Start of Question

Calculate the mean for each of the batches in Exercise 1.

End of Question

[View discussion - Exercise 2 Finding means](" \l "Session3_Discussion10)

End of Exercise

Start of Exercise

**Exercise 3 The effect of removing values on the median and mean**

Start of Question

In the data on prices for small televisions in [Activity 1](#open-u2act1-0) (Subsection 1.2), the three highest-priced televisions were considerably more expensive than all the others (which all cost under £200). Suppose that in fact these prices had been for a different, larger type of television that should not have been in the batch. (In fact that is not the case – but this is only an exercise!) Leave these three prices out of the batch and calculate the median and the mean of the remaining prices.

How do these values compare with the original median (150) and mean (162)? What does this comparison demonstrate about how resistant the median and mean are?

End of Question

[View discussion - Exercise 3 The effect of removing values on the median and mean](" \l "Session3_Discussion11)

End of Exercise

## 2 Weighted means

For goods and services, price changes vary considerably from one to another. Central to the theme question of this course, Are people getting better or worse off?, there is a need to find a fair method of calculating the average price change over a wide range of goods and services. Clearly a 10% rise in the price of bread is of greater significance to most people than a similar rise in the price of clothes pegs, say. What we need to take account of, then, are the relative weightings attached to the various price changes under consideration.

## 2.1 The mean of a combined batch

This first subsection looks at how a mean can be calculated when two unequally weighted batches are combined.

Start of Box

**Example 7 Alan’s and Beena’s biscuits**

Suppose we are conducting a survey to investigate the general level of prices in some locality. Two colleagues, Alan and Beena, have each visited several shops and collected information on the price of a standard packet of a particular brand of biscuits. They report as follows (Figure 9).

* Alan visited five shops, and calculated that the mean price of the standard packet at these shops was 81.6p.
* Beena visited eight shops, and calculated that the mean price of the standard packet at these shops was 74.0p.

Start of Figure



Figure 9 Means of biscuit prices

[View description - Figure 9 Means of biscuit prices](" \l "Session4_Description1)

End of Figure

If we had all the individual prices, five from Alan and eight from Beena, then they could be amalgamated into a single batch of 13 prices, and from this combined batch we could calculate the mean price of the standard packet at all 13 shops. However, our two investigators have unfortunately not written down, nor can they fully remember, the prices from individual shops. Is there anything we can do to calculate the mean of the combined batch?

Fortunately there is, as long as we are interested in arithmetic means. (If they had recorded the medians instead, then there would have been very little we could do.)

The mean of the combined batch of all 13 prices will be calculated as

Start of $1

fraction sum open bracket of the combined batch prices close bracket over size open bracket of the combined batch close bracket end .

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative1)

End of $1

We already know that the size of the combined batch is the sum of the sizes of the two original batches; that is, 5+8=13. The problem here is how to find the sum of the combined batch of Alan’s and Beena’s prices. The solution is to rearrange the familiar formula

Start of $1

mean = fraction sum over size end

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative3)

End of $1

so that it reads

Start of $1

sum = mean times size .

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative4)

End of $1

This will allow us to find the sums of Alan’s five prices and Beena’s eight prices separately. Adding the results will produce the sum of the combined batch prices. Finally, dividing by 13 completes the calculation of finding the combined batch mean.

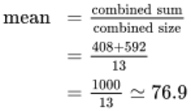
Let us call the sum of Alan’s prices ‘sum(A)’ and the sum of Beena’s prices ‘sum(B)’.

For Alan: mean = 81.6 and size = 5, so sum open bracket A close bracket = 81.6 times 5=408.

For Beena: mean =74.0 and size =8, so sum open bracket B close bracket =74.0 times 8 =592.

For the combined batch:

Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative11)

End of $1

Here, the result has been rounded to give the same number of digits as in the two original means.

End of Box

The process that we have used above is an important one. It will be used several times in the rest of this course. The box below summarises the method, using symbols.

Start of Box

**Mean of a combined batch**

The formula for the **mean** overline x subscript uppercase C end**of a combined batch** uppercase C is

Start of $1

overline x subscript uppercase C end = fraction overline x subscript uppercase A end n subscript uppercase A end + overline x subscript uppercase B end n subscript uppercase B end over n subscript uppercase A end + n subscript uppercase B end end comma

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative14)

End of $1

where batch uppercase C consists of batch uppercase A combined with batch uppercase B, and

Start of $1

overline x subscript uppercase A end = mean of batch uppercase A comma n subscript uppercase A end = size of batch uppercase A comma overline x subscript uppercase B end = mean of batch uppercase B comma n subscript uppercase B end = size of batch uppercase B.

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative18)

End of $1

End of Box

For our survey in Example 7,

Start of $1

overline x subscript uppercase A end = 81.6 comma n subscript uppercase A end = 5 comma overline x subscript uppercase B end = 74.0 comma n subscript uppercase B end = 8.

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative19)

End of $1

The formula summarises the calculations we did as

Start of $1

overline x subscript uppercase C end = fraction open bracket 81 .6 times 5 close bracket + open bracket 74 .0 times 8 close bracket over 5 +8 end .

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

End of $1

This expression is an example of a **weighted mean**. The numbers 5 and 8 are the **weights**. We call this expression the weighted mean of 81.6 and 74.0 with weights 5 and 8, respectively.

To see why the term weighted mean is used for such an expression, imagine that Figure 10 shows a horizontal bar with two weights, of sizes 5 and 8, hanging on it at the points 81.6 and 74.0, and that you need to find the point at which the bar will balance. This point is at the weighted mean: approximately 76.9.

Start of Figure

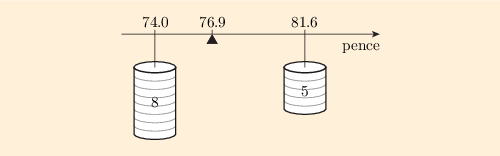


Figure 10 Point of balance at the weighted mean

[View description - Figure 10 Point of balance at the weighted mean](" \l "Session4_Description2)

End of Figure

This physical analogy illustrates several important facts about weighted means.

* It does not matter whether the weights are 5 kg and 8 kg or 5 tonnes and 8 tonnes; the point of balance will be in the same place. It will also remain in the same place if we use weights of 10 kg and 16 kg or 40 kg and 64 kg – it is only the relative sizes (i.e. the ratio) of the weights that matter.
* The point of balance must be between the points where we hang the weights, and it is nearer to the point with the larger weight.
* If the weights are equal, then the point of balance is halfway between the points.

This gives the following rules.

Start of Box

**Rules for weighted means**

**Rule 1**   The weighted mean depends on the relative sizes (i.e. the ratio) of the weights.

**Rule 2**   The weighted mean of two numbers always lies between the numbers and it is nearer the number that has the larger weight.

**Rule 3**   If the weights are equal, then the weighted mean of two numbers is the number halfway between them.

End of Box

Start of Box

**Example 8 Two batches of small televisions**

Suppose that we have two batches of prices (in pounds) for small televisions:

Start of $1

Batch uppercase A has mean 119 and size 7. Batch uppercase B has mean 185 and size 13 .

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative21)

End of $1

To find the mean of the combined batch we use the formula above, with

Start of $1

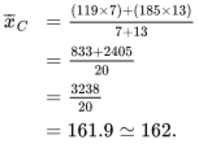
overline x subscript uppercase A end = 119 comma n subscript uppercase A end = 7 comma overline x subscript uppercase B end = 185 comma n subscript uppercase B end = 13.

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative22)

End of $1

This gives

Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative23)

End of $1

Note that this is the weighted mean of 119 and 185 with weights 7 and 13 respectively. It lies between 119 and 185 but it is nearer to 185 because this has the greater weight: 13 compared with 7.

End of Box

Example 8 is the subject of the following screencast. [Note that references to ‘the unit’ and ‘the units’ should be interpreted as ‘this course’. The original wording refers to the Open University course from which this material is adapted.]

Start of Media Content

Video content is not available in this format.

Screencast 2 Calculating a weighted mean

[View transcript - Screencast 2 Calculating a weighted mean](" \l "Session4_Transcript1)

Start of Figure



End of Figure

End of Media Content

## 2.2 Further uses of weighted means

We shall now look at another similar problem about mean prices – one which is perhaps closer to your everyday experience.

Start of Box

**Example 9 Buying petrol**

Suppose that, in a particular week in 2012, a motorist purchased petrol on two occasions. On the first she went to her usual, relatively low-priced filling station where the price of unleaded petrol was 136.9p per litre and she filled the tank; the quantity she purchased was 41.2 litres. The second occasion saw her obliged to purchase petrol at an expensive service station where the price of unleaded petrol was 148.0p per litre; she therefore purchased only 10 litres. What was the mean price, in pence per litre, of the petrol she purchased during that week?

To calculate this mean price we need to work out the total expenditure on petrol, in pence, and divide it by the total quantity of petrol purchased, in litres.

The total quantity purchased is straightforward as it is just the sum of the two quantities, so 41.2 + 10.

To find the expenditure on each occasion, we need to apply the formula:

Start of $1

cost = price times quantity .

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative25)

End of $1

This gives 136.9 times 41.2 and 148.0 times 10, respectively.

So the total expenditure, in pence, is open bracket 136.9 times 41.2 close bracket + open bracket 148.0 times 10 close bracket. The mean price, in pence per litre, for which we were asked, is this total expenditure divided by the total number of litres bought:

Start of $1

fraction open bracket 136 .9 times 41 .2 close bracket + open bracket 148 .0 times 10 close bracket over 41 .2 + 10 end .

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative29)

End of $1

We have left the answer in this form, rather than working out the individual products and sums as we went along, to show that it has the same form as the calculation of the combined batch mean. (The answer is 139.07p per litre, rounded from 139.067 97p per litre.)

End of Box

The phrase ‘goods and services’ is an awkward way of referring to the things that are relevant to the cost of living; that is, physical things you might buy, such as bread or gas, and services that you might pay someone else to do for you, such as window-cleaning. Economists sometimes use the word commodity to cover both goods and services that people pay for, and we shall use that word from time to time in this course. (Note that there are other, different, technical meanings of commodity that you might meet in different contexts.)

Start of Box

**The mean price of a quantity bought on two different occasions**

In general, if you purchase q sub 1 units of some commodity at p sub 1 pence per unit and q sub 2 units of the same commodity at p sub 2 pence per unit, then the mean price of this commodity, overline p pence per unit, can be calculated from the following formula:

Start of $1

overline p = fraction p sub 1 q sub 1 + p sub 2 q sub 2 over q sub 1 + q sub 2 end .

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative35)

End of $1

End of Box

Start of Box

**Example 10 Buying potatoes**

Suppose that, in one month, a family purchased potatoes on two occasions. On one occasion they bought 10 kg at 40p per kg, and on another they bought 6 kg at 45p per kg. We can use this formula to calculate the mean price (in pence per kg) that they paid for potatoes in that month. We have

Start of $1

. begin 2 by 2 array q sub 1 =10 next column quantity next row p sub 1 = 40 next column price end array } first occasion

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative36)

End of $1

and

Start of $1

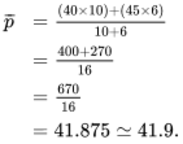
. begin 2 by 2 array q sub 2 =6 next column quantity next row p sub 2 =45 next column price end array } second occasion.

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative37)

End of $1

This gives

Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative38)

End of $1

So the mean price for that month is 41.9p per kg.

End of Box

The two formulas we have been using,

Start of $1

fraction overline x subscript uppercase A end n subscript uppercase A end + overline x subscript uppercase B end n subscript uppercase B end over n subscript uppercase A end + n subscript uppercase B end end and fraction p sub 1 q sub 1 + p sub 2 q sub 2 over q sub 1 + q sub 2 end comma

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative39)

End of $1

are basically the same; they are both examples of weighted means.

The first formula is the weighted mean of the numbers overline x subscript uppercase A end and overline x subscript uppercase B end, using the batch sizes, n subscript uppercase A end and n subscript uppercase B end, as weights.

The second formula is the weighted mean of the unit prices p sub 1 and p sub 2, using the quantities bought, q sub 1 and q sub 2, as weights.

The general form of a weighted mean of two numbers having associated weights is as follows.

Start of Box

**Weighted mean of two numbers**

The **weighted mean** of the two numbers x sub 1 and x sub 2 with corresponding weights w sub 1 and w sub 2 is

Start of $1

fraction x sub 1 w sub 1 + x sub 2 w sub 2 over w sub 1 + w sub 2 end .

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative52)

End of $1

End of Box

Weighted means have many uses, two of which you have already met. The type of weights depends on the particular use. In our uses, the weights were the following.

* The sizes of the batches, when we were calculating the combined batch mean from two batch means.
* The quantities bought, when we were calculating the mean price of a commodity bought on two separate occasions.

Another very important use is in the construction of an index, such as the Retail Prices Index; we shall therefore be making much use of weighted means in the final sections of this course.

In the next example, we do not have all the information required to calculate the mean, but we can still get a reasonable answer by using weights.

Start of Box

**Example 11 Weighted means of two gas prices**

Let us return to the gas prices in [Table 3](#open-u2table1-3) (Subsection 1.2). This has information about the price of gas for typical consumers in individual cities, but no national figure. Suppose that you want to combine these figures to get an average figure for the whole country; how could you do it? At the end of Section 1, it was suggested that weighted means could provide a solution. The complete answer to this question, using weighted means, is in Example 13 towards the end of this section. To introduce the method used there, let us now consider a similar, but simpler, question.

Here we use just two cities, London and Edinburgh, where the prices were 3.818p per kWh and 3.740p per kWh respectively. How can we combine these two values into one sensible average figure?

One possibility would be to take the simple mean of the two numbers. This gives

Start of $1

fraction 1 over 2 end open bracket 3.818 + 3.740 close bracket = 3.779.

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative53)

End of $1

However, this gives both cities equal weight. Because London is a lot larger than Edinburgh, we should expect the average to be nearer the London price than the Edinburgh price.

This suggests that we use a weighted mean of the form

Start of $1

fraction 3 .818 q sub 1 + 3 .740 q sub 2 over q sub 1 + q sub 2 end comma

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative54)

End of $1

where q sub 1 and q sub 2 are suitably chosen weights, with the weight q sub 1 of the London price larger than the weight q sub 2 of the Edinburgh price.

The best weights would be the total quantities of gas consumed in 2010 in each city. However, even if this information is not available to us, we can still find a reasonable average figure by using as weights a readily available measure of the sizes of the two cities: their populations.

The populations of the urban areas of these cities are approximately 8 300 000 and 400 000 respectively. So we could put q sub 1 = 8300000 and q sub 2 = 400 000.

However, we know that the weighted mean depends only on the ratio of the weights. Therefore, the weights q sub 1 = 83 and q sub 2 = 4 will give the same answer.

These weights give

Start of $1

fraction open bracket 3 .818 times 83 close bracket + open bracket 3 .740 times 4 close bracket over 83 + 4 end .

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative63)

End of $1

End of Box

Start of Activity

**Activity 6 Using the rules for weighted means**

Start of Question

Using the rules for weighted means, would you expect the weighted mean price to be nearer the London price or the Edinburgh price? To check, calculate the weighted mean price.

End of Question

[View discussion - Activity 6 Using the rules for weighted means](" \l "Session4_Discussion1)

End of Activity

Although we cannot think of the weighted mean price in Activity 6 as a calculation of the total cost divided by the total consumption, the answer is an estimate of the average price, in pence per kWh, for typical consumers in the two cities, and it is the best estimate we can calculate with the available information.

Sometimes the weights in a weighted mean do not have any significance in themselves: they are neither quantities, nor sizes, etc., but simply weights. This is illustrated in the following activity.

Start of Activity

**Activity 7 Weighted means of Open University marks**

Start of Question

Open University students become familiar with the combination of interactive computer-marked assignment (iCMA) and tutor-marked assignment (TMA) scores to provide an overall continuous assessment score (OCAS) for a course.

Suppose that a student obtains a score of 80 for their iCMAs and a score of 60 for their TMAs. Calculate what this student’s overall continuous assessment score will be if the weights for the two components are as follows.

End of Question

Start of Question

(a)   iCMA 50, TMA 50

End of Question

[View discussion - Part](" \l "Session4_Discussion2)

Start of Question

(b)   iCMA 40, TMA 60

End of Question

[View discussion - Part](" \l "Session4_Discussion3)

Start of Question

(c)   iCMA 65, TMA 55

End of Question

[View discussion - Part](" \l "Session4_Discussion4)

Start of Question

(d)   iCMA 25, TMA 75

End of Question

[View discussion - Part](" \l "Session4_Discussion5)

Start of Question

(e)   iCMA 30, TMA 90

End of Question

[View discussion - Part](" \l "Session4_Discussion6)

End of Activity

We have seen, in Activity 7 and in Example 11, that only the ratio of the weights affects the answer, not the individual weights. So weights are often chosen to add up to a convenient number like 100 or 1000. (This is Rule 1 for weighted means (see [Subsection 2.1](#open-u2ss2-1)).)

Activity 7 should also have reminded you of another important property of a weighted mean of two numbers: the weighted mean lies nearer to the number having the larger weight. (This is part of Rule 2 for weighted means.)

## 2.3 More than two numbers

The idea of a weighted mean can be extended to more than two numbers. To see how the calculation is done in general, remind yourself first how we calculated the weighted mean of two numbers x sub 1 and x sub 2 with corresponding weights w sub 1 and w sub 2.

1. Multiply each number by its weight to get the products x sub 1 w sub 1 and x sub 2 w sub 2.
2. Sum these products to get x sub 1 w sub 1 + x sub 2 w sub 2.
3. Sum the weights to get w sub 1 + w sub 2.
4. Divide the sum of the products by the sum of the weights.

This leads to the following formula.

Start of Box

**Weighted mean of two or more numbers**

The weighted mean of two or more numbers is

Start of $1

fraction sum of { number times weight } over sum of weights end = fraction sum of products over sum of weights end .

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative80)

End of $1

End of Box

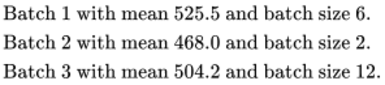
This is the formula which is used to find the weighted mean of any set of numbers, each with a corresponding weight.

Start of Box

**Example 12 A weighted mean of wine prices**

Suppose we have the following three batches of wine prices (in pence per bottle).

Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative81)

End of $1

We want to calculate the weighted mean of these three batch means using, as corresponding weights, the three batch sizes. Rather than applying the formula directly, the calculations can be set out in columns.

Start of Table

Table 4 Data on wine purchases

|  |  |  |  |
| --- | --- | --- | --- |
| **Batch** | **Number (batch mean)** | **Weight (batch size)** | **Number times weight ( = product)** |
| Batch 1 | 525.5 | 6 | 3 153.0 |
| Batch 2 | 468.0 | 2 | 936.0 |
| Batch 3 | 504.2 | 12 | 6 050.4 |
| **Sum** |  | **20** | **10 139.4** |

End of Table

The weighted mean is

Start of $1

fraction sum of products over sum of weights end = fraction 10139 .4 over 20 end = 506.97.

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative83)

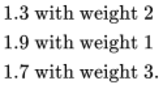
End of $1

We round this to the same accuracy as the original means, to get a weighted mean of 507.0. (Note that this lies between 468.0 and 525.5. This is a useful check, as a weighted mean always lies within the range of the original means.)

End of Box

The physical analogy in Example 12 can be extended to any set of numbers and weights. Suppose that you calculate the weighted mean for:

Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative84)

End of $1

This is given by

Start of $1

fraction open bracket 1 .3 times 2 close bracket + open bracket 1 .9 times 1 close bracket + open bracket 1 .7 times 3 close bracket over 2 + 1 + 3 end = fraction 2 .6 + 1 .9 + 5 .1 over 6 end = fraction 9 .6 over 6 end = 1.6.

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative85)

End of $1

This is pictured in Figure 11, with the point of balance for these three weights shown at 1.6.

Start of Figure



Figure 11 Point of balance for three means

[View description - Figure 11 Point of balance for three means](" \l "Session4_Description3)

End of Figure

You will meet many examples of weighted means of larger sets of numbers in Subsection 5.2, but we shall end this section with one more example.

Start of Box

**Example 13 Weighted means of many gas prices**

[Example 11](#open-u2exa2-4) showed the calculation of a weighted mean of gas prices using, for simplicity, just the two cities London and Edinburgh. We can extend Example 11 to calculate a weighted mean of all 14 gas prices from [Table 3](#open-u2table1-3), using as weights the populations of the 14 cities. The calculations are set out in Table 5.

Start of Table

Table 5 Product of gas price and weight by city

|  |  |  |  |
| --- | --- | --- | --- |
| **City** | **Price (p/kWh): x** | **Weight: w** | **Price times weight: x w** |
| Aberdeen | 3.740 | 19 | 71.060 |
| Edinburgh | 3.740 | 42 | 157.080 |
| Leeds | 3.776 | 150 | 566.400 |
| Liverpool | 3.801 | 82 | 311.682 |
| Manchester | 3.801 | 224 | 851.424 |
| Newcastle-upon-Tyne | 3.804 | 88 | 334.752 |
| Nottingham | 3.767 | 67 | 252.389 |
| Birmingham | 3.805 | 228 | 867.540 |
| Canterbury | 3.796 | 5 | 18.980 |
| Cardiff | 3.743 | 33 | 123.519 |
| Ipswich | 3.760 | 14 | 52.640 |
| London | 3.818 | 828 | 3161.304 |
| Plymouth | 3.784 | 24 | 90.816 |
| Southampton | 3.795 | 30 | 113.850 |
| **Sum** |  | **1834** | **6973.436** |

End of Table

The entries in the weight column, w, are the approximate populations, in 10 000s, of the urban areas that include each city (as measured in the 2001 Census). For each city, we multiply the price, x, by the weight, w, to get the entry in the last column, x w.

The weighted mean of the gas prices using these weights is then

Start of $1

fraction sum of products open bracket price times weight close bracket over sum of weights end

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative94)

End of $1

or, in symbols,

Start of $1

fraction sum x w over sum w end .

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative95)

End of $1

As sum x w = 6973.436 and sum w = 1834, the weighted mean is

Start of $1

fraction 6973 .436 over 1834 end =3.802310 simeq 3.802.

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative98)

End of $1

So the weighted mean of these gas prices, using approximate population figures as weights, is 3.802p per kWh.

Note that this weighted mean is larger than all but three of the gas prices for individual cities. That is because the cities with the two highest populations, London and Birmingham, also have the highest gas prices, and the weighted mean gas price is pulled towards these high prices.

End of Box

Although the details of the calculation above are written out in full in Table 5, in practice, using even a simple calculator, this is not necessary. It is usually possible to keep a running sum of both the weights and the products as the data are being entered. One way of doing this is to accumulate the sum of the weights into the calculator’s memory while the sum of the products is cumulated on the display. If you are using a specialist statistics calculator, the task is generally very straightforward. Simply enter each price and its corresponding weight using the method described in your calculator instructions for finding a weighted mean.

Start of Activity

**Activity 8 Weighted means on your calculator**

Start of Question

Use your calculator to check that the sum of weights and sum of products of the data in Table 5 are, respectively, 1834 and 6973.436, and that the weighted mean is 3.802. (No solution is given to this activity.)

End of Question

End of Activity

Start of Activity

**Activity 9 Weighted mean electricity price**

Start of Question

Table 6 is similar to Table 5, but this time it presents the average price of electricity, in pence per kilowatt hour (kWh). These data are again for the year 2010 for typical consumers on credit tariffs in the same 14 cities we have been considering for gas prices, with the addition of Belfast. Again, the weights are the approximate populations of the relevant urban areas, in 10 000s.

Start of Table

Table 6 Populations and electricity prices in 15 cities

|  |  |  |  |
| --- | --- | --- | --- |
| **City** | **Price (p/kWh): x** | **Weight: w** | **Price times weight: x w** |
| Aberdeen | 13.76 | 19 |  |
| Belfast | 15.03 | 58 |  |
| Edinburgh | 13.86 | 42 |  |
| Leeds | 12.70 | 150 |  |
| Liverpool | 13.89 | 82 |  |
| Manchester | 12.65 | 224 |  |
| Newcastle-upon-Tyne | 12.97 | 88 |  |
| Nottingham | 12.64 | 67 |  |
| Birmingham | 12.89 | 228 |  |
| Canterbury | 12.92 | 5 |  |
| Cardiff | 13.83 | 33 |  |
| Ipswich | 12.84 | 14 |  |
| London | 13.17 | 828 |  |
| Plymouth | 13.61 | 24 |  |
| Southampton | 13.41 | 30 |  |
| **Sum** |  |  |  |

End of Table

Use these data to calculate the weighted mean electricity price. (Your calculator will almost certainly allow you to do this without writing out all the values in the x w column.)

End of Question

[View discussion - Activity 9 Weighted mean electricity price](" \l "Session4_Discussion7)

End of Activity

## Exercises on Section 2

The following exercises provide extra practice on the topics covered in Section 2.

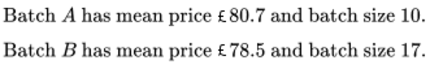
Start of Exercise

**Exercise 4 A combined batch of camera prices**

Start of Question

Find the mean price of the batch formed by combining the following two batches, uppercase A and uppercase B, of camera prices.

Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative114)

End of $1

End of Question

[View discussion - Exercise 4 A combined batch of camera prices](" \l "Session4_Discussion8)

End of Exercise

Start of Exercise

**Exercise 5 The mean price of fabric**

Start of Question

Suppose you buy 8.5 metres of fabric in a sale, at £10.95 per metre, to make some bedroom curtains. The following year you decide to make a matching bedspread and so you buy 6 metres of the same material, but the price is now £12.70 per metre. Calculate the mean price of all the material, in £ per metre.

End of Question

[View discussion - Exercise 5 The mean price of fabric](" \l "Session4_Discussion9)

End of Exercise

## 3 Measuring spread

As you have already seen, it is difficult to measure price changes when they so often vary from shop to shop and region to region. Taking some average value, such as the median or the mean, helps to simplify the problem. However, it would be a mistake to ignore the notion of spread, as averages on their own can be misleading.

Information about spread can be very important in statistical analysis, where you are often interested in comparing two or more batches. In this section we shall look first at measures of spread, and then at some methods of summarising the shape of a batch of data.

But how can spread be measured? Just as there are several ways of measuring location (mean, median, etc.), there are also several ways of measuring spread. Here, we shall examine two such measures: the range and the interquartile range. (A further, even more important, measure of spread is the standard deviation. It is, however, beyond the scope of this course.)

## 3.1 The range

The range is defined below.

Start of Box

**The range**

The range is the distance between the lower and the upper extremes. It can be calculated from the formula:

Start of $1

range = uppercase E subscript uppercase U end minus uppercase E subscript uppercase L end comma

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

End of $1

where uppercase E subscript uppercase U end is the upper extreme and uppercase E subscript uppercase L end is the lower extreme.

End of Box

Given an ordered batch of data, for example in a stemplot, the range can easily be calculated. However, the range tells us very little about how the values in the main body of the data are spread. It is also very sensitive to changes in the extreme values, like those considered in [Subsection 1.4](#open-u2ss1-4). It would be better to have a measure of spread that conveys more information about the spread of values in the main body of the data. One such measure is based upon the difference between two particular values in the batch, known as the **quartiles**. As the name suggests, the two quartiles lie one quarter of the way into the batch from either end. The major part of the next subsection describes how to find them.

## 3.2 Quartiles and the interquartile range

Finding the quartiles of a batch is very similar to finding the median.

In [Subsection 1.2](#open-u2ss1-2), we represented a batch as a V-shaped formation, with the median at the ‘hinge’ where the two arms of the V meet. The median splits the batch into two equal parts. Similarly, we can put another hinge in each side of the V and get four roughly equal parts, shaped like this: wedge wedge. For a batch of size 15, it looks like Figure 12.

Start of Figure

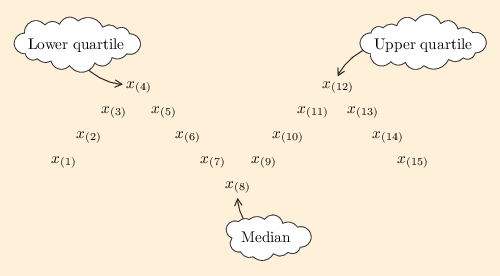


Figure 12 Median and quartiles

[View description - Figure 12 Median and quartiles](" \l "Session5_Description1)

End of Figure

The points at the side hinges, in this case x subscript open bracket 4 close bracket end and x subscript open bracket 12 close bracket end, are the quartiles. There are two quartiles which, as with the extremes, we call the **lower quartile** and the **upper quartile**. The lower quartile separates off the bottom quarter, or lowest 25%. The upper quartile separates off the top quarter, or highest 25%. They are denoted uppercase Q sub 1 and uppercase Q sub 3 respectively. (Sometimes they are referred to as the first quartile and the third quartile.)

You might be wondering, if these are uppercase Q sub 1 and uppercase Q sub 3, what happened to uppercase Q sub 2? Well, have a think about that for a moment.

uppercase Q sub 1separates the bottom quarter of the data (from the top three quarters), and uppercase Q sub 3 separates the bottom three quarters (from the top quarter). So it would make sense to say that uppercase Q sub 2 separates the bottom two quarters (from the top two quarters). But two quarters make a half, so uppercase Q sub 2 would denote the median, and since there is already a separate word for that, it’s not usual to call it the second quartile.

Usually we cannot divide the batch exactly into quarters. Indeed, this is illustrated in Figure 12 where the two central parts of the wedge wedge are larger than the outer ones. As with calculating the median for an even-sized batch, some rule is needed to tell us how many places we need to count along from the smallest value to find the quartiles. However, there are several alternatives that we could adopt and the particular rule described below is somewhat arbitrary. Different authors and different software may use slightly different rules. If your calculator can find quartiles, note that it may use a different rule.

As you might have expected, the rule involves dividing open bracket n+1 close bracket by 4, where n is the batch size (as opposed to dividing by 2 to find the median). However, the rule is slightly more complicated for the quartiles and it depends on whether n+1 is exactly divisible by 4.

Start of Box

**The quartiles**

The lower quartile uppercase Q sub 1 is at position fraction open bracket n +1 close bracket over 4 end in the ordered batch.

The upper quartile uppercase Q sub 3 is at position fraction 3 open bracket n +1 close bracket over 4 end in the ordered batch.

If open bracket n+1 close bracket is exactly divisible by 4, these positions correspond to a single value in the batch.

If open bracket n+1 close bracket is not exactly divisible by 4, then the positions are to be interpreted as follows.

* A position which is a whole number followed by fraction 1 over 2 end means ‘halfway between the two positions either side’ (as was the case for finding the median).
* A position which is a whole number followed by fraction 1 over 4 end means ‘one quarter of the way from the position below to the position above’. So for instance if a position is 5 fraction 1 over 4 end, the quartile is the number one quarter of the way from x subscript open bracket 5 close bracket end to x subscript open bracket 6 close bracket end.
* A position which is a whole number followed by fraction 3 over 4 end means ‘three quarters of the way from the position below to the position above’. So for instance if a position is 4 fraction 3 over 4 end, the quartile is the number three quarters of the way from x subscript open bracket 4 close bracket end to x subscript open bracket 5 close bracket end.

End of Box

Before we actually use these rules to find quartiles, let us look at some more examples of wedge wedge-shaped diagrams for different batch sizes n. The case where open bracket n+1 close bracket is exactly divisible by 4, so that fraction 1 over 4 end open bracket n+1 close bracket is a whole number, was shown in Figure 12. The following three figures show the three other possible scenarios, where open bracket n+1 close bracket is not exactly divisible by 4.

For n = 17, fraction 1 over 4 end open bracket n+1 close bracket = 4 fraction 1 over 2 end and fraction 3 over 4 end open bracket n+1 close bracket = 13 fraction 1 over 2 end. So uppercase Q sub 1 is halfway between x subscript open bracket 4 close bracket end and x subscript open bracket 5 close bracket end, and uppercase Q sub 3 is halfway between x subscript open bracket 13 close bracket end and x subscript open bracket 14 close bracket end.

Start of Figure

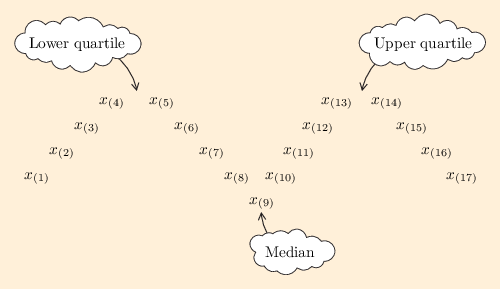


Figure 13 Quartiles for sample size n=17

[View description - Figure 13 Quartiles for sample size n=17](" \l "Session5_Description2)

End of Figure

For n = 18, fraction 1 over 4 end open bracket n+1 close bracket = 4 fraction 3 over 4 end and fraction 3 over 4 end open bracket n+1 close bracket = 14 fraction 1 over 4 end. So uppercase Q sub 1 is three quarters of the way from x subscript open bracket 4 close bracket end to x subscript open bracket 5 close bracket end, and uppercase Q sub 3 is one quarter of the way from x subscript open bracket 14 close bracket end to x subscript open bracket 15 close bracket end.

Start of Figure

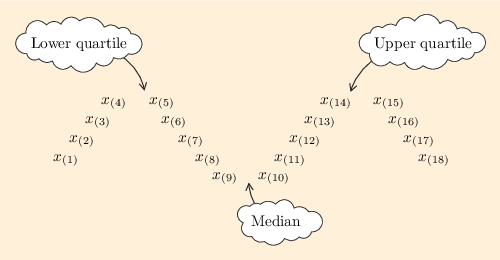


Figure 14 Quartiles for sample size n=18

[View description - Figure 14 Quartiles for sample size n=18](" \l "Session5_Description3)

End of Figure

For n = 20, fraction 1 over 4 end open bracket n+1 close bracket = 5 fraction 1 over 4 end and fraction 3 over 4 end open bracket n+1 close bracket = 15 fraction 3 over 4 end. So uppercase Q sub 1 is one quarter of the way from x subscript open bracket 5 close bracket end to x subscript open bracket 6 close bracket end, and uppercase Q sub 3 is three quarters of the way from x subscript open bracket 15 close bracket end to x subscript open bracket 16 close bracket end.

Start of Figure

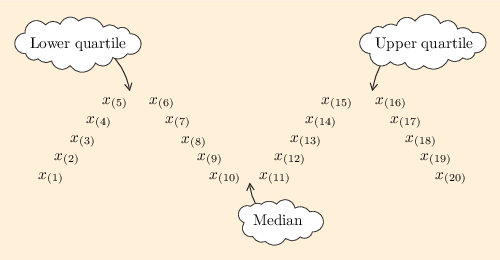


Figure 15 Quartiles for sample size n=20

[View description - Figure 15 Quartiles for sample size n=20](" \l "Session5_Description4)

End of Figure

Start of Box

**Example 14 Quartiles for the prices of small televisions**

Figure 15 showed you where the quartiles are for a batch of size 20. Let us now use the stemplot of the 20 television prices in Figure 16, which you first met in [Figure 5](#open-u2fig1-6a) (Subsection 1.2), to find the lower and upper quartiles, uppercase Q sub 1 and uppercase Q sub 3, of this batch.

Start of Figure

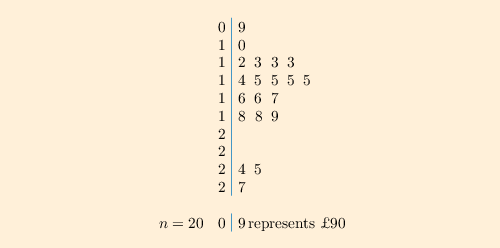


Figure 16 Prices of flat-screen televisions with a screen size of 24 inches or less

[View description - Figure 16 Prices of flat-screen televisions with a screen size of 24 inches or l ...](" \l "Session5_Description5)

End of Figure

To calculate the lower quartile uppercase Q sub 1 you need to find the number that is one quarter of the way from x subscript open bracket 5 close bracket end to x subscript open bracket 6 close bracket end. These values are both 130, so uppercase Q sub 1 is 130. To calculate the upper quartile uppercase Q sub 3 you need to find the number three quarters of the way from x subscript open bracket 15 close bracket end to x subscript open bracket 16 close bracket end. These values are both 180, so uppercase Q sub 3 is 180.

End of Box

That example was easier than it might have been, because for each quartile the two numbers we had to consider turned out to be equal!

Start of Box

**Example 15 Quartiles for the camera prices**

[Table 2](#open-u2table1-2) (Subsection 1.2) gave ten prices for a particular model of digital camera (in pounds). In order, the prices are as follows.

Start of $1

53 60 65 70 70 74 79 81 85 90

[View alternative description - Uncaptioned Equation](" \l "Session5_Alternative80)

End of $1

To find the lower and upper quartiles, uppercase Q sub 1 and uppercase Q sub 3, of this batch, first find fraction 1 over 4 end open bracket n+1 close bracket = 2 fraction 3 over 4 end and fraction 3 over 4 end open bracket n+1 close bracket = 8 fraction 1 over 4 end.

The lower quartile uppercase Q sub 1 is the number three quarters of the way from x subscript open bracket 2 close bracket end to x subscript open bracket 3 close bracket end. These values are 60 and 65. The difference between them is 65 minus 60=5, and three quarters of that difference is fraction 3 over 4 end times 5 = 3.75. Therefore uppercase Q sub 1 is 3.75 larger than 60, so it is 63.75. As with the median, in this course we will generally round the quartiles to the accuracy of the original data, so in this case we round to the nearest whole number, 64. In symbols, uppercase Q sub 1 = 60 + fraction 3 over 4 end open bracket 65 minus 60 close bracket = 63.75 simeq 64.

The upper quartile uppercase Q sub 3 is the number one quarter of the way from x subscript open bracket 8 close bracket end to x subscript open bracket 9 close bracket end. These values are 81 and 85. The difference between them is 85 minus 81=4, and one quarter of that difference is fraction 1 over 4 end times 4 = 1. Therefore uppercase Q sub 3 is 1 larger than 81, so it is 82. (No rounding necessary this time.) In symbols, uppercase Q sub 3 = 81 + fraction 1 over 4 end open bracket 85 minus 81 close bracket = 82.

End of Box

Example 15 is the subject of the following screencast. [Note that references to ‘the unit’ should be interpreted as ‘this course’. The original wording refers to the Open University course from which this material is adapted.]

Start of Media Content

Video content is not available in this format.

Screencast 3 Calculating quartiles

[View transcript - Screencast 3 Calculating quartiles](" \l "Session5_Transcript1)

Start of Figure



End of Figure

End of Media Content

Start of Activity

**Activity 10 Finding more quartiles**

Start of Question

(a)   Find the lower and upper quartiles of the batch of 15 coffee prices in Figure 17. (This batch of coffee prices was first introduced in [Table 1](#open-u2table1-1) of Subsection 1.1.)

Start of Figure

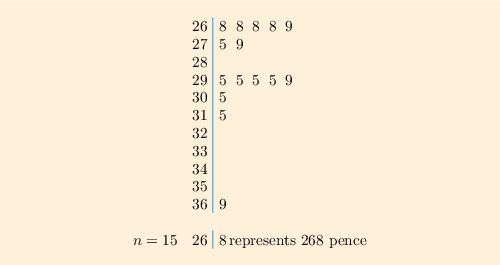


Figure 17 Stemplot of 15 coffee prices

[View description - Figure 17 Stemplot of 15 coffee prices](" \l "Session5_Description6)

End of Figure

End of Question

[View discussion - Part](" \l "Session5_Discussion1)

Start of Question

(b)   Find the lower and upper quartiles of the batch of 14 gas prices in Figure 18. (This batch of gas prices was first introduced in [Table 3](#open-u2table1-3) of Subsection 1.2.)

Start of Figure

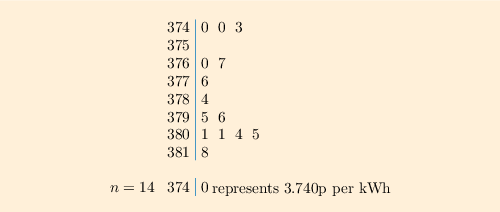


Figure 18 Stemplot of 14 gas prices

[View description - Figure 18 Stemplot of 14 gas prices](" \l "Session5_Description7)

End of Figure

End of Question

[View discussion - Part](" \l "Session5_Discussion2)

End of Activity

### A measure of spread

Now we can define a new measure of spread based entirely on the lower and upper quartiles.

Start of Box

**The interquartile range**

The interquartile range (sometimes abbreviated to **IQR**) is the distance between the lower and upper quartiles:

Start of $1

IQR = uppercase Q sub 3 minus uppercase Q sub 1 .

[View alternative description - Uncaptioned Equation](" \l "Session5_Alternative111)

End of $1

End of Box

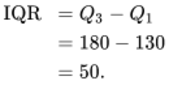
Note that this value is independent of the sizes of uppercase E subscript uppercase U end and uppercase E subscript uppercase L end.

Start of Box

**Example 16 The prices of small televisions, yet again!**

For the batch of 20 television prices in [Example 14](#open-u2exa3-0) (Subsection 3.2),

Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session5_Alternative114)

End of $1

So the interquartile range is £50.

End of Box

Start of Activity

**Activity 11 Coffee prices again**

Start of Question

Calculate both the range and the interquartile range of the batch of 15 coffee prices, last seen in [Figure 17](#open-u2fig3-1a) (Subsection 3.2).

End of Question

[View discussion - Activity 11 Coffee prices again](" \l "Session5_Discussion3)

End of Activity

Start of Activity

**Activity 12 Interquartile range of gas prices**

Start of Question

In [Activity 10](#open-u2act3-1)(b) (Subsection 3.2) you found the quartiles of the 14 gas prices from [Activity 2](#open-u2act1-1) (Subsection 1.2). Find the interquartile range.

End of Question

[View discussion - Activity 12 Interquartile range of gas prices](" \l "Session5_Discussion4)

End of Activity

You may be wondering why you are being asked to learn a new measure of spread when you already know the range. As a measure of spread, the range open bracket uppercase E subscript uppercase U end minus uppercase E subscript uppercase L end close bracket is not very satisfactory because it is not resistant to the effects of unrepresentative extreme values. (Resistant measures were explained in [Subsection 1.4](#open-u2ss1-4).) The interquartile range, by contrast, is a highly resistant measure of spread (because it is not sensitive to the effects of values lying outside the middle 50% of the batch) and it is generally the preferred choice.

Start of Box

**Example 17 Comparing the resistance of the range and the IQR**

Suppose the price of the most expensive jar of coffee is reduced from 369p to 325p. How does this affect the range and the interquartile range of the batch of coffee prices in [Figure 17](#open-u2fig3-1a) (Subsection 3.2)?

The new range is

Start of $1

uppercase E subscript uppercase U end minus uppercase E subscript uppercase L end = 325 p minus 268 p = 57 p comma

[View alternative description - Uncaptioned Equation](" \l "Session5_Alternative121)

End of $1

a lot less than the original value of 101p (found in Activity 11). The interquartile range is unchanged.

End of Box

## 3.3 The five-figure summary and boxplots

As well as giving us a new measure of spread – the interquartile range – the quartiles are important figures in themselves. Our wedge wedge-shaped diagram, Figure 19, gives five important points which help to summarise the shape of a distribution: the **median**, the **two quartiles** and the **two extremes**.

Start of Figure

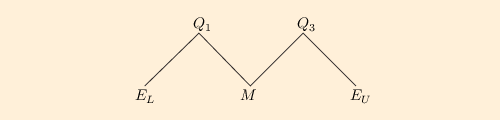


Figure 19 Values in a five-figure summary

[View description - Figure 19 Values in a five-figure summary](" \l "Session5_Description8)

End of Figure

These are conveniently displayed in the following form, called the **five-figure summary** of the batch.

Start of Box

**Five-figure summary**

Start of Figure

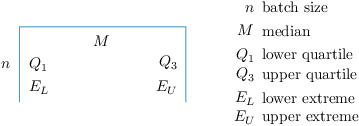


Figure 20

[View description - Figure 20](" \l "Session5_Description9)

End of Figure

End of Box

Start of Box

**Example 18 Five-figure summary for television price data**

For the television price data, we have n = 20, uppercase M=150, uppercase Q sub 1 =130, uppercase Q sub 3 =180, uppercase E subscript uppercase L end =90 and uppercase E subscript uppercase U end =270. (You last saw these data in [Figure 16](#open-u2fig3-3), Subsection 3.2.)

Therefore, the five-figure summary of this batch is

Start of Figure

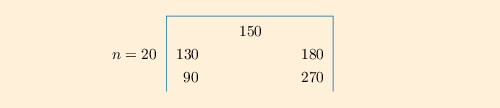


Figure 21

[View description - Figure 21](" \l "Session5_Description10)

End of Figure

This diagram contains the following information about the batch of prices.

* The general level of prices, as measured by the median, is £150.
* The individual prices vary from £90 to £270.
* About 25% of the prices were less than £130.
* About 25% of the prices were more than £180.
* About 50% of the prices were between £130 and £180.

End of Box

We hope you agree that the five-figure summary is quite an efficient way of presenting a summary of a batch of data.

The five values in a five-figure summary can be very effectively presented in a special diagram called a **boxplot**. For the 14 gas prices ([Figure 15](#open-u2fig3-1b), Subsection 3.2) the diagram looks like Figure 22.

Start of Figure

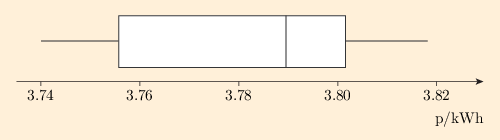


Figure 22 Boxplot of batch of 14 gas prices

[View description - Figure 22 Boxplot of batch of 14 gas prices](" \l "Session5_Description11)

End of Figure

The central feature of this diagram is a box – hence the name boxplot. The box extends from the lower quartile (at the left-hand edge of the box) to the upper quartile (the right-hand edge). This part of the diagram contains 50% of the values in the batch. The length of this box is thus the interquartile range.

Outside the box are two whiskers. (Boxplots are sometimes called box-and-whisker diagrams.) In many cases, such as in Figure 22, the whiskers extend all the way out to the extremes. Each whisker then covers the end 25% of the batch and the distance between the two whisker-ends is then the range. (You will see examples later where the whiskers do not go right out to the extremes.)

So far we have dealt with four figures from the five-figure summary: the two quartiles and the two extremes. The remaining figure is perhaps the most important: it is the median, whose position is shown by putting a vertical line through the box.

Thus a boxplot shows clearly the division of the data into four parts: the two whiskers and the two sections of the box; these are the four parts of the wedge wedge-shaped diagram and each contains (approximately) 25% of values in the batch (see Figure 21).

Start of Box

**John W. Tukey (1915–2000), inventor of the five-figure summary and boxplot**

John Tukey was a prominent and prolific US statistician, based at Princeton University and Bell Laboratories. As well as working in some very technical areas, he was a great promoter of simple ways of picturing and summarising data, and invented both the five-figure summary and the boxplot (except that he called them the ‘five-number summary’ and the ‘box-and-whisker plot’).

He had what has been described as an ‘unusual’ lecturing style. The statistician Peter McCullagh describes a lecture he gave at Imperial College, London in 1977:

Start of Quote

Tukey ambled to the podium, a great bear of a man dressed in baggy pants and a black knitted shirt. These might once have been a matching pair, but the vintage was such that it was hard to tell. …The words came …, not many, like overweight parcels, delivered at a slow unfaltering pace. …Tukey turned to face the audience …. ‘Comments, queries, suggestions?’ he asked …. As he waited for a response, he clambered onto the podium and manoeuvred until he was sitting cross-legged facing the audience. …We in the audience sat like spectators at the zoo waiting for the great bear to move or say something. But the great bear appeared to be doing the same thing, and the feeling was not comfortable. …After a long while, …he extracted from his pocket a bag of dried prunes and proceeded to eat them in silence, one by one. The war of nerves continued …four prunes, five prunes. …How many prunes would it take to end the silence?

(Source: McCullagh, P. (2003) ‘John Wilder Tukey’, Biographical Memoirs of Fellows of the Royal Society, vol. 49, pp. 537–55.)

End of Quote

End of Box

Start of Figure

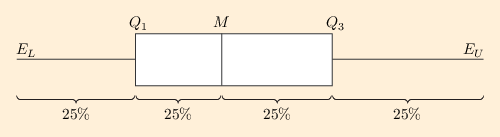


Figure 23 A standard boxplot with annotation

[View description - Figure 23 A standard boxplot with annotation](" \l "Session5_Description12)

End of Figure

A typical boxplot looks something like Figure 23 because in most batches of data the values are more densely packed in the middle of the batch and are less densely packed in the extremes. This means that each whisker is usually longer than half the length of the box. This is illustrated again in the next example.

Start of Box

**Example 19 Boxplot for the prices of small televisions**

The boxplot for the batch of 20 television prices (last worked with in Example 18) is shown in Figure 24.

Start of Figure

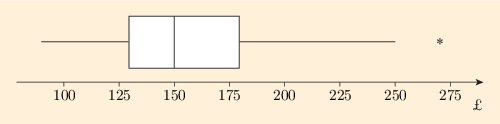


Figure 24 Boxplot of batch of 20 television prices

[View description - Figure 24 Boxplot of batch of 20 television prices](" \l "Session5_Description13)

End of Figure

You can see that each whisker is longer than half the length of the box.

However, this boxplot has a new feature. The whisker on the left goes right down to the lower extreme. But the whisker on the right does not go right to the upper extreme. The highest extreme data value, 270, which might potentially be regarded as an outlier, is marked separately with a star. Then the whisker extends only to cover the data values that are not extreme enough to be regarded as potential outliers. The highest of these values is 250.

(This course does not describe the rule to decide which data values (if any) can be regarded as potential outliers that are plotted separately on the diagram. This is another issue that may be dealt with differently by different authors and different software.)

End of Box

Example 19 is the subject of the following screencast. [Note that the reference to ‘Unit 2’ should be ‘this course’ and ‘Figure 18’ should be ‘Figure 23’. Unit 2 and Figure 18 are references to the Open University course from which this material is adapted.]

Start of Media Content

Video content is not available in this format.

Screencast 4 Interpreting a boxplot

[View transcript - Screencast 4 Interpreting a boxplot](" \l "Session5_Transcript2)

Start of Figure



End of Figure

End of Media Content

One important use of boxplots is to picture and describe the overall shape of a batch of data.

Start of Box

**Example 20 Skew televisions**

The stemplot of small television prices, last seen in [Figure 16](#open-u2fig3-3) (Subsection 3.2), shows a lack of symmetry. Since the higher values are more spread out than the lower values, the data are right-skew.

The boxplot of these data, given in [Figure 22](#open-u2fig3-9), also shows this right-skew fairly clearly. In the box, the right-hand part (corresponding to higher prices) is rather longer than the left-hand part, and the right-hand whisker is longer than the left-hand whisker.

End of Box

Start of Activity

**Activity 13 Skew gas prices?**

Start of Question

A stemplot of the gas price data from [Activity 2](#open-u2act1-1) (Subsection 1.2) is shown, yet again, in Figure 25.

Start of Figure

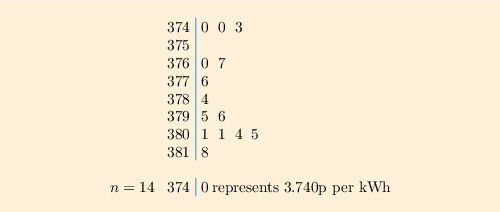


Figure 25 Stemplot of 14 gas prices

[View description - Figure 25 Stemplot of 14 gas prices](" \l "Session5_Description14)

End of Figure

End of Question

Start of Question

(a)   Prepare a five-figure summary of the batch.

End of Question

[View discussion - Part](" \l "Session5_Discussion5)

Start of Question

(b)   Figure 27 shows the boxplot of these data that you have already seen in Figure 22. What do the stemplot and boxplot tell us about the symmetry and/or skewness of the batch?

Start of Figure

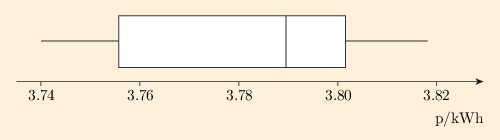


Figure 27 Boxplot of batch of 14 gas prices

[View description - Figure 27 Boxplot of batch of 14 gas prices](" \l "Session5_Description16)

End of Figure

End of Question

[View discussion - Part](" \l "Session5_Discussion6)

End of Activity

Start of Box

**Example 21 Camera prices: skew or not?**

In Example 20 and Activity 13 you saw how boxplots look for batches of data that are right-skew or left-skew. What happens in a batch that is more symmetrical?

For the small batch of camera prices from [Table 2](#open-u2table1-2) (Subsection 1.2), a (stretched) stemplot is shown in Figure 28.

Start of Figure

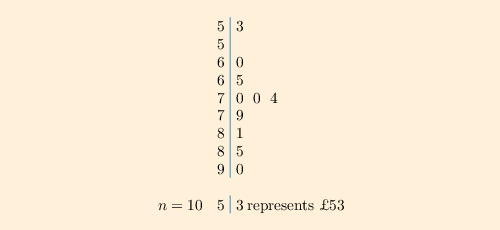


Figure 28 Stemplot of ten camera prices

[View description - Figure 28 Stemplot of ten camera prices](" \l "Session5_Description17)

End of Figure

The stemplot looks reasonably symmetric.

A boxplot of the data, Figure 29, confirms the impression of symmetry. The two parts of the box are roughly equal in length, and the two whiskers are also roughly equal in length.

Start of Figure



Figure 29 Boxplot of batch of ten camera prices

[View description - Figure 29 Boxplot of batch of ten camera prices](" \l "Session5_Description18)

End of Figure

End of Box

You have now spent quite a lot of time looking at various ways of investigating prices and, in particular, at methods of measuring the location and spread of the prices of particular commodities.

In order to begin to answer our question, Are people getting better or worse off?, we need to know not just location (and spread) of prices but also how these prices are changing from year to year. That is the subject of the rest of this course.

## Exercises on Section 3

The following exercises provide extra practice on the topics covered in Section 3.

Start of Exercise

**Exercise 6 Finding quartiles and the interquartile range**

Start of Question

(a)   For the arithmetic scores in [Exercise 1](#open-u2exe1-1) (Section 1), find the quartiles and calculate the interquartile range. The stemplot of the scores is given below.

Start of Figure

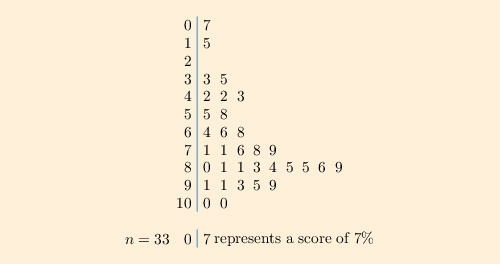


Figure 30 Stemplot of arithmetic stores

[View description - Figure 30 Stemplot of arithmetic stores](" \l "Session5_Description19)

End of Figure

End of Question

[View discussion - Part](" \l "Session5_Discussion7)

Start of Question

(b)   For the television prices in [Exercise 1](#open-u2exe1-1), find the quartiles and calculate the interquartile range. The table of prices is given below.

Start of Table

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 170 | 180 | 190 | 200 | 220 | 229 | 230 | 230 | 230 |
| 230 | 250 | 269 | 269 | 270 | 279 | 299 | 300 | 300 |
| 315 | 320 | 349 | 350 | 400 | 429 | 649 | 699 |  |

End of Table

End of Question

[View discussion - Part](" \l "Session5_Discussion8)

End of Exercise

Start of Exercise

**Exercise 7 Some five-figure summaries**

Start of Question

Prepare a five-figure summary for each of the two batches from [Exercise 1](#open-u2exe1-1).

End of Question

Start of Question

(a)   For the arithmetic scores, the median is 79% (found in [Exercise 1](#open-u2exe1-1)), and you found the quartiles and interquartile range in Exercise 6.

End of Question

[View discussion - Part](" \l "Session5_Discussion9)

Start of Question

(b)   For the television prices, the median is £270 (found in [Exercise 1](#open-u2exe1-1)), and you found the quartiles and interquartile range in Exercise 6.

End of Question

[View discussion - Part](" \l "Session5_Discussion10)

End of Exercise

Start of Exercise

**Exercise 8 Boxplots and the shape of distributions**

Start of Question

Boxplots of the two batches used in Exercises 1, 6 and 7 are shown in Figures 33 and 34. On the basis of these diagrams, comment on the symmetry and/or skewness of these data.

Start of Figure

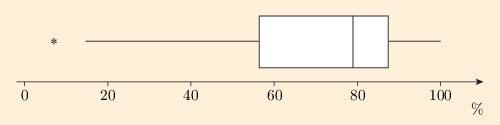


Figure 33 Boxplot of batch of 33 arithmetic scores

[View description - Figure 33 Boxplot of batch of 33 arithmetic scores](" \l "Session5_Description22)

End of Figure

Start of Figure

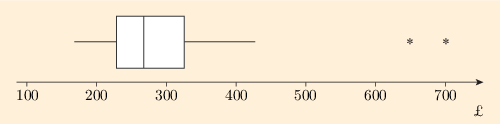


Figure 34 Boxplot of batch of 26 television prices

[View description - Figure 34 Boxplot of batch of 26 television prices](" \l "Session5_Description23)

End of Figure

End of Question

[View discussion - Exercise 8 Boxplots and the shape of distributions](" \l "Session5_Discussion11)

End of Exercise

## 4 A simple chained price index

You have already seen that it is not a simple task to measure the price of even a single commodity at a fixed time and place. Measuring the change in price of a single commodity from one year to the next will be even more complicated but, as was said in Subsection 1.1, to answer our question it is necessary to measure the changes in the prices of the whole range of goods and services which people use. Moreover, since we wish to know how all the different changes in the prices of these goods and services affect people, we need to take into account those people’s consumption patterns. For example, a large increase in the price of high-quality caviar will not affect most people’s budgets since most households’ shopping lists do not include this commodity!

This makes the task of measuring price changes and examining how they affect us seem exceedingly difficult; but such a task is carried out in the UK regularly each month, organised by the Office for National Statistics. (Most of the prices are actually collected by a market research company under contract to the Office for National Statistics.) The results of their data collection and subsequent calculations are summarised in two measures called the Consumer Prices Index (CPI) and the Retail Prices Index (RPI).

These indices do not measure prices. (‘Indices’ is the plural of ‘index’.) Each is an index of price changes over time, and one or both of these indices are commonly used when people make comparisons about the cost of living. They are highly relevant measures for those engaged in wage bargaining.

The RPI and the CPI are both ‘chained’ in the sense that the index value for each year is linked to the year before. The very first link in the chain is called the base year and it is given an index value of 100.

Start of Figure

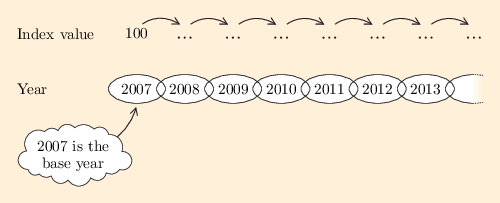


Figure 35 A chained index

[View description - Figure 35 A chained index](" \l "Session6_Description1)

End of Figure

## 4.1 A two-commodity price index

Section 5 includes an outline of how the information used to calculate the official UK price indices is collected, and describes how the indices are calculated. To introduce ideas, in this section we describe a very much simpler example of a price index calculation. It uses exactly the same basic method of calculation as the actual Retail Prices Index. (Not every index is calculated in this way.)

The context is a mythical computing company, Gradgrind Ltd.

Gradgrind Ltd uses both gas and electricity in its operations. Table 7 shows the price they paid for each fuel in 2007 and 2008. The prices are shown in £ per megawatt hour (MWh). (It is more usual, in the UK, for prices to be quoted in pence per kilowatt hour (p/kWh). Here, £/MWh have been used simply to make some of the later calculations a little more straightforward. Because there are 100 pence in £1 and 1000 kilowatts in a megawatt, £10/MWh is exactly the same price as 1p/kWh – so Gradgrind’s gas price in 2007, for instance, was 2.4p/kWh.)

Start of Table

Table 7 Gradgrind’s energy prices in 2007 and 2008

|  |  |  |
| --- | --- | --- |
| **Energy type** | **2007** | **2008** |
| Gas (£/MWh) | 24 | 29 |
| Electricity (£/MWh) | 76 | 87 |

End of Table

If we were interested in looking at the change in price of just one of these fuels, say gas, things would be relatively straightforward. For instance, it might well be appropriate to look at the increase in price as a percentage of the price in 2007.

Start of Activity

**Activity 14 Gradgrind’s gas price increase**

Start of Question

Work out the increase in Gradgrind’s gas price between 2007 and 2008 as a percentage of the 2007 price.

End of Question

[View discussion - Activity 14 Gradgrind’s gas price increase](" \l "Session6_Discussion1)

End of Activity

So we could say that, for this company at least, gas has gone up by 20.8%. In other words, for every £1 they spent on gas in 2007, they would have spent £1.208 in 2008 if they had bought the same amount of gas in each year. Or putting it another way, for every 100 units of money (pence, pounds, whatever) they spent in 2007, they would have spent 120.8 units of money in 2008 if they had bought the same amount. So a way of representing this price change would have been to define an index for the gas price such that it takes the value 100 for 2007, and 120.8 for 2008.

Notice that the value of the gas price index for 2008 could be calculated as

Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative5)

End of $1

That is, the value of the index in one year is the value of the index in the previous year multiplied by a price ratio, in this case the gas price ratio for 2008 relative to 2007. This ratio, as a number, is 1.208.

But Gradgrind did not only use gas, they used electricity as well, and the aim here is to find a representation of their overall fuel price change, not just the change in gas prices.

An electricity price ratio for 2008 relative to 2007 can be worked out, like the gas price ratio. It is fraction 87 over 76 end simeq 1.145.

Start of Activity

**Activity 15 Gradgrind’s electricity price index**

Start of Question

Use the electricity price ratio above to find the increase in Gradgrind’s electricity price between 2007 and 2008 as a percentage of the 2007 price. What would the 2008 value be for a price index of Gradgrind’s electricity price alone, calculated in the same way as the gas price index (with 2007 as the base year)?

End of Question

[View discussion - Activity 15 Gradgrind’s electricity price index](" \l "Session6_Discussion2)

End of Activity

But this has got us no further in finding a price index that simultaneously covers both fuels.

One possibility might be to look at how Gradgrind’s total expenditure on these two fuels changed from 2007 to 2008. The expenditures are given in Table 8.

Start of Table

Table 8 Gradgrind’s energy expenditure (£) in 2007 and 2008

|  |  |  |
| --- | --- | --- |
| **Energy type** | **2007** | **2008** |
| Gas | 9 298 | 8 145 |
| Electricity | 3 205 | 2 991 |
| **Total** | **12 503** | **11 136** |

End of Table

This seems not to have helped. The total expenditure went down, but you have already seen that the prices of both gas and electricity went up.

Start of Activity

**Activity 16 How much fuel did Gradgrind use?**

Start of Question

Use the data in Tables 7 and 8 to find the quantity of each fuel that Gradgrind used in 2007 and 2008 (in MWh). Hence explain why the energy expenditure fell.

End of Question

[View discussion - Activity 16 How much fuel did Gradgrind use?](" \l "Session6_Discussion3)

End of Activity

Remember the aim is to produce a measure of price changes. So looking at expenditure changes does not do the right thing, since expenditure depends on the amount of fuel consumed as well as the price.

One possibility might be as follows. We could work out how much Gradgrind would have spent on fuel in 2008 if the consumptions of both fuels had not changed from 2007. That would remove the effect of any changes in consumption. Then we could calculate an overall energy price ratio for 2008 relative to 2007 by dividing the total expenditure on energy for 2008 (using the 2007 consumption figures) by the total expenditure on energy for 2007 (again using the 2007 consumption figures).

You should have found, in Activity 16, that the quantities of gas and electricity consumed in 2007 were, respectively, 387.4 MWh and 42.2 MWh. To buy those quantities at 2008 prices would have cost (in £): 29 times 387.4 = 11234.6 for the gas and 87 times 42.2 = 3671.4 for the electricity, giving a total expenditure of

Start of $1

pounds open bracket 11234.6 + 3671.4 close bracket = pounds 14906.0.

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative14)

End of $1

So a reasonable overall energy price ratio for 2008 relative to 2007 can be found by dividing this total by the 2007 total expenditure, again calculated using the 2007 consumptions. The appropriate figure for 2007 is just the actual total expenditure, which (in £) was 9298 + 3205 =12503 (see [Table 8](#open-u2tab4-2)). This gives an overall energy price ratio for 2008 relative to 2007 as

Start of $1

fraction 14906 .0 over 12503 end simeq 1.192.

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative16)

End of $1

Now we have an appropriate price ratio, the Gradgrind energy price index can be set as 100 for the base year, 2007, and the value of the 2008 index is found by multiplying the 2007 index value by the price ratio:

Start of $1

2008 index = 100 times 1.192 = 119.2.

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative17)

End of $1

This is indeed how a chained index of this kind is calculated – but the calculations are rather messy. You might be wondering whether it would be simpler to calculate the overall energy price ratio as a weighted mean of the two price ratios for the two fuels, in much the same way that weighted means were used to combine prices in [Section 2](#open-u2s2). If you did think this, you would be right – and furthermore, the resulting overall energy price ratio is exactly the same as has just been found, if we make the right choice of weights. The overall energy price ratio for 2008 relative to 2007 is just a weighted mean of the two price ratios for gas and electricity, with the 2007 expenditures as weights.

Just to show it really does come to the same thing, let us see how it works with the numbers, using the formula for weighted means in [Subsection 2.3](#open-u2ss2-3).

Start of Table

|  |  |  |
| --- | --- | --- |
| **Energy type** | **Price ratio (2008 relative to 2007): x** | **Weight (2007 expenditure): w** |
| Gas | 1.208 | 9298 |
| Electricity | 1.145 | 3205 |

End of Table

The weighted average of these price ratios is

Start of $1

fraction open bracket 1 .208 times 9298 close bracket + open bracket 1 .145 times 3205 close bracket over 9298 +3205 end = fraction 14901 .709 over 12503 end simeq 1.192 comma

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative20)

End of $1

giving the same value for the overall energy price ratio for 2008 relative to 2007 as we found earlier. (And this is not some sort of fluke that applies only to these particular numbers; it can be shown mathematically that it always works.)

Start of Activity

**Activity 17 Gradgrind’s energy price ratio for 2009 relative to 2008**

Start of Question

Start of Table

Table 9 Gradgrind’s energy prices and expenditures for 2008 and 2009

|  |  |  |
| --- | --- | --- |
| **Energy type: price and expenditure** | **2008** | **2009** |
| Gas price (£/MWh) | 29 | 30 |
| Gas expenditure (£) | 8 145 | 23 733 |
| Electricity price (£/MWh) | 87 | 98 |
| Electricity expenditure (£) | 2 991 | 2 275 |

End of Table

End of Question

Start of Question

(a)   Using the data in Table 9, calculate the price ratios for gas and for electricity, in each case for 2009 relative to 2008.

End of Question

[View discussion - Part](" \l "Session6_Discussion4)

Start of Question

(b)   With the 2008 expenditures as weights, use your answers to part (a) to calculate the overall energy price ratio for 2009 relative to 2008.

End of Question

[View discussion - Part](" \l "Session6_Discussion5)

Start of Question

(c)   Now see what happens if you use the 2009 expenditures as weights to calculate the overall energy price ratio for 2009 relative to 2008. How do the results of the calculation differ from what you got in part (b)?

End of Question

[View discussion - Part](" \l "Session6_Discussion6)

End of Activity

The reason that the price ratios you calculated in parts (b) and (c) in Activity 17 were so different is that Gradgrind’s ‘energy mix’ changed a lot over the year. Compared with 2008, in 2009 they spent a great deal more on gas but less on electricity. The weighted mean of the gas and electricity price ratios is, in both cases, nearer the price ratio for gas than that for electricity – this is Rule 2 for weighted means – but it is even nearer the gas weighted mean when the 2009 expenditures are used. This is because the weight for gas is proportionally much greater than it is when the 2008 expenditures are used as weights.

This all shows that it does make a difference which expenditures are used as weights. In practice, it is much more common to use the expenditures from the earlier year – 2008 in this case – as weights. In some circumstances, though, there are good reasons for using the later year, or indeed some more complicated set of weights that depend on both expenditures. However, in this course we shall use the expenditures from the earlier year to provide the weights, partly because that matches more closely what is done in calculating the official UK price indices.

Another possibility for weights would have been to continue to use the 2007 expenditures. These were used to find the overall energy price ratio for 2008 relative to 2007 and could be used for later years as well. Again, in some circumstances this would make sense, but here the pattern of Gradgrind’s fuel expenditure has changed a lot over time, and weights should change in consequence. To continue to use the 2007 expenditures for all later years would mean that this change in the relative importance to Gradgrind of the two fuels would never be taken into account. Instead, to obtain the overall energy price ratio from one year to the next, we use the fuel expenditures in the earlier year as weights, so each year the weights change.

That determines the choice of weights in forming an overall price ratio. Now, how is that used to find the energy price index? Here we simply continue the ‘chaining’ that started when finding the 2008 index: the 2009 index is found by multiplying the value of the index for the previous year, 2008, by the overall energy price ratio for 2009 relative to 2008. The value of the index for 2008 was calculated earlier as 119.2, and (using the weights from the previous year) the overall energy price ratio for 2009 relative to 2008 was found in [Activity 17](#open-u2act4-4)(b) as 1.059. So the value of Gradgrind’s energy price index for 2009 is

Start of $1

119.2 times 1.059 simeq 126.2.

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative25)

End of $1

(So, in a particular kind of average way, Gradgrind’s energy prices for 2009 have risen by 26.2% since the base year, 2007.)

In general, the value index for a particular year is found by multiplying the value of the index for the previous year by the overall energy price ratio for that year relative to the previous year. This is illustrated in Figure 36.

Start of Figure

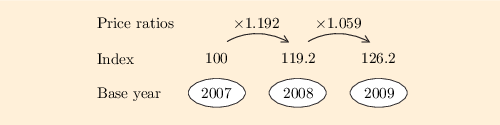


Figure 36 Determining a chained price index

[View description - Figure 36 Determining a chained price index](" \l "Session6_Description2)

End of Figure

In the process of chaining, the overall price ratio is calculated anew each year, looking back only at the previous year. The ratio is used to ‘chain’ to earlier years and hence determine the value of the index. This method of calculating a **chained price index** is summarised below. Although there were only two commodities (gas and electricity) in Gradgrind’s index, this summary is not restricted to two commodities.

Start of Box

**Procedure used to calculate a chained price index**

1. For each year calculate the following.
   * The **price ratio** for each commodity covered by the index:

Start of $1

fraction price that year over price previous year end .

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative26)

End of $1

* + The weighted mean of all these price ratios, using as weights the expenditure on each commodity in the previous year. This weighted mean is called the **all-commodities price ratio**.

1. For each year, the value of the index is

Start of $1

value of index for previous year times all minus commodities price ratio .

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative27)

End of $1

The value of the index in the first year is set at 100; this date is the **base date** of the index.

End of Box

Start of Activity

**Activity 18 Gradgrind’s energy price index for 2010**

Start of Question

Use the data in Table 10, and other necessary numbers from previous calculations, to calculate the value of Gradgrind’s energy price index for 2010.

Start of Table

Table 10 Gradgrind’s energy prices and expenditures for 2009 and 2010

|  |  |  |
| --- | --- | --- |
| **Energy type: price and expenditure** | **2008** | **2009** |
| Gas price (£/MWh) | 30 | 28 |
| Gas expenditure (£) | 23 733 | 23 969 |
| Electricity price (£/MWh) | 98 | 88 |
| Electricity expenditure (£) | 2 275 | 2 920 |

End of Table

End of Question

[View discussion - Activity 18 Gradgrind’s energy price index for 2010](" \l "Session6_Discussion7)

End of Activity

The Retail Prices Index (RPI), published by the UK Office for National Statistics, is calculated once a month rather than once a year, but the method used is basically that outlined above, though with far more than two commodities. The process of finding the weights in the Retail Prices Index is also more complicated, because it involves taking into account the expenditures of millions of people as measured in a major survey. However, the principles are the same as for Gradgrind. The calculation each January follows exactly this method. In the other 11 months of the year, the calculation is very similar but uses only the increases in prices since the previous January. (See [Subsection 5.2](#open-u2ss5-2) for the details of these calculations.) In the next section, you will learn more about how all this works.

## Exercise on Section 4

The following exercise provides extra practice on the Section 4 material.

Start of Exercise

**Exercise 9 Gradgrind’s energy price index for 2011**

Start of Question

Use the data in Table 11, and the fact that Gradgrind’s energy price index for 2010 was 117.4 (as found in [Activity 18](#open-u2act4-5)), to calculate the value of Gradgrind’s energy price index for 2011.

Start of Table

Table 11 Gradgrind’s energy prices and expenditures for 2010 and 2011

|  |  |  |
| --- | --- | --- |
| **Energy type: price and expenditure** | **2010** | **2011** |
| Gas price (£/MWh) | 28 | 30 |
| Gas expenditure (£) | 23 969 | 24 282 |
| Electricity price (£/MWh) | 88 | 86 |
| Electricity expenditure (£) | 2 920 | 3 117 |

End of Table

End of Question

[View discussion - Exercise 9 Gradgrind’s energy price index for 2011](" \l "Session6_Discussion8)

End of Exercise

## 5 The UK government price indices

Start of Quote

‘The huge squeeze on Brits was laid bare today as figures showed inflation has soared to a 20-year high.’ (The Sun, 18 October 2011)

‘Overall, prices in the economy rose 0.6% on the month from August.’ (Guardian, 18 October 2011)

‘Inflation in the UK continued to fall in February, thanks largely to lower gas and electricity bills.’ (BBC News website, 20 March 2012)

‘UK inflation rises more than expected.’ (Daily Telegraph, 16 August 2011)

End of Quote

How often have you read or heard statements like these in the media? Have you ever wondered how ‘inflation’ is measured, or precisely what is meant by a statement such as ‘prices rose by 0.6%’? In Subsection 5.3, you will see that ‘rates of inflation’ are often calculated in the UK using an index of prices paid by consumers, the Consumer Prices Index (CPI), or another slightly different index, the Retail Prices Index (RPI). These indices may be used to calculate the percentage by which prices in general have risen over any given period, and (roughly speaking) this is what is meant by inflation. But what exactly do these price indices measure, and how are they calculated? These are the questions that are addressed in this section.

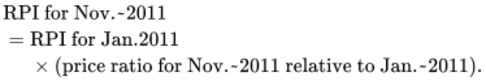
## 5.1 What are the CPI and RPI?

The CPI and the RPI are the main measures used in the UK to record changes in the level of the prices most people pay for the goods and services they buy. The RPI is intended to reflect the average spending pattern of the great majority of private households. Only two classes of private households are excluded, on the grounds that their spending patterns differ greatly from those of the others: pensioner households and high-income households. The CPI, however, has a wider remit – it is intended to reflect the spending of all UK residents, and also covers some costs incurred by foreign visitors to the UK.

The CPI and RPI are calculated in a similar way to the price index for Gradgrind Ltd’s energy in Section 4. However, they are calculated once a month rather than just once a year, and are based on a very large ‘**basket of goods**’. The contents of the basket and the weights assigned to the items in the basket are updated annually to reflect changes in spending patterns (as was the case with Gradgrind’s index for energy prices), and the index is ‘chained’ to previous values. However, once decided on at the beginning of the year, the contents of the basket and their weights remain fixed throughout the year.

For the RPI, the price ratio for the basket each month is calculated relative to the previous January. Then the value of the index is obtained by multiplying the value of the index for the previous January by this price ratio. For example,

Start of $1

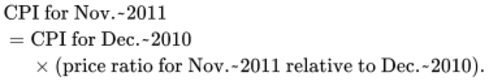


[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative1)

End of $1

The CPI works in much the same way, except that price ratios are calculated relative to the previous December. So, for example,

Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative2)

End of $1

Since these price indices are calculated from price ratios, they measure price changes in terms of the ratio of the overall level of prices in a given month to the overall level of prices at an earlier date. In practice, data on most prices are collected on a particular day near the middle of the month; the values of the RPI and CPI calculated using these data are referred to simply as the values of the RPI and CPI for the month. For example, the RPI took the value 239.9 in February 2012. This value measures the ratio of the overall level of prices in February 2012 to the overall level of prices on a date at which the index was fixed at its starting value of 100. This date, called a base date, is 13 January 1987 (at the time of writing). Thus the general level of prices in February 2012, as measured by the RPI, was 239.9/100 =2.399 times the general level of prices in January 1987. The base date has no significance other than to act as a reference point. (The CPI base date is 2005 and this refers to the average level of prices throughout 2005, not to a specific date in 2005.)

The RPI and CPI are each based on a very large ‘basket’ of goods and services. (The two baskets are similar, but not exactly the same.) Each contains around 700 items including most of the usual things people buy: food, clothes, fuel, household goods, housing, transport, services, and so on. Each basket is an ‘average’ basket for a broad range of households. The items in the baskets are often grouped into broader categories. For the RPI, the five fundamental groups are: ‘Food and catering’, ‘Alcohol and tobacco’, ‘Housing and household expenditure’, ‘Personal expenditure’ and ‘Travel and leisure’. These groups are divided into 14 more detailed subgroups (which are further divided into sections), as shown in Figure 37. The items in the CPI basket are divided into 12 broad groupings called divisions, which are further subdivided.

Start of Figure

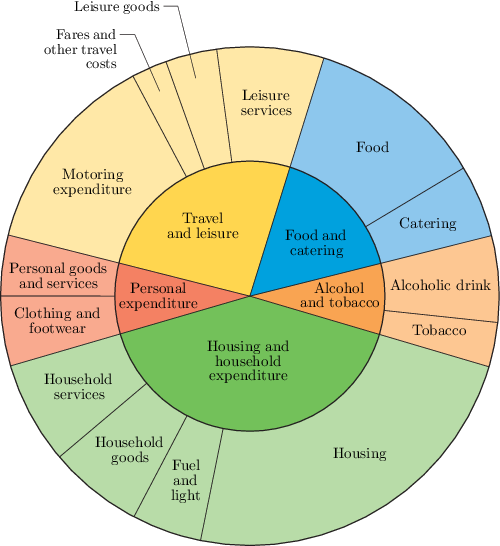


Figure 37 Structure of the RPI in 2012 (based on data from the Office for National Statistics)

[View description - Figure 37 Structure of the RPI in 2012 (based on data from the Office for National ...](" \l "Session7_Description1)

End of Figure

The inner circle shows the five groups, and the outer ring shows the 14 subgroups. Notice that in the inner circle the sector labelled ‘Food and catering’ has been drawn almost twice as large (as measured by area) as that labelled ‘Alcohol and tobacco’. This reflects the fact that the typical household spends nearly twice as much on food and catering as on alcohol and tobacco. The weight of an item or group reflects how much money is spent on it. So the weight of the ‘Food and catering’ group is almost twice that of ‘Alcohol and tobacco’.

The outer ring represents the same total expenditure as the inner circle, but in more detail. For example, in the outer ring the area labelled ‘Food’ (which mostly consists of food bought for use in the home) is more than twice as large as that labelled ‘Catering’ (which includes meals in restaurants and canteens, and take-away meals and snacks), reflecting the fact that the typical household spends more than twice as much on food as on catering; the weight of the subgroup ‘Food’ is more than double the weight of the subgroup ‘Catering’. The chart gives a good indication of average spending patterns in the UK in the early 21st century.

Start of Activity

**Activity 19 The expenditure of a typical household**

Start of Question

(a)   Using Figure 37, estimate roughly what fraction of the expenditure of a typical household is on each of the following groups and subgroups:

* Personal expenditure
* Housing and household expenditure
* Housing

End of Question

[View discussion - Part](" \l "Session7_Discussion1)

Start of Question

(b)   Suppose that a household spends a total of £540 per week on goods and services that are covered by the RPI. Use your answers to part (a) to estimate very approximately how much is spent each week on each of the groups and subgroups in part (a).

End of Question

[View discussion - Part](" \l "Session7_Discussion2)

End of Activity

To ensure that the basket of goods for the index reflects the proportion of average spending devoted to different types of goods and services, it is necessary to find out how people actually spend their money. The Living Costs and Food Survey (LCF) records the spending reported by a sample of 5000 households spread throughout the UK. Data from the LCF are used to calculate the weights of most of the items included in the RPI basket. Since 1962, the weights have been revised each year, so that the index is always based on a basket of goods and services that is as up to date as possible. Because of this regular weight revision, the index is chained (as was the Gradgrind Ltd index).

(Most of the weights for the CPI come from a different source, the UK National Accounts, though in turn this source is partly based on data from the LCF. Again, the weights are revised each year.)

The weight of a group or subgroup directly depends on the average expenditure of households on that item. In [Subsection 2.1](#open-u2ss2-1), you saw that it is only the relative size of the weights that affects the value of the weighted mean – this is Rule 1 for weighted means. So instead of using the average expenditure of an item as its weight, the expenditure figures for the items can all be multiplied by the same factor to produce a new, more convenient, set of weights. For the RPI, this factor is chosen so that the sum of the weights is 1000. Table 12 shows the 2012 weights used in the RPI for the groups and subgroups. Notice that each group weight is obtained by summing the weights for its subgroups.

Start of Table

Table 12 2012 RPI weights

|  |  |  |  |
| --- | --- | --- | --- |
| **Group** | **Subgroup** | **Weight** | **Group weight** |
| Food and catering | Food | 114 |  |
|  | Catering | 47 | 161 |
| Alcohol and tobacco | Alcoholic drink | 56 |  |
|  | Tobacco | 29 | 85 |
| Housing and household expenditure | Housing | 237 |  |
|  | Fuel and light | 46 |  |
|  | Household goods | 62 |  |
|  | Household services | 67 | 412 |
| Personal expenditure | Clothing and footwear | 45 |  |
|  | Personal goods and services | 39 | 84 |
| Travel and leisure | Motoring expenditure | 131 |  |
|  | Fares and other travel costs | 23 |  |
|  | Leisure goods | 33 |  |
|  | Leisure services | 71 | 258 |
| **All items (i.e. the sum of the weights)** | |  | **1000** |

End of Table

The following checklist provided contains the major categories of goods and services included in the RPI. In the next activity, you will be asked to complete the last three columns of this checklist to make rough estimates of your household’s group weights.

Start of Figure

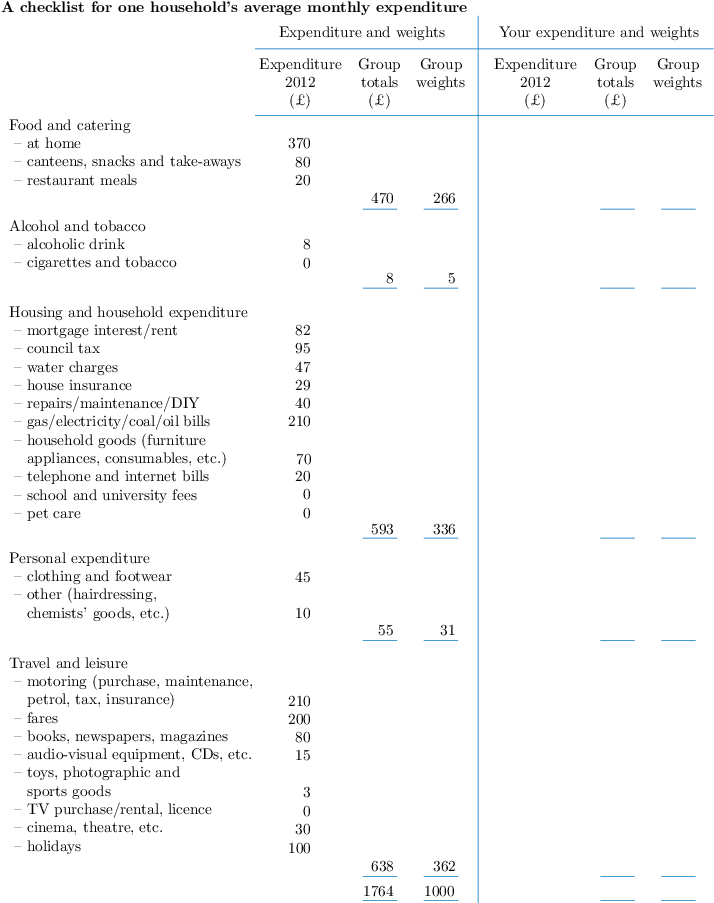


Figure 38 A checklist for one household’s average monthly expenditure

[View description - Figure 38 A checklist for one household’s average monthly expenditure](" \l "Session7_Description2)

End of Figure

The figures already in the checklist were completed for a two-person household. Some of the figures were accurate, others were necessarily very rough estimates. Nevertheless, the household’s weights give a reasonable indication of the proportion of the household’s expenditure (in 2012) on the five main groups used in the RPI.

The total expenditure was £1764. So the group weights were calculated by multiplying all the group total expenditures by a constant factor of 1000/1764, to ensure the weights sum to 1000. The weight for ‘Food and catering’, for example, is

Start of $1

470 times fraction 1000 over 1764 end simeq 266.

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative7)

End of $1

Another way to calculate this is to multiply the proportion of monthly expenditure spent on food and catering by 1000. The proportion is

Start of $1

fraction 470 over 1764 end simeq 0.266.

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative8)

End of $1

Since the total weight is 1000, the weight for ‘Food and catering’ is

Start of $1

0.266 times 1000=266.

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative9)

End of $1

Notice that the group weights for this particular household differ quite considerably from those used in the RPI in 2012 (see [Table 12](#open-u2tab5-1)). For instance, a much greater proportion of expenditure is on ‘Food and catering’ and a much smaller proportion is spent on ‘Alcohol and tobacco’.

Start of Activity

**Activity 20 Your own household’s expenditure**

Start of Question

Make rough estimates of your own household’s expenditure last year and complete the final columns of the checklist in Figure 38 ([Word version provided](https://www.open.edu/openlearn/ocw/mod/resource/view.php?id=30942)). For some categories, you may find it easier just to make a rough estimate of, say, your annual expenditure and then divide by 12. If you have no idea at all for a category, then use the corresponding figure in the checklist as a starting point for your own expenditure and adjust it up or down depending on how you think you spend your money. One way of checking that your figures are sensible is to consider how the sum of the expenditures relates to your household’s monthly income. Do not spend more than 15 minutes on estimating your expenditure; accurate figures are not needed.

Divide each group expenditure by your monthly expenditure total and then multiply by 1000 to calculate your household’s group weights.

How do your household’s weights compare with those used in the RPI in 2012?

End of Question

[View discussion - Activity 20 Your own household’s expenditure](" \l "Session7_Discussion3)

End of Activity

## 5.2 Calculating the price indices

This subsection concentrates on how the RPI is calculated. Generally the CPI is calculated in a similar way, though some of the details differ. To measure price changes in general, it is sufficient to select a limited number of representative items to indicate the price movements of a broad range of similar items. For each section of the RPI, a number of representative items are selected for pricing. The selection is made at the beginning of the year and remains the same throughout the year. It is designed in such a way that the price movements of the representative items, when combined using a weighted mean, provide a good estimate of price movements in the section as a whole.

For example, in 2012 the representative items in the ‘Bread’ section (which is contained in the ‘Food and catering’ group) were: large white sliced loaf, large white unsliced loaf, large wholemeal loaf, bread rolls, garlic bread. Changes in the prices of these types of bread are assumed to be representative of changes in bread prices as a whole. Note that although the price ratio for bread is based on this sample of five types of bread, the calculation of the appropriate weight for bread is based on all kinds of bread. This weight is calculated using data collected in the Living Costs and Food Survey.

Start of Box

**Collecting the data**

The bulk of the data on price changes required to calculate the RPI is collected by staff of a market research company and forwarded to the Office for National Statistics for processing. Collecting the prices is a major operation: well over 100 000 prices are collected each month for around 560 different items. The prices being charged at a large range of shops and other outlets throughout the UK are mostly recorded on a predetermined Tuesday near the middle of the month. Prices for the remaining items, about 140 of them, are obtained from central sources because, for example, the prices of some items do not vary from one place to another.

End of Box

One aim of the RPI is to make it possible to compare prices in any two months, and this involves calculating a value of the price index itself for every month.

Start of Box

**Changing the representative items**

The Office for National Statistics (ONS) updates the basket of goods every year, reflecting advancing technology, changing tastes and consumers’ spending habits. The media often have fun writing about the way the list of representative items changes each year.

Start of Quote

In the 1950s, the mangle, crisps and dance hall admissions were added to the basket, with soap flakes among the items taken out.

Two decades later, the cassette recorder and dried mashed potato made it in, with prunes being excluded.

Then after the turn of the century, mobile phone handsets and fruit smoothies were included. The old fashioned staples of an evening at home – gin and slippers – were removed from the basket.

So now, in 2012, it is the turn of tablet computers to be added to mark the growing popularity of this type of technology.

That received the most coverage when it was added to the basket of goods, with the ONS highlighting this digital-age addition in its media releases.

But those seafaring captains who once used the then unusual fruit as a symbol to show they were home and hosting might be astonished to find that centuries on, the pineapple has also been added to the inflation basket.

Technically, the pineapple has been added to give more varied coverage in the basket of fruit and vegetables, the prices of which can be volatile.

(Source: BBC News website, 14 March 2012)

End of Quote

End of Box

So, calculating the RPI involves two kinds of data:

* the price data, collected every month
* the weights, representing expenditure patterns, updated once a year.

Once the price data have been collected each month, various checks, such as looking for unbelievable prices, are applied and corrections made if necessary. Checking data for obvious errors is an important part of any data analysis.

Then an averaging process is used to obtain a price ratio for each item that fairly reflects how the price of the item has changed across the country. The exact details are quite complicated and are not described here. (If you want more details, they are given in the Consumer Price Indices Technical Manual, available from the ONS website. Consumer Price Indices: A brief guide is also available from the same website.)

For each item, a price ratio is calculated that compares its price with the previous January. For instance, for November 2011, the resulting price ratio for an item is an average value of

Start of $1

fraction price in November 2011 over price in January 2011 end .

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative10)

End of $1

The next steps in the process combine these price ratios, using weighted means, to obtain 14 subgroup price ratios, and then the group price ratios for the five groups. Finally, the group price ratios are combined to give the **all-item price ratio**. This is the price ratio, relative to the previous January, for the ‘basket’ of goods and services as a whole that make up the RPI.

The all-item price ratio tells us how, on average, the RPI ‘basket’ compares in price with the previous January. The value of the RPI for a given month is found by the method described in [Section 4](#open-u2s4), that is, by multiplying the value of the RPI for the previous January by the all-item price ratio for that month (relative to the previous January):

Start of $1

RPI for month x = open bracket RPI for previous January close bracket times open bracket all minus item price ratio for month x close bracket

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative11)

End of $1

Thus, to calculate the RPI for November 2011, the final step is to multiply the value of the RPI in January 2011 by the all-item price ratio for November 2011.

Start of Box

**Example 22 Calculating the RPI for November 2011**

Here are the details of the last two stages of calculation of the RPI for November 2011, after the group price ratios have been calculated, relative to January 2011. The appropriate data are in Table 13.

Start of Table

Table 13 Calculating the all-item price ratio for November 2011

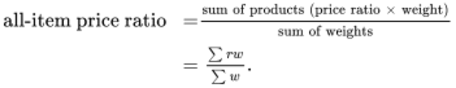
|  |  |  |  |
| --- | --- | --- | --- |
| **Group** | **Price ratio: r** | **Weight: w** | **Ratio times weight: r w** |
| Food and catering | 1.030 | 165 | 169.950 |
| Alcohol and tobacco | 1.050 | 88 | 92.400 |
| Housing and household expenditure | 1.037 | 408 | 423.096 |
| Personal expenditure | 1.128 | 82 | 92.496 |
| Travel and leisure | 1.026 | 257 | 263.682 |
| **Sum** |  | **1000** | **1041.624** |

End of Table

You may have noticed that the weights here do not exactly match those in [Table 12](#open-u2tab5-1). That is because the weights here are the 2011 weights, and those in Table 12 are the 2012 weights, and as has been explained, the weights are revised each year.

The all-item price ratio is a weighted average of the group price ratios given in the table. If the price ratios are denoted by the letter r, and the weights by w, then the weighted mean of the price ratios is the sum of the five values of rw divided by the sum of the five values of w. The formula, from [Subsection 2.3](#open-u2ss2-3), is

Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative16)

End of $1

The sums are given in Table 13. (The sum of the weights is 1000, because the RPI weights are chosen to add up to 1000.) Although Table 13 gives the individual r w values, there is no need for you to write down these individual products when finding a weighted mean (unless you are asked to do so). As mentioned previously, your calculator may enable you to calculate the weighted mean directly, or you may use its memory to store a running total of r w.

Now the all-item price ratio for November 2011 (relative to January 2011) can be calculated as

Start of $1

fraction 1041 .624 over 1000 end = 1.041624.

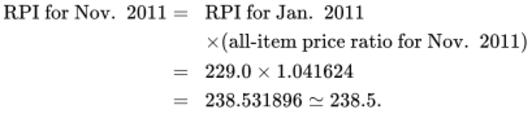
[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative19)

End of $1

This tells us that, on average, the RPI basket of goods cost 1.041 624 times as much in November 2011 as in January 2011.

The published value of the RPI for January 2011 was 229.0. So, using the formula,

Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative20)

End of $1

The final result has been rounded to one decimal place, because actual published RPI figures are rounded to one decimal place.

End of Box

Example 22 is the subject of the following screencast. [Note that references to ‘the unit’ should be interpreted as ‘this course’. The original wording refers to the Open University course from which this material is adapted.]

Start of Media Content

Video content is not available in this format.

Screencast 5 Calculating an RPI

[View transcript - Screencast 5 Calculating an RPI](" \l "Session7_Transcript1)

Start of Figure



End of Figure

End of Media Content

The same 2011 weights were used to calculate the RPI for every month from February 2011 to January 2012 inclusive. For each of these months, the price ratios were calculated relative to January 2011, and the RPI was finally calculated by multiplying the RPI for January 2011 by the all-item price ratio for the month in question. In February 2012, however, the process began again (as it does every February). A new set of weights, the 2012 weights, came into use. Price ratios were calculated relative to January 2012, and the RPI was found by multiplying the RPI value for January 2012 by the all-item price ratio. This procedure was used until January 2013, and so on.

The process of calculating the RPI can be summarised as follows.

Start of Box

**Calculating the RPI**

1. The data used are prices, collected monthly, and weights, based on the Living Costs and Food Survey, updated annually.
2. Each month, for each item, a price ratio is calculated, which gives the price of the item that month divided by its price the previous January.
3. Group price ratios are calculated from the price ratios using weighted means.
4. Weighted means are then used to calculate the all-item price ratio. Denoting the group price ratios by r and the group weights by w, the all-item price ratio is

Start of $1

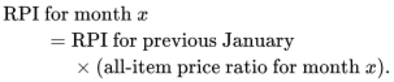
fraction sum r w over sum w end .

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative23)

End of $1

1. The value of the RPI for that month is found by multiplying the value of the RPI for the previous January by the all-item price ratio:

Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative24)

End of $1

The weights for a particular year are used in calculating the RPI for every month from February of that year to January of the following year.

End of Box

Start of Activity

**Activity 21 Calculating the RPI for July 2011**

Start of Question

Find the value of the RPI in July 2011 by completing the following table and the formulas below. The value of the RPI in January 2011 was 229.0. (The base date was January 1987.)

Start of Table

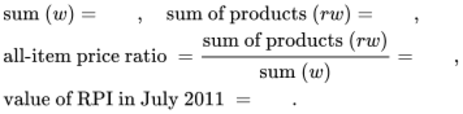
Table 14 Calculating the RPI for July 2011

|  |  |  |  |
| --- | --- | --- | --- |
| **Group** | **Price ratio for July 2011 relative to January 2011: r** | **2011 weights: w** | **Price ratio times weight: r w** |
| Food and catering | 1.024 | 165 |  |
| Alcohol and tobacco | 1.042 | 88 |  |
| Housing and household expenditure | 1.012 | 408 |  |
| Personal expenditure | 1.053 | 82 |  |
| Travel and leisure | 1.030 | 257 |  |
| **Sum** |  |  |  |

(Source: Office for National Statistics)

End of Table

Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative29)

End of $1

End of Question

[View discussion - Activity 21 Calculating the RPI for July 2011](" \l "Session7_Discussion4)

End of Activity

The published value for the RPI in July 2011 was 234.7, slightly different from the value you should have obtained in Activity 21 (that is, 234.6). The discrepancy arises because the government statisticians use more accuracy during their RPI calculations, and round only at the end before publishing the results.

The following activity is intended to help you draw together many of the ideas you have met in this section, both about what the RPI is and how it is calculated.

Start of Activity

**Activity 22 The effects of particular price changes on the RPI**

Start of Question

Between February 2011 and February 2012, the price of leisure goods fell on average by 2.3%, while the price of canteen meals rose by 2.8%. Answer the following questions about the likely effects of these changes on the value of the RPI. (No calculations are required.)

End of Question

Start of Question

(a)   Looked at in isolation (that is, supposing that no other prices changed), would the change in the price of leisure goods lead to an increase or a decrease in the value of the RPI?

Would the change in the price of canteen meals (looked at in isolation) lead to an increase or a decrease in the value of the RPI?

End of Question

[View discussion - Part](" \l "Session7_Discussion5)

Start of Question

(b)   In each case, is the size of the increase or decrease likely to be large or small?

End of Question

[View discussion - Part](" \l "Session7_Discussion6)

Start of Question

(c)   Using what you know about the structure of the RPI, decide which of ‘Leisure goods’ and ‘Canteen meals’ has the larger weight.

End of Question

[View discussion - Part](" \l "Session7_Discussion7)

Start of Question

(d)   Which of the price changes mentioned in the question will have a larger effect on the value of the RPI? Briefly explain your answer.

End of Question

[View discussion - Part](" \l "Session7_Discussion8)

End of Activity

## 5.3 Using the price indices

The RPI and CPI are intended to help measure price changes, so we shall start this section by describing how to use them for this purpose.

Start of Box

**Example 23 A news report on inflation**

The BBC News website reported (20 March 2012) ‘UK inflation rate falls to 3.4% in February’. What does that actually mean?

The rest of the BBC article makes it clear that this ‘inflation’ figure was based on the CPI rather than the RPI, but its meaning is still not obvious. What is usually meant in situations like this is the following.

**The annual rate of inflation**

In the UK, the (annual) rate of inflation is the percentage increase in the value of the CPI (or the RPI) compared to one year earlier.

(In this course, it will always be made clear whether you should use the CPI or the RPI in contexts like this.)

The annual rate of inflation is sometimes called the year-on-year rate of inflation.

In February 2012, the CPI was 121.8. Exactly a year earlier, in February 2011, the CPI was 117.8. The ratio of these two values is

Start of $1

fraction value of CPI in February 2012 over value of CPI in February 2011 end = fraction 121 .8 over 117 .8 end simeq 1.034.

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative37)

End of $1

So the value of the CPI in February 2012 was 3.4% higher than in the previous February. That is the source of the number in the BBC headline.

End of Box

Start of Activity

**Activity 23 The annual inflation rate in February 2012**

Start of Question

In February 2012, the RPI was 239.9. Exactly a year earlier, in February 2011, the RPI was 231.3. Calculate the annual inflation rate for February 2012, based on the RPI.

End of Question

[View discussion - Activity 23 The annual inflation rate in February 2012](" \l "Session7_Discussion9)

End of Activity

The fact that the inflation rates that are generally reported in the media relate to price increases (as measured in a price index) over a whole year means that one has to be careful in interpreting the figures, in several ways.

* Media reports might say that ‘inflation is falling’, but this does not mean that prices are falling. It simply means that the annual inflation rate is less than it was the previous month. So when the BBC headline said that the (annual) inflation rate had fallen to 3.4% in February 2012, it meant that the February 2012 rate was smaller than the January 2012 rate (which was 3.6%). Prices were still rising, but not quite so quickly.
* The change in price levels over one month may be, and indeed usually is, considerably different from the annual inflation rate. For instance, prices actually fell between December 2011 and January 2012: the CPI was 121.7 in December 2011 and 121.1 in January 2012. (Prices in the UK usually fall between December and January in the UK, as Christmas shopping ends and the January sales begin.) But the annual inflation rate for January 2012, measured by the CPI, was 3.6%.
* The effect of a single major cause of increased prices can persist in the annual inflation rates long after the prices originally increased. For instance, the standard rate of value added tax (VAT) in the UK went up from 17.5% to 20% at the start of January 2011, causing a one-off increase in the price (to consumers) of many goods and services. This showed up in the annual inflation rate for January 2011, where prices were 4.0% higher than a year earlier. Moreover, the annual inflation rate for every other month in 2011 was also affected by the VAT increase, because in each case the CPI was being compared to the CPI in the corresponding month in 2010, before the VAT increase.

Another important use of price indices like the RPI and CPI is for index-linking. This is used for such things as savings and pensions, as a means of safeguarding the value of money held or received in these forms.

Start of Box

**Index-linking an amount**

To index-link any amount of money, the amount in question is multiplied by the same ratio as the change in the value of the price index. Another term for this process is **indexation**.

End of Box

It is important to stress the notion of ratio in index-linking, because it is only by calculating the ratio of two indices that you can get an accurate measure of how prices have increased. For example, an increase in the RPI from 100 to 200 represents a 100% increase in price, whereas a further RPI increase from 200 to 300 represents only a further 50% increase in price.

Start of Box

**Example 24 Index-linking a pension**

The value of the RPI for February 2012 was 239.9 whereas the corresponding figure for February 2011 was 231.3. So an index-linked pension that was, say, £450 per month in February 2011, would be increased to

Start of $1

pounds 450 times fraction 239 .9 over 231 .3 end open bracket i.e. pounds 466.73 close bracket per month

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative39)

End of $1

for February 2012. The reason for index-linking the pension in this way is that the increased pension would buy the same amount of goods or services in February 2012 as the original pension bought in February 2011 – that is, it should have the same purchasing power.

End of Box

Pensions can be, and indeed increasingly are, index-linked using the CPI rather than the RPI.

Start of Activity

**Activity 24 Index-linking a pension using the CPI**

Start of Question

An index-linked pension was £120 per week in November 2010. It is index-linked using the CPI. How much should the pension be per week in November 2011? The value of the CPI was 115.6 in November 2010 and 121.2 in November 2011.

End of Question

[View discussion - Activity 24 Index-linking a pension using the CPI](" \l "Session7_Discussion10)

End of Activity

This principle leads to another much-quoted figure which can be calculated directly from the RPI: **the purchasing power of the pound**. (This is the purchasing power of the pound within this country, not its purchasing power abroad; the latter is a distinct and far more complicated concept.) The purchasing power of the pound measures how much a consumer can buy with a fixed amount of money at one point of time compared with another point of time.

The word compared here is again important; it makes sense only to talk about the purchasing power of the pound at one time compared with another. For example, if £1 worth of goods would have cost only 60p four years ago, then we say that the purchasing power of the pound is only 60p compared with four years earlier.

Start of Box

**Purchasing power of the pound**

The purchasing power (in pence) of the pound at date uppercase A compared with date uppercase B is

Start of $1

fraction value of RPI at date B over value of RPI at date A end times 100.

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative43)

End of $1

End of Box

The purchasing power of the pound could be calculated using the CPI instead, though the figures published by the Office for National Statistics do happen to use the RPI.

Start of Box

**Example 25 Calculating the purchasing power of the pound**

(a)

The purchasing power of the pound in February 2012 compared with February 2011 was

Start of $1

fraction 231 .3 over 239 .9 end times 100 p = 96.41517 p .

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative44)

End of $1

(231.3 and 239.9 are the two RPI values given in [Activity 23](#open-u2act5-3-1).)

We round this to give 96p.

(b)

The purchasing power of the pound in February 2012 compared with the base date, January 1987, was

Start of $1

fraction 100 over 239 .9 end times 100 p .

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative45)

End of $1

(At the base date, the value of the RPI is 100 by definition.)

This is, after rounding, 42p.

End of Box

Start of Activity

**Activity 25 Annual inflation and the purchasing power of the pound**

Start of Question

Start of Table

Table 15 Values of the RPI from January 2009 to December 2011

|  |  |  |  |
| --- | --- | --- | --- |
| **Month** | **2009** | **2010** | **2011** |
| January | 210.1 | 217.9 | 229.0 |
| February | 211.4 | 219.2 | 231.3 |
| March | 211.3 | 220.7 | 232.5 |
| April | 211.5 | 222.8 | 234.4 |
| May | 212.8 | 223.6 | 235.2 |
| June | 213.4 | 224.1 | 235.2 |
| July | 213.4 | 223.6 | 234.7 |
| August | 214.4 | 224.5 | 236.1 |
| September | 215.3 | 225.3 | 237.9 |
| October | 216.0 | 225.8 | 238.0 |
| November | 216.6 | 226.8 | 238.5 |
| December | 218.0 | 228.4 | 239.4 |

(Source: Office for National Statistics)

End of Table

For each of the following months, use the values of the RPI in Table 15 to calculate the annual inflation rate (based on the RPI) and to calculate the purchasing power of the pound (in pence) compared to one year previously.

End of Question

Start of Question

(a)   May 2010

End of Question

[View discussion - Part](" \l "Session7_Discussion11)

Start of Question

(b)   October 2011

End of Question

[View discussion - Part](" \l "Session7_Discussion12)

Start of Question

(c)   March 2011

End of Question

[View discussion - Part](" \l "Session7_Discussion13)

End of Activity

You have seen that the RPI can be used as a way of updating the value of a pension to take account of general increases in prices (index-linking). The RPI is used in other similar ways, for instance to update the levels of some other state benefits and investments. But the CPI could be used for these purposes.

Why are there two different indices? Let’s look at how this arose. As well as its use for index-linking, which is basically to compensate for price changes, the RPI previously played an important role in the management of the UK economy generally. The government sets targets for the rate of inflation, and the Bank of England Monetary Policy Committee adjusts interest rates to try to achieve these targets. Until the end of 2003, these inflation targets were based on the RPI, or to be precise, on another price index called RPIX which is similar to the RPI but omits owner-occupiers’ mortgage interest payments from the calculations. (There are good economic reasons for this omission, to do with the fact that in many ways the purchase of a house has the character of a long-term investment, unlike the purchase of, say, a bag of potatoes.) From 2004, the inflation targets have instead been set in terms of the CPI. The CPI is calculated in a way that matches similar inflation measures in other countries of the European Union. (So it can be used for international comparisons.)

In terms of general principles, though, and also in terms of most of the details of how the indices are calculated, the differences between the RPI and CPI are not actually very great. As mentioned in Subsection 5.1, the CPI reflects the spending of a wider population than the RPI. Partly because of this, there are certain items (e.g. university accommodation fees) that are included in the CPI but not the RPI. There are also certain items that are included in the RPI but not the CPI, notably some owner-occupiers’ housing costs such as mortgage interest payments and house-building insurance. Finally, the CPI uses a different method to the RPI for combining individual price measurements.

Because of these differences, inflation as measured by the CPI tends usually to be rather lower than that measured by the RPI. In Example 23, you saw that the annual inflation rate in February 2012 as measured by the CPI was 3.4%. The annual inflation rate in the same month, as measured by the RPI, was 3.7%, as you saw in [Activity 23](#open-u2act5-3-1). The RPI continues to be calculated and published, and to be used to index-link payments such as savings rates and some pensions. (Arguably it is rather strange to use the RPI to index pensions, given that (as was said at the beginning of Subsection 5.1) the RPI omits the expenditure of pensioner households.) However, there are reasons why the RPI is more appropriate than the CPI for some such purposes, and it seems likely to continue in use for a long time. Furthermore, changes in how index-linking is done can be politically very controversial. For instance, in 2010, the UK government announced that in future, public sector pensions would be index-linked to the CPI rather than the RPI, which caused major complaints from those affected (because inflation as measured by the CPI is usually lower than that measured using the RPI, so pensions will not increase so much in money terms).

You might be asking yourself which is the ‘correct’ measure of inflation – RPI, CPI, or something else entirely. There is no such thing as a single ‘correct’ measure. Different measures are appropriate for different purposes. That’s why it is important to understand just what is being measured and how.

## Exercises on Section 5

The following exercises provide extra practice on the topics covered in Section 5.

Start of Exercise

**Exercise 10 Calculating the RPI for February 2012**

Start of Question

Find the value of the RPI in February 2012, using the data in the table below. The value of the RPI in January 2012 was 238.0.

Start of Table

Table 16 Calculating the RPI for February 2012

|  |  |  |  |
| --- | --- | --- | --- |
| **Group** | **Price ratio for February 2012 relative to January 2012: r** | **2012 weights: w** | **Price ratio times weight: r w** |
| Food and catering | 1.009 | 161 |  |
| Alcohol and tobacco | 1.005 | 85 |  |
| Housing and household expenditure | 1.003 | 412 |  |
| Personal expenditure | 1.040 | 84 |  |
| Travel and leisure | 1.005 | 258 |  |
| **Total** |  |  |  |

(Source: Office for National Statistics)

End of Table

End of Question

[View discussion - Exercise 10 Calculating the RPI for February 2012](" \l "Session7_Discussion14)

End of Exercise

Start of Exercise

**Exercise 11 Annual inflation rates and the purchasing power of the pound**

Start of Question

For each of the following months, use [Table 15](#open-u2table5-9) (in Subsection 5.3) to calculate the annual inflation rate given by the RPI and to calculate the purchasing power of the pound (in pence) compared to one year previously.

End of Question

Start of Question

(a)   October 2010

End of Question

[View discussion - Part](" \l "Session7_Discussion15)

Start of Question

(b)   January 2011

End of Question

[View discussion - Part](" \l "Session7_Discussion16)

End of Exercise

Start of Exercise

**Exercise 12 Index-linking another pension**

Start of Question

An index-linked pension (linked to the RPI) was £800 per month in April 2010. How much should it be in April 2011? (Again, use the RPI values in [Table 15](#open-u2table5-9).)

End of Question

[View discussion - Exercise 12 Index-linking another pension](" \l "Session7_Discussion17)

End of Exercise

## Conclusion

In this free course, Prices, location and spread, you have been discovering how statistics can be used to answer questions about prices. You have learned:

* how to find a single number to summarise the price of an item at a particular point in time, even though the item might be available from a number of sources
* how to combine information on prices across a range of goods and services
* how, through the use of price ratios, changes in price over time can be quantified
* how chained price indices such as the RPI and CPI measure changes in prices over time.

In particular, you have learned how the RPI and CPI are calculated by the Office for National Statistics from a ‘basket’ of goods using weighted means to give price ratios, group price ratios and all-commodities price ratios. These all-commodity price ratios are then chained to give the value of the index relative to a base date. The RPI and CPI can be used to calculate inflation, to index-link amounts of money and to calculate the purchasing power of the pound at one time compared with another.

This course has focused on the ‘prices’ element of the question, Are people getting better or worse off?. If prices are rising, then, other things being equal, we are worse off.

Another crucial element is ‘earnings’. If our earnings are increasing, then, other things being equal, we are better off.

However, other things are usually not equal – prices and earnings are generally changing at the same time. The question of how to deal with both sorts of changes at once is beyond the scope of this particular course (although it is dealt with in the Open University course from which this free course is drawn).

Test your understanding of this OpenLearn course by working through the [end-of-course quiz](https://www.open.edu/openlearn/ocw/mod/quiz/view.php?id=28155).

This OpenLearn course is an adapted extract from the Open University course [M140 Introducing statistics](http://www3.open.ac.uk/study/undergraduate/course/m140.htm). To see if you are ready to study M140 and/or to refresh you knowledge of related topics, see the [Maths Help website](http://mathshelp.open.ac.uk/). All of the modules here, except for the Geometry one, are relevant to M140.

## Text

This free course was written by Kevin McConway.

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

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

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

Subsection 3.3 quote from McCullagh, P. (2003): The Royal Statistical Society

Subsection 5.2 quote from BBC News website, 14 March 2012: Taken from www.bbc.co.uk/news/business-17356286

Course Image:  © Leanne J in Flickr https://creativecommons.org/licenses/by-nc-nd/2.0/

Figure 37 Crown copyright material is reproduced under Class Licence Number C01W0000065 with the permission of the Controller, Office of Public Sector Information (OPSI)

Table 3 Adapted from: https://www.gov.uk/government/statistical-data-sets/annual-domestic-energy-price-statistics

Table 5 Taken from: http://en.wikipedia.org/wiki/List\_of\_conurbations\_in\_the\_United\_Kingdom. This file is licensed under the Creative Commons Attribution Licence http://creativecommons.org/licenses/by/3.0/

Table 6 Department of Energy and Climate Change

Tables 13–15 Office for National Statistics licensed under the Open Government Licence v.1.0

Table 16 Adapted from data from the Office for National Statistics licensed under the Open Government Licence v.1.0

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

**Don’t miss out**

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

## Solutions

## Activity 1 Small flat-screen televisions

#### Discussion

For a batch size of 20, the median position is fraction 1 over 2 end open bracket 20 + 1 close bracket = 10 fraction 1 over 2 end. So, the median will be halfway between x subscript open bracket 10 close bracket end and x subscript open bracket 11 close bracket end. These are both 150, so the median is £150.

[Back to - Activity 1 Small flat-screen televisions](" \l "Session3_Activity1)

## Activity 2 The price of gas in UK cities

### Part

#### Discussion

A stemplot of all 14 prices in the table is shown below.

Start of Figure

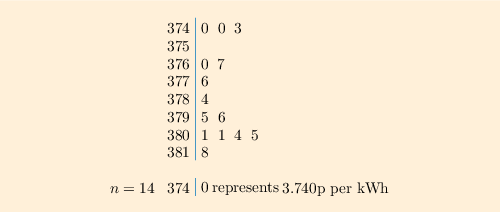


Figure 6 Stemplot of 14 gas prices

[View description - Figure 6 Stemplot of 14 gas prices](" \l "Session3_Description6)

End of Figure

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

### Part

#### Discussion

Stemplots for the prices for northern and southern cities are shown below.

Start of Figure

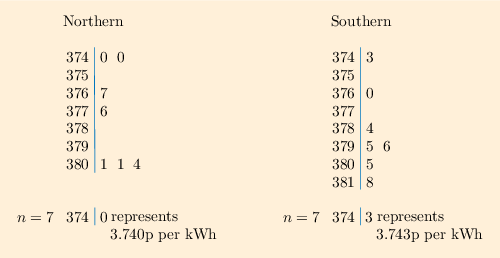


Figure 7 Stemplots for northern and southern cities separately.

[View description - Figure 7 Stemplots for northern and southern cities separately.](" \l "Session3_Description7)

End of Figure

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

### Part

#### Discussion

For a batch size of 14, the median position is fraction 1 over 2 end open bracket 14 + 1 close bracket = 7 fraction 1 over 2 end. So, the all-cities median will be halfway between x subscript open bracket 7 close bracket end and x subscript open bracket 8 close bracket end. These are 3.784 and 3.795, so the median is 3.7895, which is 3.790 when rounded to three decimal places. (The rounded median should be written as 3.790 and not 3.79, to show it is accurate to three decimal places and not just two.)

For the northern and southern batches, both of size 7, the median for each is the value of x subscript open bracket 4 close bracket end (that is, fraction 1 over 2 end open bracket 7 + 1 close bracket = 4). This is 3.776 for the northern batch and 3.795 for the southern batch.

The range is the difference between the upper extreme, uppercase E subscript uppercase U end, and the lower extreme, uppercase E subscript uppercase L end (range = uppercase E subscript uppercase U end minus uppercase E subscript uppercase L end). So the all-cities range is

Start of $1

3.818 minus 3.740=0.078 comma

[View alternative description - Uncaptioned Equation](" \l "Session3_Alternative37)

End of $1

the range for the northern batch is

Start of $1

3.804 minus 3.740=0.064 comma

[View alternative description - Uncaptioned Equation](" \l "Session3_Alternative38)

End of $1

and the range for the southern batch is

Start of $1

3.818 minus 3.743=0.075.

[View alternative description - Uncaptioned Equation](" \l "Session3_Alternative39)

End of $1

The medians and ranges are summarised below.

Start of Table

|  |  |  |
| --- | --- | --- |
| **Batch** | **Median** | **Range** |
| All cities | 3.790 | 0.078 |
| Northern cities | 3.776 | 0.064 |
| Southern cities | 3.795 | 0.075 |

End of Table

Thus the general level of gas prices in the country as a whole was about 3.790p per kWh. The average price differed by only 0.078p per kWh across the 14 cities.

The difference between the median prices for the northern and southern cities is 0.019p per kWh open bracket 3.795 minus 3.776=0.019 close bracket, with the south having the higher median.

The analysis does not clearly reveal whether the general level of gas prices for typical consumers in 2010 was higher in the south or in the north, though there is an indication that prices were a little higher in the south. The range of prices was also rather greater in the south. It is worth noting that the differences in gas prices between the cities in Table 3 were generally small, when measured in pence per kWh – although, with a typical annual gas usage of 18 000 kWh, the price difference between the most expensive city and the cheapest would amount to an annual difference in bills of about £14 on a typical bill of somewhere around £700.

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

## Activity 3 Small televisions: the mean

#### Discussion

Using the data for the prices from Activity 1:

Start of $1

mean = fraction sum over size end = fraction 90 + 100 + ellipsis +270 over 20 end = pounds 162.

[View alternative description - Uncaptioned Equation](" \l "Session3_Alternative53)

End of $1

Or using the sum notation, sum x = 90 + 100 + ellipsis +270 = 3240 and n = 20, so

Start of $1

mean = overline x = fraction sum x over n end = fraction 3240 over 20 end = pounds 162.

[View alternative description - Uncaptioned Equation](" \l "Session3_Alternative57)

End of $1

The prices were rounded to the nearest £10, so it is appropriate to keep one more significant figure for the mean, that is, to show it accurate to the nearest £1. So since the exact value is £162, it needs no further rounding.

[Back to - Activity 3 Small televisions: the mean](" \l "Session3_Activity3)

## Activity 4 Changing the gas prices

#### Discussion

The completed table is:

Start of Table

|  |  |  |
| --- | --- | --- |
| **Batch** | **Mean** | **Median** |
| Seven southern cities | 3.7859 | 3.795 |
| Five southern cities (excluding Cardiff and Ipswich) | 3.7996 | 3.796 |

End of Table

Whereas deletion of Cardiff and Ipswich has the effect of increasing the mean price by 0.0137p per kWh, the median price increases by only 0.001p per kWh. This is what we would expect as, in general, the more resistant a measure is, the less it changes when a few extreme values are deleted.

[Back to - Activity 4 Changing the gas prices](" \l "Session3_Activity4)

## Activity 5 A misprint in the gas prices

#### Discussion

The completed table is:

Start of Table

|  |  |  |
| --- | --- | --- |
| **Batch** | **Mean** | **Median** |
| Five cities (correct data) | 3.7996 | 3.796 |
| Five cities (with misprint) | 4.6996 | 3.796 |

End of Table

Here the median is completely unaffected by the misprint, although the mean changes considerably.

[Back to - Activity 5 A misprint in the gas prices](" \l "Session3_Activity5)

## Exercise 1 Finding medians

### Part

#### Discussion

For the arithmetic scores, the position of the median is fraction 1 over 2 end open bracket 33+1 close bracket = 17, so the median is 79%.

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

### Part

#### Discussion

For the television prices, the position of the median is fraction 1 over 2 end open bracket 26+1 close bracket = 13 fraction 1 over 2 end, so the median is halfway between x subscript open bracket 13 close bracket end and x subscript open bracket 14 close bracket end. Thus, the median is

Start of $1

fraction 1 over 2 end open bracket pounds 269 + pounds 270 close bracket = pounds 269.5 simeq pounds 270.

[View alternative description - Uncaptioned Equation](" \l "Session3_Alternative66)

End of $1

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

## Exercise 2 Finding means

#### Discussion

For the batch of arithmetic scores in part (a) of Exercise 1, the sum of the 33 values is 2326 and

Start of $1

fraction 2326 over 33 end simeq 70.5.

[View alternative description - Uncaptioned Equation](" \l "Session3_Alternative67)

End of $1

Therefore, the mean is 70.5%. (The original data are given to the nearest whole number, so the mean is rounded to one decimal place.)

For the batch of television prices in part (b) of Exercise 1, the sum of the 26 values is 7856 and

Start of $1

fraction 7856 over 26 end = 302.1538 simeq 302.2.

[View alternative description - Uncaptioned Equation](" \l "Session3_Alternative68)

End of $1

Therefore, the mean is £302.2.

[Back to - Exercise 2 Finding means](" \l "Session3_Exercise2)

## Exercise 3 The effect of removing values on the median and mean

#### Discussion

For the median, there are now 17 prices left in the batch, so the median is at position fraction 1 over 2 end open bracket 17+1 close bracket = 9. It is therefore 150.

The sum of the remaining 17 values is 2480, so the mean is

Start of $1

fraction 2480 over 17 end =145.8824 simeq 146.

[View alternative description - Uncaptioned Equation](" \l "Session3_Alternative70)

End of $1

In this case, removing the three highest prices has not changed the median at all, but it has reduced the mean considerably. This illustrates that the median is a more resistant measure than the mean.

[Back to - Exercise 3 The effect of removing values on the median and mean](" \l "Session3_Exercise3)

## Activity 6 Using the rules for weighted means

#### Discussion

You should expect the weighted mean price to be nearer the London price, because of Rule 2 for weighted means ([Subsection 2.1](#open-u2ss2-1)) and given that London has a much larger weight then Edinburgh.

The weighted mean price given by the formula in Example 11 is (after rounding) 3.814p per kWh, which is indeed much closer to the London price than to the Edinburgh price.

[Back to - Activity 6 Using the rules for weighted means](" \l "Session4_Activity1)

## Activity 7 Weighted means of Open University marks

### Part

#### Discussion

OCAS = fraction open bracket 80 times 50 close bracket + open bracket 60 times 50 close bracket over 50 +50 end = fraction 4000 +3000 over 100 end = fraction 7000 over 100 end = 70.

This is the same as a simple (unweighted) mean of the two scores, because the two component scores have equal weight. It lies exactly halfway between the two scores (fraction 1 over 2 end open bracket 80+60 close bracket =70).

[Back to - Part](" \l "Session4_Part2)

### Part

#### Discussion

OCAS = fraction open bracket 80 times 40 close bracket + open bracket 60 times 60 close bracket over 40 +60 end = fraction 3200 +3600 over 100 end = fraction 6800 over 100 end = 68.

This is slightly less than the simple mean in (a) because the component with the lower score (TMA) has the greater weight.

[Back to - Part](" \l "Session4_Part3)

### Part

#### Discussion

OCAS = fraction open bracket 80 times 65 close bracket + open bracket 60 times 55 close bracket over 65 +55 end = fraction 5200 +3300 over 120 end = fraction 8500 over 120 end simeq 70.8.

This is slightly higher than the simple mean in (a) because the component with the higher score (iCMA) has the greater weight.

(Note that the weights need not necessarily sum to 100, even when dealing with percentages.)

[Back to - Part](" \l "Session4_Part4)

### Part

#### Discussion

OCAS = fraction open bracket 80 times 25 close bracket + open bracket 60 times 75 close bracket over 25 +75 end = fraction 2000 +4500 over 100 end = fraction 6500 over 100 end = 65.

This is even lower than (b), so even nearer the lower score (TMA), because the TMA score has even greater weight.

[Back to - Part](" \l "Session4_Part5)

### Part

#### Discussion

OCAS = fraction open bracket 80 times 30 close bracket + open bracket 60 times 90 close bracket over 30 +90 end = fraction 2400 +5400 over 120 end = fraction 7800 over 120 end = 65.

This is the same as (d) because the ratios of the weights are the same; they are both in the ratio 1 to 3. That is, 25:75=30:90 (=1:3).

(We say this as follows: ‘the ratio 25 to 75 equals the ratio 30 to 90’.)

[Back to - Part](" \l "Session4_Part6)

## Activity 9 Weighted mean electricity price

#### Discussion

The table showing the required sums (and the values in the x w column, that you may not have had to write down), is as follows.

Start of Table

|  |  |  |  |
| --- | --- | --- | --- |
| **City** | **Price (p/kWh): x** | **Weight: w** | **Price times weight: x w** |
| Aberdeen | 13.76 | 19 | 261.44 |
| Belfast | 15.03 | 58 | 871.74 |
| Edinburgh | 13.86 | 42 | 582.12 |
| Leeds | 12.70 | 150 | 1 905.00 |
| Liverpool | 13.89 | 82 | 1 138.98 |
| Manchester | 12.65 | 224 | 2 833.60 |
| Newcastle-upon-Tyne | 12.97 | 88 | 1 141.36 |
| Nottingham | 12.64 | 67 | 846.88 |
| Birmingham | 12.89 | 228 | 2 938.92 |
| Canterbury | 12.92 | 5 | 64.60 |
| Cardiff | 13.83 | 33 | 456.39 |
| Ipswich | 12.84 | 14 | 179.76 |
| London | 13.17 | 828 | 10 904.76 |
| Plymouth | 13.61 | 24 | 326.64 |
| Southampton | 13.41 | 30 | 402.30 |
| **Sum** |  | **1892** | **24 854.49** |

End of Table

Thus sum x w = 24854.49, sum w = 1892 and

Start of $1

fraction sum x w over sum w end = fraction 24854 .49 over 1892 end = 13.136623 simeq 13.14.

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative111)

End of $1

So the weighted mean of electricity prices is 13.14p per kWh.

[Back to - Activity 9 Weighted mean electricity price](" \l "Session4_Activity4)

## Exercise 4 A combined batch of camera prices

#### Discussion

Mean price of all the cameras is

Start of $1

fraction open bracket 80 .7 times 10 close bracket + open bracket 78 .5 times 17 close bracket over 10 + 17 end = fraction 2141 .5 over 27 end comma

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative115)

End of $1

which is £79.3 (rounded to the same accuracy as the original means).

[Back to - Exercise 4 A combined batch of camera prices](" \l "Session4_Exercise1)

## Exercise 5 The mean price of fabric

#### Discussion

Mean price of all the material is

Start of $1

fraction open bracket 10 .95 times 8 .5 close bracket + open bracket 12 .70 times 6 close bracket over 8 .5 + 6 end = fraction 169 .275 over 14 .5 end comma

[View alternative description - Uncaptioned Equation](" \l "Session4_Alternative116)

End of $1

which is £11.67 (rounded to the nearest penny).

[Back to - Exercise 5 The mean price of fabric](" \l "Session4_Exercise2)

## Activity 10 Finding more quartiles

### Part

#### Discussion

Here, because n=15, an appropriate picture of the data would be [Figure 12](#open-u2fig3-1) (Subsection 3.2). To find the lower and upper quartiles, uppercase Q sub 1 and uppercase Q sub 3, of this batch, first find fraction 1 over 4 end open bracket n+1 close bracket = 4 and fraction 3 over 4 end open bracket n+1 close bracket = 12. Therefore uppercase Q sub 1 =268 p and uppercase Q sub 3 =299 p.

[Back to - Part](" \l "Session5_Part1)

### Part

#### Discussion

For this batch, n=14 so fraction 1 over 4 end open bracket n+1 close bracket = 3 fraction 3 over 4 end and fraction 3 over 4 end open bracket n+1 close bracket = 11 fraction 1 over 4 end.

Start of $1

uppercase Q sub 1 = 3.743 + fraction 3 over 4 end open bracket 3.760 minus 3.743 close bracket = 3.75575 simeq 3.756

[View alternative description - Uncaptioned Equation](" \l "Session5_Alternative109)

End of $1

and

Start of $1

uppercase Q sub 3 = 3.801 + fraction 1 over 4 end open bracket 3.804 minus 3.801 close bracket = 3.80175 simeq 3.802.

[View alternative description - Uncaptioned Equation](" \l "Session5_Alternative110)

End of $1

So the lower quartile is 3.756 p per kWh and the upper quartile is 3.802p per kWh.

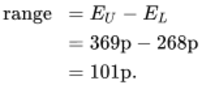
[Back to - Part](" \l "Session5_Part2)

## Activity 11 Coffee prices again

#### Discussion

The range is the distance between the extremes:

Start of $1

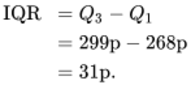


[View alternative description - Uncaptioned Equation](" \l "Session5_Alternative115)

End of $1

The interquartile range is the distance between the quartiles:

Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session5_Alternative116)

End of $1

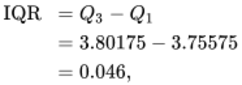
[Back to - Activity 11 Coffee prices again](" \l "Session5_Activity2)

## Activity 12 Interquartile range of gas prices

#### Discussion

The quartiles, before rounding, are uppercase Q sub 1 =3.75575 and uppercase Q sub 3 =3.80175. So

Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session5_Alternative119)

End of $1

and the interquartile range is 0.046p per kWh.

[Back to - Activity 12 Interquartile range of gas prices](" \l "Session5_Activity3)

## Activity 13 Skew gas prices?

### Part

#### Discussion

All the necessary figures have already been calculated. You found the median (3.790) in Activity 2 and the quartiles (uppercase Q sub 1 = 3.756, uppercase Q sub 3 = 3.802) in [Activity 10](#open-u2act3-1). The extremes (uppercase E subscript uppercase L end =3.740, uppercase E subscript uppercase U end =3.818) and the batch size (n=14) are clearly shown in the stemplot.

So the five-figure summary is as follows:

Start of Figure

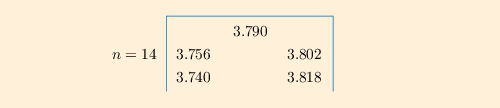


Figure 26

[View description - Figure 26](" \l "Session5_Description15)

End of Figure

[Back to - Part](" \l "Session5_Part4)

### Part

#### Discussion

Looking at the stemplot, on the whole the lower values are more spread out, indicating that the data are not symmetric and are left-skew.

The central box of the boxplot again shows left skewness, with the left-hand part of the box being clearly longer than the right-hand part. However, this skewness does not show up in the lengths of the whiskers in this batch – they are both the same length.

[Back to - Part](" \l "Session5_Part5)

## Exercise 6 Finding quartiles and the interquartile range

### Part

#### Discussion

For the arithmetic scores, n=33 so fraction 1 over 4 end open bracket n+1 close bracket = 8 fraction 1 over 2 end and fraction 3 over 4 end open bracket n+1 close bracket = 25 fraction 1 over 2 end.

The lower quartile is therefore

Start of $1

uppercase Q sub 1 = fraction 1 over 2 end open bracket 55+58 close bracket % = 56.5 % simeq 57 % .

[View alternative description - Uncaptioned Equation](" \l "Session5_Alternative138)

End of $1

The upper quartile is

Start of $1

uppercase Q sub 3 = fraction 1 over 2 end open bracket 86+89 close bracket % = 87.5 % simeq 88 % .

[View alternative description - Uncaptioned Equation](" \l "Session5_Alternative139)

End of $1

The interquartile range is

Start of $1

uppercase Q sub 3 minus uppercase Q sub 1 = 87.5 % minus 56.5 % = 31 % .

[View alternative description - Uncaptioned Equation](" \l "Session5_Alternative140)

End of $1

[Back to - Part](" \l "Session5_Part6)

### Part

#### Discussion

For the television prices, n=26 so fraction 1 over 4 end open bracket n+1 close bracket = 6 fraction 3 over 4 end and fraction 3 over 4 end open bracket n+1 close bracket = 20 fraction 1 over 4 end.

The lower quartile is therefore

Start of $1

uppercase Q sub 1 = pounds 229 + fraction 3 over 4 end open bracket pounds 230 minus pounds 229 close bracket = pounds 229.75 simeq pounds 230.

[View alternative description - Uncaptioned Equation](" \l "Session5_Alternative144)

End of $1

The upper quartile is

Start of $1

uppercase Q sub 3 = pounds 320 + fraction 1 over 4 end open bracket pounds 349 minus pounds 320 close bracket = pounds 327.25 simeq pounds 327.

[View alternative description - Uncaptioned Equation](" \l "Session5_Alternative145)

End of $1

The interquartile range is

Start of $1

uppercase Q sub 3 minus uppercase Q sub 1 = pounds 327.25 minus pounds 229.75 = pounds 97.5 simeq pounds 98.

[View alternative description - Uncaptioned Equation](" \l "Session5_Alternative146)

End of $1

[Back to - Part](" \l "Session5_Part7)

## Exercise 7 Some five-figure summaries

### Part

#### Discussion

Arithmetic scores:

From the stemplot, n=33, uppercase E subscript uppercase L end = 7 and uppercase E subscript uppercase U end = 100.

Start of Figure

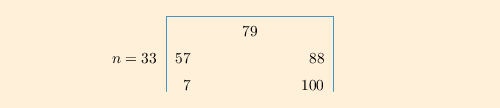


Figure 31 Five-figure summary of arithmetic scores

[View description - Figure 31 Five-figure summary of arithmetic scores](" \l "Session5_Description20)

End of Figure

[Back to - Part](" \l "Session5_Part9)

### Part

#### Discussion

Television prices:

From the data table, n=26, uppercase E subscript uppercase L end = 170 and uppercase E subscript uppercase U end = 699.

Start of Figure

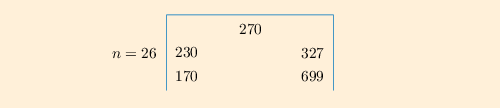


Figure 32 Five-figure summary of television prices

[View description - Figure 32 Five-figure summary of television prices](" \l "Session5_Description21)

End of Figure

[Back to - Part](" \l "Session5_Part10)

## Exercise 8 Boxplots and the shape of distributions

#### Discussion

For the boxplot of arithmetic scores, the left part of the box is longer than the right part, and the left whisker is also considerably longer than the right. This batch is left-skew.

For the boxplot of television prices, the right part of the box is rather longer than the left part. The right whisker is also rather longer than the left, and if one also takes into account the fact that two potential outliers have been marked, the top 25% of the data are clearly much more spread out than the bottom 25%. This batch is right-skew.

[Back to - Exercise 8 Boxplots and the shape of distributions](" \l "Session5_Exercise3)

## Activity 14 Gradgrind’s gas price increase

#### Discussion

The increase (in £/MWh) is 29 minus 24=5. This is fraction 5 over 24 end simeq 0.208 as a proportion of the 2007 price. That is, fraction 5 over 24 end times 100 % simeq 20.8 % of the 2007 price. Or you might have worked this out by finding that the 2008 price is fraction 29 over 24 end times 100 % simeq 120.8 % of the 2007 price, so that again the increase is 20.8% of the 2007 price.

[Back to - Activity 14 Gradgrind’s gas price increase](" \l "Session6_Activity1)

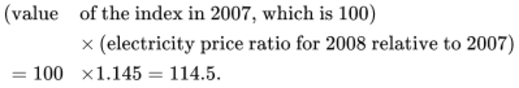
## Activity 15 Gradgrind’s electricity price index

#### Discussion

The 2008 electricity price is 1.145 times 100 % = 114.5 % of the 2007 price, so that the increase is 14.5% of the 2007 price.

The 2008 value of the electricity price index is

Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative8)

End of $1

[Back to - Activity 15 Gradgrind’s electricity price index](" \l "Session6_Activity2)

## Activity 16 How much fuel did Gradgrind use?

#### Discussion

The expenditure on a particular fuel in a particular year can be calculated as expenditure = quantity used times price. Therefore, if the expenditure and price are known, the quantity used can be calculated as

Start of $1

quantity used = fraction expenditure over price end .

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative10)

End of $1

In 2007, Gradgrind’s gas cost £24 per MWh, and they spent £9298 on gas, so the amount of gas they used in MWh was

Start of $1

fraction 9298 over 24 end simeq 387.4.

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative11)

End of $1

The other amounts, in MWh, are found in a similar way, and all are shown in the following table.

Start of Table

|  |  |  |
| --- | --- | --- |
| **Energy type** | **2007** | **2008** |
| Gas | 387.4 | 280.9 |
| Electricity | 42.2 | 34.4 |

End of Table

The reason that the expenditures went down is simply that Gradgrind used less of each fuel in 2008 than in 2007.

[Back to - Activity 16 How much fuel did Gradgrind use?](" \l "Session6_Activity3)

## Activity 17 Gradgrind’s energy price ratio for 2009 relative to 2008

### Part

#### Discussion

The gas price ratio for 2009 relative to 2008 is

Start of $1

fraction 30 over 29 end simeq 1.034.

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative21)

End of $1

The electricity price ratio for 2009 relative to 2008 is

Start of $1

fraction 98 over 87 end simeq 1.126.

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative22)

End of $1

(Over this year, electricity prices rose a lot more than gas prices.)

[Back to - Part](" \l "Session6_Part2)

### Part

#### Discussion

The overall energy price ratio for 2009 relative to 2008 is

Start of $1

fraction open bracket 1 .034 times 8145 close bracket + open bracket 1 .126 times 2991 close bracket over 8145 +2991 end = fraction 11789 .796 over 11136 end simeq 1.059.

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative23)

End of $1

[Back to - Part](" \l "Session6_Part3)

### Part

#### Discussion

Using the 2009 expenditures for weights instead of the 2008 expenditures, the overall energy price ratio for 2009 relative to 2008 is

Start of $1

fraction open bracket 1 .034 times 23733 close bracket + open bracket 1 .126 times 2275 close bracket over 23733 +2275 end = fraction 27101 .572 over 26008 end simeq 1.042.

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative24)

End of $1

This price ratio is considerably less than the one found in part (b).

(Note that if full calculator accuracy is retained throughout the calculations, the price ratio is 1.043 to three decimal places.)

[Back to - Part](" \l "Session6_Part4)

## Activity 18 Gradgrind’s energy price index for 2010

#### Discussion

The gas price ratio for 2010 relative to 2009 is

Start of $1

fraction 28 over 30 end simeq 0.933.

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative28)

End of $1

The electricity price ratio for 2010 relative to 2009 is

Start of $1

fraction 88 over 98 end simeq 0.898.

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative29)

End of $1

(Both price ratios are less than 1 because, over this year, Gradgrind’s gas and electricity prices both fell.)

The overall energy price ratio for 2010 relative to 2009 is

Start of $1

fraction open bracket 0 .933 times 23733 close bracket + open bracket 0 .898 times 2275 close bracket over 23733 +2275 end = fraction 24185 .839 over 26008 end simeq 0.930.

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative30)

End of $1

Then the value of the index for 2010 is found by multiplying the 2009 value of the index by this overall price ratio, giving

Start of $1

126.2 times 0.930 simeq 117.4.

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative31)

End of $1

[Back to - Activity 18 Gradgrind’s energy price index for 2010](" \l "Session6_Activity5)

## Exercise 9 Gradgrind’s energy price index for 2011

#### Discussion

The gas price ratio for 2011 relative to 2010 is

Start of $1

fraction 30 over 28 end simeq 1.071.

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative32)

End of $1

The electricity price ratio for 2011 relative to 2010 is

Start of $1

fraction 86 over 88 end simeq 0.977.

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative33)

End of $1

The overall energy price ratio for 2011 relative to 2010 is

Start of $1

fraction open bracket 1 .071 times 23969 close bracket + open bracket 0 .977 times 2920 close bracket over 23969 +2920 end = fraction 28523 .639 over 26889 end simeq 1.061.

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative34)

End of $1

Then the value of the index for 2011 is found by multiplying the 2010 value of the index by this overall price ratio, giving

Start of $1

117.4 times 1.061 simeq 124.6.

[View alternative description - Uncaptioned Equation](" \l "Session6_Alternative35)

End of $1

[Back to - Exercise 9 Gradgrind’s energy price index for 2011](" \l "Session6_Exercise1)

## Activity 19 The expenditure of a typical household

### Part

#### Discussion

What you need to remember here is that the size of an area represents the proportion of expenditure on that class of goods or services. (Also, it is admittedly not very easy to estimate these areas ‘by eye’! Your estimates might quite reasonably differ from those given here.)

* The sector for ‘Personal expenditure’ looks as if it is approximately a tenth of the whole inner circle – so approximately a tenth of total expenditure is personal expenditure.
* ‘Housing and household expenditure’ looks as if it is somewhere between a third and a half of the inner circle – perhaps approximately two fifths – so approximately two fifths of expenditure is on housing and household expenditure.
* The area for ‘Housing’ takes up about a quarter of the outer ring, so about a quarter of expenditure is on housing.

[Back to - Part](" \l "Session7_Part1)

### Part

#### Discussion

The amount spent each week on ‘Personal expenditure’ is approximately

Start of $1

fraction 1 over 10 end times pounds 540 = pounds 54 .

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative4)

End of $1

The amount spent each week on ‘Housing and household expenditure’ is approximately

Start of $1

fraction 2 over 5 end times pounds 540 = pounds 216 simeq pounds 220 .

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative5)

End of $1

The amount spent each week on ‘Housing’ is approximately

Start of $1

fraction 1 over 4 end times pounds 540 = pounds 135 simeq pounds 140 .

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative6)

End of $1

Recall, however, that the weights represent average proportions of expenditure, and the spending patterns of the selected household may differ from those of the ‘typical’ household.

[Back to - Part](" \l "Session7_Part2)

## Activity 20 Your own household’s expenditure

#### Discussion

Every household will be different, but think about the reasons for any large differences between your weights and those for the RPI.

[Back to - Activity 20 Your own household’s expenditure](" \l "Session7_Activity2)

## Activity 21 Calculating the RPI for July 2011

#### Discussion

Start of Table

|  |  |  |  |
| --- | --- | --- | --- |
| **Group** | **Price ratio for July 2011 relative to January 2011: r** | **2011 weights: w** | **Price ratio times weight: r w** |
| Food and catering | 1.024 | 165 | 168.960 |
| Alcohol and tobacco | 1.042 | 88 | 91.696 |
| Housing and household expenditure | 1.012 | 408 | 412.896 |
| Personal expenditure | 1.053 | 82 | 86.346 |
| Travel and leisure | 1.030 | 257 | 264.710 |
| **Sum** |  | **1000** | **1024.608** |

End of Table

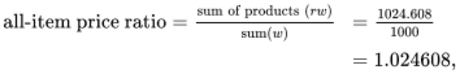
Start of $1

sum open bracket w close bracket =1000 comma sum of products open bracket r w close bracket =1024.608 comma

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative34)

End of $1

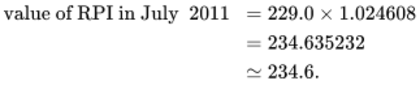
Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative35)

End of $1

Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative36)

End of $1

[Back to - Activity 21 Calculating the RPI for July 2011](" \l "Session7_Activity3)

## Activity 22 The effects of particular price changes on the RPI

### Part

#### Discussion

The RPI is calculated using the price ratio and weight of each item. Since the weights of items change very little from one year to the next, the price ratio alone will normally tell you whether a change in price is likely to lead to an increase or a decrease in the value of the RPI. If a price rises, then the price ratio is greater than one, so the RPI is likely to increase as a result. If a price falls, then the price ratio is less than one, so the RPI is likely to decrease. Therefore, since the price of leisure goods fell, this is likely to lead to a decrease in the value of the RPI. For a similar reason, the increase in the price of canteen meals is likely to lead to an increase in the value of the RPI.

[Back to - Part](" \l "Session7_Part4)

### Part

#### Discussion

Both changes are likely to be small for two reasons. First, the price changes are themselves fairly small. Second, leisure goods and canteen meals form only part of a household’s expenditure: no single group, subgroup or section will have a large effect on the RPI on its own, unless there is a very large change in its price.

[Back to - Part](" \l "Session7_Part5)

### Part

#### Discussion

The weight of ‘Leisure goods’ was 33 in 2012 (see [Table 12](#open-u2tab5-1)). Since ‘Canteen meals’ is only one section in the subgroup ‘Catering’, which had weight 47 in 2012, the weight of ‘Canteen meals’ will be much smaller than 47. (In fact it was 3.) So the weight of ‘Leisure goods’ is much larger than the weight of ‘Canteen meals’.

[Back to - Part](" \l "Session7_Part6)

### Part

#### Discussion

Since the weight of ‘Leisure goods’ is much larger than the weight of ‘Canteen meals’, and the percentage change in the prices are not too different in size, the change in the price of leisure goods is likely to have a much larger effect on the value of the RPI as a whole.

[Back to - Part](" \l "Session7_Part7)

## Activity 23 The annual inflation rate in February 2012

#### Discussion

The ratio of the two RPI values is

Start of $1

fraction value of RPI in February 2012 over value of RPI in February 2011 end = fraction 239 .9 over 231 .3 end simeq 1.037 comma

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative38)

End of $1

or 103.7%. Therefore the annual inflation rate, based on the RPI was 3.7%. (Note that this is slightly higher than the annual inflation rate measured using the CPI.)

[Back to - Activity 23 The annual inflation rate in February 2012](" \l "Session7_Activity5)

## Activity 24 Index-linking a pension using the CPI

#### Discussion

The weekly amount in November 2011 should be

Start of $1

pounds 120 times fraction 121 .2 over 115 .6 end simeq pounds 125.81.

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative40)

End of $1

[Back to - Activity 24 Index-linking a pension using the CPI](" \l "Session7_Activity6)

## Activity 25 Annual inflation and the purchasing power of the pound

### Part

#### Discussion

For May 2010, the ratio of the value of the RPI to its value one year earlier is

Start of $1

fraction 223 .6 over 212 .8 end simeq 1.051 comma

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative46)

End of $1

so the annual inflation rate is 5.1%.

The purchasing power of the pound compared to one year previously is

Start of $1

fraction 212 .8 over 223 .6 end times 100 p simeq 95 p .

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative47)

End of $1

[Back to - Part](" \l "Session7_Part9)

### Part

#### Discussion

For October 2011, the ratio of the value of the RPI to its value one year earlier is

Start of $1

fraction 238 .0 over 225 .8 end simeq 1.054 comma

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative48)

End of $1

so the annual inflation rate is 5.4%.

The purchasing power of the pound compared to one year previously is

Start of $1

fraction 225 .8 over 238 .0 end times 100 p simeq 95 p .

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative49)

End of $1

[Back to - Part](" \l "Session7_Part10)

### Part

#### Discussion

For March 2011, the ratio of the value of the RPI to its value one year earlier is

Start of $1

fraction 232 .5 over 220 .7 end simeq 1.053 comma

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative50)

End of $1

so the annual inflation rate is 5.3%.

The purchasing power of the pound compared to one year previously is

Start of $1

fraction 220 .7 over 232 .5 end times 100 p simeq 95 p .

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative51)

End of $1

[Back to - Part](" \l "Session7_Part11)

## Exercise 10 Calculating the RPI for February 2012

#### Discussion

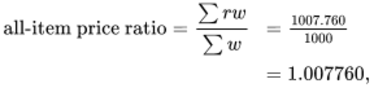
Start of $1

sum w =1000 comma sum r w=1007.760 comma

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative56)

End of $1

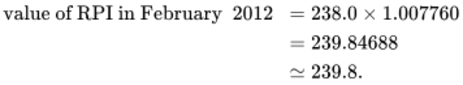
Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative57)

End of $1

Start of $1



[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative58)

End of $1

(The published index was 239.9. Again, the difference between this and your calculated value is because the ONS statisticians used more accuracy in their intermediate calculations.)

[Back to - Exercise 10 Calculating the RPI for February 2012](" \l "Session7_Exercise1)

## Exercise 11 Annual inflation rates and the purchasing power of the pound

### Part

#### Discussion

For October 2010, the ratio of the value of the RPI to its value one year earlier is

Start of $1

fraction 225 .8 over 216 .0 end simeq 1.045 comma

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative59)

End of $1

so the annual inflation rate is 4.5%.

The purchasing power of the pound compared to one year previously is

Start of $1

fraction 216 .0 over 225 .8 end times 100 p simeq 96 p .

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative60)

End of $1

[Back to - Part](" \l "Session7_Part13)

### Part

#### Discussion

For January 2011, the ratio of the value of the RPI to its value one year earlier is

Start of $1

fraction 229 .0 over 217 .9 end simeq 1.051 comma

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative61)

End of $1

so the annual inflation rate is 5.1%.

The purchasing power of the pound compared to one year previously is

Start of $1

fraction 217 .9 over 229 .0 end times 100 p simeq 95 p .

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative62)

End of $1

[Back to - Part](" \l "Session7_Part14)

## Exercise 12 Index-linking another pension

#### Discussion

The RPI for April 2011 was 234.4 and the RPI for April 2010 was 222.8. So in April 2011, the pension should be

Start of $1

pounds 800 times fraction 234 .4 over 222 .8 end simeq pounds 842 per month .

[View alternative description - Uncaptioned Equation](" \l "Session7_Alternative63)

End of $1

[Back to - Exercise 12 Index-linking another pension](" \l "Session7_Exercise3)

## Descriptions

### Figure 1 Stemplot of coffee prices from Table 1

An ordered stemplot of coffee prices from Table 1. The first column contains the numbers in the stem while the second column contains the relevant leaf values. There are 11 levels. The numbers in the stem start at 26, increase in steps of one and end at 36. The leaf values correspond to the relevant numbers in the stem, though sometimes there is more than one leaf and sometimes none at all. In this stemplot they are in numerical order.

At level 26 there are five leaves, 8, 8, 8, 8, 9. Level 27 has two leaves, 5, 9. Level 28 has no leaves while level 29 has five leaves, 5, 5, 5, 5, 9. Level 30 has a single leaf, 5, and level 31 has a single leaf, 5. Levels 32, 33, 34 and 35 have no leaves. Level 36 has a single leaf, 9. Beneath the stemplot is written n = 15, followed by 26 vertical line 8 represents 268 pence. The vertical line sits horizontally between the 26 and the 8.

[Back to - Figure 1 Stemplot of coffee prices from Table 1](" \l "Session3_Figure1)

### Figure 2 Subscript notation for ordered data

Subscript notation for ordered data. This shows a table with 2 rows and 15 columns. The first row contains x subscript 1, x subscript 2 and so on, ending at x subscript 15. The subscript numbers are in round brackets and positioned to the right of and slightly below each x. Above this row is a cloud with an arrow pointing to the term x subscript 3. In the cloud is written ‘The subscript is (3), so this is the third value in the ordered batch.’

The second row contains the corresponding values. These are 268 (shown below x subscript 1), 268, 268, 268, 269, 275, 279, 295, 295, 295, 295, 299, 305, 315 and 369 (shown below x subscript 15). Below the table are three clouds with arrows pointing upwards. The first cloud contains ‘capital E subscript capital L’ and the arrow points to the first number in the second row, 268. The middle cloud contains the word ‘Median’ and the arrow points to the first 295 in the second row, which is under x subscript 8. The last cloud contains ‘capital E subscript capital U’ and the arrow points to the last number in the second row, 369.

[Back to - Figure 2 Subscript notation for ordered data](" \l "Session3_Figure2)

### Figure 3 Median of 15 values

Median of 15 values. The letters x subscript 1, x subscript 2 and so on, each with the subscripts in round brackets, are arranged in order, but in a V-shaped formation. At the top left-hand end of the V is x subscript 1. The letters then descend in order down the slope until they reach x subscript 8. Then they begin to rise up a slope to form the right-hand side of the V, starting with x subscript 9, which is level with x subscript 7. They continue upwards to end at x subscript 15, which is level with and to the far right of x subscript 1. Below the point of the V is a cloud containing the word ‘Median’. From the cloud an arrow points to x subscript 8.

[Back to - Figure 3 Median of 15 values](" \l "Session3_Figure3)

### Figure 4 Prices of 10 digital cameras

V-shaped formation for prices of ten digital cameras. Reading down the slope from the top left, the numbers are 53, 60, 65, 70 and 70, but this V shape has no number at its point. The number 74 is to the right of but level with the preceding number 70. The remaining numbers 79, 81, 85, 90 then rise up the slope, with 90 being level with and to the far right of 53.

[Back to - Figure 4 Prices of 10 digital cameras](" \l "Session3_Figure4)

### Figure 5 Prices of all flat-screen televisions with a screen size of 24 inches or less on a major UK retailer’s website on a day in February 2012

Stemplot with 10 levels, though there are repeated numbers in the stem. The numbers in the stem are 0, 1, 1, 1, 1, 1, 2, 2, 2, 2. At level 0 there is one leaf 9. At the first level 1 there is one leaf, 0. At the second level 1 there are four leaves 2, 3, 3, 3. The third level 1 has five leaves, 4, 5, 5, 5, 5. The fourth level 1 has three leaves, 6, 6, 7. The fifth and last level 1 has three leaves, 8, 8, 9. The first and second level 2s have no leaves. The third level 2 has two leaves, 4, 5. The last level 2 has one leaf, 7. Beneath the stemplot is written n = 20, followed by 0 vertical line 9 represents 90 pounds sterling. The vertical line sits horizontally between the 0 and the 9.

[Back to - Figure 5 Prices of all flat-screen televisions with a screen size of 24 inches or less on a major UK retailer’s website on a day in February 2012](" \l "Session3_Figure5)

### Figure 6 Stemplot of 14 gas prices

A stemplot with 8 levels. The numbers in the stem start at 374, increase in steps of one and end at 381. At level 374 there are 3 leaves, 0, 0, 3. Level 375 has no leaves. Level 376 has two leaves, 0, 7 while level 377 has one leaf, 6. Level 378 has one leaf, 4. Level 379 has two leaves, 5, 6. Level 380 has four leaves, 1, 1, 4, 5. Level 381 has one leaf, 8. Beneath the stemplot is written n = 14, followed by 374 vertical line 0 represents 3.740 pence per kilowatt hour. The vertical line sits horizontally between the 374 and the 0.

[Back to - Figure 6 Stemplot of 14 gas prices](" \l "Session3_Figure6)

### Figure 7 Stemplots for northern and southern cities separately.

There are two stemplots, one for northern cities and one for southern cities. The one for the northern cities has 7 levels. The numbers in the stem start at 374, increase in steps of one and end at 380.

At level 374 there are 2 leaves, 0, 0. Level 375 has no leaves. Level 376 has one leaf, 7 and level 377 also has one leaf, 6. Levels 378 and level 379 have no leaves. Level 380 has three leaves, 1, 1, 4. Beneath the stemplot is written n = 7, followed by 374 vertical line 0 represents 3.740 pence per kilowatt hour. The vertical line sits horizontally between the 374 and the 0. The one for the southern cities has 8 levels. The numbers in the stem start at 374, increase in steps of one and end at 381. At level 374 there is one leaf, 3. Level 375 has no leaves. Level 376 has one leaf, 0 but level 377 has no leaves. Level 378 has one leaf, 4, and level 379 has two leaves, 5, 6. Level 380 has one leaf, 5 and level 381 has one leaf, 8.

Beneath the stemplot is written n = 7, followed by 374 vertical line 3 represents 3.743 pence per kilowatt hour. The vertical line sits horizontally between the 374 and the 3.

[Back to - Figure 7 Stemplots for northern and southern cities separately.](" \l "Session3_Figure7)

### Uncaptioned Equation

3.818 minus 3.740=0.078 comma

[Back to - Uncaptioned Equation](" \l "Session3_Equation1)

### Uncaptioned Equation

3.804 minus 3.740=0.064 comma

[Back to - Uncaptioned Equation](" \l "Session3_Equation2)

### Uncaptioned Equation

3.818 minus 3.743=0.075.

[Back to - Uncaptioned Equation](" \l "Session3_Equation3)

### Uncaptioned Equation

mean = fraction sum over size end .

[Back to - Uncaptioned Equation](" \l "Session3_Equation4)

### Uncaptioned Equation

fraction sum over size end = fraction 4 + 8 + 4 + 2 + 9 over 5 end = fraction 27 over 5 end = 5.4.

[Back to - Uncaptioned Equation](" \l "Session3_Equation5)

### Uncaptioned Equation

mean = fraction sum over size end

[Back to - Uncaptioned Equation](" \l "Session3_Equation6)

### Uncaptioned Equation

overline x = fraction sum x over n end .

[Back to - Uncaptioned Equation](" \l "Session3_Equation7)

### Uncaptioned Equation

mean = fraction sum over size end = fraction 90 + 100 + ellipsis +270 over 20 end = pounds 162.

[Back to - Uncaptioned Equation](" \l "Session3_Equation8)

### Uncaptioned Equation

mean = overline x = fraction sum x over n end = fraction 3240 over 20 end = pounds 162.

[Back to - Uncaptioned Equation](" \l "Session3_Equation9)

### Uncaptioned Equation

fraction 4363 p over 15 end simeq 290.9 p .

[Back to - Uncaptioned Equation](" \l "Session3_Equation10)

### Uncaptioned Equation

x subscript open bracket 1 close bracket end = 240 and x subscript open bracket 15 close bracket end = 340.

[Back to - Uncaptioned Equation](" \l "Session3_Equation11)

### Uncaptioned Equation

fraction 4306 p over 15 end simeq 287.1 p .

[Back to - Uncaptioned Equation](" \l "Session3_Equation12)

### Uncaptioned Equation

fraction pounds 3470 over 20 end = pounds 173.5 simeq pounds 174

[Back to - Uncaptioned Equation](" \l "Session3_Equation13)

### Figure 8

A stemplot with 11 levels. The numbers in the stem start at 0, increase in steps of one and end at 10. At level 0 there is one leaf, 7. Level 1 has one leaf, 5. Level 2 has no leaves, while level 3 has two leaves, 3, 5. Level 4 has three leaves, 2, 2, 3 and level 5 has two leaves, 5, 8. Level 6 has three leaves, 4, 6, 8. Level 7 has five leaves, 1, 1, 6, 8, 9. Level 8 has nine leaves, 0, 1, 1, 3, 4, 5, 5, 6, 9. Level 9 has five leaves, 1, 1, 3, 5, 9. Level 10 has two leaves which are both 0. Beneath the stemplot is written n = 33, followed by 0 vertical line 7 represents a score of 7 per cent. The vertical lines sits horizontally between the 0 and the 7.

[Back to - Figure 8](" \l "Session3_Figure9)

### Uncaptioned Equation

fraction 1 over 2 end open bracket pounds 269 + pounds 270 close bracket = pounds 269.5 simeq pounds 270.

[Back to - Uncaptioned Equation](" \l "Session3_Equation14)

### Uncaptioned Equation

fraction 2326 over 33 end simeq 70.5.

[Back to - Uncaptioned Equation](" \l "Session3_Equation15)

### Uncaptioned Equation

fraction 7856 over 26 end = 302.1538 simeq 302.2.

[Back to - Uncaptioned Equation](" \l "Session3_Equation16)

### Uncaptioned Equation

fraction 2480 over 17 end =145.8824 simeq 146.

[Back to - Uncaptioned Equation](" \l "Session3_Equation17)

### Figure 9 Means of biscuit prices

The figure shows a horizontal arrow pointing to the right and labelled pence. Part way in from the left is a marker above which is written 74.0. Farther to the right, but before the end of the arrow, is another marker, above which is written 81.6.

[Back to - Figure 9 Means of biscuit prices](" \l "Session4_Figure1)

### Uncaptioned Equation

fraction sum open bracket of the combined batch prices close bracket over size open bracket of the combined batch close bracket end .

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

### Uncaptioned Equation

mean = fraction sum over size end

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

### Uncaptioned Equation

sum = mean times size .

[Back to - Uncaptioned Equation](" \l "Session4_Equation3)

### Uncaptioned Equation

mean = fraction combined sum over combined size end = fraction 408 +592 over 13 end = fraction 1000 over 13 end simeq 76.9

[Back to - Uncaptioned Equation](" \l "Session4_Equation4)

### Uncaptioned Equation

overline x subscript uppercase C end = fraction overline x subscript uppercase A end n subscript uppercase A end + overline x subscript uppercase B end n subscript uppercase B end over n subscript uppercase A end + n subscript uppercase B end end comma

[Back to - Uncaptioned Equation](" \l "Session4_Equation5)

### Uncaptioned Equation

overline x subscript uppercase A end = mean of batch uppercase A comma n subscript uppercase A end = size of batch uppercase A comma overline x subscript uppercase B end = mean of batch uppercase B comma n subscript uppercase B end = size of batch uppercase B.

[Back to - Uncaptioned Equation](" \l "Session4_Equation6)

### Uncaptioned Equation

overline x subscript uppercase A end = 81.6 comma n subscript uppercase A end = 5 comma overline x subscript uppercase B end = 74.0 comma n subscript uppercase B end = 8.

[Back to - Uncaptioned Equation](" \l "Session4_Equation7)

### Uncaptioned Equation

overline x subscript uppercase C end = fraction open bracket 81 .6 times 5 close bracket + open bracket 74 .0 times 8 close bracket over 5 +8 end .

[Back to - Uncaptioned Equation](" \l "Session4_Equation8)

### Figure 10 Point of balance at the weighted mean

Point of balance at the weighted mean. The figure shows a horizontal arrow pointing to the right and labelled pence. Three appropriately spaced points are marked on the line: 74.0, 76.9 and 81.6. The distance between the 74.0 and 76.9 is less than the distance between 76.9 and 81.6. Immediately below 76.9 there is a solid arrow head pointing upwards, which is the fulcrum or balancing point of the balance. Suspended from the point marked 74.0 is a pile of 8 discs and suspended from the point marked 81.6 is a pile of 5 discs.

[Back to - Figure 10 Point of balance at the weighted mean](" \l "Session4_Figure2)

### Uncaptioned Equation

Batch uppercase A has mean 119 and size 7. Batch uppercase B has mean 185 and size 13 .

[Back to - Uncaptioned Equation](" \l "Session4_Equation9)

### Uncaptioned Equation

overline x subscript uppercase A end = 119 comma n subscript uppercase A end = 7 comma overline x subscript uppercase B end = 185 comma n subscript uppercase B end = 13.

[Back to - Uncaptioned Equation](" \l "Session4_Equation10)

### Uncaptioned Equation

overline x subscript uppercase C end = fraction open bracket 119 times 7 close bracket + open bracket 185 times 13 close bracket over 7 +13 end = fraction 833 +2405 over 20 end = fraction 3238 over 20 end = 161.9 simeq 162.

[Back to - Uncaptioned Equation](" \l "Session4_Equation11)

### Uncaptioned Equation

cost = price times quantity .

[Back to - Uncaptioned Equation](" \l "Session4_Equation12)

### Uncaptioned Equation

fraction open bracket 136 .9 times 41 .2 close bracket + open bracket 148 .0 times 10 close bracket over 41 .2 + 10 end .

[Back to - Uncaptioned Equation](" \l "Session4_Equation13)

### Uncaptioned Equation

overline p = fraction p sub 1 q sub 1 + p sub 2 q sub 2 over q sub 1 + q sub 2 end .

[Back to - Uncaptioned Equation](" \l "Session4_Equation14)

### Uncaptioned Equation

. begin 2 by 2 array q sub 1 =10 next column quantity next row p sub 1 = 40 next column price end array } first occasion

[Back to - Uncaptioned Equation](" \l "Session4_Equation15)

### Uncaptioned Equation

. begin 2 by 2 array q sub 2 =6 next column quantity next row p sub 2 =45 next column price end array } second occasion.

[Back to - Uncaptioned Equation](" \l "Session4_Equation16)

### Uncaptioned Equation

overline p = fraction open bracket 40 times 10 close bracket + open bracket 45 times 6 close bracket over 10 +6 end = fraction 400 + 270 over 16 end = fraction 670 over 16 end =41.875 simeq 41.9.

[Back to - Uncaptioned Equation](" \l "Session4_Equation17)

### Uncaptioned Equation

fraction overline x subscript uppercase A end n subscript uppercase A end + overline x subscript uppercase B end n subscript uppercase B end over n subscript uppercase A end + n subscript uppercase B end end and fraction p sub 1 q sub 1 + p sub 2 q sub 2 over q sub 1 + q sub 2 end comma

[Back to - Uncaptioned Equation](" \l "Session4_Equation18)

### Uncaptioned Equation

fraction x sub 1 w sub 1 + x sub 2 w sub 2 over w sub 1 + w sub 2 end .

[Back to - Uncaptioned Equation](" \l "Session4_Equation19)

### Uncaptioned Equation

fraction 1 over 2 end open bracket 3.818 + 3.740 close bracket = 3.779.

[Back to - Uncaptioned Equation](" \l "Session4_Equation20)

### Uncaptioned Equation

fraction 3 .818 q sub 1 + 3 .740 q sub 2 over q sub 1 + q sub 2 end comma

[Back to - Uncaptioned Equation](" \l "Session4_Equation21)

### Uncaptioned Equation

fraction open bracket 3 .818 times 83 close bracket + open bracket 3 .740 times 4 close bracket over 83 + 4 end .

[Back to - Uncaptioned Equation](" \l "Session4_Equation22)

### Uncaptioned Equation

fraction sum of { number times weight } over sum of weights end = fraction sum of products over sum of weights end .

[Back to - Uncaptioned Equation](" \l "Session4_Equation23)

### Uncaptioned Equation

Batch 1 with mean 525.5 and batch size 6. Batch 2 with mean 468.0 and batch size 2. Batch 3 with mean 504.2 and batch size 12.

[Back to - Uncaptioned Equation](" \l "Session4_Equation24)

### Uncaptioned Equation

fraction sum of products over sum of weights end = fraction 10139 .4 over 20 end = 506.97.

[Back to - Uncaptioned Equation](" \l "Session4_Equation25)

### Uncaptioned Equation

1.3 with weight 2 1.9 with weight 1 1.7 with weight 3.

[Back to - Uncaptioned Equation](" \l "Session4_Equation26)

### Uncaptioned Equation

fraction open bracket 1 .3 times 2 close bracket + open bracket 1 .9 times 1 close bracket + open bracket 1 .7 times 3 close bracket over 2 + 1 + 3 end = fraction 2 .6 + 1 .9 + 5 .1 over 6 end = fraction 9 .6 over 6 end = 1.6.

[Back to - Uncaptioned Equation](" \l "Session4_Equation27)

### Figure 11 Point of balance for three means

Point of balance for three means. The figure shows a horizontal arrow pointing to the right. Four suitably spaced points are marked on it, 1.3, 1.6, 1.7 and 1.9. Immediately below 1.6 there is a solid arrow head pointing upwards, which is the fulcrum or balancing point of the balance. Suspended from the point marked 1.3 is a pile of 2 discs, suspended from the point marked 1.7 is a pile of 3 discs and suspended from the point marked 1.9 is a single disc.

[Back to - Figure 11 Point of balance for three means](" \l "Session4_Figure4)

### Uncaptioned Equation

fraction sum of products open bracket price times weight close bracket over sum of weights end

[Back to - Uncaptioned Equation](" \l "Session4_Equation28)

### Uncaptioned Equation

fraction sum x w over sum w end .

[Back to - Uncaptioned Equation](" \l "Session4_Equation29)

### Uncaptioned Equation

fraction 6973 .436 over 1834 end =3.802310 simeq 3.802.

[Back to - Uncaptioned Equation](" \l "Session4_Equation30)

### Uncaptioned Equation

fraction sum x w over sum w end = fraction 24854 .49 over 1892 end = 13.136623 simeq 13.14.

[Back to - Uncaptioned Equation](" \l "Session4_Equation31)

### Uncaptioned Equation

Batch uppercase A has mean price pounds 80.7 and batch size 10. Batch uppercase B has mean price pounds 78.5 and batch size 17.

[Back to - Uncaptioned Equation](" \l "Session4_Equation32)

### Uncaptioned Equation

fraction open bracket 80 .7 times 10 close bracket + open bracket 78 .5 times 17 close bracket over 10 + 17 end = fraction 2141 .5 over 27 end comma

[Back to - Uncaptioned Equation](" \l "Session4_Equation33)

### Uncaptioned Equation

fraction open bracket 10 .95 times 8 .5 close bracket + open bracket 12 .70 times 6 close bracket over 8 .5 + 6 end = fraction 169 .275 over 14 .5 end comma

[Back to - Uncaptioned Equation](" \l "Session4_Equation34)

### Uncaptioned Equation

range = uppercase E subscript uppercase U end minus uppercase E subscript uppercase L end comma

[Back to - Uncaptioned Equation](" \l "Session5_Equation1)

### Figure 12 Median and quartiles

An extended V-shaped diagram showing the median and quartiles. The V-shape has two additional arms, one to the left and one to the right, so the shape now resembles a capital letter M, but with the sides sloping. The numbers represented by x subscript 1, x subscript 2 and so on, ending with x subscript 15 are shown.

The shape starts on the left with x subscript 1, followed, slightly higher and to the right by x subscript 2. Following the same line are x subscript 3 and x subscript 4. At that point the V formation begins, so the next entry x subscript 5 is slightly lower down and to the right and level with x subscript 3. The same line is followed by x subscript 6, x subscript 7, which is level with x subscript 1, and x subscript 8, which is lower than x subscript 1. At x subscript 8 the direction changes again and the letters move slightly to the right and begin to rise, so x subscript 9 is level with x subscript 7. They continue to rise until x subscript 12 is reached. This is level with x subscript 4. The letters then begin to fall again, moving to the right each time, ending with x subscript 15. This is on the same horizontal level as x subscript 1.

There are three clouds, each containing words. The first contains the words ‘Lower quartile’ and has an arrow pointing to the label x subscript 4, which is at the top of the first line. To the right and at the same level is another cloud, containing the words ‘Upper quartile’, from which an arrow points to the label x subscript 12. The third cloud contains the word ‘Median’. From it an arrow points to the lowest label, x subscript 8, in the middle of the diagram.

[Back to - Figure 12 Median and quartiles](" \l "Session5_Figure1)

### Figure 13 Quartiles for sample size n=17

Diagram showing the quartiles for sample size n = 17. The V-shape has two additional arms, one to the left and one to the right, so the shape now resembles a capital letter M, but with the sides sloping. The numbers represented by x subscript 1, x subscript 2 and so on, ending with x subscript 17 are shown.

The shape starts on the left with x subscript 1, followed, slightly higher and to the right by x subscript 2, Following the same line are x subscript 3 and x subscript 4. At that point the V formation begins, but there is a gap at the top. The next entry x subscript 5 is level with but to the right of x subscript 4. The same line is followed by x subscript 6, x subscript 7 and x subscript 8, which is level with x subscript 1. Continuing in the same direction leads to x subscript 9, which is at a lower level than x subscript 1. At x subscript 9 the direction changes again and the letters begin to rise, moving slightly to the right again until x subscript 13 is reached. This is level with x subscript 5. X subscript 14 is to the right of and level with x subscript 13 and forms the start of the last slope. The letters then begin to fall again, moving to the right each time, ending with x subscript 17. This is on the same horizontal level as x subscript 1.

There are three clouds, each containing words. The first contains the words ‘Lower quartile’ and has an arrow pointing to the gap between x subscript 4 and x subscript 5. To the right and at the same level is another cloud, containing the words ‘Upper quartile’, from which an arrow points to the gap between x subscript 13 and x subscript 14. The third cloud contains the word ‘Median’. From it an arrow points to the lowest label, x subscript 9, in the middle of the diagram.

[Back to - Figure 13 Quartiles for sample size n=17](" \l "Session5_Figure2)

### Figure 14 Quartiles for sample size n=18

Diagram showing the quartiles for sample size n = 18. The V-shape has two additional arms, one to the left and one to the right, so the shape now resembles a capital letter M, but with the sides sloping. The numbers represented by x subscript 1, x subscript 2 and so on, ending with x subscript 18 are shown.

The shape starts on the left with x subscript 1, followed, slightly higher and to the right by x subscript 2, Following the same line are x subscript 3 and x subscript 4. At that point the V formation begins, but there is a gap at the top. The next entry x subscript 5 is level with but to the right of x subscript 4. The same line is followed by x subscript 6, x subscript 7 and x subscript 8, which is level with x subscript 1. Continuing in the same direction leads to x subscript 9, which is at a lower level than x subscript 1. At x subscript 9 there is a gap. The direction changes again and the letters begin to rise. Slightly to the right but level with x subscript 9 is x subscript 10. They continue to rise and move to the right until x subscript 14 is reached. This is level with x subscript 5. There is then a gap and the letters then begin to fall again, starting with x subscript 15 which is level with x subscript 14. They continue down, moving to the right each time, ending with x subscript 18. This is on the same horizontal level as x subscript 1.

There are three clouds, each containing words. The first contains the words ‘Lower quartile’ and has an arrow pointing to the gap between x subscript 4 and x subscript 5 with the arrowhead closer to x subscript 5 than x subscript 4. To the right and at the same level is another cloud, containing the words ‘Upper quartile’, from which an arrow points to the gap between x subscript 14 and x subscript 15 with the arrowhead closer to x subscript 14 than x subscript 15. Below the diagram a third cloud contains the word ‘Median’. From it an arrow points to the gap between x subscript 9 and x subscript 10, in the middle of the diagram.

[Back to - Figure 14 Quartiles for sample size n=18](" \l "Session5_Figure3)

### Figure 15 Quartiles for sample size n=20

Diagram showing the quartiles for sample size n = 20. The V-shape has two additional arms, one to the left and one to the right, so the shape now resembles a capital letter M, but with the sides sloping. The numbers represented by x subscript 1, x subscript 2 and so on, ending with x subscript 20 are shown.

The shape starts with x subscript 1, followed, slightly higher and to the right by x subscript 2, Following the same line are x subscript 3, x subscript 4 and x subscript 5. At that point the V formation begins, but there is a gap at the top. The next entry x subscript 6 is level with but to the right of x subscript 5. The same line is followed by x subscript 7, x subscript 8, x subscript 9 and x subscript 10, which is level with x subscript 1. At x subscript 10 there is a gap. The direction changes again and the letters begin to rise. Slightly to the right but level with x subscript 10 is x subscript 11. The letters continue to rise and move to the right until x subscript 15 is reached. This is level with x subscript 6. There is then a gap and the letters move to the right and begin to fall again, starting with x subscript 16 which is level with x subscript 15. They continue down, moving to the right each time, ending with x subscript 20. This is on the same horizontal level as x subscript 11.

There are three clouds, each containing words. Above the diagram the first cloud contains the words ‘Lower quartile’ and has an arrow pointing to the gap between x subscript 5 and x subscript 6 with the arrowhead closer to x subscript 5 than x subscript 6. To the right and at the same level is another cloud, containing the words ‘Upper quartile’, from which an arrow points to the gap between x subscript 15 and x subscript 16 with the arrowhead closer to x subscript 16 than x subscript 15. The third cloud is below the diagram and contains the word ‘Median’. From it an arrow points to the gap between x subscript 10 and x subscript 11, in the middle of the diagram.

[Back to - Figure 15 Quartiles for sample size n=20](" \l "Session5_Figure4)

### Figure 16 Prices of flat-screen televisions with a screen size of 24 inches or less

Stemplot with 10 levels, though there are repeated numbers in the stem. The numbers in the stem are 0, 1, 1, 1, 1, 1, 2, 2, 2, 2. At level 0 there is one leaf 9. At the first level 1 there is one leaf, 0. At the second level 1 there are four leaves 2, 3, 3, 3. The third level 1 has five leaves, 4, 5, 5, 5, 5. The fourth level 1 has three leaves, 6, 6, 7. The fifth and last level 1 has three leaves, 8, 8, 9. The first level 2 and the second level 2 have no leaves. The third level 2 has two leaves, 4, 5. The last level 2 has one leaf, 7. Beneath the stemplot is written n = 20, followed by 0 vertical line 9 represents 90 pounds sterling. The vertical line sits horizontally between the 0 and the 9.

[Back to - Figure 16 Prices of flat-screen televisions with a screen size of 24 inches or less](" \l "Session5_Figure5)

### Uncaptioned Equation

53 60 65 70 70 74 79 81 85 90

[Back to - Uncaptioned Equation](" \l "Session5_Equation2)

### Figure 17 Stemplot of 15 coffee prices

A stemplot with 11 levels, which start at 26 and end at 36. At level 26 there are five leaves 8, 8, 8, 8, 9. At level 27 there two leaves 5, 9. At level 28 there are no leaves. Level 29 has five leaves, 5, 5, 5, 5, 9. Level 30 has one leaf, 5 and level 31 also has one leaf, 5. Levels 32, 33, 34, and 35 have no leaves. Level 36 has one leaf, 9. Beneath the stemplot is written n = 15, followed by 26 vertical line 8 represents 268 pence. The vertical line sits horizontally between the 26 and the 8.

[Back to - Figure 17 Stemplot of 15 coffee prices](" \l "Session5_Figure7)

### Figure 18 Stemplot of 14 gas prices

A stemplot with 8 levels, which start at 374 and end at 381. Level 374 has three leaves, 0, 0, 3. Level 375 has no leaves. Level 376 has two leaves, 0, 7. Level 377 has one leaf, 6. Level 378 also has one leaf, 4. Level 379 has two leaves, 5, 6. Level 380 has four leaves, 1, 1, 4, 5. Level 381 has one leaf, 8. Beneath the stemplot is written n = 14, followed by 374 vertical line 0 represents 3.740 pence per kilowatt hour. The vertical line sits horizontally between the 374 and the 0.

[Back to - Figure 18 Stemplot of 14 gas prices](" \l "Session5_Figure8)

### Uncaptioned Equation

uppercase Q sub 1 = 3.743 + fraction 3 over 4 end open bracket 3.760 minus 3.743 close bracket = 3.75575 simeq 3.756

[Back to - Uncaptioned Equation](" \l "Session5_Equation3)

### Uncaptioned Equation

uppercase Q sub 3 = 3.801 + fraction 1 over 4 end open bracket 3.804 minus 3.801 close bracket = 3.80175 simeq 3.802.

[Back to - Uncaptioned Equation](" \l "Session5_Equation4)

### Uncaptioned Equation

IQR = uppercase Q sub 3 minus uppercase Q sub 1 .

[Back to - Uncaptioned Equation](" \l "Session5_Equation5)

### Uncaptioned Equation

IQR = uppercase Q sub 3 minus uppercase Q sub 1 = 180 minus 130 = 50.

[Back to - Uncaptioned Equation](" \l "Session5_Equation6)

### Uncaptioned Equation

range = uppercase E subscript uppercase U end minus uppercase E subscript uppercase L end = 369 p minus 268 p = 101 p .

[Back to - Uncaptioned Equation](" \l "Session5_Equation7)

### Uncaptioned Equation

IQR = uppercase Q sub 3 minus uppercase Q sub 1 = 299 p minus 268 p = 31 p .

[Back to - Uncaptioned Equation](" \l "Session5_Equation8)

### Uncaptioned Equation

IQR = uppercase Q sub 3 minus uppercase Q sub 1 =3.80175 minus 3.75575 =0.046 comma

[Back to - Uncaptioned Equation](" \l "Session5_Equation9)

### Uncaptioned Equation

uppercase E subscript uppercase U end minus uppercase E subscript uppercase L end = 325 p minus 268 p = 57 p comma

[Back to - Uncaptioned Equation](" \l "Session5_Equation10)

### Figure 19 Values in a five-figure summary

Values in a five-figure summary. There are four lines, forming the shape of a letter M with sloping sides. At the base of the line on the left is capital E subscript capital L, indicating the lower extreme, the lowest value in the data. At the top of that line is capital Q subscript 1, indicating the lower quartile. The line then slopes down and to the right. At the end of the second line is capital M, indicating the median. This is on the same horizontal level as capital E subscript capital L. The line then rises and slopes to the right. It ends at capital Q subscript 3, indicating the upper quartile, which is at the same horizontal level as capital Q subscript 1. The line then falls, sloping to the right. At the end of the line is capital E subscript capital U, indicating the upper extreme, the highest value in the data. This is at the same horizontal level as capital M.

[Back to - Figure 19 Values in a five-figure summary](" \l "Session5_Figure9)

### Figure 20

A five-figure summary which is a diagrammatic representation showing the batch size, n, the median capital M, the lower quartile capital Q subscript 1, the upper quartile capital Q subscript 3, the lower extreme capital E subscript capital L, and the upper extreme capital E subscript capital U. The diagram forms three sides of a rectangle, with the bottom line missing. It therefore has a vertical line to the left, a horizontal line across the top and a vertical line to the right. To the left of the left vertical line is written n. Towards the bottom of the line and to its right is written capital E subscript capital L and capital Q subscript 1, with capital Q subscript 1 being above capital E subscript capital L. Beneath the middle of the horizontal line is written capital M. To the left of the second vertical line but level with capital Q subscript 1 is written capital Q subscript 3. Below that, and level with capital E subscript capital L, is written capital E subscript capital U.

[Back to - Figure 20](" \l "Session5_Figure10)

### Figure 21

A five-figure summary. The diagram forms three sides of a rectangle, with the bottom line missing. It therefore has a vertical line to the left, a horizontal line across the top and a vertical line to the right. To the left of the left vertical line is written n = 20. Towards the bottom of the left vertical line and to its right is written 90 and above that 130. Beneath the middle of the horizontal line is written 150. To the left of the second vertical line and level with 130 is written 180. Below that and level with 90 is written 270.

[Back to - Figure 21](" \l "Session5_Figure11)

### Figure 22 Boxplot of batch of 14 gas prices

A boxplot with a horizontal scale with an arrow head, pointing right, at the right-hand end. It is labelled pence per kilowatt hour. The scale starts just before the first marked point of 3.74 and is then marked in four intervals of 0.02, ending with 3.82. The boxplot is drawn above and parallel to the line. The first whisker starts at the lower extreme, level with 3.74.The box starts at the lower quartile, 3.756 and ends at the upper quartile, 3.802. Within the box but nearer the right-hand end is a vertical line indicating the median at 3.790. The second whisker starts at the midpoint of the right-hand end of the box and stretches to 3.818, as far as the upper extreme.

[Back to - Figure 22 Boxplot of batch of 14 gas prices](" \l "Session5_Figure12)

### Figure 23 A standard boxplot with annotation

A boxplot which consists of a line, a rectangle, known as a box, and a second line. The first line, known as a whisker, starts at capital E subscript capital L, the lower extreme and leads to the midpoint of the left side of the box. Immediately above and at the corner of the box is written capital Q subscript 1. This point is the lower quartile. The box stretches for some way to the right, ending at the upper quartile, with capital Q subscript 3 written above the end of the box and on the same horizontal level as capital Q subscript 1. A second horizontal line, also known as a whisker, starts at the midpoint of the right-hand end of the box and stretches as far as capital E subscript capital U, the upper extreme. The median is marked by capital M above the box at the appropriate position and by a vertical line within the box. Beneath the boxplot are 4 horizontal curly brackets, each of which has 25% written under it. The first stretches from immediately beneath capital E subscript capital L to the start of the box. The second bracket stretches from the start of the box to level with the vertical line level indicating the position of the median. The third stretches from the line level with the median to level with capital Q subscript 3 and the last stretches from level with capital Q subscript 3 to level with capital E subscript capital U.

[Back to - Figure 23 A standard boxplot with annotation](" \l "Session5_Figure13)

### Figure 24 Boxplot of batch of 20 television prices

The horizontal scale is marked from just before 100 to 275 in intervals of 25 units. The scale is labelled pounds sterling. The boxplot shows that the lower extreme is less than 100. The whisker leads to the lower quartile, at the start of the box. This occurs just past 125. The box contains a vertical line, indicating the position of the median. This occurs at 150. The box ends at the upper quartile, which occurs just after 175. The right-hand whisker ends at 250. There is then a gap. Farther on, but at the same level as the whisker, is an asterisk.

[Back to - Figure 24 Boxplot of batch of 20 television prices](" \l "Session5_Figure14)

### Figure 25 Stemplot of 14 gas prices

A stemplot with 8 levels, which start at 374 and end at 381. Level 374 has three leaves, 0, 0, 3. Level 375 has no leaves. Level 376 has two leaves, 0, 7. Level 377 has one leaf, 6. Level 378 also has one leaf, 4. Level 379 has two leaves, 5, 6. Level 380 has four leaves, 1, 1, 4, 5. Level 381 has one leaf, 8. Beneath the stemplot is written n = 14, followed by 374 vertical line 0 represents 3.740 pence per kilowatt hour. The vertical line lies horizontally between the 374 and the 0.

[Back to - Figure 25 Stemplot of 14 gas prices](" \l "Session5_Figure16)

### Figure 26

A five-figure summary. The diagram forms three sides of a rectangle, with the bottom line missing. It therefore has a vertical line to the left, a horizontal line across the top and a vertical line to the right. To the left of the left vertical line is written n = 14. Towards the bottom of the line and to the right is written 3.740 and above that 3.756. Beneath the middle of the horizontal line is written 3.790. To the left of the second vertical line and level with 3.756 is written 3.802. Below that, level with 3.740, is written 3.818.

[Back to - Figure 26](" \l "Session5_Figure17)

### Figure 27 Boxplot of batch of 14 gas prices

Boxplot of batch of 14 gas prices, previously shown as Figure 17. There is a horizontal scale with an arrow head, pointing right, at the right-hand end. It is labelled pence per kilowatt hour. The scale starts just before the first marked point of 3.74 and is then marked in four intervals of 0.02, ending with 3.82. The boxplot is drawn above and parallel to the line. The first whisker starts at the lower extreme, level with 3.74.The box starts at the lower quartile, 3.756 and ends at the upper quartile, 3.802. Within the box but nearer the right-hand end is a vertical line indicating the median at 3.790. The second whisker starts at the midpoint of the right-hand end of the box and stretches to 3.818, as far as the upper extreme.

[Back to - Figure 27 Boxplot of batch of 14 gas prices](" \l "Session5_Figure18)

### Figure 28 Stemplot of ten camera prices

A stemplot with 9 levels, which are 5, 5, 6, 6, 7, 7, 8, 8, 9. The first level 5 has one leaf, 3. The second level 5 has no leaves. The first level 6 has one leaf, 0. The second level 6 has one leaf, 5. The first level 7 has three leaves, 0, 0, 4. The second level 7 has one leaf, 9. The first level 8 has one leaf, 1 and the second level 8 also has one leaf, 5. Level 9 has one leaf, 0. Beneath the stemplot is written n = 10, followed by 5 vertical line 3 represents 53 pounds sterling. The vertical line sits horizontally between the 5 and the 3.

[Back to - Figure 28 Stemplot of ten camera prices](" \l "Session5_Figure19)

### Figure 29 Boxplot of batch of ten camera prices

A boxplot with a horizontal scale with an arrow head, pointing right, at the right-hand end. It is labelled pounds sterling. The scale starts just before the first marked point of 50 and then marked in five intervals of 10, ending with 90. The boxplot is drawn above and parallel to the line. The first whisker starts with capital E subscript capital L at 53 and ends at capital Q subscript 1, the lower quartile, 65. The box ends at capital Q subscript 3, the upper quartile, 81. Near the centre of the box is a vertical line indicating the median at 72. The second whisker starts at the midpoint of the right-hand end of the box and stretches to capital E subscript capital U at 90.

[Back to - Figure 29 Boxplot of batch of ten camera prices](" \l "Session5_Figure20)

### Figure 30 Stemplot of arithmetic stores

A stemplot with 11 levels. The numbers in the stem start at 0, increase in steps of one and end at 10. At level 0 there is one leaf, 7. Level 1 has one leaf, 5. Level 2 has no leaves, while level 3 has two leaves, 3, 5. Level 4 has three leaves, 2, 2, 3. Level 5 has two leaves, 5, 8. Level 6 has three leaves, 4, 6, 8. Level 7 has five leaves, 1, 1, 6, 8, 9. Level 8 has nine leaves, 0, 1, 1, 3, 4, 5, 5, 6, 9. Level 9 has five leaves, 1, 1, 3, 5, 9. Level 10 has two leaves which are both 0. Beneath the stemplot is written n = 33, followed by 0 vertical line 7 represents a score of 7 per cent.

[Back to - Figure 30 Stemplot of arithmetic stores](" \l "Session5_Figure21)

### Uncaptioned Equation

uppercase Q sub 1 = fraction 1 over 2 end open bracket 55+58 close bracket % = 56.5 % simeq 57 % .

[Back to - Uncaptioned Equation](" \l "Session5_Equation11)

### Uncaptioned Equation

uppercase Q sub 3 = fraction 1 over 2 end open bracket 86+89 close bracket % = 87.5 % simeq 88 % .

[Back to - Uncaptioned Equation](" \l "Session5_Equation12)

### Uncaptioned Equation

uppercase Q sub 3 minus uppercase Q sub 1 = 87.5 % minus 56.5 % = 31 % .

[Back to - Uncaptioned Equation](" \l "Session5_Equation13)

### Uncaptioned Equation

uppercase Q sub 1 = pounds 229 + fraction 3 over 4 end open bracket pounds 230 minus pounds 229 close bracket = pounds 229.75 simeq pounds 230.

[Back to - Uncaptioned Equation](" \l "Session5_Equation14)

### Uncaptioned Equation

uppercase Q sub 3 = pounds 320 + fraction 1 over 4 end open bracket pounds 349 minus pounds 320 close bracket = pounds 327.25 simeq pounds 327.

[Back to - Uncaptioned Equation](" \l "Session5_Equation15)

### Uncaptioned Equation

uppercase Q sub 3 minus uppercase Q sub 1 = pounds 327.25 minus pounds 229.75 = pounds 97.5 simeq pounds 98.

[Back to - Uncaptioned Equation](" \l "Session5_Equation16)

### Figure 31 Five-figure summary of arithmetic scores

A five-figure summary. The diagram forms three sides of a rectangle, with the bottom line missing. It therefore has a vertical line to the left, a horizontal line across the top and a vertical line to the right. To the left of the left vertical line is written n = 33. Towards the bottom of the line and to its right is written 7 and above that 57. Beneath the middle of the horizontal line is written 79. To the left of the second vertical line and level with 57 is written 88. Below that and level with 7 is written 100.

[Back to - Figure 31 Five-figure summary of arithmetic scores](" \l "Session5_Figure22)

### Figure 32 Five-figure summary of television prices

A five-figure summary. The diagram forms three sides of a rectangle, with the bottom line missing. It therefore has a vertical line to the left, a horizontal line across the top and a vertical line to the right. To the left of the left vertical line is written n = 26. Towards the bottom of the line and to its right is written 170 and above that 230. Beneath the middle of the horizontal line is written 270. To the left of the second vertical line and level with 230 is written 327. Below that and level with 170 is written 699.

[Back to - Figure 32 Five-figure summary of television prices](" \l "Session5_Figure23)

### Figure 33 Boxplot of batch of 33 arithmetic scores

A boxplot of a batch of 33 arithmetic scores. There is a horizontal scale with an arrow head, pointing right, at the right-hand end. It is labelled percentage. The scale starts at 0 and is then marked in five intervals of 20, ending with 100. Above the line, and level with the boxplot is an asterisk at 7. The first whisker on the boxplot starts at 15. The capital Q subscript 1 is at about 58, capital M is at about 79 and capital Q subscript 3 is at about 86. The second whisker stretches to the upper extreme at 100.

[Back to - Figure 33 Boxplot of batch of 33 arithmetic scores](" \l "Session5_Figure24)

### Figure 34 Boxplot of batch of 26 television prices

A boxplot of a batch of 26 television prices. There is a horizontal scale with an arrow head, pointing right, at the right-hand end. It is labelled pounds sterling. The scale starts at 100 and is then marked in six intervals of 100, ending with 700. The first whisker on the boxplot starts at 170. The box starts at capital Q subscript 1, which is about 230 and capital M is at about 270. The box ends at capital Q subscript 3 at about 320. The second whisker ends at 429. There are two asterisks. One is at 649 and the other is at 699.

[Back to - Figure 34 Boxplot of batch of 26 television prices](" \l "Session5_Figure25)

### Figure 35 A chained index

A chained index. There are two rows. The first is labelled Index value. There is then the number 100, followed by 7 sets of 3 dots. Above the 100 is an arrow, leading to the first set of dots. Then from that set is another arrow linking it to the next set, and so on to the end of the line. This gives 7 arrows in total. The next row is labelled Year. There is then a chain of 7 oval shapes, each overlapping slightly with the next one. The oval under the 100 in the first row contains the year 2007, the next one, under the first set of dots, contains 2008. The pattern continues in this way, the year increasing by 1 each time. The last entry, under the sixth set of dots, contains 2013, though it is clear that the numbers can continue further. Under the table is a cloud containing the words ‘2007 is the base year’. From it an arrow points to 2007 in the first link of the chain.

[Back to - Figure 35 A chained index](" \l "Session6_Figure1)

### Uncaptioned Equation

open bracket value of the index in 2007 comma which is taken as 100 close bracket times fraction gas price in 2008 over gas price in 2007 end .

[Back to - Uncaptioned Equation](" \l "Session6_Equation1)

### Uncaptioned Equation

open bracket value of the index in 2007 comma which is 100 close bracket times open bracket electricity price ratio for 2008 relative to 2007 close bracket = 100 times 1.145 = 114.5.

[Back to - Uncaptioned Equation](" \l "Session6_Equation2)

### Uncaptioned Equation

quantity used = fraction expenditure over price end .

[Back to - Uncaptioned Equation](" \l "Session6_Equation3)

### Uncaptioned Equation

fraction 9298 over 24 end simeq 387.4.

[Back to - Uncaptioned Equation](" \l "Session6_Equation4)

### Uncaptioned Equation

pounds open bracket 11234.6 + 3671.4 close bracket = pounds 14906.0.

[Back to - Uncaptioned Equation](" \l "Session6_Equation5)

### Uncaptioned Equation

fraction 14906 .0 over 12503 end simeq 1.192.

[Back to - Uncaptioned Equation](" \l "Session6_Equation6)

### Uncaptioned Equation

2008 index = 100 times 1.192 = 119.2.

[Back to - Uncaptioned Equation](" \l "Session6_Equation7)

### Uncaptioned Equation

fraction open bracket 1 .208 times 9298 close bracket + open bracket 1 .145 times 3205 close bracket over 9298 +3205 end = fraction 14901 .709 over 12503 end simeq 1.192 comma

[Back to - Uncaptioned Equation](" \l "Session6_Equation8)

### Uncaptioned Equation

fraction 30 over 29 end simeq 1.034.

[Back to - Uncaptioned Equation](" \l "Session6_Equation9)

### Uncaptioned Equation

fraction 98 over 87 end simeq 1.126.

[Back to - Uncaptioned Equation](" \l "Session6_Equation10)

### Uncaptioned Equation

fraction open bracket 1 .034 times 8145 close bracket + open bracket 1 .126 times 2991 close bracket over 8145 +2991 end = fraction 11789 .796 over 11136 end simeq 1.059.

[Back to - Uncaptioned Equation](" \l "Session6_Equation11)

### Uncaptioned Equation

fraction open bracket 1 .034 times 23733 close bracket + open bracket 1 .126 times 2275 close bracket over 23733 +2275 end = fraction 27101 .572 over 26008 end simeq 1.042.

[Back to - Uncaptioned Equation](" \l "Session6_Equation12)

### Uncaptioned Equation

119.2 times 1.059 simeq 126.2.

[Back to - Uncaptioned Equation](" \l "Session6_Equation13)

### Figure 36 Determining a chained price index

Determining a chained price index. There are 3 rows. The first is labelled Price ratios, the second is labelled Index and the third is labelled Base year. The row headed Index contains three numbers. These are 100, 119.2 and 126.2. Above these, the row labelled Price ratios contains two arrows. The first goes from the index number 100 to the index number 119.2. Above the arrow is written times 1.192. The second arrow goes from the index number 119.2 to the index number 126.2. Above the arrow is written times 1.059. The last row has three ovals each containing a number. The first contains 2007, below the index number 100. The second contains 2008, below the index number 119.2. The third and last contains 2009 below the index number 126.2.

[Back to - Figure 36 Determining a chained price index](" \l "Session6_Figure2)

### Uncaptioned Equation

fraction price that year over price previous year end .

[Back to - Uncaptioned Equation](" \l "Session6_Equation14)

### Uncaptioned Equation

value of index for previous year times all minus commodities price ratio .

[Back to - Uncaptioned Equation](" \l "Session6_Equation15)

### Uncaptioned Equation

fraction 28 over 30 end simeq 0.933.

[Back to - Uncaptioned Equation](" \l "Session6_Equation16)

### Uncaptioned Equation

fraction 88 over 98 end simeq 0.898.

[Back to - Uncaptioned Equation](" \l "Session6_Equation17)

### Uncaptioned Equation

fraction open bracket 0 .933 times 23733 close bracket + open bracket 0 .898 times 2275 close bracket over 23733 +2275 end = fraction 24185 .839 over 26008 end simeq 0.930.

[Back to - Uncaptioned Equation](" \l "Session6_Equation18)

### Uncaptioned Equation

126.2 times 0.930 simeq 117.4.

[Back to - Uncaptioned Equation](" \l "Session6_Equation19)

### Uncaptioned Equation

fraction 30 over 28 end simeq 1.071.

[Back to - Uncaptioned Equation](" \l "Session6_Equation20)

### Uncaptioned Equation

fraction 86 over 88 end simeq 0.977.

[Back to - Uncaptioned Equation](" \l "Session6_Equation21)

### Uncaptioned Equation

fraction open bracket 1 .071 times 23969 close bracket + open bracket 0 .977 times 2920 close bracket over 23969 +2920 end = fraction 28523 .639 over 26889 end simeq 1.061.

[Back to - Uncaptioned Equation](" \l "Session6_Equation22)

### Uncaptioned Equation

117.4 times 1.061 simeq 124.6.

[Back to - Uncaptioned Equation](" \l "Session6_Equation23)

### Uncaptioned Equation

RPI for Nov. 2011 = RPI for Jan. 2011 times open bracket price ratio for Nov. 2011 relative to Jan. 2011 close bracket .

[Back to - Uncaptioned Equation](" \l "Session7_Equation1)

### Uncaptioned Equation

CPI for Nov. 2011 = CPI for Dec. 2010 times open bracket price ratio for Nov. 2011 relative to Dec. 2010 close bracket .

[Back to - Uncaptioned Equation](" \l "Session7_Equation2)

### Figure 37 Structure of the RPI in 2012 (based on data from the Office for National Statistics)

Two concentric circles showing the structure of the Retail Prices Index in 2012, based on data from www.ons.gov.uk. The inner circle shows 5 sectors, representing the 5 fundamental groups of goods and services, each with their own colour. The biggest sector, making an angle of about 150 degrees at the centre, is coloured green and represents Housing and household expenditure. Reading round the circle clockwise from that sector there is a small sector, making an angle of about 30 degrees, coloured red and labelled Personal expenditure. The next sector, making an angle of approximately 90 degrees at the centre, is coloured yellow and labelled Travel and leisure. The next sector, making an angle of approximately 60 degrees at the centre, and coloured dark blue, is labelled Food and catering. The last sector, making an angle of about 30 degrees at the centre is labelled Alcohol and tobacco. The outer circle forms a concentric ring round the inner circle. Each of the sectors in the inner circle is subdivided in the ring, to show a breakdown of the relevant type of expenditure. The colour coding continues, but the colours in the outer ring are paler than those in the inner circle.

Housing and household expenditure is divided into four. These are 1 housing, 2 fuel and light, 3 household goods and 4 household services, with housing accounting for about half the expenditure in this sector. Personal expenditure is split into two approximately equal parts, 1 clothing and footwear and 2 personal goods and services. Travel and leisure is split into four parts. The first and largest of these, accounting for about half the expenditure in this sector, is labelled motoring expenditure. The second is fares and other travel costs and the third is leisure goods, which together amount to about the same expenditure as the fourth part, leisure services. ‘Food and catering’ is divided into two, separating food from catering, with the part labelled catering being less than half that of food. Alcohol and tobacco is split between the two, with alcohol being almost twice the size of tobacco.

[Back to - Figure 37 Structure of the RPI in 2012 (based on data from the Office for National Statistics)](" \l "Session7_Figure1)

### Uncaptioned Equation

fraction 1 over 10 end times pounds 540 = pounds 54 .

[Back to - Uncaptioned Equation](" \l "Session7_Equation3)

### Uncaptioned Equation

fraction 2 over 5 end times pounds 540 = pounds 216 simeq pounds 220 .

[Back to - Uncaptioned Equation](" \l "Session7_Equation4)

### Uncaptioned Equation

fraction 1 over 4 end times pounds 540 = pounds 135 simeq pounds 140 .

[Back to - Uncaptioned Equation](" \l "Session7_Equation5)

### Figure 38 A checklist for one household’s average monthly expenditure

A checklist for one household’s average monthly expenditure. An opportunity to compare your monthly expenditure with that of a two-person household. The table consists of the RPI groups and subgroups in the first column. There are then 3 columns under a general heading ‘Expenditure and weights’. The columns are headed ‘Expenditure 2012 in pounds sterling’, ‘Group totals in pounds sterling’ and ‘Group weights’. There is then a vertical line and there are three more columns under the general heading ‘Your expenditure and weights’. These 3 columns are also headed ‘Expenditure 2012 in pounds sterling’, ‘Group totals in pounds sterling’ and ‘Group weights’ and there are lines to indicate where entries should be made. However, there are no entries in these three columns. They have been left blank for you to input your own data.

The entries for the two-person data are as follows. Under the group heading Food and catering there are 3 subgroups. At home has an expenditure of 370. Canteen, snacks and takeaways has an expenditure of 80 and restaurant meals has an expenditure of 20. Together these give a group total of 470, which is written in the next column, on the row below. Level with this and under the column headed Group weights is written 266.

In a similar way, under the group heading Alcohol and tobacco there are 2 subgroups. Alcoholic drink has an expenditure of 8 and cigarettes and tobacco has no expenditure. The group total is therefore 8 and the weight is given as 5.

Under the group heading Housing and household expenditure there are 10 subgroups. Mortgage interest forward slash rent has an expenditure of 82. Council tax has an expenditure of 95. Water charges have an expenditure of 47 and household insurance has an expenditure of 29. Repairs, maintenance or DIY has an expenditure of 40. Gas, electricity, coal or oil bills has an expenditure of 210.

Household goods, which includes furniture, appliances and consumables etc has an expenditure of 70. Telephone and internet cost 20. There was no expenditure for school and university fees or for pet care. In total, this group Housing and household expenditure had a group total of 593 and a group weight of 336.

Under the group heading Personal expenditure clothing and footwear had an expenditure of 45 and other, which included hairdressing, chemists’ goods etc, had expenditure of 10. Together they have a group total of 55 and a group weight of 31.

There are 8 subgroups under the group Travel and leisure. Motoring, which includes purchase, maintenance, petrol, tax and insurance, had expenditure of 210. Fares had expenditure of 200. Books, newspapers and magazines had expenditure of 80. Audio-visual equipment, CDs etc had expenditure of 15. Toys, photographic and sports goods had expenditure of 3, while TV purchase or rental and licence had no expenditure. Cinema, theatre etc had expenditure of 30 and holidays had expenditure of 100. Together these give a group total of 638 and a weight of 362.

Under the group totals is a total for all the groups, which is 1764 and under the group weights the total is 1000.

[Back to - Figure 38 A checklist for one household’s average monthly expenditure](" \l "Session7_Figure2)

### Uncaptioned Equation

470 times fraction 1000 over 1764 end simeq 266.

[Back to - Uncaptioned Equation](" \l "Session7_Equation6)

### Uncaptioned Equation

fraction 470 over 1764 end simeq 0.266.

[Back to - Uncaptioned Equation](" \l "Session7_Equation7)

### Uncaptioned Equation

0.266 times 1000=266.

[Back to - Uncaptioned Equation](" \l "Session7_Equation8)

### Uncaptioned Equation

fraction price in November 2011 over price in January 2011 end .

[Back to - Uncaptioned Equation](" \l "Session7_Equation9)

### Uncaptioned Equation

RPI for month x = open bracket RPI for previous January close bracket times open bracket all minus item price ratio for month x close bracket

[Back to - Uncaptioned Equation](" \l "Session7_Equation10)

### Uncaptioned Equation

all minus item price ratio = fraction sum of products open bracket price ratio times weight close bracket over sum of weights end = fraction sum r w over sum w end .

[Back to - Uncaptioned Equation](" \l "Session7_Equation11)

### Uncaptioned Equation

fraction 1041 .624 over 1000 end = 1.041624.

[Back to - Uncaptioned Equation](" \l "Session7_Equation12)

### Uncaptioned Equation

RPI for Nov. 2011 = RPI for Jan. 2011 times open bracket all minus item price ratio for Nov. 2011 close bracket = 229.0 times 1.041624 = 238.531896 simeq 238.5.

[Back to - Uncaptioned Equation](" \l "Session7_Equation13)

### Uncaptioned Equation

fraction sum r w over sum w end .

[Back to - Uncaptioned Equation](" \l "Session7_Equation14)

### Uncaptioned Equation

RPI for month x = RPI for previous January times open bracket all minus item price ratio for month x close bracket .

[Back to - Uncaptioned Equation](" \l "Session7_Equation15)

### Uncaptioned Equation

sum open bracket w close bracket = comma sum of products open bracket r w close bracket = comma all minus item price ratio = fraction sum of products open bracket r w close bracket over sum open bracket w close bracket end = comma value of RPI in July 2011 = .

[Back to - Uncaptioned Equation](" \l "Session7_Equation16)

### Uncaptioned Equation

sum open bracket w close bracket =1000 comma sum of products open bracket r w close bracket =1024.608 comma

[Back to - Uncaptioned Equation](" \l "Session7_Equation17)

### Uncaptioned Equation

all minus item price ratio = fraction sum of products open bracket r w close bracket over sum open bracket w close bracket end = fraction 1024 .608 over 1000 end =1.024608 comma

[Back to - Uncaptioned Equation](" \l "Session7_Equation18)

### Uncaptioned Equation

value of RPI in July 2011 =229.0 times 1.024608 =234.635232 simeq 234.6.

[Back to - Uncaptioned Equation](" \l "Session7_Equation19)

### Uncaptioned Equation

fraction value of CPI in February 2012 over value of CPI in February 2011 end = fraction 121 .8 over 117 .8 end simeq 1.034.

[Back to - Uncaptioned Equation](" \l "Session7_Equation20)

### Uncaptioned Equation

fraction value of RPI in February 2012 over value of RPI in February 2011 end = fraction 239 .9 over 231 .3 end simeq 1.037 comma

[Back to - Uncaptioned Equation](" \l "Session7_Equation21)

### Uncaptioned Equation

pounds 450 times fraction 239 .9 over 231 .3 end open bracket i.e. pounds 466.73 close bracket per month

[Back to - Uncaptioned Equation](" \l "Session7_Equation22)

### Uncaptioned Equation

pounds 120 times fraction 121 .2 over 115 .6 end simeq pounds 125.81.

[Back to - Uncaptioned Equation](" \l "Session7_Equation23)

### Uncaptioned Equation

fraction value of RPI at date B over value of RPI at date A end times 100.

[Back to - Uncaptioned Equation](" \l "Session7_Equation24)

### Uncaptioned Equation

fraction 231 .3 over 239 .9 end times 100 p = 96.41517 p .

[Back to - Uncaptioned Equation](" \l "Session7_Equation25)

### Uncaptioned Equation

fraction 100 over 239 .9 end times 100 p .

[Back to - Uncaptioned Equation](" \l "Session7_Equation26)

### Uncaptioned Equation

fraction 223 .6 over 212 .8 end simeq 1.051 comma

[Back to - Uncaptioned Equation](" \l "Session7_Equation27)

### Uncaptioned Equation

fraction 212 .8 over 223 .6 end times 100 p simeq 95 p .

[Back to - Uncaptioned Equation](" \l "Session7_Equation28)

### Uncaptioned Equation

fraction 238 .0 over 225 .8 end simeq 1.054 comma

[Back to - Uncaptioned Equation](" \l "Session7_Equation29)

### Uncaptioned Equation

fraction 225 .8 over 238 .0 end times 100 p simeq 95 p .

[Back to - Uncaptioned Equation](" \l "Session7_Equation30)

### Uncaptioned Equation

fraction 232 .5 over 220 .7 end simeq 1.053 comma

[Back to - Uncaptioned Equation](" \l "Session7_Equation31)

### Uncaptioned Equation

fraction 220 .7 over 232 .5 end times 100 p simeq 95 p .

[Back to - Uncaptioned Equation](" \l "Session7_Equation32)

### Uncaptioned Equation

sum w =1000 comma sum r w=1007.760 comma

[Back to - Uncaptioned Equation](" \l "Session7_Equation33)

### Uncaptioned Equation

all minus item price ratio = fraction sum r w over sum w end = fraction 1007 .760 over 1000 end =1.007760 comma

[Back to - Uncaptioned Equation](" \l "Session7_Equation34)

### Uncaptioned Equation

value of RPI in February 2012 =238.0 times 1.007760 =239.84688 simeq 239.8.

[Back to - Uncaptioned Equation](" \l "Session7_Equation35)

### Uncaptioned Equation

fraction 225 .8 over 216 .0 end simeq 1.045 comma

[Back to - Uncaptioned Equation](" \l "Session7_Equation36)

### Uncaptioned Equation

fraction 216 .0 over 225 .8 end times 100 p simeq 96 p .

[Back to - Uncaptioned Equation](" \l "Session7_Equation37)

### Uncaptioned Equation

fraction 229 .0 over 217 .9 end simeq 1.051 comma

[Back to - Uncaptioned Equation](" \l "Session7_Equation38)

### Uncaptioned Equation

fraction 217 .9 over 229 .0 end times 100 p simeq 95 p .

[Back to - Uncaptioned Equation](" \l "Session7_Equation39)

### Uncaptioned Equation

pounds 800 times fraction 234 .4 over 222 .8 end simeq pounds 842 per month .

[Back to - Uncaptioned Equation](" \l "Session7_Equation40)

# Screencast 1 Effects on the median and mean when data points change

## Transcript

INSTRUCTOR

In this screencast, I’m going to talk about calculating the mean and the median from the stemplot and showing how the mean and the median change when some of the data changes. So I’m going to start off with a stemplot. And the stemplot I’ve got here is a stemplot of the prices of the small flat screen televisions that are shown in Activity 1 in Subsection 1.2 of Unit 2.

And the first thing I’m going to do is calculate the median. And we note first that we’ve got 20 data points in our batch. And so the median is the average of the 10th and 11th largest values.

So it’s just a question of finding out from the stemplot what the 10th and the 11th largest values are. And we can do that by counting down from the top value in our stemplot. That’s 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11.

But we could also have counted from the bottom of the stemplot. Again, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11. And notice, it’s come at the same two numbers, as it should. So for these data, the median is 150 plus 150 over 2, which is £150, taking into account this key for this stemplot.

Now the mean is just the sum of all the numbers divided by the number of numbers. And going through stemplot, we can write down what all the numbers are. So the lowest number is 90. The next one is 100. Next one is 120. And so on, so forth, until we get to 240 and 250 and 270.

And that’s all divided by the number of numbers, which is 20. And the sum on top of the fraction happens to be 3240. That’s divided by 20, and that comes to £162. So the mean of the television prices is £162, and the median is £150. So the mean is £12 bigger than the median.

Now, what happens if some of the data changes? For example, what happens if a couple of the prices for the televisions goes up? And in particular, what happens if the prices of the most expensive televisions go up? So instead of having a television that costs £250 and a television that costs £270, we actually had a couple of televisions that cost £350 and £400? What difference does this make to the median and the mean?

Well, notice that the two values we work the median out from haven’t changed. So we can just write that down immediately. The median is £150, just as it was before.

So what about the mean? Well again, we’ve got to work out the sum of all the data points and divide by the number of numbers. For most of the numbers, they haven’t changed. So most the numbers in the sum don’t change. But the last two have – so instead of 250, we’ve now got 350, and instead of 270, we’ve got 400. Again, that’s divided by 20. That’s equal to 3470 over 20, which equals 173.5, or we can say that’s £174 to the nearest pound.

So what we notice here that when the data is changed, the median has stayed the same. We say that the median is a resistant measure. It’s been resistant to a change in the data. On the other hand, the mean is bigger. We say that the mean is a sensitive measure. It has been sensitive to changes in the data.

[Back to - Screencast 1 Effects on the median and mean when data points change](" \l "Session3_MediaContent1)

# Screencast 2 Calculating a weighted mean

## Transcript

INSTRUCTOR

OK, here’s an example that you’ve seen in the units about the price of two batches of small TVs. The first batch, that’s Batch A, that’s got a mean of 119, that’s £119 actually, and size 7, there are 7 TVs in that batch. Batch B, the second batch, has got a mean of £185, good bit more expensive, and the size of that batch is 13. And what we’re asked to do here is to find the mean of the combined batch. OK, so how are we going to do that?

Well, there’s a formula in the unit which I’ll write down in a minute. But I want to build it up a bit at a time to show you where the formula comes from, and the formula looks like this. We’ve got x bar, that means mean. And c, that’s because it’s the mean of the combined batch.

Now, we usually calculate a mean by calculating the total of the values in the batch, and dividing it by the size of the batch. And essentially, that’s all that’s happening here – we just have to do it in a bit of a complicated way. So how does that work?

Well, first of all we need the sum of the values in Batch A. Now, we’ve got the mean, we haven’t got the sum. But that mean, £119, is the sum divided by the size, which is 7. So you can get back to the sum by multiplying the mean by the size. So that is it’s the mean of the values in Batch A, x bar A, times the size of Batch A.

Anyway, that deals with the TV prices in Batch A, and now we’ve got to add on the TV prices in Batch B. So you do the same trick. You add on x bar B, the mean of the values in Batch B, times the size, nB, of Batch B, and that gives you the sum part of this mean calculation.

Then you’ve got to divide that by the size, so there’s the division sign. And the size of the combined batches is the total number of TVs in this combined batch. So it’s the number you’ve got in Batch A, plus the number you’ve got in Batch B – nA plus nB. So that’s the formula, and that’s the same one as in the unit.

The next step is to convert the numbers we have to make clear how the notation we got in this formula fits with the numbers we’ve got. So let’s just translate the mean and the size of the batches into the notation we need for the formula. So for Batch A, the mean’s 119 – that is, in this notation we got x bar A is 119, and nA the size of Batch A is 7 because there’s 7 TVs. And then we do the same with Batch B, x bar B is 185 to match what it says over the left there, and nB is 13.

So we’ve got the formula here, and we’ve got the values we need to put in the formula. So now we’ve just got to go ahead and put the values in and do the arithmetic. So the first step, let’s work out what we’ve got to put on the top. Well, on the top of this expression we’ve got x bar A times nA, and that is 119, that’s x bar A, times nA, which is 7, and I’ve put the multiplication sign in there to make it clear what’s going on.

And then the other aspect of this is we’re going to add something on in a minute, but we need to do the multiplication first. That is, we need to do this multiplication before we get to this plus. So to make that absolutely clear, we put this in some brackets to indicate that’s what we’re going to do first.

So we do that, and then we add on what we get from the B batch, that is x bar B times nB. And again, we’re going to put that in brackets, so that’s 185 times 13. Close the brackets, and then divide the whole thing by nA plus nB, which is 7 plus 13. So that’s put in the values we got into the formula we had. And now it really is just a case of bashing through the arithmetic.

So I’ve got my trusty calculator here – 119 times 7, that comes to 833. And then we’ve got to add on what we get from this bit here – 185 times 13, and that comes to 2,405. So that’s the top of the thing we’ve got to calculate. We’ve got to divide by the bottom, 7 plus 13, and doing that all in my head it’s 20 and we’re nearly there.

What we’ve got now is we added 833 to 2405, that comes to 3238, and we’ve got to divide that by 20. And calculating that, that comes to 161.9. Well, we’re nearly there now. It looks as if it’s calculated the mean of the combined batch but there’s a final step we got to do, and that’s to relate it back to what’s really going on here.

If you look at these means up here, or they were here originally, they’re given in whole pounds. That is, these means are rounded to the whole pound. Well, it might have been coincidence that they actually come to whole numbers, but what’s really happened is that it’s not appropriate to express the mean to the full accuracy you get from your calculator, because that’s just too much accuracy. The data don’t support it.

So what we got to do is we got to round the means to an appropriate value. And in this case, the original means were rounded to the nearest pound, so our combined mean also needs to be rounded to the nearest pound. So let’s have look at what it is – it’s 161.9. And to round that to the nearest pound, is it going to go up or down?

Well, look at the last figure here, the one we’re going to take away in the rounding – it’s 9, that’s bigger than 5. So to round it to the nearest pound, we have to round it up to 162. That is, we finish off by saying this is £162 rounded to the nearest pound, and that really is the finish of the calculation.

But before we leave this, I just want to point out one thing to you. Look at this value here of 162 and compare it with the means of the two batches we had – Batch A: 119, Batch B: 185. The 162 we’ve got here is a good bit nearer the mean of Batch B than it is to the mean of Batch A. And that’s because Batch B is bigger. It’s got more TVs in it – it’s got 13, and this has only got 7.

So what’s happened is that the Batch B has dominated in the calculation because there’s just more TVs in Batch B. There are 13 of them compared to 7. So Batch B has had more influence on this number than Batch A has.

Now in the unit, it explains that this formula here that we had for a mean of a combined batch is a kind of weighted mean. And the unit also describes that there are various rules that weighted means follow, various properties they have. And one of the properties they have is that the value of the weighted mean of two numbers is closest to the value that had the largest weight.

Now what’s playing the role of weights here is the batch sizes, nA and nB. nB is bigger than nA, so this rule tells us that the combined value, this weighted mean, is going to be nearer to this, the mean of Batch B, than it is to this. And indeed that’s what we find – 162 is nearer to 185 than it is to 119.

So I’ll just record the fact that we found that. This is closer to the mean x bar B, the mean of Batch B, as Batch B is bigger. And that’s that.

[Back to - Screencast 2 Calculating a weighted mean](" \l "Session4_MediaContent1)

# Screencast 3 Calculating quartiles

## Transcript

INSTRUCTOR

Here’s another batch of data from the unit. This time, it’s the prices of a particular model of digital camera. And there are the prices along the top in pounds. And what we’re asked to do is to find the lower and upper quartiles of this batch of data.

So usually the first thing you have to do when you’re finding the median or the lower and upper quartiles is to put the prices in order. But if you have a look up at the top there, luckily, somebody’s already done that for us. They’re in order.

So we can bash straight on with the next step, which is actually finding the quartiles. Now, what I’m going to do to begin with is to use the notation that we use in defining what the quartiles are, just so you can relate what I’m doing to what’s in the unit on that. Remember how that works. We take the first number, the smallest number in the batch, which in this case is 53, right over the left there.

And we write that as x, within brackets, 1. The 1 in brackets means it’s the smallest value in the batch. And the next one, which is 60, that’s x, brackets, 2. And then you just carry on like that.

And the last one is x10. That means there’s 10 values in this bunch. We need to know that.

And in fact, that’s the next thing we need to worry about. There are 10 observations, 10 camera prices. So in the notation, we call that n. n’s 10.

Now, remember what you’ve got to do to work out this. We want for the quartiles a number that’s 1/4 of the way along the batch when we’ve arranged it in order and another number that’s 3/4 of the way along the batch. But of course, it’s a bit more complicated than that. You don’t look at 1/4 of n. You look at n plus 1 divided by 4.

And that gives you the position of the lower quartiles. n plus 1 divided by 4 is 11/4, which is 2 and 3/4. And that defines where the position of the lower quartile is in a way that I’ll come to.

And to get the upper quartile – well, this time, you’ve got to take 3/4 of n plus 1. That is 3 times n plus 1, over 4. And that comes to n plus 1 is 11.

So it’s 33 – 3 times 11 – over 4. And if you work that out, that’s 8 and 1/4. And that defines the position of the upper quartile.

OK, so moving on, then, Q1 is the lower quartile. And it’s defined by this number here, this 2 and 3/4 number. And what that tells us is it’s between the second number in order and the third number. And it’s 3/4 of the way between them.

So I’ll just write down what that is. Q1 is 3/4 of the way from x2 to x3. And you can work that out just in arithmetic – but it’s probably easier if we have a look and see what it looks like graphically.

Here is a bit of a number line. x2 is going to be there. And x2 is 60.

And x3’s going to be up here somewhere. And x3, if we look up here, is 65. So that one’s 65.

And what we want is a number that’s 3/4 of the way from 60 to 65. Well, that’s halfway there. That’s going to be 62 and 1/2. And 3/4 of the way is about here somewhere. And that’s going to be 63.75.

OK, but how do we write that down in algebra? The way to do it is like this. We say you start at x2, which is 60. And then you go 3/4 of the way from x2 to x3.

Well, how far is it from x2 to x3? Well, again, if you look on this, it’s 65 to 60. That is 5.

And you can calculate that as 65 minus 60. So it’s 3/4 of 65 minus 60. So that’s 60 plus – we’ve got 3 times 5 over 4. That is 15/4. And if you work that out on your calculator or something, it comes to 63.75.

So we’ve worked out the lower quartile. It’s 63.75. And luckily, that’s the same as I showed you over here.

So now we’ve got to go and find the upper quartile. Remember, we denote that as Q3. And if you look at the thing over here, that says it’s at position 8 and 1/4, which means it’s 1/4 of the way from the eighth one to the ninth one. That is, it is 1/4 of the way from x8 to x9.

And again, it’s helpful to draw what’s going on here. So here is a number line. And here we’ve got x8, which is 81. And here we’ve got x9, which is 85.

And again, we want the number that’s 1/4 of the way along, bit different. So here’s halfway. That’s going to be 83. And here is 1/4 of the way, halfway between 81 and 83. That’s 82.

So it’s going to be 82. Let’s just see how we can write that down as a bit of algebra. So again, you start from x8, which is 81. And then you go 1/4 of the way up to x9.

Now this time, the distance between x8 and x9 is 85 – that’s x9 – minus 81, which is x8. So the arithmetic’s actually a lot easier in this case. That’s 81 plus 1/4 of 4. And that just comes to 81 plus 1, which is 82. Again, it’s the same.

So just one final step to do here. We’ve got to write down the answers in an appropriate way. Now, what we have to do is round them to an appropriate accuracy. That’s all we haven’t done.

And what we’ve got is that Q1 – what accuracy do we use? Well, the usual thing we do is to round quartiles to the accuracy of the original data that we had. And the original data we had up at the top there is in whole pounds. It’s rounded to whole pounds. So we need the quartiles in whole pounds, as well.

And if you look at this one, it’s not in whole pounds. It’s 63.75. So we’ve got to round it to the nearest pound. The bit in pence, essentially, here that we’re going to get rid of in the rounding is bigger than 1/2.

So we’ve got to round up. That is, this one is– let’s write in the pound sign. It’s £64 to the nearest pound.

And what about Q3? Well, you’ve got to think about rounding that, as well. But actually, you needn’t bother. It’s already a whole number of pounds, so no rounding involved. We can just write down it’s 82. And that’s that.

[Back to - Screencast 3 Calculating quartiles](" \l "Session5_MediaContent1)

# Screencast 4 Interpreting a boxplot

## Transcript

INSTRUCTOR

In this screencast, I’m going to talk about interpreting a boxplot. And what we have here is an example of a boxplot. And it happens to be Figure 18 from Subsection 3.3 of Unit 2. And it’s a boxplot of the small television prices. And you can see here that television prices are given in pounds, and they go from just under £100 up to £275.

The first thing on the boxplot to look at is the box itself – in particular, to look at the ends of the box. The end on the left hand side shows us where the lower quartile is. And here is about £130. The end on the right hand side shows us where the upper quartile is – Q3. And this translates to about £180. So the lower quartile is about £130, and the upper quartile is about £180.

The line in the middle of the box shows us where the median is. And this is about £150. So the price of the median small television is £150. Notice in this example, it is quite clear where the line is in the middle of the box. There are some examples where the median is the same as the lower quartile or where the median is the same as the upper quartile. And then you won’t actually see a line in the box. The line will be at one end or the other.

The other thing to notice on the boxplot are the two whiskers. So there’s a whisker on the right hand side and a whisker on the left hand side. The whisker on the right hand side shows us where the values that are high but not too high are. Similarly, on the left hand side, the whisker on the left hand side shows us where the values are low but not too low are.

And finally, notice there’s one point here marked all by itself. And this is a value that we wonder whether it’s too high. In other words, we’re marking this one out as a potential outlier. This shows all the elements that are on a boxplot. And one thing we can use these elements for is to say something about the symmetry of the data.

And one thing we can look at is where the median is relative to the two ends of the box. So here we notice that the left hand side is short relative to the right hand side. We can look at the whiskers in the same way, and notice that the whisker on the left hand side is relatively short. And the whisker on the right hand side is relatively long.

And both these observations together suggest that the data are right-skew. The data tends to be more spread out on the right hand side of the median relative to the data on the left hand side the median. And notice, in doing this, we haven’t actually taken account of the outlier. If we took the outlier into account as well, this would only emphasise more that the data are right-skew. Because this adds to the impression that the data are more spread out to the right of the median relative to the left of the median.

[Back to - Screencast 4 Interpreting a boxplot](" \l "Session5_MediaContent2)

# Screencast 5 Calculating an RPI

## Transcript

INSTRUCTOR

Here’s an example out of the unit of calculating the Retail Prices Index. And what we’re asked to do is to calculate the Retail Prices Index for a particular month, November 2011.

So the first step in doing that is to collect the prices that it’s going to be based on and the weights. Well, actually, I’m not going to ask you to go out and collect hundreds and thousands of prices or anything like that. We’re just going to use the data on prices that have been collected by the Office for National Statistics.

And I’ve got some information off their website. And, similarly, I’ve got the weights from the website as well. And we’re just going to start off at that stage and write down what they are.

So here’s the sort of table that you do these calculations in. Down the side here, we’ve got just the five top-level groups that items are divided into in the calculation of the RPI. And we’ve got a column for the price ratios. We’ll come to that in a minute. And we’ve got a column for the 2011 weights. And I’m just simply going to fill in the numbers and tell you a little bit about what they mean.

So the weight for the food and catering group is 165. And carrying down, the weight for alcohol and tobacco is 88. The weight for housing and household expenditure is 408. And the weight for personal expenditure is 82. And the weight for travel and leisure is 257.

Now there’s a row we need to fill in at the bottom of this table as usual. And that’s the row for the sum. And if you add up these numbers – you can check this yourself, if you like – they come to 1000. Actually, you might not want to check. They always come to 1000 with RPI calculations because they’re designed that way.

And just to remind you of what these things mean, when the weight for food and catering is 165, that means that out of every £1000 the average household spends, 165 of those £1000 go on food and catering. And 88 go on alcohol and tobacco, and so on. And these are derived, again, by the Office for National Statistics statisticians from a survey. They’re the weights that were used in 2011. And they’re based on what people spent their money on, essentially, in the middle of the previous year, 2010.

So what’s the next step? The next step is to calculate the price ratios. You’ve already seen in the table that there’s a column for them. And, again, I’m not going to ask you to go and work these things out for yourselves. You haven’t even got the information to do it from.

So I’m just going to put in the price ratios which I found on the Office for National Statistics website. And they’re price ratios for November 2011, the month we’re interested in, relative to the previous January, that is to January 2011. With the RPI, they’re always used relative to the previous January.

So we’ll just do that again. Let’s write in the first one. The first one for food and catering is 1.030.

And that means that, on average, people spent £1.03 in November to buy the same stuff that they would have spent £1 on in January. It’s the ratio of the price in November to the price in January. And that’s simply what it is.

And you can go in and write in the other ones the same way. So I’ll do that. Alcohol and tobacco is 1.050. So alcohol and tobacco’s gone up a bit more than food and catering had. Housing and household expenditure is 1.037. So that’s gone up somewhere between the previous two.

Personal expenditure had gone up rather a lot. It’s 1.128. And travel and leisure had gone up rather less, 1.026. And we don’t need the sum of those figures. So I’ll leave this space here blank. And those are the price ratios.

So the next step is to calculate the all-item price ratio, the all-item price ratio for November 2011 relative to the previous January, January 2011. And this is usually expressed as a formula. If you call the weights w and the price ratios r, then what you do is you take the products, the price ratio times the weight. And you add them up and you divide that by the sum of the weights. So that’s what we’re going to calculate.

And it’s just a kind of formula for a weighted average. It’s a weighted average of the price ratios weighted by these weights that we had. So let’s go ahead and do that.

This is just a matter of arithmetic. So this is the price ratio times the weight. There’s the price ratio for the food and catering group. There’s the weight. So we multiply that times that. And we write the answer in there. Well, you have to do that in a calculator or something. This comes to 169.950.

We keep all of the accuracy in the calculation. As you probably realise, we’re going to round in the end. But we do that as absolutely the last step so that we don’t lose any accuracy because of rounding that we’d done in the intermediate stage. So it’s just a matter of filling in the rest of the price ratio times weight column in the same kind of way.

So this one turns out to be 92.400. And for housing and household expenditure, it’s 423.096. And the next one’s personal expenditure, that’s 92.496. And then, finally, 263.682.

And then we do need the sum of these. That was in the formula. It’s 1041.624.

And then we just got to work out this all-item price ratio. The formula is here. Sum of rw divided by sum of w, sum of ratios times weights divided by the sum of the weights. So it’s the ratio of those two sums we’ve calculated. 1041.624 divided by 1000. And since the thing we’re dividing by is a nice round number, that’s pretty easy to do. It comes to 1.041624. And, again, that’s an awful lot of decimal places. But we keep the full accuracy because we’re going to round at the end.

So back to the last step in the calculation. And that is to actually do what we wanted, to calculate the RPI. And the RPI, Retail Prices Index, for November 2011 – what we got is the all-item price ratio. We worked that out. And that’s the kind of average amount by which prices have gone up by in November 2011 compared to the previous January.

So if we actually want the Retail Prices Index, we’ve got to multiply the Retail Prices Index for that previous January by this all-item price ratio. And that gives us the Retail Prices Index, put up by the weighted average amount that prices have gone up by. So that means what we do is take the RPI for the previous January, for January 2011, and we multiply it by the all-item price ratio for November 2011. And that’s for November 2011 relative to the previous January.

And so we just need some numbers. So we worked out the all-item price ratio on the table just a bit before. We need the RPI. And, again, you’ve got to look that up on the ONS website or something like that. It was actually 229.0. So that’s the RPI.

We multiply it by the all-item price ratio, which is 1.041624, as we calculated before. And you do that calculation, and it comes to 238.531896.

Again, we’ve kept the full accuracy. But that’s clearly not justified by the accuracy of the data. The RPIs that are published always have just one decimal place. And so we need to round this to one place of decimals.

So how are we going to do that? We’re going to leave this 5 here, because that’s the last one. But the question is, do we round it up to 6? Or do we leave it where it is at 5? You have to look at the next value, which is 3. 3 is less than 5. That is, it’s less than halfway from 0.5 to 0.6. So we round it down.

And what you end up with is 238.5 correct to one decimal place. And that, 238.5, that’s the RPI for November 2011.

[Back to - Screencast 5 Calculating an RPI](" \l "Session7_MediaContent1)