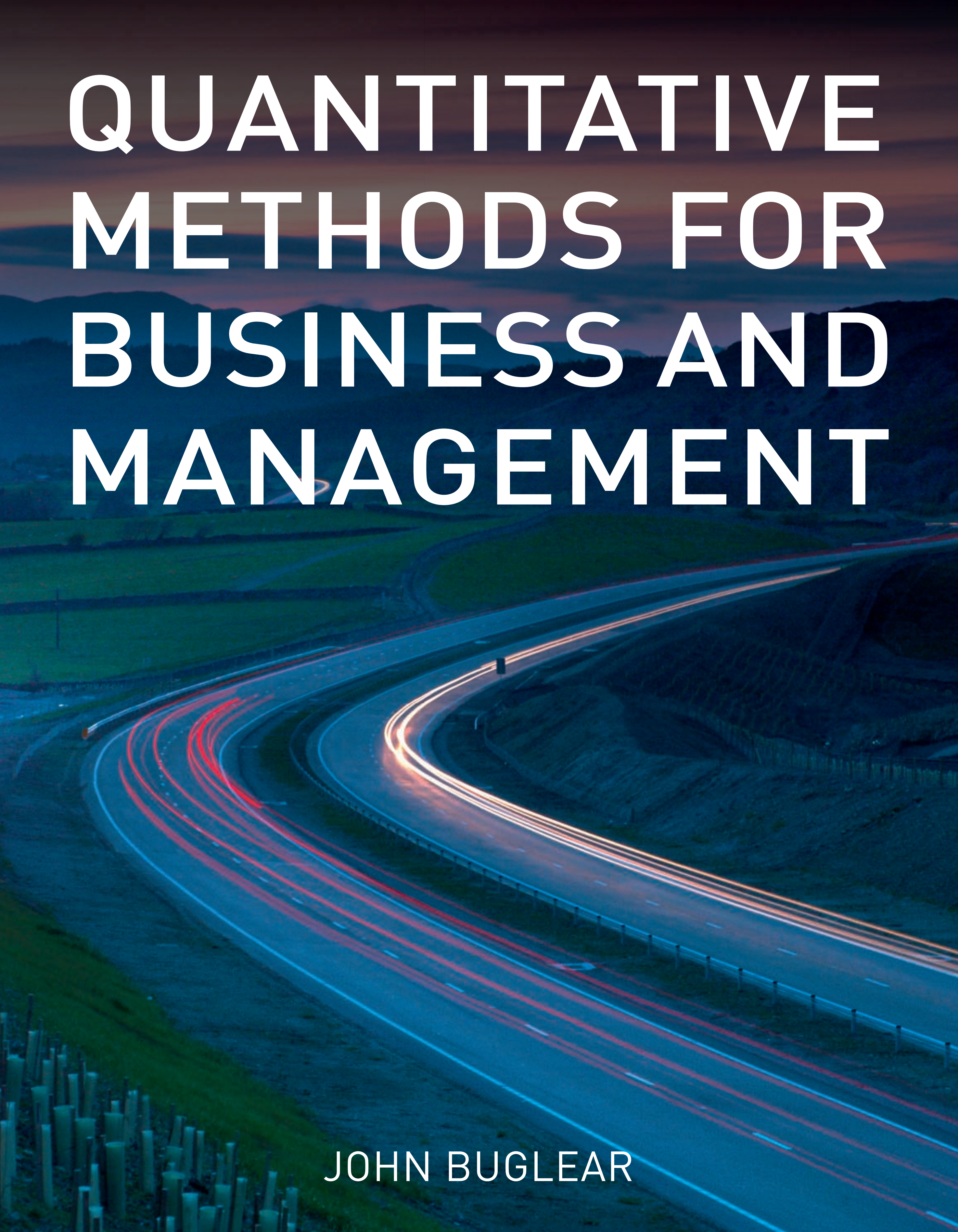
QUANTITATIVE METHODS FOR BUSINESS AND MANAGEMENT



JOHN BUGLEAR

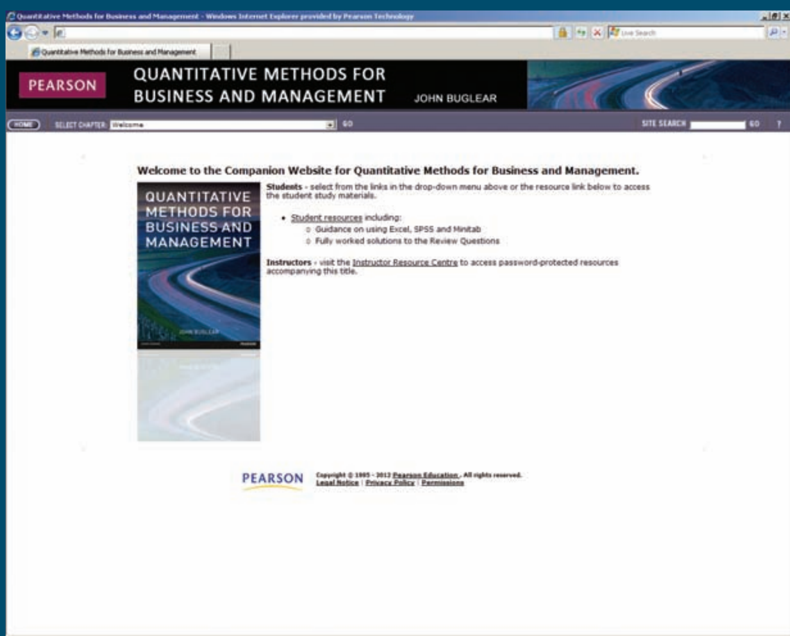
ALWAYS LEARNING

PEARSON

QUANTITATIVE METHODS FOR BUSINESS AND MANAGEMENT

Visit the *Quantitative Methods for Business and Management* Companion Website at **www.pearsoned.co.uk/buglear** to find valuable student learning material including:

- Guidance on using Excel, SPSS and Minitab
- Fully worked solutions to the Review Questions



SELF-ASSEMBLY GUIDE

Approximating a median from a grouped frequency distribution

- The general procedure is:

$$\text{Approximate median} = \text{start of MC} + \frac{(\text{median position} - \text{number of observations to MC})}{\text{frequency of MC}} * \text{width of MC}$$

- Find the median position $(n + 1)/2$. In Example 6.14 $n = 23$ so $(23 + 1)/2 = 12$.
- Find the median class (MC). This is the class that contains the observation in the median position. In Example 6.14 the median class is '£15 and under £20'.
- Find the total number of observations that are in the classes before the median class. In Example 6.14 this is the combined frequency of the first and second classes. The first has one observation and the second has five so there are six observations of less than £15.
- Find the frequency and width of the median class. In Example 6.14 these are 10 and £5 respectively.
- Enter the figures into the general procedure:

$$\text{Approximate median} = £15 + \frac{(12 - 6)}{10} * £5 = £15 + £3 = £18$$

Another way of approximating the median from a grouped frequency distribution is based on a cumulative frequency graph or a cumulative relative frequency graph of the distribution. These were introduced in section 5.3.2. The method works with either type of graph but is easier to apply with a cumulative relative frequency graph.

Plot the graph with the cumulative relative frequency scale on the vertical axis. Start from the point along this axis that represents exactly half the total cumulative relative frequency, 0.5. Draw a horizontal line across to the line depicting the cumulative relative frequency. At the point where these lines meet draw a vertical line down to the horizontal axis. The approximate value of the median is where this vertical line meets the horizontal axis.

EXAMPLE 6.15

Find the approximate value of the median monthly cost of calls data in Example 6.5 by constructing a cumulative relative frequency graph of the grouped frequency distribution in Example 6.13.

Solution

Cost (£)	Frequency	Cumulative frequency	Cumulative relative frequency
5 and under 10	1	1	0.04
10 and under 15	5	6	0.26
15 and under 20	10	16	0.70
20 and under 25	6	22	0.96
25 and under 30	1	23	1.00

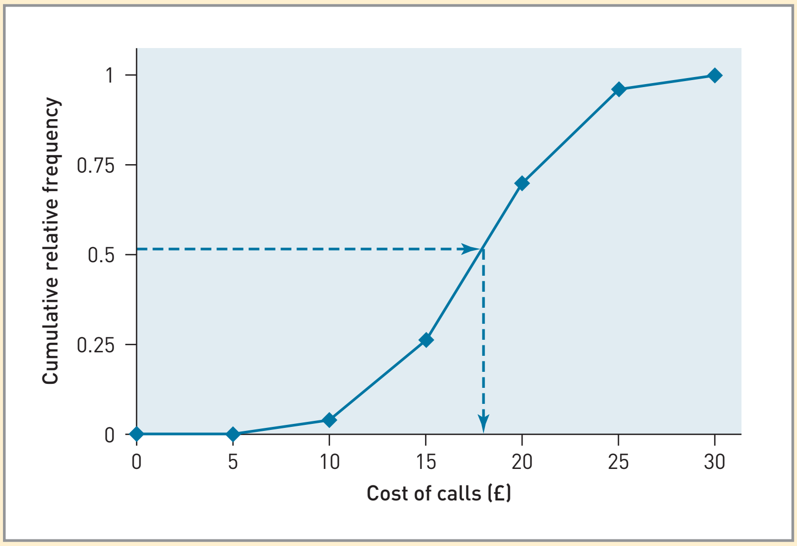


Figure 6.3 Cumulative relative frequency graph of the monthly costs of calls

The point where the vertical dotted line meets the horizontal axis is at approximately £18 on the horizontal scale (see Figure 6.3). This is the estimate of the median. From Example 6.9 we know that the actual median is £17 so our graphical approximation is quite close.

SELF-ASSEMBLY GUIDE

Approximating a median from a cumulative relative frequency graph

- Find 0.5 on the vertical axis.
- Draw a horizontal line from 0.5 on the vertical axis across to the cumulative relative frequency line.
- From the point where the two lines meet draw a vertical line down to the *x*-axis.
- The point where this vertical line meets the horizontal axis is the approximate value of the median.

To obtain an approximate value for the mean from a grouped frequency distribution we apply the same frequency-based approach as we used in Example 6.12. There we multiplied each value, x, by the number of times it occurred in the distribution, f, added up the products and divided by the total frequency of the distribution:

$$\bar{x} = \frac{\Sigma fx}{\Sigma f}$$

In Example 6.12 we worked from a frequency distribution which listed exactly how many times each value occurred. But what if we have data in a grouped frequency distribution? From this we know only the frequency of each class, not the frequency of each value. To get around this we assume that all the observations in a class take, on average, the value in the middle of the class, known as the class midpoint. The set of class midpoints is then used as the values of the variables, x, that make up the distribution. This is very unlikely to be true but the best we can do if we only have a grouped frequency distribution to work with.

EXAMPLE 6.16

Find the approximate value of the mean from the grouped frequency distribution in Example 6.13.

Solution

Cost of calls (£)	Midpoint (x)	Frequency (f)	fx
5 and under 10	7.5	1	7.5
10 and under 15	12.5	5	62.5
15 and under 20	17.5	10	175.0
20 and under 25	22.5	6	135.0
25 and under 30	27.5	1	27.5
		$\sum f = 23$	$\sum fx = 407.5$

The approximate value of the mean $= \Sigma fx/\Sigma f = 407.5/23 =$ £17.72 (to the nearest penny), which is close to the actual value in Example 6.5, £17.30 (to the nearest penny).

At this point you may find it useful to try **Review questions 6.2 and 6.5** at the end of the chapter.

6.3 Measures of spread

In section 6.2 we looked at three measures of location that in different ways tell us about the central tendency of a distribution. Similarly, there are several measures of spread that can tell us how the observations in a distribution are dispersed. They measure spread in different ways. We'll start the simplest, the range, then look at using quartiles to measure spread and finally the most important measure of spread, the standard deviation.

Measures of spread are used with measures of location to convey two key features of a distribution; where the middle is and how the observations are scattered around the middle.

6.3.1 The range

The range is the most basic measure of spread. It is simply the difference between the lowest and highest observations in a distribution:

$$\text{Range} = \text{highest observation} - \text{lowest observation}$$

In some cases the range is a sufficient way of measuring spread, especially if we need only a general idea of the spread in a set of data. Its big drawback is that it is not always a reliable enough measure. This is because it is based on just two of the observations in a distribution. It is conceivable that two distributions have the same range but the observations are spread out in very different ways.

EXAMPLE 6.17

Two employment agencies, Rabota Recruitment and Slugar Selection, each have 9 employees. The length of service that their employees have (in years) is:

Rabota	1	5	5	7	8	9	11	12	15
Slugar	1	1	5	5	8	11	11	15	15

Calculate the range of each distribution.

$$\text{Range (Rabota)} = 15 - 1 = 14 \qquad \text{Range (Slugar)} = 15 - 1 = 14$$

Plot a histogram for each distribution and use them to compare the lengths of service of the employees of the agencies.

Solution

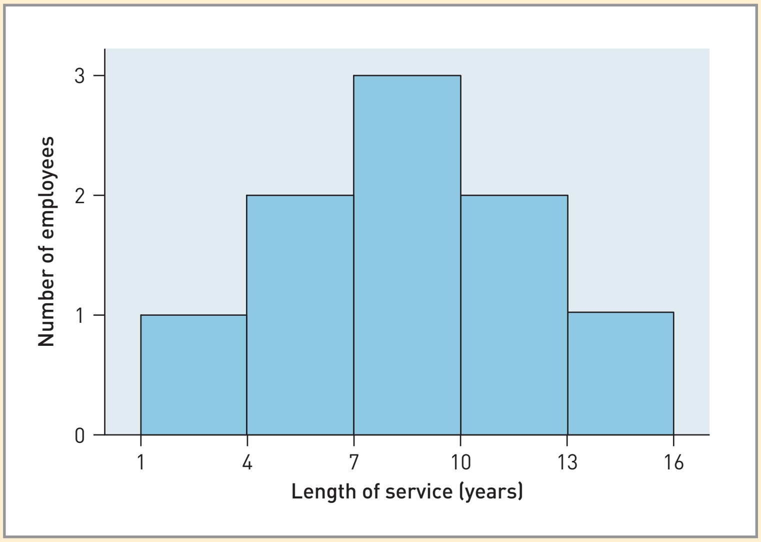


Figure 6.4 Lengths of service of Rabota staff

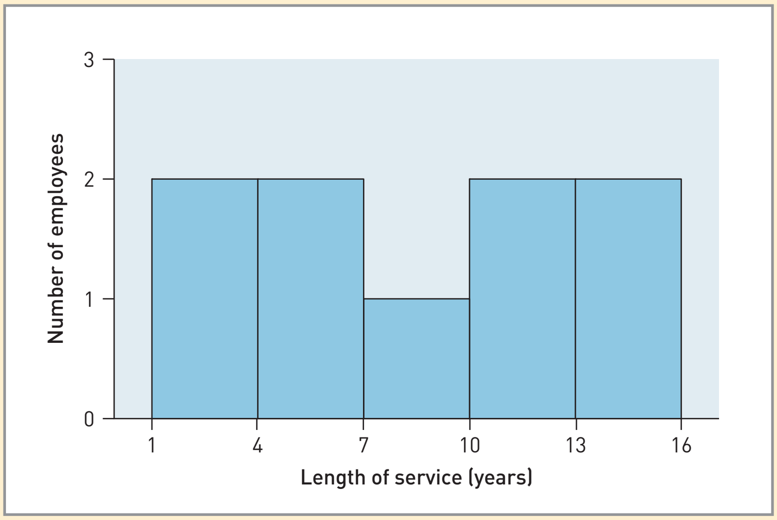


Figure 6.5 Lengths of service of Slugar staff

The ranges are exactly the same but, as Figures 6.4 and 6.5 show, the two distributions are different. The Rabota distribution has a central peak whereas the Slugar distribution dips in the centre with accumulations of observations at either end.

The ranges for the two distributions in Example 6.17 might be the same but the distributions are certainly not. The Rabota observations are more concentrated around the middle than the Slugar figures. There is more spread in the Slugar distribution but the range is not able to detect it. The range is therefore not a wholly reliable way of measuring the spread of data and this is because it is based only on the extreme observations.

6.3.2 Quartiles and the semi-interquartile range

The next measure of spread we will consider is the *semi-interquartile range*, which is often abbreviated to SIQR. We find this by using *quartiles*, which are order statistics like the median.

The median was the second measure of location we looked at in section 6.2. It is the middle point of a distribution that divides it in two. Half the observations are below the median and the other half are above it. The median cuts the distribution in half rather like a knife might cut a piece of cheese in half.

If we think of the median as cutting a distribution in two, the quartiles are the cuts that divide a distribution in four. The lower quartile cuts the lowest quarter of observations from the higher three-quarters. It is also known as the first quartile, or simply Q1. The median itself is the second of the quartiles as it cuts the lower half of the distribution (i.e. the lower two quarters of observations) away from the upper half (i.e. the upper two quarters of observations). The upper quartile cuts the highest quarter of observations from the lower three-quarters. It is also known as the third quartile, or simply Q3.

Quartiles are called *order statistics* as the values of them depend on sequence, or the order of observations in the distribution. There are other order statistics: *deciles* separate distributions into tenths, and *percentiles* separate distributions into hundredths.

Finding quartiles is similar to finding medians. We need the quartile position before we can get the quartile values.

The quartile position is midway between the median and the end of the distribution. We find it by starting with the median position, $(n + 1)/2$, in which n is the number of observations. The quartile position is the median position, which if not already a whole number must be rounded down to the nearest whole number, to which we add one then divide by two:

$$\text{Quartile position} = (\text{median position} + 1)/2$$

This is only where the quartiles are. Using an array, or a stem and leaf display, find the lower quartile by counting from the lowest observation to the quartile position. To get the upper quartile, count from the highest observation to the quartile position.

EXAMPLE 6.18

In one month the total costs (to the nearest £) of the mobile phone calls made by 23 women were:

14	5	15	6	17	10	26	10	12	17	13	29
7	27	33	16	30	9	15	7	33	28	21	

What are the median and upper and lower quartiles of this distribution?

Solution

Array:

5	6	7	7	9	10	10	12	13	14	15	15
16	17	17	21	26	27	28	29	30	33	33	

There are 23 observations so the median position = $(23 + 1)/2 = 12$th. The 12th value in the array is the second of the two 15s so the median is £15. The monthly cost of calls for half the women is less than £15, and the monthly cost for the other half is more than £15.

If the median position is 12 the quartile position = (12 + 1)/2 = 6.5th. The quartile position is therefore halfway between the 6th and 7th observations.

The lower quartile is midway between the 6th and 7th observations from the lowest. These are both 10, which means the lower quartile is £10. The monthly cost of calls for 25% of the women is less than £10.

The upper quartile is midway between the 6th and 7th observations from the highest. These are 27 and 26. The upper quartile is halfway between these, 26.5. The monthly cost of calls for 25% of the women is above £26.50.

SELF-ASSEMBLY GUIDE

Finding quartiles

- Start with the median position. This is $(n + 1)/2$, where n is however many observations there are in the distribution. In Example 6.18 there were 23 observations so the median position is (23 + 1)/2 = 12.
- Round down the median position to the nearest whole number. In Example 6.18 this is not necessary as the median location is already a whole number, 12.
- The quartile position is the median position plus 1 divided by 2. In Example 6.18 it is (12 + 1)/2 = 6.5, which puts it halfway between the 6th and 7th observations.
- This is the position of the lower and upper quartiles. The lower quartile is 6.5 observations from the lowest. The upper quartile is 6.5 observations from the highest.
- Using either an array or a stem and leaf display count along the observations to find the quartiles. In Example 6.18 the lower quartile is £10, midway between the 6th from lowest observation, 10, and the 7th from lowest, which is also 10. The lower quartile is therefore (10 + 10)/2 = 10.
- In Example 6.18 the upper quartile is £26.50, midway between the 6th from highest observation, 27, and the 7th from highest, 26. The upper quartile is therefore (27 + 26)/2 = 26.5.

The upper quartile, Q3, cuts the highest quarter of the distribution from the rest. The lower quartile, Q1, cuts off the lowest quarter from the rest. This means that between them the lower and upper quartiles span the middle half of the observations, in other words all the observations except for the lowest and highest quarters.

The difference between the lower and upper quartiles is the *interquartile range*. The semi-interquartile range (SIQR) is half the interquartile range:

$$\text{SIQR} = (Q3 - Q1)/2$$

EXAMPLE 6.19

What is the semi-interquartile range of the call costs data in Example 6.18?

Solution

The lower quartile is £10 and the upper quartile is £26.50.

$$\text{SIQR} = (£26.50 - £10)/2 = £16.50/2 = £8.25$$

The semi-interquartile range measures spread. The bigger the SIQR is the more widely spread are the observations in the distribution.