CHAPTER 4
SUMMARIZING DATA: AVERAGES AND DISTRIBUTIONS

This chapter considers the nature of datasets and the vectors within them. It suggests simple ways in which distributions of values can be described, summarized and analysed. A **distribution**, as we learned in Chapter 3, is a range of values observed for any one variable. Most column vectors in datasets consist of a distribution of values.

MEASURES OF CENTRAL TENDENCY

One of the first things that one may wish to do with a distribution of values is to calculate the **average** value. The average is an important summary characteristic but, as we shall see, averages must be chosen with care.

An **average** provides a value around which a set of data is located. It is a measure of **central tendency** in the data. Calculating the central tendency in interval or ratio data is often the first stage of an investigation. There are three commonly used measures of average:

The **arithmetic mean** usually referred to as just the **mean**;

The **median**;

The **mode**.

Each of these measures the average or central tendency of a distribution in a different way.

The choice of measure depends upon the nature of the distribution and the purpose for which the average is being calculated.

The mean

The **mean** is the average as it is most commonly understood and calculated: formed by adding all the values together and dividing by the number of observations. It is used only for interval data. The advantages of the mean are that:

(a) It takes account of all of the values;

(b) There are measures of dispersion that can be used with it (see below).

The major disadvantage is that it is sensitive to untypical extreme values: the value of the mean may be badly distorted away from the typical experience by the presence of one or two unusually large or small outlying values.

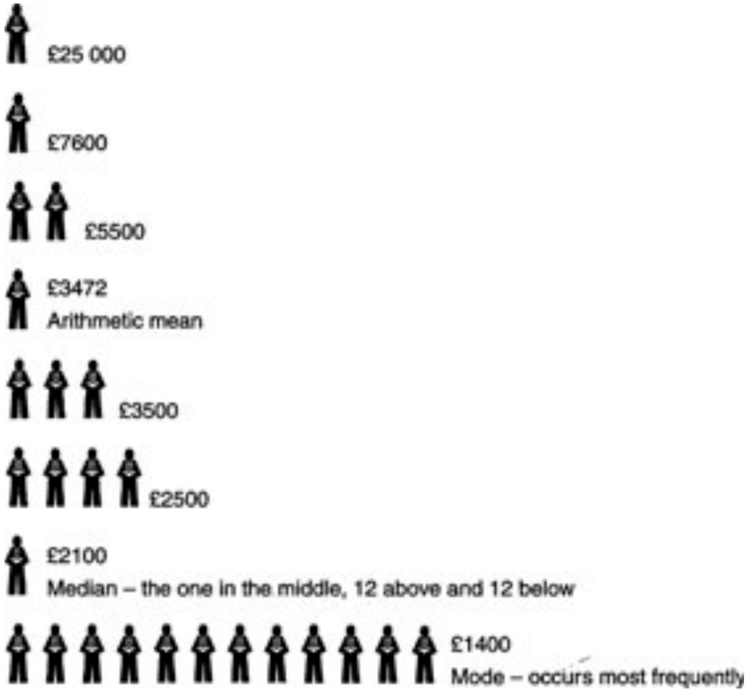

Figure 4.1 Pictogram of white-collar salaries in a firm in the 1950s.
Source: Based on Darrel Huff, *How to Lie with Statistics* (London 1973), p. 33.

In Figure 4.1, a pictogram showing average incomes in a business firm in the early 1950s, the mean would be a poor indicator of average experience because its value is inflated by the income of one man (the boss?) at the pinnacle of the earnings pyramid. The **mean** is usually represented by the symbol $\overline{X}$ and the formula for calculating the mean is given as:

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

Where
$\overline{X}$ is the mean of vector X;
X_i is the value of the variable for case i;
N is the number of observations;
Σ is 'the sum of'.

The mean of land tax payments in Table 3.17, for example, can be calculated as:

$$\text{mean} = \frac{\text{sum of all payments}}{\text{number of tax payers}} = \frac{\text{£202 10s. 0d.}}{47} = \text{£4 6s. 1d.}$$

It should be noted that in this case, the mean is again not a very good measure of the average or typical payment because of the existence of one or two atypically large payers. (Atypically large or small values in a distribution are usually referred to as **outliers**.)

The mean can also be calculated from a frequency distribution using the formula:

$$\bar{X} = \frac{\sum_{i=1}^{k} f_i X_i}{N}$$

Where

X_i is value of the variable for group *i*;
f_i is the frequency with which those values occur;
k is the number of groups;
N is the number of cases from which the frequency distribution has been compiled.

Thus, if we take the frequency distribution in Table 3.11, drawn from Table 3.10, we can calculate the mean prison sentence as:

$$\begin{aligned}\bar{X} &= \frac{\sum_{i=1}^{6} f_i X_i}{N} \\ &= \frac{[(3\times 7)+(2\times 10)+(1\times 12)+(5\times 14)+(8\times 15)+(2\times 20)]}{21} \\ &= \frac{(21+20+12+70+120+40)}{21} \\ &= \frac{283}{21} \\ &= 13.5\end{aligned}$$

Where

f_i in this case is the number of prisoners;
X_i is the sentence length;
N is the total number of prisoners;
k is the number of different sentence lengths represented.

Yet again the mean is not really the best measure of the average for this distribution because no prisoner is serving a 13.5-year sentence. All sentences are in whole years but more importantly there is an obvious candidate for the most typical experience which is 15 (this is by far the most commonly occurring experience, known as the mode; see below).

The mean provides the most justified measure of average when a distribution has few outliers that are likely to distort the mean, and when the values of the variable seem to be fairly evenly spread around a central value. Many measures occurring in nature,

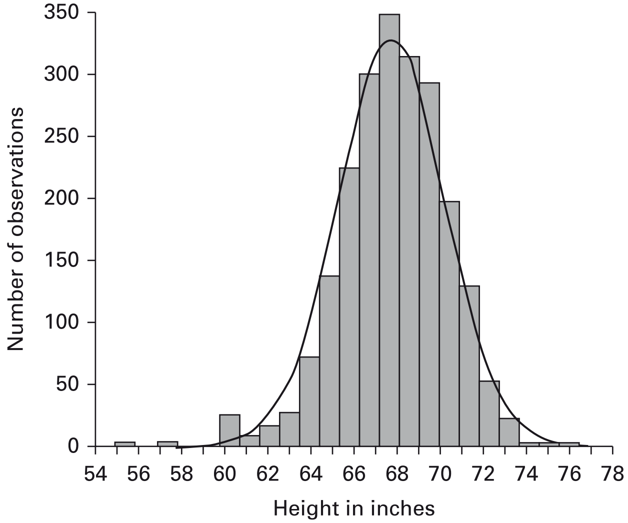

Figure 4.2 Height distribution of US passport applicants, 1830–1857.

Source: John Komlos, 'On the nature of the Malthusian threat in the eighteenth century', *Economic History Review*, 52, 4 (1999), p. 736. In this figure the histogram is compared for analytical purposes with the shape of the normal distribution (see pp. 112–113).

such as human physical attributes, other biological data and observations of social or economic characteristics such as the sizes of households, values of industrial firms, numerically scaled educational levels, social skills or wages in a population, tend to cluster evenly (above and below) an average measure, with very few observations lying outside of a certain range. This type of finite and symmetrical distribution (known as a binomial or bell-shaped distribution) is a common one in historical evidence (although the variables that historians are interested in are often distributed in other ways too).

Figure 4.2 provides an example of a bell-shaped distribution. There is a fairly even spread of observations above and below the mean, tapering off symmetrically. For further discussion of the properties of bell-shaped distributions, see p. 112.

The median

The **median** is the observation that lies at the centre or middle of a distribution when all of the observations or values are ranked in size order. It can be used with ordinal or interval data. When there is an even number of cases the median is the mean of the two middle ranking values. The advantages of using the median are:

(a) it is immune from the influence of extreme values

(b) it has some measures of dispersion associated with it (see below pp. 101–104).

The median is a better way of calculating the average level of tax paid from Table 3.17, than was the calculation of the arithmetic mean, above, because it is immune from the influence of the two largest and untypical tax payers (outliers) (Sir Wats Horton, Gentleman, who held 24 separate parcels of land and George Stansfield, merchant, who owned 15). The median tax payment is £1 0s. 6d. compared with the mean of £4 6s. 1d.

It is possible to *estimate* the median of a grouped frequency distribution which is useful if the original figures are not available. This is done by assuming that the values of items in the class containing the median are distributed evenly, that is that the median falls in the middle of that class. The probability of this occurring increases with the size of the dataset so this may be a decisive consideration in adopting this calculation. The median value of land tax paid in Sowerby in 1782 can be calculated roughly from Table 3.18 as £3, which may be a useful enough approximation depending upon the nature of the enquiry, but the size of the dataset might warn against this method.

The mode

The **mode** is the most commonly occurring observation. It can be used with nominal, ordinal or interval data and is the only average one can use with nominal data. The advantage of the mode is that it represents the most common experience or occurrence but the disadvantage is that it takes no account of other observations, has no measure of dispersal associated with it and can be entirely misleading if there is more than one commonly occurring observation (as in a bi-modal or tri-modal distribution of which more below). It is a useful measure when a distribution is not spread evenly around a central value but is of limited use when data are very dispersed.

In our example above concerning sentence lengths of prisoners, the mode is a better way of expressing the average prison sentence than the mean as the distribution is not widely spread and there is a very obvious common experience of 15 years.

In a grouped frequency distribution the **modal class** is the one with the highest frequency. In Table 3.18, for example, the modal class of land tax payers is 5s.–<£1 which contains 32 per cent of the observations.

In our pictogram in Figure 4.1 the mode (£1,400) and the median (£2,100) are both better expressions of average than the mean (£3,472) because they are less affected by the one high outlier.

The geometric mean

The **geometric mean** is another average but it is less commonly used than the mean, median or mode. The geometric mean is defined as the *N*th root of the product of the distribution (where *N* is the number of items in the distribution). The geometric mean is used only with interval data, mostly in averaging growth rates or indices of growth. (An index-plural indices-is a series expressed in percentage terms as explained in Chapter 5.)

To calculate the geometric mean one multiplies all the *N* values of a variable, X, together and then one takes the *N*th root:

$$\text{geometric mean} = \sqrt[N]{X_1 X_2 X_3 \ldots X_N}$$

This may also be written as:

geometric mean = $(X_1X_2X_3...X_N)^{1/N}$

Note that there is no need for multiplication signs: X_1X_2 is the same as. $X_1 \times X_2$.

Example

If the price of commodity A rises from £25 to £50 this is an increase of 100 per cent

If the price of commodity B rises from £80 to £100 this is an increase of 25 per cent

The mean increase of the two price rises is $= \frac{125 \text{ per cent}}{2} = 62.5$ per cent

The geometric mean = $\sqrt{100 \times 25}$ per cent=50 per cent

Which of the two measures of average growth detailed above one should use is open to debate and depends upon the researcher's purpose. The geometric mean gives less weight to extreme values than the arithmetic mean but there is no measure of dispersion associated with it. Growth rates can also be measured and expressed in other ways (see Chapter 5) and these methods are often preferred to either the arithmetic or the geometric mean.

Choice of average

It is not always easy to make a clear-cut decision about which measure of average (mean, median or mode), is the best reflection of typical experience given the character of the data. Sometimes it will depend upon the questions that one is asking about the evidence. The mode will be favoured where the most common experience is desired, the median where better knowledge of the impact of the distribution on average experience is needed, and the mean will be chosen where account must be taken of all of the observations equally (best done where there are few or no outliers and where the distribution is not markedly skewed – see below) Often two or all three of the measures are stated together. The differences between them provide a good indication of the nature of the distribution of values.

For example, in Table 4.1 taken from E. A. Wrigley's study of marriage ages in early modern Colyton, Devon, all three of the measures of average marriage age are given. This is because each highlights different features of the data upon which Wrigley comments and each gives useful additional information. Where the three measures of average are given together in this way an indication of the shape of the distribution as a whole can be visualized. The data on male and female average marriage ages in Table 4.1 can be seen immediately to be 'skewed' because for both men and women and for all of the time periods the mean is greater than the median which in turn is greater than the mode. This is characteristic of a positively skewed distribution (see below Figure 4.10a). The use of the

Table 4.1 Age at first marriage in Colyton, 1560–1837

	Number	*Mean*	*Median*	*Mode*[a]
Men				
1560–1646	258	27.2	25.8	23.0
1647–1719	109	27.7	26.4	23.8
1720–1769	90	25.7	25.1	23.9
1770–1837	219	26.5	25.8	24.4
Women				
1560–1646	371	27.0	25.9	23.7
1647–1719	136	29.6	27.5	23.3
1720–1769	104	26.8	25.7	23.5
1770–1837	275	25.1	24.0	21.8

[a] The mode is interpolated from the mean and the median and not derived directly from the data.

Source: E. A. Wrigley, 'Family limitation in pre-industrial England', *Economic History Review*, 19, 1 (1966), p. 86.

mean with the median is a common way of roughly indicating the shape of a distribution of values.

Another example of a positively skewed distribution where the mean is biased upwards, away from average experience by the presence of a relatively small number of very high values, is the distribution of wealth in most societies. Recent controversial research has suggested that inequality in terms of accumulated wealth, as well as in terms of income distribution has been growing in developed countries in the last few decades.[1] Around the turn of the millennium, the Institute of Fiscal Studies reported that although average wealth was growing in Britain, the distribution was becoming more unequal. Mean wealth (in accumulated savings and assets) of £7,136 contrasted markedly with a median of only £750! In addition, 30 per cent of the population had no savings outside of their home and pension and around 10 per cent (mostly single parents and out-of-work couples) had no savings at all.[2] The skew of the distribution both of income and of wealth is much more extreme in the United States largely because the redistributive impact of state spending, social security and healthcare spending is much more muted. The top 1 per cent of society in the USA in 2012 owned 40 per cent of the wealth (compared with 23 per cent in 1978). The top 0.1 per cent held 22 per cent (compared with 7 per cent in 1978).[3] In Figure 4.3 it can clearly be seen that wealth holding is skewed: the lowest half of the population own only 9 per cent of wealth, the top 20 per cent own 65 per cent (2008–2010 figures). (The calculation and use of percentiles and deciles is covered later in this chapter.)

Another example to illustrate the nature of skew in distributions is provided by Botticini's analysis of the marriage market in fifteenth-century Tuscany which is included as an exercise later in this volume.[4] The median and the mean are displayed together for a number of different variables as shown in Table 4.2. The means are higher than the

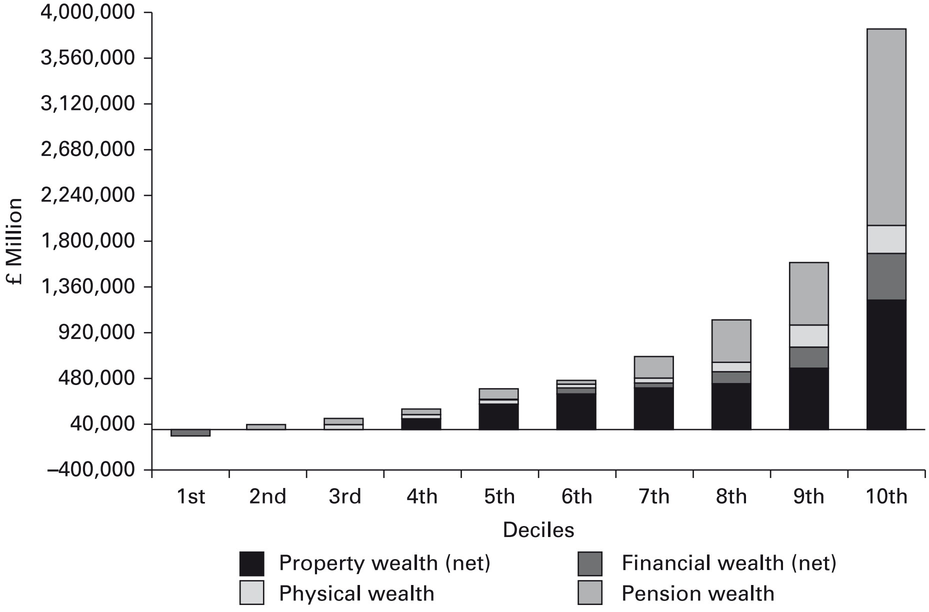

Figure 4.3 Wealth distribution in England.

Source: Office of National Statistics, Wealth and Assets Survey 2008–2010.

Table 4.2 Summary statistics of marriages in Cortona, 1415–1436

	Mean	*Median*	*Standard deviation*
Dowry (florins)	125.5	70	105.9
Groom's age (years)	28.1	27	8.3
Bride's age (years)	18.8	18	4.7
Groom household's wealth (florins)	609.7	164	1692.84
Bride household's wealth (florins)	700.7	196	1997.66
Number of children in grooms' households	2.25	2	1.87
Percentage of daughters in grooms' households	0.08	0	0.17
Number of children in brides' households	3.14	3	2.33
Percentage of daughters in brides' households	0.65	0.6	0.27
N		224	

Note: the marriages refer to households living in the town of Cortona and in 44 villages in its countryside.

Source: M. Botticini, 'A loveless economy? Intergenerational altruism and the marriage market in a Tuscan town, 1415–1436', *Journal of Economic History*, 59, 1 (1999), p. 108.

medians for all of the variables again indicating the presence of positively skewed distributions. The standard deviation, as used here, is a measure of dispersion of the data around the average and is explained in the next section.

MEASURES OF DISPERSION

An average on its own tells us very little about the entire population: in particular, it says nothing about how divergent from the average is the distribution of individual observations. All distributions are not only clustered around central points but also spread out, or dispersed, around them. The **range** is a first indication of dispersal. The range of a set of data is literally its spread: the highest value of the distribution minus the lowest. The range is often used with the mode but can be used with any interval data.

There are a number of more sophisticated measures of dispersion which can be used with the mean and the median.

Dispersion around the mean: standard deviation and variance

Many very different distributions can have the same mean. For example, all three of the distributions in Table 4.3 have a mean of 45.87 despite the fact that A is widely dispersed (range 99) whilst C is closely clustered (range 4) and B is influenced by the presence of one extreme atypical value (an 'outlier') and has the largest range because of this. Hypothetical data are used here to make the differences between the three distributions very clear.

Table 4.3 Distribution of defamation cases in three English courts, 1680–1687

	Distribution		
Year	*A*	*B*	*C*
1680	100	20	48
1681	88	28	47
1682	70	22	46
1683	50	45	45
1684	30	16	45
1685	20	167	45
1686	8	40	44
1687	1	29	47
Total	367	367	367

Source: Hypothetical data.

Because the average on its own tells us little about the entire population it is almost always used with some indication of the spread of data. A measure of dispersion tells us to what extent the values of a distribution are, or are not, bunched around the average. The measures of dispersion most commonly used with the mean are the **variance** and the **standard deviation**.

The **variance** is the average of the squares of the deviations from the mean. It is calculated by adding the square of the deviations of the individual values from the mean of the distribution together and dividing this sum by the number of items in the distribution. The following formula achieves this:

$$\text{variance} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N}$$

Where

$\overline{X}$ is the mean;

X_i is the value of the variable for row i;

N is the number of observations.

The **standard deviation** is another measure of dispersion around the mean. It is normally represented by the letter s or by the abbreviation SD. It is found by applying the formula for the variance and then taking the square root. The variables, and $\overline{X}$, X_i and N are as defined already. The variance is always equal to the square of the standard deviation (that is, s^2).

$$s = \sqrt{\left(\frac{\sum_{i-1}^{N}\left(X_i - \overline{X}\right)^2}{N}\right)}$$

In distribution A, $s = 34.77$ and the variance is 1209.11.

In distribution B, $s = 46.69$ and the variance is 2180.36.

In distribution C, $s = 1.27$ and the variance is 1.61.

See Table 4.4 for a partial breakdown of the calculations.

The greater the dispersion, the larger the standard deviation and the variance. In each case s is expressed in the original units, in this example in court cases.

The standard deviation can also be directly calculated from a grouped frequency distribution by applying the formula:

$$SD = \sqrt{\frac{\sum fD\overline{x}^2}{\sum f} - \left(\frac{\sum fD\overline{x}}{\sum f}\right)^2} \times \text{class interval}$$

Here f is the frequency and $D\bar{x}$ is the deviations from the mean (or the assumed mean). Statistical software applied to data in a matrix or an electronic spreadsheet makes this calculation very straightforward.

The formula for the standard deviation (s) takes into account the amount that each value deviates from the mean (the $X_i - \overline{X}$ part of the formula), which is what makes it so much more useful, in most cases, than the range.

Table 4.4 Statistics relating to Table 4.3

Year	*i*	*Distribution*					
		A		*B*		*C*	
		X_i	$X_i - \overline{X}$	X_i	$X_i - \overline{X}$	X	$X_i - \overline{X}$
1680	1	100	54.1	20	−25.9	48	2.1
1681	2	88	42.1	28	−17.9	47	1.1
1682	3	70	24.1	22	−23.9	46	0.1
1683	4	50	4.1	45	−0.9	45	−0.9
1684	5	30	−15.9	16	−29.9	45	−0.9
1685	6	20	−25.9	167	121.1	45	−0.9
1686	7	8	−37.9	40	−5.9	44	−1.9
1687	8	1	−44.9	29	−16.9	47	1.1
$\sum_{i=1}^{N}(X_i - X)^2$		9673		17442		12.88	
Variance		1209.12		2180.36		1.61	
Standard deviation		34.77		46.69		1.27	

Note: For all distributions, the number of observations, N, is 8, the number of court cases is 367, and the average, $\overline{X}$ is 45.9. Note also that the square of a negative number is positive (i.e. −15.9 squared = −15.9 × −15.9 = +252.81).

Source: Hypothetical data.

The Z score

Use of a measure called the **Z score** is a common method in the social science and historical literatures. A Z score is the number of standard deviations which an observation is above the mean (if it is positive) or below the mean (if it is negative). Where Z scores are used the standard deviation becomes a sort of yardstick for comparative purposes. Distributions of Z scores can be created that enable the dispersion of different distributions to be compared. The standard deviation itself is no good for this because it is expressed in the original units of measure, for example, dollars, pounds sterling, persons, cows. Z scores provide a universal unit for measuring dispersion. Because Z scores have standard values they are sometimes called standard scores.

In research on growth stunting in children born in Rwanda between 1987 and 1991 the relative impact of crop failures and the military conflict was the focus of attention in a study employing Z scores. It was found that in poor and non-poor households, boys and girls born during the Rwandan conflict, in regions experiencing fighting, were negatively affected, with height-for-age Z scores 1.05 standard deviations lower than the norm. Conversely, only girls were negatively affected by crop failure, with girls exhibiting 0.86 standard deviations lower height-for-age Z scores, the impact being worse for girls in poor households. This suggested that girls bore the brunt of dietary restriction in times of crop failure but that both sexes were similarly affected by conflict.[5] In a study of height and living standards in China between 1979 and 1995 (used as an exercise later in this volume), Chinese heights were compared to international reference standards of Z scores for such distributions at various ages.[6]

Dispersion around the mean: the coefficient of variation

The **coefficient of variation** is another measure of the extent to which a variable differs from its mean. It is simply the standard deviation (*s*), divided by the mean and is generally expressed as a percentage:

$$\text{coefficient of variation} = \frac{s}{\overline{X}} \times 100 \text{ per cent}$$

Because it is expressed as a percentage it can be used to compare the dispersion of distributions of different sorts of variables one with another. The coefficient of variation is normally only calculated for this purpose – to compare the degree to which two variables differ from their respective means. It is not possible to use standard deviations for this because standard deviations are expressed in the original units of the variable, for example, persons, exports, strikes, ploughs, hearths, looms and so on, whereas the coefficient of variation is always a percentage.

If we were told that the three distributions in Table 4.3 were not all court cases but that each distribution related to a different variable we would need, for comparative purposes, to calculate the coefficient of variation. For example, if the dataset described the assets of eight farmers in the early nineteenth century with:

- series A the number of cows
- series B the value of seed on hand in £
- series C the value of land and farm buildings in £thousands

We might wish to calculate the coefficient of variation to see the extent to which the different sorts of assets of these farmers varied from the average experience. The coefficients of variation of the three distributions are:

Distribution A: 0.76 per cent

Distribution B: 1.02 per cent

Distribution C: 0.03 per cent

The coefficient of variation is also used in comparing the variation of certain measures at different time periods or for different countries because standard measures are far easier to compare than original units. Tables 4.5 and 4.6 are drawn from an article by Jeffrey G. Williamson entitled 'Globalisation, convergence and history'. His estimates of coefficients of variation of real wages, 1854–1939 and of coefficients of variation of Gross Domestic Product (GDP) per worker hour, for the OECD, 1870–1938, support his argument that growth convergence is linked to globalization and that convergence was arrested in the period 1914–1950.[7]

Another example is provided by Lazerev and Gregory's study of vehicle allocation in the Soviet command economy of the 1930s (included as an exercise later in this volume). They show 'satisfaction rates' (the ratio of allocated to requested vehicles). They show that the Soviet system began with an enormous excess demand for vehicles but was able to satisfy nine out of ten consumers by early 1937. Especially during supply shocks, such as that in 1932, however, the Dictatorship favoured 'preferred customers' such that there was a very uneven 'actual' distribution of satisfaction shown by a coefficient of variation that reached 83.5 per cent. Some figures from the research are given in Table 4.7. (Rows 3 and 4 can be ignored for the moment. They are explained in Chapter 6.)

Table 4.5 Coefficients of variation of real wages, 1854–1939

	Full sample[a]			*Full sample less North America*[b]			*Full sample less North America and Iberia*[c]	
Year	*C(13)*	*C(17)*	*C(16)*	*C(12)*	*C(15)*	*C(14)*	*C(10)*	*C(13)*
1854	0.326			0.308			0.340	
1870	0.254	0.255		0.224	0.223		0.229	0.232
1890		0.199			0.114			0.102
1913		0.191			0.068			0.039
1914			0.103			0.085		0.068
1926			0.148			0.146		0.138
1927		0.188	0.147		0.186	0.142		0.131
1939		0.285			0.200			0.138

[a] The 'full sample' included the following 13 countries until 1870: Australia, the United States, Belgium, France, Germany, Great Britain, Ireland, Netherlands, Norway, Spain, Sweden, Brazil and Portugal; in 1870 the following four countries were added to the sample: Argentina, Canada, Denmark and Italy; Portugal dropped from the sample from 1914 to 1926 and then rejoined.

[b] 'Full sample less North America' excludes Canada and the United States, implying that we start with 12 countries and then increase to 15 in 1870; again, Portugal dropped from the sample between 1914 and 1926.

[c] 'Full sample less North America and Iberia' excludes the United States, Canada, Spain and Portugal, implying that we start with 10 countries and expand to 13 in 1870.

Note: The number of countries in the sample, *x*, is indicated by the column heading *C(x)*.

Source: J. G. Williamson, 'Globalisation, convergence and history', *Journal of Economic History,* 56, 2 (1996), p. 280.

Table 4.6 Coefficients of variation of gross domestic product (GDP) per worker-hour, 1870–1938

Year	*Full sample*[a]	*Full sample less North America*[b]
	C(15)	C(13)
1870	0.153	0.169
1890	0.118	0.122
1913	0.107	0.088
1929	0.110	0.080
1938	0.090	0.054

[a] The 'full sample' includes Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, the Netherlands, Norway, Sweden, Switzerland, the United Kingdom, and the United States; it does not include Japan.

[b] The 'full sample less North America' drops Canada and the United States from the full sample.

Source: J. G. Williamson, 'Globalisation, convergence and history', *Journal of Economic History,* 56, 2 (1996), p. 280.

Table 4.7 Vehicle allocation in Soviet Russia during supply shocks and periods of normality

	1932		*1934 3rd quarter*		*1934 4th quarter*		*1937 2nd quarter*	*1937 4th quarter*
	Planned	*Actual*	*Planned*	*Actual*	*Planned*	*Actual*	*Actual*	*Actual*
(1) Average satisfaction rate, %	31.4	15.0	41.7	39.6	53.6	51.2	87.6	46.5
(2) Coefficient of variation in satisfaction rates, %	26.4	83.5	47.8	48.5	37.5	37.8	18.3	53.3
(3) Pearson R^2 correlation	0.996	0.919	0.962	0.962	0.807	0.778	0.98	0.747
(4) Spearman rank correlation	0.976	(0.542)	0.957	0.951	0.861	0.881	0.978	0.645

Source: Valery Lazerev and Paul R. Gregory, 'The wheels of the command economy: allocating Soviet vehicles', *Economic History Review*, 55 (2), (2002), pp. 324–348, p. 333.

In the study of living standards and stature in China, mentioned above, the coefficient of variation was used to show that variation in heights amongst rural children was greater than amongst urban children and that this increased over time.

Rank order dispersal measures

These are commonly used with the median. The median is only one of a range of measures that summarize data according to their rank order. The median divides the ranked distribution into half. The first **quartile (Q$_1$)** is defined as the middle number between the smallest number and the median of the dataset. The second **quartile** (**Q$_2$**) is the median of the data. The third **quartile** (**Q$_3$**) is the middle value between the median and the highest value of the dataset. Others measures commonly used in the same way are:

quintiles

deciles

percentiles

The three **quartiles** divide the ranked distribution into 4 equal parts.

The four **quintiles** divide the ranked distribution into 5 equal parts.

The nine **deciles** divide the ranked distribution into 10 equal parts.

The 99 **percentiles** divide the ranked distribution into a hundred equal parts.

Consider the distribution of 20 observations in Table 4.8, ranked in size order, which derives from an archaeological project.

As mentioned above, the **median** can also be expressed as the second quartile **(Q$_2$)** and the measure of dispersion often used with the median is the **interquartile range.** This is the difference between the first and the third quartiles (**Q$_1$** and **Q$_3$**). In the example in Table 4.8, the interquartile range is:

$\mathbf{Q_3 - Q_1}$

$\mathbf{= 10.5 - 4.5}$

$\mathbf{= 6}$

Sometimes this is divided by two to form the **semi-interquartile range** or **quartile deviation** which would in this example be 3.

The ninth decile of the distribution in Table 4.8 is 11.5. This distribution is too small to have percentiles. Percentiles can be calculated only where there are at least a hundred observations.

Figure 4.4 is a cartogram demonstrating the inequality of Russian landholding prior to the Revolution. The distributions of 'Private holdings' and of 'All types of property' are divided into quartile ranges which is a level of detail sufficient to make the main point:

Table 4.8 Number of archaeological remains found by twenty postgraduate assistants

Number of hearths:	*Deciles*	*Quintiles*	*Quartiles Q*	
2				
2				
←	1st = 2.5			
3				
4				
←	2nd = 4.0	1st = 4.0		
4				
←			1st = 4.5	
5				
←	3rd = 5.5			
6				
6				
←	4th = 6.0	2nd = 6.0		
6				
7				
←	5th = 7.5		2nd (median) = 7.5	Interquartile range = 6 (10.5–4.5)
8				
8				
←	6th = 8.5	3rd = 8.5		
9				
9				
←	7th = 9.5			
10				
←			3rd = 10.5	
11				
←	8th = 10.5	4th = 10.5		
11				
11				
←	9th = 11.5			
12				
23				

Note: Spreadsheet functions give slightly different results from the above first approximations as they take account of any skew in the distribution. In the example above the first quartile is really 4.75. The third quartile is 10.25 and the ninth decile is 11.10. The interquartile range is 5.5.

Source: Hypothetical data.

that there were clear local concentrations of wealth and a marked variation in average wealth holding in different regions. The **Gini coefficient** is a measure of statistical dispersion intended to represent the income distribution of a nation's residents, and is the most commonly used measure of inequality. A Gini coefficient of 0 denotes a completely equal distribution and of 1 (10, or 100 depending upon choice of scale) denotes complete inequality (where one individual earns all the income). Further discussion of Gini coefficients is included in Example 3 below. In our Russian example, property was more unequally distributed around St Petersburg and Odessa, and in the north-east of the country, on the eve of the Revolution, which may have created particular tensions in those areas. As the article also shows, income inequality closely matched the geographical distribution of land- and property-holding inequality.

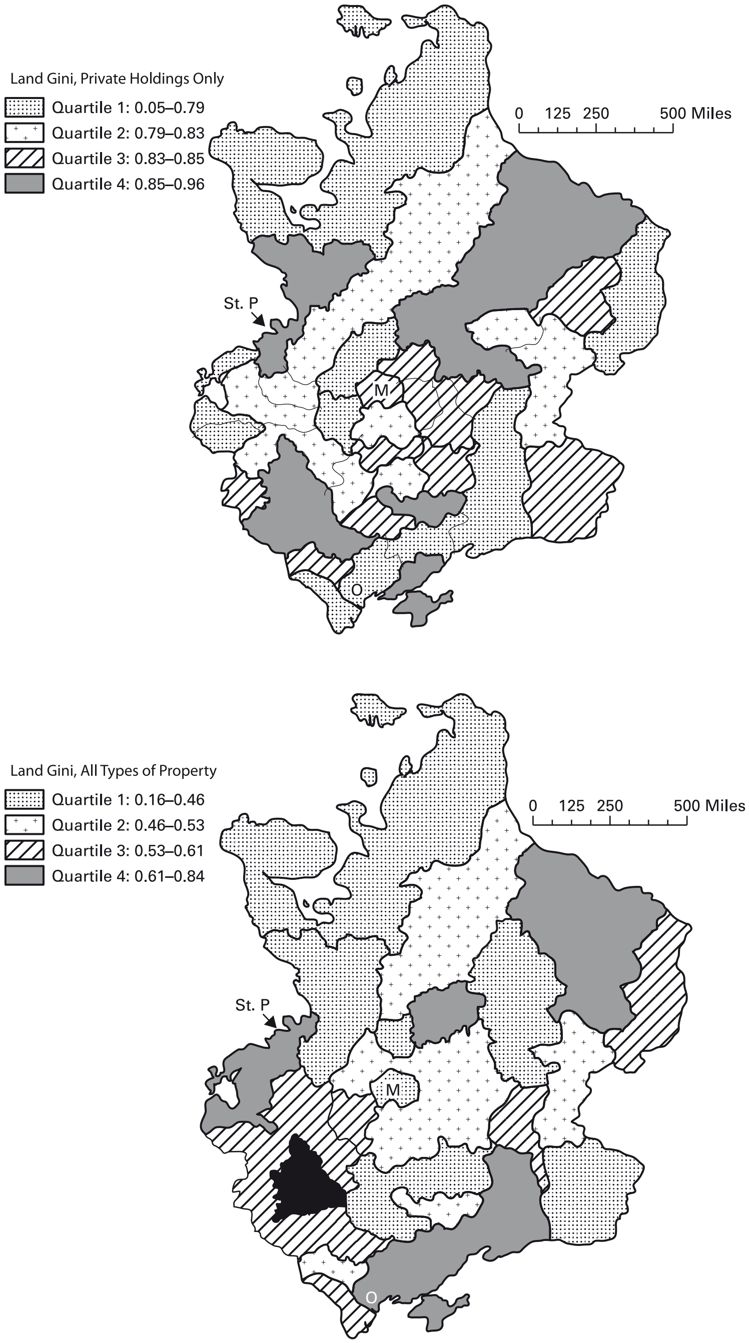

Figure 4.4 The geography of landholding inequality in Russia c. 1905.

Source: Peter H. Lindert and Steven Nafziger, 'Russian inequality on the eve of Revolution', *Journal of Economic History*, 74 (3), (2014), pp. 767–798, p. 780.

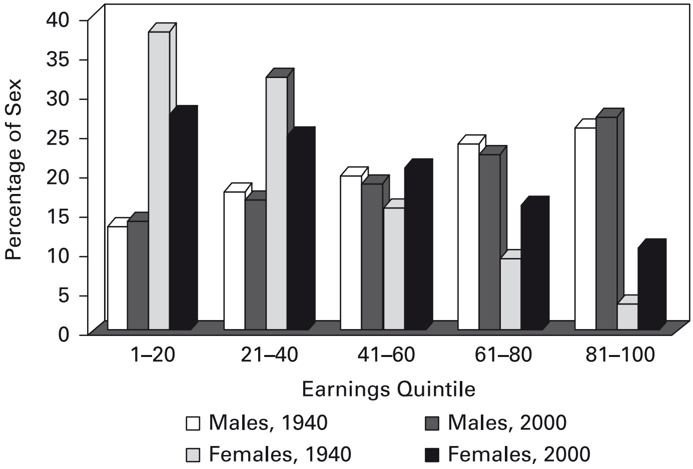

Figure 4.5 Real personal earnings quintiles for non-farm year-round workers by sex, 1940 and 2000. Source: M.B. Katz, M.J. Stern and J.J. Fader, 'Women and the paradox of economic inequality in the twentieth-century', *Journal of Social History*, 39 (1), (2005), p. 78. © Oxford University Press.

Quartiles and quintiles are often used as divisions in bar charts as in Figure 4.5. This shows the earnings of non-farm workers by sex in the USA, 1940–2000. The authors use employment data from the census to explore the 'paradox' of inequality in twentieth-century America, that is, the enduring coexistence of inequality with individual and group social mobility. Figure 4.5 shows that women's earnings were represented in a much wider range of the income scale by 2000 than ever before. The use of quintiles is effective here: the income of women working full-time is expressed in earning quintiles by gender for 1940 and 2000. The pattern immediately visible in Figure 4.5 is the persisting income disparity between men and women both in 1940 and 2000. Women were more likely to be found at the bottom quintile and much less likely to be at the top quintile of earnings. In 2000, the gender disparities are much less sharp than in 1940 but just as apparent. In their wider analysis the authors demonstrate that women had entered into a much wider variety of occupations and industries by 2000 than they had half a century earlier but that significant inequality of earnings emerged within the female workforce, as they had within the male labour force. Higher education played a key role in emerging income differences.

An **advantage** of the **interquartile range** and the **quartile deviation** is that they are immune from the influence of very small or very large values. This can be an advantage if there are just a few extreme outliers that would seriously effect alternative measures of dispersal such as the standard deviation.

More examples of analysis of distributions from history

If we are researching a historical question that hinges upon the nature of a dataset, choosing and applying the most appropriate measure of central tendency and dispersal are likely to be crucial to the arguments which may be made. We now consider four examples from historical research of such choices and applications discussing their advantages and any disadvantages or weaknesses.

Example 1: Eighteenth-century slopsellers

Beverly Lemire's analysis of male and female activity as slopsellers in London in the later eighteenth century provides our first very simple example. Slopsellers were suppliers of clothing to the Navy. The types of work clothes supplied included knitted caps, stockings, shirts, waistcoats, shoes, handkerchiefs, drawers, and 'blew' suits. Slopsellers had to bid for the contracts and once secured the clothing would be delivered to the ship in question where it was stored in slop chests under the care of the purser. During a voyage and as necessary, the purser would sell the garments to the crew keeping a tally for the slopseller.

Comparing the distribution by insured wealth of male and female slopsellers gives an indication of the difficulties that women faced in running businesses of this kind in the period. There were much fewer women running these businesses than men and although both distributions are skewed in favour of smaller sized concerns the female-headed business distribution is more skewed than the male. Social attitudes to women working at all levels but particularly as business owners, together with legal obstacles concerning the restriction upon women owning property in their own right and acting as an independent legal entity, doubtless help to account for the smaller number of female-headed businesses in the period. Indeed it is likely that many of these were run by widows who were carrying on their husbands' trade post-mortem (which was legally and socially more acceptable than a woman embarking upon and leading a business solely in their own name). The smaller scale nature of female-headed slopsellers also however points to particular difficulties of this trade. At a time when military demand for clothing was booming, it seems likely that women were held back by differential access to the credit necessary in this trade and by the nature of bidding for contracts which was often promoted by male forms of sociability.

In looking at Lemire's table (Table 4.9), it is clear that in the case of both male and female sellers, the mean is not a good indicator of average insured value. For women the mode would be the most appropriate value and, because the distribution is more evenly spread in the case of men, the median might be preferred. But remember, the choice of average is dependent upon the purpose of any research investigation and the median and mean are the only ones for which there is an accompanying set of measures of dispersion. One is prompted to ask of the table why there is no continuity in the range of insured values covered. Did this arise from some sort of sampling or from insurers only valuing to the nearest £100? It seems more likely that it is an error and that the categories should read £100–£299; £300–£499, and so on. One should also beware of the highest insured category which covers almost as great a range as the previous three categories combined. One would need to consult the original research to check on the range of values for men and women in this £1100 to £2000

Table 4.9 Insured property of female and male slopsellers in London, 1777–1796

	Slop women	*% of total*	*Slop men*	*% of total*
£100–£200	11	29	31	17
£300–£400	14	37	60	34
£500–£600	6	16	34	19
£700–£800	2	5	13	7
£900–£1000	3	8	24	14
£1100–£2000	2	5	16	9
Totals	38	100	178	100

Source: Ms 7253 Royal Exchange Insurance Registers; Ms 11936&11937, Sun Fire Insurance Registers, Guildhall Library, London. Based upon table in Beverly Lemire, *Dress Culture and Commerce. The English Clothing Trade before the Factory, 1660–1800* (Basingstoke 1997), p. 51.

category. In neither case here, with such a small dataset and with uneven categories, would it be appropriate to undertake further statistical analysis of the distribution beyond the average, the frequency tables and the associated charts. As the figures are relatively easy to input into a spreadsheet, you may wish to draw up the relevant bar charts and derive frequency polygons for male and female insured values that will immediately highlight the (positively) skewed nature of each distribution (on the same graph).

Example 2: The impact of the Black Death in Birdbrook, Essex

In considering the local impact of the Black Death in Essex and specifically the impact upon tenurial developments and the availability of customary land, Phillipp Schofield employed mean, median and standard deviation measures to demonstrate change over time. Table 4.10 shows mean length of leasehold where this indicates the period during which the tenement can be observed as remaining in the hands of the lessee by tracing it in the accounts from one year to the next. But the median is used to indicate the average term given at the inception of the lease to allow inclusion of terms granted for life or lives and to avoid replacing these with an arbitrary number of years. The standard deviation refers to leasehold lengths and relates to dispersal around the mean. The table shows that the average length of time that a lessee remained in his leasehold reduced dramatically in the first decade of the fifteenth century. The standard deviation of tenurial tenacity also declined markedly from the third quarter of the fourteenth century onwards. Schofield argues that these shifts reflected the replacement of a manorial economy based upon labour services with one based upon the money rent of farms and that a lot of the new tenants were incoming migrants.

Table 4.10 Average length of occupation leaseholds commencing in each decade, from 1350 to 1409

Decade	*Mean length of leasehold*[a]	*Standard deviation*	*Median length of term*[b]	*Number of leases entered*[c]
1350–9	17.5	13.162	12	6
1360–9	18	12.675	9	3
1370–9	11.3	11.609	7	7
1380–9	15.2	9.441	9	12
1390–9	10.6	5.795	3	13
1400–9	4	3.210	1	27

[a] This is not the term given at the inception of the lease (see note b) but is the period during which the tenement can be observed as remaining in the hand of the lessee by tracing it through the accounts from one year to the next. Note also that the length of lease has been calculated as starting and ending in the first year of each account.

[b] This is the term actually given at the inception of the lease, which, in the case of longer terms, would be recorded in the court roll or, in the case of very short terms, in the 'farms' section of the account. The median value has been used here rather than the mean so as to allow inclusion of terms granted for life or for lives without replacing these with an arbitrary number of years.

[c] Three leaseholds entered in the decade 1380–9, four in that of 1390–9 and eight in that of 1400–9 had not expired by the accounting year 1409–10. Only limited observation is possible after this date: the next surviving accounts date from accounting years 1412–3, in which year the same lessees continue to hold, and 1426–7, by which date all but one of these lessees of customary tenements had disappeared. The mean length of leasehold has been distorted as a result: in the case of 13 of the 14 lessees still holding in 1409–10 it has been assumed, for the basis of the calculation, that their tenure of the lease ended in 1412–13, and the lease of the individual still *in situ* in 1426–7 has been taken as ending in that accounting year. The effect of this is, obviously, to reduce the size of the mean, but the accuracy of the trend can be tested by artificially extending the length of those leases whose terminal date cannot be observed. By adding 3 years after 1412–13 for those leases commencing in 1380–9, 7.5 years for those commencing in 1390–9 and 10 years for those commencing in 1400–9 the following means ad standard deviations are obtained:

Decade	Mean	Standard deviation
1380–9	15.75	10.248
1390–9	13.30	7.289
1400–9	7.26	7.214

Source: Phillipp R. Schofield, 'Tenurial developments and the availability of customary land in a later medieval community', *Economic History Review*, 49, 2 (1996), p. 259.

Example 3: The impact of taxes and benefits on UK incomes in the late 1980s

In an example from more recent history, Figure 4.6 shows the effects of taxes and benefits upon quintile groups of households in Britain in 1987. It suggests that all five groups make direct and indirect tax contributions to the Welfare State and enjoy benefits in cash and kind. These taxes and benefits taken together make the distribution of final income considerably more equal than the distribution of original income.[8]

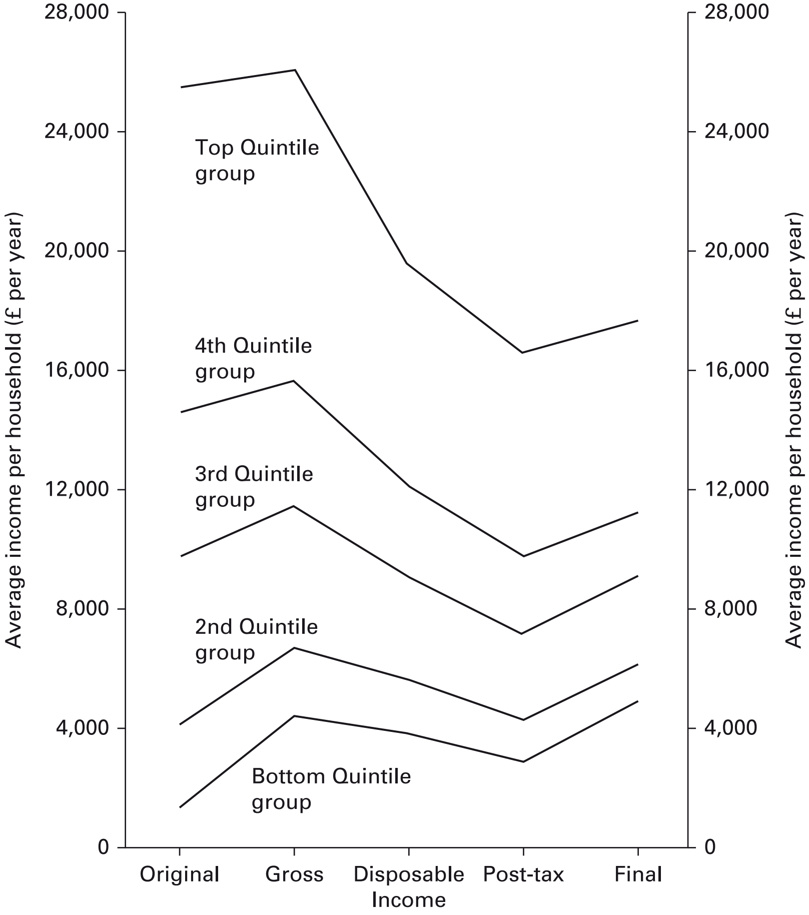

Figure 4.6 The effects of taxes and benefits on quintile groups of households, 1987.

Note: original income = employment and investment income before government intervention; gross income = original income plus cash benefits; disposable income = gross income minus direct taxes; post-tax income = disposable income minus indirect taxes; final income = post-tax income plus benefits in kind (e.g. health, education).

Source: Paul Johnson, 'The welfare state', in R. Floud and D. N. McCloskey (eds), *The Economic History of Britain Since 1700*, Volume 3, 1939–1992 (2nd edn, Cambridge 1994), p. 306. Based on *Economic Trends* (1990), no. 439, p. 88.

Quintiles are useful in Figure 4.6 in giving a clear idea of the differential effects of incomes, taxes and benefits across the spectrum of income distribution without clouding the diagram with an excessive amount of data that would add little to the point being made.[9] What is actually being measured here are the **Gini coefficients** at each stage of the income/tax/benefits process. As explained earlier, the Gini coefficient is a summary measure of distributional equality between social groups. A Gini coefficient of 0 would denote absolute equality (the top 1 per cent and the bottom 1 per cent and all percentiles

Table 4.11 Gini coefficients for the distribution of income at each stage of the tax–benefit system, 1975–1987

	Year			
Gini coefficients (%)	*1975*	*1979*	*1983*	*1987*
Income type:				
original	43	45	49	52
gross	35	35	36	40
disposable	32	33	33	36
post-tax	33	35	36	40
final	31	32	33	36

Note: For definitions of income types, see Figure 4.6

Source: Paul Johnson, 'The welfare state', in R. Floud and D. N. McCloskey (eds), *The Economic History of Britain Since 1700*, Volume 3, 1939–1992 (2nd edn, Cambridge 1994), p. 305. Based on *Economic Trends* (1990), no. 439, p. 118.

in between each receive one per cent of total income). A coefficient of 100 indicates total inequality (the top 1 per cent receive all the income, the rest get nothing).[10] The Gini coefficients relating to the data in Figure 4.6 are given in Table 4.11. They show that inequality grew between 1975 and 1987 and that this was a result of changes in original income (in turn affected by rising unemployment), rather than in the structure of taxes and benefits. Inequality has continued to grow on trend since the 1980s with the tax and benefit system having a much less positive impact on the distribution.[11] It is important to note that because Gini coefficients are independent of the original units of measurement (£, $ etc.), they are very useful for cross-national and well as cross-class comparisons.

Example 4: The age of leaving home in the USA in the twentieth century

Good use is made of medians and ranges in a study of factors determining the age of leaving home in the US in the twentieth century, by Myron Gutmann and his co-authors,[12] in an article used as an exercise for readers later in this volume. Figure 4.7 is drawn from the article. It shows changes in the median age of leaving home by gender and race, between 1880 and 1990. The data is drawn from decennial census information about the numbers of young people remaining in the parental home so the information is derived from static benchmark evidence that has limitations. The statistical evidence is given in Table 4.12. Medians are here used instead of the mean or the mode because this is likely to best reflect average experience. The age range is generally fairly narrow but there is no clear mode over time and there are outliers (those who left home very early and those who stayed permanently) which would impact upon the mean measure.

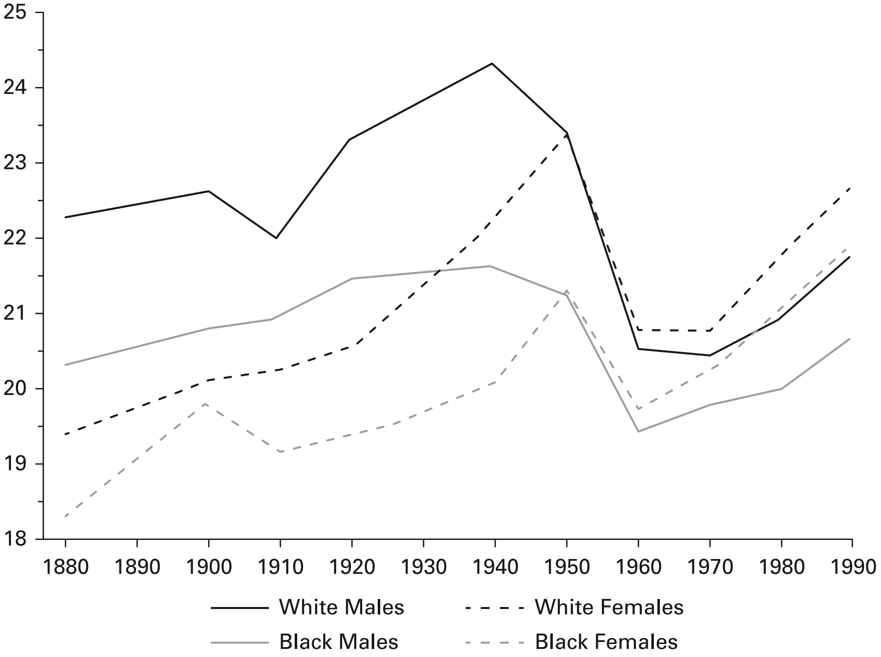

Figure 4.7 Median age at leaving home, United States, 1880–1990.
Source: M. Gutmann, S. Pullum-Piñón and T. Pullum, 'Three eras of young adult home leaving in twentieth-century America', *Journal of Social History*, 35 (3), (2002), p. 534. © Oxford University Press.

Figure 4.7 is the clearest way of demonstrating chronological change but it omits any indication of change in the range of these distributions, which is indicated in the estimations in Table 4.12. It is important to note that this table details the number of years lived with one or both parents rather than the age of leaving home because of the nature of the census evidence. Clearly in some periods and for some groups the experience of leaving home was more age-clustered than at other times. Can you spot these variations?

Studies of the age of leaving home, and the reasons for it, ideally require longitudinal evidence (such as detailed life histories) but this is rarely available for a sizeable or representative sample of the whole population so Gutmann et al. have calculated various probabilities that young people left home at particular ages based upon the ages at which most people were found co-residing with parents at the decennial census. It is not ideal but these simple statistical techniques have made it possible to use cross-sectional data to good effect. This sort of technique is called **Logit analysis**. Logit analysis was originally developed by the marketing industry to assess the scope of customer acceptance of a product, particularly a new product. It attempts to determine customers' purchase intentions and translates that into a measure of actual buying behaviour. In this case the major determinants of leaving home are compared

Table 4.12 Estimated quartiles for number of years lived with one or both parents, United States, 1880–1990

	1880	*1900*	*1910*	*1920*	*1940*	*1950*	*1960*	*1970*	*1980*	*1990*
White Males										
1st Quartile	18.5	18.8	18.9	20.0	20.9	20.6	18.5	18.6	18.8	19.1
2nd Quartile (Median)	22.3	22.6	22.0	23.3	24.3	23.4	20.5	20.4	20.9	21.7
3rd Quartile	26.2	27.0	26.5	28.0	28.9	27.5	23.3	22.9	23.6	25.3
White Females										
1st Quartile	15.2	16.6	17.4	17.9	18.4	19.5	17.8	18.0	18.7	18.5
2nd Quartile (Median)	19.4	20.1	20.2	20.5	22.2	23.3	20.7	20.7	21.7	22.6
3rd Quartile	22.9	23.7	23.7	24.1	26.5	27.9	24.6	24.4	25.9	*
Black Females										
1st Quartile	17.2	17.6	17.9	18.5	18.8	18.8	17.8	18.1	18.2	18.5
2nd Quartile (Median)	20.3	20.8	20.9	21.4	21.6	21.2	19.4	19.7	19.9	20.6
3rd Quartile	24.6	25.2	25.4	26.1	26.2	24.8	21.5	21.8	22.1	23.4
Black Females										
1st Quartile	15.9	16.2	16.6	16.9	17.1	18.0	17.2	17.8	18.4	18.5
2nd Quartile (Median)	18.3	19.8	19.1	19.3	20.0	21.3	19.7	20.2	21.0	21.8
3rd Quartile	21.3	22.8	22.6	22.6	24.3	27.1	23.1	23.3	24.6	27.2

*estimate unavailable

Source: M. Gutmann, S. Pullum-Piñón and T. Pullum, 'Three eras of young adult home leaving in twentieth-century America', *Journal of Social History*, 35 (3), (2002), pp. 533–576, p. 558. © Oxford University Press.

chronologically with the number, age and nature of those co-resident in the census data and this is used to infer real decisions about leaving home and the changes in the age of leaving home for different gender and racial groups. You can test your understanding of this article and the statistical techniques employed, as well as learning about the manifold and changing determinants of leaving home (education, orphanhood, military service, employment prospects, wage levels, changing age and importance of marriage and first cohabitation, immigration status and so on) by doing the exercise in the section following Chapter 7.

DISTRIBUTIONS

We have seen that distributions can cover a very wide range of values or they can be made up of numbers that are clustered closely together. Distributions also take on different shapes, tending towards symmetry, or a skew.

The normal distribution

There is an ideal-type of distribution, known as the **normal distribution**, which is used in statistical theorizing. The expression ideal-type is generally used to indicate a phenomenon which does not occur exactly in practice but has characteristics commonly found in real phenomena.[13] Thus normal distributions rarely occur exactly in social or historical data but in large-scale distributions and especially in the natural sciences the binomial, bell-shaped distribution is the one to which real data distributions often tend (as in Figure 4.2 which gave the heights of US passport applicants). In the normal distribution the mean, the median and the mode have the same value with an equal number of observations spread out symmetrically on either side. The normal distribution, as we shall see in Chapter 7, is also the basis of sampling theory in statistics. It is thus useful to know about the properties of the normal distribution.

In a normal distribution a constant proportion of cases lie between the mean and multipliers of the standard deviation from the mean:

68.26 per cent fall between one standard deviation above and below;

95.46 per cent fall between 2 standard deviations above and below;

99.7 per cent fall between 3 standard deviations above and below.

The **normal distribution** can be represented graphically as shown in Figure 4.8.

The bell-shaped curve of the normal distribution underlies much theorizing about statistics and probability. The 'average man' whose characteristics were described by Adolphe Quetelet (1796–1874) was conceived and recorded in this way in all his physical attributes:

> After his study of heights, Quetelet continued his measurements of other physical attributes: arms and legs, skulls and weights, for which he still observed distributions in accordance with binomial law. From this he inferred the existence of an ideal average man, in whom all average characteristics were combined and who constituted the Creator's goal – perfection.[14]

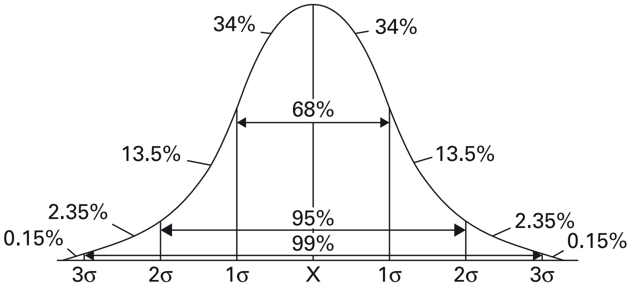

Figure 4.8 The normal distribution.

Note: σ = standard deviation; X = mean, median and mode.

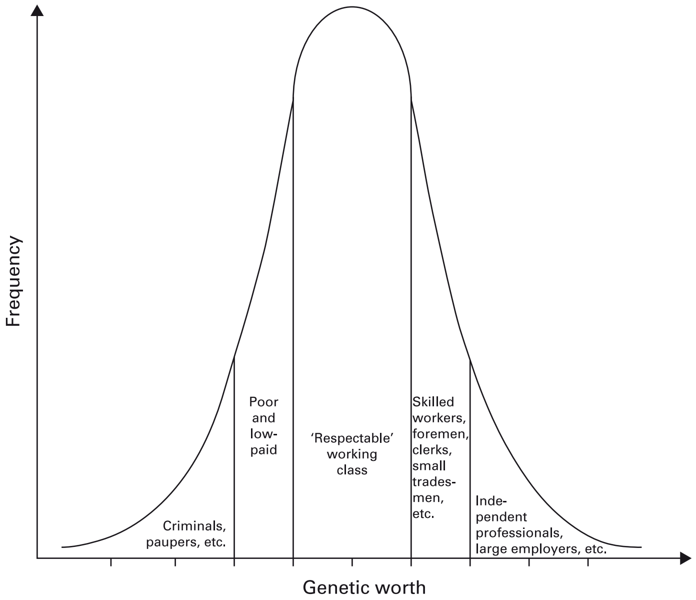
Frequency
Criminals, paupers, etc.
Poor and low-paid
'Respectable' working class
Skilled workers, foremen, clerks, small tradesmen, etc.
Independent professionals, large employers, etc.
Genetic worth

Figure 4.9 Social classes and genetic worth (Galton 1909).
Source: Alain Desrosières, *The Politics of Large Numbers: A History of Statistical Reasoning* (Cambridge, MA 1998), p. 114.

The 'perfection' approached by Quetelet in this way was the normal or binomial distribution. Galton was influenced by Quetelet, by the social investigations of Charles Booth and by Darwinian theories of evolution in theorizing the distribution of 'genetic worth' as a normal curve (see Figure 4.9).

Skewed distributions

Other distributions commonly occur with historical and social data where the spread of observations is uneven with more lying either above or below the mean. Where most observations lie below the mean the distribution is described as **positively skewed** (Figure 4.10a). Where most observations lie above the mean the distribution is described as **negatively skewed** (Figure 4.10b). These distributions are represented graphically below with the relative positions of the mode and the median as well as the mean indicated. It is easy to see why the mean is not always a good measure to use for the average of a **skewed distribution** and it is usual in these cases to give the value of all three averages.

The distribution of the land tax payers of Sowerby given in Table 3.18 is skewed in favour of those paying under £5 and could be roughly drawn as shown in Figure 4.11.

Another example of a skewed distribution is the age profile of cotton workers in the nineteenth century as indicated in Figure 4.12. Young people were favoured as employees with the female workforce appearing particularly youthful. There was however a long 'tail' of older workers.

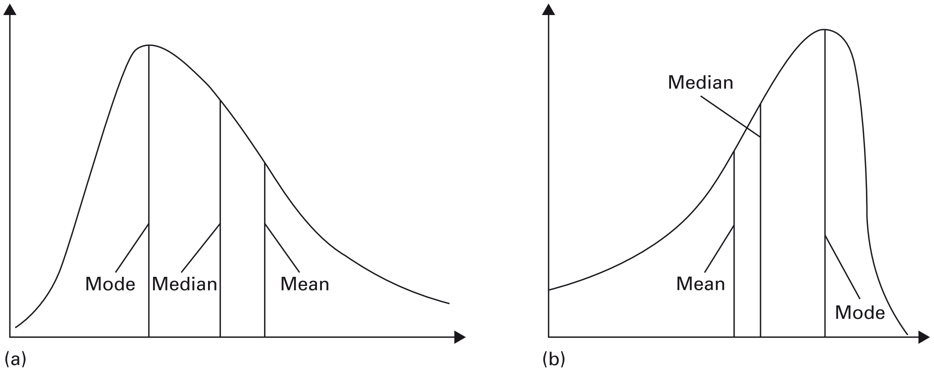

Figure 4.10 Skewed distributions: (a) positive skew (mode and median less than mean); (b) negative skew (mode and median more than mean).

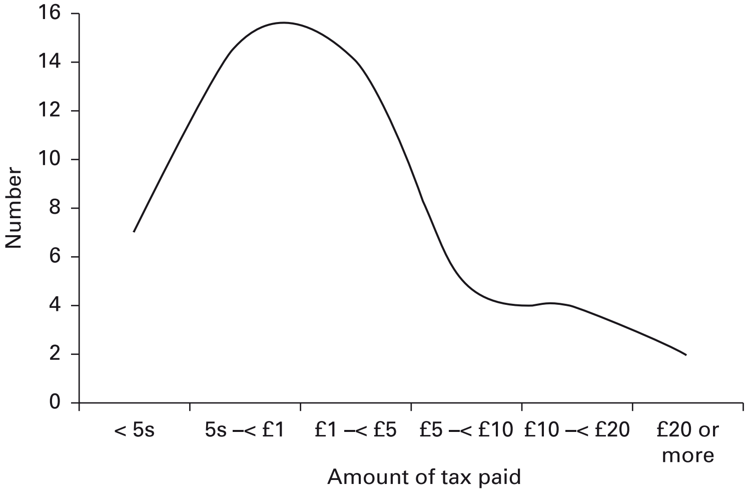

Figure 4.11 Distribution of land tax payers, Sowerby, West Yorkshire, 1782.
Source: see Table 3.18.

In our example in Chapter 3 of prisoners in Portland Prison in 1849 sentence lengths have a negative rather than a positive skew. The mean (13.5) is lower than the median (14) or the mode (15) (Figure 4.13).

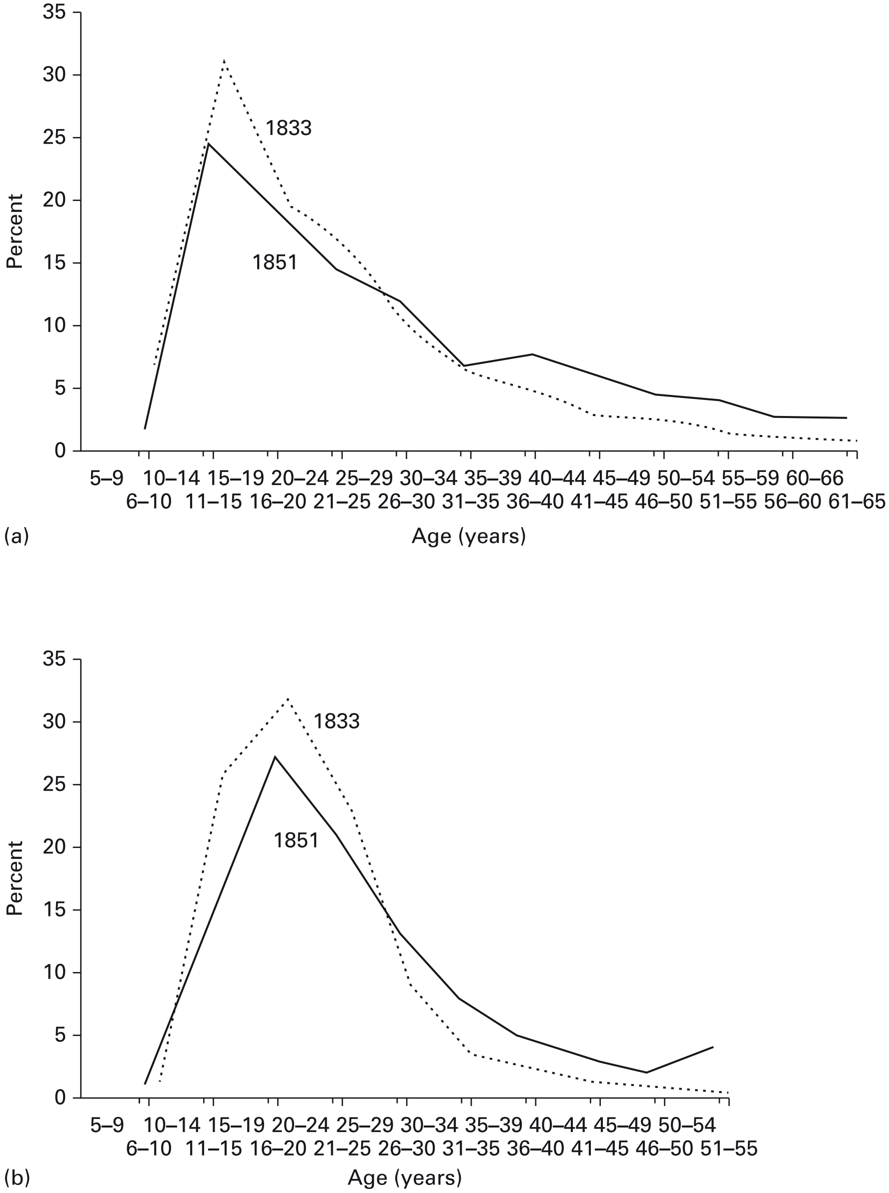

Figure 4.12 Age distribution of the Lancashire cotton industry workforce in 1833 and 1851: (a) males; (b) females.

Source: H. M. Boot, 'How skilled were Lancashire cotton factory workers in 1833?', *Economic History Review*, 48, 2 (1995), p. 286.

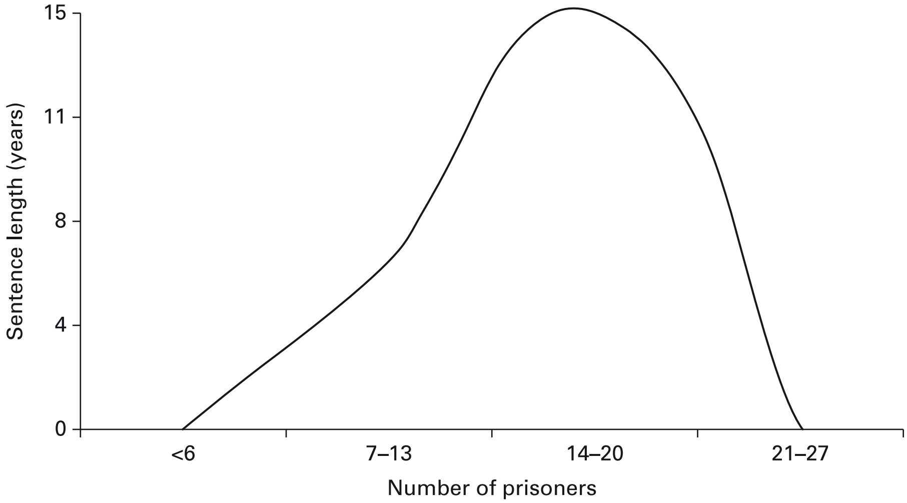

Figure 4.13 Distribution of sentence lengths, Portland Prison, 1849.
Source: Table 3.12.

Distributions with more than one mode

Sometimes distributions occur where there is more than one value around which observations cluster. The occurrence of such distributions illustrates the importance of studying the distribution carefully and perhaps graphing it or drawing a histogram or frequency polygon before rushing to select and calculate an average measure. Figures 4.14 and 4.15 show bi-modal and tri-modal distributions respectively.

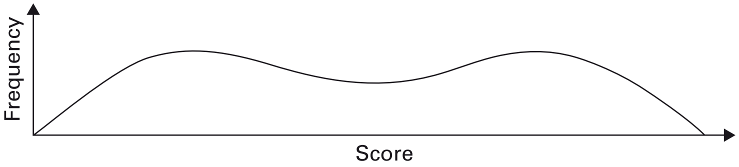

Figure 4.14 Bi-modal distribution.

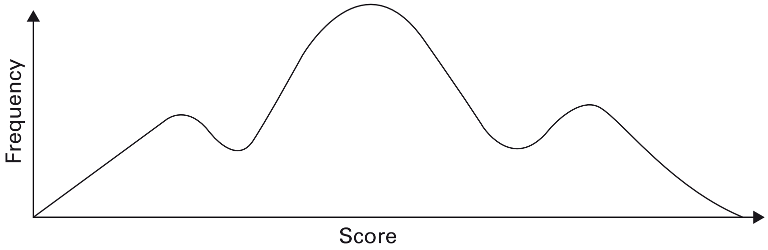

Figure 4.15 Tri-modal distribution.

Conclusion

The most common piece of elementary statistical analysis involves summarizing and considering the nature of a distribution or distributions of values. Measures of central tendency and of dispersion, together with the possibilities presented by graphing the distribution go a long way toward making sense of data and enabling one to compare one distribution with another. These calculations and techniques are important in themselves but also as a preliminary to further, more sophisticated, analysis.

Further reading

Daly, F., D. J. Hand, M. C. Jones, A. D. Lunn and K. J. McConway, *Elements of Statistics* (Harlow 1995), Chapter 1.

Feinstein, Charles, *Making History Count: A Primer in Quantitative Methods for Historians* (Cambridge 2009).

Foster, Liam, Ian Diamond and Julie Jeffries, *Beginning Statistics: An Introduction for Social Scientists,* 2nd edition (London 2015), Chapters 4–7.

Gonick, Larry and Woollcott Smith. *The Cartoon Guide to Statistics* (New York 1993).

Hanagan, T., *Mastering Statistics*, 3rd edition (London 1997), Chapters 4 and 5.

Haskins, Loren and Kirk Jeffreys, *Understanding Quantitative History* (Cambridge, MA 1991), Chapters 1–2.

Solomon, R. and P. Winch, *Calculating and Computing for Social Science and Arts Students* (Buckingham 1994), Chapter 4.

Tufte, E., *Visual Explanations: Images and Quantities, Evidence and Narrative* (Cheshire, CT 1997).