GRAPHICAL REPRESENTATIONS OF DATA

Graphic representation is often of great help in enabling us to comprehend the essential features of the data. Although a graph does not contain any new information, the pictorial presentation of the data often makes it easier to observe its important features. Two procedures for accomplishing this are discussed below.

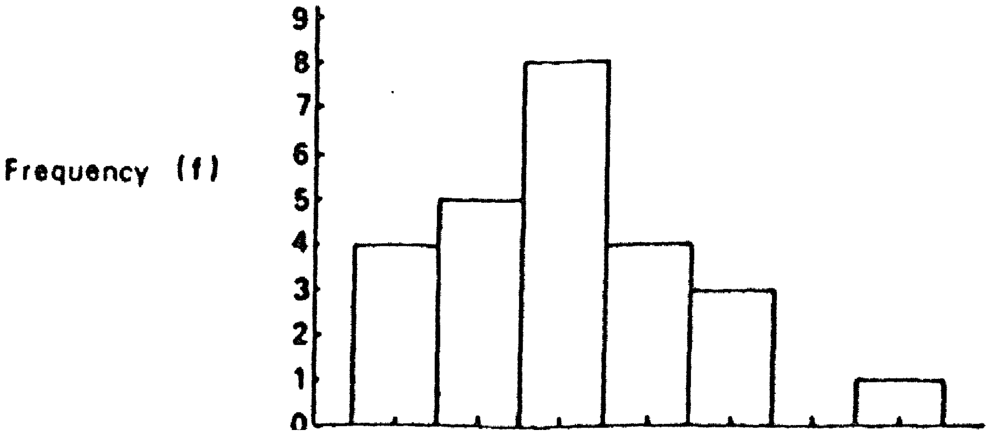
Histograms (Bar Graphs)

Suppose that a sample of 25 families is obtained and the number of children in each family is recorded. These data are presented in Table 3.1.

Table 3.1: Number of children in a sample of 25 families (hypothetical data)

x	f
7	**0**
6	**1**
5	**0**
4	**3**
3	**4**
2	**8**
1	**5**
0	**4**

A histogram, or bar graph of these data is shown in Figure 3.1. To construct the histogram, the y axis (vertical) is marked off in terms of frequencies and the x axis (horizontal) is marked off in terms of score values. To construct a histogram from a grouped frequency distribution, simply use the midpoint of each class interval as the score for that interval.



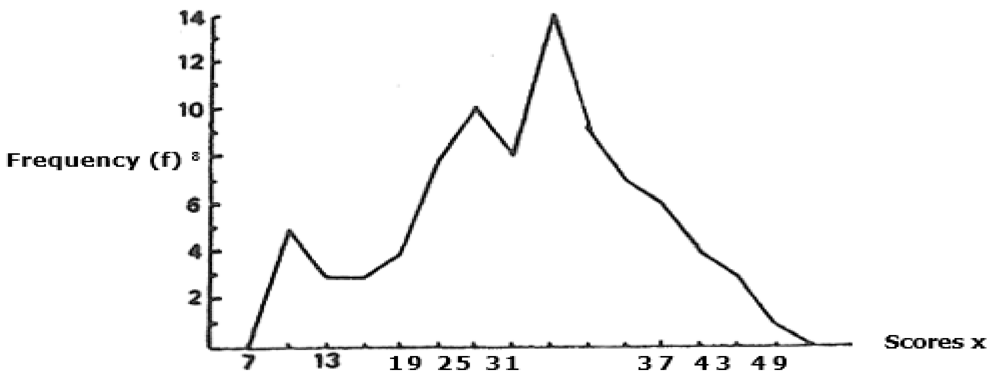
Line Graphs (Polygons)

Often, however, we wish to reflect the continuous nature of our data. In this case we would like to illustrate differences from segment to segment as they occur in a steady fashion. We could do this by placing a point at the centre of the top of each bar in Figure 3.1, and connecting these points with lines. The resulting figure is called a line graph or polygon.

As an example, go back to one of the examples already given, the data represented in Table 2.5, shown again below.

Class internal	Frequency (1)	Cumulative Frequency (cf)
51-53	0	85
48-50	1	85
45-47	3	84
42-44	4	81
39-41	6	77
36-38	7	71
33-35	9	64
30-32	14	55
27-29	8	41
24-26	10	33
21-23	8	23
18-20	4	15
15-17	3	11
12-14	3	8
9-11	5	5
6-8	0	0

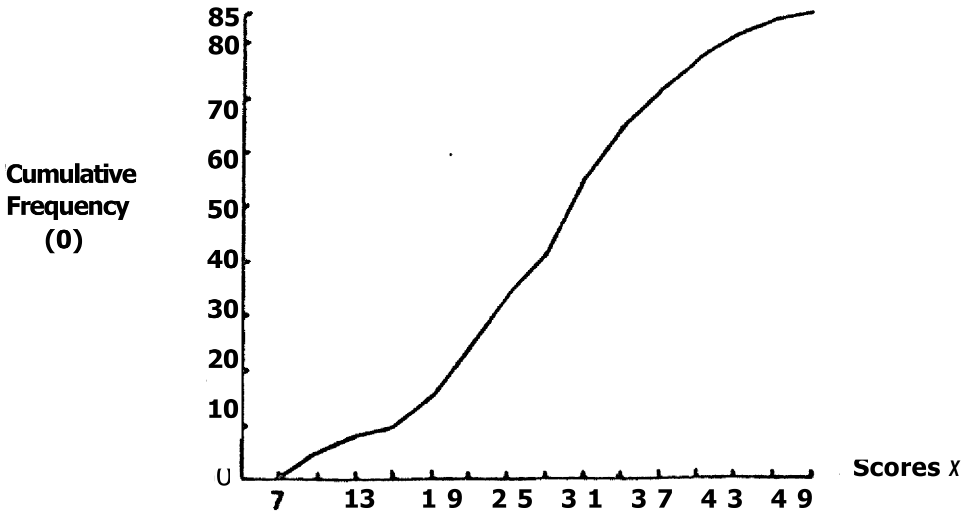
Figure 3.2 Line graph for data in Table 2.5



Cumulative frequency polygons

Cumulative frequency distributions are also commonly graphed in the form of frequency polygons and the resulting figure is called (not very surprisingly) a cumulative frequency polygon. An example, based on the data in Table 2.5 is provided in Figure 3.3. Note the reading from left to right, the cumulative frequency distribution always remains level or increases, and can never drop down towards the x axis: this is because the cumulative frequencies are formed by successive additions and the cf for an interval can be equal to, but never less than, the cf for the preceding interval.

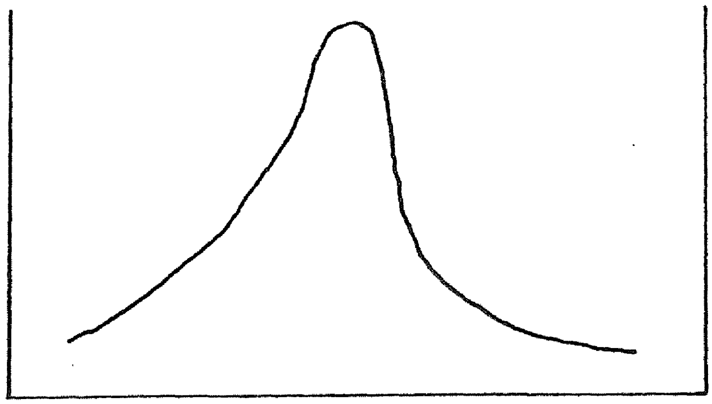
Figure 3.3: Cumulative frequency distribution



SMOOTH CURVES

It is common when graphically representing a distribution to smooth out the curve (especially if we are trying to indicate how the distribution probably looks in the population, since experience shows that population distributions in psychology are often smooth curves or fairly smooth ones). The line graph in Figure 3.2, for instance, might be smoothed to yield Figure 4.1.

Figure 4.1: A smoothed curve of the data in Figure 3.2 below



Special Properties of Smoothed Cury

Such curves have a number of properties which need to be noted:

1. They may be symmetrical or asymmetrical. **A** symmetrical curve is identical in shape on both sides of its centre line (the left-hand side is a mirror image of the right hand, and vice versa. Figure 4.2 shows all **asymmetrical** curve.

   Figure 4.2: An example of an asymmetrical curve

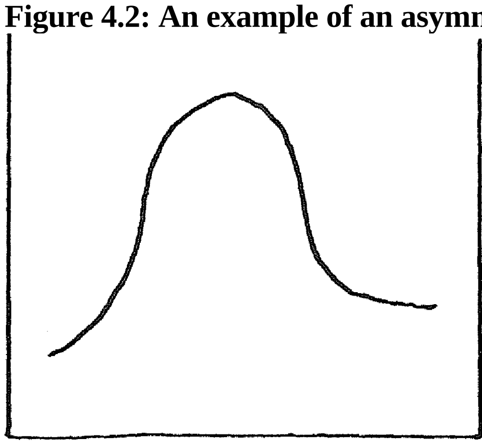
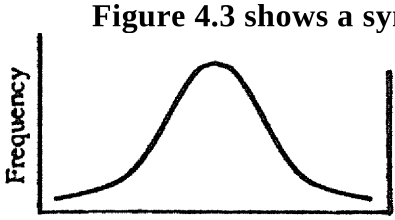


Figure 4.3 shows a symmetrical curve below.



2. They may display skew. A skewed curve has a "tail" at either the left or right hand end. If the tail is at the right hand end (where larger numbers and positive numbers are usually placed along the x axis), the curve is positively skewed, if it is at the left hand end (wherer small or negative numbers are usually placed) it is negatively skewed. The curve in Figure 4.4 **is** positively skewed, that in Figure 4.5 is negatively skewed.

Figure 4.4: A positively skewed curve

Score

Figure 4.5: A negatively skewed curve

Score

3. They may display kurtosis. A curve which is very steep with a high centre is leptokurtic, one which is flat, with a flat centre, is platykurtic, and one which is nicely proportioned is mesokurtic. Figure 4.6 shows a leptokurtic curve and Figure 4.7 a platykurtic one.

Figure 4.6: A leptokurtic curve

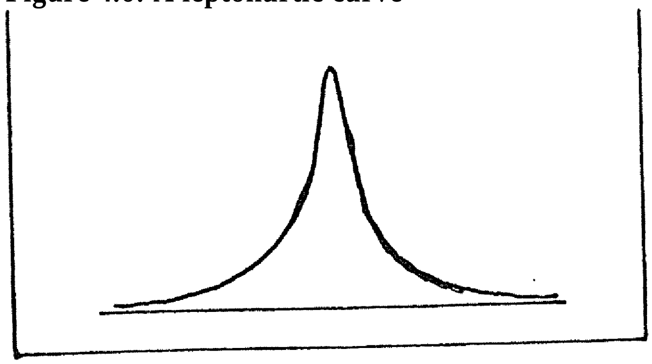
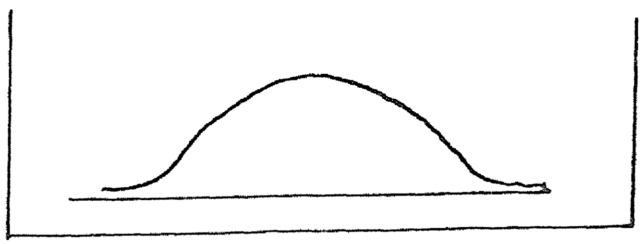


Figure 4.7: A platykurtic curve

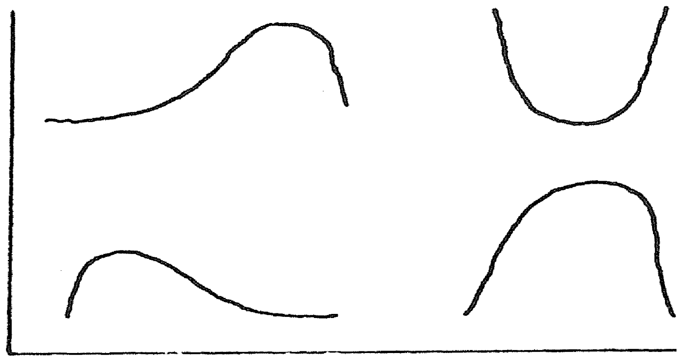


Note that a curve can be both skewed and lepto- or platykurtic at the same time, but it cannot be symmetrical and skewed. It can also be simultaneously symmetrical and lepto- or platykurtic.

Other frequently occurring curves

Several curves occur commonly when psychological data are presented graphically. Among these are the J-curve, actually an example of a skewed curve, and the U-curve, which may occur in the form of a U or an inverted U. Such curves are represented in Figure 4.8.

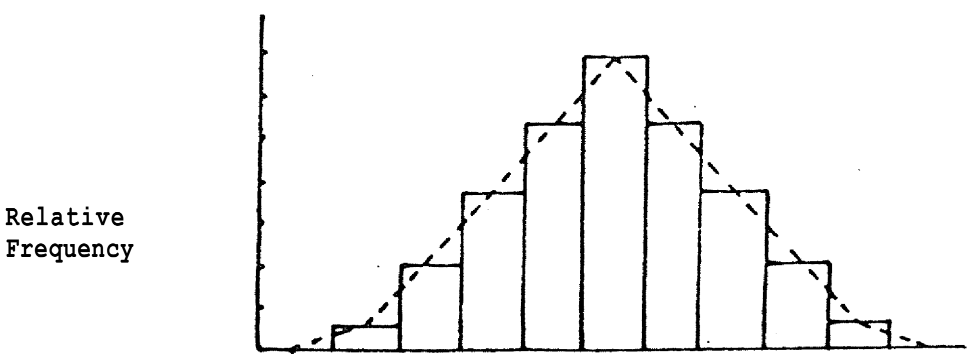
Figure 4.8: and U-curves



The normal distribution

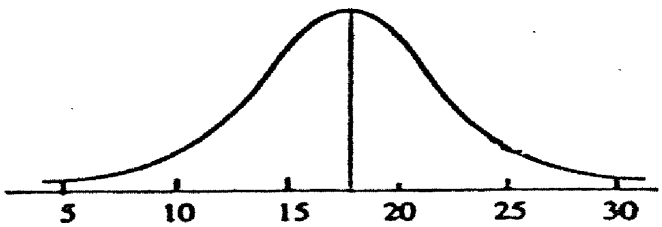
When we measure natural phenomena, we find that most individuals score around the average, or not too far from it. A few individuals, however, will deviate noticeably from the average, and a very few will be markedly different from average. In fact, when large amounts of data are collected through physical and psychological measurements, their frequency polygon often resembles a vertical cross section of **a** bell. Many naturally occurring distributions approximate this shape which is referred to as a "normal distribution". Figure 4.8 shows a distribution which is bell shaped i.e., a normal distribution.

Figure 4.8: An example of hypothetical data with a bell shaped bar chart



If the frequency polygon for a given set of observations which has a normal distribution is made into a smooth curve, the resulting curve is referred to as a normal curve (see Figure 4.9)

Figure 4.9: A normal curve



The main features of a normal distribution (as represented by the "normal curve") are:

1. It is bell shaped
2. It is symmetrical about the central value (this means automatically that it is not skewed).
3. It is mesokurtic
4. It is asymptotic to the horizontal axis; in other words, the curve gets closer and closer to the horizontal axis as it moves outward from the mean. It approaches, but does not meet the horizontal axis and extends from minus infinity to plus infinity.

   There are two ways of explaining the asymptotic nature of the distribution;

   (i) the simple explanation; it is nearly always possible to find that one more very rare and unusual case.

   (ii) the mathematical explanation; the normal distribution is a natural logarithmic function and the natural log (i.e., log to the base e) of zero is an undefined value.

   It should be noted that there is virtually no naturally occurring system which follows an exact normal distribution. This is especially so for smaller samples which often do not conform so neatly to the shape of a bell, although its general characteristics do appear; that is, most cases pile up near the centre of the distribution and fewer cases are found near the extremes. Even though the actual distribution of scores in a sample taken from some population may not be normally distributed, we can still apply the characteristics of the normal curve as long as we can assume normality in the population from which the sample was taken. For much of the data in the social sciences, this is a valid assumption. If we can assume normality, then many useful statements may be made regarding the proportions of the sample which lie in appropriate intervals.