7 The Variability of Observations

In the previous chapter we have calculated and discussed an average value for a short series of observations and for one large enough to require grouping in a frequency distribution. But the very fact that we thought it necessary to calculate an average, to define the general position of a distribution, introduces the idea of *variation* of the individual values round that average. For if there were no such variation, if, in other words, all the observations had the same value, then there would be no point in calculating an average. In introducing and using an average, usually the arithmetic mean, we therefore ignore – for the time being – that variability of the observations. It follows that, taken alone, the mean is of very limited value, for it can give no information regarding the variability with which the observations are scattered around itself, and that variability (or lack of variability) is an important characteristic of the frequency distribution. As an example Table 8 shows the frequency distributions of some recorded ages at death from two causes of death amongst women. The mean, or average, age at death, does *not* differ greatly between the two, being 37·2 years for the deaths registered as due to diseases of the Fallopian tube and 35·2 years for those attributed to abortion. But both the table and the diagram based upon it (Fig. 10; p. 76) show that the difference in the variability, or scatter, of the observations round their respective means is very considerable. With diseases of the Fallopian tube the deaths are spread over the age-groups 0–4 to 70–74, while deaths from abortion range only between 20–24 and 45–49. As a further description of the frequency distribution, we clearly need a measure of its degree of variability round the average. A measure commonly employed in medical (and other) papers is the *range*, as quoted above – i.e. the distance between the smallest and greatest observations. Though this measure is often of considerable interest, it is not very satisfactory as a description of the general variability, since it is based upon only the two extreme observations and ignores the distribution of all the observations within those limits – e.g. the remainder may be more evenly spread out over the distance between the mean and the outlying values in one dis-

TABLE 8

FREQUENCY DISTRIBUTION OF SOME RECORDED DEATHS OF WOMEN ACCORDING TO AGE FROM (1) DISEASES OF THE FALLOPIAN TUBE, AND (2) ABORTION

Age in Years	Diseases of the Fallopian Tube	Abortion
0–	1	—
5–	—	—
10–	1	—
15–	7	—
20–	12	6
25–	35	21
30–	42	22
35–	33	19
40–	24	26
45–	27	5
50–	10	—
55–	6	—
60–	5	—
65–	1	—
70–74	2	—
Total	206	99

tribution than in another. Also the occurrence of the rare outlying values will depend upon the number of observations made. The greater the number of observations the more likely is it that the rare value will appear amongst them. As a result differences between the ranges recorded in two similar investigations may arise solely from a differing total of observations. They will give a distorted view of the variability found in the two inquiries. Thus, in publishing observations, it is certainly insufficient to give only the mean and the range; as previously pointed out, the frequency distribution itself should be given when possible – even if no further calculations are made from it.

The Standard Deviation

The further calculation that the statistician invariably makes is of the Standard Deviation, which is a measure of the scatter of the observations around their mean. Put briefly, the development of this particular measure is as follows. Suppose we have, as given in Table 9, twenty observations of systolic blood pressure made on twenty different persons. The mean, or average, blood pressure is the sum of the observations divided by 20 and equals 128 mm. It is obvious from cursory inspection that the variability of the individual values around this mean is considerable. They range from 98 to 160; on the other hand, a large proportion of the values lies in the narrower range 125–135 (50 per cent of them). The mean and range are not sufficient to describe the distribution adequately. As a further step we may calculate the amount by which each

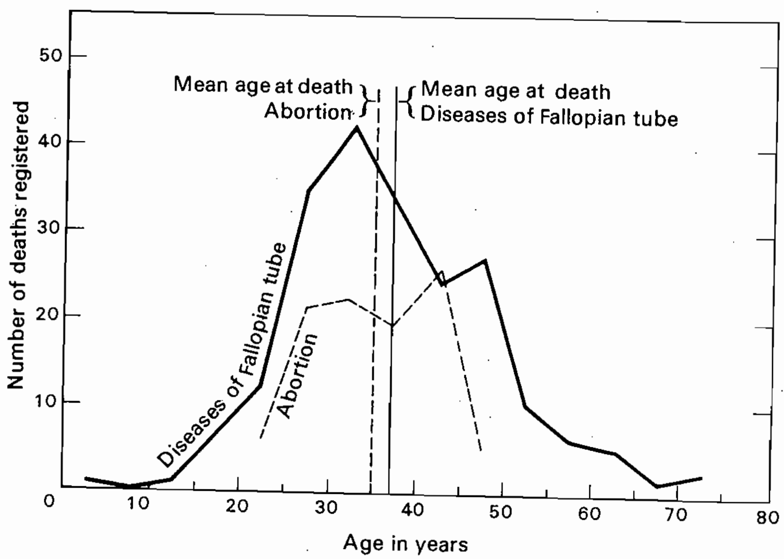

Fig. 10. Frequency distributions of some recorded deaths of women from (*a*) diseases of the Fallopian tube and (*b*) abortion.

observation differs, or deviates, from the average, as is shown in column (2). If these differences are added, ***taking their sign into account***, the sum must equal nought, for a characteristic of the arithmetic mean, or average, is that the sums of the positive and of the negative deviations of the observations from itself are equal. In this example the sum of the deviations above the mean is +93 and below the mean −93. Two ways of avoiding this difficulty are possible: we may add all the deviations, ignoring their sign, or we may square each deviation so that each becomes positive. If the deviations be added in the example, ignoring their sign,

the sum is 186 and the ***mean deviation*** is, therefore, 186 divided by 20 (the number of observations) or 9·3 mm. This is a valid measure of the variability of the observations around the mean, but it is one which, for reasons involved in the problems of sampling, discussed later, is of less value in the analysis of numerical data than the ***standard deviation***. To

TABLE 9

Twenty Observations of Systolic Blood Pressure in mm	Deviation of each Observation from the Mean (Mean = 128)	Square of each Deviation from the Mean
(1)	(2)	(3)
98	− 30	900
160	+ 32	1024
136	+ 8	64
128	0	0
130	+ 2	4
114	− 14	196
123	− 5	25
134	+ 6	36
128	0	0
107	− 21	441
123	− 5	25
125	− 3	9
129	+ 1	1
132	+ 4	16
154	+ 26	676
115	− 13	169
126	− 2	4
132	+ 4	16
136	+ 8	64
130	+ 2	4
Sum 2560	0	3674

reach the latter the squared deviations are used. Their sum is 3674, so that the mean squared deviation is this sum divided by 20, which equals 183·7. This value is known as the ***Variance***. The standard deviation is the square root of this value (for having squared the original deviations the reverse step of taking the square root must finally be made to restore the original units) and in this example is, therefore, 13·55 mm.

This is, in truth, the standard deviation of *these particular* 20 observations. We are, however, in practice almost invariably using such a set of observations to allow us to estimate the variability *in the population, or universe,* of which they merely form a sample. For this purpose it can be shown that on the average a slightly better estimate of the standard deviation in the population is reached by dividing the observed sum of the squared deviations from the mean by one less than the total number of observations in the sample, i.e. by $n - 1$ instead of by n* (see also p. 93). Thus in the present instance we should calculate as the variance 3674/19, or 193·4, and the standard deviation is the square root of 193·4 or 13·91 mm. The importance of this step is obviously greater when, as in the present instance, the number of observations is quite small.

Turning to the meaning of the result, a large standard deviation shows that the frequency distribution is widely spread out from the mean, while a small standard deviation shows that it lies closely concentrated about the mean with little variability between one observation and another. For example, the standard deviation of the widely spread age distribution of deaths attributed to diseases of the Fallopian tube (see Table 8) is 11·3 years, while of the more concentrated age distribution of deaths attributed to abortion it is only 6·8. The frequency distributions themselves clearly show this considerable difference in variability. The standard deviations have the advantage of summarising this difference by measuring the variability of each distribution in a single figure; they also enable us to test, as will be seen subsequently, whether the observed differences between two such means and between two such degrees of variability are more than would be likely to have arisen by chance.

In making a comparison of one standard deviation with another it must, however, be remembered that this criterion of variability is measured in the same units as the original observations. The mean height of a group of school-children may be 48 inches and the standard deviation 6 inches; if the observations were recorded in centimetres instead of in inches, then the mean would be 122 cm and the standard deviation 15·2 cm. It follows that it is not possible by a comparison of the standard deviations to say, for instance, that weight is a more variable characteristic than height; the two characteristics are not measured in the same units and the selection of these units – e.g. inches or centimetres, pounds or kilogrammes – must affect the comparison. In fact, it is no more helpful to compare these standard deviations than it is to compare the mean height with the mean weight. Further, a standard deviation of 10 round a mean of 40 must indicate a relatively greater degree of scatter than a standard deviation of 10 round a mean of 400, even though the units of measurement are the same.

* In Table 9 we have 20 observations and therefore 20 deviations from the mean. But when any 19 of these deviations have been calculated the twentieth is predetermined since their total, by definition, must equal 0. We have, therefore, in statistical parlance, lost one 'degree of freedom' and are left with 19.

The Coefficient of Variation

To overcome these difficulties of the comparison of the variabilities of frequency distributions measured in different units or with widely differing means, the Coefficient of Variation is utilised. This coefficient is the standard deviation of the distribution expressed as a percentage of the mean of the distribution – i.e. Coefficient of Variation = (Standard Deviation ÷ Mean) × 100. If the standard deviation is 10 round a mean of 40, then the former value is 25 per cent of the latter; if the standard deviation is 10 and the mean is 400, the former value is 2·5 per cent of the latter. These percentage values are the coefficients of variation. The original unit of measurement is immaterial for this coefficient, since it enters into both the numerator and the denominator of the fraction above. For instance, with a mean height of 48 inches and a standard deviation of 6 inches the coefficient of variation is (6/48) × 100 = 12·5 per cent. If the unit of measurement is a centimetre instead of an inch, the mean height becomes 122 cm, the standard deviation is 15·2 cm and the coefficient of variation is (15·2/122) × 100 = 12·5 per cent again. Similarly the coefficient of variation of the blood pressures of Table 9 is (13·55/128) × 100 = 10·6 per cent.

These measures of variability are just as important characteristics of a series of observations as the measures of position – i.e. the average round which the series is centred. As was said by Udny Yule one of the leading British statisticians of the early years of the 20th century, the important step is to 'get out of the habit of thinking in terms of the average, and think in terms of the frequency distribution. Unless and until he [the investigator] does this, his conclusions will always be liable to fallacy. If someone states merely that the average of something is so-and-so, it should always be the first mental question of the reader: "This is all very well, but what is the frequency distribution likely to be? How much are the observations likely to be scattered round that average? And are they likely to be more scattered in the one direction than the other, or symmetrically round the average?" To raise questions of this kind is at least to enforce the limits of the reader's knowledge, and not only to render him more cautious in drawing conclusions, but possibly also to suggest the need for further work'.

Examples of Variability

The practical application of these measures of variability may be illustrated by the figures tabulated below, which are taken from a statistical study of blood pressure in healthy adult males. In the original the full frequency distributions are also set out.

THE BLOOD PRESSURE IN 566 HEALTHY ADULT MALES
MEANS, STANDARD DEVIATIONS, COEFFICIENTS OF VARIATION, AND THE RANGE OF MEASUREMENT

	Mean	Standard Deviation	Coefficient of Variation	Range
Age (years)	23·2	4·02	17·31	18– 40
Heart-rate (beats per minute)	77·3	12·83	16·60	46–129
Systolic BP (mm)	128·8	13·05	10·13	97–168
Diastolic BP (mm)	79·7	9·39	11·78	46–108
Pulse pressure (mm)	49·1	11·14	22·69	24– 82

The variability of these physiological measurements, which is apparently compatible with good health at the time of measurement, is striking. It led the authors to conclude that we must hesitate to regard as abnormal any isolated measurements in otherwise apparently fit individuals. Some of the measurements they found are definitely within the limits usually regarded as pathological, and study is necessary to determine whether such large deviations from the 'normal' have any unfavourable prognostic significance. It is clear that the mean value alone is a very insecure guide to 'normality' (see p. 283).

As a further example of the importance of taking note of the variability of observations, the incubation period of a disease may be considered. If the day of exposure to infection is known for a number of persons we can construct a frequency distribution of the durations of time elapsing between exposure to infection and onset of disease as observed clinically. If these durations cover a relatively wide range, say 10–18 days with an average of 13 days, it is obvious that observation or isolation of those who have been exposed to infection for the *average* duration would give no high degree of security. For security we need to know the proportion of persons who develop the disease on the fourteenth, fifteenth, etc., day after exposure; if these proportions are high – i.e. the standard deviation

of the distribution is relatively large – isolation must be maintained considerably beyond the *average* incubation time. In such a case the importance of variability is indeed obvious; but there is a tendency for workers to overlook the fact that in *any* series of observations the variability, large or small, is a highly important characteristic.

For the beginner, who at first finds the standard deviation a somewhat intangible quantity, it is useful to remember that for distributions that are not very asymmetrical, six times the standard deviation includes about 99 per cent of all the observations. Thus, in the example given above, the standard deviation of the diastolic blood pressures is 9·39, and six times this, or 56 units, should include very nearly all the observations. In fact, the observations lie within $108 - 46 = 62$ units. If the distribution is symmetrical, the mean plus 3 times the standard deviation should give approximately the upper limit of the observations, and the mean minus 3 times the standard deviation should similarly give their lower limit. Thus for the diastolic blood pressure we have $79{\cdot}7 + 28$ and $79{\cdot}7 - 28$, or a range of, approximately, 52 to 108, very close to the observed range of 46 to 108. This rule also serves, it will be seen, as a check upon the calculation of a standard deviation – not, of course, to show that a small error has been made but whether some serious mistake has led to a standard deviation which is quite unreasonable. It must not, however, be expected to hold with a few observations only.

In actual practice the calculation of the standard deviation is not usually carried out by the method shown above – i.e. by computing the deviation of each observation from the mean and squaring it. Shorter methods are available both for ungrouped observations (like the twenty measurements of blood pressure in Table 9) and for grouped observations (like those in Table 8). These methods are described in the next chapter.

Symmetry

Another character of the frequency distribution is its symmetry or lack of symmetry. With a completely symmetrical distribution the frequency with which observations are recorded at each point on the graph, or within certain values below the mean, is identical with the frequency of observations at the same point, or within the same values, above the mean. With asymmetry the observations are not evenly scattered on either side of the mean but show an excess on one side or within particular values – e.g. with a mean of 50, observations below the mean may not fall below 20 units from that point, the lowest observation recorded being 30, while on the positive side of the mean observations 40 units above the mean may be observed, values of 90 being recorded. The tabulated distribution and, still more, a graph of it will afford an *indication* of this characteristic.

The 'Normal' Curve or Distribution

One particular form of a symmetrical distribution is known as the 'normal' curve, or distribution, and this curve has very great importance in statistical theory, being fundamental to the tests of significance discussed in later chapters. It should not, however, be thought that this distribution is the normal in the sense that all measurable characteristics occurring in nature, e.g. of men, animals, or plants, should conform to it. Many do, in fact, show this kind of distribution, but by no means all. Stature is a characteristic in man which does, at least very closely, show such a distribution and we may examine the curve with this as an example. (This approach, it may be noted, is adopted merely for simplicity. The normal, or Gaussian, curve and its characteristics can, of course, be derived mathematically.)

The heights of 1000 adult men (hypothetical figures) are given in Table 10 (p. 83). Calculation shows that the mean height is 172·5 cm and the standard deviation is 5 cm. Examining the frequency distribution, we can see how many men have a height which differs from the mean of all men by not more than 5 cm, i.e. whose height lies between 172·5 − 5 and 172·5 + 5 cm, or, in short, between 167·5 and 177·5 cm. There are 152 + 193 = 345 men between 167·5 and 172·5 cm and 197 + 148 = 345 men between 172·5 and 177·5 cm. There are, therefore, 690 men whose height is not more than 5 cm away from the mean. But it will be noticed that the standard deviation of the distribution is 5 cm. Instead of saying there are 690 men whose stature is within 5 cm of the mean we may therefore say that for 69·0 per cent of all the men the stature does not differ from the mean by more than once the standard deviation (either in the plus or minus direction).

Similarly we may see how many men have statures that lie within 10 cm of the mean, i.e. between 162·5 and 182·5 cm. Between 162·5 and 172·5 there are 43 + 86 + 152 + 193 = 474, and between 172·5 and 182·5 there are 197 + 148 + 91 + 45 = 481; the total is 955. But 10 cm away from the mean is twice the standard deviation and we may therefore say that for 95·5 per cent of all the men the stature does not differ from the mean by more than 2 times the standard deviation in either direction.

Lastly, we may see how many men have a stature that is not more than 15 cm distant from the mean, i.e. between 157·5 and 187·5 cm.

Between 157·5 and 172·5 there are 5 + 17 + 43 + 86 + 152 + 193 = 496 and between 172·5 and 187·5 there are 197 + 148 + 91 + 45 + 16 + 4 = 501, giving a total of 997. But 15 cm away from the mean is three times the standard deviation, and we may therefore say that, for 99·7 per cent of all the men the stature does not differ from the mean by more than 3 times the standard deviation in either direction.

This is a useful way of looking at a frequency distribution, namely to see how many of the observations lie within a given distance of the mean,

TABLE 10

EXAMPLE OF A NORMAL DISTRIBUTION

	Height in cm	Number of Men of given Height	Number of Men whose Heights lie within a given Distance of the Mean		
			Heights between the Mean and a Distance of 1 S.D. away on either Side	Heights between the Mean and a Distance of 2 S.D. away on either Side	Heights between the Mean and a Distance of 3 S.D. away on either Side
− 3 S.D.	155·0–	2			
	157·5–	5			
− 2 S.D.	160·0–	17			
	162·5–	43			
− 1 S.D.	165·0–	86			
	167·5–	152			
	170·0–	193	690	955	997
	172·5–	197			
	175·0–	148			
+ 1 S.D.	177·5–	91			
	180·0–	45			
+ 2 S.D.	182·5–	16			
	185·0–	4			
+ 3 S.D.	187·5– 190·0	1			
	Total	1000			

not in terms of the actual units of measurement but in terms of multiples of the standard deviation.

Returning to the ideal normal frequency distribution, as derived mathematically, its characteristics are: (*a*) the mean, median, and mode all coincide; (*b*) the curve is perfectly symmetrical round the mean; and (*c*) we can calculate *theoretically* how many of the observations will lie in the interval between the mean itself and the mean plus or minus *any* multiple of the standard deviation. This calculation gives the following results:—

Proportion of observations that lie within $\pm$ 1 times the S.D. from the mean.............	68·27 per cent
Proportion of observations that lie within $\pm$ 2 times the S.D. from the mean.............	95·45 per cent
Proportion of observations that lie within $\pm$ 3 times the S.D. from the mean.............	99·73 per cent

It will be seen that with a measurement that follows a normal distribution nearly one-third of the values observed will differ from the mean value by more than once the standard deviation, only about 5 per cent will differ from the mean by more than twice the standard deviation, and only some 3 in 1000 will differ from the mean by more than 3 times the standard deviation. In a normal distribution, in other words, values that differ from the mean by more than twice the standard deviation are fairly rare, for only about 1 in 20 observations will do so; values that differ from the mean by more than 3 times the standard deviation are very rare, for only about 1 in 370 will do so.

These theoretical values, it will be seen, agree very closely with the observed values given by Table 10. In fact, it is unlikely that the stature of 1000 men would follow the normal distribution so very closely and, as already stated, the hypothetical figures of the table were selected deliberately for the present demonstration of the properties of a normal curve of distribution in place of deriving those properties mathematically. The figures, therefore, must *not* be taken as quite true to life, though it is true that stature does follow a fairly normal distribution. The ideal curve and the present hypothetical figures are shown diagrammatically in Fig. 11.

Summary

As descriptions of the frequency distribution of a series of observations certain values are necessary, the most important of which are,

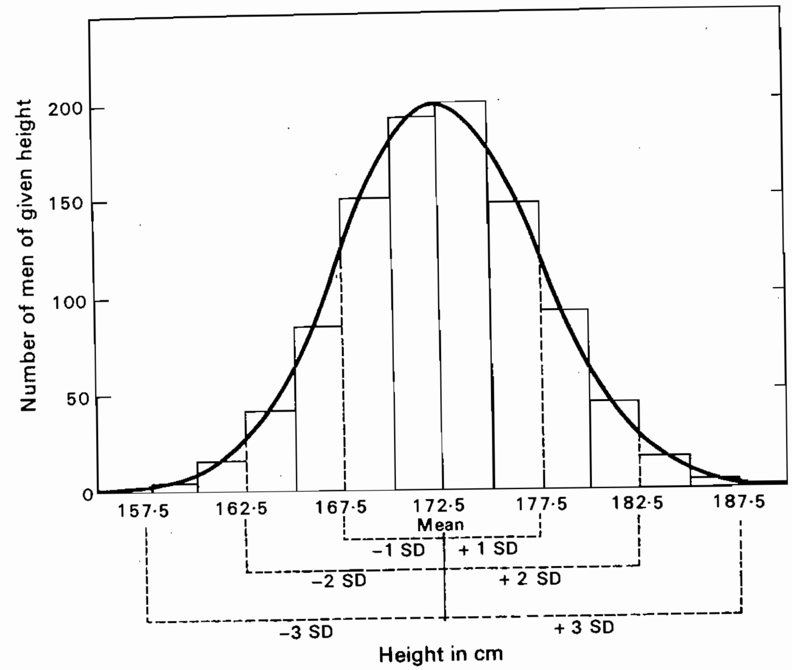

Fig. 11. Histogram of statures of 1000 men (hypothetical figures) and normal curve superimposed.

usually, the mean and standard deviation. The mean alone is rarely, if ever, sufficient. In statistical work it is necessary to think in terms of the frequency distribution as a whole, taking into account the central position round which it is spread (the mean or average), the variability it displays round that central position (the standard deviation and coefficient of variation), and the symmetry or lack of symmetry with which it is spread round the central position. The important step is to think not only of the average but also of the scatter of the observations around it. With a 'normal' distribution only 1 observation in 20 will differ from the mean by more than twice the standard deviation (plus or minus) and only some 3 in 1000 will differ from the mean by more than 3 times the standard deviation (plus or minus).