9 Problems of Sampling: Averages

The observations to which the application of statistical methods is particularly necessary are those, it has been pointed out, which are influenced by numerous causes, the object being to disentangle that multiple causation. It has also been noted that the observations utilised are nearly always only a sample of all the possible observations that might have been made. For instance, the frequency distribution of the stature of Englishmen – i.e. the number of Englishmen of different heights – is not based upon measurements of all Englishmen but only upon some sample of them. The question that immediately arises is how far is the sample representative of the population from which it was drawn, and, bound up with that question, to what extent may the values calculated from the sample – e.g. the mean and standard deviation – be regarded as true estimates of the values in the population sampled? If the mean height of 1000 men is 169 cm with a standard deviation of 7 cm, may we assert that the values of the mean and standard deviation of all the men of whom these 1000 form a sample are not likely to differ appreciably from 169 and 7? This problem is fundamental to all statistical work and reasoning; a clear conception of its importance is necessary if errors of interpretation are to be avoided, while a knowledge of the statistical techniques in determining errors of sampling will allow conclusions to be drawn with a greater degree of security.

Elimination of Bias

Consideration must first be given, as previously noted, to the presence of selection or bias in the sample. If owing to the method of collection of the observations, those observations cannot possibly be a representative sample of the total population, then clearly the values calculated from the sample cannot be regarded as true estimates of the population values, and no statistical technique can allow for that kind of error. That problem was discussed in Chapter 3. In the present discussion we will presume that the sample is 'unselected' and devote attention entirely to the problem of the variability which will be found to occur from one sample

to another in such values as means, standard deviations, and proportions, due entirely to what are sometimes known as the 'errors of sampling.' This is not in fact a very good description. We are not concerned with *error* in the usual sense of that word but with the variability that must occur through the play of *chance*. Attention may first be given to the mean.

The Mean

Let us suppose that we are taking samples from a very large population, or universe, and that we know that an individual in that universe may measure any value from 0 to 9 – e.g. we may be recording the number of attacks of the common cold suffered by each person during a specified period, presuming 9 attacks to be the maximum number possible. The mean number of attacks per person and the standard deviation in the whole population we will presume to be known; let the average number of attacks per person be 4·50 (i.e. the total attacks during the specified period divided by the number of persons in the universe) and let the standard deviation be 2·87 (as found, in the previous two chapters, by calculating how much the experience of each person deviates from the average, finding the average of the squares of these deviations, and the square root of this value). From that universe we will draw, at random, samples of 5 individuals. For each sample we can calculate the mean number of attacks suffered by the 5 individuals composing it. To what extent will these means in the small samples diverge from the real mean – i.e. the mean of the universe, 4·50?

In Table 15 are set out a hundred such samples of 5 individuals drawn at random from the universe. The 'universe' actually used for this and later demonstrations was composed of *Random Sampling Numbers* of the kind shown on pp. 305 to 312.

Sets of unit random numbers were taken in fives, tens, twenties, and fifties as required. For instance, in the first sample of Table 15 there were 3 individuals who had 2 colds each, one fortunate person who had none, and one unfortunate who had 4. From each of these samples a mean can be calculated which, in all, gives one hundred mean values; and of these means we can make a frequency distribution. In the first sample the mean is $(2 + 2 + 2 + 0 + 4) \div 5 = 2{\cdot}0$, in the second it is $(6 + 9 + 7 + 5 + 9) \div 5 = 7{\cdot}2$, and so on. The grouped distribution of these means is given in Table 16, column (2). There was one sample in which the mean was only 1·2 and one in which it was as high as 7·4 (the possible minimum and maximum values are, of course, 0 and 9). A study of this distribution shows:—

TABLE 15

NUMBER OF COLDS SUFFERED BY INDIVIDUALS, VALUES RECORDED IN SAMPLES OF 5 PERSONS (HYPOTHETICAL FIGURES)

Sample No.						Mean of each sample
1	2	4	2	0	2	2·0
2	6	9	7	5	9	7·2
3	3	1	3	1	1	1·8
4	9	1	2	6	9	5·4
5	7	7	3	6	7	6·0
6	5	5	1	8	6	5·0
7	3	1	6	4	8	4·4
8	0	0	3	3	6	2·4
9	5	2	7	8	6	5·6
10	4	3	1	6	6	4·0
11	5	9	2	9	8	6·6
12	2	2	7	2	1	2·8
13	4	3	1	9	1	3·6
14	1	5	7	3	5	4·2
15	2	8	0	2	3	3·0
16	6	6	8	8	7	7·0
17	6	5	8	1	3	4·6
18	9	8	9	5	1	6·4
19	2	2	7	6	7	4·8
20	3	3	8	5	1	4·0
21	5	1	6	2	5	3·8
22	3	6	5	2	4	4·0
23	9	1	5	4	2	4·2
24	6	6	3	7	5	5·4
25	6	2	7	2	1	3·6
26	7	5	3	1	6	4·4
27	7	0	8	5	5	5·0
28	8	0	6	0	4	3·6
29	7	2	6	5	9	5·8
30	2	5	4	9	6	5·2
31	4	2	6	8	7	5·4
32	1	6	4	4	3	3·6
33	3	5	7	4	5	4·8
34	3	9	6	9	9	7·2
35	9	7	1	5	2	4·8
36	4	1	7	3	4	3·8
37	2	5	1	5	4	3·4
38	5	1	9	2	5	4·4
39	9	7	5	3	8	6·4
40	5	8	5	7	7	6·4
41	9	3	9	9	7	7·4
42	7	6	1	9	3	5·2
43	3	6	5	2	2	3·6
44	0	9	0	7	7	4·6
45	2	3	6	4	0	3·0
46	1	1	0	2	7	2·2
47	6	1	4	6	1	3·6
48	9	8	1	7	8	6·6
49	4	4	4	2	1	3·0
50	4	1	4	7	6	4·4
51	9	9	1	0	6	5·0
52	9	9	3	4	3	5·6
53	7	8	5	4	3	5·4
54	0	9	3	1	2	3·0
55	9	4	9	1	8	6·2
56	1	0	5	1	4	2·2
57	0	6	2	5	6	3·8
58	6	7	8	3	9	6·6
59	1	5	8	4	3	4·2
60	9	2	4	7	3	5·0
61	1	7	3	7	8	5·2
62	7	8	2	4	3	4·8
63	3	7	0	5	8	4·6
64	1	5	9	6	0	4·2
65	0	7	8	6	6	5·4
66	3	6	5	8	0	4·4
67	4	1	8	3	2	3·6
68	0	6	3	3	9	4·2
69	3	3	9	9	5	5·8
70	6	4	7	4	2	4·6
71	2	0	3	7	5	3·4
72	7	2	8	0	1	3·6
73	8	9	7	7	2	6·6
74	0	6	9	2	9	5·2
75	1	9	2	2	0	2·8
76	2	1	0	4	1	1·6
77	3	9	1	2	1	3·2
78	2	8	2	4	0	3·2
79	5	1	3	3	2	2·8
80	9	4	7	8	1	5·8
81	1	3	1	1	6	2·4
82	2	4	2	7	9	4·8
83	2	7	1	5	9	4·8
84	9	6	2	7	2	5·2
85	0	9	3	9	7	5·6
86	3	3	8	0	2	3·2
87	9	4	1	7	5	5·2
88	6	9	0	2	7	4·8
89	5	3	1	5	0	2·8
90	6	5	4	2	1	3·6
91	4	0	4	4	2	2·8
92	4	6	3	9	1	4·6
93	2	3	1	7	7	4·0
94	8	5	2	3	0	3·6
95	4	4	1	2	4	3·0
96	1	1	0	0	4	1·2
97	7	9	7	3	3	5·8
98	2	9	2	5	0	3·6
99	8	6	3	8	9	6·8
100	0	4	5	2	6	3·4

(*a*) That with samples of only 5 individuals there will be, as might be expected, a very wide range in the values of the mean; the mean number of attacks of the whole 500 individuals is 4·43, which is very close to the mean of the universe sampled – namely, 4·50 (as given above) – but in the individual samples of 5 persons the values range from between 0·75 and 1·25 (the one at 1·2) to between 7·25 and 7·75 (the one at 7·4). In samples of 5, therefore, there will be instances, due to the play of chance, in which the observed mean is very far removed from the real mean.

(*b*) On the other hand, these extreme values of the observed mean are relatively rare, and a large number of the means in the samples lie fairly

TABLE 16

MEAN NUMBER OF COLDS PER PERSON IN SAMPLES OF DIFFERENT SIZE

Value of Mean in Sample	Frequency with which Mean Values, as shown in Column (1), occurred			
	Samples of 5	Samples of 10	Samples of 20	Samples of 50
(1)	(2)	(3)	(4)	(5)
0·75–	1	—	—	—
1·25–	1	—	—	—
1·75–	4	1	—	—
2·25–	2	2	—	—
2·75–	12	5	2	1
3·25–	15	8	9	5
3·75–	12	16	24	22
4·25–	10 } 39	26 } 58	31 } 77	45 } 91
4·75–	17	16	22	24
5·25–	8	15	10	3
5·75–	6	8	2	—
6·25–	7	3	—	—
6·75–	4	—	—	—
7·25–7·75	1	—	—	—
Total number of means	100	100	100	100
Total observations	500	1000	2000	5000
Grand means	4·43	4·61	4·50	4·48

close to the mean of the universe (4·50); 39 per cent of them lie within three-quarters of a unit of it (i.e. between 3·75 and 5·25).*

When in place of samples of 5 individuals a hundred samples of 10 individuals were taken at random from this universe, the distribution of the means in these samples showed a somewhat smaller scatter – as is shown in column (3) of Table 16. The extreme values obtained now lie in the groups 1·75–2·25 and 6·25–6·75, and 58 per cent of the values are within three-quarters of a unit of the real mean – i.e. the mean of the universe.

When 100 samples of 20 were taken (column 4) the scatter was still further reduced; the extreme values obtained lay in the groups 2·75–3·25 and 5·75–6·25, and 77 per cent of the values lay within three-quarters of a unit of the real mean. Finally, with samples of 50 (column 5) there were 91 per cent of the means within this distance of the true mean, and 45 per cent lay in the group 4·25–4·75 – i.e. did not differ appreciably from the real mean. Outlying values still appeared, but appeared only infrequently.

Two Factors in Precision

These results show, what is indeed intuitively obvious, that the precision of an average depends, at least in part, upon *the size of the sample.* The larger the random sample we take the more accurately are we likely to reproduce the characteristics of the universe from which it is drawn. The size of the sample, however, is not the only factor which influences the accuracy of the values calculated from it. A little thought will show that it must also depend upon the *variability of the observations in the universe.* If every individual in the universe could only have one value – e.g. in the example above every individual in the universe had exactly 3 colds – then clearly, whatever the size of the sample, the mean value reached would be the same as the true value. If on the other hand the individuals could have values ranging from 0 to 900 instead of from 0 to 9, the means of samples could, and would, have considerably more variability in the former case than in the latter. The accuracy of a value calculated from a sample depends, therefore, upon two considerations:—

(*a*) The size of the sample.

* The noticeable unevenness of the distribution with samples of 5 is artificial, being due merely to the group intervals used. With only whole numbers in the universe the mean of 5 observations *must* end in an even number (see Table 15). The groups 0·75 to 1·25, 1·75 to 2·25, etc., however, contain 3 possible values (0·80, 1·00 and 1·20, 1·80, 2·00 and 2·20, etc.) while the groups 1·25 to 1·75, 2·25 to 2·75 contain only 2 possible values (1·40 and 1·60, 2·40 and 2·60, etc.). This defect does not apply to the samples of 10, 20 and 50 and has no material effect upon the general arguments and demonstrations based upon the distribution for samples of 5.

(*b*) The variability of the characteristic within the universe from which the sample is taken.

The statistician's aim is to pass from these simple rules to more precise formulae, which will enable him to estimate, with a certain degree of confidence, the value of the mean, etc., in the universe and also to avoid drawing conclusions from differences between means or between proportions when, in fact, these differences might easily have arisen by chance.

Measuring the Variability of Means

As a first step we may return to Table 16 and measure the variability shown by the means in the samples of different sizes. So far we have illustrated that variability by drawing attention to the range of the means, and, roughly, the extent to which they are concentrated round the centre point; a better measure will be the standard deviations of the frequency distributions. The results of these calculations are shown in Table 17.

TABLE 17

VALUES COMPUTED FROM THE FREQUENCY DISTRIBUTIONS OF MEANS GIVEN IN TABLE 16

Number of Individuals in each Sample	The Mean, or Average, of the 100 Means*	The Variability or Standard Deviation of each 100 Means	The Standard Deviation of the Observations in the Population Sampled ÷ Square Root of Size of Sample
5	4·43	1·36	1·28
10	4·61	0·91	0·91
20	4·50	0·61	0·64
50	4·48	0·44	0·41

* It happens by chance that the grand mean of the 500 observations is in fact closer to the true mean than that given by the 1000 observations, while that based on 2000 is exact and that based on 5000 differs slightly. The observed standard deviations were calculated from the original ungrouped data.

The standard deviation, or scatter, of the means round the grand mean of each of the total 100 samples becomes, as is obvious from the frequency distributions, progressively smaller as the size of the sample increases. It is clear, however, that the standard deviation does not vary *directly* with the size of the sample; for instance, increasing the sample from 5 to 50 – i.e. by ten times – does not reduce the scatter of the means by ten times. The scatter is, in fact, reduced not in the ratio of 5 to 50 but of $\sqrt{5}$ to $\sqrt{50}$ – i.e. not ten times but 3·16 times (for $\sqrt{5} = 2{\cdot}24$ and $\sqrt{50} = 7{\cdot}07$ and $7{\cdot}07/2{\cdot}24 = 3{\cdot}16$). This rule is very closely fulfilled by the values of Table 17; the standard deviation for samples of 5 is 1·36, and this value is 3·09 times the standard deviation, 0·44, with samples of 50. The first more precise rule, therefore, is that *the accuracy of the mean computed from a sample does not vary directly with the size of the sample but with the square root of the size of the sample.* In other words, if the sample is increased a hundredfold the precision of the mean is increased not a hundredfold but tenfold.

As the next step we may observe how frequently in samples of different sizes means will occur at different distances from the true mean. For instance it was pointed out above that with samples of 5 individuals 39 per cent of the means lay within three-quarters of a unit of the true mean of the universe. The grand mean of these 100 samples, 4·43, is not quite identical with the true mean of the universe, 4·50, as, of course, the total 500 observations are themselves only a sample; it comes very close to it as the total observations are increased – it is 4·48 with 100 samples of 50. Instead, therefore, of measuring the number lying within three-quarters of a unit, or one unit, of the grand mean (or true mean, taking them to be, to all intents and purposes, identical), we may see how many lie within the boundary lines 'grand mean plus the value of the standard deviation' and 'grand mean minus the value of the standard deviation' – i.e. 4·43 + 1·36 = 5·79 and 4·43 – 1·36 = 3·07. The calculation can be made only approximately from Table 16, but it shows that some two-thirds of the means will lie between these limits. If we extend our limits to 'grand mean plus *twice* the standard deviation' and 'grand mean minus *twice* the standard deviation' – i.e. 4·43 + 2 (1·36) = 7·15 and 4·43 – 2 (1·36) = 1·71 – it will be seen that these include nearly all the means of the samples, only about 3 per cent lying beyond these values (according to theory we expect 5 per cent beyond ± twice the standard deviation). Roughly the same results will be reached if these methods are applied to the larger samples. Our conclusions are therefore:—

(*a*) If we take a series of samples from a universe, then the means of those samples will not all be equal to the true mean of the universe but will be scattered around it.

(*b*) We can measure that scatter by the standard deviation shown by the means of the samples; means differing from the true mean by more than twice this standard deviation, above or below the true mean, will be only infrequently observed.

The Means Show a 'Normal' Distribution

To be rather more precise, it can be proved that the means of the samples will be distributed round the mean of the universe approximately in the shape of the normal curve discussed on pp. 82–85. In other words, as shown there, it can be calculated how many of them will lie, in the long run, if we take enough samples, within certain distances of the real mean, those distances being measured as multiples of the standard deviation. To illustrate this more exactly than can be done with the figures of Table 16, a further 100 samples of 10 observations were taken from random sampling numbers. The distribution shown by these 100 means is set out in Table 18. The true mean of the universe is 4·50 and the means of the samples will be distributed round that value with a standard deviation of 0·91 (see Table 17). If they are following a normal curve we can theoretically calculate how many of them should lie in such intervals as (*a*) (real mean) to (real mean ± 1 times S.D.), (*b*) (real mean) to (real mean ± 2 times the S.D.,) etc. From Table 18 we can similarly calculate the

TABLE 18

THE DISTRIBUTION OF MEANS IN 100 SAMPLES OF 10 OBSERVATIONS, THE TRUE MEAN OF THE UNIVERSE BEING 4·50

Value of the Mean in Terms of the Actual Units of Measurement	Number of Means observed with given Value	Number of Means theoretically expected
Less than 1·770	0	0·135
1·770–	0	0·486
2·225–	1	1·654
2·685–	5	4·406
3·135–	7	9·185
3·590–	14	14·988
4·045–	21	19·146
4·500–	20	19·146
4·955–	15	14·988
5·410–	9	9·185
5·865–	5	4·406
6·320–	3	1·654
6·775–	0	0·486
7·230 or more	0	0·135
Total	100	100

actual number that did 'turn up' in these intervals, since the observations have been placed in groups with a class-interval of $\frac{1}{2}(0{\cdot}91) = 0{\cdot}455 = \frac{1}{2}$ the standard deviation.

Observation and theoretical calculation gave the results set out below where it can be seen that the values given by the experiment are extremely close to the theoretical values expected. Some very extreme values – more than $2\frac{1}{2}$ times the standard deviation away from the real mean – would in time turn up, but as there should be only 12 such values in 1000 it is obviously not surprising that no such value was noted in only 100 samples.

Number of sample means that lie between—	Observed Number	Expected Number
Real Mean of 4·50 and Real Mean ± 1 S.D., i.e. between 3·590 and 5·409	70	68·27
Real Mean of 4·50 and Real Mean ± $1\frac{1}{2}$ S.D., i.e. between 3·135 and 5·864	86	86·64
Real Mean of 4·50 and Real Mean ± 2 S.D., i.e. between 2·685 and 6·319	96	95·45
Real Mean of 4·50 and Real Mean ± $2\frac{1}{2}$ S.D., i.e. between 2·225 and 6·774	100	98·76
Real Mean of 4·50 and Real Mean ± 3 S.D., i.e. between 1·770 and 7·229	100	99·73

The numbers of means occurring within different distances from the real mean are shown in Fig. 12, together with the superimposed normal curve. It is clear that the means occurring in the experiment do follow that curve and that we should, therefore, from our knowledge of this curve, expect only about 1 in 22 to differ from the true mean by more than twice the standard deviation (for 95·45 per cent lie within that distance, and the 4·55 per cent outside these limits is 1 in 22); and we should expect only 1 in 370 to differ from the true mean by more than 3 times the standard deviation (for 99·73 per cent lie within that distance and the 0·27 per cent outside these limits is 1 in 370). *In other words, we have found once more that a mean in a sample that differs from the real mean of the universe by more than twice the standard deviation shown by the sample means is a fairly rare event, and one that differs from the real mean by more than 3 times that standard deviation is a very rare event.*

In comparison with Table 17 it may be noted that this second series of samples of 10 has a grand mean of 4·57 and a standard deviation round

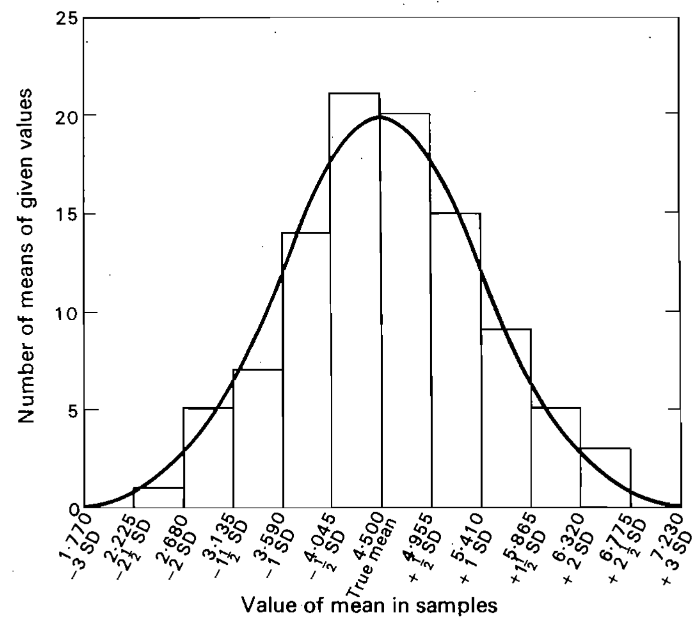

Fig. 12. Histogram of 100 means observed in samples of 10 observations and normal curve superimposed.

that mean of 0·97. Within the limits of only 100 samples, open to the play of chance, we have again reached approximately the real mean and a corresponding degree of scatter round it, as measured by the standard deviation of the sample means.

Deducing the Standard Deviation

In practice, however, we do not know this standard deviation of the means, for we do not usually take repeated samples. We take a single sample, say of patients with diabetes, and we calculate a single mean, say of their bodyweight. Our problem is this: how precise is that mean – i.e. how much would it be likely to vary if we did take another, equally random, sample of patients? What *would* be the standard deviation of the means *if* we took repeated samples? It can be shown mathematically that the standard deviation of means of samples is equal to *the standard deviation of the individuals in the population sampled* divided by the square root of the number of individuals included in the sample (usually written as $\sigma/\sqrt{n}$). These values have been added to Table 17 (right-hand column) and it will be seen that they agree very closely with the standard deviations calculated from the 100 means themselves (they do not agree exactly because 100 samples are insufficient in number to give complete accuracy). With this knowledge we can conclude as follows: the mean of the universe is 4·50 and the standard deviation of the individuals within it is 2·87 (see p. 97); if we take a large number of random samples composed of 5 persons from that universe, the means we shall observe will be grouped round 4·50 with a standard deviation of $2{\cdot}87/\sqrt{5}$; means that differ from the true mean, 4·50, by more than plus or minus twice $2{\cdot}87/\sqrt{5}$ will be rare. If we take a large number of samples of 50, then the means we shall observe will be grouped around 4·50 with a standard deviation of $2{\cdot}87/\sqrt{50}$, and means that differ from 4·50 by more than plus or minus twice $2{\cdot}87/\sqrt{50}$ will be rare.

The final step is the application of this knowledge to the single mean we observe in practice. In Chapter 7 the mean systolic blood pressue of 566 males (drawn from the area in and around Glasgow) was given as 128·8 mm. We want to determine the precision of this mean – i.e. how closely it gives the true mean blood pressure of males in this district.

Suppose that the true mean is M. Then from the reasoning developed above we know that the mean of a sample may well differ from that true mean by as much as twice $\sigma/\sqrt{n}$, where σ is the standard deviation of the blood pressures of individuals *in the universe* from which the sample was taken and n is the number of individuals in the sample; it is not likely to differ by more than that amount – i.e. our observed mean is likely to lie within the range $M \pm 2\ (\sigma/\sqrt{n})$. Clearly, however, we do not know the value of σ and as an estimate of it we must use the standard deviation of the values in our sample. It must be observed that this *is* only an estimate, for just as the mean varies from sample to sample so also will the standard deviation. But the latter varies to a slighter extent and so long as the sample is fairly large the estimate is a reasonable one, and unlikely to lead to any serious error. In the example cited the standard deviation of the 566 measures of systolic blood pressure was 13·05 mm. We therefore estimate that the standard deviation of *means* in samples of 566 would be $13{\cdot}05/\sqrt{566} = 0{\cdot}55$ mm.

We may conclude (presuming that the sample is a random one) that our observed mean may differ from the true mean by as much as ± 2 (0·55) but is unlikely to differ from it by more than that amount. In other words, if the true mean were 127·7 we might in the sample easily get a value that differed from it to any extent *up* to 127·7 + 2(0·55) = 128·8. But we should be unlikely to get a value as high as 128·8 if the true mean were lower than 127·7. Similarly if the true mean were 129·9 we might in the sample easily get a value that differed from it to any extent *down* to 129·9 – 2(0·55) = 128·8. But we should be unlikely to get a value as low as 128·8 if the true mean were higher than 129·9. In short, the true mean is likely to lie within the limits of 128·8 ± 2 (0·55) or between 127·7 and 129·9, for if it lay beyond these points we should be unlikely to reach a value of 128·8 in the sample.

The value $\sigma/\sqrt{n}$ is known as the *standard error* of the mean and is used as a measure of its precision. (In publication it should be given as Standard Error or S.E. and *not* merely by the sign ± which can be misleading).

Having calculated its value we can, as illustrated above, fix 'confidence limits' to the true mean. If the mean of the sample is $\bar{x}$, then we can estimate that the true mean of the universe, M, is not more than twice the S.E. away (plus or minus) and we can expect to be wrong in that conclusion only once in approximately 20 times. If we wish to be more 'confident,' we can estimate that the true mean of the universe is not more than two and a half times the S.E. away (i.e. wider limits) and we can then expect to be wrong in our conclusion only once in 80 times.

This estimation is clearly inapplicable if the sample is very small (say, less than 20 observations), for the substitution of the standard deviation of the few observations in the sample in place of the standard deviation of the whole universe, from which the few observations were taken, may be a serious error (see pp. 93–95, where it is shown that the standard deviations shown by samples of 5 observations often differ widely from the real standard deviation of the universe).

Summary

In medical statistical work we are, nearly always, using samples of observations taken from large populations. The values calculated from these samples will be subject to the laws of chance – e.g. the means, standard deviations, and proportions will vary from sample to sample. It follows that arguments based upon the values of a single sample must take into account the inherent variability of these values. It is idle to generalise from a sample value if this value is likely to differ materially from the true value in the population sampled. To determine how far a sample value is likely to differ from the true value, a standard error of the sample value is calculated. The standard error of a mean is dependent upon two factors – *viz.* the size of the sample, or number of individuals included in it, and the variability of the measurements in the individuals in the universe from which the sample is taken. This standard error is estimated by dividing the standard deviation of the individuals in the sample by the square root of the number of individuals in the sample. The mean of the population from which the sample is taken is unlikely to differ from the value found in the sample by more than plus or minus twice this standard error. This estimation is, however, not applicable to very small samples, of, say, less than 20 individuals, and must be interpreted with reasonable caution in samples of less than 100 individuals.

10 Problems of Sampling: Proportions

In the previous chapter the concept of the standard error was developed, and was illustrated by the calculation of the standard error of the mean. In addition it was pointed out that every statistical value calculated from a sample must have its standard error – i.e. may differ more or less from the real value in the universe that is being sampled. For example, the standard deviation, or measure of the scatter, of the observations will vary from sample to sample, and its standard error will show how much variability this value is, in fact, likely to exhibit from one sample to another taken from the same universe (see, for example, Table 14, p. 94). In practical statistical work a value which is of particular importance, owing to the frequency with which it has to be used, is the *proportion*. For example, from a sample of patients with some specific disease we calculate the proportion who die. Let us suppose that from past experience, *covering a very large body of material*, we know that the fatality-rate of such patients is 20 per cent (the actual figure, from the point of view of the development of the argument, is immaterial). We take, over a chosen period of time, a randomly selected group of a hundred patients and treat them with some drug. Then, presuming that our sample is a truly representative sample of all such patients – e.g. in age and in severity – we should observe, if the treatment is valueless, about 20 deaths (it may be noted that we are also presuming that there has been no secular change in the fatality-rate from the disease). We may observe precisely 20 deaths or owing to the play of chance we may observe more or less than that number. Suppose we observe only 10 deaths; is that an event that is likely or unlikely to occur by chance with a sample of 100 patients? If such an event is quite likely to occur by chance, then we must conclude that the drug *may* be of value, but, so far as we have gone, we must regard the evidence as insufficient and the case unproven. Before we can draw conclusions safely we must increase the size of our sample. If, on the other hand, such an event is very unlikely to occur by chance, we may reasonably conclude that the drug is of value (that is, of course, having satisfied ourselves that our sample of patients is comparable with those observed in the past in all respects except that of the treatment). Before we can answer the problem as to what is a likely