PRINCIPLES OF MEDICAL STATISTICS

V.—PROBLEMS OF SAMPLING : AVERAGES*

THE observations to which the application of statistical methods is necessary are those, it has been pointed out, which are influenced by numerous causes, the object being to disentangle that multiple causation. It must also be recognised 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 technique 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. If the average daily consumption of calories per man-value is found to be 3000 in a group of 200 families from whom particulars are collected of their week's consumption of food, it cannot be deduced that that value is likely to be the true average of all families. Housewives who are willing to undertake the task of giving such particulars may be above the average level of intelligence or be the more careful and thrifty of the population. The sample is, then, not a representative but a somewhat selected sample and there is no evidence as to the degree to which this selection affects the results. That difficulty of interpretation has been discussed in a previous section. 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 known as the "errors of sampling." 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 the standard deviation be 2·87 (as found, in the previous section, by calculating

*On line 18 of last week's article in this series the figure should be 37·2 (years) and not as printed.—ED. L.

TABLE IV.—*Number of Colds Suffered by Individuals, Values in Samples of 5. (Hypothetical Figures.)*

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

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. From 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 IV are set out a hundred such samples of 5 individuals drawn at random from the universe.[1] For instance in the first sample there were 3 individuals who had 2 colds each, one fortunate one 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

TABLE V.—*Mean Number of Colds per Person in Samples of Different Size*

(1) Value of mean in sample.	(2) Samples of 5.	(3) Samples of 10.	(4) Samples of 20.	(5) Samples of 50.
	Frequency with which mean values, as shown in column (1), occurred.			
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 mean..	4·43	4·61	4·50	4·48

sample the mean is $2+2+2+0+4 \div 5 = 2{\cdot}0$, in the second it is 7·2, and so on. The distribution of the means is given in Table V, 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 :

(*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 far removed from the real mean.

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

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

The statistician's aim is to pass from these simple rules to more precise formulæ, which will enable him to avoid drawing conclusions from differences between means or between proportions when in fact these differences might easily have arisen by chance.

MEASURING VARIABILITY

As a first step we may return to Table V 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 VI.

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 VI ; the standard deviation for samples of

[1] The " universe " actually used for this and later demonstrations was a publication entitled Random Sampling Numbers arranged by L. H. C. Tippett (Tracts for Computers, No. XV. Camb. Univ. Press, 1927). Sets of unit random numbers were taken from its columns in fives, tens, twenties, and fifties as required.

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.

TABLE VI.—*Values Computed from the Frequency Distributions of Means given in Table V*

Number of individuals in the 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

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 V, 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 $\pm$ 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.

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 body-weight. 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 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 VI (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 above); 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 $2·87/\sqrt{5}$; means that differ from the true mean, 4·50, by as much as plus or minus twice $2·87/\sqrt{5}$ 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 $2·87/\sqrt{50}$, and means that differ from 4·50 by as much as plus or minus twice $2·87/\sqrt{50}$ will be rare.

The final step is the application of this knowledge to the single mean we observe in practice. In the previous section the mean systolic blood pressure 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 x. 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 $x \pm 2\ (\sigma/\sqrt{n})$. 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. We, therefore, estimate that the standard deviation of *means* in samples of 566 would be $13·05/\sqrt{566} = 0·55$.

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 $\pm\ 2\ (0·55)$ but is unlikely to differ from it by more than that amount. In other words the true mean is likely to lie within the limits of $128·8 \pm 2\ (0·55)$ 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 $\pm$ value is known as the *standard error* of the mean and is used as a measure of its precision. The estimation is clearly inapplicable if the sample is very small, 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. With small samples different methods must be applied.

Sometimes in published work *probable errors* are given in place of standard errors. The former (which has no advantage over the latter) is merely the standard error multiplied by 0·6745. If the variability likely to be observed is $\pm$ twice the standard error it will clearly be $\pm$ thrice the probable error, for the latter is, approximately, only two-thirds of the former.

The origin of the factor 0·6745 is this. We previously calculated the number of means of samples that lay within the limits "grand mean plus once the standard deviation" and "grand mean minus once the standard deviation." If, instead, we calculated the number of means that lay within the limits "grand mean plus 0·6745 times the standard deviation" and "grand mean minus 0·6745 times the standard deviation" we should expect to find that precisely half the means lay within these limits and half lay outside them. If, therefore, we took a single sample it would be an even chance that its mean did not differ from the real mean by more than 0·6745 times the standard deviation.

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 20–30 individuals, and must be interpreted with reasonable caution in samples of less than 100 individuals.

A. B. H.

SPECIAL ARTICLES

THE NEW RESEARCH LABORATORIES AT THE ROYAL COLLEGE OF SURGEONS OF ENGLAND

ON Monday last the Hunterian Museum in Lincoln's Inn was closed for an indefinite period (perhaps till the middle of March) in order to expedite the building of laboratories, which has just begun, to house the scientific staff and research workers of the Royal College. The reasons which prompted the Bernhard Baron trustees to make their gift of £25,000 last May were the difficult conditions under which work was then being carried out, and the conviction that better accommodation for research work would lead to an increase in medical and surgical knowledge. Work is to be pushed forward with all possible speed and it is hoped to have the Bernhard Baron laboratories ready for occupation in the coming autumn. The scheme consists in the demolition and remodelling of the internal structure of the floors above the level of the College library and the erection of two towers (shown on the elevation reproduced in Fig. 1), one at each end of the existing fourth and fifth floors. The roof will be removed and re-erected in such a way that additional laboratory space will be provided on the sixth floor. Plans of these three floors as they will be when finished are given in Fig. 2.

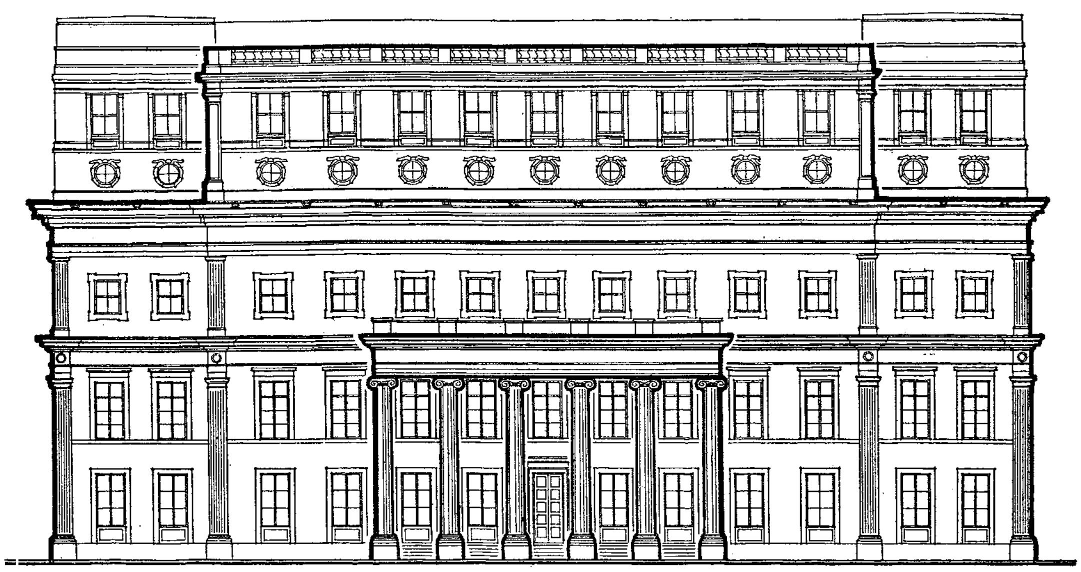

FIG. 1.—Front elevation of the College showing the new laboratories. The heavy outline encloses the original building.

SIXTH FLOOR

Commencing on the east side will be found the animal room, sufficient for the accommodation of experimental animals necessary for research work. This room is not as large as might be expected in laboratories of this size for the simple reason that stocks of experimental animals and animals recovering from operations are ordinarily housed at the Buckston Browne surgical research farm at Downe. The animal room will be finished in white tiles and will have one wall, made of glass and steel, which can be opened completely in summer weather. It faces south and is adequately ventilated by roof lights. The animal kitchen, food store, and cage sterilising room are in a corridor immediately outside the animal room. In the food store provision is made for refrigeration and bin storage of bulk foods. The cage sterilising room has a vat sufficiently large to contain and sterilise the largest cage in use. A staff kitchen is situated immediately outside the common room set apart for the use of members of the scientific staff and research workers. This room provides a central place where workers can discuss their problems and meet those interested. On the southern aspect of this room is a sliding folding window, 16 feet in length, capable of being thrown open during the summer. There will be an open fire and book-shelves