The Application *of a* Graphic Method *to* Fallible Measures.

By Francis Galton, F.R.S.

The following method of dealing statistically with values that vary in one dimension is singularly easy and of great convenience. It differs slightly in detail from what I have already published, and I shall add to it a new and very useful little table.

The observed values are supposed to be marshalled in the order of their magnitudes, and to be severally represented by vertical ordinates of lengths corresponding to the respective magnitudes, and these ordinates are supposed to be erected on a horizontal base, at equal distances apart, between two termini. The first ordinate will be separated from the first terminus by a half-space; the last ordinate will similarly be separated from the last terminus by a half-space, and each ordinate will be separated from the adjacent ordinate or ordinates by a whole space. The tops of the ordinates are then joined with a free hand to form a "curve of distribution," which always resembles more or less the peculiar ogival curve shown in fig. 1. The same curve broken in two parts, and the lowermost reversed, is shown in fig. 2:—

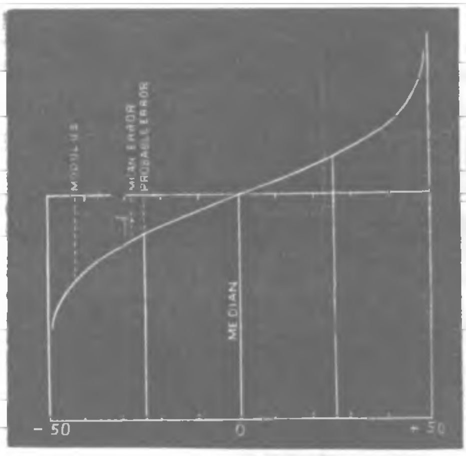

Fig. 1.

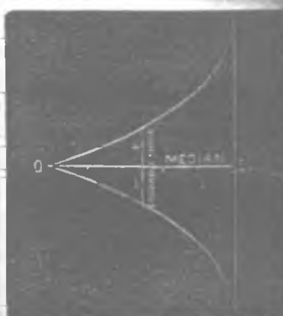

Fig. 2.

To express fig. 1 numerically, it is sufficient to measure and record the values of a very few of its ordinates corresponding to fractional lengths of its base. These values can at any subsequent time be plotted anew as ordinates to a base either of the same or of any other convenient length; then by joining the tops of those ordinates with a freely drawn line, the features of the original curve will be reproduced. The middlemost ordinate will by construction give the median value, that is to say, the value which one half of the observations exceed, and the other half fall short of. The first and third "quartile" ordinates will similarly give the values that one quarter of the observations exceed and that another quarter fall short of. These are most useful data, because the median value is practically identical with the mean value, and half the difference between the quartiles is the "probable error" of the series, which is perhaps the most convenient unit for measuring its variability.

I lately ("Journal Anthropological Institute," vol. xiv, p. 277) gave a table of statistics on this principle, in which the base was divided into 100 parts, and I named the ordinates at the middle of each of those parts, "percentiles." They were there reckoned from 1 to 100; but a considerable amount of subsequent experience has shown it to be more convenient to call the middle division 0°, and to reckon outwards on either side of it, to + 50 and to − 50 respectively. Fig. 2 admits of being expressed according to this arrangement; it is useful as showing with much distinctness the range within which any required fraction of the observed values are found to vary.

The precise method of plotting the observations is best explained by an example, Table I. The observations are first summarised as in line A (paying regard to the source of error I pointed out in the preceding paper). The meaning of A is that in a total of 775 observations there were 2 cases measuring 29 and under 30 inches, 8 cases measuring 30 and under 31 inches, and so on. The line B contains the sums of the entries in line A, reckoned from the beginning, and is to be read as follows: 2 cases under 30 inches, 10 cases(= 2 + 8) under 31 inches, 62 cases (= 2 + 8 + 52) under 32 inches, and so on. The line C contains the reduction of line B into percentages: it is these that I plot upon sectional paper, decimally divided from 0 to + and − 50.

We see in line C, that 0·2 per cent. of the observations were short of 30 inches, consequently the next higher measurement will exceed 30 inches, therefore there is no noteworthy error occasioned by accepting 0·2 as the abscissa corresponding to the exact ordinate of 30 inches. Similarly we may accept 1·3 as the abscissa corresponding to 31 inches, and so on, as shown in line D. These are the values that I plot, and from which I draw my curve of distribution, whence I measure off ordinates at 10°, 20°, 25°, &c.

TABLE I.—*Height, Sitting, of Female Adults, Aged 23—50, in Inches.*

	29—	30—	31—	32—	33—	34—	35—	36—	37—	
A	2	8	52	116	226	227	108	31	5	Total 775
B	2	10	62	178	404	631	739	770	775	Abscissæ 0 to 775
C	·2	1·3	8·0	23·0	52·2	81·4	95·4	99·4	100·0	Abscissæ 0 to 100·0
D	30	31	32	33	34	35	36	37	38	Corresponding Ordinates.

It is usually found that a series of observed values are "normally" variable, that is to say that they conform with sufficient exactitude for ordinary purposes, to the series of values calculated from the *à priori* reasonings of the law of Frequency of Error. My method of plotting the observations enables us very readily to test the presumed conformity in any given case, and, if it exists, the plotted curve not only give us the probable error of the series, but also the mean error and the modulus.

To perform this test we must shift our line of reference. The base line on which the ordinates were reared, and the ordinates themselves, must both be abandoned, and we must take instead of these the horizontal line drawn through the top of the ordinate at 0°, and a new set of ordinates drawn from that horizontal line to the curve. These new ordinates obviously represent the differences between each observed value and the mean of all the observed values, and they may be variously described as deviations, divergencies, or errors, according to the character of the observations in hand. It will be convenient to call these new ordinates by the general term of "deviates."

The accompanying table, Column B, gives the theoretical values of the deviates at specified points in the curve, in terms of the probable error. I have obtained those in column A by interpolation from the familiar table of values of the well known integral in the Law of Frequency of Error, such as will be found in an abbreviated form in Airy's "Theory of Errors," p. 22, where the values of the deviates are calculated in terms of the modulus. Column B has been deduced from A by simple proportion.

TABLE II.—*Normal Curve of Distribution of Error.*

Abscissæ, reckoned from 0° to ± 50° (Value of the Integral).		Corresponding Ordinates.	
		Value of the Deviate in which Modulus = 1. A.	Value Reduced proportionately to Probable Error = 1. B.
	10	0·179	0·38
	20	0·371	0·78
Probable error	25	0·477	1·00
Mean „	28·7	0·564	1·18
	30	0·595	1·25
	40	0·906	1·90
Modulus	42·1	1·000	2·10
	45	1·163	2·44
	47	1·330	2·79
	49	1·651	3·46
	50	Infinite	Infinite

In order to bring the observed values into a form suitable for comparison with this table, we must begin by measuring the observed deviates at ± 10°, 20°, 25°, 30°, 40°, and 45°. The mean of the ± deviates at 25° gives, by construction, the "probable "error." This must be considered as unity, and all the other observed deviates must be reduced proportionately to it. Then, if the series is "normal," the values so obtained will be identical with those in Column B, and if it is approximately normal, they will correspond approximately. In the former case the value of the deviate at 28·7 will be the value of the mean error exactly, in the latter case very nearly so. Similarly the deviate at 42·1 will give the modulus.

I have, during the last few months, had occasion to make considerable use of this graphic method of dealing with fallible measures and variable values, and have found it in every way satisfactory. The sectional paper I have latterly employed is that sold by Messrs. Letts, 36, King William Street, London Bridge, ruled in inches and tenths; which is, however, on somewhat too large a scale for the convenient plotting of percentage abscissæ, while that of millimetres is too small. On the whole, when I next have occasion to work in this way, I think it will be best to try sectional paper, decimally ruled, in which the unit of distance between line and line is one-sixteenth of an inch.