On the Value of a Mean as Calculated from a Sample.

By Professor F. Y. Edgeworth, F.B.A.

Under the above title, Dr. Isserlis has ably discussed the frequency function pertaining to the mean of a sample when the sample, numbering say n, is so large relatively to the number, say N, of the population from which the sample is taken at random that the ratio n/N is not a negligible quantity.* The constants which enter into the function—the mean, the standard-deviation and the constants known as β_1 and β_2—are deduced by Dr. Isserlis with an accuracy which leaves no room for improvement. But it may be interesting to obtain the values of these constants by a different, perhaps a shorter, method. There will come into view along with the same values of the constants a different form for the function to which they relate.

The function which represents the distribution of the *mean* or of the *sum*† presented by a large sample is commonly and unless the sample is very large (relatively to the population from which it is extracted) comprised in the class of functions which was discussed in a former number of this *Journal*, under the designation of the "Generalised Law of Error."‡

The case belongs to the—no doubt common, yet commonly ignored—species in which the elements or contributory causes through the random combination of which the law of error is set up are not perfectly independent, but slightly correlated. The correlation in the present case is of a kind specially noticed in the article cited§: that in which the form of the frequency-function pertaining to each (elementary or contributory) observation is affected by the magnitude of the preceding observations. An interesting variety of this species, that which is presented by the extraction of balls without replacement from an urn containing two kinds of balls, was discussed at some length in the said article. The present case differs from that variety only in that the frequency-function pertaining to the original population (from which the sample is taken at random) is not in general binomial.

* See *Journal of the Royal Statistical Society*, vol. lxxxi (1918), p. 75; and *Proceedings of the Royal Society*, 1915, there cited.

† The identity of the theory relating to the distribution of the *means* and to that of the *sum* may be recalled by considering the mean as the sum of the observations each divided by n (the total number thereof).

‡ *Journal of the Royal Statistical Society*, vol. lxix. (1906).

§ *Loc. cit.*, p. 506.

Let $_1y$ be the frequency-curve expressing the probability that the first observation will have an assigned value of the variable, $_1x$, measured from the mean of all the values thereof which occur in the long run, that is, the mean of the original group.* Let $_2y$ denote the probability that if $_1x$ is the first observation, the second observation (measured from the same origin, say, the mean of the initial group) will have any assigned value, $_2x$. Likewise let $_3y$ denote the probability that, if the first and second observations are respectively assigned, the third observation will have any assigned value $_3x$. And so on. Then $W = {}_ny\, {}_{n-1}y\, {}_{n-2}y \ldots , {}_2y\, {}_1y$, represents the probability that any assigned set of values for the n observations forming the sample should concur. The mean value of $f\,({}_qx, {}_px \ldots)$ any function of some of the observations, the mean in the long run of all the values enjoyed by all the variables in the course of continued sampling, is the complete integral (or more exactly sum of finite differences), between extreme limits, of the expression $W\,f\,({}_qx, {}_px, \ldots)$. This kind of mean will be called (1) the *absolute* mean of the function, and will be denoted by square brackets, as thus: $[f\,({}_qx, {}_px \ldots)]$. (2) In forming the absolute mean, there may be omitted from W, as being equatable to unity, all the y's that correspond to variables with subscripts, *e.g.*, r, t, higher than the highest, *e.g.*, q, which occurs in the function. (3) If l is the lowest subscript occurring in the function, all the y's with subscripts lower than l may be omitted from W, and for $_ly$ may be put y, the ordinate of the initial group.†

The mean value of the first observation is by definition zero. The mean value of the second observation measured from the mean value of the first, say $_2x_1$, is the mean of the N original observations with the omission of *one*, say $_1x_1$.‡ Thus the mean value of

* More correctly $y \Delta x$ is the probability of an observation occurring between $_1x$ and $_1x + \Delta x$ or $(_1x - \frac{1}{2} \Delta x$ and $_1x + \frac{1}{2} \Delta x)$. But I shall leave it to the reader to make this correction for himself throughout; the rather as the correction may be dispensed with by taking Δx as unity.

† Since, in the long run, there is no reason (other than frequency in the initial group) why one of the original set of observations rather than another should stand for $_lx$. When $_1x$ is put for $_lx$ in the function which is to be averaged it is proper to put for $_px$, $_{p-l}x$.

‡ The left subscript denotes the number of the observation (whether it is the first, second, or rth observation); the right subscript denotes the position of the origin to which the observation is referred (whether it is the mean of the first, second, or rth observation). The symbol thus subscribed is used according to the context to denote the *variable*, *e.g.*, $_2x_2$ a second observation measured from the mean of the second observation, or a *particular value* of that variable. It might have been better to designate the latter by the addition of a dash, as thus $_2x'_2$.

$_2x_1 = \frac{\Sigma\ _1x_1 - {}_1x_1}{(N-1)}$, the summation extending to *all* the values of $x_1 = -\ _1x_1/(N-1)$. Put $_2x_2$ for the second observation measured from the mean value of $_2x_1$, viz., $-\,_1x_1/(N-1)$; $_2x_1 = {}_2x_2 - {}_1x_1/(N-1)$. Whence $[_2x_1] = [_2x_2] - [_1x_1/(N-1)] = 0$; the mean value of $_2x_2$ for any assigned value of $_1x_1$, and, therefore, for all values of $_1x_1$ that occur in the long run, being zero. (*Cf.* (1) and (2).) Likewise, let $_3x_2$ denote the third observation measured from the mean value of the second; $_3x_2 = \frac{\Sigma_2x_2 - {}_2x_2}{N-2}$; the summation extending over all the $(N-1)$ values which $_2x_2$ may enjoy. Put $_3x_3$ for the third observation measured from the last written mean. Thus

$$_3x_1 = {}_3x_2 - {}_1x_1/(N-1) = {}_3x_3 - {}_2x_2/(N-2) - {}_1x_1/(N-1).$$

By parity of reasoning $[_3x_1] = 0$. Likewise (4) generally

$$_rx_1 = {}_rx_r - \frac{_{r-1}x_{r-1}}{N-(r-1)} - \frac{_{r-2}x_{r-2}}{N-(r-2)} \cdot \cdot \cdot \cdot - \frac{_1x_1}{N-1}.$$

Whence $[_rx_1] = 0$. (5) The mean of the sample, the absolute mean,

$$[_rx_1 + {}_{r-1}x_1 \cdot \cdot \cdot \cdot + {}_2x_1 + {}_1x_1]$$

$= [_rx_1] + [_{r-1}x_1] + \cdot \cdot \cdot \cdot + [_2x_1] + [_1x_1] = 0$. (6) The absolute mean of any function of the form $_tx_tf(_qx_q, {}_px_p \cdot \cdot \cdot \cdot)$ where $t > q > p \ldots$ is zero.

To compute the second moment for the sample let us write the sum of the observations $_nx_1 + {}_{n-1}x_1 + \cdot \cdot \cdot \cdot + {}_2x_1 + {}_1x_1$ in the form obtainable from (4),

$$(N-n)\left(\frac{_nx_n}{N-n} + \frac{_{n-1}x_{n-1}}{N-(n-1)} + \cdot \cdot \cdot \cdot + \frac{_2x_2}{N-2} + \frac{_1x_1}{N-1}, \text{ say } (N-n)\right)_nX.$$

The mean square of their sum, say, $_nK$,

$$= (N-n)^2 \left\{ \left[\frac{_nx^2_n}{(N-n)^2}\right] + \left[\frac{_{n-1}x^2_{n-1}}{(N-(n-1))^2}\right] + \cdot \cdot \cdot \cdot \left[\frac{_1x_1^2}{(N-1)^2}\right] \right\};$$

products of the type $[_px_p\ _qx_q]$ vanishing, by (6). The mean squares which are constituents of $[_nX^2]$ may each be expressed in terms of the mean square of deviation for the original group, say k. Thus $[_1x_1^2] = k$. The mean of $_2x_2$ for an assigned value of the first observation is found by taking the mean square of the $N-1$ members of the original group which remain after that one, $_1x_1$ has been abstracted. The required mean, say $_2k$,

$$= \frac{\Sigma\,(x_1 + x_1/(N-1))^2 - (_1x_1 + {}_1x_1/(N-1))^2}{N-1};$$

where Σ denotes summation over the whole original group numbered N. Whence $_2k = \frac{N}{N-1}\Sigma\frac{_1x_1^2}{N} + \frac{N_1x_1^2/(N-1)^2 - N^2{}_1x_1^2/(N-1)^2}{N-1}$

$$= \frac{N}{N-1}k - \frac{N}{(N-1)^2}\,_1x_1^2.$$

This is the mean value of ${}_2x_2$ relative to a particular value of the first observation, viz., ${}_1x_1$. Whence for the absolute Mean $[{}_2x_2^2]$ we have (see (1) and (2)) $\frac{N(N-2)}{(N-1)^2}\,k$; say $[{}_2k]$.

Likewise ${}_3k = \frac{N-1}{N-2}[{}_2k] - \frac{N-1}{(N-2)^2}\,{}_2x_2^2$; $[{}_3k] = \frac{(N-1)(N-3)}{(N-2)^2}\,{}_2k$.

Generally (7) ${}_{p+1}k = \frac{N-(p-1)}{N-p}[{}_pk] - \frac{N-(p-1)}{(N-p)^2}\,{}_px_p^2$.

$$(8) \qquad {}_rk = \frac{N-r}{N-(r-1)} \qquad \frac{N-(p+1)}{N-p}\,{}_pk.$$

$$(9) \qquad [{}_rk] = \frac{N-r}{N-(r-1)} \qquad \frac{N}{N-1}\,k.$$

Substituting now in each constituent (of $[{}_nX^2]$) of the type $[{}_rx_r^2]$ the value of $[{}_rk]$ given in (9) we have

$$(10)\ {}_nK = (N-n)^2\left(\frac{k}{(N-n)(N-(n-1))} + \frac{k}{N-(n-1)(N-(n-2))} + \cdots + \frac{k}{N(N-1)}\right)$$

$$= (N-n)^2\left(\frac{1}{N-n} - \frac{1}{N}\right)\frac{N}{N-1}\,k = n\,\frac{N-n}{N-1}\,k.$$

Which agrees with Dr. Isserlis' computation of the mean value of z^2; it being observed that his z is the deviation of the *mean* of the sample from that of the whole population; whereas the deviation here contemplated is the *sum* of the observations forming the sample.*

The determination of the second moment may be effected even more expeditiously if we are content with the approximation formed by neglecting powers above the first of n/N (and *a fortiori* of $1/N$.) Put $\Delta\,{}_nK$ for the difference between ${}_{n+1}K$ and ${}_nK$.

$$\Delta\,{}_nK = [{}_1x^2_{n+1}x_1] + 2\,[{}_nx_{1+1}\,{}_nX\,(N-n)].$$

The first term $= k$, by (3). The remainder may be written $2\,[({}_{n+1}x_{n+1} - {}_nX)\,{}_nX]\,(N-n)$. The part of this expression affected with ${}_{n+1}x_{n+1}$ vanishes, by (6). The remainder—$2\,[X_n^2]\,(N-n)$ is of the order n/N. Accordingly, corrections of that order relatively to the remainder (being of the second order) may be neglected. Thus (by 10), $\Delta\,{}_nK$ reduces to $k - \frac{2kn}{N}$. Whence

${}_nK = nk - \frac{n^2-n}{N}\,k = (11)\ nk\left(1 - \frac{n-1}{N}\right)$ as it ought, when terms of the order $1/N$ are neglected.

* See note † on page 624; whence it appears that the mean value of our K ought to be identical with the mean of Dr. Isserlis' $z^2 \times n^2$.

The approximate determination of the third moment, say, J, is nearly as simple. Put j for the third moment of the original group; and let ${}_2j$, ${}_3j$ . . . . be related to j as ${}_2k$, ${}_3k$ . . . . to k.

$$\Delta_n\mathrm{J} = [{}_{n+1}x^3{}_1] + 3\,[{}_{n+1}x^2{}_1\,({}_nx_1 + {}_{n-1}x_1 + \dots + {}_1x)] + 3\,[{}_{n+1}x_1\,({}_nx_1 + {}_{n+1}x_1 + \dots)^2]$$
$$= j + 3\,(\mathrm{N}-n)\,[({}_{n+1}x_{n+1} - {}_n\mathrm{X})^2{}_n\mathrm{X}] + 3\,(\mathrm{N}-n)^2\,[({}_{n+1}x_{n+1} - {}_n\mathrm{X})\,{}_n\mathrm{X}^2],$$

by (3) and (4). Whence, omitting terms affected with ${}_{n+1}x_{n+1}$, according to (6), and neglecting ${}_n\mathrm{X}^2$ as of the order $\left(\frac{n}{\mathrm{N}}\right)^2$, we have (12) $\Delta_n\mathrm{J} = j + 3\,(\mathrm{N}-n)\,[{}_{n+1}x_{n+1}{}^2\,{}_n\mathrm{X}] - 3\,(\mathrm{N}-n)^2\,[{}_n\mathrm{X}^3]$. From (7) and (8) it appears that each constituent of the second term being of the type ${}_{n+1}x_{n+1}\frac{1}{\mathrm{N}-t}\,{}_tx_t$ reduces to the type

$$\frac{1}{\mathrm{N}-t}\left[\left(\alpha_t k - \frac{\beta}{(\mathrm{N}-t)^2}\,{}_tx_t{}^2\right)\,{}_tx_t\right];$$

where α and β are co-efficients differing from unity by proportions of the order n/N. The part of this expression which is affected with the first power of ${}_tx$ vanishes. There remains for the second term of (12) $-3\,(\mathrm{N}-n)\,\beta\,{}_tx_t{}^3/(\mathrm{N}-t)^2$. This type being of the order $1/\mathrm{N}$, fractions thereof of that order may be neglected. Accordingly, the second term of (12) reduces to $-3\,nj/\mathrm{N}$. Likewise, the third term of (12) contains $3n$ constituents of the type $(\mathrm{N}-n)^2\,\frac{{}_tx_1{}^3}{(\mathrm{N}-t)^3}$, which are of the order $1/\mathrm{N}$; and other constituents of the type $(\mathrm{N}-n)^2\,{}_tx_t{}^2\,{}_qx_q$, which, as just now argued, are of a lower order. Thus, the third term also reduces to $-3\,nj/\mathrm{N}$. Altogether $\Delta_n\mathrm{J} = j - \frac{6\,nj}{\mathrm{N}}$. Whence ${}_n\mathrm{J} = nj - 3\frac{(n^2-n)}{\mathrm{N}}\,j = nj\left(1 - \frac{3\,(n-1)}{\mathrm{N}}\right)$. Which agrees with Dr. Isserlis' expression for the mean value of his z^3, viz., $\frac{(\mathrm{N}-n)\,(\mathrm{N}-2n)}{(\mathrm{N}-1)\,(\mathrm{N}-2)}\,{}_3\mu/n^2$; when it is multiplied by n^3 so as to relate to the *sum* of the observations,* and if it is expanded in ascending powers of $1/\mathrm{N}$ and powers above the first are neglected.

Of course, in order to obtain the required co-efficient not depending on the unit adopted, it is proper to divide the mean cube of deviation by the third power of the parameter, in the present case the standard deviation, which is approximately equal to $nk\left(1 - \frac{n-1}{\mathrm{N}}\right)$, by (11). We have thus for the constant which

* See note to p. 624 above.

may be called $\sqrt{B_1}\ \dfrac{nj\,(1-3)\ (n-1)/N)}{n^{\frac{3}{2}}\,k^{\frac{3}{2}}\,(1-n-1/N)^{\frac{3}{2}}}$;

$B_1 = \dfrac{1}{n}\dfrac{j^2}{k^3}\,(1-3\,(n-1)/N\,) = \dfrac{\beta_1}{n}$.* Which corresponds to Dr. Isserlis' expression for B_1 when it is expanded in ascending powers of n/N and $1/N$ and powers above the first are neglected.

A more exact investigation of ${}_nJ$ on the lines above traced will arrive exactly at Dr. Isserlis' result. It will be sufficient here to indicate some of the steps.

$${}_nJ = (N-n)^3\,[{}_nX^3]$$

$$=(13)\ \underset{r=1}{\overset{r=n}{S}}\ (N-n)^3\left[\frac{{}_rx_r{}^3}{(N-r)^3}+\frac{3{}_rx_r{}^2}{(N-r)^2}\,{}_{r-1}\,{}_{r-1}X+\frac{3{}_rx_r}{N-r}\,{}_{r-1}X^2\right].$$

Of the terms within the square brackets, the third vanishes, by (6). To compute the remainder, we obtain values of ${}_rj$ and $[{}_rj]$ analogous to those which have been obtained for ${}_rk$ and $[{}_rk]$. With the aid of these formulæ we compute the mean value of the series $\underset{p=1}{\overset{p=r-1}{S}}\ {}_rx_r{}^2\,{}_px_p/(N-p)$, say, ${}_rT$ for any assigned value of r.

Combining part of the second term of (13) thus computed with the first term we obtain $\dfrac{-2N-2r+1}{(N-(r-1))^2\,(N-r)^2}\ \dfrac{N^2}{(N-1)\,(N-2)}$ for a constituent of (13). The sum of terms of this type from $r=1$ to $r=n$ is $\left(\dfrac{1}{(N-n)^2}-\dfrac{1}{N^2}\right)\dfrac{N^2}{(N-1)\,(N-2)}$. Also the remainder of the bracketed portion of (13) being summed comes to $3\left(\dfrac{1}{N-n}-\dfrac{1}{N}\right)\dfrac{N}{(N-1)\,(N-2)}$. Adding together these two results, multiplying by $(N-n)^3$, and reducing, we obtain for (13) $n\,\dfrac{N^2-3n\,N+2n^2}{(N-1)\,(N-2)}\,j$. Which is exactly what we ought to obtain, being n^3 times Dr. Isserlis' expression for the mean value of his z^3.

An abridged value for the fourth moment may be obtained by parity of reasoning; and the exact value, if it is thought to be worth the trouble of computation. But it may be doubted whether ordinarily in the determination of the third and fourth moments and the constants depending thereon, it is worth while to proceed beyond a first correction. For it must be presumed that the said constants are small; otherwise, there would be no affinity to the

* *Loc. cit.*, p. 76. *Cf.* note to p. 624 above.

β_1, here the Pearsonian co-efficient for the original group, is closely related to the co-efficient called by the present writer in former papers j or k_1.

law of error,* and without such affinity, it is submitted, the frequency-function which is presented will not be of much avail for the purpose in hand—to estimate the accuracy of a mean obtained from a sample.†

The present purpose forms an additional consideration on the side of the law of error, or the cognate method of translation, in the issue between that system and the Pearsonian types.‡ Consider first the extremities of the frequency-group which is to be represented; say the portions lying at a distance from the Mean greater than one-and-a-half times or twice the Standard Deviation. It is objected to the types that throughout this outlying region the frequency-function which they represent does not satisfy with

* The affinity which may be expected in the case of any average where the number of the constituents is large and the correlation between them small is particularly to be expected in the present case where, as appears from Dr. Isserlis' investigation, the correlation between the constituents is such as to reduce the co-efficients of abnormality B_1 and B_2 in more than the regulation degree. What is meant by the regulation degree may be illustrated by considering what may be called the *external case* (*cf. Journal of the Royal Statistical Society*, vol. xxx, 1917, p. 425) of sampling. Beginning as before with N observations select one at random and *add* a duplicate to the group. Again in the new group of $N + 1$ observations select a second observation and add a duplicate. And so on. By parity of reasoning, with a mere change of sign, it may be shown that the mean or sum of n observations thus formed will conform to the generalised law of errors as long as n/N remains small.

† The estimate that one representative frequency-function is more accurate than another is not so simple a matter as may be supposed. We cannot hope that for every assigned deviation the probability of its occurrence should be less for one curve than for the other (the areas subtended by the curves being equal). We must resort to an estimate of the detriment due to errors in the long run; as shown by the present writer, on the lines of Laplace and Gauss. (*Journal of the Royal Society*, vol. lxxi (1908), p. 386, referring to *Philosophical Magazine*, vol. xvi (1883), p. 361.) The sort of comparison required can hardly be performed except between functions with which we are familiar; perhaps chiefly between members of the same family with different constants. These considerations supply the answer to a question which may be asked. In neglecting powers above the first of n/N in the values of the B_1 and B_2 co-efficients (or our j and i) can we be certain that we are on the safe side, that we are over-estimating the inaccuracy of the formula? The answer appears to be in the negative.

The following is a more important question. If the corrections afforded by the co-efficients B_1 and B_2 are going to be left out of account, shall we be on the safe side in using the standard-deviation of the group, viz., $\sqrt{k}$, rather than the co-efficient proper to the sample? The answer to this question is in the affirmative.

‡ As to this issue, see *Journal of the Royal Statistical Society*, vol. lxxx (1917), p. 412 *et seq.*

adequate approximation the differential equation on which it purports to be based.* Their claim rests on convenience and consistency rather than accuracy with respect to this part of the group. To be sure, no more can be said in general in favour of the rival system. At the extremities the law of error proper becomes unmeaning; and its substitute (translation) lacks credentials.† Both the rival systems, both translation and the types, are to be taken *cum grano.* They are like those apocryphal scriptures about which it is laid down that the Church doth read them as conducive to good conduct; "but yet doth not apply them to establish any doctrine." The forms employed to represent the extremities cannot be applied to establish any accurate conclusion.‡ But if, for the purpose in hand, they are to be employed to give a summary estimate of what may be called the improbable error, the method of translation has, at least, this advantage, that simply connected as it is with familiar tables it is more easily interpreted.§

A more decided advantage accrues to the law of error with respect to those portions of the group which admit of more accurate representation. The types are here open to the objection that they are based on a differential equation, which, being proper to a kind of binomial, is, in general, not known to be appropriate to the case in hand. In general, it is *not known* to be appropriate; but in the present case it is *known not* to be appropriate (usually and excepting the particular case of extraction without replacement from a medley containing only two species). But there is an even greater accession to the arguments in favour of the law of error. Not only is there reason for believing that the rival system is here not the true curve; but there is reason for believing that the (generalised) law of error is here the true curve. Ordinarily, the existence of conditions from which that law is generated is presumed as an hypothesis; but here it is known as a fact. In the limiting case when n/N becomes negligible there are presented the ideal conditions from which Laplace and Poisson deduced the law of error. True, as n/N becomes sensible, and accordingly the constituents of the average interdependent, the Laplace-Poisson proof is not applicable—or, at least, it has not been applied. But there is still available a method of proof which, according to a high authority, Poincaré, is the best proof of the law of error. Poincaré's dictum, indeed, relates only to the law of error proper, the Gaussian curve, and only as derived from *symmetrical*

* *Loc. cit.*, pp. 428, 429.

† *Cp. Journal of the Royal Statistical Society*, vol. lxxx (1917), p. 43, *et seq.*

‡ *Cp. loc. cit.*, p. 431.

§ *Cp. loc. cit.*, p. 437.

constituents.* But his authority may be fairly extended to sanction the same method of proof when applied to somewhat unsymmetrical and somewhat interdependent constituents. The true curve is the generalised law of error—not Type II_L, nor another.

This proposal to employ the constants first evaluated by Dr. Isserlis in a manner different from that which he proposes, implies no disparagement of the brilliant and original investigation by which he ascertained those constants.

* As to Poincaré's theory and its relation to the method independently pursued by the present writer, see *Philosophical Magazine*, May, 1918, p. 426.