

Hence we deduce :

$$z_1^2 = \sigma_1^2 \frac{\mu_1^2 \{1 - r_{12}^2 (1 - \mu_2^2)\}}{1 - r_{12}^2 (1 - \mu_1^2) (1 - \mu_2^2)} \dots\dots\dots (i)$$

$$z_2^2 = \sigma_2^2 \frac{\mu_2^2 \{1 - r_{12}^2 (1 - \mu_1^2)\}}{1 - r_{12}^2 (1 - \mu_1^2) (1 - \mu_2^2)} \dots\dots\dots (ii),$$

$$R_{12} = r_{12} \frac{\mu_1 \mu_2}{\sqrt{1 - r_{12}^2 (1 - \mu_1^2)} \sqrt{1 - r_{12}^2 (1 - \mu_2^2)}} \dots\dots\dots (iii),$$

$$\frac{M_1}{\sigma_1} = \frac{m_1}{\sigma_1} \frac{1 - r_{12}^2 (1 - \mu_2^2)}{1 - r_{12}^2 (1 - \mu_1^2) (1 - \mu_2^2)} + \frac{m_2}{\sigma_2} \frac{\mu_1^2 r_{12}}{1 - r_{12}^2 (1 - \mu_1^2) (1 - \mu_2^2)} \dots\dots\dots (iv),$$

$$\frac{M_2}{\sigma_2} = \frac{m_1}{\sigma_1} \frac{\mu_2^2 r_{12}}{1 - r_{12}^2 (1 - \mu_1^2) (1 - \mu_2^2)} + \frac{m_2}{\sigma_2} \frac{1 - r_{12}^2 (1 - \mu_1^2)}{1 - r_{12}^2 (1 - \mu_1^2) (1 - \mu_2^2)} \dots\dots\dots (v).$$

These formulae will be found useful (especially in deducing  $r_{12}$  from  $R_{12}$ ) in records in which there has been independent selection of two related individuals, without regard to their relationship.

## VII. On certain points concerning the Probable Error of the Standard Deviation\*.

By RAYMOND PEARL

The purpose of this paper is to discuss two problems of considerable practical importance in all biometrical investigations. These problems presented themselves in acute form in some studies of fecundity which the writer has at present under way. It was decided to be necessary to get definite answers to them before going farther with the work mentioned. In the belief that the matter is of general interest to workers in biometry it is presented here. The two points may be stated as follows:

I. It has been shown† that if in any frequency distribution  $\sigma_{\mu_q}$  be the standard deviation for errors in the  $q$ th moment coefficient  $\mu_q$ , taken about the mean, and  $\sigma$  be the standard deviation of the distribution, then

$$\sigma_{\mu_q} = \sqrt{\frac{\mu_{2q} - \frac{\mu_q^2}{n} - 2\mu_{q+1}\mu_{q-1} + q^2 \sigma^2 \mu_{q-1}^2}{n}},$$

where  $n$  denotes the number in the sample. In this expression put  $q=2$ , and then, since  $\mu_1=0$ , we have at once

$$\text{Probable error of } \mu_2 = 67449 \sqrt{\frac{\mu_4 - \mu_2^2}{n}}.$$

Further since  $\sigma = \sqrt{\mu_2}$ , we have

$$\text{P.E. of } \sigma = 67449 \sqrt{\frac{\mu_4 - \mu_2^2}{4\mu_2 n}} \dots\dots\dots (i).$$

This is the true value of the probable error designated, whatever be the type of the frequency curves. But, for the normal curve, since there  $\mu_4 = 3\mu_2^2$ , (i) reduces at once to

$$\text{P.E. of } \sigma = 67449 \sqrt{\frac{1}{2n}} \sigma,$$

or as it is usually written

$$= 67449 \frac{\sigma}{\sqrt{2n}} \dots\dots\dots (ii).$$

\* Papers from the Biological Laboratory of the Maine Agricultural Experiment Station, No. 1.

† *Biometrika*, Vol. II, p. 276.

Now this value (ii) is the one almost universally used in calculating the probable error of the standard deviation, quite without regard to the type of the distribution under discussion. Indeed one suspects, from statements made in memoirs and handbooks of biometrical methods, that the impression prevails among biologists that (ii) is the true and unique expression for this probable error in every case. Obviously (i) will not equal (ii) except in the special case of the normal curve. In the common use of (ii) for (i) it is, of course, assumed, consciously or unconsciously, that the deviation of the curve from normality is not such as to cause, for *practical* purposes, any significant error as far as these probable errors are concerned. So far as the writer is aware no one has attempted to test in a concrete way the validity of this assumption in specific instances. Yet the matter is obviously a very important one. Probable errors lie at the very basis of all biometrical reasoning. If in certain cases the probable errors are likely to be considerably in error when calculated in a certain way, any reasoning based upon them will have a most insecure foundation.

These considerations suggest the following practical problem to the working biometrician: When a frequency curve is *not* a normal curve, is the error made in the probable error of the standard deviation by assuming normality (using (ii) for (i)) ever likely to be great enough to be of *practical* significance in drawing conclusions from data? This is our first problem.

II. It has recently been pointed out by Sheppard\* that in calculating the probable errors of the mean and standard deviation it is theoretically not justifiable to use the moments to which his corrections have been applied. Instead the "raw" moments should be used. The theory of this is stated as follows by Sheppard (*loc. cit.*),  $\mu$  being used to denote corrected and  $\pi$  to denote raw moments: "Similarly if we take  $\mu_2$  to be equal to  $\pi_2' - \frac{1}{n}A^2$ , where  $\pi_2'$  is the value of  $\pi_2$  as calculated from the actual measurements, the error in  $\mu_2$ , due to the limitation of number of observations, is equal to the error in  $\pi_2$ ; and the mean square of this latter error is  $(\pi_2 - \pi_2')/n$ . Hence the probable error in the standard deviation will be, not

$$.67449 \sqrt{\frac{\mu_2 - \mu_2'}{4\mu_2}} / n \dots\dots\dots (iii),$$

but

$$.67449 \sqrt{\frac{\pi_2 - \pi_2'}{4\mu_2}} / n \dots\dots\dots (iv)."$$

Now while it is clear that *theoretically* the uncorrected moments should be used, will it *practically* make any significant difference in the results if the probable errors of the mean and standard deviation are calculated from the corrected moments? This is our second problem.

To get a "working" answer to these two questions it was decided to calculate for a series of curves the probable error of the standard deviation by each of the different methods possible. These possibilities are, of course, first to calculate the probable error by (i) with (a) corrected and (b) "raw" moments, and then by (ii) using again the corrected and "raw" values of  $\sigma$  successively. With these results in hand for a number of curves we shall be able to form a practical judgment of the necessity or desirability of using the more refined methods in ordinary biometrical investigations. In choosing curves on which to make these tests it was decided to take a rather extreme example of each of Pearson's six types of curves. We may thus expect to get some idea of the maximum effect produced by calculating the probable error under discussion in the several different ways.

The curves actually chosen for the work were the following:

*Type I.* Curve for variation in leaf number of whorls on "all branches" of certain *Ceratophyllum* plants†. This is an extreme example of a skew curve of Type I. It is figured on p. 34 of the memoir cited.

\* *Biometrika*, Vol. v. p. 455, 1907.

† Pearl, R., Pepper, O. M., and Hagle, F. J. "Variation and Differentiation in *Ceratophyllum*." Carnegie Institution of Washington, Publication No. 58, p. 32.

*Type II.* A curve from some unpublished material in the hands of the writer.

*Type III.* Curve of frequency of barometric heights, Churchstoke station\*. This curve is figured as Fig. V, of Plate II, of the memoir cited.

*Type IV.* Curve of variation in length of the carpopodite of leg III in the crayfish *Cambarus propinquus* Girard†.

*Type V.* Curve of variation in the number of lips in the medusa, *P. pentata*‡. The curve is figured on p. 456 of the memoir cited.

*Type VI.* Curve of variation in leaf number in main stem whorls of *Ceratophyllum*§. The curve is figured on p. 34 of the memoir cited.

Certain of the constants of these curves are given in Table I. From the constants it will be seen that all these curves differ widely from the normal or Gaussian curve. Taken as a whole they are fairly representative of the conditions which one finds in frequency curves in practice when large numbers of individuals are dealt with. It will be noted that in some of these curves as tabled, Sheppard's corrections of the moments have been used, while in other cases they have not.

TABLE I. *Constants of Curves.*

Constant	Curve of Type I	Curve of Type II	Curve of Type III	Curve of Type IV	Curve of Type V	Curve of Type VI
N	1954	375	4018	283	996	374
Unit of grouping	1 leaf	15	1/10 in.	3/10 mm.	1	1 leaf
$\mu_2$	1.5350	6.3761	12.6642	7.2794	.3080	.8798
$\mu_4$	5.8052	111.2316	511.4453	246.7453	1.1817	4.0691
$\beta_1$	.2112	.0008	.1258	.7276	4.1683	1.1928
$\beta_2$	2.4637	2.7360	3.1889	4.6565	12.3760	5.2568
$\kappa_1$	-1.7060	-.5303	+ .0005	+ 1.1302	+ 6.2469	+ .9354
$\kappa_2$	-.1002	-.0011	+ 198.5780	+ .5738	+ 1.0659	+ 1.2455
$\sigma^2$	1.2390	37.8765	.3559	.809	.5559	.9390
Skewness	-.6119	-.0171	+ .1773	+ .3293	-.5294	-.4270
Sheppard's corrections	Not used	Used	¶	Used	Not used	Not used

We may turn now to the consideration of our first problem. Taking the values of the moments given in Table I for the six curves I have calculated the probable error of the standard deviation for each curve, first according to equation (i) above, and then according to equation (ii). It will be noted that this procedure takes no account of whether Sheppard's corrections have been applied to the moments of the several distributions. It is here assumed that (i) is the absolutely true formula in every case. The results are shown in Table II. This table further gives the absolute differences for each pair of probable errors calculated in the two ways. Finally an expression of the relative magnitude of the error made by the use of the approximate formula was obtained by finding in each case what percentage of the true probable error the

\* Pearson, K., and Lee, A. *Phil. Trans.* Vol. 190, A, pp. 423—469, 1897.

† Pearl, R., and Clawson, A. B. "Variation and Correlation in the Crayfish." Carnegie Institution of Washington, *Publication No. 64*, p. 7, Table I.

‡ Pearson, K., *Phil. Trans.* Vol. 197, A, p. 455, 1901.

§ Pearl, Pepper, and Hagle. *Loc. cit.* p. 32.

¶ Given throughout in concrete units, not in units of grouping.

¶ The moments in this case were corrected by considering the frequency areas as trapezia. Cf. Pearson, *Phil. Trans.* Vol. 186, A, p. 350.

approximate one was and then subtracting the percentage from 100. The figures so obtained are given in the last line of the table opposite the entry "Relative Error."

TABLE II.  
*Probable Errors of Standard Deviations.*

P. E. of $\sigma$	Curve of Type I	Curve of Type II	Curve of Type III	Curve of Type IV	Curve of Type V	Curve of Type VI
Calculated by (i) ...	·0114	1·0149	·0028	·0310	·0200	·0337
Calculated by (ii) ...	·0134	1·0893	·0027	·0229	·0084	·0231
Absolute Difference ...	·0019	·0644	+·0001	+·0081	+·0116	+·0106
Relative Error ...	17 %	7 %	4 %	26 %	58 %	31 %

From this table we note at once the following facts:

(a) As was to be expected the approximate formula for P.E. of  $\sigma$  in the case of these skew curves never gives the true value.

(b) The probable error obtained by the approximate formula may be either in excess or defect of the true value.

(c) The amount of the deviation of the approximate from the true value varies in the different curves but may be very considerable. Thus in the case of Type V curve the probable error from the approximate formula is less than half as large as it should be. In the Type VI curve the approximate value is but 69 % of the true, and the Type IV curve 74 %. It will be noted that not only are these deviations of the approximate from the true values large in amount, but further they are in the direction most likely to cause serious error in practical biometrical work. It is not so serious a matter when an approximate formula makes a probable error too large as when it makes it too small.

(d) In the case of the curves of Types I, II and III the deviations of approximate from true values are all small, and in the first two cases the approximate value is in excess of the true. So far as we may judge from the curves here discussed, it would appear that the most serious difficulty from the use of the approximate probable error formula is to be expected in curves of Types IV, V and VI. Thinking that possibly the great deviations observed in the case of Type V and VI curves might be exceptional and due to the particular examples chosen, I took another Type VI curve and calculated the P.E. of  $\sigma$  as before. This curve is one discussed by Pearson (*Phil. Trans.* Vol. 197, A, p. 451). It gives the variation in age of the brides in 28,454 Italian marriages. The values of the moments indicate a curve of Type VI but the criterion  $\kappa_3$  is so nearly 1 that a Type V curve gives a good graduation. The results of the probable error calculations (using corrected values of the moments) are shown in the following table:

P. E. of $\sigma$	Curve of Type VI
Calculated by (i) ...	·0177
Calculated by (ii) ...	·0103
Absolute Difference ...	+·0074
Relative Error ...	42 %

We see here just as before that the approximate value is largely in defect of the true value.

(e) Considering Table II in connection with Table I it is evident that there is no simple relation between what I have called the "Relative Error" and any of the constants of the curves. That is to say, even though we know the ordinary constants of a given curve it will not be possible to say beforehand whether the approximate formula for the P.E. of  $\sigma$  will give a reasonably good value in that case.

Taking into account all these facts I think there can be no doubt regarding the answer to our first question. The answer definitely is that when a distribution is not closely normal, there is a considerable likelihood that the assumption of normality made in calculating the probable error of the standard deviation by the approximate formula may introduce an error great enough to be of practical importance in drawing conclusions. Whenever a piece of biometric work involves the comparison of standard deviations to determine whether they are significantly different among themselves great caution should be exercised about making conclusions depend on probable errors calculated by the approximate formula.

We may now turn to the consideration of our second problem (cf. p. 113 *supra*). In attempting to get at an answer to it the plan was to calculate for each of the curves given in Table I (with the exception of the Type III curve) the probable error of  $\sigma$  first according to equation (iv) and then according to equation (iii), using, of course, the proper values of  $\mu$  and  $\pi$ . The same calculations were also made for the curve obtained from the Italian marriage statistics (Type VI) cited above. The results of these computations are set forth in Table III. The "Relative Errors" in this case are obtained as before by taking the percentage which the approximate is of the absolutely true value and subtracting this percentage from 100.

TABLE III

*Probable Errors of Standard Deviations. Influence of Sheppard's Correction.*

	Curve of Type I	Curve of Type II	Curve of Type IV	Curve of Type V	Curve of Type VI	Curve of Type VI (Italian Marriages)
$\cdot 67449 \sqrt{\frac{\pi_1 - \pi_2^2}{4\mu_1}} / n$	$\cdot 0118$	$1\cdot 0301$	$\cdot 0312$	$\cdot 0234$	$\cdot 0355$	$\cdot 01775$
$\cdot 67449 \sqrt{\frac{\mu_1 - \mu_2^2}{4\mu_1}} / n$	$\cdot 0108$	$1\cdot 0149$	$\cdot 0310$	$\cdot 0226$	$\cdot 0340$	$\cdot 01771$
Difference ...	$+\cdot 0009$	$+\cdot 0152$	$+\cdot 0002$	$+\cdot 0009$	$+\cdot 0015$	$+\cdot 00004$
Relative Error ...	$7\cdot 8\%$	$1\cdot 5\%$	$0\cdot 6\%$	$3\cdot 8\%$	$4\cdot 2\%$	$0\cdot 2\%$

From this table we note:

(a) That the absolute differences resulting from calculating the probable errors in the two different ways are in every case very small. In only one case (the Type II curve) is the absolute difference greater than 1 in the third decimal place. Now it is, of course, obvious that in ordinary statistical work the probable errors will very rarely be tabled to more than the third place of figures. It would appear then, from the absolute differences here, that the error made by using the approximate formula (iii) instead of (iv) is for *practical purposes* insignificant.

(b) Relative to the magnitude of the probable errors themselves the differences are very small. It is seen from the last line of the table that approximately 8% is the maximum relative error made and in the majority of cases the relative error is much smaller than this.

From these results the answer to our second question is, I think, clear, and may be stated in the following way: it is *not* a matter of practical significance in ordinary statistical work that the raw moments be used to calculate the probable error of a standard deviation which has itself been deduced from a moment to which Sheppard's correction has been applied. The error made by using equation (iii) instead of equation (iv) is not likely in the great bulk of ordinary biometrical work ever to reach the magnitude where it will be of any *practical* consequence whatever. One should, of course, always keep in mind the existence of this source of error, and if very fine, close work is to be done take pains to eliminate it. It is obvious, however, that the occasions where it will have to be taken into account will be extremely rare.

Putting all our results together we may summarily state the practical conclusion as follows:

In calculating the probable error of the standard deviation of frequency distributions which deviate sensibly from the normal curve, great caution must be exercised in using the customarily applied formula for this probable error  $\left( \cdot 67449 \frac{\sigma}{\sqrt{2n}} \right)$ . The use of this approximate formula is not unlikely to lead to serious error when standard deviations are compared on the basis of their probable errors. It will be safest, in the absence of special investigation of the point, for each curve or group of curves under treatment, to calculate the probable error of  $\sigma$  by the formula  $\cdot 67449 \sqrt{\frac{\mu_4 - \mu_2^2}{4\mu_2}} / n$ . In using this formula for the probable error it is practically immaterial whether the moments used have or have not been corrected according to Sheppard's formulae.

[*Note.* The standard deviation of the standard deviation  $\sigma$  is known to be  $\frac{\sigma}{\sqrt{2n}} \sqrt{1 + \frac{1}{2}\eta}$ , where  $\eta$  is the kurtosis  $= \beta_2 - 3$ . For  $\eta$  small the multiplying factor is  $1 + \frac{1}{4}\eta$ , and supposing we may neglect for practical purposes an error of 5% in a probable error, we conclude that the ordinary formula may be applied as long as the kurtosis lies between  $\pm 2$  or for  $\beta_2 = 2.3$  to 3.2. This agrees with Dr Pearl's Table II, where only the curve of Type III has its kurtosis within this range. Thus the kurtosis alone, independently of the type, settles whether the ordinary formula is sufficient or not. A further point not yet discussed but of some importance is the significance of the "probable error" at all when we are taking the standard deviation of standard deviations in the case of very skew material. How far does the distribution of standard deviations follow a normal curve? The answer depends largely, I think, on the size of the sample, and the deduction of conclusions from the "probable error" of  $\sigma$  may lead to greater error than even neglecting  $\eta$ , unless we are fairly certain that our  $\sigma$  follows a normal distribution. See the paper by "Student" in the present number of this Journal. K. P.]

### VIII. Addendum to Memoir: "Split-Hand and Split-Foot Deformities."

*Biometrika*, Vol. VI. pp. 26—58.

Since the final proof sheets of the above paper were sent to press, several papers have appeared to which reference is necessary.

The evidence of the Bateson School to show that Mendelism is applicable to deformities in man was criticised in the memoir. At that time stress had been laid on three family trees, that of Farabee, an instance of hypophalangia, and those of Nettleship, instances of congenital cataract. The tendency for the deformity to abate in successive generations was noted, for in these families the last generation in each is the only one which shows the undeformed out-