



VIII. Exercises in the calculation of errors

Professor F. Y. Edgeworth M.A. D.C.L.

To cite this article: Professor F. Y. Edgeworth M.A. D.C.L. (1893) VIII. Exercises in the calculation of errors , Philosophical Magazine Series 5, 36:218, 98-111, DOI: [10.1080/14786449308620453](https://doi.org/10.1080/14786449308620453)

To link to this article: <http://dx.doi.org/10.1080/14786449308620453>



Published online: 08 May 2009.



Submit your article to this journal [↗](#)



Article views: 3



View related articles [↗](#)



Citing articles: 7 View citing articles [↗](#)

VIII. *Exercises in the Calculation of Errors.*
 By Professor F. Y. EDGEWORTH, M.A., D.C.L.*

AN attempt to estimate and reduce the errors in the calculation of *correlations* between organs † may prove not only useful to those who cultivate that branch of exact biology, but also interesting to the less specialized student of Probabilities, in virtue of some precepts of general application.

As shown in former papers, the calculation of correlations consists of three stages. We begin with the coefficients called by Mr. Galton r , and by the present writer $\rho_{12}, \rho_{13}, \&c.$; which determine the correlation between the members of each pair of organs; for the probability of any two deviations x and y (each measured from the corresponding average, in units of the proper modulus) being associated is equal to a constant raised to a power denoted by the following expression :

$$\frac{1}{\sqrt{1-\rho_{12}^2}}x^2 - 2\frac{\rho_{12}}{\sqrt{1-\rho_{12}^2}}xy + \frac{1}{\sqrt{1-\rho_{12}^2}}y^2.$$

The conclusion sought is the quantic of the second degree which forms the exponent of the expression for the probability that particular values of all the organs should be associated; which in the case of three variables is of the form

$$ax^2 + by^2 + cz^2 + 2fyz + 2gxz + 2hxy.$$

This result affords answers to questions like the following :—(1) Given the values of one or more of the variables, what are the values of the remaining variables which are most probably associated with the given ones? (2) Given the values of one or more of the variables, what is the dispersion of each of the other variables about its most probable value ‡?

Intermediate between the first and last stage of the calculation is the determination of the *proportionate* values of the coefficients $a, b, c, f, g, \&c.$; by which we are able to answer questions of the first kind, but not of the second §.

I propose to estimate for each of these three computations the *error* to which it is liable; that is the extent to which the results obtained from a given number of specimens are likely to differ from the results which would be obtained from an indefinitely large number of specimens.

I. The ρ -coefficients which come first are each liable to errors of two or three kinds :—

* Communicated by the Author.

† See papers by the present writer in *Phil. Mag.*, Aug., Nov., Dec. 1892, and Jan. 1893.

‡ See *Phil. Mag.* Jan. 1893.

§ *Ibid.*

(α) One source of error is at the origin or centre from which are measured the deviations, x , y , &c., which form the data for determining the required coefficients. The error in the determination of the centre will depend upon the method of determining it. For the purpose in hand I recommend the sort of method which Mr. Galton has pursued: the use of "percentiles," together with some process of "smoothing"*. The simplest variety of this generic principle is to use the *observed* quartiles and median in order to determine the *most probable* values of those points consistent with the condition that the median should be midway between the two quartiles. Thus, if q_1 , q_2 , m be the observed quartiles and median respectively, and the sought ones Q_1 , Q_2 , M ; we have to determine the latter, under the condition that the expression

$$w_1(Q_1 - q_1)^2 + w(M - m)^2 + w_2(Q_2 - q_2)^2 + 2\lambda(Q_1 + Q_2 - 2M)$$

should be a maximum; where λ is an undetermined factor, w is the *weight* pertaining to the determination of the median by putting the observed median (m) for the real one (M); and w_1 , w_2 are the corresponding weights for the quartiles respectively. Whence

$$Q_1 = q_1 - \frac{\lambda}{w_1}; \quad Q_2 = q_2 - \frac{\lambda}{w_2}; \quad M = m + \frac{2\lambda}{w}.$$

And, to determine λ , we have

$$q_1 + q_2 - \frac{\lambda}{w_1} - \frac{\lambda}{w_2} = 2m + \frac{4\lambda}{w}.$$

Now the error committed in taking the *observed* as the *real* median has for modulus $\sqrt{\frac{\pi}{2n}}$, as Laplace has proved.

And, as the present writer following his method has reasoned, the error committed in taking each observed quartile for the real one has for modulus $\sqrt{\frac{\pi}{1.7n}}$ †. Accordingly $\frac{2\lambda}{w}$, the

* Proc. Roy. Soc. vol. xlv. p. 140.

† Phil. Mag. 1886, vol. xxii. p. 375.

The scruples expressed in the passage referred to are groundless. Laplace's method is justified by the presumption that a variable, such as the position of the median, depending on a number of independent agencies, obeys the law of error (see Phil. Mag. Nov. and Dec. 1892). Also the displacements of the two quartiles are independent of each other and of that of the median; as follows from the theory that each displacement is of the order $\frac{1}{\sqrt{n}}$ (with reference to unit of modulus). This consideration shows how many percentiles—quartiles, octiles, &c.—ought to be utilized in order to employ the generic principle to the most advantage. At least

correction of m ,

$$= \frac{q_1 + q_2 - 2m}{3 \cdot 2};$$

the corrected value of m

$$= \frac{1 \cdot 2m + q_1 + q_2}{3 \cdot 2};$$

and the modulus of the error incident to this determination

$$= \sqrt{1 \cdot 44\pi \div 2 + 2\pi \div 1 \cdot 7 \div 3 \cdot 2} \sqrt{n} = \cdot 77 \div \sqrt{n}.$$

Thus by the use of the principal percentiles, the median and the quartiles, with a simple process of smoothing, a result is obtained which is better than that combination of observations which has been thought to be the best, viz., the Arithmetic mean; for which

the modulus of error is $\frac{1}{\sqrt{n}} \left(> \frac{\cdot 77}{\sqrt{n}} \right)$.

The error which has been attributed to the determination of the centre affects of course each deviation which is measured from that centre as origin. But the influence which the error of the observed deviations x and y has upon the coefficient of correlation ρ_{12} cannot be estimated, until the method of combining the former in order to determine the latter has been assigned. The methods which present themselves may be classified as (1) the most accurate, (2) the more convenient; each introducing an error additional to those which have been indicated under the heading (α).

(β) (1) Regarding each assigned, or "subject" * x divided into the associated, or "relative," y , as affording an observation-equation $\frac{y}{x} = \rho_{12}$, we see that the best combination of these data is obtained by affecting each observation $\frac{y}{x}$ with a weight inversely proportional to its modulus-squared. Now by hypothesis every y , whatever the x with which it is associated, has for the modulus of its fluctuation $\sqrt{1 - \rho_{12}^2}$ †.

as many as suffer independent displacements may with advantage be admitted; probably at least the octiles and deciles in general.

It is important to observe that the principle may be extended to "discordant" observations which do not range under a single probability-curve; in which case the w 's are to be determined according to first principles, from the height of the ordinate in each neighbourhood (Phil. Mag. *loc. cit.*).

* See Galton, Proc. Roy. Soc. 1888, p. 140.

† See the formula on p. 98 above.

Accordingly the modulus of $\frac{y}{x}$ is inversely proportional to x ; and the weight of $\frac{y}{x}$ is directly proportional to x^2 . The best value of ρ_{12} is $Sxy \div Sx^2$; and the error of this determination has for modulus

$$\sqrt{Sx^2(1-\rho^2)} \div Sx^2 = \sqrt{1-\rho^2} \div \sqrt{Sx^2} = \sqrt{1-\rho^2} \div \sqrt{\frac{1}{2}n}$$

(the modulus of x being unity by hypothesis)

$$= \sqrt{2} \times \sqrt{1-\rho^2} \div \sqrt{n}$$

(n being the number of observations utilized).

The laborious multiplication which the formula $Sxy \div Sx^2$ involves may be abridged by grouping the x 's in small fascicules*.

(2) More convenient methods of utilizing the data are (i.) that which I have recommended in a former paper †: dividing the sum of the assigned deviations Sx into the sum of the associated deviations Sy ; and (ii.) the method to which Mr. Galton's statistics lend themselves ‡: arranging the observations in small groups, taking the quotient $Sy \div Sx$ for each group, and the arithmetic mean of all these quotients as the value of ρ .

(i.) The modulus of the error incident to the expression $Sy \div Sx$ (x assigned and y observed to be associated) is

$$\sqrt{n(1-\rho^2)} \div Sx = \sqrt{\pi} \sqrt{1-\rho^2} \div \sqrt{n}$$

(the modulus for the deviation x being unity); worse than the best method in the degree $\sqrt{\pi} : \sqrt{2}$, or 1.25 times.

This result may be improved by omitting some of the data; in virtue of the following general theorem:—

When, instead of the proper weights of a set of observations $p_1^2, p_2^2, \&c.$, another set of weights, q_1^2, q_2^2 , are used as more convenient, it is in general advantageous to reject some of the given observations; the last observation admitted, say x_m , being determined by the equation

$$q_{m+1}^2 \times S_1^m q^2 = 2p_{m+1}^2 \times S_1^m \frac{q^4}{p^2}.$$

This equation is derived from the condition that we should stop taking in new observations as soon as the modulus of the

* The principles which should regulate this sort of approximation may be gathered from the present writer's paper on the "Determination of the Modulus" in the Philosophical Magazine, 1886, vol. xxi. p. 500.

† Phil. Mag. Aug. 1892.

‡ Proc. Roy. Soc. vol. xlv. p. 139.

weighted mean ceases to become smaller by the addition of a new observation ; that is, when

$$S_1^{m+1} q^4 c^2 \div (S_1^{m+1} q^2)^2$$

begins to be greater than

$$S_1^m q^4 c^2 \div (S_1^m q^2)^2, \quad \left(\text{where } c^2 = \frac{1}{p^2} \right).$$

Expanding, and neglecting terms of an inferior order, we have the equation above written.

To apply this principle to the case before us, we may, without loss of generality, regard the given values of x as ranged on one branch of a probability-curve from zero to infinity. Each x divided into the associated y gives a value for ρ_{12} of which the true weight is x^2 and the weight used in the method under consideration is x . Putting $p^2 = x^2$ and $q^2 = x$ in the general formula, we have for the limiting condition :

$$x_{m+1} \times S_1^m x \text{ begins to be greater than } 2x_{m+1}^2 \times S_1^m 1,$$

or

$$x_{m+1} \text{ begins to be less than } \frac{1}{2} S_1^m x \div n.$$

Now, as we move outwards from the centre towards infinity, the right-hand member of the above inequality converges to a constant $\frac{1}{\sqrt{\pi}}$, while the left-hand member increases indefinitely. There is therefore no upper limit.

To determine the lower limit, putting u for x_{m+1} , find a point on the x -axis of a probability-curve (of unit modulus), distant u from the origin such that u = half the distance from the origin of the centre-of-gravity of the area which is intercepted by the axis of abscissæ, the ordinate at the point u , and the curve. In symbols,

$$u = \frac{1}{2} \int_u^\infty x e^{-x^2} dx \div \int_u^\infty e^{-x^2} dx,$$

$$u \times 4 \times e^{+u^2} \times \int_u^\infty e^{-x^2} dx = 1.$$

Taking logarithms, and using the second of the tables appended to De Morgan's "Calculus of Probabilities" (*Encyc. Metrop.* vol. ii.), I find a value for u between .4 and .45 ; which is presumably the solution.

But, as it may be suspected that this result is exaggerated by the prolongation of the curve to infinity in theory, though not in fact, I have verified the solution by substituting for the *centre of gravity* the *Median* of the tract outside the ordinate

at the required point. This calculation being not disturbed by the leverage of a limb stretching to infinity, affords an inferior limit to the true value : say u' , determined from the equation

$$\int_{u'}^{\infty} \frac{1}{\sqrt{\pi}} e^{-x^2} dx = 2 \int_{2u'}^{\infty} \frac{1}{\sqrt{\pi}} e^{-x^2} dx.$$

The root of this equation is found from the Tables to lie between .35 and .4. Concluding, then, that the limit is in the neighbourhood of .4, we ought to reject about *two fifths* of the observations (that being the proportion of the deviations which fall within a distance of .38 from the centre).

The gain in accuracy is seen by comparing the modulus of error before correction, viz.

$$\sqrt{\pi} \sqrt{1-\rho^2} \div \sqrt{n},$$

with the modulus after correction, that of the expression $Sy \div Sx$ integrated between limits ∞ and (say) .4. Each y , as before, being liable to an error whose modulus is $\sqrt{1-\rho_{12}^2}$, we have for the modulus of error incident to ρ_{12} ,

$$\begin{aligned} & \sqrt{1-\rho_{12}^2} \left[n \times \frac{2}{\sqrt{\pi}} \int_{.4}^{\infty} e^{-x^2} dx \right]^{\frac{1}{2}} \div n \frac{2}{\sqrt{\pi}} \int_{.4}^{\infty} x e^{-x^2} dx \\ &= \sqrt{1-\rho_{12}^2} \times \sqrt{n} \sqrt{.572} \div \frac{n}{\sqrt{\pi}} e^{-.16} \\ &= \sqrt{1-\rho_{12}^2} \sqrt{\frac{\pi}{n}} \times .756 \times e^{+.16} = \sqrt{1-\rho_{12}^2} \times \sqrt{\frac{\pi}{n}} \times .89 \text{ nearly.} \end{aligned}$$

Thus the modulus of the purified observations is about ten per cent. smaller than the modulus of the whole set. To this slight gain in accuracy is to be added a considerable saving of trouble. As compared with the best possible method, the corrected second-best is very much less troublesome and very little less

accurate—the modulus of the former being $\sqrt{\frac{1-\rho_{12}^2}{n}} \sqrt{2}$,

the modulus of the latter $\sqrt{\frac{1-\rho_{12}^2}{n}} 1.55$.

(ii.) We have next to consider the method of treating the statistics to which Mr. Galton's tables in the paper already cited lend themselves. For each degree or small difference, *e. g.* a tenth (of the unit modulus), on the axis x take the mean of the corresponding y 's ; and put the latter, divided by the former, as a value of the *quæsitum* ρ_{12} : *e. g.*,

$$\rho'_{12} = \frac{S'y}{s'} \div x' + \cdot 05 ;$$

where $S'y$ is the sum, and s' the number, of the y 's corresponding to values of x which are between x' and $x' + \cdot 1$. Determine similarly ρ_{12}'' , ρ_{12}''' , &c. for other degrees or differences on the axis x ; and take the Arithmetic Mean of ρ_{12}' , ρ_{12}'' , &c. as the value of ρ_{12} . As each little group of observations has the same weight in this combination, it follows that there is assigned to each observation a weight inversely proportional to s_r , the size of the group to which the observation belongs. But as the values of x are distributed in conformity with a probability-curve with modulus unity, the number s_r is proportional to e^{-x^2} ; and accordingly the weight used is proportional to e^{+x^2} .

To determine the accuracy of this method uncorrected by the rejection of observations, we have for the modulus of the weighted mean

$$\sqrt{1 - \rho_{12}^2} \left[n \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-x^2} \times e^{2x^2} dx \right]^{\frac{1}{2}} \div n \int_0^\infty \frac{2}{\sqrt{\pi}} e^{-x^2} \times e^{+x^2} dx,$$

which is infinite. There is, then, evidently in this case an upper limit.

To determine the limits at which to begin rejecting observations, we have the datum that, while the weight used for each observation of the form $\frac{y}{x}$ is e^{x^2} the true weight is x^2 .

Applying the general formula above given (recollecting that the values of x are distributed in conformity with a probability-curve whose modulus is unity), we have for the upper limit, say v ,

$$\frac{1}{v^2} e^{v^2} = 2 \int_u^\infty \frac{1}{x^2} e^{x^2} dx \div (v - u), \text{ where } u \text{ is the lower limit ;}$$

and a similar equation for the lower limit, u .

To obtain an approximate solution of these equations, we may reason thus :—Observing that the curve $y = \frac{1}{x^2} e^{x^2}$ has a minimum ordinate at the point $x=1$, let us, in order to approximate to u , make abstraction of the observations outside that point, and determine a limit u' from the equation

$$\frac{1}{u'^2} e^{u'^2} = 2 \int_1^{u'} \frac{1}{x^2} e^{x^2} dx \div (1 - u').$$

The value of u' thus determined is less than the true value of u . For it may be shown that some of the observations

above the point $x=1$ had better be taken in ; and accordingly, if they are omitted, the modulus of the weighted mean is greater than if they were taken in. But, if the same weighted observations are incorporated with two averages, one of which has a larger modulus than the other, the limit at which it will cease to be advantageous to take in a new observation will be reached sooner in the case of the smaller modulus. Thus, as we move inwards towards the centre, we shall reach u sooner than u' ; in other words, u' is an inferior limit to u .

To determine u' , in the absence of tabulated integrals for the function $\frac{1}{x^2}e^{x^2}$, I have adopted a method which is in fact more appropriate to the case in hand—where the observations are broken up into little bundles :—I have plotted the ordinate of the curve $y = \frac{1}{x^2}e^{x^2}$ for the points $x=1, \cdot 9, \cdot 8, \cdot 7$, &c. ; and, joining their tops, observed the point at which $y \times (1-x)$ begins to be greater than twice the area contained between the ordinate at the point x , the abscissa, the ordinate at the point 1, and the locus joining the tops of the ordinates. This limit proves to be about $\cdot 3$. Accordingly at least all the observations below this value of x are to be rejected.

By similarly operating on the observations above the point $x=1$, I find for v' , a limit *superior* to v , the point $x=2$; and conclude that at least all the observations above that point are to be rejected.

(γ) A third source of error affecting the computation arises from the imperfect graduation of our instruments and senses, by which we are compelled to put for the true value of any object measured a value in the neighbourhood equal to an integer number of degrees, *e. g.*, tenths of an inch. Assuming that the difference between the apparent and real value is more likely to fall short of than to exceed half a degree, and is very unlikely to exceed a whole degree,—the modulus of the error from this source affecting each observation is much less than a degree ; and is therefore small, if the degree is small, as in the case of *stature*, if the degree be a tenth of an inch ; the unit in which x is measured being 3·6 inches. In the case of the *cubit* the unit is smaller, the error from source γ greater.

In computing the errors investigated above, and for other problems, the following notation will be convenient. Let the symbol *plus* written in full [and similarly *minus*] connecting *probable errors* or *moduli* denote the cumulation of the errors,

with attention *if necessary* to their *sign*. Thus, e, e_1, e_2 being the moduli of independent errors,

$$e_1 \text{ plus } e_2 = \sqrt{e_1^2 + e_2^2},$$

$$e_1 \text{ minus } e_2 = \sqrt{e_1^2 + e_2^2}.$$

But $(e_1 \text{ plus } \lambda e) \text{ plus } (e_2 \text{ minus } \mu e)$

$$= \sqrt{e_1^2 + e_2^2 + (\lambda - \mu)^2 e^2}.$$

To exemplify the combination of the errors with which ρ is affected, let us select out of the methods comprised under heading (β) that which is most easily handled, namely, the uncorrected form of $\beta(2)$ (i.); according to which $Sy \div Sx$ is put for ρ_{12} , the summation extending from positive to negative infinity. Since the ordinate of the centre from which the y 's are measured is liable to the error $\cdot77 \div \sqrt{n}^*$, Sy is liable to the error $\cdot77 \times \sqrt{n}$. By parity Sx is liable to the error $\cdot77 \times \sqrt{n}$. Also $Sy = n \div \sqrt{\pi}$ (nearly) $= Sx$ (nearly); since each set of observations ranges under a probability-curve of which the modulus is unity. Accordingly the error of ρ_{12} derived from source (α)

$$\begin{aligned} &= \frac{1}{Sx} \text{ error } Sy \text{ minus } \frac{Sy}{(Sx)^2} \text{ error } Sx \\ &= \left[\frac{\cdot58 \pi}{n} + \frac{\cdot58 \pi}{n} \right]^{\frac{1}{2}}. \end{aligned}$$

To this is to be superadded the error from source (β) †;

$$\begin{aligned} &\sqrt{\frac{1\cdot16 \pi}{n}} \text{ plus } \sqrt{\frac{(1 - \rho_{12}^2) \pi}{n}} \\ &= \sqrt{\frac{(2\cdot16 - \rho_{12}^2) \pi}{n}}. \end{aligned}$$

An illustration of this reasoning is afforded by the computation of correlation ‡ between *stature*, *cubit*, and *knee-height*, referred to in the Philosophical Magazine for December 1892, p. 523, note. There different values were obtained for each ρ , according as the positive or the negative deviations were operated on; and according as the formula used for ρ was $Sy \div Sx$ (x being taken as "subject," and y relative §) or $Sx \div Sy$ (*vice versa*).

* Above, p. 100.

† Above, p. 101.

‡ The statistical materials were supplied by Mr. Francis Galton, F.R.S. and the arithmetical work by Mrs. Bryant, D.Sc.

§ Above, p. 100.

	1.	2.	3.	4.
	Positive deviation.	Negative deviation.	x "subject."	y "subject."
ρ_{12}	·80	·80	·75	·83
ρ_{23}	·87	·82	·81	·87
ρ_{31}	·86	·87	·93	·80

Each of these results is based upon two sets of 150 observations ; but these sets are not *independent*. For instance, the basis of the "positive deviations" (column 1) consists of 150 observations (for *e. g.*, ρ_{12}) of the form $\frac{y}{x}$, and 150 of the form $\frac{x}{y}$; taking the average of the two values (of ρ_{12}) respectively determined from the two sets of 150 observations. Now the error under head (α) affecting the sum of these is,—if ϵ_1 is put for the error of the abscissa of the Median, and ϵ_2 for that of its ordinate—

$$\sqrt{\frac{\pi}{n}} (\epsilon_2 \text{ minus } \epsilon_1) \text{ plus } \sqrt{\frac{\pi}{n}} (\epsilon_1 \text{ minus } \epsilon_2);$$

that is, zero. The result therefore is affected only with the error (β), that is,

$$\sqrt{\frac{\pi(1-\rho_{12}^2)}{n}};$$

or, putting $\rho = \cdot 8$, and $n = 300$, $\cdot 06$. This result again has to be diminished by about ten per cent.; considering that in the calculation the *corrected* method ((i.)) was employed*. Similar remarks apply to the "negative deviation"—the results in the second column of the table. The *probable error* of the difference between these two results is

$$\cdot 06 \times \cdot 9 \times \sqrt{2} \times \cdot 477, \text{ or } \cdot 04 \text{ nearly.}$$

Similarly the basis of the results in the third column is one set of 150 positive observations and another set of 150 negative observations. The errors under head (α) which affect each of these sets cut each other out; and accordingly the probable error of the difference between an entry in column 3 and the corresponding entry in column 4 is the

* Above, p. 101.

same as that for the differences between the entries in columns 1 and 2, namely $\cdot 04$.

To these estimates an addition is to be made in virtue of the error of the species (γ). It will be found that the errors of this species are nearly cancelled in both 1 and 2, but not so in 3 or 4; a circumstance which perhaps accounts for the considerable divergence between 3 and 4 as compared with that between 1 and 2.

The differences actually occurring are 0, $\cdot 01$, $\cdot 05$, $\cdot 06$, $\cdot 07$, $\cdot 08$.

II. We have next to consider the errors incident to the *proportional* values of the coefficients a, b, c, f, g, h —the case of three variables being taken as an example. It will be recollected that these proportions are thus obtained. The ratios $\frac{a}{\Delta}, \frac{b}{\Delta}, \frac{c}{\Delta}$ are respectively equal to the principal minors, and the ratios $\frac{f}{\Delta}, \frac{g}{\Delta}, \frac{h}{\Delta}$ to the other minors, of the determinant

$$\begin{vmatrix} 1 & \rho_{12} & \rho_{31} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{31} & \rho_{23} & 1 \end{vmatrix}$$

Accordingly the first set of results, the proportional coefficients of the squares of the variables, have the least possible relative error when the ρ 's are all zero (the deviations of the organs independent); and the greatest possible relative error when the ρ 's are each unity (the correlation a case of simple law unmixed with chance). Contrariwise, the relative error of $\frac{f}{\Delta}, \frac{g}{\Delta}, \frac{h}{\Delta}$, the proportional coefficients of the products yz, zx, xy , is greater, the less the coefficients ρ are.

The relative error of $\frac{f}{\Delta}, \frac{g}{\Delta}$, &c. is apt to be greater than that of the ρ 's from which they are calculated. For, put e_{12}, e_{23} , &c. as the absolute errors of ρ_{12}, ρ_{23} , &c. Then the relative errors of ρ_{12}, ρ_{23} , &c. are $e_{12} \div \rho_{12}, e_{23} \div \rho_{23}$, &c.; while the relative error of, for example, $\frac{f}{\Delta}$ ($= \rho_{12}\rho_{13} - \rho_{23}$) is

$$(\rho_{12}e_{13} \text{ plus } \rho_{13}e_{12} \text{ minus } e_{23}) \div (\rho_{12}\rho_{13} - \rho_{23});$$

which is made up of three terms, each of which seems as likely as not to be of the same order as $e_{23} \div \rho_{23}$ (it being recollected that all the ρ 's are proper fractions).

III. We pass to the *absolute* values of the coefficients by

multiplying each of the proportional coefficients above written by Δ , the determinant of the quantic

$$ax^2 + by^2 + cz^2 + 2fyz + 2gzx + 2hxy.$$

The relative error of the divisor $\frac{1}{\Delta}$ has the following elegant expression :—

$$2[fe_{23} \text{ plus } ge_{31} \text{ plus } he_{12}].$$

The relative error—or briefly *r.e.*—of any of the principal coefficients a

$$\begin{aligned} &= r.e. \left(\frac{a}{\Delta} \div \frac{1}{\Delta} \right) = r.e. \frac{a}{\Delta} \text{ minus } r.e. \frac{1}{\Delta} \\ &= \left(\text{since } \frac{a}{\Delta} = 1 - \rho_{23}^2 \right) - \frac{\Delta}{a} 2\rho_{12}e_{12} \text{ minus } r.e. \frac{1}{\Delta} \\ &= (\text{by the formula of last paragraph}) \\ &2 \left[\left(f + \frac{\Delta\rho_{12}}{a} \right) e_{23} \text{ plus } ge_{31} \text{ plus } he_{12} \right]. \end{aligned}$$

If the ρ 's are determined as above on p. 104, the α element of error disappears altogether, and the γ element in part. Accordingly, putting $f + \frac{\Delta\rho_{12}}{a} = k$, and expanding, we have

$$r.e. a = 2\sqrt{\frac{\pi}{n}} \left[k \sqrt{1 - \rho_{23}^2} \text{ plus } g \sqrt{1 - \rho_{31}^2} \text{ plus } h \sqrt{1 - \rho_{12}^2} \right]$$

nearly.

Put as approximate values $\rho_{12} = .8$, $\rho_{23} = .8$, $\rho_{31} = .9$; and evaluate the coefficients $a, b \dots h$. They are approximately $a = c = 6$, $f = h = 1$, $b = 3$, $g = 4$. Substitute these values in the expression for *r.e. a*; and we have, when $n = 300$, modulus of *r.e. a* = about .5. The probable error for the difference between two determinations = $.5 \times .477 \times \sqrt{2} = .33$, nearly.

The relative errors of the other coefficients may be similarly determined, and may be expected to be equally precarious while the number of observations is limited to 300.

This anticipation is fully borne out by the following experience. From each of the four sets of ρ 's above cited there has been calculated* the exponential quantic of the second degree, or ellipsoid of equal probability. The comparative results are exhibited in the following Table.

* By Mrs. Bryant.

TABLE showing Coefficients of Correlation obtained from four batches of 300 observations each.

	ρ_{12}	ρ_{23}	ρ_{31}	a	b	c	f	g	h
Positive and negative observations; x, y, z respectively "subjects."	.75	.81	.93	7.401	2.907	9.419	-6.943	-2.637	+ .07105
Positive and negative observations; y, z, x respectively "subjects."	.83	.87	.80	3.494	5.177	4.472	-1.122	-2.962	-1.927
Positive observations only.	.80	.87	.86	4.024	4.306	5.962	-2.716	-3.014	-.861
Negative observations only.	.80	.82	.86	4.733	3.752	5.195	-1.371	-1.905	-2.944

Taking a and c as each nearly 6, we should expect *à priori* (according to the last paragraph but one) the probable error between any two determinations of each coefficient to be about 2; and, *à posteriori*, we observe that of the six differences presented by the four values of a , three are within and three without 2. A similar statement is true of the six values of c .

The precariousness of the results becomes much greater, as the number of the variables is increased, as may be seen by considering the expression for $\frac{1}{\Delta}$ in the case of four, compared with that of three, variables.

I do not think it necessary to exhibit the work in full. What has been proved of arithmetical observations may be true also of algebraic formulæ—a better general result is sometimes obtained by not working up all the particulars.

To sum up: we have estimated the error incident to each coefficient employed in determining the correlation between organs. We have shown how this error becomes greater with the number of the organs; the instability of the construction increases rapidly with its height.

In reaching these conclusions respecting correlated averages, we have come upon two principles of wider application. (1) When observations are combined according to a system of weights different from that which is known to be the best, it is in general advantageous to *reject* a certain class of the given observations.

(2) When, as usual, the observations range under a probability-curve, the median m corrected by the quartiles q_1 and q_2 affords a formula for the Mean, viz. $(1.2m + q_1 + q_2) \div 3.2$, which is more accurate than that method of combining such observations which has hitherto been supposed to be the most accurate, viz. the Arithmetic Mean. The principle may be applied with great ease and advantage to Discordant Observations*.

IX. The Hydrates of Hydrogen Bromide.

By SPENCER U. PICKERING, F.R.S.†

IN the various criticisms which have appeared in this Magazine of the conclusions which I drew from my study of the properties of sulphuric-acid solutions, no allusion has been made to one of the strongest of the arguments adduced in favour of the reality of the changes of curvature

* See the papers by the present writer in the 'Philosophical Magazine' for 1886 and 1887.

† Communicated by the Author.