TABLE I.
Constants of $z = f(x_1, x_2)$.

$z = f(x_1, x_2)$	Mean of z	Standard Deviation of z
$z = x_1 + x_2$	$m_1 + m_2$	$\sqrt{(\sigma_1^2 + \sigma_2^2 + 2r_{12}\sigma_1\sigma_2)}$
$z = x_1 + x_2 + x_3$	$m_1 + m_2 + m_3$	$\sqrt{(\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + 2r_{12}\sigma_1\sigma_2 + 2r_{13}\sigma_1\sigma_3 + 2r_{23}\sigma_2\sigma_3)}$
$z = x_1 - x_2$	$m_1 - m_2$	$\sqrt{(\sigma_1^2 + \sigma_2^2 - 2r_{12}\sigma_1\sigma_2)}$
$z = x_1 \cdot x_2$	$m_1 m_2 + r_{12}\sigma_1\sigma_2$	$m_1 m_2 [v_1^2 + v_2^2 + 2r_{12}v_1v_2 + v_1^2v_2^2(1 + r_{12})]^{\frac{1}{2}},^*$ or approximately $m_1 m_2 [v_1^2 + v_2^2 + 2r_{12}v_1v_2]^{\frac{1}{2}}$
$z = x_1 \cdot x_2 \cdot x_3$	$m_1 m_2 m_3 [1 + r_{12}v_1v_2 + r_{13}v_1v_3 + r_{23}v_2v_3]$	$m_1 m_2 m_3 [v_1^2 + v_2^2 + v_3^2 + 2r_{12}v_1v_2 + 2r_{13}v_1v_3 + 2r_{23}v_2v_3]^{\frac{1}{2}}$ approximately
$z = \frac{x_1}{x_2}$	$\frac{m_1}{m_2} (1 + v_2^2 - r_{12}v_1v_2)$	$\frac{m_1}{m_2} \sqrt{(v_1^2 + v_2^2 - 2r_{12}v_1v_2)}$

III. The Correlation between a Variable and the Deviation of a Dependent Variable from its Probable Value.

By J. ARTHUR HARRIS, PH.D., Carnegie Institute, Station for Experimental Evolution, Long Island, U.S., and Biometric Laboratory, University College, London.

The degree of interdependence of two quantitatively measurable variables is conveniently expressed by the correlation coefficient, r , or the correlation ratio, η . These constants will both be used with increasing frequency to describe many relationships in evolutionary, morphological and physiological work.

There appear to be, however, certain problems in which the correlation coefficient, while indispensable, does not yield all the information which we require.

Take such a case as the relationship between the number of eggs per clutch and the number of birds hatching. The limits of the number of birds hatching per nest are fixed by the number of eggs laid, and it may be designated as the dependent variable. It is quite clear that if a fixed proportion of the eggs hatched the correlation between the number of eggs and the number of birds would be perfect.

The biologist can suggest many analogous cases; for instance number of flowers per inflorescence and number of fruits developing, number of ovules formed and number of seeds maturing.

[* This formula, given by me to Mr J. F. Tocher, depends on the assumption of normal correlation, see *Biometrika*, Vol. iv. p. 820. The approximate value depends on neglecting higher powers of the coefficients of variation. The formula for the mean of the double product (Tocher, *loc. cit.*) is exact. The formula for the mean of the triple product is not exact, any more than the formula for the s.d. of the triple product (see Tocher, *loc. cit.* p. 821). The formulae for the mean and s.d. of an index are only true to the lowest powers in v_1 and v_2 , and must not be applied if v_1 and v_2 are large. The formulae for $z = x_1 \pm x_2$, the sums or differences of any number of variables, are of course exact for both mean and s.d. Ed.]

The interpretation of the correlations between such characters as these demands a little caution. Suppose we have a series of coefficients of correlation for number of ovules formed and number of seeds developing per pod in various species. It will be quite clear that the species in which the value of r is the highest are those in which the number of ovules per pod has the largest influence upon the number of seeds developing per pod, but we are not at all justified in concluding for any individual series that the pods with the larger number of ovules have a greater, or less, capacity, relative to the number of ovules they produce, for maturing seed.

Suppose that we have a population of individuals in which we recognize the two characters, x and y , and that y is merely a proportion of x which varies from individual to individual. Our problem is to determine whether the relative value of y changes from the lower to the higher grades of x .

Clearly the most probable value of y for any individual is px , where

$$p = \bar{y}/\bar{x},$$

and $\bar{x}$ and $\bar{y}$ are the mean values of the two characters in the whole population. Now we may take $s = y - px$, or the deviation of the y of any individual from its most probable value. If we can discover the correlation between x and s we shall have a convenient measure of any organic influence which x may have upon the relative value of y . It is undesirable to correlate x with the ratio y/x because of the spurious correlation which it introduces.

Using the differential notation to represent variations, we get

$$\delta x \delta s = \delta x (\delta y - p \delta x),$$

and summing and dividing by the number of pairs, we have

$$\sigma_x \sigma_s r_{xs} = \sigma_x \sigma_y r_{xy} - p \sigma_x^2.$$

Hence

$$r_{xs} = \frac{\sigma_y r_{xy} - p \sigma_x}{\sqrt{\sigma_y^2 + p^2 \sigma_x^2 - 2p \sigma_x \sigma_y r_{xy}}} \dots\dots\dots(i).$$

If $r_{xs} = 0$, or y be proportional to x in every case, then

$$r_{xy} = \frac{p \sigma_x}{\sigma_y} = \frac{V_x}{V_y} \dots\dots\dots(ii).$$

Of course (ii) will never be absolutely true in practice. It gives, however, a quick test of the sign of the correlation coefficient r_{xs} . When $V_x/V_y < r_{xy}$ the correlation r_{xs} is positive and when $V_x/V_y > r_{xy}$ it is negative.

(i) may be written

$$r_{xs} = \frac{V_y r_{xy} - V_x}{\sqrt{V_x^2 + V_y^2 - 2V_x V_y r_{xy}}} = \frac{r_{xy} - V_x/V_y}{\sqrt{1 + (V_x/V_y)^2 - 2(V_x/V_y) r_{xy}}} = \frac{r_{xy} - V_x/V_y}{\sqrt{1 - r_{xy}^2 + (r_{xy} - V_x/V_y)^2}} \dots\dots\dots(iii).$$

For practical purposes (iii) is the most convenient formula to use in calculations. V_x and V_y are usually wanted for other purposes and will have been calculated already. The value for $1 - r_{xy}^2$ will have been calculated already in obtaining the probable error of r_{xy} . Beyond this the calculation of r_{xs} can be completed in a few minutes if formula (iii) be adopted.

This formula has already been rather extensively used in series of data which will be published shortly, but I give the following illustrations of the kind of problems to which it may be applied.

Illustration I. Correlation between the number of flowers per inflorescence and the number of developing fruits in *Staphylea trifolia*.

Table I shows the correlation between the number of flowers formed and the number of immature fruits at the time of counting for 270 inflorescences of *Staphylea trifolia* from the North American Tract of the Missouri Botanical Garden. The countings of the number of flowers and fruits were made incidentally while an investigation of the development of the fruit was being carried on. The series is a short one and represents the inflorescences at

TABLE I.

Staphylea trifolia.

Flowers per Inflorescence.

Developing Fruits																																	
	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	26	27	28	29	30	31	Totals	
1	3	—	3	2	2	3	3	2	1	4	1	2	2	1	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	38
2	—	10	4	5	5	5	15	7	1	6	3	4	2	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	70
3	—	—	14	6	3	6	9	7	4	4	—	1	2	—	—	—	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	60
4	—	—	—	10	2	3	4	3	2	5	3	4	2	1	1	1	1	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	41
5	—	—	—	—	4	2	4	2	2	3	3	4	1	1	1	1	1	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	33
6	—	—	—	—	—	4	1	—	2	1	—	1	—	1	1	1	2	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	15
7	—	—	—	—	—	—	—	2	—	—	—	1	1	1	1	1	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	11
8	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	6
9	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
10	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	2
Totals	3	10	21	23	10	23	36	23	13	23	8	19	9	3	5	3	7	2	2	4	3	2	3	1	0	0	1	3	1	1	2	270	

a time when no fruits had attained a length of over 20 mm. It is quite probable that many of the fruits which were still developing when the material was examined would have fallen later. Thus while the investigation of the relationship between the number of flowers and the number of fruits developing per inflorescence during the time that the ovaries which fail to develop are being eliminated presents some points of particular interest, conclusions drawn from one series cannot be applied to any other of the same species. The series is here given merely as an illustration of a statistical method.

Number of flowers furnishes our x character and number of fruits developing our y character. For constants we get

Mean $x = 9.193 \pm .243$,	Mean $y = 3.396 \pm .074$,
S. D. $x = 5.924 \pm .172$,	S. D. $y = 1.801 \pm .052$,
$V_x = 64.444$,	$V_y = 53.038$,
$p = .369$,	$V_x/V_y = 1.215$,
$r_{xy} = .457 \pm .033$.	$r_{yx} = -.649 \pm .024$.

These figures seem to indicate that while there is a positive correlation of .457 between the number of flowers per inflorescence and the number of fruits developing, the inflorescences are not all affected alike in the large elimination of ovaries which takes place. Only a portion of the flowers formed can develop into fruits, and the negative correlation of -.649 seems to show that the larger inflorescences lose not only actually but *relatively* more of their ovaries than do the smaller ones. But this conclusion can be applied only to this particular series of data. It may be found later that after the elimination is complete the correlation between the number of flowers and the deviation of the number of fruits per inflorescence from their probable number will rise to sensibly 0, or take a substantial positive value.

Illustration II. Correlation between the number of ovules per pod and the number of seeds developing per pod in *Robinia Pseudacacia*.

The relationship between the number of ovules and the number of seeds in a sample of 1427 pods of the black locust collected from 12 trees near Lawrence, Kansas, in the fall of 1905, is shown in Table II. The constants which interest us here are :

x = Ovules,	y = Seeds,
Mean $x = 12.1794$,	Mean $y = 7.6874$,
S. D. $x = 2.2763$,	S. D. $y = 3.4938$,
$V_x = 18.690$,	$V_y = 45.448$,
	$V_x/V_y = .4112$,
$r_{xy} = .693 \pm .009$.	$r_{yx} = .365 \pm .015$.

Here it appears from the substantial positive value of r_{xy} that the pods with the larger numbers of ovules are relatively more capable of maturing their seeds than those with fewer. But I may add that this is not true in the case of each individual tree, so that more data are necessary before we are justified in extending this conclusion to the species at large. Evidence from another leguminous plant will be of interest in this connection.

Illustration III. Correlation between the number of ovules per pod and the number of seeds developing per pod in *Cercis Canadensis*.

Table III is extracted from a forthcoming memoir on fertility in *Cercis Canadensis*. The correlations are :

x = Ovules,	y = Seeds,
$r_{xy} = .6855 \pm .0046$.	$r_{yx} = .0070 \pm .0087$.

Here r_{yx} is so close to 0 that there seems little doubt that there is essentially no relationship between the number of ovules per pod and the capacity of the pod for maturing its seeds. This point will be discussed in detail in the memoir.

Biometrika vi

TABLE II.
Robinia Pseudacacia.

Seeds per Pod.

Ovules per Pod.		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Totals
	6	—	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	2
	7	1	1	3	2	4	—	—	—	—	—	—	—	—	—	—	—	—	—	11
	8	1	5	4	8	4	6	2	—	—	—	—	—	—	—	—	—	—	—	30
	9	9	9	28	15	7	19	6	10	—	—	—	—	—	—	—	—	—	—	103
	10	6	33	27	38	38	23	17	12	5	—	—	—	—	—	—	—	—	—	199
	11	5	25	36	39	34	41	36	12	11	6	—	—	—	—	—	—	—	—	245
	12	3	16	35	33	39	32	36	19	21	15	9	—	—	—	—	—	—	—	258
	13	2	8	14	17	31	28	33	23	19	20	10	10	—	—	—	—	—	—	215
	14	1	1	4	4	8	10	15	12	23	14	18	9	16	—	—	—	—	—	135
15	1	1	—	3	1	3	8	6	12	18	14	15	8	9	—	—	—	—	99	
16	—	—	1	—	—	—	2	3	4	9	14	7	8	9	14	—	—	—	71	
17	—	—	—	2	1	—	—	1	1	3	2	—	4	4	5	9	—	—	32	
18	—	—	—	—	—	—	—	—	1	3	2	1	1	1	3	2	6	—	20	
19	—	—	—	—	—	—	—	—	—	—	—	—	2	1	—	—	1	2	7	
Totals	29	100	152	161	168	162	155	98	97	88	69	44	38	23	22	12	7	2	1427	

TABLE III.
Cercis Canadensis.

Seeds per Pod.

Ovules per Pod.									Totals
	1	2	3	4	5	6	7	8	
2	3	31	—	—	—	—	—	—	34
3	34	176	300	—	—	—	—	—	510
4	37	226	885	1093	—	—	—	—	2241
5	17	71	310	784	1040	—	—	—	2222
6	4	14	54	153	288	375	—	—	888
7	—	1	4	9	15	33	39	—	101
8	—	—	—	1	—	1	—	2	4
Totals	95	519	1553	2040	1343	409	39	2	6000

Illustration IV. Correlation between the number of flowers per inflorescence and the number of synanthous flowers in *Spiraea Van Houtii*.

"Excess of nutrition" has been the explanation offered hundreds of times for plant abnormalities. It seems desirable, therefore, to try to correlate the presence of abnormalities with the magnitude of other organs of the plant in order to see whether they are more frequent in cases in which the organs of the individual have reached an abnormal size. An example of this kind of problem is taken from my notes on abnormalities in the inflorescence of the garden species, *Spiraea Van Houtii*.

The particular abnormality here considered is that frequently designated as synanthly in teratological literature. In the inflorescence of this species the pedicels of two or more flowers are frequently found fused for their entire length. The flowers themselves may have a more or less common receptacle, but the essential floral organs are usually so distinct that they would be counted as belonging to two flowers. The details of the abnormalities do not concern us here, and will be discussed in due time. The essential question which we wish to answer on the

basis of data already tabled may be stated: Is an individual flower more (or less) likely to be abnormal if it is borne on a large* inflorescence than if it is produced on a small one?

If the chances of an individual flower being abnormal were in no way influenced by the number of flowers per inflorescence, but the chances of belonging to a synanthly were equal for all flowers, abnormalities would be distributed among the inflorescences in proportion to their number of flowers, and a correlation between the number of flowers and the number of abnormal flowers would result. Therefore this correlation coefficient will not give us the information we desire. What we need to know is the correlation between the number of flowers per inflorescence and the deviation of the number of abnormal flowers from the probable number if abnormalities were distributed among the inflorescences in proportion to the number of flowers they produce.

TABLE IV.

Spiraea Van Houtii.

Number of Flowers per Inflorescence.

Abnormal Flowers.		5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	Totals
	0	1	—	5	5	8	6	9	13	16	22	17	17	12	18	21	12	14	17	9	9	3	—	234
	1	—	—	1	1	2	1	9	5	3	3	5	7	9	10	7	10	6	6	4	—	1	1	91
	2	—	—	—	—	—	—	1	1	2	3	7	3	3	6	10	2	1	—	3	1	1	1	45
	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	8
	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Totals	1	0	6	6	10	7	19	19	21	28	29	29	25	35	41	24	21	24	16	10	6	2	379	

Table IV. shows the number of flowers per inflorescence and the number of flowers counted as abnormal in each inflorescence. Since the abnormality with which we are dealing is the fusion† of two flowers there can be no group of "one flower abnormal"; each abnormality was considered as two or more abnormal flowers. Our series of data is not large but it has the advantage of being taken all from one clump of plants, probably all arising from the same individual, at the Missouri Botanical Garden.

The correlations are:

$$x = \text{All flowers,} \quad y = \text{Abnormal flowers,}$$

$$r_{xy} = .121 \pm .034. \quad r_{yx} = -.071 \pm .034.$$

With probable errors so large as these any trustworthy conclusion is impossible. But the negative sign of the r_{yx} coefficient will be sufficient to warn one against concluding that the larger (and so presumably the more vigorous, or "better nourished") inflorescences are more apt to produce abnormal flowers. They produce *actually* more, but this is only to be expected; the question is whether they produce *relatively* more.

Finally, I would again warn the reader against any generalization from the above illustrations. They are selected to show the kind of problems to which the biometric method described may be profitably applied.

* Here the only measure of the size of the inflorescence is the number of flowers which it produces.

† I wish to express no opinion whatever on the embryological origin of the anomaly, and the term fusion is used simply because it is the conventional and convenient term for such cases as these.