

## Biometrika Trust

---

On the Influence of Double Selection on the Variation and Correlation of Two Characters

Author(s): Karl Pearson

Source: *Biometrika*, Vol. 6, No. 1 (Mar., 1908), pp. 111-112

Published by: [Biometrika Trust](#)

Stable URL: <http://www.jstor.org/stable/2331563>

Accessed: 17/06/2014 04:15

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at  
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



*Biometrika Trust* is collaborating with JSTOR to digitize, preserve and extend access to *Biometrika*.

<http://www.jstor.org>

general population, or there is no variability in the sex-ratio due to individuality. All the variation observed is actually due to the random sampling involved in taking small families.

It appears to me therefore that Mr Cobb's note is of value, as it enables Mr Heron to confirm his results from another standpoint. For practical purposes the main problem is: Are the actual sex-ratios in parent and offspring related? The answer is: Not sensibly. Mr Cobb suggests that this result is only due to the apparent family not representing the gametic reality. This raised the question of whether any individuality at all exists for the sex-ratio. The answer appears to be none, because the standard deviation observed is precisely that which would be found were the sex-ratio for any individual that due to a random selection from the general population of a group equal in number to his family.

## VI. On the Influence of Double Selection on the Variation and Correlation of two Characters.

By KARL PEARSON, F.R.S.

In a memoir published in the *Phil. Trans.* Vol. 200 A, pp. 1—66, I have dealt with the problem of the influence of selecting  $q$  characters on the variability and correlation of  $n-q$  non-selected characters\*. My present problem is somewhat different, and as the solution provided has been in use, and given in lectures for some years past, it may be desirable to publish it.

There are two correlated variables, 1 and 2, with standard deviations  $\sigma_1$  and  $\sigma_2$  and correlation  $r_{12}$ . In entering up a record a selection is made of the variable 1, so that its mean is shifted to  $m_1$  and its standard deviation to  $s_1$ ; let  $\mu_1 = s_1/\sigma_1$ . In entering for the variable 2, a selection is made so that its mean is shifted to  $m_2$  and its standard deviation is changed to  $s_2$ ; let  $\mu_2 = s_2/\sigma_2$ . As illustration suppose cards formed of characters in parent and offspring; we pass through the cards, putting a red line through every card not one of the selected parents; we pass through them again and put a blue line through every card not one of the selected offspring; the result is that cards may be thrown out because they carry a red line or a blue line or because they carry *both*. The problem is what are the standard deviations  $\Sigma_1$  and  $\Sigma_2$  and the correlation  $R_{12}$  of the material left after this double selection. If we assume the material normal the following values are readily found by considering the probability of selecting a definite individual pair  $x_1, x_2$  to be of the form

$$\text{const.} \times e^{-\frac{1}{2} \frac{1}{1-r_{12}^2} \left( \frac{x_1^2}{\sigma_1^2} - \frac{2rx_1x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2} \right)} \times \frac{e^{-\frac{(x_1-m_1)^2}{2s_1^2}}}{e^{-\frac{x_1^2}{2\sigma_1^2}}} \times \frac{e^{-\frac{(x_2-m_2)^2}{2s_2^2}}}{e^{-\frac{x_2^2}{2\sigma_2^2}}}.$$

Thus the correlation surface is:

$$z = \text{const.} \times e^{-\frac{1}{2} \frac{1}{1-R_{12}^2} \left\{ \frac{(x_1-M_1)^2}{\Sigma_1^2} - \frac{2R_{12}(x_1-M_1)(x_2-M_2)}{\Sigma_1\Sigma_2} + \frac{(x_2-M_2)^2}{\Sigma_2^2} \right\}}$$

\* It may be of interest to note that the whole of the formulae developed in that memoir are true far beyond the range of the Gaussian distributions for which they were proved. They are universally true provided we start with a generalised notion of correlation as involving the maximum dependence of one variable on an arbitrary linear function of  $(n-1)$  other variables.

Hence we deduce :

$$\Sigma_1^2 = \sigma_1^2 \frac{\mu_1^2 \{1 - r_{12}^2 (1 - \mu_2^2)\}}{1 - r_{12}^2 (1 - \mu_1^2) (1 - \mu_2^2)} \dots\dots\dots (i)$$

$$\Sigma_2^2 = \sigma_2^2 \frac{\mu_2^2 \{1 - r_{12}^2 (1 - \mu_1^2)\}}{1 - r_{12}^2 (1 - \mu_1^2) (1 - \mu_2^2)} \dots\dots\dots (ii),$$

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

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

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

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

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

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

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

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.