OXFORD JOURNALS
OXFORD UNIVERSITY PRESS

Biometrika Trust

The Correlation Coefficient of a Polychoric Table

Author(s): A. Ritchie-Scott

Source: *Biometrika*, Vol. 12, No. 1/2 (Nov., 1918), pp. 93-133

Published by: Oxford University Press on behalf of Biometrika Trust

Stable URL: <http://www.jstor.org/stable/2331931>

Accessed: 20-06-2016 13:27 UTC

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

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, Oxford University Press are collaborating with JSTOR to digitize, preserve and extend access to *Biometrika*

THE CORRELATION COEFFICIENT OF A POLYCHORIC TABLE.

By A. RITCHIE-SCOTT, B.Sc.

§ 1. INTRODUCTION.

We have at our disposal a considerable number of methods for finding the coefficient of correlation between two characters from a table of frequencies. These methods may be summarily named and classified as follows:

1. Product Moment.
2. Tetrachoric r .
3. Marginal centroids.
4. Biserial r .
5. Three Row η .
6. Variate difference methods including the correlation of grades and ranks.
7. Equiprobable tetrachoric r .
8. Mean contingency.
9. Mean square contingency.

Each of these methods has its own specially appropriate field of usefulness, but there still remains one class of table for which no entirely satisfactory methods have been devised, namely those which contain more than 2×2 cells and fewer than 4×4 , to which the tetrachoric and mean square contingency methods respectively may be applied.

It was with a view to investigating satisfactory methods for such tables that the following work was undertaken. Such tables arise under many circumstances, particularly when we can, as in many psychological investigations which depend upon the instinctive judgment of some character, definitely assign individuals with pronounced characters to either end of a scale, but are compelled to relegate doubtful cases to an intermediate but somewhat indefinite category. We have, in a word, good, indifferent, bad; present, doubtful, absent—classifications resulting in a frequency table with three categories for one or both characters.

In the present memoir a normal distribution has been assumed as it has been found to be not infrequently applicable and its assumption has given fairly satisfactory results even with distributions which are not strictly normal.

§ 2. NOTATION.

Let the normal surface (when standard deviations are used as units)

$$Nz(x, y) = \frac{N}{2\pi\sqrt{1-r^2}} e^{-\frac{x^2+y^2-2rxy}{2(1-r^2)}}$$

be divided as in the diagram by planes drawn parallel to the yz plane at the points $x = h_1, x = h_2 \dots$ and by planes parallel to the xz plane at the points $y = k_1, y = k_2 \dots$. Let the planes intersect in lines whose projections are the points $(h_1, k_1), (h_2, k_2), \text{etc.}$, contracted to 11, 12, etc., where the first figure is the suffix of h and the second figure is the suffix of k .

	h_1	h_2	h_3	h_4		h_{p-1}	h_p	
	n_{11}	n_{21}	n_{31}				n_{p1}	$n_{.1}$
k_1	11	21	31	41				$\left. \begin{matrix} n_{p1} \\ n_{.1} \end{matrix} \right\} m_{.1}$
	n_{12}	n_{22}	n_{32}				n_{p2}	$n_{.2}$
k_2	12	22	32	42				$\left. \begin{matrix} n_{p2} \\ n_{.2} \end{matrix} \right\} m_{.2}$
	n_{13}	n_{23}	n_{33}					$n_{.3}$
k_3	13	23	33	43				$\left. \begin{matrix} n_{p3} \\ n_{.3} \end{matrix} \right\} m_{.3}$
	14	24	...	...				
k_{q-1}	n_{1q}	n_{2q}	...				n_{pq}	$n_{.}$
k_q	$n_{1.}$	$n_{2.}$	$n_{3.}$	...			$n_{p.}$	$N.$
	$m_{1.}$							
	$m_{2.}$							
	$m_{3.}$							

The frequencies in the cells and the marginal totals are indicated by $n_{11}, n_{12}, \text{etc.}$, and $n_{1.}, n_{2.} \dots n_{.1}, n_{.2}, \text{etc.}$, as shown in the diagram.

The surface may also be regarded as divided at each point 11, 12, ..., into four quadrants. One of these quadrants is shown by the dotted lines in the diagram. The quadrant in the position shown, viz. the left upper or $(- -)$ quadrant will be

regarded in what follows as the leading quadrant and its frequency denoted by m . Thus the quadrant shown will have the frequency m_{43} .

In the ordinary scheme for a tetrachoric table, the quadrants are denoted by a, b, c, d , and when necessary these letters will be used with the appropriate suffix. Thus the division at the point $s \cdot t$ would be represented as in the diagram,

		h_s	
	(m_{st})		
	a_{st}	b_{st}	$m_{\cdot t}$
k_t	st		
	c_{st}	d_{st}	$m_{\cdot t'}$
	$m_{s\cdot}$	$m_{s'\cdot}$	N

The marginal totals corresponding to the leading quadrant are denoted by $m_{s\cdot}, m_{\cdot t}$ and the complementary totals by $m_{s'\cdot}, m_{\cdot t'}$.

Clearly any cell frequency may be expressed in terms of quadrant frequencies since

$$n_{st} = m_{st} - m_{s-1, t} - m_{s, t-1} + m_{s-1, t-1}.$$

§ 3. ENNEACHORIC METHOD.

In order to determine r , since we assume the distribution to be normal and we know the marginal totals, only one more datum is required. This for example may be a frequency block (or the total frequency on a continuous system of cells). As special cases we have the "briquette" or frequency on a rectangle of cells or again the quadrant frequency. The block may be the frequency contents of a group of corner, marginal or internal cells*. Consider (for future use) the general case of a quadrant frequency.

Let
$$\frac{m_{s\cdot}}{N} = {}_s\tau_0, \quad \frac{m_{\cdot t}}{N} = {}_t\theta_0,$$

where ${}_s\tau$ and ${}_t\theta$ are the tetrachoric coefficients. Then

$$\begin{aligned} \frac{m_{st}}{N} &= {}_s\tau_0 {}_t\theta_0 + {}_s\tau_1 {}_t\theta_1 r + {}_s\tau_2 {}_t\theta_2 r^2 + \dots \\ &= {}_s\tau_0 {}_t\theta_0 + \mathfrak{T}({}_s\tau, {}_t\theta, r) \dots\dots\dots(1). \end{aligned}$$

In using Everitt's Tables of the tetrachoric functions in which τ_0 and θ_0 must be less than $\frac{1}{2}$ we must either rearrange the table or adjust the above formula for the position of the mean with regard to the quadrants. It is more convenient to adjust the formula as follows, dropping the suffixes as we are dealing with any point of division.

* A "cell" is the least element for which the frequency is provided in the original data. Cells grouped together for any working purposes are collectively termed a "block."

Let $1 - \tau_0 = \tau_0', \quad 1 - \theta_0 = \theta_0'.$

Mean in a ,

$$\begin{aligned} \frac{m}{N} &= \tau_0 + \theta_0 - 1 + \frac{d}{N} \\ &= \tau_0 + \theta_0 - 1 + \tau_0' \theta_0' + \tau_1' \theta_1' r + \tau_2' \theta_2' r^2 + \dots \\ &= \tau_0 + \theta_0 - 1 - (1 - \tau_0)(1 - \theta_0) + \mathfrak{T}(\tau', \theta', r) \\ &= \tau_0 \theta_0 + \mathfrak{T}(\tau', \theta', r) \dots \dots \dots (2). \end{aligned}$$

Mean in b ,

$$\begin{aligned} \frac{m}{N} &= \tau_0 - \frac{c}{N} \\ &= \tau_0 - (\tau_0 \theta_0' + \tau_1 \theta_1' (-r) - \tau_2 \theta_2' (-r)^2 + \dots) \\ &= \tau_0 - \tau_0 (1 - \theta_0) - \mathfrak{T}(\tau, \theta', -r) \\ &= \tau_0 \theta_0 - \mathfrak{T}(\tau, \theta', -r) \dots \dots \dots (3). \end{aligned}$$

Mean in c ,

$$\begin{aligned} \frac{m}{N} &= \theta_0 - \frac{b}{N} \\ &= \theta_0 - (\tau_0' \theta_0 + \tau_1' \theta_1 (-r) + \tau_2' \theta_2' (-r)^2 + \dots) \\ &= \theta_0 - (1 - \tau_0) \theta_0 - \mathfrak{T}(\tau', \theta, -r) \\ &= \tau_0 \theta_0 - \mathfrak{T}(\tau', \theta, -r) \dots \dots \dots (4). \end{aligned}$$

$$\text{Mean in } d, \quad \frac{m}{N} = \tau_0 \theta_0 + \mathfrak{T}(\tau, \theta, r) \dots \dots \dots (5).$$

In place of taking a quadrant we may take a marginal or internal block. I shall only consider the latter as the case of a marginal block may be deduced from that of an internal block by removing one of the bounding planes to an infinite distance.

In discussing the central block we in effect reduce any table to a 3×3 (ennea-choric) table as we consider it to be constituted of a central block (or group of cells) and 4 marginal and 4 corner blocks. I may therefore use the nomenclature for a 3×3 table without any loss of generality.

$$\begin{aligned} n_{22} &= m_{22} - m_{12} - m_{21} + m_{11}, \\ \therefore \frac{n_{22}}{N} &= {}_2\tau_0 {}_2\theta_0 + \mathfrak{T}({}_2\tau, {}_2\theta, r) - {}_1\tau_0 {}_2\theta_0 - \mathfrak{T}({}_1\tau, {}_2\theta, r) \\ &\quad - {}_2\tau_0 {}_1\theta_0 - \mathfrak{T}({}_2\tau, {}_1\theta, r) + {}_1\tau_0 {}_1\theta_0 + \mathfrak{T}({}_1\tau, {}_1\theta, r) \\ &= ({}_2\tau_0 - {}_1\tau_0)({}_2\theta_0 - {}_1\theta_0) + ({}_2\tau_1 - {}_1\tau_1)({}_2\theta_1 - {}_1\theta_1)r \\ &\quad + ({}_2\tau_2 - {}_1\tau_2)({}_2\theta_2 - {}_1\theta_2)r^2 + \text{etc.} \\ &= \frac{n_{2 \cdot} n_{\cdot 2}}{N} + ({}_2\tau_1 - {}_1\tau_1)({}_2\theta_1 - {}_1\theta_1)r + ({}_2\tau_2 - {}_1\tau_2)({}_2\theta_2 - {}_1\theta_2)r^2 + \text{etc.} \\ &\quad \dots \dots \dots (6). \end{aligned}$$

As an example of the rearrangement of the formula for computation consider the case when the mean is in n_{22} :

$$(\text{mean in } a) \quad \frac{m_{22}}{N} = {}_2\tau_0 {}_2\theta_0 + {}_2\tau_1' {}_2\theta_1' r + {}_2\tau_2' {}_2\theta_2' r^2 + {}_2\tau_3' {}_2\theta_3' r^3 + \dots,$$

$$(\text{mean in } b) \quad \frac{m_{12}}{N} = {}_1\tau_0 {}_2\theta_0 + {}_1\tau_1 {}_2\theta_1' r - {}_1\tau_2 {}_2\theta_2' r^2 + {}_1\tau_3 {}_2\theta_3' r^3 - \dots,$$

$$(\text{mean in } c) \quad \frac{m_{21}}{N} = {}_2\tau_0 {}_1\theta_0 + {}_2\tau_1' {}_1\theta_1 r - {}_2\tau_2' {}_1\theta_2 r^2 + {}_2\tau_3' {}_1\theta_3 r^3 - \dots,$$

$$(\text{mean in } d) \quad \frac{m_{11}}{N} = {}_1\tau_0 {}_1\theta_0 + {}_1\tau_1 {}_1\theta_1 r + {}_1\tau_2 {}_1\theta_2 r^2 + {}_1\tau_3 {}_1\theta_3 r^3 + \dots$$

$$\begin{aligned} \therefore \frac{n_{22}}{N} &= \frac{1}{N} (m_{22} - m_{12} - m_{21} + m_{11}) \\ &= \frac{n_2 \cdot n_{\cdot 2}}{N} + ({}_2\tau_1' - {}_1\tau_1) ({}_2\theta_1' - {}_1\theta_1) r + ({}_2\tau_2' + {}_1\tau_2) ({}_2\theta_2' + {}_1\theta_2) r^2 \\ &\quad + ({}_2\tau_3' - {}_1\tau_3) ({}_2\theta_3' - {}_1\theta_3) r^3 + \text{etc.} \dots\dots\dots(7). \end{aligned}$$

It will be noted that when one set of categories is symmetrical about the mean, i.e. when say ${}_2\tau_1' = {}_1\tau_1$ all the terms of odd degree in r vanish. This corresponds to the fact that symmetrical categories may be reversed without altering the numerical value of the marginal totals and their relation to the central frequency; but such reversal will change the sign of r .

§ 4. STANDARD DEVIATION OF r BY ENNEACHORIC METHOD.

We have now to determine the probable error of r found in this manner.

Throughout what follows differentials will be used to indicate random sample variations, i.e. it is always supposed that the variations are small as compared to any quantity varied so that all the dn 's are small, or all the n 's are large quantities.

$$m_{st} = N \int_{-\infty}^{h_s} \int_{-\infty}^{k_t} z(x, y, r) dx \cdot dy = f(h_s, k_t, r) \dots\dots\dots(8).$$

Since the variations of the means and the standard deviations are, in this form of m_{st} , involved in the variations of h_s and k_t , we have

$$dm_{st} = \frac{\partial f}{\partial h_s} dh_s + \frac{\partial f}{\partial k_t} dk_t + \frac{\partial f}{\partial r} dr \dots\dots\dots(9).$$

Evaluating the differential coefficients,

$$\frac{\partial f}{\partial h_s} = N \int_{-\infty}^{k_t} z(h_s, y, r) dy = N \frac{e^{-h_s^2/2}}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{k_t - rh_s} \sqrt{1-r^2} e^{-y^2/2} dy \dots\dots\dots(10).$$

This is the area of that portion of the dichotomic plane $x = h_s$ which bounds the quadrant m_{st} . But the area of the whole dichotomic plane is

$$N \int_{-\infty}^{+\infty} z(h_s, y, r) dy = \frac{N}{\sqrt{2\pi}} e^{-h_s^2/2} = NH_s \dots\dots\dots(11),$$

$$\text{so that if we write} \quad \frac{\partial f}{\partial h_s} = NH_s A_{st} \dots\dots\dots(12),$$

$$\text{where} \quad A_{st} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{k_t - rh_s} \sqrt{1-r^2} e^{-y^2/2} dy \dots\dots\dots(13),$$

the factor A_{st} will be that fraction of the whole dichotomic plane section, which bounds the quadrant m_{st} and will have no dimensions. The value of A_{st} may be taken directly from Sheppard's tables of the probability integral entering with

the argument $\frac{k_t - rh_s}{\sqrt{1 - r^2}}$. It will be convenient to refer to this tabled integral as $\mathfrak{E}$, so that

$$A_{st} = \mathfrak{E} \left(\frac{k_t - rh_s}{\sqrt{1 - r^2}} \right) \dots\dots\dots(14)$$

It is convenient to note that with this notation

$$\mathfrak{E}(-x) = 1 - \mathfrak{E}(x).$$

Further since

$$m_{s.} = \frac{N}{\sqrt{2\pi}} \int_{-\infty}^{h_s} e^{-\frac{1}{2}x^2} dx,$$

$$\frac{dm_{s.}}{dh_s} = \frac{N}{\sqrt{2\pi}} e^{-\frac{1}{2}h_s^2} = NH_{s.},$$

and

$$dh_s = \frac{dm_{s.}}{NH_{s.}}.$$

Hence

$$\frac{\partial f}{\partial h_s} dh_s = NA_{st}H_{s.} \frac{dm_{s.}}{NH_{s.}} = A_{st}dm_{s.} \dots\dots\dots(15).$$

Similarly

$$\frac{\partial f}{\partial k_t} dk_t = B_{st}dm_{.t} \dots\dots\dots(16),$$

where

$$B_{st} = \mathfrak{E} \left(\frac{h_s - rk_t}{\sqrt{1 - r^2}} \right) \dots\dots\dots(17).$$

Lastly

$$\begin{aligned} \frac{\partial f}{\partial r} &= \frac{\partial N}{\partial r} \int_{-\infty}^{h_s} \int_{-\infty}^{k_t} z(x, y, r) dx dy \\ &= N \int_{-\infty}^{h_s} \int_{-\infty}^{k_t} \frac{d}{dr} z(x, y, r) dx dy \\ &= N \int_{-\infty}^{h_s} \int_{-\infty}^{k_t} \frac{d^2}{dx dy} z(x, y, r) dx dy \\ &= Nz(h_s, k_t, r) \\ &= \chi_{st} \dots\dots\dots(18), \end{aligned}$$

which is the length of the ordinate at the point (s, t) . We may now write

$$dm_{st} = A_{st}dm_{s.} + B_{st}dm_{.t} + \chi_{st}dr \dots\dots\dots(19),$$

and

$$- \chi_{st}dr = A_{st}dm_{s.} + B_{st}dm_{.t} - dm_{st} \dots\dots\dots(20).$$

Considerable use will be made of this formula later and the following abbreviated notation will be used:

$$A_{st}dm_{s.} + B_{st}dm_{.t} - dm_{st} = \delta P_{st} = - \chi_{st}dr \dots\dots\dots(21),$$

and

$$A_{st}m_{s.} + B_{st}m_{.t} - m_{st} = P_{st} \dots\dots\dots(22).$$

The reader must be careful to note that δP_{st} is, not dP_{st} but only a part of it, and this symbol is used here as at once a conventional abbreviation and a *memoria technica*.

Since

$$\begin{aligned} n_{22} &= m_{22} - m_{12} - m_{21} + m_{11}, \\ \therefore dn_{22} &= dm_{22} - dm_{12} - dm_{21} + dm_{11} \\ &= A_{22}dm_{2.} + B_{22}dm_{.2} + \chi_{22}dr - A_{12}dm_{1.} - B_{12}dm_{.2} - \chi_{12}dr \\ &\quad - A_{21}dm_{2.} - B_{21}dm_{.1} - \chi_{21}dr + A_{11}dm_{1.} + B_{11}dm_{.1} + \chi_{11}dr \dots\dots(23), \end{aligned}$$

$$\therefore -(\chi_{22} - \chi_{12} - \chi_{21} + \chi_{11}) dr = (A_{22} - A_{21}) dm_{2\cdot} - (A_{12} - A_{11}) dm_{1\cdot} \\ + (B_{22} - B_{12}) dm_{\cdot 2} - (B_{21} - B_{11}) dm_{\cdot 1} - dn_{22} \dots \dots \dots (24).$$

Reference to the diagram shows that $A_{22} - A_{21}$ and the other coefficients are the proportions of the areas of the trapezettes bounding the briquette of volume n_{22} . These areas may be systematically named for the whole table thus,

$$\begin{array}{c} | \\ \alpha_{st} \\ \text{---} \beta_{st} \text{---} | \\ st \end{array}$$

that is, the areas of the planes meeting in the line of which the point s, t is the projection, in the direction shown, are named from the point so that

$$A_{s, t} - A_{s, t-1} = \alpha_{st}, \\ B_{s, t} - B_{s-1, t} = \beta_{st}.$$

Hence we may write

$$A_{22} - A_{21} = \alpha_{22}, \\ A_{12} - A_{11} = \alpha_{12}, \\ B_{22} - B_{12} = \beta_{22}, \\ B_{21} - B_{11} = \beta_{21}.$$

If now we notice that since $m_{2\cdot} = N - n_{3\cdot}$ etc., and $m_{1\cdot} = n_{1\cdot}$,

$$\therefore dm_{2\cdot} = -dn_{3\cdot}, \\ dm_{\cdot 2} = -dn_{\cdot 3}, \\ dm_{1\cdot} = dn_{1\cdot}, \\ dm_{\cdot 1} = dn_{\cdot 1},$$

we then have

$$-(\chi_{11} - \chi_{12} - \chi_{21} + \chi_{22}) dr = -(\alpha_{22} dn_{3\cdot} + \alpha_{12} dn_{1\cdot} + \beta_{22} dn_{\cdot 3} + \beta_{21} dn_{\cdot 1} + dn_{22}) \dots \dots \dots (25).$$

Expanding this in terms of frequency volumes this becomes

$$(\chi_{11} - \chi_{12} - \chi_{21} + \chi_{22}) dr = (\alpha_{12} + \beta_{21}) dn_{11} + \beta_{21} dn_{21} + (\alpha_{22} + \beta_{21}) dn_{31} \\ + \alpha_{12} dn_{12} + dn_{22} + \alpha_{22} dn_{32} + (\alpha_{12} + \beta_{22}) dn_{13} + \beta_{22} dn_{23} + (\alpha_{22} + \beta_{22}) dn_{33} \dots \dots \dots (26).$$

It has already been shown by Pearson (*Biometrika*, vol. ix, p. 1) that when random samples are taken from a population so large that its composition is not appreciably affected by removing the samples we have the following relations:

$$\sigma_{n_{ss'}} = n_{ss'} \left(1 - \frac{n_{ss'}}{N}\right) \dots \dots \dots (27),$$

$$\text{Mean } (dn_{ss'} dn_{tt'}) = -\frac{n_{ss'} n_{tt'}}{N} \dots \dots \dots (28a),$$

$$\text{Mean } (dn_s \cdot dn_{t\cdot}) = -\frac{n_s \cdot n_{t\cdot}}{N} \dots \dots \dots (28b),$$

$$\text{Mean } (dn_s \cdot dn_{\cdot s'}) = n_{ss'} - \frac{n_s \cdot n_{\cdot s'}}{N} \dots \dots \dots (28c),$$

$$\text{Mean } (dn_{ss'} dn_{\cdot t'}) = -\frac{n_{ss'} n_{\cdot t'}}{N} \dots \dots \dots (28d),$$

$$\text{Mean } (dn_{ss'} dn_{t.}) = - \frac{n_{ss'} n_{t.}}{N} \dots\dots\dots (28e),$$

$$\text{Mean } (dn_{ss'} dn_{s.}) = n_{ss'} \left(1 - \frac{n_{s.}}{N} \right) \dots\dots\dots (28f).$$

Hence squaring both sides, summing for all possible samples and dividing by the number of samples we have,

$$\begin{aligned} (\chi_{11} - \chi_{12} - \chi_{21} + \chi_{22})^2 \sigma_r^2 = & (\alpha_{12} + \beta_{21})^2 n_{11} + \beta_{21}^2 n_{21} + (\alpha_{22} + \beta_{21})^2 n_{31} \\ & + \alpha_{12}^2 n_{12} + n_{22} + \alpha_{22}^2 n_{32} + (\alpha_{12} + \beta_{22})^2 n_{13} + \beta_{22}^2 n_{23} + (\alpha_{22} + \beta_{22})^2 n_{33} \\ & - \frac{1}{N} \left\{ \begin{aligned} & (\alpha_{12} + \beta_{21}) n_{11} + \beta_{21} n_{21} + (\alpha_{22} + \beta_{21}) n_{31} \\ & + \alpha_{12} n_{12} + n_{22} + \alpha_{22} n_{32} \\ & + (\alpha_{12} + \beta_{22}) n_{13} + \beta_{21} n_{23} + (\alpha_{22} + \beta_{22}) n_{22} \end{aligned} \right\}^2 \dots\dots\dots (29). \end{aligned}$$

The expression within the large brackets

$$= \alpha_{12} n_{1.} + \alpha_{22} n_{3.} + \beta_{21} n_{.1} + \beta_{22} n_{.3} + n_{22}.$$

Calling this Nm we have

$$\begin{aligned} (\chi_{11} - \chi_{12} - \chi_{21} + \chi_{22})^2 \sigma_r^2 = & (\alpha_{12} + \beta_{21} - m)^2 n_{11} + (\beta_{21} - m)^2 n_{21} \\ & + (\alpha_{22} + \beta_{21} - m)^2 n_{31} + (\alpha_{12} - m)^2 n_{12} + (1 - m)^2 n_{22} + (\alpha_{22} - m)^2 n_{32} \\ & + (\alpha_{12} + \beta_{22} - m)^2 n_{13} + (\beta_{22} - m)^2 n_{23} + (\alpha_{22} + \beta_{22} - m)^2 n_{33} \dots\dots (30). \end{aligned}$$

The following form for m is instructive although giving an apparently less symmetrical form than the above,

$$\begin{aligned} m = & \frac{1}{N} (\alpha_{12} n_{1.} + \alpha_{22} n_{3.} + \beta_{21} n_{.1} + \beta_{22} n_{.3} + n_{22}) \\ = & \frac{1}{N} \left\{ (A_{12} - A_{11}) m_{1.} + (B_{22} - B_{12}) (N - m_{.2}) + (B_{21} - B_{11}) m_{.1} \right. \\ & \left. + (B_{22} - B_{12}) (N - m_{.2}) + m_{22} - m_{21} - m_{12} + m_{11} \right\} \\ = & \frac{1}{N} \left\{ A_{12} m_{1.} + B_{12} m_{.2} - m_{12} - A_{11} m_{1.} - B_{11} m_{.1} + m_{11} - A_{22} m_{2.} \right. \\ & \left. - B_{22} m_{.2} + m_{22} + A_{21} m_{2.} + B_{21} m_{.1} - m_{21} + N (A_{22} - A_{21} + B_{22} - B_{12}) \right\} \\ = & \alpha_{22} + \beta_{32} - \frac{P_{11} - P_{12} - P_{21} + P_{22}}{N} \\ = & \alpha_{22} + \beta_{22} - \frac{\mathfrak{P}\mathfrak{P}}{N} \dots\dots\dots (31). \end{aligned}$$

We may then write

$$\begin{aligned} (\chi_{11} - \chi_{12} - \chi_{21} + \chi_{22})^2 \sigma_r^2 = & \left(\alpha_{12} - \alpha_{22} + \beta_{21} - \beta_{22} + \frac{\mathfrak{P}\mathfrak{P}}{N} \right)^2 n_{11} + \left(0 - \alpha_{22} + \beta_{21} - \beta_{22} + \frac{\mathfrak{P}\mathfrak{P}}{N} \right)^2 n_{21} \\ & + \left(\beta_{21} - \beta_{22} + \frac{\mathfrak{P}\mathfrak{P}}{N} \right)^2 n_{31} + \left(\alpha_{12} - \alpha_{22} + 0 - \beta_{22} + \frac{\mathfrak{P}\mathfrak{P}}{N} \right)^2 n_{12} \\ & + \left(\frac{1}{2} - \alpha_{22} + \frac{1}{2} - \beta_{22} + \frac{\mathfrak{P}\mathfrak{P}}{N} \right)^2 n_{22} + \left(0 - \beta_{22} + \frac{\mathfrak{P}\mathfrak{P}}{N} \right)^2 n_{32} \\ & + \left(\alpha_{12} - \alpha_{22} + \frac{\mathfrak{P}\mathfrak{P}}{N} \right)^2 n_{13} + \left(0 - \alpha_{22} + \frac{\mathfrak{P}\mathfrak{P}}{N} \right)^2 n_{23} + \left(\frac{\mathfrak{P}\mathfrak{P}}{N} \right)^2 n_{33} \dots\dots (32). \end{aligned}$$

It will be shown in a later paper that the coefficients of the cell frequencies are functions of moments of the frequencies about the mean.

If any relation between h_1, h_2, k_1, k_2 makes

$$\chi_{11} - \chi_{12} - \chi_{21} + \chi_{22} = 0,$$

while the right side of the σ_r^2 equation remains finite the value of the expression for σ_r^2 will become infinite. $\chi_{11} - \chi_{12} - \chi_{21} + \chi_{22}$ obviously vanishes (1) when $h_1 = h_2$ or $k_1 = k_2$, i.e. when either of the central categories becomes vanishingly small, and (2) when $h_1 = -h_2, k_1 = -k_2$ and $r = 0$, i.e. when both sets of the extreme categories have equal frequencies and the correlation is very small.

(1) When $h_1 = h_2$. Then $\chi_{11} = \chi_{12}, \chi_{21} = \chi_{22}, \alpha_{12} = \alpha_{22} = \alpha$ say, $\beta_{21} = \beta_{22} = 0, n_{21} = n_{22} = n_{23} = 0, n_{2\cdot} = 0$.

$$\text{Then} \quad m = \frac{\alpha (n_{1\cdot} + n_{3\cdot})}{N} = \alpha,$$

and the right side of the σ_r^2 equation reduces to zero giving the indeterminate form $\frac{0}{0}$ for σ_r^2 .

(2) When $h_1 = -h_2, k_1 = -k_2$ and $r = 0$. This case will be discussed in the next section.

§ 5. STANDARD DEVIATION OF ENNEACHORIC r IN SPECIAL CASES.

Two particular cases are of interest, (1) when $r = 0$, (2) when the table degenerates into a 2×2 table.

$$\begin{aligned} (1) \quad \text{When } r = 0 \quad A_{st} &= \mathfrak{E}(k_t) = \frac{m_{\cdot t}}{N}, \\ B_{st} &= \mathfrak{E}(h_s) = \frac{m_{s \cdot}}{N}, \\ P_{st} &= A_{st}m_{s \cdot} + B_{st}m_{\cdot t} - m_{st} \\ &= \frac{m_{\cdot t}m_{s \cdot}}{N} + \frac{m_{s \cdot}m_{\cdot t}}{N} - m_{st} \\ &= m_{st}. \end{aligned}$$

Hence

$$\begin{aligned} \alpha_{22} &= A_{22} - A_{21} = \frac{m_{\cdot 2} - m_{\cdot 1}}{N} = \frac{n_{\cdot 2}}{N}, \\ \alpha_{12} &= A_{12} - A_{11} = \frac{m_{\cdot 2} - m_{\cdot 1}}{N} = \frac{n_{\cdot 2}}{N}. \end{aligned}$$

Similarly

$$\begin{aligned} \beta_{22} &= \beta_{21} = \frac{n_{2 \cdot}}{N}, \\ m &= \alpha_{22} + \beta_{22} - \frac{P_{11} - P_{12} - P_{21} + P_{22}}{N} \\ &= \frac{n_{\cdot 2}}{N} + \frac{n_{2 \cdot}}{N} - \frac{1}{N} (m_{11} - m_{12} - m_{21} + m_{22}) \\ &= \frac{n_{\cdot 2}}{N} + \frac{n_{2 \cdot}}{N} - \frac{n_{22}}{N}. \end{aligned}$$

$$\begin{aligned}\chi_{11} - \chi_{12} - \chi_{21} + \chi_{22} &= N (H_1 K_1 - H_1 K_2 - H_2 K_1 - H_2 K_2) \\ &= N (H_1 - H_2) (K_1 - K_2),\end{aligned}$$

and the right-hand side of the equation reduces to

$$\begin{aligned}&\left(\frac{n_{22}}{N}\right)^2 (n_{11} + n_{13} + n_{31} + n_{33}) + \left(\frac{n_{22} - n_{2\cdot}}{N}\right)^2 (n_{21} + n_{23}) \\ &+ \left(\frac{n_{22} - n_{2\cdot}}{N}\right)^2 (n_{12} + n_{32}) + \left(1 - \frac{n_{2\cdot}}{N} - \frac{n_{2\cdot}}{N} + \frac{n_{22}}{N}\right)^2 n_{22}.\end{aligned}$$

Remembering that $\frac{n_{2\cdot} n_{\cdot 2}}{N} = n_{22}$ this resolves into

$$\begin{aligned}&\left(\frac{n_{2\cdot} n_{\cdot 2}}{N}\right)^2 \frac{(N - n_{2\cdot})(N - n_{\cdot 2})}{N} + \left(\frac{N - n_{2\cdot}}{N}\right)^2 \left(\frac{n_{2\cdot}}{N}\right)^2 \frac{n_{2\cdot}}{N} (N - n_{\cdot 2}) \\ &+ \left(\frac{n_{2\cdot}}{N}\right)^2 \left(\frac{N - n_{\cdot 2}}{N}\right)^2 \frac{n_{2\cdot}}{N} (N - n_{2\cdot}) + \left(\frac{N - n_{2\cdot}}{N}\right)^2 \left(\frac{N - n_{\cdot 2}}{N}\right)^2 \frac{n_{2\cdot} n_{\cdot 2}}{N} \\ &= \frac{(N - n_{2\cdot})(N - n_{\cdot 2}) n_{2\cdot} n_{\cdot 2}}{N^5} \{n_{2\cdot} n_{\cdot 2} + (N - n_{2\cdot}) n_{\cdot 2} + (N - n_{\cdot 2}) n_{2\cdot} \\ &\quad + (N - n_{2\cdot})(N - n_{\cdot 2})\} \\ &= \frac{n_{2\cdot} (N - n_{2\cdot}) n_{\cdot 2} (N - n_{\cdot 2})}{N^3} \dots\dots\dots(33).\end{aligned}$$

Hence

$$N^2 (H_1 - H_2)^2 (K_1 - K_2)^2 \sigma_r^2 = n_{2\cdot} \frac{(N - n_{2\cdot})}{N} \cdot \frac{n_{\cdot 2} (N - n_{\cdot 2})}{N} \cdot \frac{1}{N} = \frac{\sigma_{n_{2\cdot}}^2 \sigma_{n_{\cdot 2}}^2}{N}$$

and

$$\sigma_r = \frac{1}{\sqrt{N}} \cdot N \frac{\sigma_{n_{2\cdot}} \sigma_{n_{\cdot 2}}}{(H_1 - H_2) (K_1 - K_2)} \dots\dots\dots(34).$$

This may also be expressed as follows:

$$\begin{aligned}\sigma_{n_{2\cdot}}^2 \sigma_{n_{\cdot 2}}^2 &= \frac{n_{2\cdot} (N - n_{2\cdot})}{N} \frac{n_{\cdot 2} (N - n_{\cdot 2})}{N} \\ &= \frac{n_{2\cdot} n_{\cdot 2}}{N} \cdot \frac{n_{1\cdot} + n_{3\cdot}}{N} \cdot \frac{n_{\cdot 1} + n_{\cdot 3}}{N} \\ &= n_{22} (n_{11} + n_{13} + n_{31} + n_{33}) \dots\dots\dots(35)\end{aligned}$$

and

$$\sigma_r = \frac{\sqrt{\frac{1}{N} n_{22} (n_{11} + n_{13} + n_{31} + n_{33})}}{N (H_1 - H_2) (K_1 - K_2)} \dots\dots\dots(36).$$

When the extreme categories are nearly equal and r is zero $H_1 = H_2$ and $K_1 = K_2$ and the value for σ_r^2 becomes infinite. It is necessary then to keep second powers of differences to determine σ_r^2 . Keeping second order terms we have:

$$\begin{aligned}dm &= \frac{\partial f}{\partial h} dh + \frac{\partial f}{\partial k} dk + \frac{\partial f}{\partial r} dr + \frac{1}{2} \left\{ \frac{\partial^2 f}{\partial h^2} (dh)^2 + \frac{\partial^2 f}{\partial k^2} (dk)^2 + \frac{\partial^2 f}{\partial r^2} (dr)^2 \right. \\ &\quad \left. + 2 \frac{\partial^2 f}{\partial h \partial k} (dh dk) + 2 \frac{\partial^2 f}{\partial h \partial r} (dh dr) + \frac{\partial^2 f}{\partial k \partial r} (dk dr) \right\} \dots\dots\dots(37).\end{aligned}$$

To find $\frac{\partial^2 f}{\partial h^2}$ etc. we may proceed as follows.

By straightforward differentiation we find that

$$\begin{aligned}\frac{\partial^2 f}{\partial h^2} &= \frac{\partial}{\partial h} \left\{ \frac{N}{2\pi\sqrt{1-r^2}} \int_{-\infty}^k e^{-\frac{y^2+h^2-2rhy}{2(1-r^2)}} dy \right\} \\ &= -\frac{hN}{2\pi\sqrt{1-r^2}} \int_{-\infty}^k e^{-\frac{y^2+h^2-2rhy}{2(1-r^2)}} dy + \frac{rN}{2\pi\sqrt{1-r^2}} e^{-\frac{h^2+k^2-2rkh}{2(1-r^2)}} \\ &= N \left(-hHA - \frac{r\chi}{N} \right) \dots\dots\dots(38),\end{aligned}$$

$$\frac{\partial^2 f}{\partial k^2} = N \left(-kKB - \frac{r\chi}{N} \right) \dots\dots\dots(39),$$

$$\frac{\partial^2 f}{\partial r^2} = \frac{\partial \chi}{\partial r} = \frac{N}{1-r^2} \left\{ r + \frac{(h-kr)(k-rh)}{1-r^2} \right\} \chi \dots\dots\dots(40),$$

$$\frac{\partial^2 f}{\partial h \partial r} = \frac{\partial \chi}{\partial h} = -\frac{h-kr}{1-r^2} \chi \dots\dots\dots(41),$$

$$\frac{\partial^2 f}{\partial k \partial r} = \frac{\partial \chi}{\partial k} = -\frac{k-rh}{1-r^2} \chi \dots\dots\dots(42),$$

$$\frac{\partial^2 f}{\partial h \partial k} = \frac{\partial f}{\partial r} = \chi \dots\dots\dots(43).$$

Hence summing for all samples and dividing by the number of samples and denoting this operation by $\mathfrak{S}$, we have, since quantities of the first order disappear, an equation involving $\mathfrak{S}(dh)^2$, $\mathfrak{S}(dhdk)$, etc. We may determine these values thus:

$$\begin{aligned}dh_s &= \frac{dm_{s\cdot}}{NH_s}, \\ \therefore \mathfrak{S}(dh_s)^2 &= \mathfrak{S} \left(\frac{dm_{s\cdot}}{NH_s} \right)^2 = \frac{m_{s\cdot} \left(1 - \frac{m_{s\cdot}}{N} \right)}{N^2 H_s^2} \dots\dots\dots(44).\end{aligned}$$

$$\text{Similarly} \quad \mathfrak{S}(dk_t)^2 = \frac{m_{\cdot t} \left(1 - \frac{m_{\cdot t}}{N} \right)}{N^2 K_t^2} \dots\dots\dots(45),$$

$$\mathfrak{S}(dh_s dk_t) = \mathfrak{S} \left(\frac{dm_{s\cdot} dm_{\cdot t}}{N^2 HK} \right) = \frac{m_{st} - \frac{m_{s\cdot} m_{\cdot t}}{N}}{N^2 HK} \dots\dots\dots(46).$$

To find $(dhdr)$ we have

$$\begin{aligned}-\chi dr &= A_{st} dm_{s\cdot} + B_{st} dm_{\cdot t} - dm_{st}, \\ \therefore -\chi dr dh_s &= -\frac{\chi dr dm_{s\cdot}}{NH_s}, \\ \therefore -\chi NH_s \mathfrak{S}(dr dh_s) &= A_{st} \mathfrak{S}(dm_{s\cdot})^2 + B_{st} \mathfrak{S}(dm_{s\cdot} dm_{\cdot t}) - \mathfrak{S}(dm_{st} dm_{\cdot t}) \\ &= A_{st} m_{s\cdot} \left(1 - \frac{m_{s\cdot}}{N} \right) + B_{st} \left(m_{st} - \frac{m_{s\cdot} m_{st}}{N} \right) - m_{st} \left(1 - \frac{m_{s\cdot}}{N} \right) \dots\dots\dots(47).\end{aligned}$$

After simplification this becomes

$$\left(1 - \frac{m_{s\cdot}}{N} \right) P_{st} - B_{st} (m_{\cdot t} - m_{st}),$$

$$\therefore \mathfrak{S}(drdh) = -\frac{1}{NH\chi} \left\{ \left(1 - \frac{m_{s\cdot}}{N}\right) P_{st} - B(m_{\cdot t} - m_{st}) \right\} \dots\dots\dots(48).$$

Similarly $\mathfrak{S}(drdk) = -\frac{1}{NK\chi} \left\{ \left(1 - \frac{m_{\cdot t}}{N}\right) P_{st} - A(m_{s\cdot} - m_{st}) \right\} \dots\dots\dots(49).$

Equation (37) then becomes

$$\begin{aligned} 0 = & N \left(-hHA - \frac{r\chi}{N} \right) \mathfrak{S}(dh)^2 + N \left(-kKB - \frac{r\chi}{N} \right) \mathfrak{S}(dk)^2 \\ & + \frac{\chi}{1-r^2} \left\{ r + \frac{(h-kr)(k-rh)}{1-r^2} \right\} \mathfrak{S}(dr)^2 - 2 \frac{h-rk}{1-r^2} \chi \cdot \mathfrak{S}(dhdr) \\ & - 2 \frac{k-rh}{1-r^2} \mathfrak{S}(dkdr) + 2\chi \mathfrak{S}(dhdk) \dots\dots\dots(50). \end{aligned}$$

$\therefore$ Substituting the values of the $\mathfrak{S}$'s, and transposing σ_r^2 we find

$$\begin{aligned} \frac{\chi}{1-r^2} \left\{ r + \frac{(h-kr)(k-rh)}{1-r^2} \right\} \sigma_r^2 = & -N \left(hHA + r \frac{\chi}{N} \right) \frac{m_{s\cdot} \left(1 - \frac{m_{s\cdot}}{N}\right)}{N^2 H_s^2} \\ & - N \left(kKB + r \frac{\chi}{N} \right) \frac{m_{\cdot t} \left(1 - \frac{m_{\cdot t}}{N}\right)}{N^2 K_t^2} \\ & + 2 \frac{h-rk}{1-r^2} \chi \cdot \frac{1}{NH\chi} \left\{ \left(1 - \frac{m_{s\cdot}}{N}\right) P_{st} - B_{st}(m_{\cdot t} - m_{st}) \right\} \\ & + 2 \frac{k-rh}{1-r^2} \chi \cdot \frac{1}{NH\chi} \left\{ \left(1 - \frac{m_{\cdot t}}{N}\right) P_{st} - A_{st}(m_{s\cdot} - m_{st}) \right\} \\ & + 2\chi \cdot \frac{m_{st} - \frac{m_{s\cdot} m_{\cdot t}}{N}}{N^2 HK} \dots\dots\dots(51). \end{aligned}$$

When $r = 0$ and $h_1 = -h_2$, $k_1 = -k_2$, χ reduces to NHK , etc., as above and the equation reduces to

$$-NHKhk\sigma_r^2 = -\frac{hm_{st}}{NH} \left(1 - \frac{m_{s\cdot}}{N}\right) - \frac{km_{st}}{NK} \left(1 - \frac{m_{\cdot t}}{N}\right) \dots\dots\dots(52).$$

If we take this equation for each of the four points 11, 12, 21, 22 and combine them according to the scheme $m_{22} = m_{11} - m_{12} - m_{21} + m_{22}$ as before, the left member of the equation becomes

$$\begin{aligned} -NHK(h_1k_1 - h_1k_2 - h_2k_1 + h_2k_2)\sigma_r^2 = & -NHK(h_1k_1 + h_1k_1 + h_1k_1 + h_1k_1)\sigma_r^2 \\ = & -4NHKh_1k_1\sigma_r^2 \dots\dots\dots(53). \end{aligned}$$

In a similar manner the right side reduces to

$$\frac{h_1}{NH_1} \left(n_{12} + \frac{m_{1\cdot}}{N} n_{22} \right) + \frac{k_1}{NK_1} \left(n_{21} + \frac{m_{\cdot 1}}{N} n_{22} \right) \dots\dots\dots(54),$$

which may be further simplified as follows

$$\begin{aligned} n_{12} + \frac{m_{1\cdot}}{N} n_{22} &= n_{12} + \frac{m_{1\cdot}}{N} (m_{\cdot 2} - 2n_{12}) \\ &= 2n_{12} \left(1 - \frac{m_{1\cdot}}{N} \right) \end{aligned}$$

$$\begin{aligned}
&= 2 \frac{m_{1\cdot} m_{\cdot 2}}{N} \left(1 - \frac{N - m_{2\cdot}}{N}\right) \\
&= 2 \frac{m_{1\cdot} m_{\cdot 2}}{N} \frac{m_{2\cdot}}{N} \dots\dots\dots(55).
\end{aligned}$$

Similarly
$$n_{21} + \frac{m_{\cdot 1}}{N} n_{22} = 2 \frac{m_{\cdot 1} m_{2\cdot} m_{\cdot 2}}{NN} \dots\dots\dots(56).$$

Hence substituting we have

$$-4NHKh_1 k_1 \sigma_r^2 = 2 \frac{n_{22}}{N^2} \left\{ \frac{h_1}{H_1} m_{1\cdot} + \frac{k_1}{K_1} m_{\cdot 1} \right\} \dots\dots\dots(57),$$

and

$$\sigma_r^2 = - \frac{n_{22}}{2N^3 H K h_1 k_1} \left\{ \frac{h_1}{H_1} m_{1\cdot} + \frac{k_1}{K_1} m_{\cdot 1} \right\} \dots\dots\dots(58)$$

(h_1 and k_1 are of course negative numbers) which remains finite when the extreme categories are equal to $r = 0$.

(2) When we put $h_2 = k_2 = \infty$ the enneachoric table degenerates into a tetrachoric table and we have

$$n_{13} = n_{23} = n_{33} = 0,$$

$$n_{31} = n_{32} = n_{33} = 0,$$

$$n_{3\cdot} = n_{\cdot 3} = 0,$$

$$\chi_{12} = \chi_{21} = \chi_{22} = 0,$$

and

$$m = \frac{\alpha_{12} n_{1\cdot} + \beta_{21} n_{\cdot 1} + n_{22}}{N}.$$

But

$$\begin{aligned}
\alpha_{12} &= A_{12} - A_{11} \\
&= \mathfrak{E} \left(\frac{\infty - r h_1}{\sqrt{1 - r^2}} \right) - \mathfrak{E} \left(\frac{k_1 - r h_1}{\sqrt{1 - r^2}} \right) = 1 - A_{11}.
\end{aligned}$$

Similarly

$$\beta_{21} = 1 - B_{11}.$$

Hence
$$m = \frac{(1 - A_{11}) n_{1\cdot} + (1 - B_{11}) n_{\cdot 1} + N - n_{1\cdot} - n_{\cdot 1} + n_{11}}{N}$$

$$= \frac{N - (A_{11} n_{1\cdot} + B_{11} n_{\cdot 1} - n_{11})}{N}$$

$$= 1 - \frac{P_{11}}{N}$$

$$\begin{aligned}
\therefore \chi_{11}^2 \sigma_r^2 &= \left(\frac{P_{11}}{N} - A_{11} - B_{11} + 1 \right)^2 n_{11} + \left(\frac{P_{11}}{N} - B_{11} \right)^2 n_{21} \\
&\quad + \left(\frac{P_{11}}{N} - A_{11} \right)^2 n_{12} + \left(\frac{P_{11}}{N} \right)^2 n_{22} \dots\dots\dots(59).
\end{aligned}$$

This form of the standard deviation of a tetrachoric correlation will be referred to again later, and will then be reduced to a still more symmetrical form.

From the above it will be seen that we can determine the correlation coefficient from any frequency block, assuming a normal distribution, but the accuracy of this determination varies with the position and size of the frequency block. The probable error ($\cdot 67449\sigma$.) varies from cell to cell, and an unlucky choice of the working cell may lead to a correlation coefficient with large probable error. A correction

of this latitude might in some cases be obtained by using another cell. But as the r from this cell would probably differ from that previously found, and as neither of them would be identical with that of the normal surface from which they are supposed to be sampled, we must find some means of approximating to the "best" r . The most general method of doing this, following on the above, would seem to be to weight each of the frequency blocks and determine the weights so that the resulting probable error of the weighted r is a minimum. In doing this we must have regard to the fact that the variates we are dealing with are not independent but correlated. We must consider this method.

Let the polychoric table have p rows and q columns so that the indices of the last row and column are $1q, 2q, \dots pq$ and $p1, p2, \dots pq$ respectively, and let each of the frequency volumes be weighted by an arbitrary weight $w_{11}, w_{12}, \dots$ indicated by the same suffix as their respective cells.

$$\begin{aligned} & \text{Then} \quad w_{11}n_{11} + w_{12}n_{12} + \dots \\ &= w_{11}m_{11} + w_{12}(m_{12} - m_{11}) + w_{21}(m_{21} - m_{11}) + w_{22}(m_{22} - m_{12} - m_{21} + m_{11}) + \dots \\ & \quad + w_{st}(m_{st} - m_{s-1,t} - m_{s,t-1} + m_{s-1,t-1}) + \dots \\ &= (w_{11} - w_{12} - w_{21} + w_{22})m_{11} + \dots + (w_{st} - w_{s+1,t} - w_{s,t+1} + w_{s+1,t+1})m_{st} + \dots \\ &= \omega_{11}m_{11} + \omega_{12}m_{12} + \dots + \omega_{st}m_{st} \dots\dots\dots(60). \end{aligned}$$

$$\begin{aligned} & \text{Then} \quad \frac{1}{N}(w_{11}n_{11} + w_{12}n_{12} + \dots) = \frac{1}{N}(\omega_{11}m_{11} + \omega_{12}m_{12} + \dots) \\ & \quad = \omega_{11}({}_1\tau_{01}\theta_0 + \mathfrak{T}_{11}) + \omega_{12}({}_1\tau_{02}\theta_0 + \mathfrak{T}_{12}) + \dots\dots\dots(61), \end{aligned}$$

which is an equation to find r , using $\mathfrak{T}_{st}$ as an abbreviation for

$${}_s\tau_{1,t}\theta_1r + {}_s\tau_{2,t}\theta_2r^2 + {}_s\tau_{3,t}\theta_3r^3 + \text{etc.}$$

(Compare the usage in equations (2) to (5).)

$m_{1q}, m_{2q}, \dots m_{p1}, m_{p2} \dots$ are complete segments of the normal solid and are independent of r , i.e. $\frac{m_{sq}}{N} = {}_s\tau_{0q}\theta_0$, and these terms disappear from both sides of the equation. The w 's of the cells in the last row and last column of the table may therefore be dropped and we have $(p-1)(q-1)$ w 's to determine. The probable error may be written down in the manner already shown and as there are $(p-1)(q-1)$ independent frequencies there is sufficient data to determine the w 's so that the probable error is a minimum.

There is no essential difficulty in carrying this out except that the coefficients rapidly become very cumbersome as tables increase beyond 3×3 for then the simplification $dm_2 = -dn_{.3}$ is no longer available. It will be found however that the same result may be derived more simply from the method discussed below.

§ 6. POLYCHORIC METHOD.

The frequency surface divided into p columns and q rows is divided at each point $11, 12 \dots$ into four quadrants, and for each of these divisions a value for r , viz., $r_{11}, r_{12} \dots$, may be found by the tetrachoric method. These may be regarded as approximations to the true value of r , and their weighted mean found, the weights

being determined so that the probable error of the mean r so found shall be a minimum.

$$\text{Let } r = \frac{C_{11}r_{11} + C_{12}r_{12} + \dots}{C_{11} + C_{12} + \dots} \dots\dots\dots(62).$$

$$\text{Then } (C_{11} + C_{12} + \dots) dr = C_{11}dr_{11} + C_{12}dr_{12} + \dots \dots\dots(63).$$

Squaring, summing for all possible values and dividing by the number of samples,

$$(\Sigma C_{st})^2 \sigma_r^2 = \Sigma (C_{st} \sigma_{st})^2 + 2 \Sigma (C_{st} C_{s't'} \sigma_{st} \sigma_{s't'} R_{st, s't'}) \dots\dots\dots(64),$$

where

$$\sigma_{st} = \sigma_{r_{st}}, \quad R_{st, s't'} = r_{r_{st}, r_{s't'}}.$$

$$\text{Let } S = \Sigma (C_{st} \sigma_{st})^2 + 2 \Sigma (C_{st} C_{s't'} \sigma_{st} \sigma_{s't'} R_{st, s't'}),$$

$$C = \Sigma (C_{st}).$$

Then for a minimum

$$dS = \frac{\partial S}{\partial C_{11}} dC_{11} + \frac{\partial S}{\partial C_{12}} dC_{12} + \dots = 0,$$

$$dC = dC_{11} + dC_{12} + \dots = 0,$$

$$\therefore \left(\frac{\partial S}{\partial C_{11}} - \lambda \right) dC_{11} = \left(\frac{\partial S}{\partial C_{12}} - \lambda \right) dC_{12} = \dots = 0.$$

$$\therefore C_{11} \sigma_{11}^2 + C_{12} \sigma_{11} \sigma_{12} R_{11, 12} + C_{13} \sigma_{11} \sigma_{13} R_{11, 13} + \dots = \lambda,$$

$$C_{11} \sigma_{11} \sigma_{12} R_{11, 12} + C_{12} \sigma_{12}^2 + C_{13} \sigma_{12} \sigma_{13} R_{12, 13} + \dots = \lambda \dots\dots\dots(65).$$

Let

$$\Delta = \begin{vmatrix} 1 & \frac{1}{\sigma_{11}} & \frac{1}{\sigma_{12}} & \frac{1}{\sigma_{13}} & \dots \\ \frac{1}{\sigma_{11}} & 1 & R_{11, 12} & R_{11, 13} & \dots \\ \frac{1}{\sigma_{12}} & R_{11, 12} & 1 & R_{12, 13} & \dots \\ \frac{1}{\sigma_{13}} & R_{11, 13} & R_{12, 13} & 1 & \dots \end{vmatrix} \dots\dots\dots(66).$$

Then

$$C_{11} \sigma_{11} = - \frac{\lambda \Delta_{0011}}{\Delta_{0000}},$$

$$C_{12} \sigma_{12} = - \frac{\lambda \Delta_{0012}}{\Delta_{0000}} \dots\dots\dots(67).$$

Since λ is arbitrary we may put $\Sigma (C_{st}) = 1$,

$$\begin{aligned} \Sigma (C_{st}) &= - \frac{\lambda}{\Delta_{0000}} \left\{ \frac{\Delta_{0000}}{1} + \frac{\Delta_{0011}}{\sigma_{11}} + \frac{\Delta_{0012}}{\sigma_{12}} + \dots - \Delta_{0000} \right\} \\ &= - \frac{\lambda (\Delta - \Delta_{0000})}{\Delta_{0000}} = 1 \dots\dots\dots(68), \end{aligned}$$

$$\therefore \lambda = - \frac{\Delta_{0000}}{\Delta - \Delta_{0000}} \dots\dots\dots(69).$$

$$\therefore C_{11} \sigma_{11} = \frac{\Delta_{0011}}{\Delta - \Delta_{0000}} \dots\dots\dots(70).$$

$$\begin{aligned} \therefore \Sigma (C_{st}\sigma_{st})^2 + 2\Sigma (C_{st}C_{s't'}\sigma_{st}\sigma_{s't'}R_{st,s't'}) \\ = \frac{1}{\Delta - \Delta_{0000}} \Sigma \left\{ C_{st}\sigma_{st} \left(\frac{\Delta_{0000}}{\sigma_{st}} + \Delta_{0011}R_{11st} + \Delta_{0012}R_{12\cdot st} \right. \right. \\ \left. \left. + \Delta_{00st} + \dots - \frac{\Delta_{0000}}{\sigma_{st}} \right) \right\} \\ = \frac{1}{\Delta - \Delta_{0000}} \Sigma \left\{ C_{st}\sigma_{st} \left(-\frac{\Delta_{0000}}{\sigma_{st}} \right) \right\} \\ = -\frac{\Sigma C_{st}\cdot\Delta_{0000}}{\Delta - \Delta_{0000}} = \frac{\Delta_{0000}}{\Delta_{0000} - \Delta} \dots\dots\dots(71), \end{aligned}$$

and since $\Sigma C_{st} = 1$ $\sigma_r^2 = \frac{\Delta_{0000}}{\Delta_{0000} - \Delta} \dots\dots\dots(72).$

§ 7. COMPARISON OF POLYCHORIC AND ENNEACHORIC COEFFICIENTS OF CORRELATION.

We may now compare the polychoric coefficient (r_p) with the enneachoric coefficient (r_e) previously found.

Since $\Sigma (C_{st}) = 1,$
 $r_p = C_{11}r_{11} + C_{12}r_{12} + \dots \dots\dots(73),$

$\therefore dr_p = C_{11}dr_{11} + C_{12}dr_{12} + \dots = -C_{11}\frac{\delta P_{11}}{\chi_{11}} - C_{12}\frac{\delta P_{12}}{\chi_{12}} \dots\dots\dots(74),$

$\therefore -(C_{11}dr_{11} + C_{12}dr_{12} + \dots)$
 $= \frac{C_{11}}{\chi_{11}} (A_{11}dm_{1\cdot} + B_{11}dm_{\cdot 1} - dm_{11}) + \frac{C_{12}}{\chi_{12}} (A_{12}dm_{1\cdot} + B_{12}dm_{\cdot 2} - dm_{12}) + \text{etc.}$
 $\dots\dots\dots(75).$

Transposing and rearranging we find

$$\begin{aligned} \frac{C_{11}}{\chi_{11}} dm_{11} + \frac{C_{12}}{\chi_{12}} dm_{12} + \dots \\ = \frac{C_{11}}{\chi_{11}} (A_{11}dm_{1\cdot} + B_{11}dm_{\cdot 1} + \chi_{11}dr_{11}) + \frac{C_{12}}{\chi_{12}} (A_{12}dm_{1\cdot} + B_{12}dm_{\cdot 2} + \chi_{12}dr_{12}) + \text{etc.} \\ = \frac{C_{11}}{\chi_{11}} \left(\frac{\partial f_{11}}{\partial h_1} dh_1 + \frac{\partial f_{11}}{\partial k_1} dk_1 + \frac{\partial f_{11}}{\partial r_{11}} dr_{11} \right) + \frac{C_{12}}{\chi_{12}} \left(\frac{\partial f_{12}}{\partial h_1} dh_1 + \frac{\partial f_{12}}{\partial k_2} dk_2 + \frac{\partial f_{12}}{\partial r_{12}} dr_{12} \right) + \dots \end{aligned}$$

$\dots\dots\dots(76),$

where $f_{st} = N \int_{-\infty}^{h_s} \int_{-\infty}^{k_t} z(x, y, r) dx dy = m_{st},$

$$\frac{C_{11}}{\chi_{11}} m_{11} + \frac{C_{12}}{\chi_{12}} m_{12} + \dots = \frac{C_{11}}{\chi_{11}} ({}_1\tau_0{}_1\theta_0 + \mathfrak{T}_{10}) + \frac{C_{12}}{\chi_{12}} ({}_1\tau_0{}_2\theta_0 + \mathfrak{T}_{12}) + \dots$$

$\dots\dots\dots(77),$

using $\mathfrak{T}_{st}$ as in equation (61), page 106, which is an equation of which r_p is a root. If in this equation we put $C_{11} = \chi_{11}, C_{12} = -\chi_{12}, C_{21} = -\chi_{21}, C_{22} = \chi_{22},$ and the remaining C 's = 0 we have the equation for the enneachoric coefficient and r_e now appears to be the weighted mean of four tetrachoric r 's,

$$r_e = \frac{\chi_{11}r_{11} - \chi_{12}r_{12} - \chi_{21}r_{21} + \chi_{22}r_{22}}{\chi_{11} - \chi_{12} - \chi_{21} + \chi_{22}}.$$

Comparing this with the generalized equation (61) for enneachoric r previously found in which all the frequency volumes were weighted, we have

$$\begin{aligned}\omega_{11} &= \frac{C_{11}}{\chi_{11}}, \\ \omega_{12} &= \frac{C_{12}}{\chi_{12}}, \\ &\dots\dots\dots \\ \omega_{st} &= \frac{C_{st}}{\chi_{st}},\end{aligned}$$

and since $\omega_{st} = w_{st} - w_{s+1, t} - w_{s, t+1} + w_{s+1, t+1}$,
it can be easily shown* if there are p rows and q columns that

$$\begin{aligned}w_{st} &= \omega_{st} + \omega_{s, t+1} + \omega_{s, t+2} + \dots \omega_{s, q-1} + \omega_{s+1, t} + \omega_{s+1, t+1} + \dots \omega_{s+1, q-1} \\ &\dots\dots\dots \\ &+ \omega_{p-1, t} + \omega_{p-2, t+1} + \dots \omega_{p, q} \dots\dots\dots (78),\end{aligned}$$

that is the sum of all the weights having the same suffixes as the points contained within the d quadrant of which n_{st} occupies the corner.

* Consider a two-fold extension ruled and named in a manner similar to the polychoric scheme on page 94.

Then

$$\begin{aligned}m_{11} &= n_{11} \\ m_{12} &= m_{11} + n_{12} \\ &\dots\dots\dots \\ m_{st} &= n_{11} + n_{12} + n_{13} \dots\dots + m_{1t} \\ &\quad + n_{21} + n_{22} + n_{23} \dots\dots + m_{2t} \\ &\quad \dots\dots\dots \\ &\quad + n_{s1} + n_{s2} + n_{s3} \dots\dots + n_{st} \\ &\quad \dots\dots\dots\end{aligned}$$

Hence $\omega_{11}m_{11} + \omega_{12}m_{12} + \dots\dots$

$$\begin{aligned}&= \omega_{11}n_{11} + \omega_{12}(n_{11} + n_{12}) + \omega_{13}(n_{11} + n_{12} + n_{13}) \\ &\quad + \dots\dots \\ &\quad + \omega_{st}(n_{11} + n_{12} + \dots\dots n_{1t} \\ &\quad + n_{21} + n_{22} + \dots\dots n_{2t} \\ &\quad + \dots\dots \\ &\quad + n_{s1} + n_{s2} + \dots\dots n_{st}).\end{aligned}$$

If we rearrange this in terms of n_{11}, n_{12} we shall have

$$\begin{aligned}&n_{11}(\omega_{11} + \omega_{12} + \dots\dots \omega_{st}) \\ &+ n_{12}(\omega_{12} + \omega_{13} + \dots\dots \omega_{st}) \\ &= w_{11}n_{11} + w_{12}n_{12} + \dots\dots w_{st}n_{st} + \dots\dots w_{pq}n_{pq}.\end{aligned}$$

It is clear that w_{st} will be the sum of the w 's belonging to all the m 's of which n_{st} is a constituent. that is, from the figure, all the n 's whose boundary lines lie beyond the lines $h=s-1, k=t-1$, i.e.

$$w_{st} = \sum_{i=1}^s \sum_{j=1}^t \omega_{st}.$$

The relation

$$n_{st} = m_{st} - m_{s-1, t} - m_{s, t-1} + m_{s-1, t-1}$$

may be compared to the partial finite difference in two variables

$$\Delta_x \Delta_y U_{x, y} = U_{x+1, y+1} - U_{x+1, y} - U_{x, y+1} + U_{x, y},$$

which may help to make the above relation clearer.

§ 8. COMPUTATION OF R 'S AND σ 'S.

As the computation of r_p even for an enneachoric table is somewhat lengthy, it is necessary to have a definite scheme to work to. In addition to this the values of the R 's when resolved into their constituents present some interesting features.

A new expression for tetrachoric r has already been deduced from the degeneration of an enneachoric table. The following is a derivation of it directly and in a more symmetrical form.

Consider the tetrachoric table

	D		
	a	b	$n_{.1}$
F	E	L	
	c	d	$n_{.2}$
	G		
$n_{1.}$		$n_{2.}$	N

Let A and B have the same significance as before, i.e. A is the fraction which the area of the plane DE is of the whole dichotomic plane and B the same for FE , and write

$${}_aP = An_{1.} + Bn_{.1} - a \dots\dots\dots(79),$$

where the a suffix is used to indicate that it is the P of the leading or a quadrant.

Then since the fractional area of EG will be $1 - A$ and of EL , $1 - B$, the corresponding P for b quadrant will be

$$\begin{aligned} {}_bP &= An_{2.} + (1 - B) n_{.1} - b \\ &= A (N - n_{1.}) + (1 - B) n_{.1} - (n_{.1} - a) \\ &= AN - (An_{1.} + Bn_{.1} - a) \\ &= AN - {}_aP \dots\dots\dots(80). \end{aligned}$$

Similarly ${}_cP = BN - {}_aP \dots\dots\dots(81),$

$${}_dP = -N (A + B - 1) + {}_aP \dots\dots\dots(82).$$

Hence we have ${}_aP + {}_bP + {}_cP + {}_dP = N \dots\dots\dots(82) \text{ bis.}$

We have already seen (20) that

$$-\chi^2 dr = Adm_{1.} + Bdm_{.1} - dm_{11} = \delta {}_aP_{11} \dots\dots\dots(83).$$

Using the symbol $\mathfrak{S}$ as before to denote the operation "sum for all possible samples and divide by the number of samples"* we have

* It would be useful to have a distinctive name for this operation, verb as well as substantive.

$$\begin{aligned}
\chi^2 \sigma_r^2 &= \sum (\delta_a P)^2 \\
&= \sum (A d m_{1.} + B d m_{.1} - d m_{11})^2 \\
&= A^2 m_{1.} + B^2 m_{.1} + m_{11} + 2AB m_{11} - 2A m_{11} - 2B m_{11} - \frac{a P^2}{N} \\
&= A^2 (a + c) + B^2 (a + b) + a + (2AB - 2A - 2B) a - \frac{a P^2}{N} \\
&= (A + B - 1)^2 a + A^2 c + B^2 b - \frac{a P^2}{N} \dots\dots\dots(84).
\end{aligned}$$

But

$${}_a P = A(a + c) + B(a + b) - a = (A + B - 1)a + Ac + Bb + 0 \cdot d \dots\dots(85)$$

and

$$a + b + c + d = N,$$

$$\begin{aligned}
\therefore \sum (\delta_a P)^2 &= \left(A + B + 1 - \frac{aP}{N}\right)^2 a + \left(B - \frac{aP}{N}\right)^2 b + \left(A - \frac{aP}{N}\right)^2 c + \left(0 - \frac{aP}{N}\right)^2 d \\
&= \left(-\frac{aP}{N}\right)^2 a + \left(\frac{cP}{N}\right)^2 b + \left(\frac{bP}{N}\right)^2 c + \left(-\frac{aP}{N}\right)^2 d \\
&\quad \frac{1}{N^2} \left\{ a^2 P^2 a + c^2 P^2 b + b^2 P^2 c + a^2 P^2 d \right\} \dots\dots\dots(86).
\end{aligned}$$

Further

$$\begin{aligned}
&(-{}_a P) a + ({}_c P) b + ({}_b P) c + (-{}_a P) a \\
&{}_a \{N(A + B - 1) - {}_a P\} + b(BN - {}_a P) + c(AN - {}_a P) + d(-{}_a P) \\
&= N\{(A + B - 1)a + Bb + Ac\} - N {}_a P \\
&= N\{A(a + c) + B(a + b) - a\} - N {}_a P \\
&= N {}_a P - N {}_a P = 0 \dots\dots\dots(87).
\end{aligned}$$

The above form (86) of the square of the standard deviation (omitting factor) is interesting as involving only the squares of the P 's. Since the P 's are connected by the relation (82) *bis* and (87) their values may be determined from any two of them.

The R 's.

Since

$$-\chi_{st} \delta r_{st} = \delta P_{st}$$

and

$$-\chi_{s't'} \delta r_{s't'} = \delta P_{s't'},$$

$$\chi_{st} \chi_{s't'} \delta r_{st} \delta r_{s't'} = \delta P_{st} \delta P_{s't'}$$

and

$$\chi_{st} \chi_{s't'} \sigma_{st} \sigma_{s't'} R_{st.s't'} = \sum (\delta P_{st} \delta P_{s't'}) = S_{st.s't'} \text{ (say).}$$

In conformity with this notation

$$\chi_{st}^2 \sigma_{st}^2 = S_{st.st}.$$

It is useful to have a verbal rule for writing down such mean products as

$$\sum (\delta P_{st} \delta P_{s't'}).$$

The following will serve.

Multiply the detached coefficients of the differentials in the δP 's as in ordinary multiplication; strike out the products in which the related frequencies have no common frequency and insert the common part of the frequencies after the related coefficients. From the whole subtract the full products of the P 's divided by the total frequency. This may be proved as follows:

Let p and q be any frequencies in a given distribution in which the population N is so large that sampling does not alter its composition; then we have the well-known results (p. 99)

$$\sigma_p^2 = p \left(1 - \frac{p}{N} \right),$$

$$\mathfrak{S}(dp dq) = -\frac{pq}{N}.$$

Now let p and q have a common part c so that

$$p = p' + c,$$

$$q = q' + c.$$

$$\begin{aligned} \text{Then} \quad \mathfrak{S}(dp dq) &= \mathfrak{S}d(p' + c) \cdot d(q' + c) \\ &= \mathfrak{S}(dp' dq' + dp' dc + dq' dc + dc^2) \\ &= -\frac{p'q'}{N} - \frac{p'c}{N} - \frac{q'c}{N} + c \left(1 - \frac{c}{N} \right) \\ &= -\frac{(q' + c)(p' + c)}{N} + c \\ &= c - \frac{pq}{N}. \end{aligned}$$

Now the mean product of any two linear functions of p 's and q 's,

$$S = \mathfrak{S}d(i_1 p_1 + i_2 p_2 + \dots) \cdot d(k_1 q_1 + k_2 q_2 + \dots),$$

will consist of the sum of the mean products of terms such as

$$i_s dp_s \cdot k_t dq_t.$$

$$\begin{aligned} \text{But} \quad \mathfrak{S}(i_s dp_s \cdot k_t dq_t) &= i_s k_t \mathfrak{S}(dp_s \cdot dq_t) \\ &= i_s k_t \left(c_{st} - \frac{p_s q_t}{N} \right), \end{aligned}$$

where c is the common part of p_s and q_t .

$$\begin{aligned} \text{Therefore} \quad S &= \Sigma \left\{ i_s k_t \left(c_{st} - \frac{p_s q_t}{N} \right) \right\} \\ &= \Sigma i_s k_t c_{st} - \Sigma i_s k_t \frac{p_s q_t}{N}. \end{aligned}$$

Hence the rule.

As an example consider $S_{11 \cdot 21}$,

$$\begin{aligned} S_{11 \cdot 21} &= \mathfrak{S}(\delta P_{11} \delta P_{21}) \\ &= \mathfrak{S}(A_{11} dm_{1 \cdot} + B_{11} dm_{\cdot 1} - dm_{11})(A_{21} dm_{2 \cdot} + B_{21} dm_{\cdot 1} - dm_{21}) \\ &= A_{11} A_{21} m_{1 \cdot} + A_{11} B_{21} m_{11} - A_{11} m_{11} + B_{11} A_{21} m_{21} + B_{11} B_{21} m_{\cdot 1} - B_{11} m_{21} \\ &\quad - A_{21} m_{11} - B_{21} m_{11} + m_{11} - \frac{P_{11} P_{21}}{N} \\ &= (A_{11} + B_{11} - 1)(A_{21} + B_{21} - 1)n_{11} + B_{11}(A_{21} + B_{21} - 1)n_{21} + B_{11} B_{21} n_{31} \\ &\quad + A_{11} A_{21} n_{12} \qquad \qquad \qquad + 0 \cdot A_{21} \cdot n_{22} \qquad \qquad \qquad + 0 \cdot 0 \cdot n_{32} \\ &\quad + A_{11} A_{21} n_{13} \qquad \qquad \qquad + 0 \cdot A_{21} \cdot n_{23} \qquad \qquad \qquad + 0 \cdot 0 \cdot n_{33} \end{aligned}$$

$$\begin{aligned}
& -\frac{1}{N} \left\{ \begin{array}{l} (A_{11} + B_{11} - 1) n_{11} + B_{11} n_{21} + B_{11} n_{31} \\ + \quad A_{11} n_{12} + 0 \cdot n_{22} + 0 \cdot n_{32} \\ + \quad A_{11} n_{13} + 0 \cdot n_{23} + 0 \cdot n_{33} \end{array} \right\} \\
& \times \left\{ \begin{array}{l} (A_{21} + B_{21} - 1) n_{11} + (A_{21} + B_{21} - 1) n_{21} + B_{21} n_{31} \\ + \quad A_{21} n_{12} + A_{21} n_{22} + 0 \cdot n_{32} \\ + \quad A_{21} n_{13} + A_{21} n_{23} + 0 \cdot n_{33} \end{array} \right\} \\
& = \left(A_{11} + B_{11} - 1 - \frac{P_{11}}{N} \right) \left(A_{21} + B_{21} - 1 - \frac{P_{21}}{N} \right) n_{11} \\
& \quad + \left(B_{11} - \frac{P_{11}}{N} \right) \left(A_{21} + B_{21} - 1 - \frac{P_{21}}{N} \right) n_{21} + \left(B_{11} - \frac{P_{11}}{N} \right) \left(B_{21} - \frac{P_{21}}{N} \right) n_{31} \\
& \quad + \left(A_{11} - \frac{P_{11}}{N} \right) \left(A_{21} - \frac{P_{21}}{N} \right) n_{12} + \left(0 - \frac{P_{11}}{N} \right) \left(A_{21} - \frac{P_{21}}{N} \right) n_{22} \\
& \quad + \left(0 - \frac{P_{11}}{N} \right) \left(0 - \frac{P_{21}}{N} \right) n_{32} + \left(A_{11} - \frac{P_{11}}{N} \right) \left(A_{21} - \frac{P_{21}}{N} \right) n_{13} \\
& \quad + \left(0 - \frac{P_{11}}{N} \right) \left(A_{21} - \frac{P_{21}}{N} \right) n_{23} + \left(0 - \frac{P_{11}}{N} \right) \left(0 - \frac{P_{21}}{N} \right) n_{33} \dots\dots\dots (88).
\end{aligned}$$

In the above the P 's are ${}_aP$'s and remembering that

$$A_{11} + B_{11} - 1 - \frac{{}_aP_{11}}{N} = -\frac{{}_dP_{11}}{N}, \text{ etc.,}$$

we have

$$\begin{aligned}
S_{11 \cdot 21} &= \left(-\frac{{}_dP_{11}}{N} \right) \left(-\frac{{}_dP_{21}}{N} \right) n_{11} + \left(\frac{{}_cP_{11}}{N} \right) \left(-\frac{{}_dP_{21}}{N} \right) n_{21} + \left(\frac{{}_cP_{11}}{N} \right) \left(\frac{{}_cP_{21}}{N} \right) n_{31} \\
& \quad + \left(\frac{{}_bP_{11}}{N} \right) \left(\frac{{}_bP_{21}}{N} \right) n_{12} + \left(-\frac{{}_aP_{11}}{N} \right) \left(\frac{{}_bP_{21}}{N} \right) n_{22} + \left(-\frac{{}_aP_{11}}{N} \right) \left(-\frac{{}_aP_{12}}{N} \right) n_{32} \\
& \quad + \left(\frac{{}_bP_{11}}{N} \right) \left(\frac{{}_bP_{21}}{N} \right) n_{13} + \left(-\frac{{}_aP_{11}}{N} \right) \left(\frac{{}_bP_{21}}{N} \right) n_{23} + \left(-\frac{{}_aP_{11}}{N} \right) \left(-\frac{{}_aP_{21}}{N} \right) n_{33} \\
&= \frac{1}{N^2} \{ {}_dP_{11} {}_dP_{21} n_{11} - {}_cP_{11} {}_dP_{21} n_{21} + {}_cP_{11} {}_cP_{21} n_{31} \\
& \quad + {}_bP_{11} {}_bP_{21} (n_{12} + n_{13}) - {}_aP_{11} {}_bP_{21} (n_{22} + n_{23}) + {}_aP_{11} {}_aP_{21} (n_{32} + n_{33}) \} \\
& \quad \dots\dots\dots (89).
\end{aligned}$$

The relation between the coefficients in the above expression is very simple. We have already seen that

$$N^2 \chi^2 \sigma_r^2 = a (-{}_dP)^2 + b ({}_cP)^2 + c ({}_bP)^2 + d (-{}_aP)^2 \dots\dots\dots (90).$$

In the quadrants of a tetrachoric table write the P coefficients. Thus

$$\begin{array}{c|c}
-{}_dP & +{}_cP \\
\hline
+{}_bP & -{}_aP
\end{array}$$

The a frequency is related to $-{}_aP$, etc.

Consider now the empty scheme of an enneachoric table regarded as a tetrachoric table with the point of division first at, say, 11 and second at 12, and write in the P coefficients as above.

Divided at 11.

$- {}_aP_{11}$	${}_cP_{11}$	${}_cP_{11}$
${}_bP_{11}$	$- {}_aP_{11}$	$- {}_aP_{11}$
${}_bP_{11}$	$- {}_aP_{11}$	$- {}_aP_{11}$

Divided at 21.

$- {}_aP_{21}$	$- {}_aP_{21}$	${}_cP_{21}$
${}_bP_{21}$	${}_bP_{21}$	$- {}_aP_{21}$
${}_bP_{21}$	${}_bP_{21}$	$- {}_aP_{21}$

Now if we superpose these two schemes upon an enneachoric table with a frequency in each cell, each cell will then contain the P coefficient and frequency of each term of the expansion of $S_{11 \cdot 21}$ (with the omission of the factor $\frac{1}{N^2}$) thus

$(- {}_aP_{11}) (- {}_aP_{21}) n_{11}$	$({}_cP_{11}) (- {}_aP_{21}) n_{21}$	$({}_cP_{11}) ({}_cP_{21}) n_{31}$
$({}_bP_{11}) ({}_bP_{21}) n_{12}$	$(- {}_aP_{11}) ({}_bP_{21}) n_{22}$	$(- {}_aP_{11}) (- {}_aP_{21}) n_{32}$
$({}_bP_{11}) ({}_bP_{21}) n_{13}$	$(- {}_aP_{11}) ({}_bP_{21}) n_{23}$	$(- {}_aP_{11}) (- {}_aP_{21}) n_{33}$

When 11 coincides with 21, R becomes = 1 and the mean product degenerates into the square of the standard deviation.

This may be summarised in the following table in which the letter, a , b , etc. gives the suffix and the sign gives the sign of the P required.

	P_{11}	P_{12}	P_{21}	P_{22}
n_{11}	$-d$	$-d$	$-d$	$-d$
n_{12}	$+b$	$-d$	$+b$	$-d$
n_{13}	$+b$	$+b$	$+b$	$+b$
n_{21}	$+c$	$+c$	$-d$	$-d$
n_{22}	$-a$	$+c$	$+b$	$-d$
n_{23}	$-a$	$-a$	$+b$	$+b$
n_{31}	$+c$	$+c$	$+c$	$+c$
n_{32}	$-a$	$+c$	$-a$	$+c$
n_{33}	$-a$	$-a$	$-a$	$-a$

Thus the coefficient of n_{32} in $S_{12 \cdot 21}$ is $(+ {}_cP_{12}) (- {}_aP_{21})$.

This table is sufficient for a polychoric table of any size since any two cross points st . $s't'$ in the table, with the planes through them divide it into nine portions or groups of cells, each of which is represented by one of the above cells.

The relations between two superimposed tetrachoric divisions involve the determination of ten constants, four σ 's, σ_{11} , σ_{12} , σ_{21} , σ_{22} , and six R 's, $R_{11 \cdot 12}$, $R_{11 \cdot 21}$, $R_{11 \cdot 22}$, $R_{12 \cdot 21}$, $R_{12 \cdot 22}$, $R_{21 \cdot 22}$. The σ 's follow the example already given, the proper suffixes being attached. The value of $S_{11 \cdot 21}$ has already been given. The remaining five S 's are as follows:

$$S_{11 \cdot 12} = \frac{1}{N^2} \{ {}_aP_{11a}P_{12}n_{11} + {}_cP_{11c}P_{12}(n_{21} + n_{31}) - {}_bP_{11b}P_{12}n_{12} - {}_aP_{11a}P_{12}(n_{22} + n_{32}) \\ {}_bP_{11b}P_{12}n_{13} + {}_aP_{11a}P_{12}(n_{23} + n_{33}) \} \dots\dots\dots(91).$$

$$S_{11 \cdot 22} = \frac{1}{N^2} \{ {}_aP_{11a}P_{22}n_{11} - {}_cP_{11c}P_{22}n_{21} + {}_cP_{11c}P_{22}n_{31} - {}_bP_{11b}P_{22}n_{12} + {}_aP_{11a}P_{22}n_{22} \\ - {}_aP_{11a}P_{22}n_{32} + {}_bP_{11b}P_{22}n_{13} - {}_aP_{11a}P_{22}n_{23} + {}_aP_{11a}P_{22}n_{33} \} \dots\dots\dots(92).$$

$$S_{12 \cdot 21} = \frac{1}{N^2} \{ {}_aP_{12a}P_{21}n_{11} - {}_cP_{12c}P_{21}n_{21} + {}_cP_{12c}P_{21}n_{31} - {}_aP_{12a}P_{21}n_{12} + {}_cP_{12c}P_{21}n_{22} \\ - {}_cP_{12c}P_{21}n_{32} + {}_bP_{12b}P_{21}n_{13} - {}_aP_{12a}P_{21}n_{23} + {}_aP_{12a}P_{21}n_{33} \} \dots\dots\dots(93).$$

$$S_{12 \cdot 22} = \frac{1}{N^2} \{ {}_aP_{12a}P_{22}(n_{11} + n_{12}) - {}_cP_{12c}P_{22}(n_{21} + n_{22}) + {}_cP_{12c}P_{22}(n_{31} + n_{32}) \\ + {}_bP_{12b}P_{22}n_{13} - {}_aP_{12a}P_{22}n_{23} + {}_aP_{12a}P_{22}(n_{33}) \} \dots\dots\dots(94).$$

$$S_{21 \cdot 22} = \frac{1}{N^2} \{ {}_aP_{21a}P_{22}(n_{11} + n_{21}) + {}_cP_{21c}P_{22}n_{31} - {}_bP_{21b}P_{22}(n_{12} + n_{22}) - {}_aP_{21a}P_{22}n_{32} \\ + {}_bP_{21b}P_{22}(n_{13} + n_{23}) + {}_aP_{21a}P_{22}n_{33} \} \dots\dots\dots(95).$$

A more convenient form of the above for actual computation purposes will be found on page 120.

We may now by means of the P 's express the standard deviation r_p in a form consisting of sums of squares.

$$-(\Sigma C_{st}) dr = \Sigma \left(\frac{C_{st}}{\chi_{st}} \delta P_{st} \right) \dots\dots\dots(96),$$

$$\therefore (\Sigma C_{st})^2 \sigma_r^2 = \mathfrak{Z} \left\{ \Sigma \left(\frac{C_{st}}{\chi_{st}} \delta P_{st} \right) \right\}^2 = \Sigma \left(\frac{C_{st}}{\chi_{st}} \right)^2 S_{st, st} + 2 \Sigma \frac{C_{st} C_{s't'}}{\chi_{st} \chi_{s't'}} S_{st, s't'} \\ = \frac{C_{11}}{\chi_{11}} \left(\frac{C_{11}}{\chi_{11}} S_{11, 11} + \frac{C_{12}}{\chi_{12}} S_{11, 12} + \frac{C_{13}}{\chi_{13}} S_{11, 13} + \dots \right) \\ + \frac{C_{12}}{\chi_{12}} \left(\frac{C_{11}}{\chi_{11}} S_{11, 12} + \frac{C_{12}}{\chi_{12}} S_{12, 12} + \frac{C_{13}}{\chi_{13}} S_{12, 13} + \dots \right) + \text{etc.} \dots\dots\dots(97).$$

Now consider the S 's to be expanded in terms of the frequencies and pick out all the coefficients of the frequency n_{st} say. The coefficient of the n_{st} taken from $S_{lm, l'm'}$ say will be : $P_{lm} : P_{l'm'}$ in which the quadrant suffixes will be determined by the relative position of n_{st} to lm and $l'm'$. Let these undetermined P 's be denoted by p . We shall then have as the complete coefficient of n_{st}

$$\frac{C_{11}}{\chi_{11}} \left(\frac{C_{11}}{\chi_{11}} p_{11} \cdot p_{11} + \frac{C_{12}}{\chi_{12}} p_{11} \cdot p_{12} + \frac{C_{13}}{\chi_{13}} p_{11} \cdot p_{13} + \dots \right) \\ + \frac{C_{12}}{\chi_{12}} \left(\frac{C_{11}}{\chi_{11}} p_{11} \cdot p_{12} + \frac{C_{12}}{\chi_{12}} p_{12} \cdot p_{12} + \frac{C_{13}}{\chi_{13}} p_{12} \cdot p_{13} + \dots \right) + \text{etc.}$$

Since we are dealing throughout with the cell n_{st} the quadrantal suffix (i.e. the

a, b , etc.) for any l, m will be the same throughout. Hence we may write the complete coefficient of n_{st} as

$$\begin{aligned} & \frac{C_{11}}{\chi_{11}} p_{11} \left(\frac{C_{11}}{\chi_{11}} p_{11} + \frac{C_{12}}{\chi_{12}} p_{12} + \frac{C_{13}}{\chi_{13}} p_{13} + \dots \right) \\ & + \frac{C_{12}}{\chi_{12}} p_{12} \left(\frac{C_{11}}{\chi_{11}} p_{11} + \frac{C_{12}}{\chi_{12}} p_{12} + \frac{C_{13}}{\chi_{13}} p_{13} + \dots \right) + \text{etc.} \\ & = \left(\frac{C_{11}}{\chi_{11}} p_{11} + \frac{C_{12}}{\chi_{12}} p_{12} + \frac{C_{13}}{\chi_{13}} p_{13} + \dots \right)^2 \dots\dots\dots(98). \end{aligned}$$

In the case of an enneachoric table for example we have, ΣC_{st} being = 1,

$$\begin{aligned} \sigma_r^2 = & \left(-\frac{C_{11}}{\chi_{11}} {}_aP_{11} - \frac{C_{12}}{\chi_{12}} {}_aP_{12} - \frac{C_{21}}{\chi_{13}} {}_aP_{21} - \frac{C_{22}}{\chi_{22}} {}_aP_{22} \right)^2 n_{11} \\ & + \left(\frac{C_{11}}{\chi_{11}} {}_bP_{11} - \frac{C_{12}}{\chi_{12}} {}_bP_{12} + \frac{C_{21}}{\chi_{11}} {}_bP_{21} - \frac{C_{22}}{\chi_{22}} {}_bP_{22} \right)^2 n_{12} \\ & + \left(\frac{C_{11}}{\chi_{11}} {}_cP_{11} + \frac{C_{12}}{\chi_{12}} {}_cP_{12} + \frac{C_{21}}{\chi_{21}} {}_cP_{21} + \frac{C_{22}}{\chi_{22}} {}_cP_{22} \right)^2 n_{13} + \text{etc.} \dots\dots(99), \end{aligned}$$

the P 's being at once written down from the table on page 114.

Or more generally thus:

Since the P of any cell n_{lm} with reference to any cross point (st) is invariable it may be written generally as ${}_{lm}P_{st}$. This notation gives up the recognition of the equality of the P 's in any given quadrant but gains in generality. The quadrantal suffix and sign may be supplied by inspection. We have then the following lemma:

$$\begin{aligned} S_{st, s't'} &= \mathfrak{S}(\delta P_{st} \delta P_{s't'}) \\ &= {}_{11}P_{st \cdot 11} P_{s't'} n_{11} + {}_{21}P_{st \cdot 21} P_{s't'} n_{21} + \dots + {}_{12}P_{st \cdot 12} P_{s't'} n_{12} + \dots \\ &\dots\dots\dots \\ &= \sum_{lm} ({}_{lm}P_{st \cdot lm} P_{s't'} n_{lm}) \dots\dots\dots(100). \end{aligned}$$

The standard deviation of r_p may then be developed as follows:

$$\begin{aligned} (\Sigma C_{st}) dr &= \Sigma (C_{st} dr_{st}) \\ &= \Sigma \left(C_{st} \frac{\delta P_{st}}{\chi_{st}} \right) \dots\dots\dots(101). \end{aligned}$$

$$\begin{aligned} \therefore (\Sigma C_{st})^2 \sigma_r^2 &= \Sigma \left(\frac{C_{st}^2}{\chi_{st}^2} \delta P_{st} \right) + 2 \Sigma \left(\frac{C_{st} C_{s't'}}{\chi_{st} \chi_{s't'}} \delta P_{st} \delta P_{s't'} \right) \\ &= \sum_{lm} \left[\left\{ \sum_{st} \left(\frac{C_{st}}{\chi_{st}} \right)^2 {}_{lm}P_{st}^2 + 2 \sum_{st} \frac{C_{st} C_{s't'}}{\chi_{st} \chi_{s't'}} {}_{lm}P_{st \cdot lm} P_{s't'} \right\} \right] n_{lm} \\ &= \sum_{lm} \left\{ \left[\sum_{st} \left(\frac{C_{st}}{\chi_{st}} {}_{lm}P_{st} \right) \right]^2 n_{lm} \right\} \dots\dots\dots(102). \end{aligned}$$

More fully written this is

$$\begin{aligned} (\Sigma C_{st})^2 \sigma_r^2 &= \left(\frac{C_{11}}{\chi_{11}} {}_{11}P_{11} + \frac{C_{12}}{\chi_{12}} {}_{11}P_{12} + \frac{C_{13}}{\chi_{13}} {}_{11}P_{13} + \dots \right)^2 n_{11} \\ &+ \left(\frac{C_{11}}{\chi_{11}} {}_{12}P_{11} + \frac{C_{12}}{\chi_{12}} {}_{12}P_{12} + \frac{C_{13}}{\chi_{13}} {}_{12}P_{13} + \dots \right)^2 n_{12} + \dots \dots(103). \end{aligned}$$

§ 9. THE STANDARD DEVIATION OF POLYCHORIC r IN SPECIAL CASES.

The value of the standard deviation when $r = 0$ is of interest and may be got as follows.

Assuming $r = 0$ throughout:

$$A_{st} = \mathfrak{E}(k_t) = \frac{m_{\cdot t}}{N},$$

$$B_{st} = \mathfrak{E}(h_s) = \frac{m_{s \cdot}}{N},$$

and writing $m'_{s \cdot} = N - m_{s \cdot}$ and $m'_{\cdot t} = N - m_{\cdot t}$, then

$${}_a P_{st} = \frac{m_{s \cdot} m_{\cdot t}}{N},$$

$${}_b P_{st} = \frac{m'_{s \cdot} m_{\cdot t}}{N},$$

$${}_c P_{st} = \frac{m_{s \cdot} m'_{\cdot t}}{N},$$

$${}_d P_{st} = \frac{m'_{s \cdot} m'_{\cdot t}}{N}.$$

Substituting these values in the S 's we have after reduction

$$S_{11 \cdot 11} = \frac{1}{N^3} m_{1 \cdot} m_{\cdot 1} m'_{1 \cdot} m'_{\cdot 1} \dots\dots\dots (104),$$

$$S_{12 \cdot 12} = \frac{1}{N^3} m_{1 \cdot} m_{\cdot 2} m'_{1 \cdot} m'_{\cdot 2} \dots\dots\dots (105),$$

$$S_{21 \cdot 21} = \frac{1}{N^3} m_{2 \cdot} m_{\cdot 1} m'_{2 \cdot} m'_{\cdot 1} \dots\dots\dots (106),$$

$$S_{22 \cdot 22} = \frac{1}{N^3} m_{2 \cdot} m_{\cdot 2} m'_{2 \cdot} m'_{\cdot 2} \dots\dots\dots (107),$$

$$S_{11 \cdot 12} = \frac{1}{N^3} m_{1 \cdot} m_{\cdot 1} m'_{1 \cdot} m'_{\cdot 2} \dots\dots\dots (108),$$

$$S_{11 \cdot 21} = \frac{1}{N^3} m_{1 \cdot} m_{\cdot 1} m'_{2 \cdot} m'_{\cdot 1} \dots\dots\dots (109),$$

$$S_{12 \cdot 21} = S_{11 \cdot 22} = \frac{1}{N^3} m_{1 \cdot} m_{\cdot 1} m'_{2 \cdot} m'_{\cdot 2} \dots\dots\dots (110),$$

$$S_{12 \cdot 22} = \frac{1}{N^3} m_{1 \cdot} m_{\cdot 2} m'_{2 \cdot} m'_{\cdot 2} \dots\dots\dots (111),$$

$$S_{21 \cdot 22} = \frac{1}{N^3} m_{2 \cdot} m_{\cdot 1} m'_{2 \cdot} m'_{\cdot 2} \dots\dots\dots (112).$$

From these values we get the σ 's and R 's,

$$\sigma_{11} = \frac{1}{N^2 H_1 K_1} \sqrt{\frac{m_{1 \cdot} m_{\cdot 1} m'_{1 \cdot} m'_{\cdot 1}}{N}} \left(= \frac{1}{q_{11}} \right) \dots\dots\dots (113),$$

$$\sigma_{12} = \frac{1}{N^2 H_1 K_2} \sqrt{\frac{m_{1.} m_{.2} m'_{1.} m'_{.2}}{N}} \left(= \frac{1}{q_{12}} \right) \dots\dots\dots (114),$$

$$\sigma_{21} = \frac{1}{N^2 H_2 K_1} \sqrt{\frac{m_{2.} m_{.1} m'_{2.} m'_{.1}}{N}} \left(= \frac{1}{q_{21}} \right) \dots\dots\dots (115),$$

$$\sigma_{22} = \frac{1}{N^2 H_2 K_2} \sqrt{\frac{m_{2.} m_{.2} m'_{2.} m'_{.2}}{N}} \left(= \frac{1}{q_{22}} \right) \dots\dots\dots (116).$$

Since $m_{1.} m'_{.1} = m_{1.} (N - m_{1.}) = N \sigma_{m_1}^2$.

$$\sigma_{11} = \frac{1}{N \sqrt{N}} \frac{\sigma_{m_1} \cdot \sigma_{m_1}}{H_1 K_1},$$

and similarly for the others

$$R_{11 \cdot 12} = R_{21 \cdot 22} = \sqrt{\frac{m_{1.} m'_{.2}}{m'_{.1} m_{.2}}} \quad (= \epsilon) \dots\dots\dots (117),$$

$$R_{11 \cdot 21} = R_{12 \cdot 22} = \sqrt{\frac{m_{1.} m'_{2.}}{m'_{1.} m_{2.}}} \quad (= \epsilon') \dots\dots\dots (118),$$

$$R_{11 \cdot 22} = R_{12 \cdot 21} = \sqrt{\frac{m'_{1.} m'_{.1} m'_{2.} m'_{.2}}{m'_{1.} m'_{.1} m_{2.} m_{.2}}} \quad (= \epsilon \epsilon') \dots\dots\dots (119).$$

With these values we have

$$\Delta = \begin{vmatrix} 1 & q_{11} & q_{12} & q_{21} & q_{22} \\ q_{11} & 1 & \epsilon & \epsilon' & \epsilon \epsilon' \\ q_{12} & \epsilon & 1 & \epsilon \epsilon' & \epsilon' \\ q_{21} & \epsilon' & \epsilon \epsilon' & 1 & \epsilon \\ q_{22} & \epsilon \epsilon' & \epsilon' & \epsilon & 1 \end{vmatrix} \dots\dots\dots (120).$$

Δ_{0000} is symmetrical with respect to the centre of the square, hence*

$$= \begin{vmatrix} 1 + \epsilon \epsilon' & \epsilon + \epsilon' \\ \epsilon + \epsilon' & 1 + \epsilon \epsilon' \end{vmatrix} \cdot \begin{vmatrix} 1 - \epsilon \epsilon' & \epsilon - \epsilon' \\ \epsilon - \epsilon' & 1 - \epsilon \epsilon' \end{vmatrix} = \{(1 - \epsilon^2) (1 - \epsilon'^2)\}^2 = \frac{N^3 n_{22}}{m'_{1.} m'_{.1} m_{2.} m_{.2}} \dots\dots\dots (121).$$

The remaining minors are easily reduced and we have after reduction and substitution

$$\frac{\Delta_{0011}}{\sigma_{11}} = - (1 - \epsilon^2) (1 - \epsilon'^2) N^5 \frac{H_1}{m_{1.} m'_{1.}} \left(H_1 - \frac{m_{1.}}{m_{2.}} H_2 \right) \frac{K_1}{m_{.1} m'_{.1}} \left(K_1 - \frac{m_{.1}}{m_{.2}} K_2 \right) \dots\dots\dots (122),$$

$$\frac{\Delta_{0012}}{\sigma_{12}} = (1 - \epsilon^2) (1 - \epsilon'^2) N^5 \frac{H_1}{m_{1.} m'_{1.}} \left(H_1 - \frac{m_{1.}}{m_{2.}} H_2 \right) \frac{K_2}{m_{.2} m'_{.2}} \left(\frac{m'_{.2}}{m'_{.1}} K_1 - K_2 \right) \dots\dots\dots (123),$$

$$\frac{\Delta_{0021}}{\sigma_{21}} = (1 - \epsilon^2) (1 - \epsilon'^2) \frac{H_2}{m_{2.} m'_{2.}} \left(\frac{m'_{2.}}{m'_{1.}} H_1 - H_2 \right) \frac{K_1}{m_{.1} m'_{.1}} \left(K_1 - \frac{m_{.1}}{m_{.2}} K_2 \right) \dots\dots\dots (124),$$

$$\frac{\Delta_{0022}}{\sigma_{22}} = - (1 - \epsilon^2) (1 - \epsilon'^2) \frac{H_2}{m_{2.} m'_{2.}} \left(\frac{m'_{2.}}{m'_{1.}} H_1 - H_2 \right) \frac{K_2}{m_{.2} m'_{.2}} \left(\frac{m'_{.2}}{m'_{.1}} K_1 - K_2 \right) \dots\dots\dots (125).$$

* See Scott and Mathews, *Theory of Determinants* (2nd ed.), p. 89.

Summing these four latter expressions we have

$$\begin{aligned}\Delta - \Delta_{0000} &= - (1 - \epsilon^2) (1 - \epsilon'^2) N^5 \left\{ \frac{H_1^2}{m_{1.} m'_{1.}} - \frac{2H_1 H_2}{m_{2.} m'_{1.}} + \frac{H_2^2}{m_{2.} m'_{2.}} \right\} \\ &\quad \times \left\{ \frac{K_1^2}{m_{1.} m'_{1.}} - \frac{2K_1 K_2}{m_{2.} m'_{1.}} + \frac{K_2^2}{m_{2.} m'_{2.}} \right\} \dots\dots\dots(126), \\ \therefore \sigma_r^2 &= \frac{\Delta_{0000}}{\Delta_{0000} - \Delta} \\ &= \frac{n_{22}/m'_{1.} m'_{1.} m_{2.} m_{2.}}{N^2 \left\{ \frac{H_1^2}{m_{1.} m'_{1.}} - \frac{2H_1 H_2}{m_{2.} m'_{1.}} + \frac{H_2^2}{m_{2.} m'_{2.}} \right\} \times \left\{ \frac{K_1^2}{m_{1.} m'_{1.}} - \frac{2K_1 K_2}{m_{2.} m'_{1.}} + \frac{K_2^2}{m_{2.} m'_{2.}} \right\}} \\ &= \frac{n_{22}}{N^2 \left\{ H_1^2 \frac{m_{2.}}{m_{1.}} - 2H_1 H_2 + H_2^2 \frac{m'_{1.}}{m'_{2.}} \right\} \times \left\{ K_1^2 \frac{m_{2.}}{m_{1.}} - 2K_1 K_2 + K_1^2 \frac{m'_{1.}}{m'_{2.}} \right\}} \dots\dots\dots(127).\end{aligned}$$

When the table is symmetrically divided in both categories and $r = 0$ we have $m_{2.} = m'_{1.}$, $H_1 = H_2$, $K_1 = K_2$, etc. and the above reduces to

$$\sigma_r^2 = \frac{n_{22}}{N^2 H^2 K^2 \cdot 4 \left(\frac{m_{2.}}{m_{1.}} - 1 \right) \left(\frac{m_{2.}}{m_{1.}} - 1 \right)} = \frac{n_{22}}{4N^2 H^2 K^2 \frac{n_{2.}}{n_{1.}} \cdot \frac{n_{2.}}{n_{1.}}} = \frac{n_{11}}{4N^2 H^2 K^2}$$

and

$$\sigma_r = \frac{\sqrt{n_{11}}}{2NHK} \dots\dots\dots(128).$$

§ 10. COMPARISON OF THE STANDARD DEVIATIONS OF POLYCHORIC r AND ENNEACHORIC r .

We may now compare the standard deviations of r_e and r_p .

$$-(\chi_{11} - \chi_{12} - \chi_{21} + \chi_{22}) \delta r_e = \delta P_{11} - \delta P_{12} - \delta P_{21} + \delta P_{22} \dots\dots\dots(129),$$

$$-(C_{11} + C_{12} + C_{21} + C_{22}) \delta r_p = \frac{C_{11}}{\chi_{11}} \delta P_{11} - \frac{C_{12}}{\chi_{12}} \delta P_{12} - \frac{C_{21}}{\chi_{21}} \delta P_{21} + \frac{C_{22}}{\chi_{22}} \delta P_{22} \dots\dots\dots(130),$$

from which it appears as before (p. 108) that the enneachoric r is equivalent to a polychoric r in which the weights of the r 's are χ_{11} , $-\chi_{12}$, $-\chi_{21}$, χ_{22} , i.e.

$$r_e = \frac{\chi_{11} r_{11} - \chi_{12} r_{12} - \chi_{21} r_{21} + \chi_{22} r_{22}}{\chi_{11} - \chi_{12} - \chi_{21} + \chi_{22}} \dots\dots\dots(131).$$

Hence also the standard deviation of the enneachoric r may be written

$$\sigma_r^2 = (-aP_{11} + aP_{12} + aP_{21} - aP_{22})^2 n_{11} + \text{etc.} \dots\dots\dots(132).$$

Upon expansion this reduces to

$$\begin{aligned}\sigma_r^2 &= \left[-aP_{11} + N(A_{11} + B_{11} - 1) + aP_{12} - N(A_{12} + B_{12} - 1) \right]^2 n_{11} + \text{etc.} \\ &= \left[N\{(A_{22} - A_{21}) - (A_{12} - A_{11}) + (B_{22} - B_{12})\} - (B_{21} - B_{11}) \right]^2 n_{11} + \text{etc.} \\ &= \left[N\left\{ a_{22} - a_{12} + \beta_{22} - \beta_{21} - \frac{P_{11} - P_{12} - P_{21} + P_{22}}{N} \right\} \right]^2 n_{11} + \text{etc.} \dots\dots\dots(133)\end{aligned}$$

(using α , β in the sense of p. 99), which is identical with the corresponding coefficient in σ_r^2 as given in equation (30).

§ 11. FORMULAE FOR COMPUTATION.

The forms found for S are not convenient for computation. The following have been found more expeditious. Various other formulae are also collected for reference :

$$\chi_{st} = Nz(h_s, k_t, r) = \frac{N}{2\pi\sqrt{1-r^2}} e^{-\frac{h_s^2 + k_t^2 - 2rh_s k_t}{2(1-r^2)}} \dots\dots\dots(134),$$

$$A_{st} = \mathfrak{E}\left(\frac{k_t - rh_s}{\sqrt{1-r^2}}\right),$$

$$B_{st} = \mathfrak{E}\left(\frac{h_s - rk_t}{\sqrt{1-r^2}}\right),$$

$$\Pi_{st} = A_{st} + B_{st} - 1,$$

$$P_{st} = A_{st}m_{s\cdot} + B_{st}m_{\cdot t} - m_{st},$$

when no quadrantal suffix is used a is understood.

$$S_{11\cdot 11} = \Pi_{11}^2 a_{11} + A_{11}^2 c_{11} + B_{11}^2 b_{11} - \frac{P_{11}^2}{N} \dots\dots\dots(135),$$

and similarly for $S_{12\cdot 12}$, etc.,

$$\begin{aligned} S_{11\cdot 12} = & \Pi_{11}\Pi_{12}n_{11} + B_{11}B_{12}(n_{21} + n_{31}) \\ & + A_{11}\Pi_{12}n_{12} - \frac{P_{11}P_{12}}{N} \dots\dots\dots(136), \\ & + A_{11}A_{12}n_{13} \end{aligned}$$

$$\begin{aligned} S_{11\cdot 21} = & \Pi_{11}\Pi_{21}n_{11} + B_{11}\Pi_{21}n_{21} + B_{11}B_{21}n_{31} \\ & + A_{11}\Pi_{22}n_{12} - \frac{P_{11}P_{21}}{N} \dots\dots\dots(137), \\ & + A_{11}A_{22}n_{13} \end{aligned}$$

$$\begin{aligned} S_{11\cdot 22} = & \Pi_{11}\Pi_{22}n_{11} + B_{11}\Pi_{22}n_{21} + B_{11}B_{22}n_{31} \\ & + A_{11}\Pi_{22}n_{12} - \frac{P_{11}P_{22}}{N} \dots\dots\dots(138), \\ & + A_{11}A_{22}n_{13} \end{aligned}$$

$$\begin{aligned} S_{12\cdot 21} = & \Pi_{12}\Pi_{21}n_{11} + B_{12}\Pi_{21}n_{21} + B_{12}B_{21}n_{31} \\ & + \Pi_{12}A_{21}n_{12} + B_{12}A_{21}n_{22} - \frac{P_{12}P_{21}}{N} \dots\dots\dots(139), \\ & + A_{12}A_{21}n_{13} \end{aligned}$$

$$\begin{aligned} S_{12\cdot 22} = & \Pi_{12}\Pi_{22}(n_{11} + n_{12}) + B_{12}\Pi_{22}(n_{21} + n_{22}) + B_{12}B_{22}(n_{31} + n_{32}) \\ & + A_{12}A_{22}n_{13} - \frac{P_{12}P_{22}}{N} \dots\dots\dots(140), \end{aligned}$$

$$\begin{aligned} S_{21\cdot 22} = & \Pi_{21}\Pi_{22}(n_{11} + n_{21}) + B_{21}B_{22}n_{31} \\ & + A_{21}\Pi_{22}(n_{12} + n_{22}) - \frac{P_{21}P_{22}}{N} \dots\dots\dots(141), \\ & + A_{21}A_{22}(n_{13} + n_{23}) \end{aligned}$$

$$\chi_{st}^2 \sigma_{st}^2 = S_{st\cdot st\cdot}$$

$$\therefore \sigma_{st} = \frac{\sqrt{S_{st\cdot st\cdot}}}{\chi_{st}}.$$

$$\chi_{st}\chi_{s't'}\sigma_{st}\sigma_{s't'}R_{st\cdot s't'} = S_{st\cdot s't'} \dots\dots\dots(142),$$

$$\begin{aligned} \therefore R_{st\cdot s't'} &= \frac{S_{st\cdot s't'}}{\chi_{st}\sigma_{st}\chi_{s't'}\sigma_{s't'}} \\ &= \frac{S_{st\cdot s't'}}{\sqrt{S_{st\cdot st\cdot}S_{s't'\cdot s't'}}} \dots\dots\dots(143). \end{aligned}$$

In place of calculating σ and R it will be found easier to employ the S 's directly by writing the equations for C in the form

$$\left. \begin{aligned} C_{11} \frac{S_{11 \cdot 11}}{\chi_{11}^2} + C_{12} \frac{S_{11 \cdot 12}}{\chi_{11} \chi_{12}} + C_{13} \frac{S_{11 \cdot 13}}{\chi_{11} \chi_{13}} + \dots = \lambda \\ C_{11} \frac{S_{11 \cdot 12}}{\chi_{11} \chi_{12}} + C_{12} \frac{S_{12 \cdot 12}}{\chi_{12}^2} + C_{13} \frac{S_{12 \cdot 13}}{\chi_{12} \chi_{13}} + \dots = \lambda \end{aligned} \right\} + \text{etc.(144)}.$$

Eliminate the λ by subtracting one equation from each of the others; put $C_{11} = 1$ and solve by successive elimination for the remaining C 's. This is preferable to using the determinant as it is at least no more laborious and lends itself to various checks for accuracy. The λ should be determined from each of the equations as a further check. Then we have

$$\sigma_r^2 = \frac{\lambda}{\sum C} \dots\dots\dots(145).$$

By putting $C_{11} = \chi_{11}$, $C_{12} = -\chi_{12}$, $C_{21} = -\chi_{21}$, $C_{22} = \chi_{22}$, we may derive the enneachoric standard deviation from the polychoric r in a form convenient for computation in terms of the S 's,

$$(\chi_{11} - \chi_{12} - \chi_{21} + \chi_{22})^2 \sigma_{r_e}^2 = S_{11 \cdot 11} + S_{12 \cdot 12} + S_{21 \cdot 21} + S_{22 \cdot 22} + 2(S_{11 \cdot 22} + S_{12 \cdot 21}) - 2(S_{11 \cdot 12} + S_{11 \cdot 21} + S_{12 \cdot 22} + S_{21 \cdot 22}) \dots\dots(146).$$

§ 12. COMPARATIVE RESULTS OF VARIOUS METHODS OF FINDING r FROM A 3×3 TABLE.

In testing the methods developed in this paper upon actual material it was thought desirable to try them side by side with all the other methods of finding the correlation coefficient so that some indication could be got of their comparative accuracy. Each of the tables was therefore dealt with by nine methods which are indicated in § 13. These tables were selected at the beginning of the investigation, and had the course which the research has taken been foreseen probably a different selection might have been made. Two of them, I and III, are normal tables with an arbitrary population of 1000. In Table I the frequencies have been taken to the nearest integer and in III to the nearest two places of decimals, so that any irregularity in them is due to the roughness of the approximation to the true figures. In the r_ψ , we have an additional lack of approximation in taking r_ψ from the curve* for determining r_ψ and also in r_ψ , r_ϕ and r_e from finding the class index correlation from a small number of marginal groups. In II and IV we have actual samples.

A rough test of the value of the various methods may be made by finding the mean square deviation of the calculated from the "observed" value of r , each constituent being merely weighted with its total frequency, regarding the product moment values of r as the "observed" value.

Thus let $n_1, n_2 =$ total frequency in Tables I, II; $R_1, R_2 =$ product moment value of the correlation coefficient in Tables I, II; $r_1, r_2 =$ correlation coefficient calculated by one of the methods, then writing

$$(n_1 + n_2 + \dots) \Sigma^2 = n_1 (R_1 - r_1)^2 + n_2 (R_2 - r_2)^2 + \dots,$$

we shall have in Σ^2 a measure of the goodness of the various methods. This gives the following values of Σ^2 .

* *Tables for Statisticians and Biometricians*, p. lvii and p. 65.

Mean weighted square deviation of calculated from "observed" or product moment values of r .

	Σ^2	Σ^2 (omitting H)
Mean contingency r_ψ	·00138	·00102
Mean square contingency r_ϕ	·00089	·00060
Enneachoric r r_e	·00364	·00036
Polychoric r r_p	·00004	·00002
Tetrachoric r r_t	·00018	·00016
Mean tetrachoric r r_m	·00005	·00003
Mean weighted tetrachoric r r_w	·00002	·00002
Three row η from mean dispersion* ... η_m	·00020	·00019
Three row η from "individual" dispersion η_n	·00151	·00144
Marginal centroids r_c	·00215	·00255

I have given the value of Σ^2 including and omitting Table H, which gives very anomalous results, as yet unexplained. Broadly the best results are given by r_p , r_m and r_w , and, Table H aside, the best result is by r_p . In the case of r_ψ the results are not quite satisfactory. The figure given was arrived at by taking the mean of the raw figure from the curve and the same corrected for broad categories as suggested in *Tables for Statisticians and Biometricians*. An attempt was made to find an empirical formula which would give better results with the tables here described, but the result was not worthy of record. With three row η , although strictly the method is quite inapplicable to 3×3 tables, it may be useful to notice that when so applied the best results on the whole were got from assuming the distribution to be homoscedastic and using the mean dispersion of the arrays. This was largely due to several of the tables being divided so that some of the arrays contained very small frequencies which had therefore large probable errors, giving an undue effect on the result when squared. When such small frequencies are avoided the results appear to be about equally good. Of course our theory fails, as we have already pointed out, when any cell frequency is of the same order as its variation.

Comparing the probable errors of r_m , r_w and r_p (tabulated for convenience in the Appendix on page 133) it will be seen that on the whole they are in descending order of magnitude. They differ very little from each other and, considering the labour involved in finding r_p , r_m , would in most cases give a result with a sufficiently low probable error.

The method of marginal centroids as already known is unsuited for tables with so few categories.

An interesting and important relation which is not shown in the tables of numerical results (§ 13) is the degree of correlation between r_{11} , r_{12} , etc., viz. $R_{11 \cdot 12}$, $R_{11 \cdot 21}$, etc. These are collected in the table on p. 123.

All the enneachoric tables are arranged so that reading from r_{11} to the right, and downwards, r is positive so that the values R may be compared among each other. It will be seen on examination that $R_{11 \cdot 12}$, $R_{11 \cdot 21}$, $R_{12 \cdot 22}$, $R_{21 \cdot 22}$ are on the whole greater than $R_{11 \cdot 22}$ and $R_{12 \cdot 21}$ and of the two latter $R_{12 \cdot 21}$ is usually the greater.

* See § 13, 8.

	$R_{11 \cdot 12}$	$R_{11 \cdot 21}$	$R_{12 \cdot 22}$	$R_{21 \cdot 22}$	$R_{11 \cdot 22}$	$R_{12 \cdot 21}$
A	·3418	·3041	·3196	·3450	·0608	·1466
B	·6977	·6040	·6132	·7050	·4048	·5231
C	·4880	·4550	·4751	·4821	·1872	·2614
D	·5100	·5678	·6038	·4907	·2620	·3221
E	·2180	·4953	·5378	·1813	·0830	·1217
F	·4252	·4422	·4133	·4307	·1209	·2392
G	·6633	·6723	·6565	·6701	·3414	·4769
H	·4395	·6592	·6567	·4367	·2391	·3219
K	·3842	·3488	·3656	·3882	·1124	·1632
L	·5014	·4831	·4798	·5162	·2276	·2611
M	·5282	·2820	·2961	·5213	·0865	·2128
Pairs of brothers	·8203	·8203	·8732	·8732	·8837	·8756

With regard to the computation of r_p it will be seen from example appended to Table A that the amount of labour involved in dealing even with a 3×3 table is considerable and will rapidly increase with the number of cells, and it is very desirable that some short method of approximating to the weights (C 's) of the r 's be devised. For the present it may be of interest to give here the C 's for the various tables used.

	C_{11}	C_{12}	C_{21}	C_{22}
A	1	·51333	·38587	·71892
B	1	·26872	·19654	·55240
C	1	·27000	·26993	·34266
D	1	·65737	·15786	·85445
E	1	3·79397	·16186	3·81496
F	1	·56452	·43667	·72594
G	1	−·02973	−·13835	·69769
H	1	·72919	·33838	·79708
K	1	·59665	·48857	·66355
L	1	·39951	·34626	·27119
M	1	·36146	·38718	·74467
Pairs of brothers	1	−·26399	−·26399	·82793

The case of Table G, with negative weight for r_{12} and r_{21} is suggestive and needs further study. The table has the characteristic that the mean is in n_{11} and the marginal frequencies are decreasing in magnitude and nearly equal in both sets of categories. The table "Pairs of brothers" which is accompanied by similar weights is taken from *Biometrika*, vol. III, 1904, p. 182, and is given below. It compares the athletic capacities of pairs of brothers.

First Brother

Second Brother		Athletic	Betwixt	Non-athletic	Total
	Athletic	906	20	140	1066
	Betwixt	20	76	9	105
	Non-athletic	140	9	370	519
	Total	1066	105	519	1690

$$r_{11} = .8046 \pm .0126,$$

$$r_{12} = .7190 \pm .0162,$$

$$r_{21} = .7190 \pm .0162,$$

$$r_{22} = .8028 \pm .0132.$$

$$C_{11} = 1, \quad C_{12} = -.26399, \quad C_{21} = -.26399, \quad C_{22} = .82793.$$

$$r_p = .8382 \pm .0122.$$

These negative weights require further investigation, particularly the conditions for the existence of zero weights, but it is clear that certain divisions are to be avoided in determining r from a 3×3 table.

On the whole C_{11} , C_{22} , C_{12} , C_{21} are in descending order of magnitude.

§ 13. PRÉCIS OF THE METHODS OF FINDING THE COEFFICIENT OF CORRELATION.

1. r_ψ . Mean contingency, corrected for class index correlation.
2. r_ϕ . Mean square contingency, corrected for class index correlation and where necessary for the number of cells.
3. r_e . By selecting the central cell, the method first described in this paper. As its use treats any table as virtually 3×3 , it may be called enneachoric r .
4. r_p . By weighting the r 's so that the P.E. shall be a minimum, the second method described in this paper. As it is applicable to tables of any size it may be called polychoric r .
5. $r_{11}, r_{12}, r_{21}, r_{22}$. Tetrachoric r of the various quadrants. The probable errors were calculated by the complete formula (P.E.) and also by the approximate method (A.P.E.). (*Tables for Statisticians and Biometricians*, p. xl.)
6. r_m . The unweighted mean of $r_{11}, r_{12}, r_{21}, r_{22}$.
7. r_w . The mean of the r_{11} , etc., weighted by the reciprocals of the squares of their standard deviation.
8. $\eta_{k_1}, \eta_{k_2}, \eta_{h_1}, \eta_{h_2}$. Three row η calculated from each of the dividing planes as planes of reference with a class index correction on the foot of the columns. Since the standard deviation may be found in this case from the individual arrays or, assuming the distribution sufficiently homoscedastic, may be given the mean value $\sigma\sqrt{1-\eta^2}$, I have used both methods for the purpose of comparison. These are distinguished by the headings "individual dispersion" and "mean dispersion" respectively*.

9. r_c . By marginal centroids.

The probable error of r_m and r_w was obtained as follows:

Let the correlation coefficients	$r_{11}, r_{12}, \dots,$
have the S.D.'s	$\sigma_{11}, \sigma_{12}, \dots,$
and the weights	$t_{12}, t_{12}, \dots$

* The probable error of Biserial (or three row η) has now been given (*Biometrika*, Vol. ix, part iv), but too late for use in the present paper.

$$\text{Then } r_t = \frac{\Sigma t_{11} r_{11}}{\Sigma t_{11}}, \quad (dr_t)^2 = \frac{\Sigma t_{11}^2 (dr_{11})^2 + 2 \Sigma t_{11} t_{12} dr_{11} dr_{12}}{(\Sigma t_{11})^2},$$

$$\sigma_{r_t}^2 = \frac{\Sigma t_{11}^2 \sigma_{11}^2 + 2 \Sigma t_{11} t_{12} \sigma_{11} \sigma_{12} R_{11 \cdot 12}}{(\Sigma t_{11})^2} \dots\dots\dots (147).$$

When $t_{11} = t_{12} = \dots = 1$ we have the mean r , r_m , and if there are l r 's

$$\sigma_{r_m}^2 = \frac{\Sigma \sigma_{11}^2 + 2 \Sigma \sigma_{11} \sigma_{12} R_{11 \cdot 12}}{l^2} \dots\dots\dots (148),$$

which for convenient computation may be written

$$= \frac{\Sigma \frac{S_{11 \cdot 11}}{\chi_{11}^2} + 2 \Sigma \frac{S_{11 \cdot 12}}{\chi_{11} \chi_{12}}}{l^2}.$$

In finding the mean weighted r we may regard r_{11} as the mean of t_{11} uncorrelated values of r of equal weight each having the s.d. σ_0 . Hence $\sigma_{11} = \frac{\sigma_0}{\sqrt{t_{11}}}$ and $t_{11} = \frac{\sigma_0^2}{\sigma_{11}^2}$, i.e. the weights are proportional to the reciprocals of the squares of the s.d.'s. Putting this value in (147) we have

$$\sigma_{r_w}^2 = \frac{\Sigma \left(\frac{1}{\sigma_{11}^2} \right) + 2 \Sigma \frac{R_{11 \cdot 12}}{\sigma_{11} \sigma_{12}}}{\Sigma \left(\frac{1}{\sigma_{11}^2} \right)^2} \dots\dots\dots (149).$$

§ 14. DETAILS OF TABLES AND SUMMARY OF NUMERICAL RESULTS.

I. The first table examined was taken from Pearson and Heron's paper "On Theories of Association," *Biometrika*, vol. ix, p. 220, Table XIV, and is a Gaussian surface for $r = .5$ adjusted to give whole units in the cells.

	1	2	3	4	5 + 6	7	8	Total
1	7	20	5	2	—	—	—	34
2	21	145	79	36	10	9	1	301
3	6	94	85	54	19	22	4	284
4	2	32	39	31	12	17	4	137
5 + 6	—	18	28	25	11	18	5	105
7	—	11	22	24	12	22	7	98
8	—	2	6	8	5	13	7	41
Total	36	322	264	180	69	101	28	1000

The frequency in heavy type contains the mean of the surface.

A. Table I divided so that the mean falls in cell n_{22} .

	1 + 2	3 + 4	5 + 6 + 7 + 8	Total
1 + 2	193	122	20	335
3 + 4	134	209	78	421
5 + 6 + 7 + 8	31	113	100	244
Total	358	444	198	1000

$r_\psi = .482$		
$r_\phi = .4840 \pm .0170$		A.P.E.
$r_p = .50346 \pm .02094$	$\left. \begin{array}{l} r_{11} = .498 \pm .02872 \quad (.0290) \\ r_{12} = .510 \pm .03210 \quad (.0303) \\ r_{21} = .508 \pm .03505 \quad (.0321) \\ r_{22} = .504 \pm .03259 \quad (.0340) \end{array} \right\}$	
$r_e = .48594 \pm .04918$		
$r_m = .5050 \pm .0246$		
$r_w = .5045 \pm .0211$		
$r_e = .5145$		
	Mean dispersion	Individual dispersion
η_{k_1}	.5031	.4950
η_{k_2}	.5057	.4975
η_{h_1}	.5045	.4955
η_{h_2}	.5058	.4942

I here insert as an illustration of the new method the constants required in finding r_p for the above table, and the calculation of $S_{12 \cdot 21}$, P_{11} and Π_{11} , and the equations to find the O 's.

Table A	$h_1 k_1$	$h_1 k_2$	$h_2 k_1$	$h_2 k_2$
r	.498	.510	.508	.504
h	-.36381	-.36381	.84879	.84879
rh	-.18118	-.18554	.43117	.42779
Eh	.3733945	.3733945	.2782707	.2782707
k	-.42615	.69349	-.42615	.69349
rk	-.21222	.35368	-.21648	.34952
$E k$	.3643145	.3136735	.3643145	.3136735
$h - rk$	-.15159	-.71749	1.06527	.49927
$\frac{h - rk}{\sqrt{1 - r^2}}$	-.1748086	-.8341218	1.236735	.5780570
$E \frac{h - rk}{\sqrt{1 - r^2}}$	.3928931	.2817266	.1856865	.3375594
$B = E \frac{h - rk}{\sqrt{1 - r^2}}$	.4306151	.2021063	.8919072	.7183872
$k - rh$	-.24497	.87903	-.85732	.26570
$\frac{k - rh}{\sqrt{1 - r^2}}$	-.2824913	1.021920	-.9953132	.3076287
$E \frac{k - rh}{\sqrt{1 - r^2}}$	.3833377	.2366676	.2431048	.3805048
$A = E \frac{k - rh}{\sqrt{1 - r^2}}$	.3887835	.8465906	.1597920	.6208174
χ	165.0602	102.7353	78.53720	122.5923
Π	-.1806014	.0486969	.0516992	.3392046
P	90.44050	128.8716	111.9422	383.0020
$S_{11}/\chi : \chi : *$	.001812696	.000692554	.0006728030	.0001333663
$S_{12}/\chi : \chi :$	—	.002265285	.0003626002	.0007349840
$S_{21}/\chi : \chi :$	—	—	.002700292	.0008662932
$S_{22}/\chi : \chi :$	—	—	—	.002335123
$G :$	1	.51333	.3859	.71892

* The suffix : indicates that appropriate suffix is to be taken from the column.

Calculation of $S_{12 \cdot 21}$.

Π_{12}	·0486969	
Π_{21}	·0516992	
n_{11}	193	·485895
B_{12}	·2021063	
Π_{21}	·0516992	
n_{21}	122	1·274745
B_{12}	·2021063	
B_{21}	·8919072	
n_{21}	20	3·605201
Π_{12}	·0486969	
A_{21}	·1597920	
n_{12}	134	1·042704

B_{12}	·2021063	
A_{21}	·1597920	
n_{22}	209	6·749650
A_{12}	·8465906	
A_{21}	·1597920	
n_{13}	31	4·193630
		17·351825
P_{12}	128·8716	
P_{21}	111·9422	
$\div N$	1000	14·426170
		2·925655
χ_{12}	102·7353	
χ_{21}	78·5370	$\div 8068·543$
		= ·0003626002

Calculation of P_{11} .

A_{11}	·3887835	
$m_{1 \cdot}$	358	139·1845
B_{11}	·4306151	
$m_{\cdot 1}$	335	144·2560
		283·4405
m_{11}		193
$P_{11} =$		90·4405

Calculation of Π_{11} .

A_{11}	·3887835
B_{11}	·4306151
	·8193986
	1
$\Pi_{11} = -$	·1806014

Equations to find C ,

$$\begin{aligned} \cdot 001812696C_{11} + \cdot 000692554C_{12} + \cdot 000672803C_{21} + \cdot 000133363C_{22} &= \lambda, \\ \cdot 000692554C_{11} + \cdot 002265285C_{12} + \cdot 000362600C_{21} + \cdot 000734984C_{22} &= \lambda, \\ \cdot 000672803C_{11} + \cdot 000362600C_{12} + \cdot 002700292C_{21} + \cdot 000866293C_{22} &= \lambda, \\ \cdot 000133366C_{11} + \cdot 000734984C_{12} + \cdot 000866293C_{21} + \cdot 002335123C_{22} &= \lambda. \end{aligned}$$

The solution of these equations gives the first row of figures in the C table on page 123.

B. Table I divided so that the mean falls in cell n_{11} .

	1 + 2 + 3	4	5 + 6 + 7 + 8	Total
1 + 2 + 3	462	92	65	619
4	73	31	33	137
5 + 6 + 7 + 8	87	57	100	244
Total	622	180	198	1000

$r_\psi = .533$ $r_\phi = .510 \pm .0161$ $r_p = .5007 \pm .0250$ $r_e = .5073 \pm .1437$ $r_m = .5010 \pm .0254$ $r_w = .5008 \pm .0253$ $r_c = .5445$

$r_{11} = .499 \pm .028$ $r_{12} = .501 \pm .030$ $r_{21} = .500 \pm .031$ $r_{22} = .504 \pm .033$

A.P.E.
(.028)
(.030)
(.032)
(.034)

	Mean dispersion	Individual dispersion
η_{k_1}	.4921	.4873
η_{k_2}	.4917	.4763
η_{h_1}	.4858	.4802
η_{h_2}	.4881	.4671

C. Table I divided so that the mean falls in cell n_{11} .

	1 + 2 + 3	4 + 5 + 6	7 + 8	Total
1 + 2 + 3	462	121	36	619
4 + 5 + 6	119	79	44	242
7 + 8	41	49	49	139
Total	622	249	129	1000

$r_\psi = .537$ $r_\phi = .5183 \pm .0202$ $r_p = .4988 \pm .0241$ $r_e = .5055 \pm .0652$ $r_m = .4985 \pm .0253$ $r_w = .4973 \pm .0311$ $r_c = .5480$

$r_{11} = .499 \pm .028$ $r_{12} = .490 \pm .035$ $r_{21} = .505 \pm .035$ $r_{22} = .500 \pm .039$

A.P.E.
(.028)
(.035)
(.035)
(.043)

	Mean dispersion	Individual dispersion
η_{k_1}	.4895	.5041
η_{k_2}	.4934	.4615
η_{h_1}	.4846	.4730
η_{h_2}	.4947	.4496

D. Table I divided so that the mean falls in cell n_{12} .

	1 + 2 + 3	4	5 + 6 + 7 + 8	Total
1 + 2	277	38	20	335
3	185	54	45	284
4 + 5 + 6 + 7 + 8	160	88	133	381
Total	622	180	198	1000

$$\begin{aligned}
 r_\psi &= .524 \\
 r_\phi &= .482 \pm .0631 \\
 r_p &= .5018 \pm .0235 \\
 r_e &= .5173 \pm .2330 \\
 r_m &= .5030 \pm .0239 \\
 r_w &= .5025 \pm .0236 \\
 r_c &= .5098
 \end{aligned}
 \left. \begin{array}{l} \\ \\ \\ \\ \\ \end{array} \right\} \begin{array}{ll} r_{11} = .501 \pm .030 & (.030) \\ r_{12} = .499 \pm .028 & (.029) \\ r_{21} = .508 \pm .035 & (.032) \\ r_{22} = .504 \pm .031 & (.032) \end{array} \quad \text{A.P.E.}$$

	Mean dispersion	Individual dispersion
η_{k_1}	.5211	.4798
η_{k_2}	.4885	.4907
η_{h_1}	.4956	.4709
η_{h_2}	.5115	.4553

E. Table I divided so that the frequency of n_{22} differs very little from the frequency of a table with the same marginal frequencies but of zero correlation. The mean is in cell n_{23} .

	1 + 2	3	4 + 5 + 6 + 7 + 8	Total
1	27	5	2	34
2	166	79	56	301
3 + 4 + 5 + 6 + 7 + 8	165	180	320	665
Total	358	264	378	1000

Here $\frac{301 \times 264}{1000} = 79.464$ so that the constant term in the equation for r is a small quantity and any error of sampling will have an excessive weight. It will be found as one might expect that the P.E. of r_e is very large.

The very large value of η_{h_2} is due to the column having the marginal total 378, for the frequency 2 in it is the nearest whole number to a true value and being so small, a small absolute difference makes a large fractional value resulting in a large difference between the true and apparent standard deviations of this particular array. Actually the method applied is inapplicable to a frequency of this order.

$$\begin{aligned}
 r_\psi &= .502 \\
 r_\phi &= .4827 \pm .0246 \\
 r_p &= .4991 \pm .0245 \\
 r_e &= .4658 \pm .4371 \\
 r_m &= .4995 \pm .0327 \\
 r_w &= .4995 \pm .0247 \\
 r_c &= .5065
 \end{aligned}
 \left. \begin{array}{l} \\ \\ \\ \\ \\ \end{array} \right\} \begin{array}{ll} r_{11} = .500 \pm .056 & (.054) \\ r_{12} = .498 \pm .029 & (.029) \\ r_{21} = .500 \pm .072 & (.054) \\ r_{22} = .500 \pm .030 & (.029) \end{array} \quad \text{A.P.E.}$$

	Mean dispersion	Individual dispersion
η_{k_1}	.4862	.4906
η_{k_2}	.5169	.4991
η_{h_1}	.4950	.5022
η_{h_2}	.4966	.7150

II. The second table examined was taken from Macdonell's paper "On Criminal Anthropometry," *Biometrika*, vol. I, p. 216. The original table is too extensive to be given here, but may be found in *loc. cit.* The horizontal categories are the heights of 3000 criminals in feet and inches, and the vertical categories the lengths of their left middle fingers in millimetres. The correlation coefficient found by the product moment method is $\cdot6608 \pm \cdot0069$.

F. Table II divided so that the mean falls in cell n_{22} .

	$55\frac{9}{16}''-64\frac{9}{16}''$	$64\frac{9}{16}''-66\frac{9}{16}''$	$66\frac{9}{16}''-77''$	Total
9.4-11.3 mm.	682	270	101	1053
11.4-11.7 mm.	282	351	286	919
11.8-13.5 mm.	90	299	639	1028
Total	1054	920	1026	3000

$$\begin{aligned} r_{\psi} &= \cdot6635 \\ r_{\phi} &= \cdot6170 \pm \cdot0075 && \text{A.P.E.} \\ r_p &= \cdot6544 \pm \cdot0101 \\ r_e &= \cdot6316 \pm \cdot0301 \\ r_m &= \cdot6530 \pm \cdot0101 \\ r_w &= \cdot6538 \pm \cdot0101 \\ r_c &= \cdot6911 \end{aligned} \quad \left. \begin{aligned} r_{11} &= \cdot667 \pm \cdot013 \quad (\cdot014) \\ r_{12} &= \cdot670 \pm \cdot014 \quad (\cdot013) \\ r_{21} &= \cdot644 \pm \cdot015 \quad (\cdot014) \\ r_{22} &= \cdot631 \pm \cdot014 \quad (\cdot014) \end{aligned} \right\}$$

	Mean dispersion	Individual dispersion
η_{k_1}	$\cdot6477$	$\cdot6295$
η_{k_2}	$\cdot6548$	$\cdot6306$
η_{h_1}	$\cdot6647$	$\cdot6151$
η_{h_2}	$\cdot6345$	$\cdot6510$

G. Table II divided so that the mean falls in cell n_{11} .

	$55\frac{9}{16}''-65\frac{9}{16}''$	$65\frac{9}{16}''-66\frac{9}{16}''$	$66\frac{9}{16}''-77''$	Total
9.4-11.5 mm.	1122	176	216	1514
11.6-11.7 mm.	191	96	171	458
11.8-13.5 mm.	203	186	639	1028
Total	1516	458	1026	3000

$$\begin{aligned} r_{\psi} &= \cdot731 \\ r_{\phi} &= \cdot6426 \pm \cdot0077 && \text{A.P.E.} \\ r_p &= \cdot6613 \pm \cdot0108 \\ r_e &= \cdot6808 \pm \cdot0390 \\ r_m &= \cdot6553 \pm \cdot0111 \\ r_w &= \cdot6573 \pm \cdot0112 \\ r_c &= \cdot7182 \end{aligned} \quad \left. \begin{aligned} r_{11} &= \cdot680 \pm \cdot012 \quad (\cdot013) \\ r_{12} &= \cdot668 \pm \cdot013 \quad (\cdot013) \\ r_{21} &= \cdot642 \pm \cdot014 \quad (\cdot014) \\ r_{22} &= \cdot631 \pm \cdot014 \quad (\cdot014) \end{aligned} \right\}$$

	Mean dispersion	Individual dispersion
η_{k_1}	·6553	·6421
η_{k_2}	·6473	·6546
η_{h_1}	·6657	·6669
η_{h_2}	·6277	·7113

H. Table II divided so that the mean falls in cell n_{12} .

	$55 \frac{9}{16}'' - 65 \frac{9}{16}''$	$65 \frac{9}{16}'' - 66 \frac{9}{16}''$	$66 \frac{9}{16}'' - 77''$	Total
9·4-11·3 mm.	840	112	101	1053
11·4-11·7 mm.	473	160	286	919
11·8-13·5 mm.	203	186	639	1028
Total	1516	458	1026	3000

$$r_\psi = \cdot669$$

$$r_\phi = \cdot6172 \pm \cdot0088$$

A.P.E.

$$r_p = \cdot6479 \pm \cdot0107$$

$$r_e = \cdot5162 \pm \cdot0438$$

$$r_m = \cdot6478 \pm \cdot0108$$

$$r_w = \cdot6591 \pm \cdot0107$$

$$r_c = \cdot7920$$

$$r_{11} = \cdot648 \pm \cdot014 \quad (\cdot014)$$

$$r_{12} = \cdot668 \pm \cdot013 \quad (\cdot013)$$

$$r_{21} = \cdot644 \pm \cdot015 \quad (\cdot014)$$

$$r_{22} = \cdot631 \pm \cdot014 \quad (\cdot014)$$

	Mean dispersion	Individual dispersion
η_{k_1}	·6539	·6279
η_{k_2}	·6472	·6156
η_{h_1}	·6479	·6103
η_{h_2}	·6345	·6418

III. The third table examined was taken from Pearson and Heron's paper "On Theories of Association," *Biometrika*, vol. ix, p. 219, Table XIII, and is a normal surface having $r = \cdot3$. The values of the frequencies to two places of decimals were used.

	1	2	3	4	5 + 6	7	8	Total
1	4·04	17·16	7·55	3·30	0·91	0·92	0·12	34
2	17·41	123·59	79·76	46·64	14·61	17·67	3·32	301
3	8·86	93·00	78·31	52·04	19·20	26·40	6·19	284
4	2·83	37·73	37·24	27·51	10·95	16·31	4·43	137
5 + 6	1·62	25·21	27·75	22·09	9·26	14·64	4·43	105
7	1·02	19·50	24·47	21·39	9·58	16·36	5·68	98
8	0·22	5·81	8·92	9·03	4·49	8·70	3·83	41
Total	36	322	264	180	69	101	28	1000

K. Table III divided so that the mean falls in cell n_{22} .

	1 + 2	3 + 4	5 + 6 + 7 + 8	Total
1 + 2	162.20	135.25	37.55	335
3 + 4	142.42	195.10	83.48	421
5 + 6 + 7 + 8	53.38	113.65	76.97	244
Total	358	444	198	1000

$$\begin{aligned} r_{\psi} &= .3088 \\ r_{\phi} &= .2960 \pm .0199 && \text{A.P.E.} \\ r_p &= .3000 \pm .0246 \\ r_e &= .3000 \pm .0991 \\ r_m &= .3000 \pm .0249 \\ r_w &= .3000 \pm .0247 \\ r_c &= .3025 \end{aligned} \left\{ \begin{array}{l} r_{11} = .300 \pm .033 \quad (.034) \\ r_{12} = .300 \pm .036 \quad (.035) \\ r_{21} = .300 \pm .038 \quad (.037) \\ r_{22} = .300 \pm .038 \quad (.039) \end{array} \right.$$

	Mean dispersion	Individual dispersion		Mean dispersion	Individual dispersion
η_{k_1}	.3011	.2988	η_{h_1}	.2998	.2997
η_{k_2}	.2999	.3003	η_{h_2}	.2989	.2989

L. Table III divided so that the mean falls in cell n_{11} .

	1 + 2 + 3	4 + 5 + 6	7 + 8	Total
1 + 2 + 3	429.68	134.70	54.62	619
4 + 5 + 6	132.38	69.81	39.81	242
7 + 8	59.94	44.49	34.57	139
Total	622	249	129	1000

$$\begin{aligned} r_{\psi} &= .330 \\ r_{\phi} &= .3095 \pm .0225 && \text{A.P.E.} \\ r_p &= .3000 \pm .0282 \\ r_e &= .3000 \pm .1031 \\ r_m &= .3000 \pm .0326 \\ r_w &= .3000 \pm .0285 \\ r_c &= .3152 \end{aligned} \left\{ \begin{array}{l} r_{11} = .300 \pm .032 \quad (.033) \\ r_{12} = .300 \pm .039 \quad (.040) \\ r_{21} = .300 \pm .040 \quad (.041) \\ r_{22} = .300 \pm .046 \quad (.050) \end{array} \right.$$

	Mean dispersion	Individual dispersion
η_{k_1}	.2965	.2975
η_{k_2}	.2910	.2898
η_{h_1}	.2953	.2975
η_{h_2}	.2909	.2871

IV. The fourth table examined was from Pearson and Lee's paper "On the Distribution of Frequency (Variation and Correlation) of Barometric Heights at Divers Stations," *Phil. Trans. A*, 1897, vol. 190, p. 453, Table IX. The original

table is too extensive for reproduction and may be found in *loc. cit.* A condensed form of it will be found in *Biometrika*, vol. ix, 1913, p. 223, Table XVIII.

This was selected as an example of a very skew distribution. The correlation coefficient found by Product Moments is .780 (*Biometrika*, vol. ix, p. 223).

M. Table IV divided so as to give a reasonably large frequency in the cell n_{22} . The mean falls in the cell n_{21} .

	30.1'' and over	30''-29.8''	29.7'' and under	Total
29.9'' and over	1086.5	412	43	1541.5
29.8''-29.7''	144.5	275	103	522.5
29.6'' and under	56.5	323	478.5	858.0
Total	1287.5	1010	624.5	2922

$$r_{\psi} = .789$$

$$r_{\phi} = .7504 \pm .0210$$

A.P.E.

$$r_p = .7864 \pm .0077 \quad \left. \begin{array}{l} r_{11} = .780 \pm .010 \quad (.010) \\ r_{12} = .787 \pm .011 \quad (.011) \\ r_{21} = .795 \pm .012 \quad (.011) \\ r_{22} = .785 \pm .011 \quad (.012) \end{array} \right\}$$

$$r_e = .7745 \pm .0151$$

$$r_m = .7877 \pm .0078$$

$$r_w = .7858 \pm .0077$$

$$r_c = .8770$$

	Mean dispersion	Individual dispersion		Mean dispersion	Individual dispersion
η_{k_1}	.7857	.7116	η_{h_1}	.7962	.7417
η_{k_2}	.7812	.6951	η_{h_2}	.8065	.6841

Appendix.

Probable errors of r_m , r_w , r_p .

	P.E. of arithmetic mean (r_m)	P.E. of weighted mean (r_w)	P.E. of polychoric r (r_p)
A	.0246	.0211	.0209
B	.0254	.0253	.0250
C	.0253	.0311	.0241
D	.0239	.0236	.0235
E	.0327	.0247	.0245
F	.0101	.0101	.0101
G	.0111	.0112	.0108
H	.0108	.0107	.0107
K	.0249	.0247	.0246
L	.0326	.0285	.0282
M	.0078	.0077	.0077

My thanks are due to Professor Pearson, who suggested the enquiry, for his ever ready help and advice throughout the work. I have also to thank Miss Alison Robertson for assistance in reading the proofs.