Comparative Pathology. — *The relation of head-length and head-index of* JOHANNSEN *and the spurious correlation of* PEARSON. By G. P. FRETS. (Communicated by Prof. J. BOEKE).

(Communicated at the meeting of April 24, 1937).

In a publication in 1907 JOHANNSEN discussed at length the fact that, if head measurements are arranged in groups according to the length, the average head-index decreases as the length increases. This observation was first made by him with beans.

The same fact was demonstrated by BOAS in 1899 with Indian material.

Similar calculations were made by me in 1922, following on JOHANNSEN and BOAS, and corresponding results obtained for the length and the head-index, also respectively for the breadth and the height and the head-index, for my own material and that of TOCHER.

In 1897 a communication was made in the Proceedings of the Royal Society by K. PEARSON, regarding what he calls "spurious correlation". He defines spurious correlation as "the correlation which will be found between indices, when the absolute values of the organs have been selected purely at random."

PEARSON also deduces a formula for this spurious correlation. Different statisticians have taken up this question. (YULE, GUMBEL).

Spurious correlation appears with the correlation of indices and, in general, if magnitudes of which the correlation is calculated are compounds and contain an equal component part; thus with indices, e.g., if one calculates the correlation of LB and LD, of LB and BD, of LD and BD; also with the correlation of L and LB, L and LD, but not with that of L and BD.

This spurious correlation is large. If spurious correlation appears, the correlation calculated can be taken as total or gross correlation and this is equal to the sum of the organic (true) correlation and the spurious correlation. Spurious correlation may be greater than the gross correlation.

In 1897 PEARSON stated "that the difficulty and danger which arise from the use of indices was brought home (to him) in an endeavour to deal with a considerable series of personal equation data" and that it was some time before he realised that "this spurious correlation had nothing to do with the manner of judging." He concludes that "the method which judges of the intensity of organic correlation by the reduction of all absolute measures to indices, is not free from obscurity."

YULE (1910) concurs with PEARSON in thinking that "the interpretation of correlation between indices is not free from obscurity", but in his view

"this obscurity is no less for correlations between absolute measurements." YULE then points out that we "at present have no real knowledge of the process of growth to guide us. We do not know the factors which determine the ultimate form of the individual."

With reference to PEARSON's publication, GALTON wrote a note, "believing that it might be useful in enabling others to realise the genesis of spurious correlation." (1897, p. 498). "The results arrived at," he says, "which are of serious interest, have at first sight a somewhat paradoxical appearance."[1])

Collaborators and pupils of PEARSON have determined the spurious correlation in the total correlation found by them. (WELDON, PEARL with the lower animals; FAWCETT, MACDONELL with the human skull).

FAWCETT (1902) found for Naqada skulls (Egypt) for 130 male skulls as correlation between L and LB the gross correlation equals $-.551 \pm -.04$ and the spurious correlation $-.770 \pm .023$ (according to MACDONELL $-.694 \pm .031$); for 169 female skulls these figures are $-.560 \pm .037$ (acc. to MACDONELL $-.613 \pm .032$) and $-.624 \pm .033$ (acc. to M. $-.673 \pm .028$) respectively.

MACDONELL (1904) found for 131 English male crania the gross correlation of L and LB equals $-.547 \pm .041$ and the spurious correlation $-.658 \pm .033$ and for 130 English female crania $-.541 \pm .042$ and $-.689 \pm .031$ respectively. In all these cases the spurious correlation is thus found to be greater than the gross correlation. As the organic correlation equals gross correlation-spurious correlation, the organic correlation is here positive.

JOHANNSEN does not mention the investigations of PEARSON regarding spurious correlation, nor PEARSON those of JOHANNSEN concerning the relation of head-length and index; PEARSON is acquainted with the publication of BOAS (1899).

In my investigations regarding the heredity of the dimensions and the indices of the seed of Phaseolus vulgaris, I found, the same as JOHANNSEN, that the average index decreases as the length of the beans increases (1934). I also noticed that, if the beans are grouped according to weight classes, there is a negative correlation of the length and the breadth for beans of the same weight class. Earlier (1922) I had found that there is a negative correlation of length and breadth for heads, of which $L + B + H$ (as expressing the size of the head) are the same. Like PEARSON and his pupils, I also found a negative correlation of L and LB.

In recent years (1932) TIMOFEEFF-RESSOVSKY and ZARAPKIN have determined with material of beetles that, besides dimensions and indices, the direction of the variations (das Gerichtetsein der Variationen) must be studied as well. As an example they also discuss the relation of the

[1]) PEARSON (1924) defends spurious correlation against criticisms. "It seems to be the appropiate word to describe these correlations which are due, or largely due to "covering" factors" (p. 356).

skull-length and the skull-index. These authors make no communication from the literature.

If we endeavour to determine the significance of the findings of JOHANNSEN, that with increasing length the average index decreases, then it becomes clear that this relation includes spurious correlation.

We may replace this relation by the correlation of length and breadth for heads of the same capacity and then find that this correlation is negative, likewise by the correlation of the length and the index, which is also negative.

If in the correlation table of L and index we calculate the average index for each length class, we get the result obtained by JOHANNSEN, viz. with

TABLE 1. Classification of $L\,B$-indices according to head-length classes. 2765 measurements of the head. Twenty years and older. The material till 1929.

Length of the head	$L\,B$-Index					
	N	Males		N	Females	
		$M \pm m$	$\sigma \pm m$		$M \pm m$	$\sigma \pm m$
16.5				5	86.2	
17	2	80.8		29	84.42 ± 0.65	3.56 ± 0.47
17.5	8	82		197	83.1 ± 0.17	2.76 ± 0.14
18	62	82.41 ± 0.35	2.77 ± 0.25	483	81.29 ± 0.1	2.27 ± 0.07
18.5	150	81.65 ± 0.235	2.87 ± 0.165	506	80.28 ± 0.11	2.49 ± 0.08
19	318	80.05 ± 0.125	2.29 ± 0.09	296	79.1 ± 0.17	2.4 ± 0.1
19.5	317	79.17 ± 0.155	2.74 ± 0.115	80	78.3 ± 0.22	2 ± 0.16
20	199	78.39 ± 0.2	2.73 ± 0.165	17	77.4	
20.5	76	77.19 ± 0.3	2.58 ± 0.21	3	77.7	
21	17	75.74				
	1149			1616		

increasing length a decreasing average index. (FRETS, 1922, tab. 14—16, p. 512.)

The correlation of length and breadth of a head-material selected at random is positive. As stated above, I have already calculated before that for heads of equal size (expressed by $L + B + H$) and for beans of the same weight there exists a negative correlation of the length and the breadth.

In the following manner I have demonstrated the spurious correlation in my results.

PEARSON, when he regards the notion spurious correlation, says approximately: if an imp mixed up the observations belonging to each other and

if then the indices of these numbers not belonging to each other were calculated, and from these, two by two, the correlation, then one would obtain a correlation, and a rather high one.

I have done the same as PEARSON's imp, in order to demonstrate empirically the spurious correlation. I copied out the lengths and the breadths of the heads of adults of part of my material, arranged chronologically, and the indices and also the lengths, but now also showing the breadths in the reverse order in which they belong to the lengths; thus I now obtained side by side lengths and breadths, which do not belong to the same head, and from these dimensions calculated the "indices"[1]).

If, in accordance with JOHANNSEN, I calculate from these indices the average indices for the different length classes, then I also find here a decrease of the average indices with increasing lengths (tab. 2).

TABLE 2. Classification as in table 1 for "spurious heads", which are composed of the same material as of table 1.

Length of the head	$L\,B$-Index					
	N	Males		N	Females	
		$M \pm m$	$\sigma \pm m$		$M \pm m$	$\sigma \pm m$
16.5				6	89.2	
17	2	92 ± —		36	87.32 ± 0.5	3.03 ± 0.36
17.5	9	88.6 + —		187	84.27 ± 0.22	2.47 ± 0.16
18	63	85.3 ± 0.515	4.13 ± 0.365	478	82 ± 0.11	2.36 ± 0.1
18.5	150	83.27 ± 0.255	3.15 ± 0.18	505	80.1 ± 0.13	2.36 ± 0.9
19	315	80.83 ± 0.18	3.215 ± 0.13	294	78.12 ± 0.17	2.35 ± 0.12
19.5	322	78.9 ± 0.16	2.85 ± 0.11	82	76.9 ± 0.3	2.71 ± 0.22
20	199	76.65 ± 0.19	2.7 ± 0.135	18	74.2	
20.5	80	74.77 ± 0.71	3.21 ± 0.56	3	71.7	
21	17	74.1 ± —				
	1157			1609		

I likewise find a negative correlation of length and index, L and $L\,B$, (tab. 3).

A comparison of tables 1—3 shows that the spurious correlation is greater than the gross correlation. As gross correlation equals organic correlation plus spurious correlation, the organic correlation of L and $L\,B$ is positive. (tab. 3).

---

[1]) I see that WELDON determined the spurious correlation of his material in a corresponding manner.

TABLE 3. Correlation.

Correlation	Males		Females	
	N	$r \pm m$	N	$r \pm m$
Gross correlation of Land $L\,B$	1149	-0.426 ± 0.01	1616	-0.506 ± 0.02
Spurious „ „ „ „	1157	-0.698 ± 0.05	1609	-0.65 ± 0.014
Organic „ „ „ „		$+0.272$		$+0.144$
Stature and head-length (Material of TOCHER)	787	$+0.23 \pm 0.03$		

This is a remarkable result. JOHANNSEN (1907) does not discuss the significance of the relation found by him. He points out, it is true, that the length is a unit of measurement in the index and he mentions that for different bean species the results are not the same. He gives some examples in another field.

BOAS (1899) studies the biological significance of the index and, if he finds that with increasing head-length the average index falls, he points out that the cephalic index is (also) greatly influenced by causes other than the length and breadth of the head. No more than JOHANNSEN does BOAS recognise in his results the element of spurious correlation. He concludes (p. 460) "Among skulls belonging to the same type, a breath above the average is compensated by a height and length below the average. The law of compensation of VIRCHOW holds good also in normal skulls." "The correlation between length and breadth is not an expression of a biological relation between the two measurements, but an effect of the changes which both undergo when the capacity of the skull increases or decreases."

The negative correlation of head-length and head-breadth with skulls of the same capacity (or of length and breadth of beans of the same weight) was also explained by me (1922, 1934) as an expression of compensational growth in the sense of TSCHEPOURKOWSKY (1905). I would further mention that ORENSTEIN (1916) was struck by the "fact that the correlation coefficient for the length and the breadth of the head varies much from race to race" and he is inclined to believe that "a correlation really exists, but that its variation in value is very probably due to special factors, probably working independently and tending to modify the real measure of relationship. A close study of such factors is being undertaken" this author concluded.

TIMOFEEFF-RESSOVSKY and ZARAPKIN (1932a) put the problem as follows: "es kann die durch einen bestimmten Index charakterisierte Form in Abhängigkeit von der Grösze des betreffenden Merkmals variïren: das wird in den Fällen stattfinden, wenn die Raten der Grószenzunahme der zwei in Index in Relation gebrachte Masze verschieden sind."

TIMOFEEFF constructs "Formtendenzkurven", e.g. for the skull material of MACDONELL. These authors conclude (1932b) regarding "das Gerichtetsein der Formvariation" "(dass) die Variationsrichtung (die Spezifizität der Manifestierung) vorwiegend, wenn nicht ausschliesslich erblich bedingt ist und gegenüber Milieueinflüszen sehr resistent."

It is clear that TIMOFEEFF overlooks the spurious correlation. An analysis of his results is also necessary and the share of the spurious correlation must be determined.

The correlation of L and I, i.e. $\frac{100\,B}{L}$ is negative; with increasing L, I becomes smaller, because in I, i.e. in $\frac{100\,B}{L}$, L appears in the denominator. This is the spurious correlation, which is an arithmetical phenomenon arising as a result of our comparing length and $L\,B$-index, and these possess the same factor.

Spurious correlation has nothing to do with compensational growth, as was assumed by BOAS (1899) and FRETS (1922) in explanation of the negative correlation found by them, nor with heredity, as TIMOFEEFF-RESSOVSKY assumes in explanation of the Gerichtetsein der Variabilität. Only when the spurious correlation is removed from the gross correlation found by these writers can the possibilities of explanation for the then remaining organic correlation be admitted.

As stated above, in so far as the correlation of L and $L\,B$ differs from the spurious correlation, there is an organic correlation; thus organic correlation equals gross correlation minus spurious correlation. With increasing length the average index changes, because L appears in $\frac{100\,B}{L}$; B is a varying percentage of L.

Because the numbers ascend in arithmetical progression, the spurious correlation arises. This must be removed from the gross correlation found. In our case there remains then a small positive correlation.

These correlations are low and the results of others vary. (Cf. PEARSON and his pupils; PEARSON p. 495, FAWCETT p. 454, MACDONELL p. 244).

If the positive correlation has any significance, the explanation by compensational growth lapses [1]). A small positive correlation may be regarded as an expression of adaptability. From parents with very varying shapes and sizes of heads, a potentially very great length with a potentially very small breadth may be found together in the shape of the head of the children; it is possible that as an expression of adaptation the potentially

[1]) D'ARCY W. THOMPSON in his book „Growth and Form" (1917) attaches importance to this principle. He considers the volume primary; "the volume is less affected than are the linear dimensions." (p. 776). "If thus a dimension changes, there will be an inclination with the other dimension to change in the opposite manner, in order to keep the volume unchanged." "The skull varies as a whole." These hypotheses are only to be proved empirically by further investigations.

very small breadth will be somewhat less small. Consequently for heads of the same group, so of equal length or of equal capacity, a small positive organic correlation of length and breadth would arise. (This small positive correlation of length and breadth is something else than the positive correlation of length and breadth of an arbitrary head-material, so of heads of different capacity; it is the difference of "interracial" and "intraracial" correlation). Still further investigation is necessary here with large and reliable material. PEARL also, on the lines of DRIESCH, takes up the general significance of the small positive correlation found by him in his material. He argues that DRIESCH's law of the proportionality of the parts, that is independent of the size, does not hold. There is a significant correlation between shape of body and absolute size.

Finally an observation regarding the negative correlation of stature and *L B*-index may be made. This negative correlation also, in my opinion, contains spurious correlation. The head-length is partly an extension of the body axis, i.e. a component part of the stature; there is a positive correlation of stature and head-length (tab. 3). The negative correlation of stature and *L B*-index is thus partly the same as that of head-length and *L B*-index and consequently contains spurious correlation. (Vide also JOHANNSEN).

In this communication I have aimed at showing that different authors, partly in other terms, have discussed the correlation of head-length and *L B*-index and that this correlation contains an untrue or spurious correlation not belonging to the organic correlation, which spurious correlation is fairly great, in our case greater than the total correlation.

It is necessary that this spurious correlation be taken into consideration in anthropology. Besides the index, also the dimensions of the material investigated must, therefore, always be mentioned.

LITERATURE.

1917. D'ARCY W. THOMSON: Growth and Form.
1899. F. BOAS: The Cephalic Index. American Anthropologist. Vol. **1**, p. 448.
1912. J. I. CRAIG: Anthropometry of Modern Egyptians. Biom. Vol. **8**, pag. 67.
1922. G. P. FRETS: The Index Cephalicus. Genetica Vol. **4**, pag. 481. (also published in "The Cephalic Index and its Heredity". Mart. Nijhoff. The Hague, 1925).
1934. G. P. FRETS & G. WANROOY: Die Erblichkeit der Bohnenform usw. Genetica, Bd. **16**. p. 45.
1935. G. P. FRETS: Pure Lines in Beans. Eugenical News, p. 33.
1897. F. GALTON: Note to the Memoir by Prof. K. PEARSON on Spurious Correlation. Proc. Roy. Soc. Vol. **60**, p. 498.
1926. E. J. GUMBEL: Spurious correlation and its significance to Physiology. Journ. Amer. Statst. Assoc. Vol. **21**. N. 153. pag. 192.
1907 W. JOHANNSEN: Ueber Dolichocephalie und Brachycephalie. Arch. f. Rass. u. Ges. Biol. **4** Jhrg., p. 171.
1926. ——— Elemente der Exakten Erblichkeitslehre. 3 Aufl.

1901 A. LEE and K. PEARSON: A First Study of the correlation of the Human Skull. Phil. Trans. Ser. A. Vol. **196**, p. 225.
1904. W. R. MACDONELL: A study of the Variation and Correlation of the Human Skull with ref. to English Crania. Biom. III, p. 191.
1916. M. M. ORENSTEEN: Correlation of Anthropometrical Measurements in Cairo-born Natives. Biom. Vol. **11**, p. 67.
1906. R. PEARL: Variation in Chilomonas under Favourable and Unfavourable Conditions. Biometrika Vol. **5**, p. 53.
1897. K. PEARSON: On a Form of Spurious Correlation which may arise when Indices are used in the Measurement of Organs. Proc. Roy. Soc. Vol. **60**, p. 489.
1924. K. PEARSON and A. G. DAVIN: On the biometric constants of the Human Skull. Biometrika, Vol. **16**, p. 328.
1911. E. SCHUSTER: First Results from the Oxford Anthropometric Laboratory. Biom. Vol. **8**, pag. 80.
1932a. N. W. TIMOFEEFF-RESSOVSKY u. S. R. ZARAPKIN: Zur Analyse der Formvariationen, usw. Biol. Zentr. **52**. Bd. 138.
1932b. ——— Zur Analyse der Formvariationen. Einige Gesetzmässigkeiten in der Variabilität der Fleckenform bei Epilachna chrysomelina F. Die Naturwissenschaften, 20. Jhrg. S. 384.
1908. J. F. TOCHER: Measurements of Scottish Soldiers. The Henderson Trust Reports II.
1905. E. TSCHEPOURKOWSKY: Contributions to the Study of Interracial correlation. Biom. Vol. **4**, pag. 286.
1897. W. F. R. WELDON: Note (to the publication of Prof. PEARSON on spurious Correlation) Proc. Roy. Soc. Vol. **60**, p. 498.
1929. G. U. YULE: An introduction to the Theory of Statistics. q. ed. London.
1910. G. U. YULE: On the interpretation of Correlations between Indices or Ratios. Journ. of the Roy. Statist. Soc. Vol. **73**, p. 644.