QH
431
T45
XX

AA000868244

UC SOUTHERN REGIONAL LIBRARY FACILITY

QH

431

T45

XX

Qx 101 1410

EMPIRICAL STUDIES IN THE THEORY OF MEASUREMENT

BY

EDWARD L. THORNDIKE,

Professor of Educational Psychology in Teachers College, Columbia University.

ARCHIVES OF PSYCHOLOGY

EDITED BY R. S. WOODWORTH

NO. 3, APRIL, 1907

Columbia University Contributions to Philosophy and Psychology, Vol. XV. No. 3

NEW YORK
THE SCIENCE PRESS

PRINTED BY
THE NEW ERA PRINTING COMPANY,
LANCASTER, PA.

CONTENTS

MEASUREMENTS OF TYPE AND VARIABILITY

§ 1. <i>The Comparative Accuracy of the Average and the Median</i>	1
§ 2. <i>The Comparative Accuracy of the Mean Square Deviation and the Average Deviation</i>	5
§ 3. <i>The Divergencies of the Obtained from the True Measures by Theory and by Experiment</i>	8
§ 4. <i>The Relation between the Amount of a Central Tendency and the Amount of the Variability of the Group about the Central Tendency</i>	9

MEASUREMENTS OF RELATIONSHIPS

§ 5. <i>The Meaning of Typical Measures of Relationship</i>	15
§ 6. <i>The Presuppositions of Measures of Relationship</i>	25
§ 7. <i>The Advantages of the Different Measures</i>	25
§ 8. <i>The Attenuation of Measurements of Relationship</i>	35
§ 9. <i>Minor Advice to Students of Mental and Social Relationships</i> . .	41

EMPIRICAL STUDIES IN THE THEORY OF MEASUREMENT

IN the present condition of psychology, sociology and education, convenience, economy and directness are as important desiderata in methods of measurement as refinement with respect to precision. The results of these studies justify certain methods which have the decided advantage of giving measures which are direct functions of the data, independent of any hypothesis about the prevalence of the so-called 'normal' distribution, but which have been somewhat discountenanced or at least neglected in both the theory and the practice of statistics.

The section on correlation attempts also to make clear just what is measured by a coefficient of correlation and what the dangers are in the application of correlation formulas without constant supervision by an adequate sense for the concrete individual facts to be related.

MEASUREMENTS OF TYPE AND VARIABILITY

§ 1. *The Comparative Accuracy of the Average and the Median*

The median as a measure of the central tendency of a series of measures has the advantages of greater quickness of calculation, freedom from the influence of erroneous measurements, ease of interpretation and often greater practical significance. It is, therefore, important to know whether the accuracy, with which the median actually obtained from a small sampling of a series conforms to the true median of the total series, is much less than the similar accuracy in the case of the more commonly used measure, the average.

It is possible with any given form of distribution to calculate on the basis of the theory of probability the accuracy in either case. Trusting that some one will soon do this for typical forms of distribution other than the so-called 'normal' I have chosen to get empirical data on the same question from actual experiments with random samplings from certain large series of measures.

The median was calculated for each sampling by regarding the total series as measures of a continuous variable, quantity 61, for instance, equalling from 60.0 up to 62.0, quantity 63 equalling from 62.0 up to 64.0, etc. Where the median fell within a unit of the scale, as of course it usually did, the fractional part was taken

which would be correct, supposing the cases within that unit of the scale to be equally frequent in all equal subdivisions of that unit of scale.

The series used were the four presented in Table I. A is an almost perfect representative of the so-called 'normal' surface of frequency, limited at about $+3.2\sigma$ and -3.2σ . B is also a symmetrical distribution following, but not so closely, the so-called 'normal' type. C is a skewed distribution of the kind so frequently found in mental and social measurements. D is a flattened and rather sharply cut-off type of distribution, such as occurs often in facts subject to conventional regulation. The number of cases was for A 1,000, for B 1,307, for C 1,250 and for D 600. The mechanical arrangement of each series was simply so many small cards or slips of paper each with a number written on it. In each series these cards were approximately of the same size, shape and weight. From such a series, properly shuffled in a large bowl, drawings were made.

The total number of cases in any series is of course of no significance. Whether a series contains 1,000, 1,100, 1,426, 13,982 or 160,000 cases makes no appreciable difference to any of the matters to be investigated here, and in the case of a distribution of the type of D, drawings of 100 from 6,000 cases would not differ appreciably from drawings from 600. The reason for the particular sizes of the total series was economy of time.

It is most convenient to arrange series for such experiments with measures $+$ and $-$ from the central tendency, as in B and D; the time of recording the results of draws is lessened and also the likelihood of errors. Thus in A -31 , -37 , -35 , etc., would be better than 61, 63, 65, etc. I give the series, however, in just the way they were made and used.

Every drawing of 10 or 50 or 55 or whatever number of cases was made from the full series. However, a draw of 10 having been made and recorded, a draw of 50 was obtained by adding 40 to the 10 and one of 100 by adding 50 to the 50. The 100 is thus from the full series, but is obtained with a saving of time.

As a rule drawings of 10 or 11, 50 or 55, 100 or 110 and 275 were made, but, with the larger drawings, if not exactly 50 or 100 were drawn, the drawing was still utilized. Of course exact similarity in the size of the drawings is of no consequence whatever to any of the conclusions drawn.

TABLE I.

A		B		C		D	
Quan- tity	Fre- quency	Quan- tity	Fre- quency	Quan- tity	Fre- quency	Quan- tity	Fre- quency
61	1			1	30	— 7	20
3	1						
5	1			2	80	— 5	80
7	2						
9	3			3	140	— 3	100
71	5						
3	6	— 27	1	4	175	— 1	100
5	9	— 25					
7	12	— 23	2	5	200	+ 1	110
9	15	— 21	2				
81	20	— 19	8	6	160	+ 3	90
3	26	— 17	10				
5	31	— 15	26	7	120	+ 5	70
7	37	— 13	28				
9	43	— 11	58	8	95	+ 7	30
91	50	— 9	62				
3	54	— 7	98	9	80		
5	59	— 5	102				
7	62	— 3	128	10	60		
9	63	— 1	129				
101	63	+ 1	132	11	45		
3	62	+ 3	125				
5	59	+ 5	102	12	35		
7	54	+ 7	98				
9	50	+ 9	64	13	20		
111	43	+ 11	56				
3	37	+ 13	28	14	10		
5	31	+ 15	26				
7	26	+ 17	11				
9	20	+ 19	7				
121	15	+ 21	2				
3	12	+ 23	1				
5	9	+ 25	1				
7	6	+ 27					
9	5						
131	3						
3	2						
5	1						
7	1						
9	1						
Av.	100		0		6.0		.0
Med.	100		0		5.5		.0
A.D.	10.0		6.2		2.3		3.13
σ	12.4		7.8		2.9		3.68
Q.	8.4		5.4		2.0		2.94

A.D. = the average deviation from the average.

σ = the mean square deviation from the average.

Q. = one half the difference between the 25 percentile and 75 percentile measures.

The results of these drawings are summarized in Table II. In Table II., Nt = the number of sets drawn; Nc = the number of cases in each set; $Av.$ = the average divergence of the obtained¹ from the true² average; $Med.$ = the average divergence of the obtained from the true median; $A.D.$ = the average divergence of the obtained from the true average deviation; σ = the average divergence of the obtained from the true mean square deviation; $Q.$ = the average divergence of the obtained from the true (75 percentile - 25 percentile)/2.

The figures for the last three divergences under 'Actual' are the direct results; the figures under 'Percentile' are these divergences in percents of the true fact.

The table shows that there is not enough superiority in accuracy in any case to outweigh the practical advantages which the median has as a measure of such quantities as prevail in the mental sciences.³ The divergences of the medians are on the whole only about 22 per cent. greater than those of the averages.*

TABLE II.
DIVERGENCE OF OBTAINED FROM TRUE

SERIES	Nt.	Nc.	Av.	Med.	ACTUAL			PERCENTILE		
					A.D.	σ	Q.	A.D.	σ	Q.
A	3	102	.47	.33	.33	.53	.77	.33	.43	.92
	4	53	.92	1.07	.78	1.08	.80	.78	.87	.95
	10	10	3.04	3.60	1.13	1.59	1.22	1.13	1.28	1.45
B	3	100	.57	.47	.27	.37	.13	.44	.47	.24
	4	50	1.25	1.15	.88	.65	1.05	1.42	.83	1.94
	5	10	1.80	2.14	1.28	1.50	.70	2.06	1.92	1.30
C	2	275	.045	.055	.05	.075	.055	.22	.26	.28
	3	110	.11	.17	.04	.04	.21	.17	.14	1.05
	5	55	.33	.43	.18	.18	.22	.78	.62	1.10
	25	11	.62	.77						
D	7	100	.33	.53	.097	.081	.14	.31	.22	.48
	6	50	.34	.59	.19	.16	.21	.61	.44	.71
	16	10	.86	1.04						

¹ Obtained, that is, from the limited number of cases in the drawing in question.

² 'True' meaning here that obtainable if all the measures of the total series are taken.

³ The general merits of the median are not discussed in this report; so also in the case of the general merits of the average deviation and of percentile measures of variability. To thoughtful students of mental measurements they will be obvious. The matter is briefly discussed in my 'Mental and Social Measurements' ('04) pp. 37 ff. and *passim*.

figures would indicate no practical diff. when $N = 50$ or more.

§ 2. *The Comparative Accuracy of the Mean Square Deviation (called by various authors σ , μ , ϵ , S.D., or Standard Deviation) and the Average Deviation*

The most burdensome of the ordinary statistical operations is the calculation of the square root of the average of the squares of the deviations of a series of measures from their central tendency, that is, of the mean square deviation. In the case of the author, at least, practical judgment has long rebelled against the imposition of this measure upon workers with mental measurements by the experts in the theory of measurement of variable facts. To call it the *standard* deviation has seemed to him objectionable. There is apparently no reason whatever for its use except its supposedly greater accuracy. Perhaps because of lack of knowledge of the purely mathematical side of statistics, I am not aware that this greater amount of accuracy has been calculated from theory in the case of typical forms of distribution other than the so-called 'normal.' At all events it will be useful to the non-mathematical student to learn the facts in the case of empirical samplings from known series.

The series were A, B, C and D of Table I. The facts concerning the divergences from the true average deviation of the total series of average deviations obtained from random samplings, and similarly for mean square deviations, are given in Table II.

The average deviation and the mean square deviation were calculated from an approximate average never over a half of the unit of the scale from the actual average and as a rule from an approximate average less than a fourth of the unit of the scale from the actual average. The Q. was calculated on the basis of the same suppositions as the median.

So far as these samplings go, the average deviation is nearly as accurate as the mean square deviation. The latter is on the whole 5 per cent. more accurate, with about one chance in eighteen that an infinite number of drawings from these series would raise this superiority to 15 per cent. There surely can not be enough superiority of the latter to recommend its use in even 10 per cent. of the operations involved in present researches in psychology, sociology or education. Indeed it is a question whether the mathematical statisticians ought not to recognize the average deviation as approximately equal in accuracy and vastly superior in practical serviceableness, and hence as the measure to be recommended to students.

There is something to be said in favor of a still simpler measure of variability, the percentile. Galton's quartile (Q.), for instance (one half the distance between the 25th percentile and the 75th

percentile), is for a sampling of 100 or more of a series scored on a reasonably fine scale nearly as accurate as the measures that take account of the amount of every deviation. The facts for my series are given in Table II. In general the arguments that support the median as a measure of central tendency, support also the Quartile as a measure of variability in the case of large samplings from finely scaled series (say of 100 or more cases of a series with 20 or more steps). In the case of smaller samplings the calculation of Q. is often as long as that of the A.D.

If the report of an investigation gives somewhere the entire distributions the author may properly compute only the medians and Q.'s. If the average is used as the measure of central tendency the Q. is of course not very advantageous, since an approximate A.D. will have been calculated in getting the average.

Table III. gives the results of the individual sets drawn. It is not necessary to examine it to follow the discussion past or to come, but I insert it for the sake of any student who may wish to make calculations from its facts other than those which I have made in Table II.

TABLE III.

SERIES A						SERIES B					
N.	Av.	M.	A.D.	σ	Q.	N.	Av.	M.	A.D.	σ	Q.
101	100.2	99.8	10.0	12.0	9.3	101	— .5	— 1.0	6.6	8.3	5.2
105	101.2	100.1	10.6	13.0	8.9	100	+ 1.0	+ .4	6.3	7.7	5.6
101	100.0	99.3	10.4	13.0	9.3	100	+ .2	0.0	5.9	7.3	5.4
Sum of						Sum of					
Dev.	1.4	1.0	1.0	1.6	2.3	Dev.	1.7	1.4	.8	1.1	.4
52	99.1	98.8	11.3	14.9	8.6	50	— .3	+ .3	6.0	7.6	5.3
54	100.9	101.6	9.3	12.0	6.1	51	— 2.4	— 2.5	7.9	9.7	6.6
54	98.6	99.0	9.1	11.2	7.9	50	+ 1.7	+ 1.6	5.7	7.4	4.1
51	100.5	99.5	10.2	12.6	8.6	50	+ .6	+ .2	5.1	6.7	3.8
Sum of						Sum of					
Dev.	3.7	4.3	3.1	4.3	3.2	Dev.	5.0	4.6	3.5	2.6	4.2
10	95.0	94.0	10.5	11.4	9.8	10	— .8	+ 1.0	5.0	5.7	4.3
10	96.2	95.0	7.6	9.7	6.0	10	+ .6	+ .7	4.4	5.7	4.3
10	106.6	108.0	10.0	12.1	10.0	10	— 3.2	— 2.0	6.2	8.3	4.8
10	94.6	96.0	7.6	9.0	8.0	10	+ 1.6	+ 5.0	8.8	11.1	6.0
10	102.4	99.0	13.0	17.3	8.2	10	— 2.8	— 2.0	7.0	8.3	5.5
10	99.6	104.0	10.4	11.4	10.3	Sum of					
10	102.8	99.0	10.6	11.9	9.3	Dev.	9.0	10.7	6.4	7.5	3.5
10	102.4	101.0	10.6	12.4	8.3						
10	100.4	104.0	10.8	12.7	9.3						
10	101.2	102.0	9.4	14.2	6.0						
Sum of											
Dev.	30.4	36.0	11.3	15.9	12.2						

TABLE III. (continued)

SERIES C

Divergence in 25 sets of 11 cases

Av. Obt.-Av. True.	M. Obt.-M. True	In 5 sets of 55 cases				
		Av. Obt.- Tr.	M. Obt.- Tr.	A.D. Obt.- Tr.	σ Obt.- Tr.	Q. Obt.- Tr.
— 1.0	— 1.2					
— .8	— .8					
— .6	— .5					
— .6	— .5	— .20	— .23	.06	.12	.10
— .5	— .4	+ .09	+ .21	.12	.13	.13
— .4	— .3	+ .27	+ .32	.15	.16	.14
— .4	— .3	+ .42	+ .50	.21	.17	.27
— .3	— .2	+ .69	+ .79	.35	.35	.44
— .2	— .2					
0	+ .2					
+ .1	+ .2					
+ .2	+ .3					
+ .3	+ .4					
+ .4	+ .5					
+ .4	+ .5					
+ .4	+ .5					
+ .5	+ .5					
+ .5	+ .8					
+ .6	+ .8					
+ .8	+ 1.2					
+ 1.0	+ 1.5					
+ 1.1	+ 1.5					
+ 1.2	+ 1.5					
+ 1.6	+ 1.5					
+ 1.8	+ 3.1					
Av. Div. from true =		.33	.43	.18	.18	.22
.62	.77					

In 3 sets of 110 cases

In 2 sets of 275 cases

SERIES D

In 16 sets of 10 cases

Av. Obt.-Av. True.	M. Obt.-M. True.	In 6 sets of 50 cases				
		Av. Obt.- Tr.	M. Obt.- Tr.	A.D. Obt.- Tr.	σ Obt.- Tr.	Q. Obt.- Tr.
— 1.6	— 2.5					
— 1.4	— 2.0					
— 1.4	— 2.0	+ .09	— .37	.09	.12	.09
— 1.2	— 1.0	— .76	— .84	.43	.30	.40
— 1.2	— 1.0	+ .16	— .40	.11	.10	.26
— 1.0	— 1.0	— .10	— .50	.01	.07	.01
— .8	— 1.0	+ .36	+ .32	.27	.14	.35
— .6	— 1.0	— .56	+ 1.10	.25	.24	.18
— .4	0					
— .4	0					
0	0					
0	0					
+ .6	+ 1.0					
+ .6	+ 1.0					
+ .8	+ 1.0					
+ 1.8	+ 2.0					
Av. Div. from true =		.34	.59	.19	.16	.21
.86	1.04					

In 7 sets of 100 cases

Av. Div.
from true

§ 3. *The Divergences of the Obtained from the True Measures by Theory and by Experiment*

It is always interesting to compare the result of experiments in chance with the expectations derived from the theory of probability. Accordingly, I give the facts in Table IV. so as to save the reader interested in this matter the time of collation and calculation from the data of Tables II. and III.

The figures under Theory were calculated not from the A.D.'s of all the separate samplings, but once for all from the A.D. of the total series.

The figure under Theory is not in any case exactly the amount to be expected under strictly correct theory, but is the amount to be expected from the formula $A.D._{tr. - obt. Av.} = \frac{A.D._{dis.}}{\sqrt{n}}$.

This formula, applicable to cases of random sampling from a distribution of the so-called normal type, will of course not suit exactly distributions limited in extent and irregular in form. Comparison with it is however the important matter practically, since it is the formula in universal use.

TABLE IV.

AV. DEV. OF OBTAINED FROM TRUE AVERAGE

Ser.	N	n	Theory.	Exper.	$\frac{E}{T}$
A	102	3	1.00	.47	47 %
B	100	3	.62	.57	92 "
C	110	3	.22	.11	50 "
D	100	7	.31	.33	107 "
A	53	4	1.37	.93	68 "
B	50	4	.87	1.25	144 "
C	55	5	.31	.33	107 "
D	50	6	.44	.34	77 "
A	10	10	3.20	3.00	94 "
B	10	5	1.96	1.80	92 "
C	11	25	.69	.62	90 "
D	10	16	.99	.86	87 "

§ 4. *The Relation Between the Amount of a Central Tendency and the Amount of the Variability of the Group about the Central Tendency*

In comparing groups with respect to variability allowance must be made for the fact that, in certain cases at least, the amounts of the central tendency influence the amounts of the variabilities. Thus the A.D. of men in weight is hundreds of times that of butterflies, yet the former are of course not really a hundred times as variable. Thus the A.D. of a group in a test of addition was, for trials of 40 seconds, 2.18; for trials of 80 seconds, 3.41; and for trials of 120 seconds, 5.18. It would obviously be silly if we had tested men with trials of 80 seconds and women with trials of 40 seconds, and obtained these results, to infer that men are 50 per cent. more variable in ability to add than are women.

In using the so-called coefficient of variation (proposed by Pearson) one makes allowance for the possible influence of the central tendencies' amounts by dividing through the gross variabilities each by the amount of its corresponding central tendency. I have elsewhere shown that for mental and social measurements no one such rule can be always or even often right and suggested that in any case a division through by the square root of the corresponding central tendency is more in accord with both theory and facts.¹

In this section enough data will be presented to practically demonstrate both of these assertions. It is not important to investigate the matter exhaustively for the very reason that no one general rule for comparing groups with respect to variability can be found. All that is needed is a clear enough proof of the inadequacy of the practice of comparing groups after dividing through the gross variabilities by the corresponding means—clear enough to stop the spread of the practice and to warn readers against conclusions based on such comparisons.

If we take the arrays of y in a case where y is positively correlated with x we have a series of groups with central tendencies varying from lower to higher which are selected at random so far as concerns any influence on the variability except the influence of the amount of the mean. The differences in variability found for these arrays give, then, in connection with the differences in the amounts of their central tendencies, the answer to our problem for the case of comparisons of groups with respect to their variability in the same trait. If we find that even in such cases there is no constant relation of difference in central tendency to difference in

¹ Mental and Social Measurements, pp. 102-103.

variability, but that one law obtains for stature and another for span or finger length, then *a fortiori* no constant relation can be presupposed when the variability of a group in one trait is to be compared with its variability (or that of a second group) in a different trait.

The first facts to which I call the reader's attention are the comparison of arrays of y corresponding to very low values of x with arrays of y corresponding to very high values of x in the case of ten correlations chosen at random (so far as this issue is concerned) from Vols. I. and II. of *Biometrika*. The number of cases ranged from 49 to 319. The results are given in Table V. in the form of (1) the variability of arrays related to high central tendencies of x (and consequently having high central tendencies of y) divided by the variability of arrays related to low central tendencies of x , under the heading 'Gross'; (2) the Pearson coefficient of variability for the former divided by the Pearson coefficient of variability for the latter under the heading $\frac{\text{Gross}}{\text{C.T.}}$; and (3) the similar ratio for the two variabilities each having been divided by the square root of the amount of the corresponding central tendency, under the heading $\frac{\text{Gross}}{\sqrt{\text{C.T.}}}$. A perfect method would give values of 100 throughout.

TABLE V. (a)

	Gross.	$\frac{\text{Gross}}{\text{C.T.}}$	$\frac{\text{Gross}}{\sqrt{\text{C.T.}}}$	Nearest to Equality.
Width of head	101.5	95.7	98.5	$\frac{\text{Gross}}{\sqrt{\text{C.T.}}}$
Length of left middle finger	94.4			$\frac{\text{Gross}}{\sqrt{\text{C.T.}}}$
Number of stamens	112.3	89.5	96.3	$\frac{\text{Gross}}{\sqrt{\text{C.T.}}}$
Number of stamens	164.7	110	135	$\frac{\text{Gross}}{\text{C.T.}}$
Frontal breadth	106.8	86	96.1	$\frac{\text{Gross}}{\sqrt{\text{C.T.}}}$
Length of right antenna (aphis)	88.2			$\frac{\text{Gross}}{\sqrt{\text{C.T.}}}$
Number of stamens	111.6	90.8	100.8	$\frac{\text{Gross}}{\sqrt{\text{C.T.}}}$
Number of stamens (lesser celandine)	132.6	106	119	$\frac{\text{Gross}}{\text{C.T.}}$
Span	115	99	107	$\frac{\text{Gross}}{\text{C.T.}}$
Forearm length	95.7			$\frac{\text{Gross}}{\sqrt{\text{C.T.}}}$
Median	109.2	90.1	97.5	$\frac{\text{Gross}}{\sqrt{\text{C.T.}}}$

The detailed facts from which these ratios come are given in Table V. (b).

TABLE V. (b)

VARIABILITIES OF ARRAYS OF y RELATED TO LOW AND HIGH VALUES OF x . IN TERMS OF A.D.

(Each case measured is recorded in three lines: the first line gives the values of x ; the second line gives the variabilities of the related arrays of y ; the third line gives the numbers of cases in the arrays. The volume and page numbers refer always to *Biometrika*.)

I., 214	x = Head length	18.0	18.1	18.2	20.1	20.2	20.3	20.4			
	y = Head breadth	3.2	3.1	3.0	3.2	3.9	3.5	3.0			
		35	38	51	53	54	33	30			
I., 216	x = Height	58.6	59.6	60.6	69.6	70.6	71.6				
	y = Left middle finger length	3.6	4.0	3.0	3.2	3.3	3.2				
		23	48	90	97	46	16				
I., 126	x = No. of pistils	12	13	14	20	21	22	23			
Table I.	y = No. of stamens	2.1	2.3	2.4	2.8	2.6	2.3	2.4			
		13	12	22	19	13	15	10			
I., 126	x = No. of pistils	6	7	8	16	17	18				
Table II.	y = No. of stamens	1.8	1.1	1.1	2.0	1.8	2.2				
		6	16	35	23	16	11				
I., 152	x = Frontal breadth (aphis) 1st brother	13.5			19.5						
	y = Frontal breadth (aphis) 2d brother	1.46			1.56						
		57			50						
I., 153	x = Length of antenna (aphis) 1st brother	26	28		48	50					
	y = Length of antenna (aphis) 2d brother	1.6	1.6		1.2	3.3					
		14	71		43	12					
II., 161	x = No. of pistils (lesser celandine)	13			22						
	y = No. of stamens (lesser celandine)	1.8			2.3						
		24			25						
II., 162	x = No. of pistils (lesser celandine)	14	15	16	17	28	29	30	31	32	33
	y = No. of stamens (lesser celandine)	1.9	2.6	2.1	4.7	3.6	3.4	3.2	3.4	4.0	3.8
		10	17	16	28	20	16	13	11	19	15
II., 399	x = Height	61	62		72	73					
	y = Span	1.2	1.4		1.7	1.5					
		8.5	32.5		33	13					
II., 403	x = Span	63	64		75	76					
	y = Forearm	.83	1.28		1.03	1.28					
		13	32		28	11.5					

The gross variabilities often increase as we would expect with higher central tendencies, though by no means always. Seven out of ten do so, giving a median value of 109.2 instead of 100. The Pearson coefficient of variation makes too much of a deduction for an increase in the amount of the central tendency in all but three cases, giving a median value of 90.1 instead of 100. The square root deduction, with a median value of 97.5, makes the least error of any one single method. These facts alone disqualify the so-called 'coefficient of variation' as a means of comparing variabilities. But more detailed studies of the cases of length of finger, span and stature will be still clearer.

The facts for length of left middle finger are as given in Table VI.

TABLE VI.

RELATION OF AMOUNT OF VARIABILITY TO AMOUNT OF CENTRAL TENDENCY.
FINGER LENGTH. (*Biometrika*, Vol. I., p. 216)

Array.	Value of x to Which the Array is Related.	No. of Cases in the Array.	Central Ten- dency of the Array.	Variability (A.D.) of the Array.
1	581	6	103	167
2	591	23	107	357
3	601	48	108	404
4	611	90	109	309
5	621	175	111	325
6	631	317	112	347
7	641	393	114	312
8	651	920	116	331
9	661			
10	671	413	118	339
11	681	264	119	345
12	691	177	120	334
13	701	97	122	322
14	711	46	124	333
15	721	17	126	318
16	731	7	128	386
17	741	4	128	275

In the case of finger length increase in the amount of the central tendency does *not* imply an appreciable increase in the amount of variability. No allowance is needed.

In the case of span it would be equally absurd *not* to make an allowance and one as great or nearly as great as the Pearson method makes. For the preliminary study of the variability of span reported in Table V. is confirmed by the facts in the case of three other span series. These facts (given in Table VII.) abundantly prove that the influence of the amount of the central tendency on the amount of the variability follows totally different laws in the case of span and of finger length.

TABLE VII.

RELATION OF AMOUNT OF VARIABILITY TO AMOUNT OF CENTRAL TENDENCY.
SPAN. (*Biometrika*, Vol. II., pp. 399-401)

Fathers.			Sons.			Mothers.			Daughters.		
N.	C.T.	Var.	N.	C.T.	Var.	N.	C.T.	Var.	N.	C.T.	Var.
32.5	642	578	31	657	519	18	571	428	15.5	585	471
42.5	651	587	56	660	505	34.5	582	596	52	595	447
71.5	658	560	78.5	670	516	79.5	593	600	101	600	466
122.5	667	666	127	677	580	135.5	600	524	150	613	515
142.5	675	662	178.5	687	608	163	609	608	199	620	485
136.5	687	593	189	700	600	183	619	573	438	625	510
154.5	692	574	137	707	636	163	627	554	169.5	650	492
118.5	702	658	137	715	505	114.5	637	542	151.5	660	605
102.5	713	698	93	720	601	78.5	640	624	81.5	660	601
56.5	720	601	52.5	735	503	41	647	588	40.5	665	481
33	735	678	39	745	595	16	655	881	19.5	680	436

As a final case let us take stature. Here the variability is slightly *less* as the amount of the central tendency increases. The facts are given in Table VIII. constructed on the same plan as Table VI.

TABLE VIII.

RELATION OF AMOUNT OF VARIABILITY TO AMOUNT OF CENTRAL TENDENCY IN
GROUPS DIFFERING IN CENTRAL TENDENCY. STATURE.
(*Biometrika*, Vol. I., p. 216)

Related to <i>x</i> .	N.	C.T.	A.D.
10.0	7	61.1	286
.1	7	61.1	146
.2	10	60.6	190
.3	17	60.7	179
.4	20	62.3	170
.5	30	62.8	183
.6	44	62.8	132
.7	74	62.1	153
.8			
.9	177	63.6	137
11.0			
.1	315	64.1	162
.2			
.3	347	64.6	155
.4			
.5	461	65.1	152
.6			
.7	458	66.1	158
.8			
.9	346	66.1	156
12.0			
.1	289	67.1	147
.2			
.3	180	67.1	157
.4	44	67.6	170
.5	52	69.6	158
.6	35	69.6	147
.7	31	68.6	127
.8	25	70.1	148
.9	7	69.1	136
13.0	8	69.1	264

MEASUREMENTS OF RELATIONSHIPS

The importance to any science of exact and convenient methods of measuring the relationships of the facts it studies should be obvious. It is therefore unfortunate that students of psychology and the social sciences have with few exceptions neglected both the theoretical problem of correlated variations and the careful measurement of such relationships as they have in fact found.

The failure to utilize the methods devised by Galton, Pearson, Sheppard, Spearman and others is due partly to an ignorant and partly to an intelligent suspicion aroused by the mathematical derivations of these methods. Ignorance of the rationale of their derivations cooperating with ignorance of the conditions which require their use and of the necessity of some such refined methods has caused the stupid suspicion and aversion. Inability to follow the mathematics of the derivation of formulæ, at least in detail, cooperating with the rational expectation that too abstract methods will fit the concrete cases imperfectly and with the equally rational confidence that proofs resting upon the assumption of close approximation of actual variations in mental and social facts to the probability curve distribution are always unsafe and, perhaps, usually misleading, has caused the intelligent suspicion.

It is probable that unless these methods are soon subjected to a review by some one who can both make perfectly clear their presuppositions to the rank and file of investigators in psychology and the social sciences and prove their applicability to actual cases of relations to be measured, there will be damage done in two ways. Many investigators will as in the past use hopelessly crude methods and misinterpret relationships; and also many investigators will learn off the formulas of the mathematical statisticians and apply them to cases where they are out of place and give inadequate and misleading results. To both of these errors the writer, for instance, confesses himself guilty in the past.

I am unable to make such a review but as no one of those who are able seems willing,¹ I have made a partial and inferior substitute for it which I hope may, in so far as it is sound, be instructive to students of mental measurements and, in so far as it is unsound,

¹ Perhaps Mr. C. Spearman's article on 'The Proof and Measurement of Association between Two Things' (in the *Am. J. of Psy.*, Vol. XV.) may be considered as filling the need, but I fear that it is too technical in parts and not inquisitive enough concerning the actual relations between (1) the individual relationships, from which all our computations ought to start, and (2) the general expressions or summaries of them. At all events I am not trying to do over again, for better or worse, what Mr. Spearman has done, but something which is needed as introductory and accessory to his work.

may provoke some capable student to give the adequate review that is so much needed.

This report will presuppose in the reader knowledge of the bare elements of the theory of measurement of variable facts such as is given for instance in the writer's *Introduction to the Theory of Mental and Social Measurements*. It will deal in order with the following topics:

I. What is actually measured by typical measures of the relationship between first and second member of a pair in a series of pairs of values, each first-member value being a deviation from the central tendency of one series and each second-member value being a related deviation from the central tendency of a second series?

II. What are the respective presuppositions of each of these typical measures?

III. What are the advantages and disadvantages of each of these typical measures?

The only original contributions which this discussion contains are (1) the investigation of certain artificially constructed cases of correlation and (2) a laborious but not very important experimental testing of the comparative reliability of different measures of relationship, and (3) a similar experimental testing of methods for correcting measures of relationship for the 'attenuation' due to inaccurate original data.

§ 5. I. *What is actually measured by typical measures of the relationship between first and second member of a pair in a series of pairs of values, each first-member value being a deviation from the central tendency of one series and each second-member value being a related deviation from the central tendency of a second series*

Consider the following series of paired values of *A* and *B*:

<i>A</i>	<i>B</i>	<i>A</i>	<i>B</i>	<i>A</i>	<i>B</i>
-7	-5	-1	-3	+1	+7
-5	-5	-1	-1	+3	-1
-5	-3	-1	+1	+3	+1
-5	-1	-1	+1	+3	+1
-3	-7	-1	+3	+3	+3
-3	-3	+1	-3	+3	+5
-3	-3	+1	-1	+5	-1
-3	-1	+1	+1	+5	+3
-3	+1	+1	+3	+5	+3
-1	-5	+1	+5	+7	+5

Pearson Coefficient = .634.

Median Ratio *B/A* = .65.

Average of Ratios = .902.

The average of ratios is valueless because it overweights positive values of 1, 2 pairs, etc.

Per cent. unlike signs = .267, *r* as calculated therefrom being .665. (cf. pp. 21)

Outline

30/800
26
200

Each of these pairs represents a relationship, the entire series reading: A deviation in *A* of -7 from the central tendency of *A* brought with it a deviation in *B* of -5 from the central tendency of *B*; a deviation of -5 brought in one case a deviation in *B* of -5 , in a second case one of -3 , and in a third case of -1 , etc.

Consider now two measures each expressing an important fact concerning this series of 30 individual relationships. The first is,

$\frac{\Sigma(AB)}{\sqrt{\Sigma A^2} \sqrt{\Sigma B^2}} = .634$. The second is, The median of the 30 *B/A* ratios = .65. The former is of course the Pearson Coefficient of correlation for *A*—*B*; the latter is the Median or Mid Ratio *B/A*.

What the former measures can not be stated except in terms not yet given by the individual relationships themselves. Professor Pearson's own statements for instance are in terms of certain facts of a correlation diagram such as Fig. 1, not in terms of the individual relationships.

It is clear that in the case of Fig. 1, which represents our 30 relationships graphically, the slope of the straight line *LL*¹ through

O so drawn that the sum of the deviations of the individual dots from it is zero (measuring deviations in the direction of the *B* line and calling deviations above the line in the left hand half of the surface and below the line in the right hand half of the surface +, and calling deviations below the line in the left half and above the line in the right hand half —) is a measure of an important fact about the series of relationships.

The Pearson Coefficient does not, however, measure the slope of just such a line as we have supposed to be drawn in Fig. 1 and described in the last paragraph. Its line is not so calculated as to

¹ In this case the slope is roughly 73 per cent. of 45°, the slope which would be found were correlation perfect. The slope for the *A*'s taken as dependent on the *B*'s is roughly 64 per cent. of 45°.

make the deviations from it toward closer correlation equal to the deviations from it towards less correlation, but is so calculated as to make the sum of the squares of the deviations from it least.

This of course weights the extreme deviations much more than those near the center of the surface, for the same change in the slope of the line alters the sum of the squares of the deviations from the line near the center of the surface far less than that of the remote deviations. This is a possibly questionable feature of the Pearson Coefficient.

Moreover it is calculated as the slope of this line of so-called 'regression' as found when the two traits are reduced to equivalence of variability and double entries are made in the correlation table, i. e., *B*'s as related to *A*'s and *A*'s as related to *B*'s, the two sets of entries being so superposed that the intersection of the means in the one case coincides with the intersection of the means in the other case.

Professor Pearson gives many readers the impression that his coefficient of correlation is calculated as the slope of the straight line

Fig. 2.

Fig. 3.

through 0 to fit the points in the correlation diagram that represent the means of the arrays¹ (the two related series being reduced to an equivalence in variability and entered doubly), but in fact it is the slope of the line from which the sum of the squares of the deviations of all the dots each representing one relationship is least, not the slope of the line from which the sum of the squares of the deviations of the dots representing each the mean of one array is least. It is in our illustration a line to fit the dots of Fig. 3, not those of Fig. 2. That is, an array of 100 cases is (quite properly) given greater weight than one of 2 cases.

¹ See, for instance, 'Grammar of Science,' 2d edition, 1900, p. 393 and p. 396.

Consider now the Pearson Coefficient from another point of view. Let us for the present restrict relationships to those between two series of the same form of distribution, and also define perfect correlation as a relationship such that any deviation of A from its central tendency will imply a deviation of B from B 's central tendency which shall be the same fraction of B 's variability that the deviation of A is of A 's variability. That is,

$$\frac{B_1}{\text{Var. of } B \text{ series}} = \frac{A_1}{\text{Var. of } A \text{ series}}, \quad \frac{B_2}{\text{Var. of } B} = \frac{A_2}{\text{Var. of } A}, \quad \text{etc.}$$

If then all values of B are divided by $\frac{\text{Var. of } B \text{ series}}{\text{Var. of } A \text{ series}}$, we should in perfect correlation find each deviation of A accompanied by an identical deviation of B . The sum of the AB products would be equal to the sum of the A^2 , or to the sum of the B^2 , or to $\sqrt{\Sigma A^2} \sqrt{\Sigma B^2}$.

In the case of two series of the same form of distribution and of equal variability the Pearson Coefficient formula then measures the proportion which the sum of the series A_1B_1, A_2B_2 , etc., is of what it would be with perfect correlation as defined.

It can be shown that without reducing B or A to equivalence in variability perfect correlation as defined would give for the sum of the AB products $\sqrt{\Sigma A^2} \sqrt{\Sigma B^2}$, provided the form of distribution of A is the same as that of B .

The Pearson Coefficient measures, then, in cases where the form of distribution of the two facts to be related is the same, the proportion which the sum of the AB products is of what it would be were correlation perfect.

There is no ambiguity as to what is measured by the median of the B/A ratios. Whatever the distributions may be or the ratios, the median means always a definite thing: the ratio B/A which is exceeded in magnitude by as many of the ratios as it exceeds. We have only to note that the median of the B/A 's and the median of the A/B 's are two different things and that if we are interested in representing in one number both what a given A deviation implies with respect to B and what a given B deviation implies with respect to A , we must use both the B/A and the A/B median.

Certain other measures deserve mention. The directly calculated average of all the individual relationships B/A or A/B is a perfectly comprehensible measure but rather a useless one. The Modal Ratio B/A or A/B is also a perfectly clear conception and, in cases where it can be easily and accurately determined, a very valuable one.

The per cent. of direct or the per cent. of inverse relationships is equally comprehensible, and is an important function of the closeness of relationship.

TABLE IX.

-39 etc.	X $\rightarrow$		-1 +1	+39
39	1			1
37				1
35				1
				2
				3
				5
				6
				9
				12
				15
				20
				26
				31
				37
				44
				49
				55
				59
				62
				63
				63
				62
				59
				54
				50
				43
				37
				31
				26
				20
				15
				12
				9
				6
				5
				3
				2
				1
				1
				1
39				1

When the individual values of A and B are not measured as amounts of deviation from their central tendencies, but only as so many A^1 's known to be less than Z and so many A^2 's greater than Z , and as so many B^1 's less than W and so many B^2 's greater than W , the per cents. of A^1B^1 pairs, A^1B^2 pairs, A^2B^1 pairs and A^2B^2 pairs give important information.

The number and amount of the divergences of the ranks of the second members from the ranks of their related first members also give important information.

If the two related facts are of the so-called normal distribution and the relationship is uniform for all amounts of A and each array is also a normal distribution, the Median Ratio, the Modal Ratio and

FIG. 4.

the Pearson Coefficient will, if the two series are reduced to equivalence in variability, coincide and will equal $\cosine \pi U$.¹ This is the case of so-called normal correlation approximated in many organic and hereditary anatomical relationships. It is of course only one of many possible types of relationship. The extent to which it prevails in mental and social relationships is not known. Its prevalence in the case of anatomical facts has probably been over-estimated.

Table IX. gives the facts of the relationship between two series both of the same form of distribution, almost exactly the so-called normal, and of the same variability, the relationship being devised artificially so that the average of each array of y is $.5 \times$ the corresponding value of x . This regression of y on x is shown graphically in Fig. 4, which gives the average of each array of the y 's. The regression of x on y is shown graphically in Fig. 5, which gives the average of each array of the x 's. The Pearson Coefficient for this case is .53. The Median Ratio is much higher (.60 for the y/x

¹ U equalling the per cent. of unlike-signed pairs.

and x/y ratios together) because the correlation is much closer for mediocre values of x and y than for extreme values (see especially the regression of x on y). U is .292 and r from $\cos \pi U$ is accordingly .61.

This case illustrates the fact that the relation of y to x may not be the same as that of x to y even when the form of distribution and variability is the same for both cases. It also illustrates a rather close approach to the so-called 'normal' correlation.

FIG. 5.

Table X. gives graphically the correlation in the case of age at death of husband with age at death of wife in 935 pairs from records of the Society of Friends. This is taken from the table on p. 498 of Vol. I. of *Biometrika*, the table being due to Mary Beeton in cooperation with Karl Pearson. This case shows a relationship between two series neither of which is anything like normal in form of distribution, which are not of the same form of distribution and which therefore are in strictness incomparable in variability.

TABLE X.

Fig. 6 gives the regression of y (wife's age) on x (husband's age) in terms of averages of arrays of y and also of medians of arrays of y . To give the regression by single modes for the arrays would be fallacious, for each array is more or less clearly a bimodal distribution. This is shown in Fig. 7, where the y 's are grouped in four large arrays. It should be clear that any single figure is inadequate to express this relationship. The Pearson Coefficient of correlation is .20 and the regression of y (wife's age) on x (husband's age) calculated from it is .25. But this would lead one far astray concerning the real regression, as we see by Fig. 6. The relationship is closer for early deaths than for late. The form of distribution of the relationship is, apart from this, skewed in general from a mode of close resemblance toward very great diversity, and is in the third place complicated by the submodal tendency of a wife to die at about 35 more often than at 30 or 40. Such a case illustrates the fact that each type of measure of a relationship measures some particular aspect thereof and also the fact of the extreme abstractness from reality of the Pearson Coefficient, which in this

case measures neither a uniform tendency nor a central tendency of the series of individual relationships.

The reader will obtain concrete information about the meaning of the different measures of relationship and of their merits in actual practise if he will calculate them for a score of representative relationships and examine them in the light of the entire correlation tables. I have done this for the cosine πU and Median Ratio (or rather,

FIG. 6. The dotted line is from averages; the continuous line from medians. The dash line is the regression as calculated from the Pearson Coefficient.

in order to have the resulting figure comparable directly with the cosine πU and the Pearson r , for the median of all the ratios $\frac{y \text{ var. } x}{x \text{ var. } y}$ and $\frac{x \text{ var. } y}{y \text{ var. } x}$ in the case of nine relationships representing organic and hereditary and conjugal relations, relations in animals and in plants, relations of definite structural features and complex properties. The results are given in Table XI. They show that the

median ratio method gives results as close to the unlike-signs method as does the Pearson method. The reader who will examine Table XI. in connection with the original¹ correlation tables in *Biometrika*

FIG. 7.

will find also that where the Pearson Coefficient r and the Median Ratio r diverge at all widely it is the latter which better fulfils Pearson's criterion of telling how much nearer the most probable value of a second member of a pair is to the value of the first member than it would be with no relationship at all.

TABLE XI.

Vol.	Page	Traits to be Related	x	y	N	Mutual Relationship By Pearson Coef.	By Median Ratio	By Cosine πU
I.,	84	Longevity of adult brothers			2000	.2853	.479	.3763
I.,	126	No. of stamens with No. of pistils in late flowers of <i>Ficaria ranunculoides</i>	Pistils	Stamens	373	.7489	.80	.7815
I.,	214	Human head length with head width	Length	Width	3000	.4016	.415	.3875
I.,	216	Human height with left middle finger length	Height	L.M.F.	3000	.6608	.69	.6747
II.,	97	Capsule height of brother plants (Shirley poppies)			13800	.3782	.48	.5030
II.,	97	Stigmata of brother plants (Shirley poppies)			4716	.2561	.253	.2160
II.,	163	No. of stamens with No. of pistils in lesser celandine from Surrey	Pistils	Stamens	500	.6601	.55	.5570
II.,	498	Longevity of husband with longevity of wife. Friends' records	Husband	Wife	935	.1909	.41	.2560
III.,	170	Cephalic index of brothers			1982	.49	.53	.5090
Average difference of r by Pearson Coefficient from r by $\cos \pi U$						.055.		
Average difference of r by Median Ratio from r by $\cos \pi U$						.045.		

¹ These examples are all taken from the first three volumes of *Biometrika*, the 'Vol.' and 'Page' of the table referring to that journal.

§ 6. *The Presuppositions of Measures of Relationship*

The Pearson Coefficient.

Taken at its mere face value, $\frac{\Sigma(xy)}{\sqrt{\Sigma x^2} \sqrt{\Sigma y^2}}$ or $\frac{\Sigma(xy)}{n\sigma_1\sigma_2}$, the Pearson Coefficient has of course no presuppositions, but if it means the proportion that the $\Sigma(xy)$ is of what it would be with perfect correlation it presupposes sameness of form of distribution in the two series. If it means the proportion which the slope of a certain straight line is of the slope of the line of perfect correlation, the certain line being so drawn that the sum of the squares of the divergences from it of the given y values (in double entry) toward greater correlation equals the sum of the squares of those toward less, it presupposes the 'normal' distribution in the case of both series.

The Median Ratio.

The Median Ratio need have no presuppositions. It is simply one of the obtained individual relationships. When, however, we come to draw inferences from it about the entire series of relationships, we must state certain additional facts or use certain presuppositions.

The Modal Ratio and the Percentage of Like-signed or of Unlike-signed pairs are also directly drawn from the series of individual relationships themselves. In calculating the general trend of relationship, r , from $r = \cosine \pi U$ (U being the per cent. of unlike-signed pairs) we presuppose (if I understand Mr. Sheppard correctly) that the correlation surface is transformable into a surface of revolution by a slide and two stretches.

§ 7. *The Advantages of the Different Measures*

The two previous sections are preliminary to the main topic which forms the title of this section.

I shall first compare the conventional measure, the Pearson Coefficient, with the Median Ratio and later deal very briefly with some of the other measures.

The main desiderata in any measure are that it measure some real fact and that this fact be important. Other desiderata in the case of a measure of relationship are that the measure be comparable with other measures of other relationships, that it be conveniently and easily calculated and that it diverge little from the corresponding measure of the total series from a random sampling of which it is calculated. These desiderata we will consider in the above order.

Reality.

The Median Ratio is a clear statement of a real fact, an observed

relationship, such that the number of relationships closer than it equals the number less close. It gives the amount of y 's difference from its central tendency implied by such difference in x for this mid-case.

The Pearson r is not an observed relationship but a measure inferred from certain features of the observed relationships on the basis of certain presuppositions about them and the distribution of the facts from which they come. It is of course real in the sense of being the most probable real central tendency of the relationships if these various presuppositions are true, but in fact they never are except by chance more than approximately true, and in the majority of the cases in which students of the mental and social sciences need to measure relationships, they are far from true.

The 'regression,' that is the relation between actual amounts of y and actual amounts of x , is the reality at the basis of all measures of the relationship. The Median Ratio expresses it directly. It can be ascertained from the Pearson r only indirectly and on the hypothesis that certain very questionable conditions are realized.

Importance of the Fact Measured.

There is no great advantage either way in this respect. Neither the Pearson Coefficient nor the Median Ratio gives the entire fact of the relationship. Only the total distribution of the relationship does that. For 'normal' correlation where the relationship is the same regardless of the amount of x and where all of the arrays are distributed in normal surfaces of frequency the Pearson Coefficient and the Median Ratio both give the central tendency of the relationship. In other cases than this the Median Ratio is a trifle more important because less misleading and because it is nearer the modal relationship if the distribution of the relationship is skewed.

It is also worthy of note that our thinking about relationships should for practical reasons usually be in terms of the actual y/x or x/y ratios, that is the 'regressions,' since what we usually need to know is the implication of some actual deviation of one concerning the related deviation of the other. It seems better then to calculate the y/x or x/y ratio directly and when necessary to infer the r (that is the ratio when both traits are reduced to an equivalence in variability and the correlation table is one of double entry) rather than to calculate the r and infer the y/x or x/y ratio.

Comparability.

To compare the relationship between A and B with that between C and D adequately, we must compare the total distribution of the relationship $A-B$ with the total distribution of the relationship $C-D$. The Pearson Coefficients of $A-B$ and $C-D$ are per-

fectly fit to compare only when the form of distribution of the relationship $A - B$ is the same as that of the relationship $C - D$. So also of the median B/A and median D/C , or of the median A/B and median C/D , or of any measure of the central tendency of relationship which may be inferred from them. In so far as what we wish to compare is the modal relationship, however, there is a smaller error as a rule in inferring from the comparison of the Median Ratios of unlike distributions of relationships than in inferring from the comparison of their Pearson Coefficients.

Convenience of Calculation.

Provided the original measures are on a sufficiently fine scale, as they ought for every reason to be where relationships are to be measured by a Pearson Coefficient or a Median Ratio or a Modal Ratio, the Median Ratio is of course far more convenient than the Pearson Coefficient. Once a correlation table is written out the Median Ratios can be obtained with very little computation or eye strain. Inspection of the correlation table will tell about what they will be and only a few of the ratios will need to be ranged in order. I append a sample calculation (Fig. 8).

(1) First one makes an exact median sectioning of the x 's and the y 's and then counts the cases that give negative ratios.

By inspection one then chooses for the y/x ratios an approximate median (here of about .25) and for convenience draws a line to include these cases and counts them. One then increases their number by adding the cases of the next smallest ratios not included or by taking away the cases of the largest ratios included until one reaches the Median Ratio (here .333). One then repeats the process of guessing at an approximate median for the y/x ratios and correcting it.

In making comparisons on the basis of the median ratios we must of course bear in mind the variabilities of our A , B , C and D . In the Pearson Coefficients the series concerned are reduced to an equivalence in variability in the process of calculation. With the Median Ratios, if we wish to make this reduction to terms of the variability as a unit we must do it as a separate operation. For instance let A , B , C and D be series with variabilities 1, 2, 4 and 5. If then the Median Ratios found are

$$B/A = 1.00, \quad A/B = .25, \quad D/C = .625 \quad \text{and} \quad C/D = .40,$$

the Median Ratios that would be found if the differences in variability were eliminated would be $B/A \div 2/1$, $A/B \div 1/2$, etc., that is .50, .50, .50 and .50. If we wish to compare the mutual implication of A and B with the mutual implication of C and D we must go further still and combine the median $\frac{B}{A} \cdot \frac{\text{var. } A}{\text{var. } B}$ with the median

$$N = 1308 \quad \frac{1}{2}N = 654$$

$$\text{Neg. } (< 0) = 526$$

$$\frac{y}{x} \quad \begin{matrix} < 5/21 = 104 \\ 2/5 \quad 5/11 \quad 3/11 = 15 \end{matrix}$$

$$5/17 = 5$$

$$1/3 = 51$$

$$\text{Median } y/x = .333$$

$$\frac{x}{y} \quad \begin{matrix} < 7/13 = 100 \\ 1/3 = 46 \end{matrix}$$

$$\text{Median } x/y = .333$$

FIG. 8.

$\frac{A}{B} \cdot \frac{\text{var. } B}{\text{var. } A}$ and similarly the median $\frac{D}{C} \cdot \frac{\text{var. } C}{\text{var. } D}$ with the median $\frac{C}{D} \cdot \frac{\text{var. } D}{\text{var. } C}$. This combination would be made by taking the median of both the $\frac{B}{A} \cdot \frac{\text{var. } A}{\text{var. } B}$ and $\frac{A}{B} \cdot \frac{\text{var. } B}{\text{var. } A}$ ratios, an equivalent of the double entering involved in the Pearson Coefficient, or more easily still by taking
$$\frac{\frac{B}{A} \cdot \frac{\text{var. } A}{\text{var. } B} + \frac{A}{B} \cdot \frac{\text{var. } B}{\text{var. } A}}{2}.$$

Comparison is thus more awkward with the Median Ratios than with the Pearson Coefficients, because the latter method automatically both divides through by the variability and gets a measure of mutual implication. The superiority of the Pearson Coefficients is to some extent specious for it makes comparison easy not by removing difficulties but by presupposing that they do not exist. The obvious additional steps needed in the case of comparison of Median Ratios witness and emphasize the hypotheses on the basis of which we do compare. They may also prevent us from inadequate comparison. For instance from the facts that the Pearson r for adult brother's longevity with adult brother's longevity is .2853 and that the Pearson r for stature with left middle finger length is .6608, we have no right to conclude that the latter relationship is 2.3 times as close. Any one who will study the individual relationships in these two cases¹ will see that no single ratio can express the comparison of the two relationships.

Speed of Calculation.

Once the correlation-table is written out the Median Ratio can be calculated in from one tenth to one hundredth of the time taken for the Pearson Coefficient.

Divergence of Results Obtained from a Partial Sampling from the Results from the Entire Series Sampled.

The Pearson Coefficient is for normal correlation by the theory of error the more reliable. Whether in the actual cases of relationship with which we work, where the distributions and correlations are not exactly normal and where the theory of error does not apply without modification, it is more reliable, is a matter to be determined. Its use of the exact amount of every case of the relationships makes for superior reliability, but its weighting of extreme cases may somewhat counterbalance this.

The reason given by Professor Pearson for replacing Galton's method of obtaining the Median Ratio by this product-moment method was this superior reliability. No other reason has so far as I am aware ever been advanced. It is doubtful if Professor Pearson now would lay so much stress on greater reliability in the case of normal correlation of normal distributions, since he has so emphatically shown the rarity of both of these, and has been at some pains to test empirically certain measures which are valid regardless of the normality of distribution of the two facts.

Since in almost every other respect the Median Ratio is a more advantageous measure, it seems worth while to determine empirically, for some typical relationships, the comparative freedom from

¹ See *Biometrika*, Vol. I., p. 84, and Vol. I., p. 216.

chance error of it and of the Pearson Coefficient. I have also tested the influence of the number of cases on the per cent. of unlike-signed pairs (which I have called U) because at least for preliminary investigations of mental and social relationships the formula $r = \cosine \pi U$ (where U = the per cent. of unlike-signed pairs, deviations being calculated from an exact median sectioning, with no zero deviations) will often possess great advantages.

The accuracy with which the Pearson r , the Median Ratio and the cosine πU calculated from a random sampling of a series of individual relationships approximate the true r , the true Median Ratio, and the true cosine πU of the entire series was experimentally determined in the case of the series A , B and C (shown in Tables IX., XII. and XIII.).¹ These reliabilities could, I suppose, be calculated by theory for any given series of relationships but it seemed wise to determine them also by experiments with real cases. In calculating the results for each draw of 200, 100 or of 50 cases the deviations were reckoned always from the *true* central tendencies of the total series, not from the obtained central tendencies of the draw itself. This saves much time and introduces no error relevant to the problem. The Median Ratio was taken simply as the observed ratio of which it was a case. That is, if the distribution of ratios was:

Less than 1.00 — 49
 1.00 — 12
 over 1.00 — 39,

the Median Ratio would be taken as 1.00. If one took as the Median Ratio the average of this observed ratio and the ratio halfway between the 40 and 60 percentiles, the divergences for the Median Ratio would be reduced. The results are given in Table XIV. In every case the Median Ratio means the median of all the ratios (y/x and x/y), the two series being reduced to an equivalence in variability.

The relationships as calculated from the entire series are:

	Series A	Series B	Series C
Pearson Coefficient	.51	.27	.73
Median Ratio	.60	.33	.83
Cosine πU	.61	.30	.79

It is clear from Table XIV. that if N is as great as 100, there is no great loss in precision from the use of the Median Ratio method or even of the unlike-signed pairs method.

¹Table IX. is on page 19.

TABLE XIV.

AVERAGE DIVERGENCE OF OBTAINED FROM TRUE MEASURE OF RELATIONSHIP¹

(Figures in parentheses give the ranks of the three methods in freedom from chance error.)

	No. of Trials	No. of Cases	Pearson Coefficient	Median Ratio (Double-entry)	Cosine πU
Series A	10	200	.039 (1)	.053 (2)	.058 (3)
	10	100	.065 (2)	.062 (1)	.101 (3)
	10	50	.100 (1)	.155 (3)	.135 (2)
Series B	5	200	.064 (2)	.063 (1)	.082 (3)
	5	100	.105 (3)	.072 (1)	.075 (2)
	10	50	.153 (1)	.192 (2)	.197 (3)
Series C	3	200	.044 (2)	.072 (3)	.013 (1)
	3	100	.032 (1-2)	.050 (3)	.032 (1-2)
	5	50	.119 (2)	.120 (3)	.077 (1)

The Advantages of Certain Other Measures.

The Average Ratio has no advantage over the Median Ratio and suffers from the disadvantage of taking an enormous amount of time and being influenced so much by extreme ratios. No experienced worker with relationships would favor its use.

The Modal Ratio is in some respects the most important single feature of the entire series of relationships, and is probably a better basis of comparison between different relationships when either is not normally distributed than the Pearson Coefficient or the Median Ratio. The observed Modal Ratio from a small sampling diverges so much from the true Modal Ratio of the total series, however, that it can not be well used alone unless the number of ratios is 500 or more. The scale should also be fine. The most probable true Modal Ratio inferred from a large part of the total distribution of

¹ It is hardly worth while to compare the empirical divergences of Table XIV. for the Pearson Coefficients with the divergences to be expected from the formula A.D. $\text{true } r - \text{obtained } r = \frac{.7979(1-r^2)}{\sqrt{n}}$, for this formula, calculated for 'normal' correlation, would not be expected to fit very closely any of the three sets, A, B and C, or to fit C at all closely. A certain interest does attach to the comparison from the fact that the formula A.D. $\text{true } r - \text{obtained } r = \frac{.7979(1-r^2)}{\sqrt{n(1+r^2)}}$ has also been proposed as the valid one. So far as my drawings go, the former is surely the better. They vary from it, moreover, with a constant deviation toward a larger divergence, the divergences by theory being:

Series A	Series B	Series C
.042	.053	.027
.059	.074	.038
.083	.105	.054

the relationship is a very valuable measure but one the calculation of which takes a long time and involves presuppositions about the form of distribution of the relationship.

In all cases the investigator of a relationship should be observant of the form of distribution of the individual relationships and of their approximate mode. Where the correlation table shows any marked eccentricity in the distribution of the relationships the observed modal relationship at least should probably be stated, even though the more reliable Median Ratio or Pearson Coefficient has been calculated.

The correlation (in the sense of the slope of the line which the Pearson Coefficient measures) may be inferred from the frequencies of certain types of pairs, as in the case, $r = \cos. \pi U$ (U equalling the percentage of unlike-signed pairs with median sectioning).

The methods of making this inference are especially valuable when we wish to compare two relationships, one (or both) of which is measured very crudely, for instance, the relation between health and cheerfulness and the relation between intellect and morality. From such measures as the following:

		Health			Intellect	
		Sickly	Healthy		Dull	Bright
Cheerfulness	Much	150	150	Morality	315	285
	Little	250	450		145	265
					Inferior	Superior

one can not compare directly the closeness of relationship of health and cheerfulness with that of intellect and morality.

The following formulas, suggested by Pearson, are probably the best available for dealing with such cases. In all $N =$ the total

FIG. 9.

number of pairs; a , b , c and d mean respectively the numbers of x_1y_1 , x_2y_1 , x_1y_2 and x_2y_2 pairs where x_1 means measures above any given degree of x and x_2 , measures below it, and similarly for y_1 and y_2 (see Fig. 9).

$$\text{I. } r = \sin \frac{\pi}{2} \frac{1}{\sqrt{1+k^2}} \quad \text{where } k^2 = \frac{4abcdN^2}{(ad-bc)^2(a+d)(b+c)}.$$

$$\text{II. } r = \sin \frac{\pi}{2} \frac{1}{1 + \frac{2bcN}{(ad-bc)(b+c)}}$$

cases being so chosen that $ad > bc$.

$$\text{III. } r = \sin \frac{\pi}{2} \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}.$$

$$\text{IV. } \frac{ad-bc}{N^2HK} = r + \frac{r^2}{2}hk + \frac{r^3}{6}(h^2-1)(k^2-1) + \frac{r^4}{24}h(h^2-3)k(k^2-3), \text{ etc.}$$

Since

$$\frac{(a+c) - (b+d)}{N} = \sqrt{\frac{2}{\pi}} \int_0^h e^{-\frac{1}{2}x^2} dx$$

and

$$\frac{(a+b) - (c+d)}{N} = \sqrt{\frac{2}{\pi}} \int_0^k e^{-\frac{1}{2}y^2} dy,$$

h and k are found from tables of the probability integral, a , b , c and d being known.

$$H \text{ is taken as } \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}h^2}$$

$$K \text{ " " " } \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}k^2},$$

H and K are thus found from tables.

Of these formulas IV. is for 'normal' correlation the most accurate. It presupposes 'normal' correlation; I., II. and III. do not.

When the facts to be related are measured on a fine scale but in terms of relative position only, not of amount, the relationship may be measured, as Spearman has shown, by the degree of conformity of the second member's position to that of the first member.¹ This method suffers from the disadvantage of giving results only with much difficulty comparable with other methods and of taking much more time without being much more reliable than the cosine πU method.

From the reduction in variability of an array of y related to a given value of x below the variability of the total series of y , the correlation may be inferred on the supposition that the correlation is 'normal' and that the variabilities of all arrays of y are equal.

The infrequency of 'normal' correlation and the fact that, as

¹ See *American Journal of Psychology*, Vol. XV., p. 86 ff.

shown in § 4, the variabilities of all arrays of y are usually *not* equal make this method of no great practical service except for the few cases where no better method can be used.¹

Section 4 tested the hypothesis of equal variability of all arrays of y and found it true in some cases and false for others. It is somewhat extraordinary that Professor Pearson should in support of his coefficient of variability argue that the gross variability depends on the size of the mean from which the variability is measured, being proportioned to it, and yet not recognize that, since the means of the arrays of y in positive correlation would then increase as we pass from arrays related to low values of x to arrays related to high values of x , the variability of one of the latter arrays should be greater than that of one of the former.

§ 8. *The Attenuation of Measurements of Relationship*

Attenuation

Chance inaccuracies in the original measures make the relationship obtained therefrom vary toward zero from the relationship that would be found with accurate measures. C. Spearman announced in the *American Journal of Psychology*, Vol. XV., pp. 89-91, that the following formulas gave the necessary correction:

$$(1) \quad r_{pq} = \frac{r_{p'q'}}{\sqrt{r_{p'p'} \cdot r_{q'q'}}$$

where $r_{p'q'}$ = the mean of the correlations between each series of values obtained for p with each series obtained for q ;

$r_{p'p'}$ = the average correlation between one and another of these several independently obtained series of values of p ;

$r_{q'q'}$ = the same as regards q ;

and r_{pq} = the required real correlation between the true objective values of p and q .

$$(2) \quad r_{pq} = \frac{\sqrt[m \cdot n]{m \cdot n r_{p'q'} - r_{p'q'}}}{\sqrt[m \cdot n]{m \cdot n - 1}}$$

where m and n = the number of independent gradings for p and q respectively;

¹ Cases, that is, where we know the variability of a related array but lack the data needed for the use of the better methods. For instance, we may find the variability of 100 men eminent in engineering science in early liking for arithmetic to be only 30 per cent. as great as the variability of men in general and so infer the amount of relationship between early liking of arithmetic and engineering ability. The actual rating of a random sampling of men in both early liking for arithmetic and engineering ability would be hardly possible.

$r_{p'q'}$ = the mean correlation between the various gradings for p and those for q ;

and $r_{p''q''}$ = the correlation of the amalgamated series for p with the amalgamated series for q .

He has been criticized with some venom by Karl Pearson (*Biometrika*, Vol. III., p. 160), who believes these formulas wrong, and concludes that "Perhaps the best thing at present would be for Mr. Spearman to write a paper giving algebraical proofs of all the formulas he has used, and if he did not discover their erroneous character in the process, he would at least provide tangible material for definite criticism, which it is difficult to apply to mere unproven assertions."

These formulas of "Spearman's, if correct, are of importance. They should be proved valid or replaced by formulas that are valid.

The first formula may be replaced by

Formula. (3)
$$r_{pq} = r_{p'q'} \frac{\sigma_{p'}\sigma_{q'}}{\sqrt{\sigma_{p'}^2 - \sigma_{ep'}^2} \sqrt{\sigma_{q'}^2 - \sigma_{eq'}^2}}$$

where r_{pq} and $r_{p'q'}$ are as above and

$\sigma_{p'}$ = the mean square deviation of the series of measures of p ;

$\sigma_{q'}$ = the mean square deviation of the series of measures of q ;

$\sigma_{ep'}$ = the mean square deviation of the different measures of p in the same individuals;

$\sigma_{eq'}$ = the mean square deviation of the different measures of q in the same individuals.¹

The presupposition of this formula and of Spearman's first formula is, that the attenuation is due to chance errors. Dr. Clark Wissler has called attention to the fact that, where practise, fatigue and other constant influences help to cause the different observations of a fact to vary, these formulas will, therefore, give inaccurate results.²

(1) & (3)

(3) Of these two formulas, Spearman's possesses the advantage of being usable in cases where the two traits are not measured in units of amount, such as allow the variabilities of the two traits to be calculated; the formula of Boas has the advantage of being more rapid and convenient in cases where the variabilities of the two traits can be calculated.

(2) No active attention has so far as the writer knows been yet given to formula (2) above.³ Practical necessity seems to justify the labor

¹ This formula is due to Professor Franz Boas. See also the note by Dr. C. Wissler in *Science*, Vol. XXII., p. 309 ff.

² *Loc. cit.* in note 1.

³ Spearman's second formula has the advantage of measuring the probable true correlation by the actual changes produced in the obtained 'raw' correlation by a certain increase in accuracy. The nature and validity of the presuppositions upon which it is based I am not competent to discuss.

of testing it (and in a measure the first formula also) inductively. This I have done to some extent for values of r where the r 's from accurate measures are from .70 to .80 in connection with my 'Measurements of Twins' (*Archives of Philosophy, Psychology and Scientific Methods*, No. 1, September, 1905).

I had records from 50 pairs of twins in 5 tests of efficiency of perception; (1) in marking A's on a sheet of printed capitals, (2) in marking A's on a second sheet of printed capitals, (3) in marking words containing e and r on a page of Spanish, (4) in marking words containing a and t on a page of Spanish and (5) in marking misspelled words on a page of narrative, 100 of whose words were misspelled. I had also 6 tests in efficiency of controlled association, tests 6 and 7 being addition, 8 and 9 being multiplication and 10 and 11 being writing the opposites of two lists of words.

If we combine all 5 of the tests of efficiency of perception allowing approximately equal weight to each, we have a measure which is presumably close to the true measure of a child's capacity at a certain day and hour to pick out small details efficiently. The correlation between twin and twin is for this combined score .697. Similarly the combined measure for addition, multiplication and opposites gives a measure presumably close to the true measure of a child's ability at a certain day and hour to make proper mental connections. The correlation between twin and twin is .815. The .697 and .815 are presumably only slightly below the true r 's.

Now the correlations for twin and twin in tests 1-11 were in order .607, .633, .595, .428, .754, .645, .644, .653, .579, .734 and .560.

Subjecting these values to correction by Spearman's formulas, taking, as he does, the mean of both corrected r 's I obtained for the perception tests: Marking A's, true r = .69; marking letters in words, true r = .71; misspelled words, not corrected because only one test was given. The Spearman correction thus produced results in accord with the expectation derived from the value r = .697 for the combined mark. For the association tests I obtained after correction: Addition, true r = .75; multiplication, true r = .84; opposites, true r = .90. The average of these, .83, is again closely in accord with the .815 from the combined measure. In both cases the result by correction is slightly higher than the result empirically obtained from the more accurate data, as of course it should be.

I have made a test *ad hoc* in the case of a series of 100 pairs drawn at random from Series B which give a true r of .281. These 100 pairs of accurate measures I made inaccurate artificially. I then calculated the r 's obtained from such inaccurate measures, applied the Spearman formulas and in so far tested their validity.

Special precautions were taken to have the errors artificially induced in the 200 measures such as would come in reality from variable errors of apparatus, observation and record. The errors were in fact a random sampling of the errors actually made by a psychologist in estimating areas. A series of 121 rectangles of approximately the same shape, 40, 41, 42 . . . 160 sq. cm. with also many duplicates were used. The area of each was estimated, the slips being drawn in a random order, and the error + or — from the true area was recorded. The errors used by me were those made after from 3 to 5 trials with the series and were little influenced by practice (the sums of the errors regardless of signs were for successive repetitions of the series 605, 614, 563, 613, 587, 637, 531, 542, 578, 581). I used the deviation from the standard if the constant error for the given area was less than 1 sq. cm. and the approximate deviation from the subject's own average judgment if the constant error was over 1 sq. cm. The errors taken were those (10 in each case) made with areas 43 sq. cm. up through 122 sq. cm., four errors being taken for each of the 200 accurate measures. These errors were assigned to the accurate measures so that the magnitude of the area with which the error of estimation was made corresponded roughly to the magnitude of the measure to which the error was assigned. Thus errors from areas 43–53 would be put with measures — 27, — 25, — 23 and the like, and errors from areas 110–122 would be put with measures + 17, + 19, + 27 and the like. The true measures and the errors assigned to each are given in Table XV.

If now to each true measure is added (regarding signs) its assigned error, we have (four errors having been assigned to each) four series of inaccurate measures of two series whose true values and true correlation are known. These facts give the data for testing the Spearman formulas.¹

¹ These errors can of course be used with any series of 400 or less measures to test Spearman's formulæ, as I have done for this series ($r = .281$ of Series B).

TABLE XV.

True x					Errors Assigned					True y					Errors Assigned				
a	-19	-3	+2	-4	-2					a	-11	-2	-4	-4	+5				
b	-17	-2	+5	-1	-4					b	-1	+12	-13	-9	-7				
c	-15	+9	+4	0	+1					c	-27	+7	+2	-4	-3				
d	-15	+6	-2	-2	-3					d	-9	0	+3	-8	0				
etc.	-15	-3	-1	-1	-3					etc.	+3	+7	+7	-2	-3				
	-15	-8	+1	+2	+2						+7	+3	+10	-7	+7				
	-13	+1	+6	-2	-3						-11	+6	-5	+9	0				
	-13	-8	+1	-1	+4						-3	+4	+3	+3	+6				
	-13	+1	+2	+1	+3						+13	+1	-1	-1	-6				
	-11	+1	-5	+11	-4						-13	-4	-3	+7	-5				
	-11	-2	0	-6	+4						-5	+7	-4	+2	+4				
	-11	0	-2	0	+7						-5	+7	-1	0	-6				
	-11	-12	-1	0	-3						-3	-1	-6	-8	-9				
	-9	-9	+7	+9	0						-5	+3	0	0	+9				
	-9	+13	+8	+3	-12						-3	+1	-1	+1	+6				
	-9	+2	+1	-3	+2						-1	-9	+6	-5	-3				
	-9	-4	-1	-4	-3						-1	-5	+6	+3	-6				
	-9	-6	+6	+4	+12						+5	-2	+3	-6	+2				
	-9	-3	-2	-2	-3						+13	-8	-1	+6	+4				
	-7	0	+2	-2	+5						-1	+5	-5	+1	-5				
	-7	-2	-6	+3	+2						+7	+1	+2	-4	+3				
	-7	+7	+4	+5	-2						+9	+2	-5	-3	-5				
	-5	-6	-2	-2	+3						-13	+1	+1	+4	+6				
	-5	+13	-4	-4	+3						-9	-6	0	+3	-3				
	-5	+8	-3	-7	0						-7	-6	+2	-2	+2				
	-5	-3	+9	-4	-3						-7	+1	+3	-5	+2				
	-5	-7	-3	-5	+2						-3	-2	+6	-2	-7				
	-5	+7	+2	0	-1						-3	+7	+1	+4	-2				
	-5	-1	+4	0	-2						-3	+5	+10	-5	+4				
	-5	+4	-2	+4	-3						-3	-14	+6	-3	-5				
	-5	+1	-3	-3	-3						-3	-1	-11	-3	-2				
	-5	+6	+3	-3	-9						+13	-3	-4	+3	+11				
	-3	+3	+7	+6	-5						-11	+11	+7	-2	-11				
	-3	+7	+3	+6	-2						-9	+5	-2	-3	+3				
	-3	-1	-7	-2	-7						-9	-6	-3	-5	+5				
	-3	+3	-9	+3	-2						-7	+7	-4	+8	+7				
	-3	+3	-4	+3	-3						-5	+4	+4	-1	+3				
	-3	0	+2	+5	0						-5	+6	-5	-4	+1				
	-3	-1	-12	-5	+6						-5	+5	-9	+5	-3				
	-3	-10	+2	+9	+2						-5	0	0	-5	+3				
	-3	+1	-5	+4	-5						-1	0	+4	-1	-3				
	-3	+1	+1	0	-8						+1	-11	-5	+12	+7				
	-3	+5	+5	-6	0						+1	-2	-12	-10	+6				
	-3	0	-3	+3	0						+1	+13	-2	+6	-8				
	-3	+3	+5	+3	-1						+5	-4	+11	-1	-6				
	-3	-11	+6	+4	-9						+9	+5	+7	-1	-11				
	-1	+2	-6	+13	-1						-9	-6	-2	-1	-1				
	-1	-4	+8	+7	+4						-5	-2	+4	-7	+1				
	-1	+1	-8	-4	-7						-3	+4	+7	-2	-3				
	-1	+2	+8	-2	+2						-3	-1	0	+12	+11				

TABLE XV. (continued)

True x	λ_a	λ_b	λ_c	λ_d	True y	y_a	y_b	y_c	y_d
	Errors Assigned					Errors Assigned			
-1	-9	+9	-1	0	+1	-5	+5	+9	+3
-1	+8	-5	+5	0	+3	0	+1	+1	0
-1	+1	-5	-5	-1	+5	-3	+1	+2	-3
-1	+2	+7	-1	+4	+7	+3	-2	+14	-6
+1	+3	-5	-6	0	+17	0	+4	-1	-8
+1	-6	+3	-15	-2	+11	-13	+7	+4	+1
+1	+1	-2	+6	+1	+11	+2	-4	+4	+12
+1	+6	+8	+2	0	+5	+7	+5	+4	-7
+1	+1	0	+2	0	+1	-7	-10	-9	+8
+1	+1	+3	+2	-8	+1	+2	+6	-9	-3
+1	+17	+1	-9	-2	+1	+5	0	+3	+7
+1	-3	+4	-5	-6	-3	-9	-1	+4	-6
+1	0	+7	+2	+4	-5	+3	+7	-1	+6
+1	-2	+1	+7	-9	-7	+11	+3	-2	-3
+1	+3	-6	+4	-1	-9	-1	-7	-2	-3
+1	0	+3	-5	-4	-9	+3	-1	+10	-8
+1	+5	+1	+5	+4	-13	-6	+6	+6	-1
+3	-3	-9	-3	+6	+1	-4	+3	-3	0
+3	+1	+3	+1	-1	+1	0	-3	-3	-2
+3	-2	-3	-2	-2	-1	+4	0	0	+7
+3	+7	0	+3	+5	-5	+1	-2	+3	-2
+5	+3	+2	+4	-1	+11	+1	-1	+1	+2
+5	+5	+3	+5	-11	+5	+8	-2	-12	-4
+5	+3	+2	0	-9	+1	+7	-3	+16	+4
+5	-4	-8	-4	+3	+1	-1	-2	-2	-6
+5	+10	+9	-2	+3	-7	-2	-7	+6	+5
+5	-5	-5	+4	+3	-9	+1	-3	-4	+1
+5	+10	-6	-8	+12	-19	+2	+5	+1	-4
+7	-6	-4	+2	-4	+11	-4	-3	-8	-5
+7	+5	+10	-5	-3	+7	+7	+15	-7	-7
+7	-2	-7	+2	-5	+5	-3	-3	-3	+14
+7	+7	+1	+15	-1	+1	+2	+1	-4	-4
+7	-8	+6	-6	+10	-3	-2	+7	+3	-11
+7	+8	-7	-10	-8	-5	+7	-2	-6	-4
+7	+7	+6	-3	-4	-7	+5	-8	-5	+8
+9	-10	-3	+1	+7	+9	0	-3	+1	+5
+9	+7	+15	-7	-7	+7	+3	-7	+6	+6
+9	+3	-7	+6	+6	+3	+10	-3	0	-4
+9	-10	-9	+6	-6	+1	+1	-13	-2	-3
+9	+7	+6	-3	-4	-5	+7	+6	+5	-13
+9	-10	-3	+1	+7	-7	-1	+1	-6	+4
+11	-1	0	0	+4	+5	-13	+4	-6	+6
+13	+10	-2	0	+11	+13	-3	-6	+4	0
+13	-16	-5	-9	+11	-11	+5	-11	0	+9
+15	+3	-6	+1	-5	+1	0	+5	+1	0
+17	-1	+3	-4	+8	+13	-1	-6	+2	0
+17	+2	+5	-3	-4	+11	+7	-9	-1	+14
+17	-3	+6	+5	-11	+1	0	+7	-4	-15
+17	+3	-3	+3	-1	+1	+3	-2	+5	-3
+25	+2	-1	+2	+2	+7	-10	-9	+6	-6

Let us call the four series of inaccurate measures obtained with the four errors, Xa , Xb , Xc , Xd , and Ya , Yb , Yc , Yd .

Call the series obtained by averaging each member of Xa with its correspondent in Xb , Xab .

Let Xcd , Yab and Ycd have similar meanings.

Call the series obtained by averaging each member of Xa with the corresponding Xb , Xc and Xd , $Xabcd$.

Let $Yabcd$ have a similar meaning.

We have then 4 very inaccurate measures of X in every one of the 100 pairs; so also of Y . We have two less inaccurate measures of X and also of Y in each pair. We have one still better measure, the best obtainable from our data.

We may then calculate the corrected r according to Spearman, using many different combinations of the r 's obtained from the above series. The combinations which I have used and the results follow in Table XVI.¹

The correspondence of the coefficients corrected by Spearman's formulas with the actual coefficient from accurate measures is satisfactory.

TABLE XVI.

r_{xa} with ya	= .142	r_{xa} with xb	= .731	r_1	= .239	} Form
r_{xb} with yb	= .208	r_{ya} with yb	= .468	$\sqrt{r_{xa} w. xb \ r_{ya} w. yb}$		
r_{xa} with yb	= .083			r_2	= .277	} Form
r_{xb} with ya	= .243			$\sqrt{r_{xab} w. xcd \ r_{yab} w. ycd}$		
r_1 (the average of the four)	= .169			$\sqrt[4]{2.2(r_{xab} w. yab) - r_1}$	= .313	} Form
r_{xab} with yab	= .212	r_{xab} with xcd	= .803	$\sqrt[4]{2.2(r_2) - r_2}$	= .380	
r_{xcd} with yed	= .221	r_{yab} with yed	= .717	Average by all formulæ	.315	
r_{xab} with yed	= .239			Median by all formulæ	.301	
r_{xcd} with yab	= .170			True relationship	.281	
r_2 (the average of the four)	= .2105					
r_{xabcd} with $yabcd$	= .260					
i. e.						
r_3	= .260					

§ 9. Minor Advice to Students of Mental and Social Relationships

As a rule nothing should be taken for granted about any relationship and the result of any calculation should be to express, not to replace, the comprehension of a fact about the series of individual relationships.

¹In all the calculations I have assumed the original 0 as the central tendency from which to reckon deviation values. To have turned each of the 200 values of each of the fourteen series into a new deviation measure would have added practically nothing to the general result in the way of accuracy. The labor of 2,800 little sums in addition and 2,800 copyings of numbers could be more profitably spent. My figures are on this basis.

Measurements should be on the finest scale that can be recorded without special difficulty. The attenuation by chance error is thus diminished and the time taken in making a more elaborate correlation table can be saved ten times over by the use of the Median Ratio.

The central tendency from which one measures deviations should be chosen with care so that it stands for some reality divergence from which is significant.

In the relationship given in Table XVII.,¹ for instance, from what point should one reckon deviations? The authors take the mean, 56.568. But there is much to be said for taking the modal adult life (at about 70), since that represents an important real tendency and the force of heredity in determining departures from that tendency is perhaps more important than its form in determining departures from the rather arbitrary age, 56.568. The re-

TABLE XVII.

	23	28	33	38	43	48	53	58	63	68	73	78	83	88	93	98	Totals of Arrays
23	10	20	8	14	9	8	5	10	4	15	11	6	7	2			133
28	20	18	15	6	9	13	8	4	3	7	5	6	1	9	2		126
33	8	15	18	12	14	8	8	9	3	11	8	3	10	7	2	1	137
38	14	6	12	12	8	11	9	4	2	11	10	15	5	6	2		127
43	9	9	14	8	8	8	13	5	3	7	12	6	8	3	2		115
48	8	13	8	11	8	16	6	11	4	17	11	6	9	7	2		137
53	5	8	8	9	13	6	8	7	3	6	9	11	10	9	3	1	116
58	10	4	9	4	5	11	7	5	3	8	15	4	12	7	2	1	107
	4	3	3	2	3	4	3	3	1	3	6	4	7	4	1	1	52
63	4	7	11	11	7	17	6	8	3	16	18	22	11	10	3		154
68	15	5	8	10	12	11	9	15	6	18	28	31	19	12	9	4	212
73	11	6	3	15	6	6	11	4	4	22	31	40	16	13	3	1	193
78	6	1	10	5	8	9	10	12	7	11	19	16	28	17	12	3	176
83	7	9	7	6	3	7	9	7	4	10	12	13	17	12	8	3	134
88	2	2	2	2	2	2	3	2	1	3	9	3	12	8	8	1	62
93								1	1		4	1	3	3			13
98		1					1				1	2		1			6
Medians of Arrays	49	43	47	51	52	55	57	59	62	65	67	68	65	73	74	76	

Pearson Coefficient = .2853, C.T. being 56.568.

Median Ratio = .479, C.T. being 59.4.

¹ The relationship between brother and brother in length of life in cases where both brothers are 21 or over, from 'The Inheritance of the Duration of Life' by M. Beeton and K. Pearson, *Biometrika*, Vol. I., p. 84. I have divided the array of 58 so as to make a median sectioning of the series. In the original the array for 58 is given simply as 14, 7, 12, 6, 8, 15, 10, 12, 11, 21, 8, 19, 11, 3, 2. I have also added approximate medians of arrays.

lationship in the latter case (70.5 being taken as the central tendency) is closer, the Median Ratio being .54 or about 7 higher than the Median Ratio when divergences are calculated from 56.5. The Modal Ratio is unchanged.

It should be evident from the facts stated in previous sections that it is out-and-out folly to be content with calculating for every relationship studied the same type of coefficient. Nothing short of the entire correlation table is the adequate measure of the relationship in question. Any measure of one central tendency of relationship may be misleading, for the relationship may be bimodal. When the observed modal relationship is clearly not near the Pearson Coefficient the latter should be accompanied by the former. So also if the modal relationship is clearly not near the median relationship.

The averages or medians or modes of the arrays should be cal-

FIG. 19.

culated and stated, and unless the relationship is uniform (within the limits of chance error) throughout the course of the series a most probable curved line to fit the entire series should be calculated instead of the slope of a straight line.

For instance, the Pearson Coefficient for the relation between adult brother and adult brother in longevity is given by Beeton and Pearson as .2853. The relation is sufficiently close to uniformity for all values of x to make a linear relation at least approximately true (if we consider also the similar relation between sister and sister). The relation is, however, by no means identical with other relations giving a similar coefficient, for the modal relationship is *approximately 1.00*. This can be seen at a glance from the graphic representation of the correlation table (Fig. 10) or the distribution of the ratios (deviations are reckoned from 59.4 and 59.4 as central tendencies) in Fig. 11.

The .2853 then does not mean that the most likely value of $B - \text{C.T. of } B$ is near $.2853 \times (A - \text{C.T. of } A)$, nor that the forces producing correlation tend to make $B/A = .2853$, divergencies from

FIG. 11. Frequencies of different degrees of relationship in the case of fraternal longevity. The numbers stand for the ratios in per cents, the heights for their relative frequency. The mode is at very, very close resemblance, or ratios of 90 to 115 per cent.

this being due to minor causes producing variations in the correlation. On the contrary the .2853 represents a most ambiguous summation of the force of a tendency to identical longevity and many other forces. If the authors had not given the full correlation table, the .2853 would evidently have been definitely misleading.

The determination of the most likely law of relationship for a series of pairs may then be theoretically and practically a different problem for each particular case, a problem to solve which we need not only certain mathematical technique but also abundant knowledge of other similar relationships and of the entire body of facts relevant to the relationship in question. Thus the same set of pairs could properly be interpreted on the basis of a linear relationship

when they were male brothers' first-rib lengths, and could not properly be so interpreted if they were related body-strengths and earnings in dollars. For we have evidence from cephalic index, stature and the like to justify some expectation of linear correlation for fraternal relationships in features of anatomy, whereas what evidence we have concerning the relationship between body-strength and earning capacity in individuals goes to show that it is far less close for those of high earning capacity than for those of very low earning capacity.

4591

UC SOUTHERN REGIONAL LIBRARY FACILITY

AA 000 868 824 4