'FOOTRULE' FOR MEASURING CORRELATION.

BY C. SPEARMAN.

§ 1. *Need of a very easy, but yet precise, method of measuring correlation. Estimate by intuition is vague and illusive.*

§ 2. *Necessity of comparison by rank, as absolute measurements are not properly comparable with one another.*

§ 3. *Existing methods of calculating correlation. The standard 'r' method. Various methods proposed by mathematicians and investigators respectively.*

§ 4. *Proposed easy 'R' method, with an actual illustration. Usage of the 'probable error.' The two kinds of probable errors.*

§ 5. *Relation of the R to the r method. Latter can be simplified, and then only differs from former in that certain data are squared. $r = \sin\left(\frac{\pi}{2}R\right)$. Evidence that psychological correlations approximate to one constant form, and therefore are really measurable by a single coefficient. Summary.*

Appendix. Mathematical proofs in support of the R method.

§ 1. *Need of a 'footrule' for correlation.*

NOWADAYS little apology is required for any attempt to improve the psychological methods of dealing with quantity. The investigation of quantity is not intended, as some would seem to imagine, to replace that of quality, but to complement it. Qualitative and quantitative phenomena are, as it were, the flesh and bones of psychology; if this branch of science is meaningless without fine qualitative self-observation, it is at least rickety and crank without a frame-work of hard quantitative facts.

The particular domain of quantitative fact to be considered here is one which has but recently been fully annexed by science, and which seems of an unparalleled extent and richness, only waiting to be exploited. Previously science had almost exclusively devoted itself to regularities or 'laws'; but lately it has been learning to deal even with tendency to regularity, or, as it is called, to measure *the degree of correlation* between any two things.

Now, every one knows that the length of a line may be estimated with very varying grades of exactitude; many purposes are served perfectly well by a single glance of the eye; at other times, on the contrary, we require tediously repeated observations with an elaborate micrometer; while between these two extremes, there are a great number of cases where the object in view is much best met by the simple application, say, of a footrule. In this paper, it is desired to emphasize the fact, that precisely the same holds good in correlation also; mere casual observation is enough, for instance, to reveal the general fact, that warm damp weather has some connection or other with the growth of fungi; on the other hand, certain correspondences between biological events are taxing the utmost resources of statistical mathematics; while here too, in correlation, there is a middle region, where we are much in need of a plain handy 'footrule.'

The want of some such convenient measure is indeed, I believe, largely responsible for the slow progress and innumerable apparent contradictions of psychological research. In the course of almost every investigation, or perusal of investigations, one has many times occasion to compare one series of figures with another, to see whether and how much they vary concurrently. And this comparison is still almost universally performed on the primitive plan of just inspecting the figures and forming a general impression. Hereby the mind is called upon to perform a feat intuitively which it is only fitted to execute discursively. If a person does not happen to be a mathematical prodigy, just looking over the figures will no more suffice for him to determine accurately a correlation, than it will enable him to effect any other similar quantitative operation, say, addition, multiplication, or division. And even if the mind did possess a fine general sense of correlation, the present resources of language could not bring it to expression; such vague generalities as 'large,' 'small,' 'medium,' etc. are far removed from the precise terms required for scientific usefulness.

Nor is this vagueness all, or perhaps even the worst, that is to be said against estimating correlation by the sole light of intuition. This sort of estimate is further exposed to various formidable illusions; dozens of instances could be quoted, where authors have published arrays of figures as being an irrefragable proof of correlation, whereas actual reckoning quickly shows none whatever to exist; and even more frequent is the opposite illusion, which has led many of our best and soundest workers to deny positively any correlation, whereas the

reckoning easily establishes one beyond all doubt. Some, but by no means all, of these illusions are dependent upon the manner in which the figures happen to be arranged. For instance, it has frequently occurred, that one author announces little or no correlation between two things, because he has found the ratio between them to vary irregularly from case to case; while another author declares the correlation between the same things to be very high, because he has been dealing with averages, in which of course individual fluctuations cancel one another.

And if the intuitive estimate breaks down when it only has to judge the actual proportionality between any given values, still far more incompetent is it to decide whether any actually observed proportionality should be taken as a *general* fact, and not a mere accidental coincidence between the particular cases given.

§ 2. *Comparison by rank.*

Attention must here be drawn to an important point, which does not seem to have hitherto met with sufficient regard. In order that a precise measure of correlation between two series of values may have full significance, these two species of values should be *properly comparable with one another*.

This requisite does not make itself felt as a difficulty, so long as the two species of values are homogeneous.

In physical and physiological matters this generally occurs, in that both values are usually spatial dimensions; such is the case, for instance, in the correlation between the lengths of the arm and the leg respectively, or in that between the length and breadth of the skull.

But the matter changes as soon as we turn to psychology. Suppose that we wanted to measure the correlation between the skin's 'spatial sense'¹ and its sensitivity to pain; we should measure both senses at a great many different places all over the body; the spatial sense would probably be measured by ascertaining the threshold at which two points could just be distinguished from one another, while the sensitivity to pain might be tested by determining the amount of pressure that just begins to be painful. But then arises the obvious question, how far are the variations of distance between the two points of the aesthesiometer legitimately comparable with the variations of pressure of the algometer? Unquestionably, indeed, the varying distance between the points is some mathematical function of the

¹ The German 'Raumsinn'.

varying spatial sense, and similarly as regards the pressure and the sensitivity to pain; but we can never hope to determine such a function, nor even to show that it is the same in both cases¹.

All that we can properly assert seems to be that a greater distance between the points implies a duller spatial sense, and a stronger pressure indicates less sensitivity to pain. On this assumption it is clear that, if the two senses were perfectly correlated with one another, that part of the skin which required the least distance between the points would also need the smallest pressure; the part that wanted the second smallest distance would also want the second smallest pressure, and so on. But when we confine ourselves to this sole legitimate assumption, we are doing neither more nor less than converting the experimentally obtained measurements from absolute quantities into *comparative ranks*.

The disadvantage of conversion into rank is that thereby a certain amount of the experimental information is sacrificed. We merely call *A* one rank higher than *B* and *B* one higher than *C*, although we may actually know that the intervals *A—B* and *B—C* are of different magnitudes. But the amount of sacrifice is only of the same order as occurs in the very general procedure of grouping the cases into classes; here the insignificant difference between the lowest value in one class and the highest value in the class just beneath is handled in quite the same way as the considerable difference between the highest value in the higher class and the lowest in the lower class. True enough, in grouping by classes the extent of any single such injustice cannot possibly exceed the range of two classes, whereas in using ranks there is no limit. But this point of view is based upon a wrong principle; in the matter of correlation we have not to consider the conceivable vagaries of individual cases, but the amount of disturbance actually to be feared according to the laws of probability in the finally deduced coefficient. From this truer point of view, theoretically it would not be hard to demonstrate that the sacrifice of information by using ranks must really be as insignificant as that by grouping into classes; and practically, experience has convinced me that the results obtained by using ranks, instead of being less constant and reliable, are on the contrary (owing to certain other moments²) even more so.

¹ Obviously, instead of obtaining 'thresholds,' we might from the same experimental data equally well derive 'obtuseness values' ('Stumpfheitswerte'), as was done in all extensive investigations of the Vierordt school. Thereupon all the proportions, and with these all the correlations, would become altered.

² Cf. *Amer. J. Psych.* 1904, Vol. XIII. pp. 79-82.

§ 3. *Existing methods of calculating correlation.*

Let us now consider for a moment how far our requirements are met by any of the methods of calculating correlation that are already known.

The present, as one may fairly call it, 'standard' method is the one chiefly elaborated by Bravais, Galton and Pearson. The total degree of proportionality between the two compared series of values is summed up in a numerical coefficient, symbolized by the letter r . When the proportionality is complete throughout, $r=1$; the more incomplete the general proportionality, the smaller becomes r ; and when there is no tendency whatever to proportionality, $r=0$. A further formula gives the 'probable error,' showing just how much the value of r may reasonably be expected to be disturbed by mere chance.

There is not the least difficulty in applying this r method to correlations by rank. I have worked out about 40 coefficients, first with absolute measurements, and then again after converting these into relative ranks. The mean value of correlation reckoned in the two respective ways turned out to be identical up to the 3rd decimal place inclusive. Of course under special circumstances the values arrived at in the two different ways might differ to a considerable extent; but then, in accordance with the preceding section, we must consider the value given by means of ranks as being the truer of the two; ranks will always express perfect correlation by $r=1$, as intended by the formula; whereas absolute measurements, if heterogeneous, will usually *not* do so. Even the formula for the probable error does not seem to become appreciably inaccurate when ranks are used, except in the case of a very short series and a very high correlation (say, over 0.90), when the value given by the formula becomes much too small.

The great objection to this r method is that it involves very laborious calculations. Despite all other merits, this quite unfits it for the 'footrule,' of which we are now in quest.

The same objection applies, of course, even more forcibly to the ingenious complications introduced by Pearson, in order to measure 'skew' correlation.

Some continental mathematicians have recently endeavoured to express correlations more accurately by means of a series of terms, instead of a single one. Not only is the calculation more tedious than ever, but even the results are of very doubtful serviceability in actual research.

Most of the other hitherto proposed methods of reckoning correlation have been devised on emergency in the course of particular investigations. These rarely at all answer the needs of a scientific method. Even when the correlation has been summed up in a definite expression, this has usually had no upper limit marking perfect correspondence; or, on the other hand, it has had no lower limit marking perfect absence of correspondence; in either case, we learn little as to the size of the correlation. More often the expression has been such as to have no constant limits at all, but to vary according to the number of cases tested; or it has even depended on special arrangements of the data, that would not apply to correlations between other things, or between the same things under other experimental circumstances; here the expressions of correlation are well nigh useless, as they cannot be compared with one another. Moreover, all these methods have lacked the fundamental requisite, a valuation of the 'probable error.'

Of the more scientifically devised methods, very few meet the requirement urged in the second section of this paper,—that of measuring the correlation between two series of ranks¹. And these few suffer from the fatal defect, that their coefficients are not in any way comparable with the r of the standard method. This defect is fatal, partly because the numerous valuable results already obtained by the r method are thereby shut out from comparison, and partly because the r values have an intrinsic significance that can ill be dispensed with².

§ 4. *Proposed method.*

The last-named defect vitiated among others a method suggested by myself (but unaccompanied by proofs) two years ago³. But I believe now to have overcome this difficulty; the respective results given by this and the r methods turn out to be connected by such a simple relation, that translation from one measure into the other is almost as easy as turning a few pounds into their value in shillings.

The method proposed by me seems also to meet every other main requirement of a 'footrule'; above all, the calculation is so extremely

¹ At any rate without an enormous and quite unnecessary increase of the probable error.

² These r values admit further of certain corrective formulae, without which, I believe, psychological correlation-research must remain in large measure delusive (cf. *Amer. J. Psych.* 1904, xiii. pp. 88-101.

³ *Loc. cit.* p. 86.

short, that it can often be completed under a minute, and sometimes even in one's head. Such a trifling expenditure of labour could often have saved an investigator from completely misapprehending his own results, the labour perhaps of many months.

This short method will first be explained (by an actual example); then its relation to the r method will be demonstrated; and finally the requisite mathematical proofs will be given (in Appendix).

The example is picked from some experiments recently done by Professor Krueger and myself on the correlation between various intellectual faculties. Eleven persons (21—38 years of age, all students of psychology) were individually tested by Krueger as to their power of adding numbers quickly and accurately. Each person, a week after being tested by Krueger in addition, was examined by myself as to his power of discriminating pitch; two tones were sounded successively and he had to decide which was the higher¹. When all the tests had been completed, Krueger and I showed each other our respective results. These are given below.

Reagent	Rank for addition (Krueger)	Rank for discrimination of pitch (Spearman)	Gains in rank for pitch as compared with addition
A	7	6	1
B	4	4½	
C	10	10	
D	1	3	
E	6	11	
F	9	8	1
G	11	7	4
H	3	1	2
I	2	2	
J	5	4½	½
K	8	9	
Observed sum of the gains (denoted by Σg) = 8½			

Our method of calculating the correlation is as follows. Let the observed sum of gains be denoted by Σg^* . Let the sum of gains to be expected on an average by mere chance be denoted by M ; this amounts to $\frac{n^2 - 1}{6}$, where n is the number of cases in each series (for

¹ It is hoped to publish soon a full account of these experiments.

• Where Σ , as usual, denotes summation. We need only consider the gains, not the losses also, because the sum of the one must necessarily be equal to the sum of the other.

proof, see App.)¹. Then the correlational coefficient, say

$$R, = 1 - \frac{\Sigma g}{M} *.$$

In the above case $n = 11$, so that $M = \frac{11^2 - 1}{6} = 20$. Thus the correlation between addition and pitch

$$= R = 1 - \frac{\Sigma g}{M} = 1 - \frac{8.5}{20} = \frac{11.5}{20} = 0.57.$$

The probable error may be taken with sufficient nearness as being

$$= \frac{0.43}{\sqrt{n}} \text{ (for proof, see App.)}.$$

In the above case, this becomes

$$\frac{0.43}{\sqrt{11}} = 0.13.$$

To utilize these two values, the coefficient R and its probable error, it must be remembered that in this (and in almost every) probability formula, *any experimental result such as R has no scientific significance—except negatively—unless it be at least twice as great as its probable error*; for otherwise, it is almost as likely as not to be a mere chance coincidence. To be fairly good evidence, the R (or other analogous result) must be over three times greater than its probable error. For a perfectly satisfactory demonstration, the R should be not far short of five times greater².

In this way we learn that our correlation between the faculty of adding numbers and that of discriminating pitch is just about large enough to be beyond all reasonable suspicion of mere chance coincidence. Had it been only three times greater than the probable

¹ In practice one usually has to consider several series of the same length. In such case, however much Σg may vary, M remains always the same and only requires to be reckoned once for all.

* Any *minus* result to this equation means that the correlation is inverse. Inverse correlations large enough to be appreciable should be avoided; this is easily done by ranking one of the series in the reverse order (this reversal of ranks usually modifies the result, but only to an extent that is much smaller than the probable error and therefore negligible).

² These relations are deduced from the fact, that the larger accidental deviations of a result from its normal value are much rarer than the smaller ones. With values distributed according to the common probability curve, a deviation twice as great as the probable error will occur by mere chance no less than once out of six trials, on an average; a deviation three times greater than the probable error occurs only once out of twenty-three trials; a deviation five times greater, not once out of 1,000 trials.

error, *i.e.* 0.39, the danger of mere chance would have been small, but not quite negligible.

The probability relations brought to light by the formulae teach several facts of general interest for experimental methodics. One sees that, when a high correlation (say, $R = 0.50$) exists between two things, its existence may be satisfactorily proved by—as a minimum—about a dozen cases.

But, on the other hand, it is clearly no good attempting to demonstrate a low correlation (say, $R = 0.20$) without first collecting for this purpose at least a hundred cases. And, similarly, about a hundred cases would be required to prove that *no* correlation over 0.20 existed. While to prove with certainty the absence of any correlation over 0.02, one would need ten thousand cases.

One may further note the danger of the common custom of gathering all the data into a small number of groups. Suppose, for instance, that one were determining whether the 'spatial sense' of children alters with their age. A not unusual procedure would be to test the spatial sense of about 100 children, group these, say, into three classes according to age, and then simply see whether the average spatial sense and the average age varied concurrently. But the mere observation of any such concurrent variation would be scientifically worthless, for the odds are only two to one against it occurring just by chance. Now, had we calculated the correlation direct from the individual cases, '*n*' would have been, not 3, but 100, and the reliability of our result would have risen accordingly; we could have determined the existence and precise amount of any correlation, with the small probable error of $\frac{0.43}{\sqrt{100}} = 0.04^*$.

Moreover, the correlation between averages of classes varies enormously with the size of the classes; and as such size differs from one experiment to another, the respective correlations admit of no comparison together. Whereas, if the correlation be calculated from the individual cases, its value is wholly unaffected by the quantity of the cases¹, and all coefficients are strictly comparable.

It should be mentioned that the large probable error given by the

* There are, it is true, other formulae, by which one may reckon whether the differences between the respective averages could have arisen by mere chance; but in any case the probable error is much larger than when calculating the correlation between the individual cases.

¹ Except, of course, for merely accidental fluctuations, which will be confined within the limits indicated by the probable errors.

above formula for use with R is only necessary when weighing the evidence as to the *existence* of any real correlation at all. When once this existence has been sufficiently proved and when our sole aim is to estimate the *accuracy* of the coefficient, another and much less severe probable error becomes applicable. It depends not only on the number of cases but also on the size of the observed correlation; the rule is that, the larger the correlation, the smaller its probable error. Of this nature is the probable error according to the formula devised by Pearson for the r method. On the other hand, that for the R formula has the great advantage of remaining sufficiently accurate even for correlations derived from no more than half a dozen cases, a fact that has not been demonstrated for the Pearson formula. On the r and R methods being brought into relation with one another (see next section), each becomes able to avail itself to some extent of the special advantages of the other's probable error.

§ 5. *Relation of the short method to the standard method.*

Pearson's formula for the standard method is:

$$r = \frac{\Sigma xy}{n \sigma_1 \sigma_2},$$

where x and y are the deviations of any corresponding pair of characteristics from their respective averages, and Σxy denotes the sum of all the values of xy ; σ_1 and σ_2 are the 'standard deviations'¹ of x and y respectively; and n is the number of pairs. The practical working out involves, as is well known, very tedious arithmetical operations.

Now, when the same method is applied to a correlation between two series of ranks, the formula admits of a remarkable simplification. We get

$$r = 1 - \frac{\Sigma d^2}{m},$$

where d is the numerical difference between any corresponding pair of ranks, and m is the mean value of Σd^2 by mere chance, being

$$= \frac{n(n^2 - 1)}{6} *.$$

¹ The 'standard deviation' is the square root of the average of the squares of the deviations.

* The identity (in case of ranks), $\frac{\Sigma xy}{n \sigma_1 \sigma_2} = 1 - \frac{6 \Sigma d^2}{n(n^2 - 1)}$, can be so easily tested practically, that a mathematical demonstration seems here unnecessary. A few 'ties' of rank will produce a discrepancy, but so slight as to be wholly negligible.

By this simplification the calculations are cut down in a way that should be welcome enough to those whose time is being lavished upon the usual procedure.

In the case of any ranks 'tying,' fractions may be conveniently avoided by doubling each difference before squaring. The formula then becomes

$$r = 1 - \frac{\Sigma (2d)^2}{4n}$$

If we apply this formula to the same illustration as before, we get

Reagent	Rank for addition	Rank for discrimination of pitch	$(2d)^2$
A	7	6	4
B	4	4½	1
C	10	10	0
D	1	3	16
E	6	11	100
F	9	8	4
G	11	7	64
H	3	1	16
I	2	2	0
J	5	4½	1
K	8	9	4
$\Sigma (2d)^2 = 210$			

Then
$$4m = \frac{4n(n^2 - 1)}{6} = \frac{4 \times 11(11^2 - 1)}{6} = 880.$$

So that $r = 1 - \frac{210}{880} = 0.76$ (the next two decimals are 14). If r be worked out by the usual laborious method, it again comes to 0.76 (the next two decimals being 10). Hence the discrepancy caused by the tie amounts only to 0.0004.

In this simplified form, the standard or r method shows itself to be solely distinguished from our short or R method by the fact that the differences of rank are squared¹. The effect of squaring is to give more 'weight' to the extreme differences as compared with the medium ones. This is probably a considerable advantage in most physical measurements. But in other fields of research, and perhaps above all

¹ In the R method we needed to consider the sum of the gains only, because this must always be exactly the same as that of the losses. But such equality no longer holds good after squaring; hence here both sorts of difference, losses as well as gains, must be taken into reckoning.

in psychology, these extreme cases are just the ones of most suspicious validity, so that the squaring is here more likely to do harm than good.

But the discovery of this simple relation between the processes of arriving at R and r respectively is far from telling us at once the relation between the magnitudes of R and r themselves. Turning now to this task, we first of all have to choose whether our proof shall be theoretical or practical. At first sight, a theoretical proof might seem much more complete and scientific. But two important facts have to be taken into consideration. The one is that correlations may conceivably have very *varied forms*; while some may be 'rectilinear' (*i.e.* based on simple proportionality between the two species of values compared), others may possess any kind of 'skewness.' The second fact is that on such variation of the correlation's form, a change must also be expected in the relation between the magnitudes of any two coefficients based—as is the case with R and r —on different powers of the data; the changing relation of such coefficients is in fact the recognized mathematical means of precisely determining the form of a correlation. It is thus clear that any theoretical proof of a constant relation between R and r is quite impossible, except so far as it can be supported by an empirical proof that the class of correlations in question does not vary much in form. But this required empirical proof of a constant form will itself consist, as just mentioned, in an empirical proof of a constant relation between such coefficients as R and r , based on different powers of the data. Thus any attempt at a theoretical proof of a constant relation between R and r is at once thrown back upon an empirical proof of the same relation; whichever route we choose, the validity of our argument always remains coextensive with the validity of the empirical proof. The present paper, therefore, may well content itself with such empirical evidence; this will be the most direct demonstration possible, that our proposed formula connecting R and r actually holds good in all the very various types of psychological correlation from which our observations have been derived.

Incidentally, such evidence will perform a much more important office than the originally intended mere enhancement of the usefulness of the R method. For until any class of correlations has been shown to possess some constancy of form, such correlations can scarcely be said to be *measurable at all*. This may be most readily seen by assuming the extreme case, that any correlation may be of any conceivable form whatever; then, obviously, any actually observed relations whatever between any two compared series may—for all we can know—represent

a perfect correlation. To measure any correlation quantitatively, it is indispensable to make some assumption as to the ideal form, and then to determine how much the observed facts fall short thereof.

The present discussion has, it is hoped, helped to justify the requirement urged in section 2, that psychological correlations should always be based upon values properly comparable with one another (such as are furnished, in general, by comparative ranks). No constancy of correlational form, nor indeed any other kind of regularity, can fairly be expected so long as the calculations are based upon the absolute measurements, whose quantitative relation to the things to be measured are unascertainable—and even, in large measure, arbitrary.

So much for the nature and importance of our task; it is to consist of an empirical proof of some constant relation between the magnitude of R and that of r . First of all we may note that the values of these two symbols already by definition coincide with one another as regards their extreme upper and lower limits, since both coefficients become 1 for perfect correlation and 0 for entire absence of correlation. The problem is to find out what relation connects them for all intermediate values. For this purpose I have worked out both the R and the r of one hundred and eleven correlations that I have actually come across¹. These have then been gathered into nine groups according to magnitude, so as to obtain at every intermediate tenth an average R and its corresponding r , whereby we can at once inspect the relation between the two throughout the whole range of possible correlation². In the following graph, the base line represents the entire range of R from 0

to 1; the heights of the dots above the base line represent the excesses of r over R actually given by the above-mentioned averages; and the

¹ In experiments, school class lists, etc. For calculation of both R and r , any series originally expressed in absolute measurements were converted into series of comparative ranks.

² There is evidently a slight error involved in applying the formula to averages instead of to each correlation singly, but this is so slight as to be negligible.

height of the curve above the base line shows how much any r would be larger than its corresponding R , if the following relation were true:

$$r = \sin \left(\frac{\pi}{2} R \right)^*.$$

As may be seen, the correspondence between the curve and the dots—that is, between theory and observation—is all that could be desired.

The same result is shown, less vividly but more precisely, in the following table. Here the crucial test is furnished by the right-hand column, whose values ought, by the theory, to remain generally constant from top to bottom, and whose total average should closely approximate to 1.

The correlation R lies between the limits	No. of correlations observed	Average of R	Average of r	$\sin \left(\frac{\pi}{2} R \right)$	$\sin \left(\frac{\pi}{2} R \right)$ divided by r
·06—·15	9	·107	·169	·167	·988
·16—·25	16	·211	·321	·325	1·001
·26—·35	24	·311	·460	·469	1·019
·36—·45	25	·400	·579	·588	1·016
·46—·55	17	·498	·693	·705	1·016
·56—·65	5	·598	·816	·806	·985
·66—·75	9	·691	·862	·884	1·024
·76—·85	4	·787	·945	·945	1·000
·86—·95	2	·915	·990	·991	1·001
Total = 111		Total Average = 1·009 (± 0·003)†			

The point perhaps of greatest practical interest is the size of the discrepancy between r and $\sin \left(\frac{\pi}{2} R \right)$ in the individual correlations; this amounts, on an average, to 0·034. As the average probable error of r

* Two years ago I described another relation between R and r , so long as r was between 0·20 and 0·60 (*Amer. J. Psych.* xiii. p. 88, 1904); this I still find to be true enough within such limits, but it turns out to be very inaccurate either above or below them.

† Let the deviations from unity shown by the respective values in the right-hand column be denoted by λ , the number of such values by μ , and the number of correlations entering into each value by p , then, in the usual probability notation, the probable error of the total average

$$= 0\cdot6745 \sqrt{\frac{[p\lambda\lambda]}{[p](\mu-1)}} = 0\cdot6745 \sqrt{\frac{0\cdot014}{111(9-1)}} = 0\cdot003.$$

Thus this discrepancy from unity, although so minute, is still three times greater than its probable error, and therefore rather large for mere chance.

is 0.09 (n and r averaging a little over 21 and 0.50 respectively), the discrepancy is not much more than a third of it. Only three times out of all the 111 correlations did the discrepancy actually exceed the probable error (of r), and then barely.

It is particularly important to note that this discrepancy, small as it is, seems to be wholly dependent on the size of n . In the great majority of our correlations, the series were of very nearly equal length, n averaging 20.6; for these the average between r and $\sin\left(\frac{\pi}{2}R\right)$ was 0.035; if we multiply the discrepancy by the square root of n , we get $0.035 \times \sqrt{20.6} = 0.16$. In only nine correlations was the length distinctly different, n here averaging 31.8; for these the average discrepancy between the two formulae sank to 0.029; again multiplying the discrepancy by $\sqrt{n}$, we get $0.029 \times \sqrt{31.8}$, which once more = 0.16.

Thus the average discrepancy between r and $\sin\left(\frac{\pi}{2}R\right)$ seems to diminish, just like the probable error, in proportion to the increase of $\sqrt{n}$, that is to say, in proportion to the reliability of the evidence¹. Like the probable error itself, the discrepancy between the two formulae seems to be wholly due to the accidental deviations of the actual correlational coefficients from their true value (that which would be obtained from series of infinite length). No discrepancy remains over to be explained by any real variation of the *form* of the correlation. Hence, this variation of form would seem to be vanishingly small; all these psychological correlations which form the basis of the present investigation, although derived from very varying sources, approximate exceedingly closely to one constant form.

In general, we seem entitled to consider the discrepancies between the two formulae as being too small to be sensible². As far as they do occur, they are due to the mere accidental errors of the two correlational coefficients, and are as much the fault of the standard as of the short method. When we have obtained R —a calculation performed in a minute or so—we can easily enough get r , either by working out the formula, or by using the table given below. In my own experience the r thus derived from R has proved an actually more reliable value than r calculated directly in the usual manner; that is to say, the former has

¹ This result has been confirmed by all later observations of the present writer.

² Should such discrepancy ever in any exceptional case become really marked, it performs the valuable office of indicating that the correlation in question departs from the common form.

accorded more closely than the latter with the r found (even by direct calculation) in other quite similar experiments. (If, as is usually the case, R does not exceed 0.50, r may simply be taken as $= 1\frac{1}{2} R$, without error of importance for most psychological purposes.)

To sum up, the R method seems to proffer a remarkably handy and yet quite accurate enough 'footrule' for measuring correlations. By using it, we obtain a precise quantitative value, which can be compared with that found for any other correlation, under any other circumstances or between any other things, either by the same or by the standard r method; we completely free ourselves from various illusions, which are otherwise almost irresistible; we get a reliable estimate as to the danger of our result being merely an accidental coincidence; and we even learn how to plan out our experiments from the outset in a manner properly adapted to the object in view.

Further, incidentally, some evidence has been brought to show that psychological correlations in general approximate to one constant form. Hereby they prove themselves to be really measurable quantitatively (and, what almost comes to the same thing, to be properly expressible by one single correlational coefficient).

Table to compare together the two correlational coefficients R and r , the latter being taken as $= \sin \left(\frac{\pi}{2} R \right)$.

R	r	R	r	R	r	R	r	R	r
.00	.00	.20	.31	.40	.59	.60	.81	.80	.95
.01	.01	.21	.32	.41	.60	.61	.82	.81	.96
.02	.03	.22	.34	.42	.61	.62	.83	.82	.96
.03	.05	.23	.35	.43	.62	.63	.84	.83	.96
.04	.06	.24	.37	.44	.64	.64	.84	.84	.97
.05	.07	.25	.38	.45	.65	.65	.85	.85	.97
.06	.08	.26	.40	.46	.66	.66	.86	.86	.98
.07	.11	.27	.41	.47	.67	.67	.87	.87	.98
.08	.13	.28	.43	.48	.69	.68	.88	.88	.98
.09	.14	.29	.44	.49	.70	.69	.88	.89	.99
.10	.16	.30	.45	.50	.71	.70	.89	.90	.99
.11	.17	.31	.47	.51	.72	.71	.90	.91	.99
.12	.19	.32	.48	.52	.73	.72	.90	.92	.99
.13	.20	.33	.50	.53	.74	.73	.91	.93	.99
.14	.22	.34	.51	.54	.75	.74	.92	.94	1.00
.15	.23	.35	.52	.55	.76	.75	.93	.95	1.00
.16	.25	.36	.54	.56	.77	.76	.93	.96	1.00
.17	.26	.37	.55	.57	.78	.77	.94	.97	1.00
.18	.28	.38	.56	.58	.79	.78	.94	.98	1.00
.19	.29	.39	.57	.59	.80	.79	.95	.99	1.00

APPENDIX.

I. *Proof that M , the mean value of Σg by pure chance, $= \frac{n^2 - 1}{6}$.*

Let $S = 2 \Sigma g$, denote the sum of the differences of rank between any two series consisting each of n terms (see p. 95); let $M(S)$ denote the mean value of all the possible combinations of S ; then,

$$S = |h_1 - 1| + |h_2 - 2| + \dots + |h_n - n| = \sum_{\alpha} |h_{\alpha} - \alpha|$$

and
$$M(S) = M(\sum_{\alpha} |h_{\alpha} - \alpha|) = \frac{1}{n_{\alpha\beta}} \sum |\beta - \alpha|$$

(since the number of possible combinations between α and $\beta = n^2 = n$ times the number of terms in S)

$$\begin{aligned} &= \frac{1}{n} \sum_{\alpha} \{|1 - \alpha| + |2 - \alpha| + \dots + |n - \alpha|\} \\ &= \frac{1}{n} \sum_{\alpha} \{(\alpha - 1) + (\alpha - 2) + \dots + 1 + 0 + 1 + 2 + \dots + (n - \alpha)\} \\ &= \frac{1}{n} \sum_{\alpha} \left\{ \binom{\alpha}{2} + \binom{n - \alpha + 1}{2} \right\}^* \\ &= \frac{1}{n} \left\{ \binom{n + 1}{3} + \binom{n + 1}{3} \right\} \\ &= \frac{2}{n} \cdot \frac{(n + 1)n(n - 1)}{1 \times 2 \times 3} = \frac{n^2 - 1}{3}. \dots\dots\dots(a)^1 \end{aligned}$$

And as all the possible combinations are *a priori* equally likely to occur, this $\frac{n^2 - 1}{3}$ may be taken as the mean value of S by pure

chance, so that
$$M = \frac{n^2 - 1}{6}. \dots\dots\dots(b)$$

* Where $\binom{\alpha}{2}$ denotes, as usual, the binomial coefficient, $\frac{\alpha(\alpha - 1)}{1 \cdot 2}$.

¹ This equation was arrived at in a different way by G. Lipps (see *Amer. J. Psych.* 1894, xiii. p. 86).

II. Proof that the 'probable error' of R approximately $= \frac{0.43}{\sqrt{n}}$.

i. Let us commence by determining the mean square deviation of $2 \Sigma g$ from $M(S)$, that is, of S from $\frac{n^2-1}{3}$,

$$\begin{aligned} M\left(\frac{n^2-1}{3} - S\right)^2 &= \left(\frac{n^2-1}{3}\right)^2 - 2M(S) \cdot \frac{n^2-1}{3} + M(S^2) \\ &= -\left(\frac{n^2-1}{3}\right)^2 + M(S^2). \dots\dots\dots(c) \end{aligned}$$

We require, then, to calculate $M(S^2)$,

$$\begin{aligned} S^2 &= \sum_{\alpha\beta} |h_\alpha - \alpha| |h_\beta - \beta| \\ &= \sum_\alpha (h_\alpha - \alpha)^2 + \sum_{\alpha\beta}^* |h_\alpha - \alpha| |h_\beta - \beta| \end{aligned}$$

where the star over Σ indicates summation only of pairs of values $\alpha \geq \beta$ (whereas Σ without a star shows that α and β run independently of one another from 1 to n).

Therefore

$$M(S^2) = \frac{1}{n} \sum_{\alpha\beta} (\beta - \alpha)^2 + \frac{1}{n(n-1)} \sum_{\alpha\beta\gamma\delta}^* |\gamma - \alpha| |\delta - \beta|. \dots(d)$$

The first component,

$$\begin{aligned} \frac{1}{n} \sum_{\alpha\beta} (\beta - \alpha)^2 &= 2 \sum \alpha^2 - \frac{2}{n} (\Sigma \alpha)^2 \\ &= \frac{n(n+1)(2n+1)}{3} - \frac{n(n+1)^2}{2} = \frac{n(n^2-1)}{6}. \dots\dots\dots(e) \end{aligned}$$

As regards the second component of the equation (d), $\sum_{\alpha\beta}^*$ may be transformed into $\sum_{\alpha\beta} - \sum_\alpha (\beta = \alpha)$.

$$\begin{aligned} \text{Then } \sum_{\alpha\beta\gamma\delta}^* |\gamma - \alpha| |\delta - \beta| &= \sum_{\alpha\beta\gamma\delta} |\gamma - \alpha| |\delta - \beta| - \sum_{\alpha\gamma\delta} |\gamma - \alpha| |\delta - \alpha| \\ &= \sum_{\alpha\beta\gamma\delta} |\gamma - \alpha| |\delta - \beta| - \sum_{\alpha\beta\gamma} |\gamma - \alpha| |\delta - \beta| - \sum_{\alpha\gamma\delta} |\gamma - \alpha| |\delta - \alpha| \\ &\quad + \sum_{\alpha\gamma} |\gamma - \alpha| \\ &= \sum_{\alpha\beta\gamma\delta} |\gamma - \alpha| |\delta - \beta| - 2 \sum_{\alpha\beta\gamma} |\gamma - \alpha| |\gamma - \beta| + \sum_{\alpha\beta} (\beta - \alpha)^2 \\ &= \left\{ \sum_{\alpha\gamma} (\gamma - \alpha) \right\}^2 - 2 \sum_\gamma \left\{ \sum_\alpha |\gamma - \alpha| \right\}^2 + \sum_{\alpha\beta} (\beta - \alpha)^2. \end{aligned}$$

The first term on the right of the above equation is obviously

$$= \{nM(S)\}^2, \text{ which by (a) } = \frac{n^2(n^2-1)^2}{9}. \dots\dots\dots(f)$$

The third term is by (e)

$$= \frac{n^2(n^2-1)}{6} \dots\dots\dots(g)$$

Lastly, the second term

$$\begin{aligned} 2 \sum_{\gamma} \left\{ \sum_{\alpha} |\gamma - \alpha| \right\}^2 &= 2 \sum_{\gamma} \left\{ \binom{\gamma}{2} + \binom{n+1-\gamma}{2} \right\}^2 \\ &= 2 \sum_{\gamma} \left\{ \gamma^2 - (n+1)\gamma + \frac{n(n+1)}{2} \right\}^2 \\ &= 2 \sum_{\gamma} \left\{ \gamma^4 - 2(n+1)\gamma^3 + (n+1)(2n+1)\gamma^2 - n(n+1)^2\gamma + \frac{n^2(n+1)^2}{4} \right\} \\ &= \frac{n(n+1)(2n+1)(3n^2+3n-1)}{15} - n^2(n+1)^3 + \frac{n(n+1)^2(2n+1)^2}{3} \\ &\quad - n^2(n+1)^3 + \frac{n^3(n+1)^2}{2} \\ &= \frac{n(n^2-1)(7n^2-8)}{30}, \text{ on reduction. } \dots\dots\dots(h) \end{aligned}$$

Combining equations (c), (d), (e), (f), (g), and (h), we get

$$\begin{aligned} M \left(\frac{n^2-1}{3} - S \right)^2 &= - \frac{(n^2-1)^2}{9} + \frac{n(n^2-1)}{6} \\ &\quad + \frac{1}{n(n-1)} \left\{ \frac{n^2(n^2-1)^2}{9} - \frac{n(n^2-1)(7n^2-8)}{30} + \frac{n^2(n^2-1)}{6} \right\}, \\ \text{which reduces to } &\frac{(n+1)(2n^2+7)}{45} \dots\dots\dots(i) \end{aligned}$$

As the term on the left of the above equation is the mean square deviation of $2\Sigma g$ from $M(S)$, we get the 'standard deviation' of Σg from M

$$= \frac{1}{2} \sqrt{\frac{(n+1)(2n^2+7)}{45}} \dots\dots\dots(j)$$

ii. To find thence the *median* deviation of Σg , I was obliged to proceed empirically¹. Many series of numbers were obtained in such a way, that the distribution of values could be considered that of pure chance. These series were then taken by pairs; for each pair the median deviation of Σg from M was actually observed, and then com-

¹ After I had unsuccessfully tried to determine the probable error directly, Prof. Hansdorff kindly suggested, as being a much easier task, the determination of the mean square error. He had taken the trouble to work this out, but then had unfortunately mislaid his calculations; however in his letter to me, he indicated his result and the general course of his proof. Thus, while the credit of this part of the proof is wholly his (as already mentioned by me in the *Amer. J. Psych.*), any faults in the calculations given here lie only with the present writer.

pared with the standard deviation according to equation (j); the results were as follows :

Number of pairs of series	Size of n	Median deviation of Σg actually observed ¹	Standard deviation of Σg by equation (j)	Median $\div$ standard deviation
12	11	2.8	4.3	0.66
47	24	9.1	13.1	0.70
16	48	22.4	35.8	0.63
Average = 0.66*				

It is thus evident that the ratio of the median to the standard deviation (given in right-hand column) may be taken—at any rate with sufficient approximation for our present purpose—as being the same as that for a normal probability curve, *i.e.* 0.6745. The median deviation of Σg becomes then, by equation (j),

$$= 0.6745 \times \frac{1}{2} \sqrt{\frac{(n+1)(2n^2+7)}{45}} = P, \text{ say.} \quad \dots\dots(k)$$

iii. We get finally, as the sought probable error of R ,

$$\begin{aligned} & 1 - \frac{M \pm P}{M} \\ &= 1 - 1 \pm 0.6745 \times \frac{1}{2} \sqrt{\frac{(n+1)(2n^2+7)}{45}} \cdot \frac{6}{n^2-1} \\ &= \pm 0.6745 \sqrt{\frac{2}{5n}} \left(1 + \frac{1}{n}\right)^{\frac{1}{2}} \left(1 + \frac{7}{2n^2}\right)^{\frac{1}{2}} \left(1 - \frac{1}{n^2}\right)^{-1} \\ &= \pm \frac{0.4266}{\sqrt{n}} \left(1 + \frac{1}{2n} + \text{terms divided by higher powers of } n\right). \end{aligned}$$

The terms on the right of the preceding equation form a very rapidly convergent series when n is not exceedingly small, so that the

$$\text{probable error of } R = \pm \frac{0.43}{\sqrt{n}},$$

with two correct decimals when n is not less than 10, and with sufficient accuracy for practical purposes when n is not less than, say, 6.

¹ The plus and minus deviations were very nearly equal.

* This result is not appreciably altered when the values in this (right-hand) column are 'weighted,' according to the amount of observational data upon which they are respectively based.