THE ESSENTIALS OF MENTAL MEASUREMENT

CAMBRIDGE UNIVERSITY PRESS
London: FETTER LANE, E.C.
C. F. CLAY, Manager

Edinburgh: 100, PRINCES STREET
London: H. K. LEWIS, 136, GOWER STREET, W.C.
Berlin: A. ASHER AND CO.
Leipzig: F. A. BROCKHAUS
New York: G. P. PUTNAM'S SONS
Bombay and Calcutta: MACMILLAN AND CO., Ltd.

THE ESSENTIALS OF MENTAL MEASUREMENT

BY

WILLIAM BROWN, M.A. (OXON.), D.SC. (LOND.)
Lecturer on Psychology, University of London, King's College

Cambridge:
at the University Press
1911

Cambridge:
PRINTED BY JOHN CLAY, M.A.
AT THE UNIVERSITY PRESS

PREFACE

THE composition of this book requires some explanation, Chapters I, II, and III of Part II originally formed a doctorate thesis* on "The Use of the Theory of Correlation in Psychology," and were privately printed as such in August of last year. The other chapters of the book have, as it were, been written round this nucleus, with a view to producing a concise yet fairly complete account of the various quantitative methods employed in psychology. The slightly controversial nature of the thesis is intentionally retained, since the interests of quantitative psychology would seem to be best served thus—at least until the biometric methods of Professor Karl Pearson and his school become more generally known among psychologists than they are at present. Perhaps it is as well to add that I am alone responsible for the views therein expressed.

Although written primarily for the professed psychologist, it is hoped that the book will also prove of use to the educationist. The theory of correlation is likely to be found especially valuable in educational psychology, if only its presuppositions are clearly grasped and the requisite precautions in its application exercised; and now that a real training in psychology is beginning to be regarded as an essential part of a teacher's equipment, we may

* Approved for the Degree of Doctor of Science in the University of London.

expect valuable contributions to psychology from teachers in the immediate future. If this book is found helpful on the quantitative side to such research students, its existence will be justified.

My thanks are due to the Editors of the *British Journal of Psychology*, who have very kindly allowed me to re-publish an article from that periodical.

W. B.

May, 1911.

CONTENTS

PART I. PSYCHO-PHYSICS.

PART II. CORRELATION.

APPENDICES.

PART I

PSYCHO-PHYSICS

CHAPTER I

MENTAL MEASUREMENT

The pre-conditions of measurement in any sphere of experience are (1) the *homogeneity* of the phenomenon, or of any particular aspect of it, to be measured, (2) the possibility of fixing a *unit* in terms of which the measurement may be made, and of which the total magnitude may be regarded as a mere multiple or sub-multiple. These pre-requisites are satisfied in the cases of spatial and temporal magnitudes, in terms of which, directly or indirectly, all the measurements of the physical sciences are expressed. It was thought by Fechner that they also applied in the case of the strictly psychical phenomenon of sensation-intensity, i.e. it was assumed that any given sensation-intensity might be regarded as made up of a sum of unit sensation-intensities. This view has been definitely rejected by later psychologists. Every sensation-intensity is qualitatively distinct from every other sensation-intensity. No addition or subtraction of intensities is possible. "To introspection, our feeling of pink is surely not a portion of our feeling of scarlet; nor does the light of an electric arc seem to contain that of a tallow-candle in itself" (James)*. Fechner's mistake was due to a confusion of sensation-intensities with the (physical) stimulus-values required to produce them. The latter can be measured, the former cannot.

Nevertheless, purely psychical measurement is not entirely impossible. Within any one series of sensation-intensities,

* *Principles of Psychology*, Vol. I. p. 546.

e.g. a series of greys, the contrasts or "distances" separating different pairs of intensities are perfectly homogeneous with one another and can be measured in terms of one another or in terms of an arbitrarily chosen unit of "sense-distance." Given two brightness-intensities a and b, it is quite possible to find, within limits of error, a brightness-intensity c which is as much higher than b in the scale of intensities as b is than a, i.e. such that the sense-distance $\overline{bc}$ = the sense-distance $\overline{ab}$; or, again, it is quite possible, theoretically, to find a brightness-intensity d which bisects the sense-distance $\overline{ab}$, i.e. which is such that it is as far removed from a in the scale of intensities as b is from it—in symbols, $\overline{ad} = \overline{db}$. Hence the "distance," or

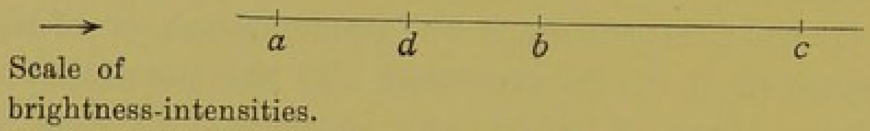


Fig. 1.

disparity of b from a is twice that of d from a, the distance of c from a is four times that of d from a. If, now, $\overline{ad}$, or the distance of d from a, be taken as a conventional unit, the values of $\overline{ab}$ and $\overline{ac}$ will be 2 and 4 respectively*.

A scale of intensities may in this way be formed rising by "equal-appearing intervals" or sense-distances, and the magnitude of any given interval, may then, theoretically, be read off in terms of the unit-distance employed in the construction of the scale. In practice, however, it is found more convenient to fix the successive scale-marks, the successive members of the intensity-series, in terms of their corresponding stimulus-values. It has been found by experiment that in the case of light-intensities and sound-intensities the successive stimulus-values form, with fair approximation, a geometric progression, or, in other words, each stimulus value divided by the immediately preceding one gives approximately the same quotient. From the stimulus-values corresponding to an ascending series of

* This view of mental measurement in terms of sense-distance first originated with J. R. L. Delboeuf: *Revue Philosophique* v. 1878, p. 53. His term for sense-distance was "contraste sensible."

eight equidistant brightness-values, Ebbinghaus obtained the following series of quotients:

2·3 2·1 2·1 1·8 1·7 1·7 2·0

The quotient-value is not entirely constant, being slightly greater towards the two ends of the scale than it is at or about the middle. For this central region of the scale, then, the general relation of the sense-distances to the stimulus-values is given by the logarithmic formula:

$$\overline{SS_0} = k \log \frac{R}{R_0} \quad \text{.....................(1)*},$$

where $\overline{SS_0}$ represents a sense-distance, and S_0 is any finite intensity of sensation taken as the starting point or conventional zero (N.B. it is not necessarily liminal); R, R_0 are corresponding stimulus-values.

A mode of procedure which not only admits of much wider practical application than the above-mentioned "method of mean gradations," but also possesses a peculiar historical importance, is that which is concerned with the determination of the stimulus-increments corresponding to just-noticeable increments of sensation-intensity in different parts of the intensity scale. *Weber* found, in a series of experiments chiefly with lifted weights, that this stimulus-increment was relatively, not absolutely, constant for different regions of the intensity scale, i.e. that the stimulus corresponding to any original sensation-intensity had always to be increased by a constant proportion to arouse a just-noticeable increment of the sensation-intensity. If a 103 grms. wt. is just noticeably heavier than a 100 grms. wt., then the wt. just noticeably heavier than a 200 grms. wt. will be a 206 grms. wt., that just noticeably heavier than 300 grms. will be 309 grms., etc.

Mathematically formulated, Weber's Law is:

$$\frac{dR}{R} = \text{a constant} \quad \text{.....................(2)}.$$

The quantity $\frac{dR}{R}$,—$\frac{3}{100}$ or ·03 in our example of weight-lifting—is known as the "relative difference limen" (*DL*).

* Ebbinghaus.

It is of course the average value of a considerable number of determinations.

Fechner verified Weber's Law in many different realms of sensation-intensity, and made it the basis of his own system of mental measurement. This he did by making the following three assumptions:

(1) that a sensation-intensity is a measurable magnitude and may therefore be regarded as a sum of unit-intensities;

(2) that just-noticeable differences of sensation-intensity are equal at different parts of the stimulus scale, and may therefore conveniently serve as the unit-intensities above-mentioned;

(3) that the just-noticeable difference of sensation may be treated as a difference of two sensations, or at least that if Weber's Law applies to the former ("sensed difference") it will also apply to the latter ("difference sensation").

On the basis of Weber's Law and these added assumptions, Fechner obtains the following formula, viz.

$$dS = c\frac{dR}{R} \qquad \text{.........................(3)}$$

which he calls *the fundamental formula for mental measurement.* Integrating, this becomes

$$S = c \log_e R + C.$$

Putting $R = r$, the "stimulus limen" for which S is just below the threshold of consciousness, i.e. $= 0$, and we have

$$0 = c \log_e r + C.$$

Subtracting, $$S = c \log_e \frac{R}{r}.$$

Putting $r = 1$, and transferring to the ordinary logarithm system, we get

$$S = k \log R \qquad \text{.......................(4).}$$

All the assumptions involved are questionable. The first one has already been considered at some length. It is not the single sensation-intensity which is measurable, but the distinctness, disparity or distance of one sensation-intensity from another. Any intensity taken by itself is 0, it is at a zero distance from itself, $\overline{S_0S_0} = 0$ [see equation (1) above].

We must therefore regard the just-noticeable difference, not as a difference of two sensation-intensities but as a minimal sense-distance, if we are to be able to make use of it in our scheme of mental measurement. This modification, however, still leaves us involved in the difficulties of assumptions (2) and (3).

Fechner's own reason for regarding all just-noticeable differences belonging to any one scale of intensities as equal was that they appear equal to introspection. Introspection in a case like this is obviously difficult, even for the most skilful observers, and its verdict cannot, therefore, be greatly relied upon. Theoretically it is quite conceivable that just-noticeable differences, though equivalent to one another as being all just noticeable, i.e. as being sense-distances so small that the slightest diminution of them would cause them all, equally, to cease to be noticeable, yet as noticed or perceived would appear of different magnitude one from another. Ebbinghaus points to the analogous case of differentials or infinitesimals in mathematics. These are all equivalent to one another as being all equally negligible as compared with finite magnitudes, yet are by no means necessarily equal to one another. If they belong to different "orders," those of a higher order are negligible as compared with those of a lower, etc. Again, "the least distances perceived as such at different parts of the skin or in direct and indirect vision do not by any means all appear as equal magnitudes. On the contrary, so soon as they come to consciousness as distances they are at once perceived as distances of varying size, in a certain approximation to their objective differences*." In spite of these considerations, Ebbinghaus regards the correspondence of the stimulus results obtained for equal appearing intervals and for just perceptible intervals in the case of brightness-intensities (in the middle region of the scale both series of stimuli form a geometrical progression) as sufficient evidence for the approximate equality of the latter intervals. Müller and Wundt had previously advanced the same argument.

* Ebbinghaus: *Grundzüge der Psychologie*, 2nd ed., 1905, p. 524.

Several experimental investigations have been made with the express purpose of testing the relation of the methods of minimal change and mean gradations. Titchener* sums up the theoretical basis of such experiments concisely as follows: "There are in reality two possible ways of working. (1) We might take a series of R-values, corresponding to a series, say, of eight successive j. n. d. of S, and thereafter directly compare the two half-distances, of four j. n. d. each, and decide upon their equality or inequality. This would be a direct method of experiment: determine the R-differences as $a-b$, $b-c$, $c-d$, $d-e$, $e-f$, $f-g$, $g-h$, $h-i$, and then see if $a-e$ is equal in S to $e-i$. Or (2) we might determine a few j. n. d. of S, at different parts of the R-scale, in order to establish the constancy of the relative DL, and thereafter work with supraliminal d., and decide whether the same uniformity holds. This would be an indirect method: it is the method indicated by the authors just cited [Müller, Wundt, Köhler, and Tannery]. Let $a-b$ and $h-i$ be R-differences corresponding to the j. n. d. of S, and let there be constancy of the relative DL, i.e., let $\frac{b-a}{a}=\frac{i-h}{h}$. If the j. n. d. are equal, then the point e, which lies midway for S between a and i, must be such that $\frac{e-a}{a}=\frac{i-e}{e}$, or $e=\sqrt{ai}$. Either of these two methods would, presumably, take us to our goal. The experimental work would be exceedingly difficult. Liminal determinations are always and intrinsically difficult; and, further, the judgments passed upon j. n. d. and upon supraliminal d. are, even under the most favourable conditions, the expressions of radically different mental attitudes."

It should be mentioned here that the principal rival to Fechner's hypothesis—the "difference hypothesis"—is that first formulated by Plateau, and generally known as the "quotient hypothesis." Plateau adopted the psycho-physical formula

$$S=c\,R^k \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

* "Experimental Psychology," Vol. II. *Instructor's Manual*, p. lxxvii.

on the basis of experiments by the method of mean gradations. This implies, in the place of Fechner's fundamental formula, the formula

$$\frac{dS}{S} = k\frac{dR}{R} \quad \text{.................................(6)},$$

in other words, it assumes that j. n. d. are relatively, not absolutely equal *S*-magnitudes. Although Plateau himself withdrew his formula later, the "quotient hypothesis" still remains as the rival of Fechner's "difference hypothesis."

To return to the experiments. Merkel (1888) worked with brightnesses, pressures, and noises, and found that the *R*-value of the *S* bisecting a supraliminal sense-distance was the arithmetic mean of *R*-values of the two terminal sensations, i.e., in Titchener's notation, $e = \frac{a+i}{2}$, not $\sqrt{ai}$. This would seem to support the quotient hypothesis, although such an inference is not entirely free from objection. Angell (1892) worked with noise-intensities by the method of mean gradations, avoided certain sources of error present in Merkel's form of procedure, and obtained results supporting the difference hypothesis, $e = \sqrt{ai}$.

A more thorough investigation was carried out by W. Ament* in 1900. He used a series of Marbe greys for the brightnesses and employed the direct method; for the noise-intensities he used a Fechner sound pendulum and worked principally by the indirect method. The result reached was that "the *DL* as a magnitude increases with increasing *R*," i.e. it supported the quotient hypothesis. Titchener† summarises Ament's conclusions as follows:

"The result is illustrated by the following Figure.

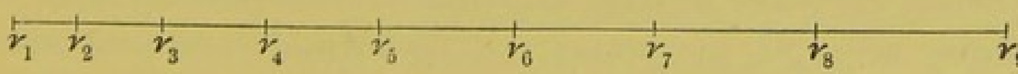


Fig. 2.

* W. Ament: "Ueber das Verhältnis der ebenmerklichen zu den übermerklichen Unterschieden bei Licht- und Schallintensitäten," *Phil. Studien*, XVI. 1900, pp. 135 ff.

† *Op. cit.* pp. lxxxv, lxxxvi.

The values r_1, r_2, r_3, etc., are *R*-intensities, arranged in ascending order: the distance between every pair of successive r represents a *DL*. The distances $r_1—r_2$, $r_2—r_3$, etc., increase by equal increments. It is clear (1) that the distance $r_5—r_9$, though it contains the same number of j. n. d., will appear greater than the distance $r_1—r_5$; (2) that the difference decreases, the nearer the compared distances approach to r_5,—so that a comparison of $r_4—r_5$ with $r_5—r_6$ may even evoke the judgment 'equal'; and (3) that a comparison of $r_1—r_2$ with $r_8—r_9$ will evoke a decided judgment of difference, although the two distances are alike j. n. d."

Ament's work has not escaped criticism. A repetition of his experiments on brightness-intensities by Fröbes* has failed to confirm his results. Ebbinghaus also has found, in careful experiments with rotating sectors, that just-noticeable differences in different parts of the scale of brightnesses are equal to one another. On the whole, therefore, the balance of evidence seems to be in favour of the "difference hypothesis."

Fechner's third assumption—see above, p. 4—brings us to the question of the interpretation of Weber's Law.

There are three general forms of interpretation:

(1) the psycho-physical (Fechner),
(2) the physiological (Müller, Ebbinghaus, James, etc.),
(3) the psychological (Wundt).

According to (1), the logarithmic transition takes place in passing from the physiological changes in the sensory centres of the cerebral cortex to the corresponding sensation-intensities. The chief objection to this view of Fechner's is that Weber's Law is not exact. This same consideration supports (2), or the physiological view, according to which the transition occurs either at the inception of the stimulus in the sense organ, or somewhere between the nerve endings and their central connections in the sensory areas of the cortex. Experimental results obtained by Waller† and Steinach‡ are in favour of this view. Waller

* *Zeitschrift für Psychologie*, XXXVI. p. 344.

† Waller: "Points relating to the Weber-Fechner Law," *Brain*, Vol. XVIII. p. 200, 1895.

‡ Steinach: "Elecktromotorische Erscheinungen an Hautsinnesnerven bei adäquater Reizung," *Pflüger's Arch.* Bd. 63, S. 495, 1896.

stimulated a frog's eye with light of different intensities, and found that the corresponding "negative variations" set up in the optic nerve varied in intensity as the logarithm of the stimulus-values (approx.). Steinach obtained similar results on stimulating the skin of a frog's thigh with weights and noting the negative variation in the attached nerve. If the negative variation may be assumed to be proportional to the intensity of the nerve-current passing along the nerve, these results point to the conclusion that the logarithmic transition occurs in the sense organ and its sensory nerve endings.

The effect would thus be analogous to that of friction and inertia in machinery. Ebbinghaus* has constructed a theory based upon the conception of varying degrees of dissociability of complex molecules to account for the law and also for the deviations from it towards the two extremes of the intensity scale.

The psychological view, (3), of the law, held by Wundt, regards it as a special case of the general psychological "law of relativity." Stimulus, physiological process, and pure sensation-intensity increase in simple proportion to one another. The logarithmic transition occurs in passing from mere sensation and sensation difference to apperceived sensation and apperceived sensation difference, and the intensities are apperceived always in relation to one another. In addition to the objection that it regards the sensation-intensities themselves, not their distances from one another, as measureable magnitudes, this view is also open to the criticism that it furnishes no explanation of the widely varying size of the relative difference limen in different sense-departments (e.g. *DL* for brightness-intensities $=\frac{1}{100}$, *DL* for sound-intensities = about $\frac{1}{3}$); the physiological view sees in the varying structure of the different sense-organs the adequate explanation of this. Moreover, the psychological view has no completely satisfactory explanation to give of the deviations from Weber's Law so frequently met with.

These objections and difficulties make the view improbable, but by no means prove it to be impossible. It has the great

* Ebbinghaus: "Über den Grund der Abweichungen von dem Weber'schen Gesetz bei Lichtempfindungen," *Pflüger's Arch.* Bd. 54, S. 113, 1889.

merit of emphasizing more definitely than was heretofore the case the importance of the more purely psychological factors in psycho-physical experiments,—in particular, it brings into prominence the distinction between mere disparity of sensation-intensities present simultaneously or in immediate succession in the same consciousness and the perception of this disparity, the *discrimination* of the intensities one from the other. In psycho-physical experiments the subject's consciousness is not limited to the mere sensational level.

Although, with continuous increase of stimulus-intensities the corresponding sensation-intensities rise in "steps," each representing a just-noticeable difference, the psycho-physical relation is really a strictly continuous one. This becomes at once obvious if we consider a special case. The sensation-intensity aroused in lifting a wt. of 100 grms. is indistinguishable from that aroused by 102 grms., the sensation aroused by 102 grms. is indistinguishable from that aroused by 104 grms.; yet the sensation-intensity aroused by 100 grms. is perceptibly less than that aroused by 104 grms. Thus the sensation-intensity increases continuously, and the reason why this is not immediately apparent is probably to be looked for in the physiological mechanism of the pyscho-physical organism. Delboeuf held that the limen has no psychological importance whatever. If this is an extreme view, the importance which Fechner attributed to the limen is equally extreme in the other direction, and farther removed from the truth.

The absolute or stimulus limen, *RL*, is similar in kind to the difference limen, since consciousness is never empty of sensation-intensities when such a limen is being determined. Here again, Fechner's rigid distinction of the two was a fundamental error.

Indirect Methods of Measurement.

The preceding account has probably sufficed to show that purely psychical measurement is a possibility. Its practical application, however, has been more detailed than extensive. A more generally useful method in quantitative psychology has been found to be that which measures the external, physical or physiological, causes and effects of mental process. The measurements are made in terms of the physical units of space and time, yet they are not merely physical measurements, since they derive all their significance from the correlated psychical processes. They are indirect psychical measurements*. Measurements of reaction-times, memory, fatigue, illusions, etc. are all of this nature. Their varieties are innumerable, and are illustrated by the accounts in any good text-book of experimental psychology (Sanford, Titchener, Myers). In all cases full introspective accounts are essential, and when correlated with the measurements make the latter essentially psychical measurements. Measurements of limina, referred to in the previous section, are of the same nature. They are of some special importance as being measures of sensory acuity, etc.—aspects of the total mental ability of psycho-physical organisms. They figure prominently in many researches based on the use of "mental tests."

A method which makes a partial return to the more purely psychical form of measurement in terms of "distance" is the method of ranks or grades. Suppose we are considering the relative abilities of, say, 100 boys in English Composition. We should find it difficult to mark their essays individually in terms of any constant unit but might find it possible to arrange them in order of merit, especially if we had sufficient time at our disposal to employ the method of "paired comparisons." According to the procedure of this latter method, the essays would be taken in pairs, quite at random, and the better essay of each pair would be given a "preference mark." This procedure would be repeated again and again until every essay had been compared with every other essay. The order of merit is then given by the number of preferences attaching to each essay. In this order, however, we cannot assume that the "ability-distance" from one boy to the next is a constant quantity. The boys near the extreme ends of the series will be farther removed from one another than the boys near the middle. We could only adjust for this if we knew the law of frequency-distribution for this kind of ability in this particular

* See Ebbinghaus, *Grundzüge der Psychologie*, 1905, pp. 75, 76.

species of boy, and theoretically the determination of this distribution depends upon a prior fixing of the psychological unit, the unit of "ability-distance"; so that strictly speaking the problem is insoluble. Since, however, under certain definite and indefinite conditions, the form of distribution in a large number of cases of physical measurement has been found to be Gaussian or "normal," we might, with some probability of being near the truth, *assume* this form of distribution in the given case and so obtain a quantitative measure for the ability of each particular boy. A direct psychological determination and application of the (conventional) unit-distance is not, perhaps, an entirely impossible problem, and work in this direction may be expected in the future. The method, if achieved, would certainly be much more scientific and psychological than the present one which measures in terms of external quantum of work done.

Finally, the interrelations of different mental abilities within any well-defined group of individuals situated within any definite environment may be determined by means of the technical method of "correlation." A correlation coefficient or other constant (e.g. correlation ratio) measures the *tendency* towards concomitant variation of two mental (or other) abilities within a group of individuals. The result may be transferred to any single individual within the group as measuring the degree of probability of connection or the closeness of connection of the two abilities in the particular case. The correlation between the two abilities may be due to an actual direct relation of the abilities to one another, or, indirectly, to the influence of a common external environment upon them both. The first of these two cases is perhaps the more important, but the possibility of the second should not be lost sight of, and it also has a special interest of its own. The problem of correlation will be considered more fully in a later chapter.

CHAPTER II

THE PSYCHO-PHYSICAL METHODS

The experimental determination of absolute and difference thresholds or limina is complicated and difficult. A considerable number of physical and psychological factors are involved, of varying relative importance in different cases. The result is that different methods of procedure have been found most suitable for different cases. These methods have been traditionally grouped under three (or four) distinct headings, and called the Psycho-physical Methods. They are:

(1) the Method of Limits (Method of Minimal Changes),

(2) the Method of Average Error (Method of Production),

(3) the Constant Method (Method of Right and Wrong Cases).

A fourth method is generally added to the list, viz.:

(4) the Method of Equal-Appearing Intervals, or Method of Mean Gradations, but this is really no new method. It owes its special name to the nature of the task which it fulfils, viz. the determination of equal-appearing (übermerklich) sense-distances as distinguished from just perceptible (ebenmerklich) sense-distances. The method which it employs falls under one or other of the first three headings.

1. The Method of Limits.

In using this method for the determination of difference limina the following mode of procedure is adopted. The variable stimulus, V, is first made equal or slightly larger than the standard, S, and then gradually increased by small increments until the subject finds it just noticeably greater than S. V is increased still more, and then gradually diminished until it just ceases to appear greater than S. The mean of the two values of $V-S$ thus obtained is the upper difference limen, D_u. Four values, instead of two, might also be obtained, viz. for (i) V just not perceptibly greater than S, (ii) V just perceptibly greater than S, (iii) V just perceptibly greater than S, (iv) V just not perceptibly greater than S, (i) and (ii) being for ascending values of V, (iii) and (iv) for descending. D_u will be the mean of these four values of $V-S$.

The lower difference limen, D_l, is obtained in a similar way. Both limina may be obtained in the same series of experiments by the "method of complete descent and ascent."

A series of determinations of each limen is made and the average taken. The reliability of this average measure is conveniently given by the "mean variation" (M.V.), which is the average of the deviations of the individual determinations (these deviations all being reckoned as positive) from their average. If there are n individual determinations $x_1, x_2, \ldots x_n$, with average $\bar{x}$, then

$$\text{M.V.} = \frac{(\bar{x} \sim x_1) + (\bar{x} \sim x_2) + \ldots + (\bar{x} \sim x_n)}{n}.$$

M.V. thus measures the "scatter" of the individual measures about their mean. Other measures of "scatter" are the "measure of precision" (h) (inverse), the standard deviation (σ) and the "probable error" (P.E.). The general question of the measure of "scatter" will be considered later on in this book.

The number and size of the increments employed must be adjusted to the particular conditions of the experiment. The subject of the experiment should be given a certain amount of preliminary practice before being started upon the work, and introspective reports should be asked of him.

Among the various possible sources of error which deflect the subject's judgment, two are of special importance. They arise from the temporal and spatial arrangement of the compared stimuli, and are called the "time error" and "space error" respectively. Thus, in a determination of the difference limen for sound-intensities, the two stimuli, S and V, cannot be presented simultaneously. One must precede the other, and in this way it may produce a slight degree of fatigue which causes an over-estimation of the other, or it may produce the reverse effect of sharpening the attention to the second. Thus a time error arises. Again, in experiments with brightness-intensities or visual extents, where the stimuli can be presented simultaneously, a space error arises from the fact that the one stimulus must be presented either to the right or to the left of the other stimulus and the subject's judgment varies accordingly. In experiments with lifted weights both sources of error may be involved. These so-called *constant errors* may be approximately neutralised by arranging that in the course of the experiment the standard shall precede the variable or stand to the right of the variable in half the cases and follow or stand to the left in the other half—the time or space order, or both, being of course changed quite at random in the successive limen-determinations. A better plan is to evaluate the limen (and its M.V.) for each time and space order separately. This gives us the values of the time error and space error, which are of interest for their own sakes. Fechner's theory of these errors and their measurement is based upon the assumption that the time and space order of the two stimuli, S and V, exert an influence upon the result which is equivalent to a definite increase or diminution in the stimulus-value of one or the other, and thus increase or diminish the value of $S \backsim V$ by the amount of the time error, p, or that of the space error, q, or that of both together, $p + q$, which $= c$, the constant total error. The time error, p, is positive or negative according as the effect of the time order is to increase or diminish the apparent value of the *first-presented* stimulus. The space error, q, is positive or negative according as the effect of the space order is to increase or diminish the apparent value of the *left-hand* stimulus.

Four principal cases of time and space order are possible, and are conventionally numbered as follows:

I. Standard presented first and to the right,
II. " " second " " "
III. " " first " " left,
IV. " " second " " "

Employing these numbers as suffixes, we have the equations*

$$2\,D_u = D_{u\,\mathrm{I}} + D_{u\,\mathrm{IV}} = D_{u\,\mathrm{II}} + D_{u\,\mathrm{III}}$$
$$2\,D_l = D_{l\,\mathrm{I}} + D_{l\,\mathrm{IV}} = D_{l\,\mathrm{II}} + D_{l\,\mathrm{III}}$$
$$2\,c_{\mathrm{I}} = D_{l\,\mathrm{I}} - D_{l\,\mathrm{IV}} = D_{u\,\mathrm{IV}} - D_{u\,\mathrm{I}}$$
$$2\,c_{\mathrm{II}} = D_{l\,\mathrm{II}} - D_{l\,\mathrm{III}} = D_{u\,\mathrm{III}} - D_{u\,\mathrm{II}}$$

In the case where a difference of time order alone is possible, the above equations reduce to

$$2\,D_u = D_{u\,\mathrm{I}} + D_{u\,\mathrm{II}}$$
$$2\,D_l = D_{l\,\mathrm{I}} + D_{l\,\mathrm{II}}$$
$$2\,p = D_{u\,\mathrm{I}} - D_{u\,\mathrm{II}} = D_{l\,\mathrm{II}} - D_{l\,\mathrm{I}}.$$

Corresponding equations are valid for the space error q, in a case where a difference of space order alone is present.

The errors due to expectation, habituation, fatigue, etc., are neutralised or at least reduced to a minimum by determining the limen by means of both ascending and descending values of V and averaging, and by other special precautions in applying the method.

A modification of the method which effects a more complete elimination of the expectation error by the use of occasional "catch experiments," where the variable given is equal to the standard, is that known as the *Method of Serial Groups*†.

Each fixed value of V is presented with the standard ten times, not in immediate succession but interspersed at random among ten other values of V equal to S. The percentage of correct answers given by the subject is noted, and the experimenter passes on to the next value of V which is presented along with catch stimuli in a similar manner. The value of V which, as presented in this way, gives 80 % right answers is arbitrarily chosen as measuring the limen. This method combines with

* G. E. Müller: *Die Gesichtspunkte und die Tatsachen der Psychophysischen Methodik*, 1904, pp. 67, 71.

† See C. S. Myers: *Textbook of Experimental Psychology*, p. 209.

the principle of the method of limits that of the constant method (yet to be described) and is a very convenient one to use in measuring a large number of subjects in "mental test" experiments, where economisation of time is essential.

2. The Method of Average Error.

In this method the subject is required to adjust a variable stimulus, v, so that it seems subjectively equal to a given standard stimulus, s. The experiment is repeated a large number of times—at least 100—and the arithmetical mean, v_m, of all the obtained stimulus-values is calculated. The difference between this mean value and the standard stimulus is known as the *crude constant error*, e. It may be either positive or negative. The mean variation, M.V. $= \frac{S(v_m \sim v)}{n}$, is also calculated.

The crude constant error may be partly due to a space error (the time error cannot occur in this method), partly to other constant conditions*. Let us assume that the experiment is to adjust the length of a variable line until it seems to be equal in length to a standard line. In this case, the standard should be situated to the right in one-half the number of adjustments and to the left in the other half, either alternately or in haphazard order. The results are tabulated in two columns, I standard to right, II standard to left, and $v_{m\text{I}}$, $v_{m\text{II}}$ are calculated from these two columns separately. Then, if q be the required space error and k the residual constant error, we have

$$v_{m\text{I}} - s = -q + k, \quad v_{m\text{II}} - s = +q + k,$$

i.e.

$$q = \frac{v_{m\text{II}} - v_{m\text{I}}}{2}, \quad k = \frac{v_{m\text{II}} + v_{m\text{I}}}{2} - s.$$

In order to give as much definiteness as possible to the task of adjustment, the variable should start sometimes shorter than the standard, sometimes (an equal number) longer, and the adjustment be made by lengthening and shortening respectively. Again, the requisite amount of lengthening or

* Cf. E. B. Titchener, *Experimental Psychology*, Vol. II. *Student's Manual*, p. 74. "A 'constant' error is simply an error whose *conditions* are constant; its *amount* may vary, quite considerably, from stage to stage of a long series of experiments."

shortening should be arranged to be different on different occasions, but alternating with some degree of uniformity.

The value of M.V. (mean variable error) obtained in this method is from a general point of view a more important result than the value of the constant errors, since it has often been regarded as proportional to the value of the difference threshold as determined by the other two psycho-physical methods. The truth is that although there is a certain amount of proportionality between the values, this proportionality is not complete. The correlation of mean variable errors and corresponding just noticeable differences is less than unity. Under certain conditions the two values vary in opposite directions. It is hardly necessary to point out that the M.V. obtained in the Method of Limits has but little relation to the mean variable error obtained by the Average Error Method. They are not entirely unrelated, however.

The closest correspondence of any is that between the "frequency distribution" of the errors in the present method, and that of the judgments "equal" obtained by the Constant Method.

3. The Constant Method.

This is generally considered to be the most satisfactory of the psycho-physical methods. It can be employed with equal convenience for the determination of absolute thresholds, difference thresholds, equally-appearing sense-distances, and other measurements of psychological importance. The different values of the variable stimulus to be employed are fixed once for all at the beginning of the investigation, and are presented to the subject a large number of times (say 100 applications of each variable stimulus value) in irregular order, or in a prearranged order, unknown to the subject, corresponding to certain precautions to be mentioned later. If an absolute threshold is being determined, the variable is presented alone, if a difference threshold, it is on each occasion preceded, accompanied or followed by the standard stimulus. In the latter case, the subject returns the replies, "greater," "uncertain" or "equal," "less" with reference either to the standard, or to the variable, or on some other prearranged principle. The percentage of each of these three types of answers is determined for each value of the variable employed, and recorded. The required threshold is taken as that corresponding to the variable which evokes 50 % objectively correct answers, and consequently also 50 % of incorrect and "uncertain" (or "equal") answers.

Taking as an example of the determination of an absolute threshold that of the "spatial threshold," we may say that it is given by 50 % two-point judgments; and as an illustration of the detailed working of the method, we shall find it convenient to refer to a series of results obtained by Riecker (*Zeitschr. f. Biologie*, Bd. 10) which has already served in the descriptions of the method given by Müller, Titchener, and Myers. Riecker obtained the following results in an investigation of the "spatial threshold" or threshold of two-point judgments of the skin of the lower eyelid:

D (distance between the points of the aesthesometer in Paris lines*)	0	0·5	1	1·5	2	3	4	5	6
Z (% of two-pt. judgments)	30	10	14	40	65	80	87	96	100
	(−20)	(4)	(26)	(25)	(15)	(7)	(9)	(4)	

It will be observed that, with two exceptions, the series of percentages follows a general law of increase, the rate of increase itself increasing at first and then diminishing. The numbers within the brackets are formed by subtracting each percentage from the immediately succeeding one, and show this uniformity more clearly. The two exceptions are at $D=0$ and at $D=5$. At $D=0$ the percentage is greater than at the immediately succeeding interval. This irregularity is known as an *inversion of the first order*. The cause of it is doubtless to be looked for in the exceptional way in which the stimulus (one point) may have been applied, the pressure or general nature of the contact may have been different from what they were with two-point contact, or some misleading suggestion may have accompanied these particular experiments.

* A Paris line = 2·25 mm.

At $D=5$, the percentage is indeed larger than that for $D=4$ and smaller than that for $D=6$, but reference to the bracketed numbers shows that the increase from 4 to 5 is greater than the increase from 3 to 4, thus breaking the general rule as regards rate of increase. This irregularity is known as an *inversion of the second order*, and being in the present case slight is probably to be explained as due to an insufficiency in the number of applications of the stimulus.

It is important to realise clearly the fact, which G. E. Müller* was the first to point out and emphasize, that a limen is a variable magnitude following a certain law of frequency-distribution†. There is no fixed limen, only an average limen, a most frequent limen or a most representative limen. The percentages in Riecker's table represent the relative frequencies of limina for distances between the compass-points less than the corresponding D's. Thus there are 65 % limina less than 2 Paris lines, 80 % limina less than 3 lines; i.e. there are 80 − 65, or 15 % limina between the limits of stimulus values 2 lines and 3 lines.

This suggests the plan of plotting a "frequency-polygon" for the limina. For the present case (omitting the anomalous percentage value for 0) the frequency-polygon is given by the accompanying figure (Fig. 3).

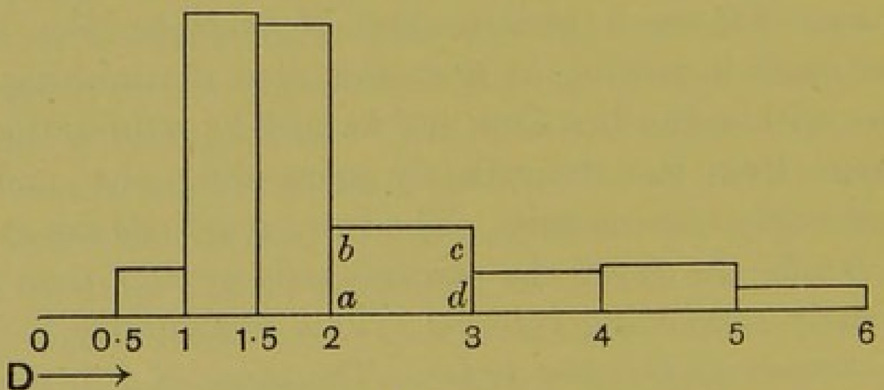


Fig. 3.

* This, at least, is the view to which Müller himself inclines (*op. cit.* p. 59), but he deduces his formulae on the assumption, supported by Fechner and Bruns, that the threshold has a single definite value, D, subject in the course of an experimental determination to variable apparent increase or decrease by random influences, $\pm\delta$, which obey the Gaussian "Law." He points out that both views lead to the same formulae.

† The term "frequency-distribution" is here used in a loose sense. Its technical significance involves the conception of "random samples" taken from

The area of each rectangle represents the percentage or relative frequency of the limina lying between the stimulus-limits indicated by the abscissae of the feet of the two perpendiculars forming its vertical sides. Thus the area of *abcd* is 15. The irregularity of the frequency-polygon or "curve" is probably due to the relatively small number of observations recorded. Taking the limit of an exceeding large number of (hypothetical) observations or cases we might have obtained one or other of the two general types of frequency-curve depicted in Fig. 4:

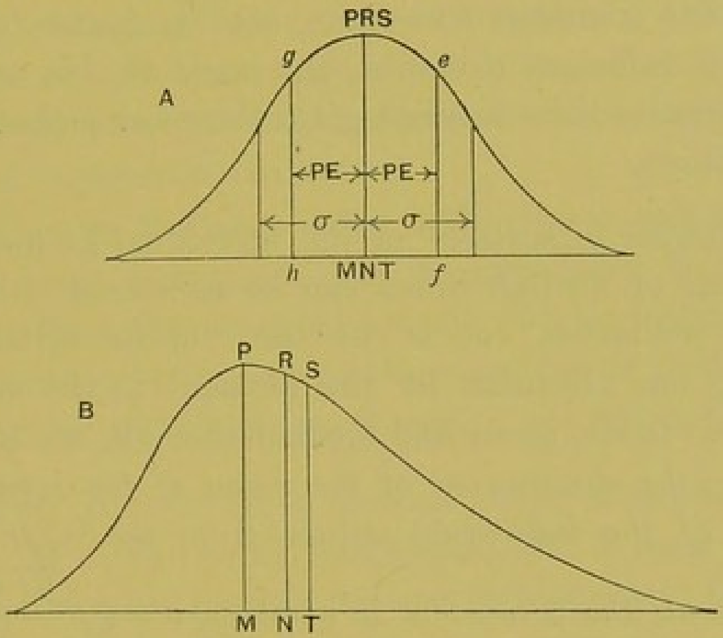


Fig. 4.

A is a symmetrical curve, *B* is known as a skew curve. In both cases the distribution of cases (spatial limina in the present instance) is represented by the area enclosed by the curved line and the horizontal line or axis of abscissae.

There are three ordinates of a frequency-curve which are of special interest. They are:

1. the maximum ordinate, whose abscissa is the *mode*;

2. the centroid vertical, the ordinate passing through the "centre of gravity" of the surface, whose abscissa is the *mean*;

3. the ordinate dividing the surface into two equal areas, whose abscissa is the *median*.

---

a "total population," and is explained more fully in the last section of this chapter. The description on pp. 21, 22 really belongs to this latter account, but is inserted here because necessary for a proper understanding of what follows.

In symmetrical curves, these three ordinates coincide. In asymmetrical or skew* curves they do not. Thus, in Fig. 4, *B*, *PM* is the ordinate corresponding to the mode, *RN* that corresponding to the median, and *ST* that corresponding to the mean. The abscissa of any one of these three ordinates would give us a definite representative liminal value. In fixing upon 50 % two-point judgments as the measure we choose the median. Where interpolation is justifiable, i.e. in results where the number of limina determined is sufficiently large, so that a continuous curve can be found to "fit" the results with sufficient closeness, the *mode* of this curve is the best measure since its abscissa gives the most probable or most frequent limen.

The "degree of scatter" of the individual limina about the modal, mean, or median limen can be measured either by the "standard deviation," σ, or by the "mean variation," M.V. Restricting our attention for the moment to the symmetrical curve, where mode, mean and median coincide, we have as the value of σ, *the square-root of the mean of the squares of the deviations of the individual values from the mean* (*mode or median*) *of all the values*; or in symbols, $\sigma = \sqrt{\frac{S(fx^2)}{N}}$, where N is the total frequency (total number of limina), x is the deviation of any individual value from the mean, f the frequency of each group of equal values and S is the symbol for summation (i.e. $S(fx^2) = f_1x_1^2 + f_2x_2^2 + \ldots + f_Nx_N^2$). The mean variation has already been explained (p. 14). Another measure, called the "probable error," P.E., is in common use. It $= \cdot 67449\,\sigma$, and when employed in a general way, i.e. for all kinds of skew and all kinds of symmetrical curves, may be regarded as merely a conventional reduction of the standard deviation. It only has a peculiar significance of its own in the case of the Gaussian or "normal" curve, in which it corresponds to the ordinate which divides either half of this symmetrical curve into equal areas, so that half the total number of cases are included within

* $\text{Skewness} = \frac{\text{Mean-Mode}}{\text{Standard Deviation}}$.

the limits ± P.E. In Fig. 4, curve A, the area of the curve between the ordinates ef and gh would be half the entire area if the curve were a "normal" curve.

In the case of a skew curve, these measures of "scatter" can be taken about any one of the three "central tendency" measures. They are also different for each half of the curve, although it is usually sufficient to lump the two values together.

It is most important to distinguish carefully the measure of "central tendency" from the measure of "scatter." Both should be evaluated in any piece of research. The second is a measure of the reliability of the first, or at least of the variability of the individual values of which the first is the average or representative value.

To return to Riecker's table. The limen corresponding to 50 % two-point judgments may be calculated in two general ways: (A) from the observed values, (B) by finding the best-fitting smooth curve, adjusting the observations to this, and then calculating the constants (mode, etc., σ, etc.) from the curve.

A. The required limen obviously falls somewhere between 1·5 lines (40 % two-point judgments) and 2 lines (65 %). To determine its value roughly we may substitute in the following formula, devised by Wundt in another connection:

$$D=\frac{D_a(50-Z_b)+D_b(Z_a-50)}{Z_a-Z_b},$$

where D_a = distance giving Z_a % two-point judgments

D_b = " " Z_b % " " "

and Z_a is the percentage just above 50, in the table,

Z_b " " " " below 50, " " "

Substituting,

$$D=\frac{2(50-40)+1{\cdot}5(65-50)}{65-40}$$

$$=\underline{1{\cdot}7}\text{ Paris lines.}$$

This method is open to several objections:

(1) It does not employ all the data; it uses two of the percentages only.

(2) It assumes that between the values D_a and D_b the frequency-curve runs parallel to the axis of abscissae.

(3) It gives no measure of scatter.

A value which *can* be determined by a use of all the data is that of the mean or average limen. The formula to be employed is, of course, the well-known formula for the mean,

$$D = \frac{S(fx)}{N},$$

where x = abscissa of centre of base of any column,

f = area of corresponding column, i.e. the relative frequency,

and N = total relative frequency.

Applied to our example, this gives

$$D = \frac{\cdot 75 \times 4 + 1\cdot 25 \times 26 + 1\cdot 75 \times 25 + 2\cdot 5 \times 15 + 3\cdot 5 \times 7 + 4\cdot 5 \times 9 + 5\cdot 5 \times 4}{4 + 26 + 25 + 15 + 7 + 9 + 4}$$

$$= \frac{203\cdot 75}{90} = \underline{2\cdot 26.}$$

It has been recently suggested by Dr Spearman* that the most representative ordinate of a column (that about which the limina of the column may be most accurately assumed to be concentrated) is not the mid-ordinate, as is assumed in the above formula, but an ordinate shifted slightly towards the higher of the two neighbouring columns (see footnote†). He therefore proposes to replace f of the preceding formula by a, where

$$a_k = D_k + \frac{D_{k+1} - D_k}{3} \cdot \frac{2f_{k+1} + 3f_k + f_{k-1}}{f_{k+1} + 2f_k + f_{k-1}}.$$

* The Method of "Right and Wrong Cases" ("Constant Stimuli") without Gauss's Formulae, *Brit. Journ. of Psychology*, Vol. II. 1908, pp. 227—242.

† *Figure given by Spearman, op. cit.* p. 241 (slightly simplified).

The mid-ordinate c_k is replaced by the ordinate a_k which passes through the centroid of the trapezium $D_k y_k y_{k+1} D_{k+1}$.

y_k is the mid-point of $af_k \{= \frac{1}{2}(f_{k-1} + f_k)\}$,

y_{k+1} ,, ,, ,, bf_{k+1},

f_k = average relative frequency for kth column.

There is clearly some serious misconception here, since it has long been known that the mean of any frequency-distribution needs no correction*. A glance at the proof (see footnote) reveals the source of the error. Spearman replaces each rectangle by a trapezium of *different* area and then multiplies the abscissa of the centroid vertical of this new area by the *old* area, which area, however arranged, could not possibly have the centroid of the new one. The unadjusted value is the best value of the mean, but it is doubtful whether this is of much significance as a measure of the limen.

B. *Assuming "normal" distribution* (Fechner, Müller).

The classical description of this method is given in G. E. Müller, *Die Gesichtspunkte und die Tatsachen der psychophysischen Methodik*, pp. 45—50, 56—58, and the following

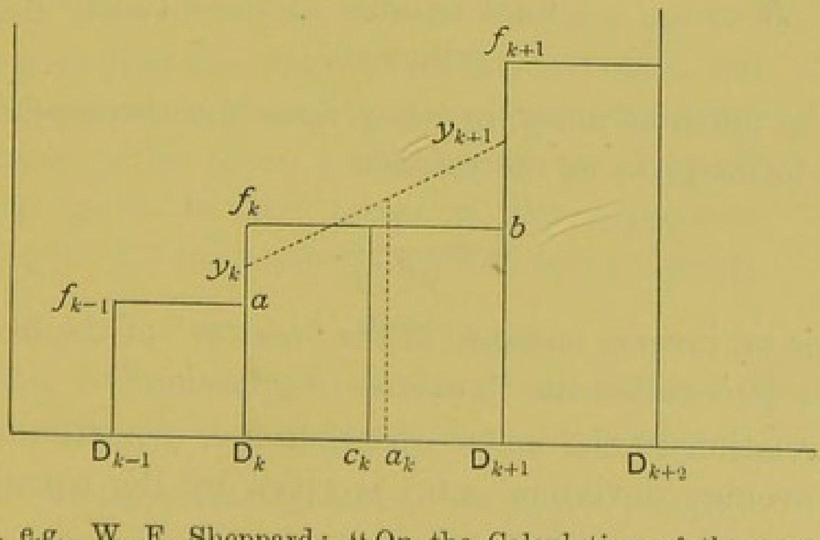
---

This gives at once the abscissa

$$a_k = D_k + \frac{\int_0^{D_{k+1}-D_k} xy\,dx}{\int_0^{D_{k+1}-D_k} y\,dx},$$

which reduces to the required formula,

$$a_k = D_k + \frac{D_{k+1}-D_k}{3} \cdot \frac{2f_{k+1}+3f_k+f_{k-1}}{f_{k+1}+2f_k+f_{k-1}}.$$



* Cf., e.g., W. F. Sheppard: "On the Calculation of the most Probable Values of Frequency Constants, etc.," *Proc. Lond. Math. Soc.*, Vol. XXIX. 353 ff.; and also below, p. 38.

account aims at being merely a condensation of this description. But first a few words of preliminary explanation. The form of

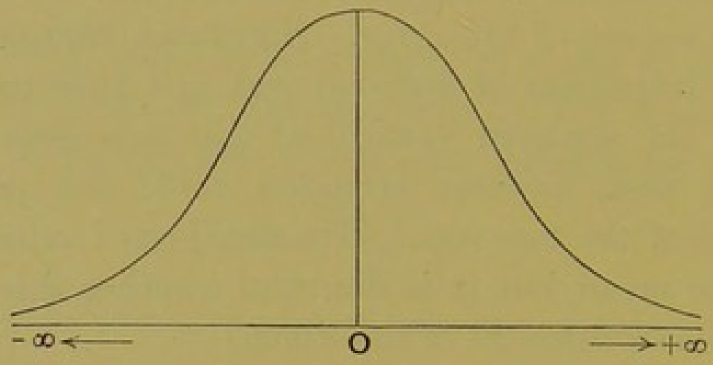


Fig. 5.

distribution of the individual limina is assumed to approximate with a fair degree of closeness to the so-called Gaussian or normal curve, or "curve of error," the equation to which is generally written in the form

$$y = \frac{h}{\sqrt{\pi}} e^{-h^2x^2}$$

(δ is sometimes put in place of x, to represent "error": see later).

A more convenient form, however, is that employed by the biometricians, viz.

$$y = \frac{N}{\sqrt{2\pi} . \sigma} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}},$$

in which σ, as usual, = the standard deviation

and N = total number of cases (area) of the distribution.

N may be taken as unity, in which case the correspondence of the two forms gives us the relation

$$h = \frac{1}{\sqrt{2} . \sigma}.$$

Thus h is an inverse measure of the "scatter" of the frequency-surface. It is called the "measure of precision."

Its relation to the mean variation M.V. (sometimes known as the average deviation, A.D.) is given by the equation

$$\text{M.V.} = \frac{1}{h\sqrt{\pi}}.$$

Thus, again, in the case of the normal curve,

$$\sigma = \sqrt{\frac{\pi}{2}} \cdot \text{M.V.} = 1{\cdot}2533 \text{ M.V.}$$

Finally, $$\text{P.E.} = {\cdot}67449\,\sigma = \frac{{\cdot}4769}{h}.$$

Obviously the mode, mean, and median coincide.

The ordinates corresponding to $x = \pm\,\sigma$ cut the curve in the two points of inflexion (i.e. the points where the curve changes from concavity to convexity, or vice versa).

The mathematical deduction of the equation of the probability curve is based upon the general theory of independent probability.

It is developed from the binomial expansion of $(p+q)^n$, viz.

$$p^n + np^{n-1}q + \frac{n(n-1)}{1\,.\,2}p^{n-2}q^2 + \ldots + \frac{n(n-1)}{1\,.\,2}p^2q^{n-2} + npq^{n-1} + q^n,$$

the terms of which represent the distribution of chances of n independent events, where the probability of occurrence of each single event in one trial is p and the probability against its occurrence is q (i.e. $p+q=1$). The first term is the probability that all will happen, the second that $n-1$ will happen and 1 fail, the third that $n-2$ will happen and 2 fail, and so on. The single events must be independent of one another. Thus if n teetotums are taken, each of which is marked in the ratio of p to q, so that the chance that any teetotum will fall on the marked piece is p, then the distribution of chances when these n teetotums are spun N times is given by the terms of the expansion, $N(p+q)^n$. In the stock example of coin spinning,

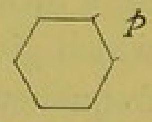


Fig. 6.

$$p = q = \tfrac{1}{2}.$$

It can be proved that the Gauss curve is an extremely close approximation to the binomial frequency polygon

(1) whatever be the value of n, if $p = q = \frac{1}{2}$,

(2) with inequality of p and q, provided that neither p nor q is very small, and that n is very large.

It is *not* in agreement

(1) if p or q is very small,

(2) if there is correlation between the events or contributory causes, i.e. if the events are not completely independent.

The standard deviation (σ) of the binomial frequency $= \sqrt{npq}$, that of the Gaussian $= \sqrt{(n+1)pq}$.

The above brief account must suffice for the theoretical discussion of the Gauss curve*. Returning to Müller's method, we find it *assumed* that the successively increasing percentages of two-point judgments correspond to the increasing area of the portion of a Gaussian frequency-distribution† included between a limiting ordinate OP (see accompanying figure), the axis of x, the curve and the moving ordinate pD. Thus the area Pp_1D_1O represents the percentage of two-point judgments given for an

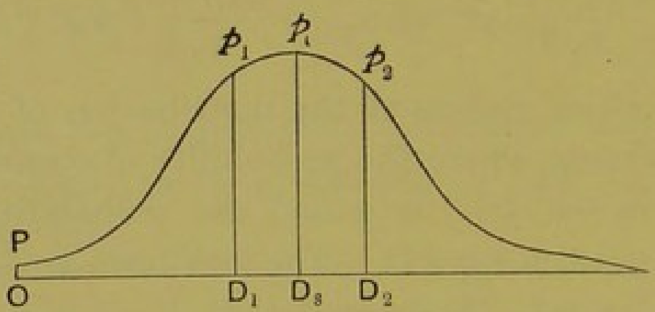


Fig. 7.

objective distance between the compass points represented by the abscissa OD_1, the area Pp_2D_2O represents the percentage for a distance OD_2. Our task is to determine the distance OD_s which gives 50 % two-point judgments, represented by half the frequency-surface, viz. the area Pp_sD_sO. Let us call the average limen (represented by OD_s in the figure) S, and any particular stimulus-value employed, D. Then, in a sufficiently large number of experiments with $D > S$ (e.g. $= OD_2$ in figure), so that

* For a very clear elementary exposition, see Titchener, *Experimental Psychology*, Vol. II. *Student's Manual*, pp. 38—55, on "The Law of Error." Also Merriman, *Method of Least Squares*, pp. 6—25.

† As stated above, p. 20, footnote, there is, strictly speaking, no justification for treating these percentages as parts of a "frequency-distribution." This will be made more clear later on.

the percentage, $Z, > 50$, taking the whole area of the surface as 1, we have

$$Z = \text{area } Pp_2D_2O = \text{area } Pp_sD_sO + \text{area } p_sp_2D_2D_s$$
$$= \tfrac{1}{2} + \frac{h}{\sqrt{\pi}} \int_0^{D-S} e^{-h^2\delta^2}\, d\delta$$

or, putting $h\delta = t$,

$$Z = \tfrac{1}{2} + \frac{1}{\sqrt{\pi}} \int_0^{(D-S)\,h} e^{-t^2} dt.$$

Similarly, with $D < S$ (e.g. $= OD_1$ in figure), we have

$$Z = \text{area } Pp_1D_1O = \text{area } Pp_sD_sO - \text{area } p_1p_sD_sD_1$$
$$= \tfrac{1}{2} - \frac{h}{\sqrt{\pi}} \int_{-(S-D)}^{0} e^{-h^2\delta^2} d\delta$$
$$= \tfrac{1}{2} - \frac{1}{\sqrt{\pi}} \int_{-(S-D)h}^{0} e^{-t^2} dt.$$

Turning now to our experimental results, if we omit the first and last series ($D = 0$ and $D = 6$) for reasons previously given, viz. that 30 % for $D = 0$ is an inversion of the first order, and 100 % for $D = 6$ shows that the limits of fluctuation of S have been passed, and call the other D's $D_1, D_2, \dots D_7$, and the corresponding percentages $Z_1, Z_2, \dots Z_7$, then the values $t_1, t_2, \dots t_7$ corresponding to the Z's are given by *Fechner's Fundamental Table* (see Appendix I, Table I); and we have the equations:

$$t_1 = (D_1 - S)\, h,$$
$$t_2 = (D_2 - S)\, h,$$
$$\dots\dots\dots\dots\dots\dots$$
$$t_7 = (D_7 - S)\, h,$$

which, if t_1, t_2, etc. were "observed" values, could be solved for S and h by the Method of Least Squares.

But these t-values are not observed values; they are theoretically deduced from the observed Z-values. Hence we must, as it were, change them into observed values by "weighting" them. They are to be weighted both by w', a weight proportional to the number of observations upon the basis of which the corresponding Z's were determined and also by w''

to change them into observed values. Thus t_1 is to be multiplied by $w_1'w_1''$, t_2 by $w_2'w_2''$, and so on. Since, however, the number of observations is almost universally equal in each series of experiments (giving the different Z's), w' is in almost all cases = 1, so that w'' is alone needed. This is given by *Müller's Table of Coefficients of Weights** (see Appendix I, Table II). We thus have the equations:

$$w_1 t_1 = (D_1 - S)\,h,$$
$$w_2 t_2 = (D_2 - S)\,h,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$w_7 t_7 = (D_7 - S)\,h,$$

which, solved by the Method of Least Squares, give the "normal" equations:

$$\Sigma\,(D^2 w)\,.\,h - \Sigma\,(Dw)\,.\,Sh = \Sigma\,(Dtw),$$
$$-\,\Sigma\,(Dw)\,.\,h + \Sigma\,(w)\,.\,Sh = -\,\Sigma\,(tw),$$

where $\Sigma\,(D^2w) = (D_1{}^2 w_1 + D_2{}^2 w_2 + \ldots + D_7{}^2 w_7)$, and similarly with the others.

Titchener has worked out Riecker's results by this method, and gets

$$\underline{h = 0{\cdot}49}, \quad \underline{S = 1{\cdot}88.}$$

In Fechner's original discussion of this method, there was no sufficiently clear distinction made between the quantities S and h. It was assumed that S and $\frac{1}{h}$ measure much the same quantity, i.e. that the product Sh is a constant. Müller was the first to distinguish clearly between the two measures in psycho-physics and to point out that Sh is by no means always a constant quantity, although approximating to it in certain cases. S and h do not always vary inversely with one another, in a series of determinations of a particular threshold. In some cases they show a tendency to vary directly.

The same method is applicable to the determination of a *difference-limen*. In Müller's form of the method the most

* A more correct table is that suggested by F. M. Urban, which is also given in Appendix I.

representative limen is taken to be that corresponding to 50 % "right answers" (and therefore also 50 % wrong + uncertain or "equal" answers). The formulae required are:

$$g = \tfrac{1}{2} - \frac{1}{\sqrt{\pi}} \int_0^{(DL_l \pm D) h_l} e^{-t^2}\, dt,$$

$$l = \tfrac{1}{2} + \frac{1}{\sqrt{\pi}} \int_0^{(\pm D - DL_u) h_u} e^{-t^2} dt,$$

$$u = \frac{1}{\sqrt{\pi}} \int_{(\pm D - DL_u) h_u}^{(DL_l \pm D) h_l} e^{-t^2}\, dt\,;$$

where $g = \frac{\%}{100}$ answers "greater," l = less, u = equal or uncertain, DL_u = upper difference-limen, DL_l = lower difference-limen, D = any particular stimulus-difference ($V - S$) employed, and is $+^{ve}$ or $-^{ve}$ according as the variable, V, is greater or less than the standard, S.

h_u, h_l are the measures of precision of the upper and lower limen, respectively.

The results—*DL*'s and *h*'s—are obtained, as before, by means of Fechner's Fundamental Table and Müller's Table of Coefficients of Weights*.

Where time-errors or space-errors† or both are possible, the values of *DL* and *h* are determined separately for each time or space order. In these cases the distributions often tend to be "skew," even when the results all lumped together give an approximation to the Gaussian curve. Indeed Müller himself emphasizes the unjustifiability of the assumption that all distributions obtained by the constant method must be normal or even must be symmetrical, and shows that for certain of Merkel's results‡ a "zweiteiliges Gaussches Gesetz" gives a very good fit. This "bipartite" or compound Gaussian law differs from the ordinary one simply in having two *h*'s or measures of precision, one for the positive deviations of the individual liminal values from their central value, and another, a different one, for the negative deviations. For a brief

* *Better*, Urban's Table.

† See above, pp. 15, 16.

‡ G. E. Müller, *op. cit.* pp. 92, 93.

criticism of the law of distribution suggested by Bruns, and known by his name, the reader is referred to Müller, *op. cit.* pp. 90, 91. The whole question, though important, is too difficult to come within the scope of this book.

Correspondence of Results obtained by the Three Psycho-Physical Methods.

It has already been pointed out that the average error or mean variation of the Method of Average Error corresponds closely to the mean variation obtained from the distribution of "equal" or "uncertain" answers in the Constant Method. The relation existing between determinations of liminal-values by the Limiting Method and Constant Method, respectively, is more complex, and may be discussed from two somewhat distinct points of view, viz. the mathematical and the psychological. Assuming identical or approximately identical psychological conditions in the employment of the two methods, the result (value of the limen) will be the same if the chance values of the limen form a normal frequency-distribution, different if this frequency-distribution is skew. In the latter eventuality the limen given by the Limiting Method will, it is clear, be greater or less than the limen given by the Constant Method according as the frequency-distribution has its mode towards the left (zero end) or towards the right. In the ordinary forms of the two methods, however, the psychological conditions must be very different. They can be approximated to one another by getting the subject, in the Constant Method, to reply "much larger," "much smaller" as well as "larger," "smaller*." In calculating the result only the last two series of answers are employed.

It has generally been held hitherto that the Constant Method gives more reliable results than the Method of Limits, but F. M. Urban† has recently brought forward reasons for

* Cf. Ebbinghaus: *Grundzüge der Psychologie*, 1905, p. 89; but see Müller, *op. cit.* p. 160.

† F. M. Urban: "On the Method of Just Perceptible Differences," *Psychological Review*, Vol. XIV. 1907, pp. 244—253. See, too, Vol. XVII. 1910, "The Method of Constant Stimuli and its Generalizations," pp. 229—259, by the same author.

believing that, under certain conditions of experimentation, the converse is the case. He employs the following formula for the measure (M) of the limen:

$$M = \frac{1}{N}(r_1 P_1 N + r_2 P_2 N + \dots + r_n P_n N) = r_1 P_1 + r_2 P_2 + \dots + r_n P_n,$$

where $r_1 < r_2 < \dots < r_n$ are the successive values of the variable-stimulus (V) employed, and

$$P_1 = p_1,$$
$$P_2 = q_1 p_2,$$
$$P_3 = q_1 q_2 p_3,$$
$$\dots\dots\dots$$
$$P_n = q_1 q_2 \dots q_{n-1} p_n,$$

where, again, the p's and q's need to be explained; thus

p_k is the probability that in the comparison of the stimulus r_k with the standard the judgment "greater" will be given,

and

q_k is the probability that a judgment will be given which is not a "greater" judgment, and therefore $= 1 - p_k$.

$N =$ the number of series of experiments made.

This modified form of the method evidently approximates closely to the Constant Method. The p's are the percentage-values obtained in the latter. The P's are more reliable, though smaller, than the p's. P_k is a measure of the component probability that r_k is judged "greater" and that on all the smaller r ($r_1 \dots r_{k-1}$) judgments are given which are not "greater" judgments. Since the value of the P's is not changed by the order in which the r's are presented, these latter may be presented in any order, thus eliminating the influence of expectation, and bringing the method very close indeed to the Constant Method. Another important point is "to vary the steps 'by which one approaches the threshold,' because otherwise one can not make the supposition of a symmetrical distribution" (p. 247). Urban compared the results obtained by means of his formula, using lifted weights, with the results obtained empirically by (as it were) directly observing P_k, or something proportional to P_k, viz. N_k, "the number of times it occurred that each weight [in this case, r_k] was the lightest weight of the entire series to be judged 'heavier'" (p. 249). The agreement of observed results and theoretical results was found to be very close in all the cases. The article concludes as follows: "The theoretical basis of the method of just perceptible differences is the same as that of the error method, namely empirical determinations of the probabilities of judgments of different types on given differences of intensity. The result of the so-called method of just perceptible differences is that amount of difference for which there exists the probability one half that it will be recognised"* (p. 253).

One aspect of the psycho-physical methods has been scarcely mentioned in all the preceding description, viz. the attitude of the subject. Its importance is hardly likely to be overlooked in the present state of opinion in psychology. It is superfluous to emphasize the need of careful introspective reports (given either at the end of the experimental series or after each observation) in all quantitative determinations, liminal and supraliminal alike.

The degree of "fore-knowledge" of the subject, the form in which he is asked to give his judgment—in terms of standard or variable, of right-hand or left-hand stimulus, etc.—and other recognisable psychological factors directly bearing upon the quantitative determination, should be varied in as many ways as possible and the corresponding results noted; the effects of practice, fatigue, "contrast," "side-comparisons," "absolute impression," etc. should be reduced to a minimum or else rendered constant and, where possible, measured; in short, the quantitative determinations should always be made for the sake of the psychological processes involved, in the fullest sense of the phrase, and not conversely. The problems of the psychological processes of discrimination, judgment, etc. thus raised are of the first importance, but their discussion would take us beyond

* Cf. Külpe: the probable value of the limen is "that stimulus difference which is just as often cognized (correctly judged) as not cognized (incorrectly judged)"—quoted from Sanford; *Experimental Psychology*, footnote, p. 351.

the scope of this book. We can only just mention them here, emphasize their importance, and indicate the relation in which they stand to mental measurement in the narrowest sense.

Frequency-Distributions.

We have seen above that the graphical representation of the percentages of any one of the three kinds of judgments obtained in the Constant Method, although sometimes approximating to the form of a Gaussian frequency-distribution, is in many cases decidedly "skew." In these latter cases the most representative limen would be given by the mode of the smooth curve which best fits the figure. Unfortunately the determination of the equation of such a best-fitting curve must be almost entirely empirical in nature, since we have so slight a theoretical basis on which to build. Urban's notion of the probability of a judgment, illustrated in the preceding section, seems to be the most promising as a starting-point, and there is little doubt that his publications* are the most important for the modern student of psycho-physics. Much still remains to be done, however, and great caution is needed in the theoretical estimation of results. We have already suggested that the identification, or even the approximation of graphical representations of "percentage" judgments to true frequency-distributions is of extremely doubtful validity. This will become clearer on a consideration of the nature of the latter. True frequency-curves represent the distribution of "random samples" taken from a "total population," and therefore have a definite mathematical basis in the theory of probability. They correspond to the results obtained in drawing balls from a bag, or in teetotum-spinning, dice-throwing, etc. Cases of this kind obviously occur in great numbers in psychology,— e.g. the distribution of a particular kind of mental ability in

* In addition to articles already quoted, see "Die psychophysischen Massmethoden als Grundlagen empirischer Messungen," *Archiv f. d. ges. Psychologie*, Vols. xv. and xvi.; also, *The Application of Statistical Methods to the Problems of Psychophysics*, Philadelphia, 1908.

a group of individuals,—and hence, though not strictly relevant to the subject of the present chapter, they are relevant to the general problem of mental measurement, and are most conveniently discussed at this point.

The general theory of curve-fitting has been worked out in great detail by Professor Karl Pearson*. A good account of his method is given in a book by W. Palin Elderton, *Frequency-Curves and Correlation*, C. and E. Layton, London, 1906, to which the mathematical reader is referred for fairly complete theoretical and practical information. The following short sketch of the outline of Pearson's system will perhaps serve as a reminder for readers already acquainted with the methods, and in any case will be of some little use in indicating the general significance of biometric notation which will be employed later on in this book.

Frequency-curves of data not involving a mixture of species tend to commence at zero, rise to a maximum and then fall either at the same or at a different rate. There is often high contact at one or both ends of the distribution. An equation of the general form

$$\frac{dy}{dx} = \frac{(x+a)\,y}{f(x)}$$

satisfies these conditions, since if $y = 0$, $\frac{dy}{dx} = 0$ (high contact) and if $x = -a$, $\frac{dy}{dx} = 0$ (maximum; for a maximum, again, the second differential coefficient must be negative). Expanding $f(x)$ by Maclaurin's theorem, we have

$$\frac{dy}{dx} = \frac{(x+a)\,y}{c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \dots}.$$

* Karl Pearson, "Skew Variation in Homogeneous Material," *Phil. Trans.* Vol. 186 A, 343 ff.; "On the Systematic Fitting of Curves to Observations and Measurements," *Biometrika*, Vol. I. pp. 265 ff. and Vol. II. pp. 1 ff., 1901—1903; "On the Curves which are most suitable for describing the frequency of Random Samples of a Population," *Biometrika*, Vol. V. 1906, 172—175 (an exceedingly clear summary of the principles involved).

(1) Putting $c_1 = c_2 = c_3 = \ldots = 0$, we have

$$\frac{1}{y}\frac{dy}{dx} = \frac{x+a}{c_0},$$

which is the Gaussian or normal curve. It fits the symmetrical binomial, $(\frac{1}{2}+\frac{1}{2})^n$, e.g. in coin-tossing, where the chances for and against are equal ($p=q$), and the contributory causes are independent of one another.

(2) Putting $c_2 = c_3 = \ldots = 0$, we have

$$\frac{1}{y}\frac{dy}{dx} = \frac{x+a}{c_0 + c_1 x},$$

which represents a class of curves varying from the Gaussian curve to the *J*-curve. It fits the asymmetrical or "point" binomial, $(p+q)^n$, e.g. in teetotum-spinning or dice-throwing, where the chances for and against are not equal, $p \neq q$, but the contributory causes are independent of one another.

(3) Putting $c_3 = c_4 = \ldots = 0$, we have

$$\frac{1}{y}\frac{dy}{dx} = \frac{x+a}{c_0 + c_1 x + c_2 x^2},$$

which can be made to represent almost all the frequency distributions which may arise. It fits the "hypergeometrical series," the successive terms of which give, e.g., the chances of getting $r, r-1, \ldots 0$ black balls from a bag containing pn black and qn white balls when r balls are drawn*.

Here the contributory causes are *not* independent of one another.

There is no advantage in employing equations which involve c_3 and higher constants, because their use involves the calculation of the 6th and higher "moments," and these have very high probable errors.

Definition. The *nth moment coefficient* (μ'_n) of any distribution about *any* ordinate is the sum of the products of the

* The series is:

$$\frac{pn(pn-1)\ldots(pn-r+1)}{n(n-1)\ldots(n-r+1)}\left\{1 + \frac{rqn}{pn-r+1} + \frac{r(r-1)}{\underline{|2}}\cdot\frac{qn(qn-1)}{(pn-r+1)(pn-r+2)} + \ldots\right\}$$

Other series may arise.

partial-frequencies and the nth power of the distances of these frequencies from the ordinate, divided by the total frequency.

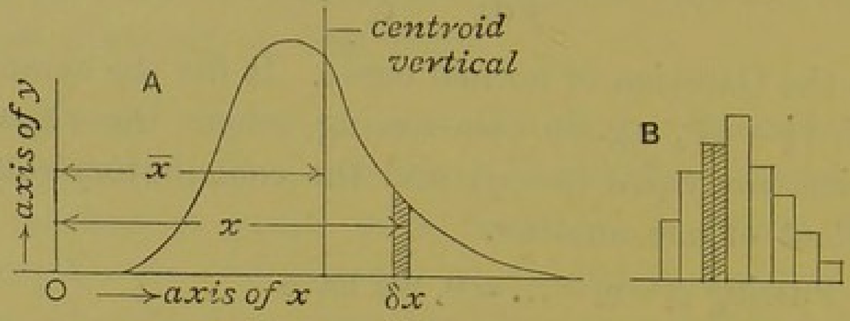


Fig. 8.

In symbols, if N be the total frequency,

$$N\mu'_n = \int x^n y \,\delta x.$$

The moments are, in practice, first calculated about any arbitrary ordinate that is most convenient, and then reduced to moments about the centroid in the following way*:

$$N\mu_n = \int (x - \bar{x})^n y \,\delta x$$
$$= N\mu'_n - n\bar{x}N\mu'_{n-1} + \frac{n(n-1)}{1.2}\bar{x}^2 N\mu'_{n-2} - \ldots.$$

This gives the general reduction formula

$$\mu_n = \mu'_n - n\mu'_1\mu'_{n-1} + \frac{n(n-1)}{1.2}\mu'^2_1\mu'_{n-2} - \frac{n(n-1)(n-2)}{1.2.3}\mu'^3_1\mu'_{n-3} + \ldots$$

which enables us to transfer any moment from an arbitrary ordinate to the mean. Thus we have:

$$\mu_2 = \mu'_2 - \mu'^2_1,$$
$$\mu_3 = \mu'_3 - 3\mu'_1\mu'_2 + 2\mu'^3_1,$$
$$\mu_4 = \mu'_4 - 4\mu'_1\mu'_3 + 6\mu'^2_1\mu'_2 - 3\mu'^4_1.$$

μ's represent moments of the *curve*; but we have to start with grouped frequencies, where the frequencies are assumed to be concentrated along the mid-ordinates of the rectangles (cf. Fig. 8, B). The moments obtained from these grouped frequencies are denoted by ν's (dashed and undashed), and corrections are necessary. These have been deduced by Sheppard† and are consequently known as Sheppard's adjustments.

* *Dashed* μ's represent moments about an arbitrary ordinate,
undashed μ's ,, ,, ,, the centroid vertical.

† W. F. Sheppard: "On the Calculation of the most Probable Values of Frequency Constants, for Data arranged according to Equidistant Divisions of

They are: $\mu_2 = \nu_2 - \frac{1}{12}$,

$\mu_3 = \nu_3$,

$\mu_4 = \nu_4 - \frac{1}{2}\nu_2 + \frac{7}{240}$.

It is generally said that they are only valid when there is "high contact" at the ends of the frequencies, but the equations for μ_2 and μ_3 are probably still valid, even without high contact, if the terminal frequencies are zero.

[N.B. In working, ν''s are changed into ν's *before* applying Sheppard's corrections.]

It is obvious that $\nu_0 = 1$ and $\nu_1 = 0$. ν'_1 is the distance of the mean or centroid vertical from the arbitrary ordinate about which the moments are first taken, and is conveniently known as d. $\sigma = \sqrt{\mu_2}$.

Two very important constants in curve-fitting are

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}, \text{ and } \beta_2 = \frac{\mu_4}{\mu_2^2}.$$

The values of these are always to be calculated, and, within the limits of their probable errors (for which see Rhind's Tables, *Biometrika*, Vol. VII. 1909), they fix the type to which the curve belongs. The general frequency-curve equation, written in terms of moments, is

$$\frac{1}{y}\frac{dy}{dx} = -\frac{x + \sigma\dfrac{\sqrt{\beta_1}}{2}\dfrac{(\beta_2+3)}{5\beta_2 - 6\beta_1 - 9}}{\sigma^2\left\{\dfrac{4\beta_2 - 3\beta_1}{10\beta_2 - 12\beta_1 - 18} + \dfrac{\sqrt{\beta_1}}{2}\dfrac{(\beta_2+3)}{5\beta_2 - 6\beta_1 - 9}\cdot\dfrac{x}{\sigma} + \dfrac{2\beta_2 - 3\beta_1 - 6}{10\beta_2 - 12\beta_1 - 18}\cdot\left(\dfrac{x}{\sigma}\right)^2\right\}}.$$

This gives at once, for the distance between the mean and mode,

$$x = -\frac{\sigma\sqrt{\beta_1}(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)}. \quad \text{(Origin is at mean.)}$$

Hence the curve is symmetrical if $\beta_1 = 0$. If $\beta_1 = 0$, and $\beta_2 = 3$, the curve reduces to the Gaussian, since the terms involving x in the denominator of the right-hand side of the general frequency-curve equation vanish.

In using the Pearsonian method, then, the order of procedure to be adopted is:

a Scale," *Proc. Lond. Math. Soc.* Vol. XXIX. pp. 353 ff. Karl Pearson: "On an Elementary Proof of Sheppard's Formulae for correcting Raw Moments and on other allied Points," *Biometrika*, Vol. III. 1904, pp. 308 ff.

(1) calculate the moment-coefficients ν'_1, ν'_2, ν'_3, ν'_4 about a convenient arbitrary ordinate,

(2) transfer to the mean by the equations

$$\nu_2 = \nu'_2 - \nu'^2_1,$$
$$\nu_3 = \nu'_3 - 3\nu'_1\,\nu'_2 + 2\nu'^3_1,$$
$$\nu_4 = \nu'_4 - 4\nu'_1\,\nu'_3 + 6\nu'^2_1\,\nu'_2 - 3\nu'^4_1$$

(ν'_1, or d, is the distance of the mean from the arbitrary ordinate),

(3) determine the corresponding moments for the curve by the equations

$$\left.\begin{aligned} \mu_2 &= \nu_2 - \tfrac{1}{12} \\ \mu_3 &= \nu_3 \\ \mu_4 &= \nu_4 - \tfrac{1}{2}\nu_2 + \tfrac{7}{240} \end{aligned}\right\} \text{ Sheppard's corrections.}$$

[N.B. For these corrections to be applicable, two conditions must be fulfilled:

(i) there must be high contact,

(ii) the grouping of the frequencies must be *equal*.]

(4) Calculate β_1 and β_2 by the equations

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3},\ \beta_2 = \frac{\mu_4}{\mu_2^2}.$$

These results give the distance of the *mean* (ν'_1 or d) from the arbitrary ordinate, the *standard deviation* ($\sqrt{\mu_2}$), and the *mode*,

$$\text{mean} - \frac{\sigma\sqrt{\beta_1}\,(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)}.$$

The investigator may then proceed to determine to which of the seven Pearsonian "types" the particular curve belongs, to find its equation, and to plot it. For instructions on these points the reader is referred to Palin Elderton's book already mentioned.

PART II

CORRELATION

INTRODUCTION

A somewhat detailed account of the mathematical theory of correlation and of the way in which it may be usefully applied to psychological measurements will be found in the later chapters of this Part. The object of the following introductory pages is to give the reader a general preliminary view of the method, free from mathematical complications, and to illustrate it by means of a simple example.

Correlation may be briefly defined as "tendency towards concomitant variation," and a so-called correlation coefficient (or, again, correlation ratio) is simply a measure of such tendency, more or less adequate according to the circumstances of the case. J. S. Mill, in his "System of Logic," distinguished a special scientific "Method of Concomitant Variations," which he based upon the following principle:

"Whatever phenomenon varies in any manner whenever another phenomenon varies in some particular manner, is either a cause or an effect of that phenomenon, or is connected with it through some fact of causation*."

The instances of this principle which Mill had in mind were mainly cases of approximately "complete" concomitance of variation, such as those usually met with in the domain of Physics. In such cases, the conditions of an experiment admit of a high degree of simplification, the phenomenon, or series of

* *Logic*, Bk. III. Ch. viii. § 6.

phenomena, under investigation can be isolated with tolerably complete success, the "irrelevant" factors can be reduced to a minimum. Under such conditions, when the degree of concomitance of the different corresponding measures of the two phenomena is found to be very high, the slight deviations from complete correspondence are put down to "errors of observation" or other unavoidable imperfections in the experimental method employed.

If the correspondence is one of simple proportionality, so that the graphical representation of it (one phenomenon being measured along the axis of x, the other along the axis of y) is a straight line*, the correlation coefficient, r, will be $= 1$. *Example*: the variation of the length of a metal rod with temperature.

If the correspondence, although still approximately complete, is not one of simple proportionality, the graphical representation of it will be, not a straight line, but a curve of greater or less degree of complexity†, and the correlation, also complete, will be measured not by the correlation coefficient, r, but by the correlation *ratio*, η. η in this case will be $= 1$. *Example*: the variation (inverse) of the volume of a certain quantity of gas with the pressure to which it is subjected, the temperature remaining constant. A number of pairs of values, P_1, V_1; P_2, V_2; P_3, V_3; etc. are obtained, and when plotted are found to give a "scatter diagram" of the approximate form of Figure 9.

In this figure, the dots represent the individual pairs of observations, P, V. They cluster very closely about the hyperbola, $PV = k$, represented by the broken curve. The curve is assumed to represent the "real" or "true" relation of the two "variates" (as we call such quantities as P, V), and the slight deviations of the observed values from this curve are explained as due to errors of observation and to other factors irrelevant to the relation under investigation. However this may be, the interesting point about the figure so far as our

* Hence the correlation is said to be "linear."

† Correlation said to be "non-linear" or "skew."

present purpose of explaining correlation is concerned is that any definite observed P-value is "correlated" with a plurality or "array" of observed V-values, and that, similarly, any

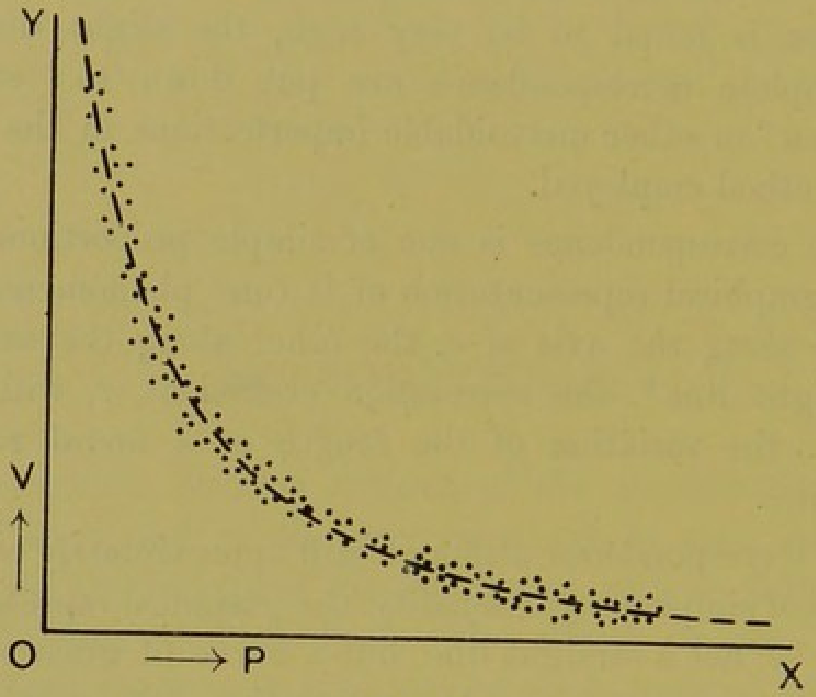


Fig. 9.

definite observed V-value is correlated with a plurality of P-values. These arrays of observed values cluster extremely closely about their means (situated on the curve), i.e. their "scatter" or "variability," as measured by their standard deviations (σ), is extremely small.

The modern theory of correlation is directed towards the manipulation of observations made upon phenomena of a much greater degree of variability than that found in the case of isolated physical phenomena. The increased variability is no doubt due, in the main, to the complexity of factors involved. The elementary factors do not admit of isolation, and with reference to the concomitance of variation of the two series of phenomena under consideration they, as it were, pull in different directions. The correlation coefficient and correlation ratio measure, in these cases, the average extent of the concomitance. As will be explained more fully in the next chapter, r can only be taken as a measure of correlation when the average relation between the two variates is *linear*, and in this case its value is identical with that of η. When the relation is non-linear r is practically meaningless, but η still measures the relation accurately.

The general problem will become clearer by reference to the accompanying figure.

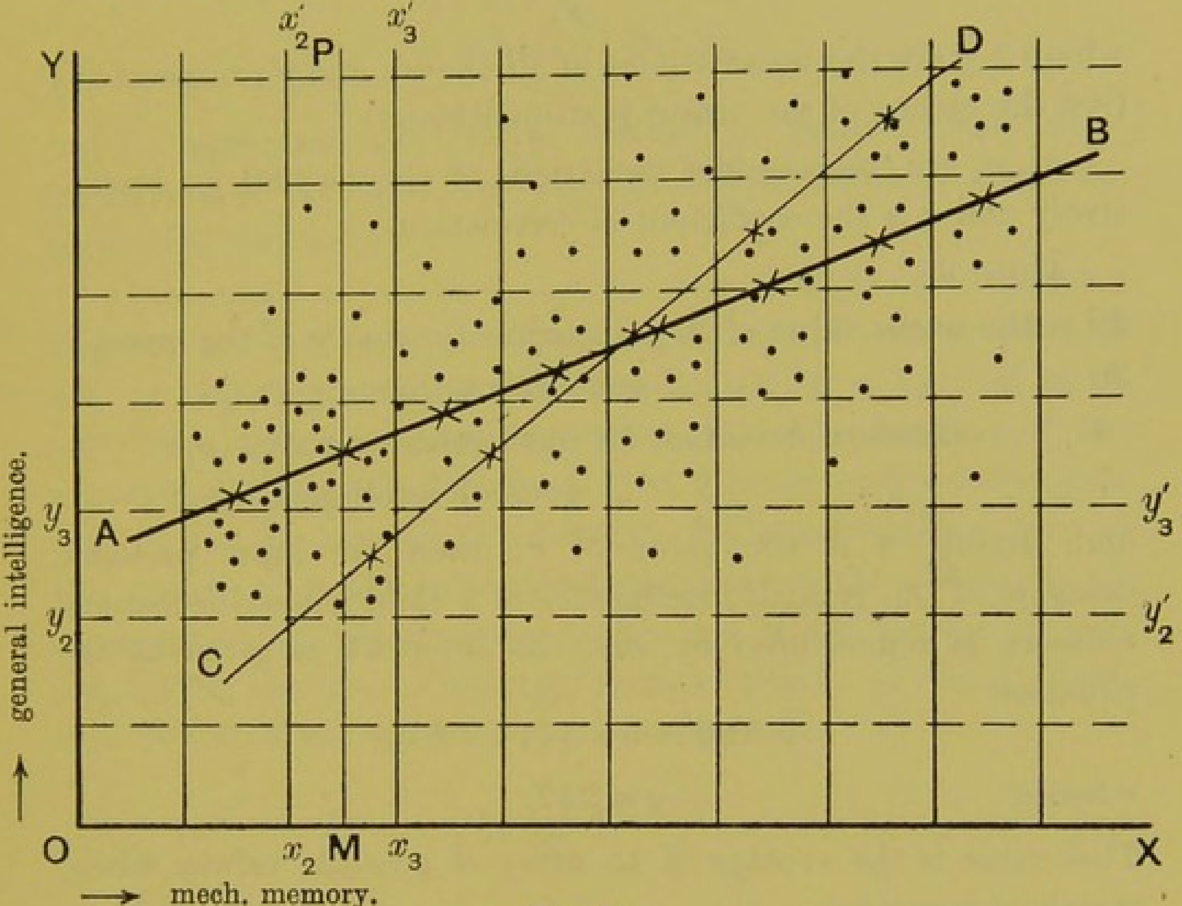


Fig. 10.

Let us assume that we have a group of 200 school-children and have measured each of them for mechanical memory (x) and for general intelligence (y). Each of the dots in the figure represents a child. Then if we determine the mean y-values corresponding to each successive "group" of x-values, e.g. x_2 to x_3, by assuming the observations concentrated on the mid-ordinate PM*, the line AB† drawn through these mean y-values (marked by crosses in heavy type) represents the law of change of mean y-value with increase of x and gives the

* The true centroid ordinate is slightly nearer the denser part of the scatter-diagram, here slightly towards the right of PM. The correction is made later by means of Sheppard's formulae (see above p. 39, and Appendix II, p. 137).

† I have placed the means on the straight line for the sake of convenience of exposition. Actually, they will occur irregularly on either side of it, and AB will be the "best-fitting" straight line, determined by an application of the Method of Least Squares. See next chapter and also Fig. 16, Appendix II, p. 140.

"most probable" value of y for any particular value of x. If the line is straight or approximately straight the "regression" is said to be linear, and the equation to the line is

$$y - \bar{y} = r\frac{\sigma_2}{\sigma_1}(x - \bar{x}),$$

where $\bar{x}, \bar{y}$ are the mean values of all the x's and y's respectively (*not* the means of the arrays just mentioned),

σ_1, σ_2 are the standard deviations of the x's and y's respectively and r is the coefficient of correlation.

Thus, if

20 is the mean value of the mechanical memory of the group
30 " " " " " " general intelligence " " "
4 " " standard deviation for mechanical memory (σ_1)
7 " " " " " general intelligence (σ_2)

and, finally, ·6 is the value of r; then the most probable measure of the general intelligence of a child whose mechanical memory is represented by, say, the value 14, is given by the equation

$$y - 30 = \cdot 6 \times \tfrac{7}{4}(14 - 20),$$

whence $$y = \underline{23 \cdot 7}.$$

This value is the average of an array of possible values, whose standard deviation

$$= \sigma_2\sqrt{1 - r^2}$$
$$= \underline{5 \cdot 6}.$$

It will be proved in the next chapter that

$$r = \frac{S(xy)}{N\sigma_1\sigma_2}*,$$

where x and y are *deviations from the mean* (not absolute values as assumed above), $S(\ )$ indicates summation, i.e.

$$S(xy) = x_1y_1 + x_2y_2 + \ldots + x_Ny_N,$$

and N = the total number of cases (children measured).

* First suggested by Bravais; shown to be the best measure by Karl Pearson, who gave it the name of the "product-moment formula."

A. Bravais: "Analyse mathématique sur les probabilités des erreurs de situation d'un point," *Acad. des Sciences; Mémoires présentés par divers savants*, II^e Série, IX. 1846, p. 255.

Karl Pearson: "Regression, Heredity, and Panmixia," *Phil. Trans. Roy. Soc.*, Series A, Vol. CLXXXVII. 1896, pp. 253 ff.

It is important to note that by starting from y instead of from x, and determining the means of the y-arrays (such as the array within the limits $y_2y'_2$, $y_3y'_3$), another regression line, CD, is obtained *different* from the first. Its equation is

$$x - \bar{x} = r\frac{\sigma_1}{\sigma_2}(y - \bar{y}),$$

and it represents the law of change of mean x-value with increase of y. It gives the "most probable" value of x for any particular value of y.

If the series of means do not lie on a straight line (approx.) but on a curve of greater or less complexity, the above calculation is meaningless. In such a case, called a case of skew correlation and non-linear regression, the only measure of the correlation of the two variates is that given by η, the correlation ratio. η is the ratio of the standard deviation of the means of the arrays (Σ) to the total standard deviation (of either the x's or the y's). Thus there are two values of η, one for the x's, and another for the y's. They approximate closely to one another, as a rule, so that only one need be calculated.

$$\eta = \frac{\Sigma_1}{\sigma_1}, \text{ or } \frac{\Sigma_2}{\sigma_2}.$$

When the regression is linear, $\eta = r$; otherwise $\eta > r$. r ranges between the values ± 1, η between 0 and 1. η is always positive.

It will now have become clear that the correlation ratio, η (always) and the correlation coefficient, r (when regression is linear) are measures of the tendency towards concomitant variation exhibited by two series of phenomena, and hence throw some light upon the causal relations of these phenomena. Exactly what kind of causal relation we are justified in inferring from them will become clearer in the course of the next few chapters.

We may illustrate the significance of the idea of correlation in a slightly different (and more elementary) way. Let us suppose that the 200 children have been arranged in order of merit, as regards mechanical memory, on the one hand, and as regards general intelligence on the other. If now it were found that each child's order was the same in both, i.e. that the child first in mechanical memory was first in general intelligence, the child second in mechanical memory was second in general intelligence, and so on, the correspondence between the two series would be complete and r would be $+1$. Or if, on a second supposition, the child first in the one was last in the other, second in the one was next to last in the other, and so on, the correspondence between the two series would again be complete, but inverse, and r would be -1. Finally, if there is no correspondence whatever between the two series, r will be 0. A value of r between 0 and $+1$ will express a tendency, greater or less according to r's size, for children above the average or mean position in the one ability to be above the mean position in the other, and for children below the mean position in the one to be below the mean position in the other. A value of r between 0 and -1 will express a tendency, greater or less according as r is numerically greater or less, for the children above the mean position in the one ability to be below the mean position in the other, and conversely. Now if order or "rank" be taken as an (inverse) measure of ability, the value of

$$\frac{S(xy)}{N\sigma_1\sigma_2} \text{ or } r$$

becomes

$$1 - \frac{6S(d^2)}{N(N^2-1)},$$

where $S(\ )$ is the symbol for summation, and d is the difference between the rank of an individual in the one series and his rank in the other. This form gives us a general impression of its appropriateness for the purpose in view, since the greater the disparity between the two series of ranks the greater is $S(d^2)$ and hence the smaller is r. If there is no relation at all between the two series, $S(d^2)$ acquires the value it would have according to pure chance, and this can be shown to be $= \frac{1}{6}N(N^2-1)$, which makes the whole expression 0, as it should do.

The one objection to the formula is that it assumes the difference between any two neighbouring ranks to be equal at all parts of the scale. This is obviously a false assumption; the distance of individual from individual at the two extreme ends of the scale must be considerably greater than that between individuals near the middle. A correction for this, based on the assumption that the form of distribution of the abilities in each of the cases is Gaussian, has been calculated by Prof. K. Pearson. It is

$$r = 2 \sin\left(\frac{\pi}{6}\rho\right),$$

where
$$\rho = 1 - \frac{6S(d^2)}{N(N^2-1)}.$$

At the end of this chapter is given a table whereby ρ-values may be at once converted into corresponding r-values, according to Pearson's equation.

Finally, there is the question of the "probable error" (P.E.). Like every other constant calculated from a limited sample of variable material, the coefficient of correlation varies in value from sample to sample, and a measure is needed of the limits within which it may be expected with a fair degree of probability to lie. This measure is given by the "probable error." In the case of r determined by the product-moment formula, when N is sufficiently large,

$$\text{P.E.} = \frac{\cdot 67449}{\sqrt{N}}(1 - r^2),$$

which means that it is an even chance that the true value of r lies between the limits $r \pm \dfrac{\cdot 67449\,(1 - r^2)}{\sqrt{N}}$.

The chances are 16 to 1 against the value falling outside the limits

$$r \pm 3\ \text{P.E.}$$

For r determined by the rank-formula, the probable error is slightly larger, being $\dfrac{\cdot 7063\,(1 - r^2)}{\sqrt{N}}$.

If N, the number of cases, be small (say, less than 30), the probable error is larger. Its exact size under such conditions is not known.

The following is an example of the way in which a correlation coefficient may be obtained by means of ranks. The subjects were boys in the Fourth Form of a Public School, and the correlation to be obtained is that between ability in Classics and ability in Drawing.

	Form Order		
	Classics	Drawing	d^2
R. C. O.	1	9	$(1 \sim 9)^2 =$ 64
H. G. M.	2	2	0
B. L.	9	16	49
F. L. S.	7	6	1
C. M. S.	3	15	144
C. J. L. H.	5	4	1
A. L. P.	6	17	121
E. G. T.	4	3	1
F. C. F.	8	5	9
N. P. R. N.	11	14	9
H. B. D.	10	12	4
S. H. T.	14	7	49
H. B. M.	12	1	121
L. H. S.	13	8	25
J. P. C.	15	10	25
E. W.	16	18	4
C. C. M.	17	11	36
L. H. W.	18	13	25
E. M. J.	19	19	0
$N=19$			$688 = S(d^2)$

$$\rho = 1 - \frac{6S(d^2)}{N(N^2-1)} = 1 - \frac{6 \times 688}{19 \times 360} = \cdot 40,$$

$$\therefore \ r = 2 \sin \left(\frac{\pi}{6} \rho \right) = \cdot 416,$$

$$\text{P.E.} = \frac{\cdot 7063 (1 - r^2)}{\sqrt{N}} = \cdot 134.$$

r is here just over three times its probable error, and we might therefore feel inclined to conclude that it proves a real correlation between the two series. We must remember, however, that 19 is a very small number of cases, and that therefore the real probable error is considerably larger than that given by the formula. Hence the reality of the correlation is not so certain. Our caution is proved to be justified when we turn to the next higher form, the Remove, and find that, with the same number of boys, the correlation between Classical ability and Drawing ability works out as $-\cdot 313\ (\pm \cdot 14)$, quite a different result. It might be objected that other factors than mere smallness in the number of cases were responsible for the difference; e.g. that the tendency to specialise in Classics was greater in the Remove than in the Fourth, and that the consequent neglect of Drawing by the abler boys lowered the correlation. To this it may be replied, firstly, that the drawing-master was the same for both forms, and was likely to get as much out of the boys as possible in each case, and, secondly, that the difference between the two forms in respect of the degree of specialising tendency was insufficient to account for the disparity of the results.

The correct way to compare the results mathematically is to determine the *probable error of their difference.* This = the square root of the sum of the squares of the probable errors of each, i.e.

$$\text{P.E.}_{a-b} = \sqrt{\text{P.E.}_a{}^2 + \text{P.E.}_b{}^2},$$

which, in this case, $= \sqrt{\cdot 13^2 + \cdot 14^2}$

$= \cdot 19.$

The difference $= \cdot 416 + \cdot 313 = \cdot 73$,

its probable error $= \cdot 19$,

nearly four times the size of the difference.

A very important extension of the theory of correlation is the conception of "partial" correlation. If, e.g., three mental abilities are correlated with one another, it is of interest to know how closely any two of them are correlated with one another *for a constant value of the third.* Such a coefficient is written, in Yule's notation, $r_{12.3}$, and measures the closeness of interrelation of the first two abilities independent of the relation between them due to their respective relations to the third ability.

This may be illustrated from our example by taking the form-order for English into consideration in addition to that for Classics and for Drawing. The correlation between Classics and English works out as ·78, that between Drawing and English, as ·21.

Then the correlation between Classics and Drawing for "English constant"

$$r_{CD.E} = \frac{r_{CD} - r_{CE}r_{DE}}{\sqrt{(1 - r^2_{CE})(1 - r^2_{DE})}}$$

$$= \frac{\cdot 42 - \cdot 78 \times \cdot 21}{\sqrt{(1 - \cdot 78^2)(1 - \cdot 21^2)}}$$

$$= \underline{\cdot 42}.$$

Thus, in this particular case, the "partial" coefficient is practically identical with the "entire" coefficient, showing that the relation between Classics and Drawing is practically independent of their common relation to English. Calculating the other two partial coefficients, we get

$$r_{CE.D} = \underline{\cdot 78}, \quad r_{ED.C} = \underline{\cdot 58}.$$

The only coefficient showing an appreciable difference between entire and partial correlation is that between English and Drawing: entire = ·21, partial = ·58. This result means that that part of ability in English which is not identical with ability in Classics is related more closely to that part of ability in Drawing which is not identical with ability in Classics than is entire English ability to entire Drawing ability.

The result might be illustrated by the intersection of three spheres representing Classical, Drawing, and English ability respectively, after the analogy of Euler's circles in Logic, the volume common to any two of the spheres being proportional to the "entire" correlation coefficient for the two abilities, and the volume common to the same two *exclusive* of any part also common to the third sphere being proportional to the "partial" correlation coefficient for the two abilities. This cannot, of course, be taken as any *more* than an illustration, since the exact significance of the correlation coefficient, in this sense, is not known (see below p. 80).

The principle of partial correlation can be extended to include an indefinite number of variables, and general formulae for this purpose will be given in the next chapter.

It will be obvious that the same formulae may be employed for the elimination of "irrelevant" factors, e.g. age, sex, etc.

Table for converting ρ *into* r $(r = 2 \sin \frac{\pi}{6} \rho)$*.

ρ	r	ρ	r	ρ	r	ρ	r
·05	·052	·30	·313	·55	·568	·80	·813
·10	·105	·35	·364	·60	·618	·85	·861
·15	·157	·40	·416	·65	·668	·90	·908
·20	·209	·45	·467	·70	·717	·95	·954
·25	·261	·50	·518	·75	·765	1·00	1·000

* Quoted from K. Pearson, *Drapers' Company Research Memoirs*, Biometric Series IV. 1907, p. 18.

CHAPTER I*

THE MATHEMATICAL THEORY OF CORRELATION

In the present chapter an attempt will be made to summarize briefly the principal methods in use for obtaining a measure of the correlation, or tendency towards concomitant variation, of two or more variates.

Let the coordinates of the dots in the accompanying diagram —commonly known as a "scatter diagram"—represent the measures of two separate characteristics, e.g. speed of adding figures (x) and accuracy of adding figures (y), in a number of individuals (N).

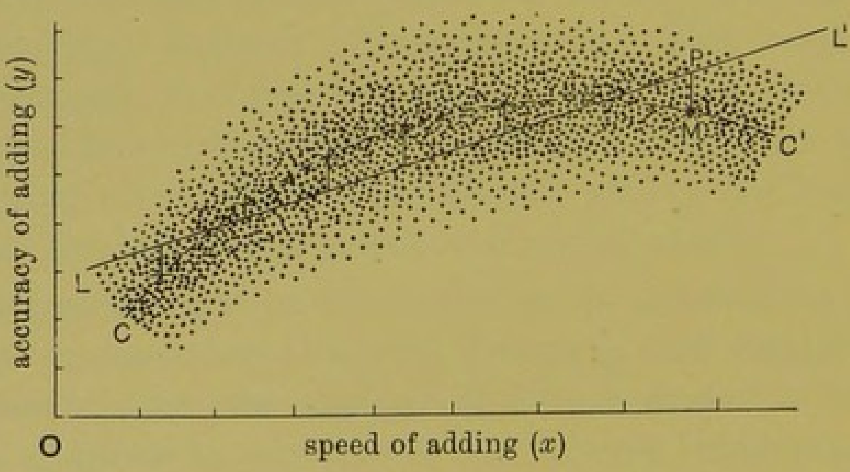


Fig. 11.

Let the crosses represent the mean values of y corresponding to values of x lying between the limits of pairs of successive units of measurement. Then the broken curve CC' passing through

* The proofs of the formulae for r, η, and the "two-row table" to be found in this chapter are abbreviations of those given by Professor Karl Pearson in his lectures.

these crosses represents the most probable law of relationship between speed of adding and accuracy of adding, and is known as the *regression curve*.

I. **Correlation coefficient** (r).

Let us now find the "best fitting" straight line, LL', to this curve. To do so we apply the method of least squares, merely from motives of convenience; i.e. we choose a line such that the sum of the squares of all distances like PM, each weighted with the number of cases (n_x) from which the mean value of the corresponding array of y's was determined, is a minimum.

Let ordinate of P be Y and ordinate of M be $\bar{y}_x$ (the mean of an array).

Then $S\{n_x(Y-\bar{y}_x)^2\} = V$ (say), which is to be a minimum.

Let equation to line LL' be $Y = Ax + B$.

Then $$V = S\{n_x(Ax + B - \bar{y}_x)^2\} \quad \ldots\ldots\ldots\ldots\ldots(\alpha).$$

For V a minimum, this gives*

$$A = \frac{\frac{S(xy)}{N} - \bar{x}\bar{y}}{\sigma_1^2} = \frac{S(x-\bar{x})(y-\bar{y})}{N\sigma_1^2},$$

and also $$\bar{y} = A\bar{x} + B$$

($\bar{x}$, $\bar{y}$ being the mean values of the x's and y's respectively, and σ_1, σ_2 their standard deviations).

* $V = S\{n_x(Ax+B-\bar{y}_x)^2\} = A^2 . S(n_x x^2) + B^2 . S(n_x) + S(n_x \bar{y}_x^2) + 2AB . S(n_x x) - 2A . S(n_x x\bar{y}_x) - 2B . S(n_x \bar{y}_x)$,

$$\therefore \frac{V}{N} = A^2(\sigma_1^2 + \bar{x}^2) + B^2 + \frac{S(n_x\bar{y}_x^2)}{N} + 2AB\bar{x} - 2A . \frac{S(n_x x\bar{y}_x)}{N} - 2B\bar{y}.$$

For this to be a minimum, $\frac{d}{dA}\left(\frac{V}{N}\right)$ and $\frac{d}{dB}\left(\frac{V}{N}\right)$ must both be zero, and $\frac{d^2}{dA}\left(\frac{V}{N}\right)$, $\frac{d^2}{dB}\left(\frac{V}{N}\right)$ must both be positive.

$$\therefore \begin{cases} A(\sigma_1^2+\bar{x}^2) + B\bar{x} - \frac{S(n_x x\bar{y}_x)}{N} = 0 & \ldots\ldots\ldots\ldots\ldots(1), \\ \text{and} \quad B + A\bar{x} - \bar{y} = 0 & \ldots\ldots\ldots\ldots\ldots(2). \end{cases}$$

Solving these equations for A and B, we have

$$A = \frac{\frac{S(n_x x\bar{y}_x)}{N} - \bar{x}\bar{y}}{\sigma_1^2} = \frac{\frac{S(xy)}{N} - \bar{x}\bar{y}}{\sigma_1^2} = \frac{S(x-\bar{x})(y-\bar{y})}{\sigma_1^2},$$

and $$B = \bar{y} - A\bar{x}.$$

Therefore the line passes through the means of the two values, and its equation is

$$y - \bar{y} = A\,(x - \bar{x}).$$

Let us define r as $\quad = \dfrac{S\,(x - \bar{x})\,(y - \bar{y})}{N\sigma_1\sigma_2}.$

Then $\quad A = \dfrac{S\,(x - \bar{x})\,(y - \bar{y})}{N\sigma_1^2} = r\,\dfrac{\sigma_2}{\sigma_1},$

and the equation to the best-fitting straight line, LL', is

$$y - \bar{y} = r\,\frac{\sigma_2}{\sigma_1}(x - \bar{x}) \quad \ldots\ldots\ldots\ldots\ldots\ldots(\beta).$$

LL' is known as the *regression line*, and $r\dfrac{\sigma_2}{\sigma_1}$ is called the *coefficient of regression* of y on x, being the tangent of the angle which this line makes with the axis of x.

Let $y - Y$ measure the distance of any individual point from this line. Then the average of the sum of the squares of all such distances

$$= \frac{S\,(y - Y)^2}{N}$$

$$= \frac{S\{y - \bar{y} - A\,(x - \bar{x})\}^2 *}{N}$$

$$= \sigma_2^2 + A^2\sigma_1^2 - 2Ar\sigma_1\sigma_2$$

$$= \sigma_2^2\,(1 - r^2).$$

Hence the standard error (standard deviation) made in estimating the value of y most probably associated with any particular value of x from equation (β) is

$$\sigma_2\sqrt{1 - r^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(\gamma).$$

r is known as the coefficient of correlation, and evidently must lie between the values $+1$ and -1. If the regression line coincides with the regression curve, within the limits of errors of random sampling,—in other words, if the regression is linear —r is a measure of the degree of dependence between x and y. When $r = \pm 1$, the points close up upon the line and the "scatter diagram" contracts to become the line itself.

* Y is the y of equation (β), and therefore $= \bar{y} + A\,(x - \bar{x})$.

If x and y be measured from their mean values, instead of from zero or some other arbitrary origin, we get

$$r = \frac{S(xy)}{N\sigma_1\sigma_2},$$

and the regression of y on x is given by the equation

$$y = r\frac{\sigma_2}{\sigma_1}x \qquad (\delta).$$

An analogous equation, $x = r\frac{\sigma_1}{\sigma_2}y$, gives the regression of x on y. There are thus *two* regression lines. The coefficients of regression, $r\frac{\sigma_1}{\sigma_2}$ and $r\frac{\sigma_2}{\sigma_1}$, are conveniently denoted by the symbols b_{12} and b_{21} respectively.

If x and y be measured from their means and also in terms of their respective standard deviations as unity, the regression equations become

$$x = ry \text{ and } y = rx,$$

and r is then itself the coefficient of regression of y on x and of x on y, the two regressions being equal.

II. **Correlation Ratio** (η)*.

It is clear that if the regression is not linear r ceases to be a satisfactory measure of the relation between the two characters under consideration. In an extreme case, such as that shown

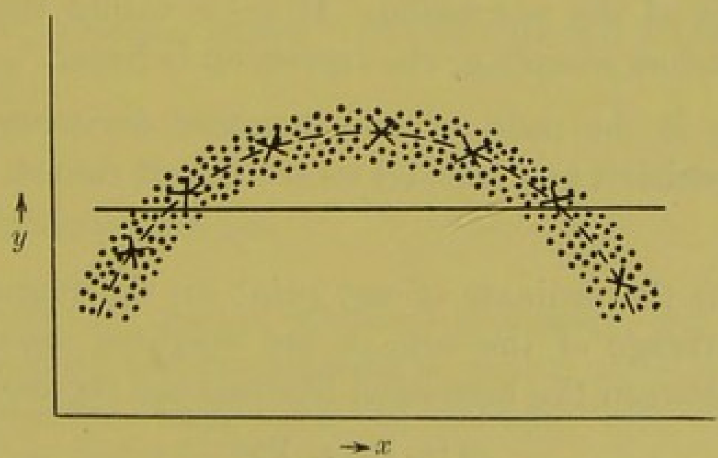


Fig. 12.

* See *Drapers' Company Research Memoirs*, Biometric Series II. pp. 9 et seq. Karl Pearson "On the Theory of Skew-Correlation and Non-Linear Regression."

in the accompanying diagram, r may be zero while there is yet a very close relation between the two characters.

Let $y-\bar{y}_x$ be the distance of any point of the "scatter diagram" from the regression *curve.* Then the average of the sum of the squares of all such distances, each x-array of y's being weighted with the number of individuals (n_x) in it,

$$=\frac{S\{n_x(y-\bar{y}_x)^2\}}{N},$$

and this expression, on reduction,

$$=\sigma_2^2-\Sigma^2,$$

where $\Sigma=$ the standard deviation of the means of the arrays, each array being weighted with the number in it;

i.e. $$=\sqrt{\frac{S(n_x\bar{y}_x^2)}{N}-\bar{y}^2}=\sqrt{\frac{S\{n_x(\bar{y}_x-\bar{y})^2\}}{N}}.$$

Let $$\eta=\frac{\Sigma}{\sigma_2}.$$

Then $$\frac{S\{n_x(y-\bar{y}_x)^2\}}{N}=\sigma_2^2(1-\eta^2),$$

and the mean square of the distance of points from the regression curve

$$=\sigma_2\sqrt{1-\eta^2} \quad \text{.......................}(\epsilon).$$

Comparing expressions (ϵ) and (γ), we see that η is a test of the linearity of the regression. If $\eta=r$, within the limits of errors of random sampling, the regression is linear.

Since η is the ratio of two standard deviations, it must always be positive, and from (ϵ) we see that it cannot be greater than 1.

Let Y be the ordinate of any point on the regression *line.* Then the average of the sum of the weighted squares of the distances between the regression line and the regression curve

$$=\frac{S\{n_x(\bar{y}_x-Y)^2\}}{N},$$

which reduces to

$$\sigma_2^2(\eta^2-r^2) \quad \text{.......................}(\zeta).$$

Thus η must always be numerically greater than r, except in the case of linear regression, when it is numerically equal to r.

In examining the relationship between two measurable characters, η should be calculated as well as r, since it serves as a test of the linearity or non-linearity of the regression, and is also a better measure of causal relation than r.

A simple criterion for linearity which is very generally applicable is that

$$\frac{\sqrt{N}}{\cdot 67449} \cdot \tfrac{1}{2}\sqrt{\eta^2 - r^2} < 2\cdot 5\,^{*}.$$

For very exact work, more complicated formulae need to be employed.

The results obtained above are all *independent of the forms of distribution* of the variates.

In Appendix II. will be found an example of the evaluation of r and of η between speed of adding single digits and accuracy in doing so, the individuals measured being 86 boys between the ages of 11 and 12 years from two L.C.C. elementary schools. The two groups could be thrown together for this purpose, since the means and standard deviations calculated from them separately were in very close agreement—well within the limits of the probable errors. The results will be discussed in Chapter III.

III. Probable Errors.

In determining means, standard deviations, and other frequency constants, the investigator is unable to work from the "total population" and must be content with the results obtained from "random samples" of greater or less size taken from this (in some cases, hypothetical) total population. Evidently

* J. Blakeman: *Biometrika*, Vol. IV. pp. 349, 350.

the results obtained from different random samples will not as a rule coincide, and some measure is needed of the variability of these results about their average, which latter quantity, for a large number of samples, may be regarded as coinciding approximately with the value of the frequency constant for the total population. Such a measure is given by the standard deviation, σ, of the various results obtained from the different

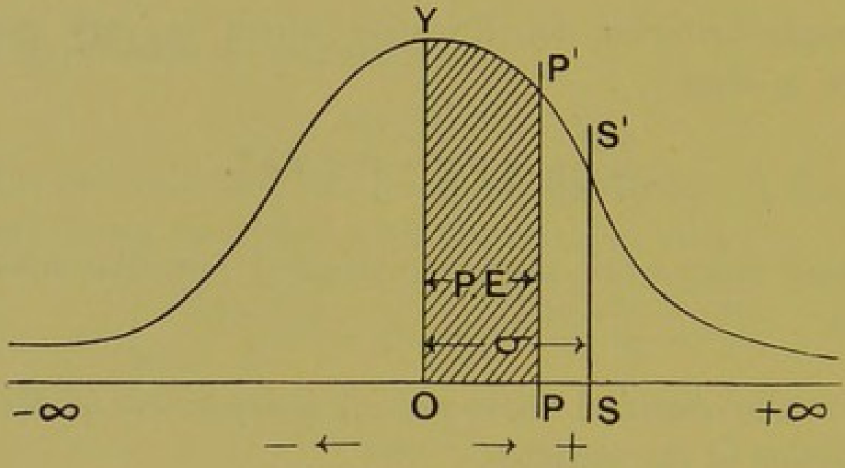


Fig. 13.

samples, but, since the theory of probable errors was originally developed on the assumption of the "normal" or Gaussian form of distribution, the value universally taken is $\cdot 67449 \times \sigma$, which is conventionally termed *the* probable error (P.E.). In the normal curve (see Fig. 13) it is the distance OP, the position of P being such that the ordinate PP' divides the positive half of the curve into two equal areas. Thus, if the total area of the curve is taken as unity, the value of P.E. is given by the equation

$$\int_0^{\text{P.E.}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}} dx = 0\cdot 25,$$

whence $\text{P.E.} = \cdot 67449\sigma$.

In the language of chance, a deviation of $\pm$ P.E. is as likely to occur as not, when the distribution of frequencies is "normal." The value of a frequency constant determined from a random sample may be considered fairly reliable within the limits of $\pm$ (2—3 times its P.E.).

When the number of cases (n) in the random sample is fairly large—so large that fractions containing certain higher powers of n in the denominator can be neglected—the probable errors are found to be as follows*:

$$\text{P.E. of a mean} = \cdot 67449 \frac{\sigma}{\sqrt{n}},$$

$$\text{,,} \quad \text{,,} \qquad \sigma = \cdot 67449 \frac{\sigma}{\sqrt{2n}},$$

$$\text{,,} \quad \text{,,} \qquad r = \cdot 67449 \frac{1 - r^2}{\sqrt{n}}.$$

The second and third of these values are only correct when the frequency distribution is normal or approximately normal. In particular, for large values of r the true P.E. may be considerably different from that given by the above formula unless the distribution is normal.

P.E. of $\eta = \cdot 67449 \dfrac{1 - \eta^2}{\sqrt{n}}$, for linear regression, and also, as a rough measure, for cases of skew-correlation. If greater exactitude is needed in the latter cases, more complicated formulae have to be employed†.

Another frequency constant in common use is the *coefficient of variation* V, which $= \dfrac{100\sigma}{\text{mean}}$.

$$\text{Its P.E.} \quad = \cdot 67449 V \left\{1 + 2\left(\frac{V}{100}\right)^2\right\}^{\frac{1}{2}} \Big/ \sqrt{2n}\ \ddagger.$$

As stated above, the values just given for the probable errors only apply in cases where n is fairly large. In cases where n is so small that certain higher powers of its reciprocal cannot be neglected in comparison with the rest of the expressions involving them, the values cannot be used. For such cases no theoretical formulae have hitherto been devised.

* See W. Gibson: "Tables for Facilitating the Computation of Probable Errors," *Biometrika*, Vol. IV. pp. 385 et seq.

† Karl Pearson: *op. cit.* (Biometric Series II.) p. 19.

‡ Calculated values for different values of n given in Gibson's Tables, just quoted.

Quite recently, however, an *empirical* investigation has been made on samples of 4, 8, and 30 cases, taken from a "total population" of 3000 pairs of measurements (height and left middle finger measurements of 3000 criminals; "real" correlation, ·66)*. In the case of correlation it was found that when the original pairs of measurements were arranged so that the "real" correlation was 0, the distribution of the values of r obtained from samples of n was given approximately by the formula

$$y = y_0 (1 - x^2)^{\frac{n-4}{2}}.$$

Five distributions in all were worked out, giving the following moment coefficients:

	Mean	S.D.	μ_2	μ_3	μ_4	β_1	β_2
Samples of 4 ($r=0$)	—	·5512	·3038	—	·1768	—	1·918
" 8 ($r=0$)	—	·3731	·1392	—	·0454	—	2·336
" 4 ($r=·66$)	·5609	·4680	·2190	− ·1570	·2152	2·245	4·489
" 8 ($r=·66$)	·6139	·2684	·07202	− ·02634	·02714	1·857	5·232
" 30 ($r=·66$)	·661	·1001	·01003	− ·000882	·000461	·7713	4·580

[Unit of grouping for 4's and 8's = ·04.]

Correlation results, for real value of r = ·66, were

samples of 4 ...... ·561 ± ·011,
" " 8 ...... ·614 ± ·065,
" " 30 ...... ·6609 ± ·0067.

Hence it may be concluded that, although in the case of such small samples as 4 or 8 the ordinary formula for the P.E. of r gives much too low a value, yet in the case of as many as 30, the formula applies with tolerable accuracy. We must, however, bear in mind that this result has only been proved (empirically) to hold in the single case where the actual correlation was ·66.

* "Student": "The Probable Error of a Coefficient of Correlation," *Biometrika*, Vol. VI. pp. 302–310. 1908–1909.

IV. Multiple Correlation*.

The regression equation for the two variates, x and y, both measured from their means, was found, on p. 57, to be of the form

$$x = b_{12} y,$$

the equation being obtained by an application of the method of least squares, on the assumption that the regression was linear.

It is evident that the method thus applied to the case of two variates might, without further assumptions, be used in the general case of n variates. Thus, if $x_1, x_2, \dots x_n$ denote deviations from means, the equation expressing the regression of x_1 on $x_2 \dots x_n$ could be written

$$x_1 = b_{12.34\dots n} x_2 + b_{13.24\dots n} x_3 + \dots + b_{1n.23\dots n-1} x_n \quad \dots(\alpha).$$

"In this notation the suffix of each regression coefficient completely defines it. The first subscript gives the dependent variable, the second the variable of which the given regression is the coefficient, and the subscripts after the period show the remaining independent variables which enter into the equation." (Yule, *op. cit.* p. 182.)

In terms of the same notation, write

$$r_{12.34\dots n} = (b_{12.34\dots n} \cdot b_{21.34\dots n})^{\frac{1}{2}} \dots\dots\dots\dots\dots (\beta),$$

$$x_{1.23\dots n} = x_1 - (b_{12.34\dots n} x_2 + b_{13.24\dots n} x_3 + \dots + b_{1n.23\dots n-1} x_n) \dots (\gamma),$$

$$N \cdot \sigma^2_{1.23\dots n} = S(x^2_{1.23\dots n}) \dots\dots\dots\dots\dots\dots (\delta).$$

Then the normal equations, from which the regressions are determined by the method of least squares, may be written

$$S(x_2 \cdot x_{1.23\dots n}) = S(x_3 \cdot x_{1.23\dots n}) = \dots = S(x_n \cdot x_{1.23\dots n}) = 0 \dots (\epsilon).$$

From these equations Mr Yule has deduced the following results:

$$b_{12.34\dots n} = r_{12.34\dots n} \frac{\sigma_{1.34\dots n}}{\sigma_{2.34\dots n}} \dots\dots\dots\dots\dots\dots (\zeta),$$

$$\sigma^2_{1.23\dots n} = \sigma_1^2 (1 - r^2_{12})(1 - r^2_{13.2})(1 - r^2_{14.23}) \dots (1 - r^2_{1n.23\dots n-1}). \ (\eta),$$

$$r_{12.34\dots n} = \frac{r_{12.34\dots n-1} - r_{1n.34\dots n-1} r_{2n.34\dots n-1}}{(1 - r^2_{1n.34\dots n-1})^{\frac{1}{2}} (1 - r^2_{2n.34\dots n-1})^{\frac{1}{2}}} \dots\dots (\theta).$$

* See G. Udny Yule: "On the Theory of Correlation for any Number of Variables, treated by a New System of Notation," *Proc. Roy. Soc.* Vol. 79 A, pp. 182–193, 1907, of which the present section is merely a much condensed account.

$r_{12.34...n}$ is known as a "partial" correlation coefficient, being the value of the correlation between 1 and 2 for constant values of 3, 4 ... n. Similarly, $b_{12.34...n}$ is a "partial" regression coefficient (cf. "partial differentiation" in the Differential Calculus). Knowing the "total" correlations r_{12}, r_{13}, r_{23}, etc., equation (θ) enables us to obtain the various partial coefficients by successive substitutions. Thus, in the case of 3 variables, 1, 2, 3, we have

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{1 - r^2_{13}}\,\sqrt{1 - r^2_{23}}},$$

and two similar equations, expressing the value of the correlation between two of the variables for a constant value of the third*. If a fourth variable be added, we have the further set of equations

$$r_{12.34} = \frac{r_{12.3} - r_{14.3}\, r_{24.3}}{(1 - r^2_{14.3})^{\frac{1}{2}}\,(1 - r^2_{24.3})^{\frac{1}{2}}}.$$

Perhaps a more convenient formula for obtaining the partial correlation of two variables for constant values of a third and fourth is

$$r_{12.34} = \frac{r_{12}\,(1 - r^2_{34}) - r_{13}\,(r_{23} - r_{24}\, r_{34}) - r_{14}\,(r_{24} - r_{23}\, r_{34})}{\sqrt{1 - r^2_{13} - r^2_{14} - r^2_{34} + 2r_{13} r_{14} r_{34}}\,\sqrt{1 - r^2_{23} - r^2_{34} - r^2_{24} + 2r_{23}\, r_{34}\, r_{24}}}.$$

When the partial correlation coefficients have been determined, the regressions can be found by substituting in equations (η) and (ζ), and these give at once the regression equations. As explained before, a regression equation gives the most probable value of one variable for given values of the remaining variables, the standard error in such a prediction being $\sigma_{1.23...n}$ etc.

An example of the method of applying the above formulae is given in Appendix III., where the partial correlations, regressions, etc. in the case of four interrelated psychical capacities are worked out on lines identical with those illustrated by Mr Yule in his paper.

* For an interesting representation of correlation between three variables by a model showing the distribution of points in space, see G. Udny Yule, "An Introduction to the Theory of Statistics," C. Griffin and Co., London, 1911, pp. 241–243.

The probable error of a partial correlation coefficient $r_{12.34...n}$

$$= \cdot 67449 \frac{1 - r^2_{12.34...n}}{\sqrt{n}},$$

i.e. is similar in form to that for a "total" coefficient.

V. **Other methods for determining Correlation.**

1. *Four-fold Table.*

	x_1	x_2	
y_1	a	b	$a+b$
y_2	c	d	$c+d$
	$a+c$	$b+d$	N

(*a*) When the divisions pass through the means of both characters

$$r = \sin \frac{\pi}{2}\left(\frac{a-b}{a+b}\right).$$

This formula (Sheppard's) is of little practical use, since the mean values, in cases where the four-fold table is the only method which can be used, are generally unknown.

(*b*) In cases where the dividing lines do not pass through the means (means not known), the equations for determining the correlation are:

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

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

* All this is rendered much easier by P. F. Everitt's Tables of Tetrachoric Functions just published in *Biometrika* (Vol. VII. pp. 437–451).

from which h and k can be obtained by Sheppard's Tables; and finally

$$\frac{ad-bc}{N^2HK} = r + \frac{r^2}{2}hk + \frac{r^3}{6}(h^2-1)(k^2-1) + \frac{r^4}{4}h(h^2-3)k(k^2-3)$$
$$+ \frac{r^5}{120}(h^4-6h^2+3)(k^4-6k^2+3)$$
$$+ \frac{r^6}{720}h(h^4-10h^2+15)k(k^4-10k^2+15)$$
$$+ \frac{r^7}{5040}(h^6-15h^4+45h^2-15)(k^6-15k^4+45k^2-15)$$
$$+ \text{etc.},$$

where $H = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}h^2}$, and $K = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}k^2}$.

In obtaining these equations, a *normal* correlation surface is assumed; equation:

$$z = \frac{N}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}e^{-\frac{1}{2}\frac{1}{1-r^2}\left(\frac{x^2}{\sigma_1^2}-\frac{2rxy}{\sigma_1\sigma_2}+\frac{y^2}{\sigma_2^2}\right)}.$$

The probable error of r obtained by the four-fold table method is much larger than that given by the formula of p. 61. The correct formula is too complicated to insert here.

2. *Method of Contingency**.

The following is an example of a contingency table†.

Fathers.

Sons.		Merry	Melancholy	Alternating	Even	Totals
	Merry ...	122	8	81	67	278
	Melancholy	10	2	7	10	29
	Alternating	70	9	101	68	248
	Even ...	58	6	66	45	175
	Totals ...	260	25	255	190	730

* Karl Pearson: "On the Theory of Contingency and its Relation to Association and Normal Correlation," *Drapers' Company Research Memoirs*, Biometric Series I. 1904; Dulan and Co., London.

† Taken from a paper by E. Schuster and E. M. Elderton on "The Inheritance of Psychical Characters (being a further Statistical Treatment of Material Collected and Analysed by Messrs G. Heymans and E. Wiersma)," *Biometrika*, Vol. v. 1906–1907, pp. 460–469.

The method is employed when the grouping is merely by class, and the different classes have no quantitative relation to one another,—in other words, when the grouping is merely qualitative. The relationship between the two variables is measured by the differences between the numbers actually found in the various compartments of the table and the numbers that might be expected there by pure chance.

To state the rule:

The total mean square contingency, ϕ^2, of the table is given by

$$\phi^2 = \frac{1}{N} S_{pq} \left\{ \frac{\left(n_{pq} - \frac{n_p n_q}{N}\right)^2}{\frac{n_p n_q}{N}} \right\},$$

where

n_p = total frequency in pth row,

n_q = " " " qth column,

n_{pq} = frequency of constituent common to pth row and qth column,

N = total frequency of table.

Then the coefficient of mean-square contingency,

$$c = \sqrt{\frac{\phi^2}{1 + \phi^2}} *.$$

In the above case of fathers and sons

$$c = 0{\cdot}16.$$

The probable error of c is very complicated†.

For $c = 0$, $$\text{P.E.} = \frac{{\cdot}67449}{\sqrt{N}}.$$

3. *Two-row Table.* (New method‡.)

This method, recently devised by Professor K. Pearson, gives a unique value of r in the case of two variates one of which is both quantitative and continuous (e.g. intelligence)

* There are certain corrections on ϕ^2, not mentioned here, which often make much difference to the result.

† J. Blakeman and Karl Pearson: "On the Probable Error of Mean-Square Contingency," *Biometrika*, Vol. v. pp. 191–197, 1906.

‡ Karl Pearson: *Biometrika*, Vol. VII. 1909, p. 97.

while the other, though quantitative, admits of only two subdivisions (e.g. into good and bad visualisers), or, in more technical language, is "alternative."

→ x (intelligence).

↓ y

Good visualisers									
Bad visualisers									
									N

The assumptions made are two in number:

(1) that the regression is *linear*:

(2) that the distribution of the alternative variate is approximately normal or Gaussian.

The accompanying diagram (Fig. 14) will make the method clearer. LL' is the regression line of x on y, its equation being

$$x - \bar{x} = r\frac{\sigma_1}{\sigma_2}(y - \bar{y}),$$

$$\therefore \quad r = \frac{\dfrac{\bar{x} - x}{\sigma_1}}{\dfrac{\bar{y} - y}{\sigma_2}}.$$

Putting $\bar{x}$, the mean of all the x's $= \bar{p}$, $x = p'$, the mean value of the array of x's corresponding to the smaller group of y's (good visualisers), and $\bar{y} - y = q$, the distance of the centroid of the area RST from the mean MM of the Gaussian curve, we have the formula:

$$r = \frac{\dfrac{\bar{p} - p'}{\sigma_1}}{\dfrac{q}{\sigma_2}}.$$

The numerator of this is known, and the denominator

$$\frac{q}{\sigma_2} = \frac{y_0 \int_{x_0}^{\infty} e^{-\frac{1}{2}\frac{x^2}{\sigma_2^2}} x\,dx}{\sigma_2 \times \text{area of part } RST},$$

$$= \frac{N}{n_t} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\frac{x_0^2}{\sigma_2^2}},$$

where n_t = no. of cases in upper group (area RST),

and N = total no. of cases.

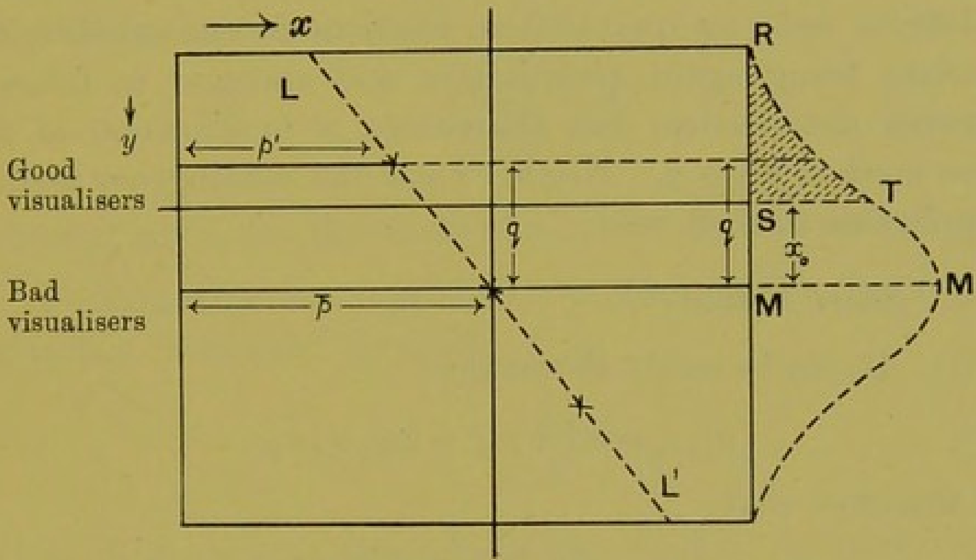


Fig. 14.

In this expression $\frac{n_t}{N} = \frac{1}{2}(1 - \alpha)$ in Sheppard's Tables (*Biometrika,* Vol. II. p. 182) from which the corresponding value of $\frac{x_0}{\sigma_2}$ can be read; and the value of z corresponding to this latter result gives

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_0}{\sigma_2}\right)^2};$$

$$\therefore\ r = \frac{\dfrac{\bar{p} - p'}{\sigma_1}}{\dfrac{N}{n_t} \times \left(\begin{array}{c}\text{the value of } z \text{ in Sh.'s Tables corresponding} \\ \text{to the } \frac{1}{2}(1-\alpha) \text{ of } \frac{n_t}{N}\end{array}\right)}.$$

The above method has not yet been applied in psychology, but might advantageously be used in determining the relationship of two characters, such as those suggested (intelligence and visual imagery), which are both quantitative, but of which one only admits of reliable division into two groups. It is the smaller of these groups (the "tail" of the distribution) which gives the n_t of the formula. In the suggested case, it might be represented by the number of individuals, out of the entire group (N) measured, who successfully pass certain somewhat difficult tests of visual imagery.

In cases where the x-variate (continuous and quantitative in our example above) can only be divided into classes, showing no definite order or quantitative relations to one another, the y-variate being again quantitative and assumed to follow a Gaussian distribution, but alternative, a modification of the above method gives η. Neither r nor the contingency method is applicable in such cases.

4. *Short Methods*.*

(i) It can be easily shown that

$$\sigma^2_{x-y} = \sigma_x^2 + \sigma_y^2 - 2r_{xy}\,\sigma_x\,\sigma_y,$$

and therefore that

$$r_{xy} = \frac{\sigma_x^2 + \sigma_y^2 - \sigma^2_{x-y}}{2\sigma_x\,\sigma_y}.$$

(ii) If the distributions are both normal, and if both variates have the same mean, m, and the same standard deviation, σ, then

$$r = 1 - \frac{\pi\,\{S\,(\overline{m+x} - \overline{m+y})\}^2}{N^2\,\sigma^2},$$

where $S(\overline{m+x} - \overline{m+y})$ is the sum of the *positive* differences only. In Professor Pearson's words, the rule is:

"Subtract from unity π times the square of the mean sum of the positive differences of paired variates divided by their common standard deviation."

This method might sometimes be conveniently used in determining the "individual" correlation (see below p. 86)

* Karl Pearson: "On Further Methods of Determining Correlation," *Drapers' Company Research Memoirs*, Biometric Series IV. 1907.

between performances of the same individuals in the same mental tests on different occasions.

(iii) *The Method of Ranks.*

It has recently been suggested by Dr C. Spearman* that measurements of psychical performance may conveniently—nay, preferably—be replaced by the numbers representing the rank or order of merit of the individuals in the group. On this basis the ordinary product-moment formula for r, $\frac{S(xy)}{N\sigma_1\sigma_2}$, can be easily shown to reduce to the form

$$1-\frac{6S(\nu_1-\nu_2)^2}{N(N^2-1)} \quad\ldots\ldots\ldots\ldots\ldots\ldots(\alpha),$$

where ν_1 and ν_2 are the ranks of an individual in the two series.

Dr Spearman has suggested a still simpler formula, which he calls a "foot-rule" formula. It is

$$R=1-\frac{S(g)}{\frac{1}{6}(N^2-1)}\ldots\ldots\ldots\ldots\ldots\ldots(\beta),$$

where $S(g)$ denotes the sum of the "gains" in rank (sum of positive differences) of the second series on the first, and $\frac{1}{6}(N^2-1)$ is the value of the sum of such gains which may be expected by chance.

In order to make the determination of correlation by R comparable with that obtained by the ordinary formula—presumably, in Dr Spearman's view, formula (α)—he has suggested the formula

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

This method of using ranks, and the formulae suggested therefor, were vigorously criticised by Karl Pearson in the paper quoted on the preceding page.

* C. Spearman: "Measurement of Association between Two Things," *Am. J. P.* Vol. xv. 1904; "'Foot-rule' for Measuring Correlation," *Brit. Journ. of Psychology*, Vol. II. Pt. I. July 1906.

Pearson's Criticisms.

(*a*) The *direct* use of ranks, or order in a series, as a quantitative measure of character, is inadmissible. It assumes that the unit of rank is equal throughout the scale. But this is very obviously not the case. "Between mediocrities the unit of rank treated as a measure of a variate is practically zero, between extreme individuals it is very large indeed." Some form of frequency distribution must be assumed, and the method of ranks assumes that form to be a *rectangle*. The assumption made could hardly be a more improbable one. On the assumption of "normal" or Gaussian distribution, Pearson shows that the correlation of grades (not quite the same thing as ranks) is

$$r = 2 \sin \left(\frac{\pi}{6} \rho\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (\delta),$$

where $\rho = 1 - \frac{6S(\nu_1 - \nu_2)^2}{N(N^2 - 1)}$, the expression ($\alpha$) above.

The *grade* of an individual in a group is measured by the number of individuals above him, represented by the shaded portion of the following normal curve, thus:

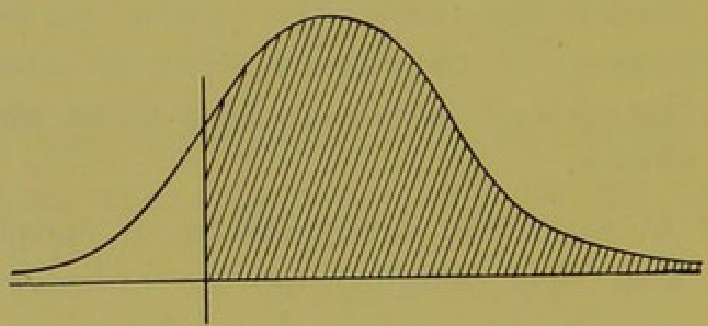


Fig. 15.

Hence the quantities to be correlated are

$$x\text{-grade, } g_1 = \tfrac{1}{2}N + \frac{N}{\sqrt{2\pi}\sigma_1}\int_0^x e^{-\frac{1}{2}\frac{x^2}{\sigma_1^2}} dx,$$

and

$$y\text{-grade, } g_2 = \tfrac{1}{2}N + \frac{N}{\sqrt{2\pi}\sigma_2}\int_0^y e^{-\frac{1}{2}\frac{y^2}{\sigma_2^2}} dy,$$

giving

$$r = 2 \sin \left\{\frac{\pi}{6}\left[1 - \frac{6S(\nu_1 - \nu_2)^2}{N(N^2 - 1)}\right]\right\} \ldots\ldots (\alpha \text{ and } \delta).$$

In terms of the sum of positive differences of ranks ("gains" in rank) the formula is

$$r = 2\cos 2\pi \left\{\frac{S(g)}{N^2-1}\right\} - 1$$

$$= 2\cos\frac{\pi}{3}(1-R) - 1 \quad \ldots\ldots\ldots\ldots\ldots\ldots(\epsilon).$$

This, then, is the formula which should take the place of Spearman's formula (γ).

(b) With small values of N, R retains the same value for wide variations in ρ [expression (α) above].

(c) Spearman obtains the value $\frac{\cdot 43}{\sqrt{n}}$ for the probable error of R for zero correlation, empirically, by noting the distribution of a large number of chance values of $S(g)$*. In the case of some of these, the corresponding values of R are negative. Now the negative values of R only range from 0 to $-\cdot 5$. Hence the value which he deduces for the probable error, being that of a quantity ranging from $+1$ to $-\cdot 5$, is not comparable with the probable error for the product-moment formula (range, $+1$ to -1). The comparative accuracy of the different methods may be deduced from Pearson's formulae by using differentials, and Pearson has proved that the probable error of ρ for zero correlation $= \frac{\cdot 67449}{\sqrt{n-1}}$. He is thus able to show that, for zero correlation,

$$\text{P.E. of } r \text{ deduced from } \rho = \frac{\cdot 7063}{\sqrt{n-1}},$$

and

$$\text{P.E. ,, } r \text{ ,, ,, } R = \frac{\cdot 7738}{\sqrt{n-1}}.$$

Thus the rank methods are less accurate than the product-moment method. "In particular it requires about 30 % more observations by the R method to obtain r with the same degree of certainty when $r = 0$."

(d) When R works out negative, Spearman recommends that one of the series be reversed. But since R is not a

* *Brit. Journ. of Psychology*, Vol. II. Pt. I. 1906, pp. 107, 108.

symmetrical function of ρ, this is not a sound course to adopt. In fact, if the numbers are small, reversal of one of the series may yet leave R negative. Pearson's own formula (ϵ) is free from this drawback*.

An important argument brought forward by Spearman† is one to the effect that absolute measurements are often *heterogeneous* and that ranking removes this heterogeneity. There is an assumption underlying this argument which would seem to be of somewhat doubtful validity. All, or almost all, so-called mental measurement is measurement of the object. It is physical, not mental at all. The physical objects, the physical results *can* be measured—in terms of space and time—and they can be correlated. Psychical measurement has perhaps some plausibility in the case of sensation intensities, in the form of the "sense-distances" of Delboeuf; but this furnishes small justification for the claims made above on behalf of ranks. The unit of rank means nothing until the form of distribution is fixed, and this can hardly be determined on a purely psychical basis; at least it is difficult at present to conceive of even an attempt at such a determination.

Another important passage occurs on p. 99 of the previously-quoted article. "In this simplified form the standard or r method [Spearman apparently means the formula $1 - \frac{6S(\nu_1 - \nu_2)^2}{N(N^2-1)}$, which is ρ, *not* the standard or r method; $r = 2\sin\left(\frac{\pi}{6}\rho\right)$] shows itself to be solely distinguished from our short or R method by the fact that the differences of rank *are squared.* 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 in psychology, these extreme cases are just the ones of most

* Spearman endeavours to meet some of these criticisms in a recent paper; *British Journal of Psychology*, Vol. III. Pt. 3, Oct. 1910.

† C. Spearman: "'Foot-rule' for Measuring Correlation," *British Journal of Psychology*, Vol. II. Pt. I. July 1906, p. 93.

suspicious validity [why ?] so that squaring is here more likely to do harm than good."

Such an argument is sound as applied to actual measurements, but as soon as those measurements have been translated into *ranks*, it loses its force. The extreme ranks are *more* reliable than the medium ones since they are farther apart from one another, and therefore the extreme differences are *more*—not less—reliable than the medium ones.

Finally, Spearman says that he deduced his formula

$$r = \sin\left(\frac{\pi}{2} R\right),$$

empirically from 111 correlations, the average number of cases in each being 21. Now, for $n = 21$ we have seen that the P.E. is large, even larger than that given by the usual formula, $\cdot 67449 \frac{1 - r^2}{\sqrt{n}}$. Hence, the chance that the selected formula was the best one can not have been great. Doubtless many others would have fitted equally well; and if Spearman wishes to support his formula as against Pearson's on the strength of the fact that it was deduced empirically from actual cases of psychical correlation and therefore was likely to be best suited to the forms of frequency distribution commonly followed by psychical measurements, such an argument would be of little cogency.

In all work where accuracy, and, above all, thoroughness are required, the method of ranks should be avoided since it gives neither the mean nor the standard deviation—quantities important in themselves and valuable as a means of obtaining the coefficient of variation of the group in the particular mental functions under consideration. Ranks hide variations of great psychological importance, and introduce a spurious homogeneity which can well be dispensed with. When the number of cases is large, ranking takes too long a time, and the grouping of the individuals in a correlation table gives much more information as to the nature of the correlation. Even when the numbers worked with are comparatively small (say 30—100), means, standard deviations, and coefficients of variation are results of considerable intrinsic value, and in finding the standard deviation the deviations from the means have to be written down, so that everything needful for the calculation of r is already determined, and the evaluation of correlation coefficients between the various series of measurements taken in pairs is an operation quickly performed. Another reason for drawing up a correlation table whenever possible—and this of course is impossible with ranks—is that only thereby can the correlation ratio, η, be calculated, a result which tests the linearity or non-linearity of the regression, and is moreover a better measure of causal relationship than r itself.

Only when n is small (< 30) are ranks to be recommended in preference to actual measurements (since means, S.D.'s, etc. are in these cases of little value), and then ρ is as easily calculated as R, and more reliable. The results with such small numbers could do little more than indicate the *existence* of correlation in certain cases. They could give practically no information as to its *size* owing to the large amount of probable error.

VI. Spurious Correlation.

Correlation is said to be "spurious" when it is due to extraneous conditions and does not arise directly out of the two functions under consideration. The term is one of relative rather than absolute significance, but its appropriateness will become apparent after a consideration of the following two examples:

1. Spurious correlation arising from *heterogeneity of material.*

Let us suppose that two distinct groups of children have been measured for two characters A and B, and that the mean abilities, in both A and B, are higher in the one group than in the other. Then, even if there is no correlation between the two characters, as estimated from each group, separately, a positive correlation will be obtained from the two groups taken together. On the other hand, if the mean ability in A is higher and the mean ability in B lower, in the one group than in the other, a negative correlation will be obtained by taking the two groups together. The correlation in each case will be due simply to the *heterogeneity* of the material employed. The difference in the mean values for the two groups must of course have some cause, such as a difference of nationality, sex, or even locality (within any one town or district) from which the children are drawn; or, again, such a cause as the difference of discipline to which the two groups have been subjected in the past. These are extraneous conditions and, if measurable, can be allowed for by employing the method of partial correlation. As a rule, however, they are not easy to determine quantitatively; hence their dangerous character.

2. *Index Correlation**.

This is a form of spurious correlation which arises from the use of ratios for measurements. Thus if

$$z_1 = \frac{x_1}{x_3},\ z_2 = \frac{x_2}{x_3},$$

and x_1, x_2, x_3 are uncorrelated with one another, it can be shown that the correlation between z_1 and z_2,

$$r_{z_1 z_2} = \frac{\left(\frac{\sigma_{x_3}}{\bar{x}_3}\right)^2}{\sqrt{\left(\frac{\sigma_{x_1}}{\bar{x}_1}\right)^2 + \left(\frac{\sigma_{x_3}}{\bar{x}_3}\right)^2}\sqrt{\left(\frac{\sigma_{x_2}}{\bar{x}_2}\right)^2 + \left(\frac{\sigma_{x_3}}{\bar{x}_3}\right)^2}}$$

a quantity which may be as large as ·5.

As an illustration from psychology, we may mention the correlation of errors of observation, either of different individuals or of the same individual at different times. Even if the absolute errors are entirely uncorrelated, the percentage errors (errors expressed as percentages or other fractions of the observed quantity) will often show a large amount of correlation. In such a case as this, the important question arises whether we are, from other considerations, justified in using ratios (or indices) rather than absolute measures. If we are, the index correlation is not to be regarded as spurious†.

* Karl Pearson: "On a form of Spurious Correlation which may arise when Indices are used in the Measurement of Organs," *Proc. Roy. Soc.* Vol. LX. 1897, p. 489.

† Spurious correlation of this (as of any) kind can be eliminated by partial correlation.

VII. The Significance of the Correlation Coefficient.

A correlation between two series of measurable characters may be regarded quite generally as due to some sort and degree of community or identity of causation, and the coefficient of correlation, r, is some sort of measure of this identity of causation. We may, then, quite legitimately inquire what numerical relation r bears to the identity. To take an example from Psychology: when we find that the (partial) correlation between performance in Algebra and performance in Arithmetic is ·75, does this mean that 75 % of combined algebraical ability and arithmetical ability is common to the two, the mental operations being practically identical in the two subjects for this portion, or does the ·75 measure the degree of identity in some more complicated way?

The general solution of the problem appears to be impossible. Its meaning may, however, become clearer after a consideration of a particular case in which the significance of the correlation coefficient *is* apparent. The case referred to is an experiment in dice-throwing devised by the late Professor Weldon*, and called after him *Weldon's experiment*†. Let a dozen dice be shaken up in a box and thrown again and again, the number of dice showing four or more spots on the uppermost face in each throw being recorded. Then if the results of the 1st, 3rd, 5th, etc. throws, and those of the 2nd, 4th, 6th, etc., be written down in the form of two series, the 2nd opposite the 1st, the 4th opposite the 3rd, etc., these two series will obviously show no correlation, since the result of the 2nd throw is in these circumstances quite independent of the result of the 1st, and likewise with the 4th and 3rd, the 6th and 5th, etc. Let now the second of each of these pairs of throws be made dependent on the first of each pair, in the following way. Let a certain proportion of the dice, say 6, be stained red, and then, after throwing all the dice and counting indiscriminately

* W. F. R. Weldon: "Inheritance in Animals and Plants" in *Lectures on the Method of Science*, edited by T. B. Strong, Oxford, Clarendon Press, 1906, pp. 100–106.

† A. D. Darbishire: "Some Tables for illustrating Statistical Correlation," *Memoirs and Proceedings of the Manchester Literary and Philosophical Society*, Vol. LI. Pt. III. 1906–1907.

all those showing 4 or more spots on the uppermost face, for the second throw gather up only the white dice, leaving the red upon the table, and throw, counting however both white and red dice when computing the score. The number due to the red dice will thus be common to the two scores. Successive pairs of throws are then made in the same way. If now the two series of values be compared, they will be found to be correlated together; a result to be expected, since the score of the second throw of each pair was to a certain extent dependent on that of the first. The actual value of the correlation coefficient will be found to be, in the case taken, $=\frac{6}{12}$ or ·5. Quite generally, if n is the total number of dice thrown, and l the number left on the table after the first of each pair of throws,

$$r=\frac{l}{n}.$$

This result admits of a theoretical proof as follows:

Let the successive scores of the pairs of throws be

$$\begin{array}{ll} a_1+c_1 & c_1+b_1 \\ a_2+c_2 & c_2+b_2 \\ a_3+c_3 & c_3+b_3 \\ \vdots & \vdots \\ a_N+c_N & c_N+b_N \end{array}$$

and let $x=a+c, \quad y=b+c.$

Then $$r_{xy}=\frac{\sigma_x^2+\sigma_y^2-\sigma^2_{x-y}}{2\sigma_x\sigma_y},$$

$$\sigma^2_{x-y}=\frac{S(x-y)^2}{N}=\frac{S(a-b)^2}{N}=\frac{S(a^2)}{N}+\frac{S(b^2)}{N}=2\sigma_1^2,$$

$$\sigma_x^2=\frac{S(a+c)^2}{N}=\frac{S(a^2)}{N}+\frac{S(c^2)}{N}=\sigma_1^2+\sigma_2^2,$$

$$\sigma_y^2=\frac{S(b+c)^2}{N}=\frac{S(b^2)}{N}+\frac{S(c^2)}{N}=\sigma_1^2+\sigma_2^2,$$

$$\therefore\ r_{xy}=\frac{2\sigma_2^2}{2(\sigma_1^2+\sigma_2^2)}=\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}.$$

Let $l=$ no. of dice left on table
„ $m=$ „ „ „ thrown twice } , $l+m=n.$

Then, since the distribution of results in the throwing of n dice N times is given by the successive terms of the binomial expansion of

$$N(p+q)^n,$$

of which the S.D. $= \sqrt{npq}$ (putting N equal to unity),

$$\therefore \quad \sigma_2^2 = lpq, \qquad \sigma_1^2 = mpq,$$

and
$$r_{xy} = \frac{l}{l+m} = \frac{l}{n}.$$

In the case of Weldon's experiment, then, the proportion of factors common to the two series is given by the correlation coefficient itself, but in the cases ordinarily met with in biology and psychology the relations of the factors at work must be much more complex. The extent of identity of causation is measured by $\phi(r)$, some function of r, but the form of this function is unknown to us. It is, for example, an unwarranted assumption to say, as Spearman does, that the *square* of the correlation coefficient measures "the relative influence of the factors in A tending towards any observed correspondence (with another variable B) as compared with the remaining components of A tending in other directions*."

* C. Spearman: "The Proof and Measurement of Association between Two Things," *American Journal of Psychology*, Vol. xv. Jan. 1904, p. 75.

CHAPTER II

HISTORICAL

The history of the use of the theory of correlation in Psychology can hardly be said to have begun earlier than the commencement of the present century. During the previous twenty years, indeed, a great deal of work had been done by many observers in measuring simple mental abilities (by the "mental test" method) in larger or smaller groups of subjects, and attempts had even been made to determine in what way these abilities were related to one another and to more general mental ability, or "general intelligence."

Owing, however, to a universal lack of knowledge of the mathematical theory of correlation among psychologists during this period, the results were not obtained in a form suitable for comparison with one another, so that it is not surprising to find that they hopelessly contradict one another. The heterogeneity of the material worked with, the non-elimination of irrelevant factors, and the absence of any measure of the "probable error" of the results make the conclusions drawn by the investigators themselves from their researches utterly unreliable.

The first investigation showing any mathematical precision was that published by Clark Wissler* in 1901. It contained (*inter alia*) an account of the careful application of a large number of simple mental tests upon over 200 college students, and a correlation of the results with one another and with the students' marks in the various subjects of the college curriculum. The mental tests were found to correlate but slightly with one another or with ability in college subjects of study,

* Clark Wissler: "The Correlation of Mental and Physical Tests," *Psychological Review, Monograph Supplement*, Vol. III. No. 16, June 1901.

though these latter showed considerable correlation with one another (·30—·75).

In the following year Aikins and Thorndike* published results which were in a sense confirmatory of those of Wissler, since, notwithstanding the greater similarity to one another of the functions investigated than in Wissler's research, the correlations were again found to be *low*. For example, different tests devised for the measurement of "speed of association" were found to show hardly any correlation—a result which seems to furnish some justification for the authors' statement that "quickness of association as an ability determining the speed of all one's associations is a myth" (*op. cit.* p. 375). A similar lack of close relationship was found in the case of other mental functions which would, on the evidence of introspection alone, be confidently classed as particular instances of the same general mental function.

In C. Spearman's† researches, published in 1904 and 1906, a change in tactics appears. Instead of measuring large numbers of individuals, as his predecessors had done, Spearman contents himself with small numbers, groups of less than 40 in his first research, and as few as 11,—in the case of one of the tests, 7—in the second; but he proposes to make up for the unreliability thus introduced into the results by a more careful measurement of his cases, and the application of "corrections" to the "raw" values of his correlation coefficients by means of appropriate mathematical formulae.

The "corrections" are of two kinds. One is that for the "elimination of irrelevant factors" and is nothing new in the theory of correlation, being simply the method of "partial" correlation (see p. 63 above).

For example: To eliminate effects of difference of *age* in the group experimented on one would determine the "partial"

* "Correlations in the Perceptive and Associative Processes," *Psychological Review*, Vol. IX.

† C. Spearman: "'General Intelligence' Objectively Determined and Measured," *American Journal of Psychology*, Vol. XV. 1904, pp. 201–292.

F. Krueger and C. Spearman: "Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten," *Zeitschrift für Psychologie*, Bd. XLIV. 1906.

correlation between the two characters under consideration, for "age constant." Thus, in Yule's notation, if 1 and 2 represent the characters to be compared and 3 represents "age,"

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{(1 - r^2_{13})(1 - r^2_{23})}}.$$

A similar procedure is needed for eliminating the effects of differences of sex, etc. A preferable course, however, would be to dispense as far as possible with the necessity for such correcting by selecting groups of individuals of the same age, sex, etc.

The other method of correction advocated is that for the "elimination of observational errors." Obviously, errors of observation must make any correlation, worked from measurements containing them, different from (generally, though not universally, *less* than) the true value of the correlation. The formula, in its full form, which Spearman has proposed as a means of correcting for this, and of which he has given a proof in the *Am. J. P.**, is as follows:

$$r_{xy} = \frac{\sqrt[4]{r_{x_1 y_1} r_{x_1 y_2} r_{x_2 y_1} r_{x_2 y_2}}}{\sqrt{r_{x_1 x_2} r_{y_1 y_2}}},$$

where x, y are the *true* values to be correlated, and x_1, x_2, y_1, y_2 are two pairs of *obtained* values.

$r_{x_1 x_2}$, and similar coefficients, are called by him "reliability coefficients." They represent the correlation of two distinct series of measurements of the same mental capacity.

The proof of Spearman's formula is only valid on the assumption that the errors of measurement are uncorrelated with each other or with x or y†. Thus, if x, y, δ, ϵ represent deviations from means, and

$$x_1 = x + \delta_1, \qquad x_2 = x + \delta_2,$$
$$y_1 = y + \epsilon_1, \qquad y_2 = y + \epsilon_2,$$

* C. Spearman: "Demonstration of Formulae for True Measurement of Correlation," *American Journal of Psychology*, Vol. XVII. 1906.

† See Mr G. Udny Yule's short proof of the formula, quoted in my pamphlet on "Some Experimental Results in Correlation," *Comptes Rendus du VI^{me} Congrès International de Psychologie*, Genève, Août 1909.

then $\qquad S(x\delta)$, etc. $=0, \qquad S(\delta_1\delta_2)$, etc. $=0$;

also $$S(x_1y_1) = S(xy) = S(x_2y_2).$$

[All this is involved in Mr Yule's proof.]

Now, these are very large assumptions to make. Even in cases where the quantities δ, ϵ are genuine errors of measurement, there are strong reasons for assuming (on general principles and also from experimental evidence)* that they *will* be correlated. But in the case of almost all the simpler mental tests the quantities δ and ϵ are not errors of measurement at all. They are the deviations of the particular performances from the hypothetical average performance of the several individuals under consideration. Thus they represent the *variability* of performance of function *within* the individual. When an individual in the course of three minutes succeeds in striking through 100 *e*'s and *r*'s in a page of print on one day, and 94 under the same conditions a fortnight later, there is no error of observation involved. The numbers 100 and 94 are the actual true measures of ability on the two occasions. The average or mean ability, which is the more interesting measure for the purposes of correlation, is doubtless different from either, but that does not make the other two measures erroneous. Evidently in these cases δ and ϵ represent *individual variability*, and to assume them uncorrelated with one another or with the mean values of the functions is to indulge in somewhat *à priori* reasoning.

There are two comparatively simple ways of testing the assumption:

$$(1) \quad S(x_1y_1) = S(xy) = S(x_2y_2),$$

$\therefore$ $S(x_1y_1) - S(x_2y_2)$ should $=0$ within the limits of the probable error of the difference.

I have applied this test to the case of correlation between accuracy in bisecting lines and accuracy in trisecting them in 43 adult subjects.

* See Karl Pearson: "On the Mathematical Theory of Errors of Judgment, with special reference to the Personal Equation," *Phil. Trans.* A, Vol. 198, pp. 235–299.

Here $S(b_1t_1) - S(b_2t_2) = 137780 - 60036$
$= 77744,$

$$\text{P.E. of } S(xy) = \cdot 67449 \sqrt{\frac{p_{22} - p_{20}p_{02}}{n}},$$

in Pearson's notation.

$$= \frac{\cdot 67449}{\sqrt{n}} \sqrt{\frac{S(xy)^2}{n} - \frac{S(x^2)\,S(y^2)}{n^2}},$$

$$\text{P.E. of } S(b_1t_1) = 687\,; \qquad \text{P.E. of } S(b_2t_2) = 365\,;$$

$$\therefore \ \text{P.E. of } S(b_1t_1) - S(b_2t_2) = \sqrt{687^2 + 365^2}$$
$$= 778.$$

Since 778 is less than one-third of 77744, the formula cannot be employed to obtain the correlation between mean abilities in bisecting and trisecting lines.

$$(2) \quad r_{\substack{X_1 - X_2 \\ Y_1 - Y_2}} = \frac{S\{(x_1 - x_2)(y_1 - y_2)\}}{\sqrt{S(x_1 - x_2)^2 \, . \, S(y_1 - y_2)^2}}$$
$$= \frac{S\{(\delta_1 - \delta_2)(\epsilon_1 - \epsilon_2)\}}{\sqrt{S(\delta_1 - \delta_2)^2 \, . \, S(\epsilon_1 - \epsilon_2)^2}}$$
$$= 0$$

if errors are uncorrelated with one another (since numerator then $= 0$).

Applying this test to the same case of bisection and trisection, I get

$$r_{\substack{B_1 - B_2 \\ T_1 - T_2}} = 0{\cdot}30 \pm 0{\cdot}09,$$

which proves once more the inapplicability of the formula.

I applied test (2) also to the case of correlation between speed of addition of figures and accuracy of addition in a group of 38 school children (girls between the ages of 11 and 12, Group II below) and found

$$r_{\substack{S_1 - S_2 \\ A_1 - A_2}} = 0{\cdot}35 \pm 0{\cdot}09.$$

Even when test (2) does give the value 0, we can only conclude from this that

$$S(\delta_1\epsilon_1) + S(\delta_2\epsilon_2) = S(\delta_1\epsilon_2) + S(\delta_2\epsilon_1).$$

We cannot conclude that the formula is applicable, unless we have further independent evidence*.

For the reasons presented above, I should prefer to avoid the use of Spearman's formula—by increasing the number of the original measurements of each ability†—and would also suggest that his so-called "reliability" coefficients might in some cases be more appropriately termed "coefficients of *individual* correlation," since they are more analogous to Karl Pearson's "correlation of undifferentiated like parts" than to anything else‡.

The above discussion raises the interesting question as to the relation between ability and variability, and the correlation coefficient between mean ability and the standard deviation would be the best measure of this relation. When the measurements have been made on only two separate occasions, the expression $r_{\substack{X_1+X_2 \\ X_1 \sim X_2}}$ might be regarded as a rough measure of the relation, and I would suggest that it be called a "variability coefficient"—not to be confused with the "coefficient of variation," which $= \frac{100\sigma}{\text{mean}}$. If x_1 and x_2 are chance values, and if the distribution of abilities at the given task within one and the same individual is approximately normal, then

$$x_1 \sim x_2 = \tfrac{26}{23}\sigma \text{ (approx.)§},$$

so that there is sufficient justification for this value.

To return to Spearman's researches. In his first series of investigations (*Am. J. P.* Vol. XV.) the following groups of subjects were measured: 24 village school children of both sexes,

* In a recent paper (*British Journal of Psychology*, Vol. III. Oct. 1910, pp. 271 ff.) Spearman admits the difficulties attending the use of his "correction formula," and brings forward another formula to correct for "accidental" as distinct from "constant" errors. Since no experimental evidence is given as to the validity of the assumptions which this new formula involves, I merely content myself with mentioning the fact of its publication here.

† This plan would also have the advantage of keeping the probable error low. Correction by Spearman's formula while raising the value of the coefficient raises the size of the P.E. in the same proportion.

‡ *Grammar of Science*, 2nd Ed., pp. 393, 397.

§ Karl Pearson: "Francis Galton's Difference Problem," *Biometrika*, Vol. I. p. 399.

age limits 10·0—13·10; 23 boys of a high class preparatory school, age limits 9·5—13·7; and 27 adults of both sexes, age limits 21—78. The tests employed were those for pitch discrimination, weight discrimination and discrimination of light intensities; and measures of intelligence were obtained, in the case of the children, from results of school examinations, grading by teachers, and grading of one another by the children themselves (measure of common sense). Only in the cases of "common sense" (village children) and "final school order" were reliability coefficients calculated. Consequently the reliability coefficient for common sense (·64) was used to bring up the various sensory discrimination coefficients to their true values. Corrections were made for age and sex, and the final results were then found to be comparatively high in the village school group.

In the preparatory school group the highest raw coefficient was 0·32 ± 0·08, that between weight discrimination and light discrimination; in the group of adults that between pitch discrimination and light discrimination was the highest, being likewise 0·32 ± 0·08. The various school subjects in the preparatory school (Classics, French, English, Mathematics) were found to correlate highly with one another, and when, with the inclusion of pitch discrimination and music, they were arranged in rows and columns in order, from left to right and from above downwards, of the average degree of the correlation of each with all the rest, they were found to form a *hierarchy* such that every coefficient was greater than any to the right of it (in the same row) or below it (in the same column). The order was Classics, French, English, Mathematics, Pitch Discrimination, Music. Spearman takes this to be one form of evidence for the unity of the intellective function, defined as follows:—"Wherever branches of intellectual activity are at all dissimilar, then their correlations with one another appear wholly due to their being all variously saturated with some common fundamental function (or group of functions)*." He considers that by the use of his formula for the correction of observational errors he can demonstrate the same thing (the existence of a "central function")

* *Am. J. P.* Vol. xv. *op. cit.* p. 273.

and can in particular show "that the common and essential element in the intelligences wholly coincides with the common and essential element in the sensory functions*." The method of proof is as follows:

Let x_1, x_2 be two distinct measures of sensory discrimination, and y_1, y_2 two distinct measures of intelligence.

Then the correlation of the function common to the functions measured by x_1, x_2 with the function common to the functions measured by y_1, y_2 is equal to

$$\frac{\sqrt[4]{r_{x_1y_1} r_{x_1y_2} r_{x_2y_1} r_{x_2y_2}}}{\sqrt{r_{x_1x_2} r_{y_1y_2}}},$$

and if the two functions referred to are identical this expression should be equal to *unity*.

In the present article, Spearman uses a simplified formula,

$$\frac{r_{xy}}{\sqrt{r_{x_1x_2} r_{y_1y_2}}},$$

and puts for the numerator the *average* of the various correlations evaluated between the intelligences and the discriminations, and in the denominator puts

$r_{x_1x_2}$ = the average correlation of the intellective gradings with one another,

$r_{y_1y_2}$ = the average correlation of the gradings in discrimination with one another,

and in this way gets results approximately equal to 1 in the different groups tested.

One or two remarks may appropriately be made here. In the first place, the full formula is the only one that can be used with any meaning or justice since it is the only one which issues logically from the mathematical proof. In the second place, the applicability of the true formula must be considered in the light of its presuppositions (mentioned above, p. 83).

$S(x_1y_1)$ should $= S(x_2y_2)$ within the limits of probable error, and $r_{\substack{x_1-x_2\\ y_1-y_2}}$ should be zero. Indeed, the assumption that δ and ϵ are uncorrelated with each other or with x or y seems even

* *Op. cit.* p. 269.

more unwarrantable here than in the case of "correcting" coefficients, for which the formula was originally devised.

In Spearman's second research*, the subjects experimented on were *eleven* adults, and the measurements made were:

(1) the spatial threshold,
(2) addition of figures,
(3) memory of the order of a series of figures previously seen,
(4) combination method of Ebbinghaus (only 7 subjects took this test), and
(5) discrimination of pitch.

After corrections for disturbing but irrelevant factors it was found that (1) and (3) showed no correlation with each other or with (2), (4) and (5), but that these latter correlated highly with one another ($r_{24} = \cdot 70$, $r_{25} = \cdot 68$, $r_{45} = \cdot 64$), and the coefficients could be arranged to show a hierarchical order. The raw values were found to give evidence of a "central factor" when tested by Spearman's correction formula, though it is true that the arithmetic mean instead of the geometric mean was evaluated in the numerator (an entirely unjustifiable procedure), thus evading the difficulty that would have otherwise arisen owing to the presence in it of a *zero* value†. After working up some further data, taken from Oehrn's writings (number of subjects employed, 10), and finding the results in agreement with the hypothesis of a "central factor," and finding too that the correlation coefficients for the same pair of abilities *increase* with increase of practice and thus contradict the view of Binet‡ that correlation in mental tests is the result of lack of practice, and *diminishes* in size with increase of practice, Spearman concludes his paper with the suggestion (made "with the greatest reserve") that the central factor is a plastic function—"plastische Funktion"—of the nervous system which varies from individual to individual and exists as a determining influence for all the various psycho-physiological functions of the individual. Whatever it is, it is *not* voluntary attention (p. 106), as Binet believed it to be§.

* Krueger and Spearman: *Zeitschr. f. Psych.* Bd. XLIV., *op. cit.*

† *Op. cit.* p. 85.

‡ *L'Année Psychologique*, VI. p. 395, 1899.

§ It is perhaps hardly necessary to point out that the detailed consideration of Dr Spearman's work given here and elsewhere in the present book is, and is

Spearman's theory of an identical or absolute correspondence between "general discrimination" and "general intelligence," has been recently put to the test by Messrs Thorndike, Lay and Dean*. The subjects examined were 37 young women students and 25 high school boys. The tests for sensory discrimination were:

(1) drawing lines equal to given lines, and

(2) filling boxes with shot to equal in weight standard weights;

those for intelligence were:

(3) judgment of fellow-students, and

(4) judgment of teachers.

For the high school boys (3) and (4) were combined teachers' and fellow-students' judgments, and school marks, respectively.

In the first case, Spearman's formula gave for the correlation of the factor common to (1) and (2) with that common to (3) and (4) *the value* 0·26 *instead of* 1·00. In the second case, the value was 0·29. Moreover, Thorndike found a much higher correlation between discrimination of lengths and discrimination of weights than between either one of them and general intelligence, the coefficients being

Accuracy in drawing lines, intelligence	0·15,
Accuracy in making up weights, intelligence ...	0·25,
Accuracy in drawing lines and making up weights	0·50.

Thus the results were in decided conflict with both parts of Spearman's concluding statement "that all branches of intellectual activity have in common one fundamental function (or group of functions) whereas the remaining or specific elements of the activity seem in every case to be wholly different from that in all the others†."

meant to be, a tribute to the outstanding importance of that work. No one can appreciate more vividly than does the writer the valuable suggestiveness of all Dr Spearman's writings, above all in respect to the special precautions therein emphasized as requisite in measuring *psychical* (as distinct from anatomical, etc.) correlation.

* Thorndike, Lay and Dean: "The Relation of Accuracy in Sensory Discrimination to General Intelligence," *Am. J. P.* Vol. xx. July 1909, pp. 364–369.

† C. Spearman, *Am. J. P.* Vol. xv. *op. cit.* p. 284.

Thorndike sums up as follows: "In general there is evidence of a complex set of bonds between the psychological equivalents of both what we call the formal side of thought and what we call its content, so that one is almost tempted to replace Spearman's statement by the equally extravagant one that there is *nothing whatever* common to all mental functions, or to any half of them*."

Quite recently (Dec. 1909) there has appeared a detailed account by Mr Cyril Burt† of some experiments on children, "commenced with a view to testing in practice the mathematical methods of Dr Spearman, and to verifying the experimental results both of Dr Spearman and of Prof. Meumann" (*op. cit.* p. 95).

Two groups of subjects were experimented on: (1) a group of 30 boys (age limits, 12·6—13·6) from an elementary school, and (2) a group of 13 boys (same age limits) from a preparatory school. The mental tests employed were discrimination of (1) weights, (2) pitch, and (3) lengths of lines (W. H. R. Rivers' adjustable apparatus used), measurement of (4) "spatial threshold," (5) rate of tapping (modification of Binet's "petits points"), (6) card-dealing, (7) card-sorting, (8) alphabet-finding, (9) immediate memory, (10) "mirror test," (11) "spot pattern" test (tachistoscopic: McDougall's apparatus), (12) "dotting" test (apparatus used: "dotting apparatus" of McDougall, devised for testing and recording continued maximal voluntary concentration of attention). A measure of "imputed intelligence" was also obtained, from judgments of teachers and school-fellows.

Tests (1), (2), (3) and (4) were classed as sensory tests.

Tests (5) and (6) were classed as motor tests.

Tests (7) and (8) were classed as sensori-motor tests.

Tests (9), (10) and (11) were classed as association tests.

Test (12) was classed as a test of voluntary attention.

The tests were applied to the children individually, and repeated at least once. The reliability coefficients were in most

* Thorndike, Lay and Dean: *op. cit.* p. 368.

† Cyril Burt: "Experimental Tests of General Intelligence," *British Journal of Psychology*, Vol. III. Parts I. and II. Dec. 1909, pp. 94–177.

cases high, only those of tests 1 (preparatory school), 3, 5, 7 and 8 (preparatory school), 10, and 11 being under ·6, and of these only 7 and 8 (preparatory school) were lower than ·5; so that the average ability of each individual tested was determined with fairly close approximation.

The sensory tests were found to correlate but slightly with one another (see p. 177, foot of page), and the highest correlation with imputed intelligence was that of pitch-discrimination in the elementary school, which $= 0{\cdot}40 \pm 0{\cdot}10$. The remaining tests correlated much more closely with one another and with imputed intelligence. The chief general result of importance which the author considers himself to have demonstrated conclusively is that the coefficients form a "*hierarchy*" (see p. 87 above) in both groups of subjects tested. Considering this result due to the existence of a "central factor," he suggests that the central factor indicated is "voluntary attention," since the dotting test (no. 12) heads the hierarchy in both cases, and also "the intelligence coefficients for successive series with the same test" *diminish* from the first onwards (except in the case of tests 5 and 11).

This conclusion contradicts Spearman's view*, and coincides with the view of Binet†.

Psychologically the research is excellent. The tests are in many instances original, and the possible psychological factors involved in each are debated with exhaustive completeness. With regard to the "statistical" side of the work, I should like to make a few very brief remarks:

(1) The number of individuals tested in the preparatory school (13) is *far too small.* The size of probable errors is not known for such small numbers, though it must be large. Again, the two groups tested are not comparable to one another. It is a pity that more groups of each kind of school were not examined, in order to obtain results that would be comparable and would serve as "controls" of one another.

(2) It is a pity that Spearman's formula, $r = \sin\left(\frac{\pi}{2} R\right)$,

* See above, p. 89.

† *L'Année Psychologique*, VI. p. 395. See also above, p. 89.

should have been employed (see above, pp. 71—76). In particular, the proof of the probable error of this value given at the foot of p. 109 is incorrect. The true value of the P.E. of r calculated from R (for zero value of the coefft.) is $\frac{\cdot 7738}{\sqrt{n-1}}$, not $\frac{\cdot 6745}{\sqrt{n}}$ (see above, p. 73).

(3) It is very doubtful whether the results do prove the existence of a "hierarchical" order. The orders for the elementary school and for the preparatory school do not correspond at all closely, the rank-correlation, ρ, working out to 0·56 only. The plan of comparing *average* deviations and *average* P.E.'s on p. 163 is open to serious objection (apart from the fact mentioned in footnote 2 of the same page).

(4) The standard method for investigating the inter-relations of the several coefficients would of course be that of multiple or "partial" correlation*; but this method was not employed.

(5) The probable errors of the quantities in Table VIII, p. 172, should have been given, and taken into consideration in drawing inferences from them.

These are the few difficulties that occur to one on a rapid reading of the paper, though they detract but little from its general usefulness.

The results of a very detailed investigation into the relationship of intelligence to the size and shape of the head, and to other physical and mental characters, by Prof. Karl Pearson†, appear in *Biometrika*, Vol. v. 1906—1907. The subjects measured were 1000 Cambridge graduates and considerably more than 5000 school children. Special care was taken in drawing up a quantitative scale of intelligence, adjustments being made so that the results fitted a "normal" or Gaussian distribution. The correlations were worked out in several different ways,—correlation coefficient (r), correlation ratio (η), coefficient of mean square contingency, and the method, first

* See above pp. 51, 52 and 63, 64; also Appendix III.

† Karl Pearson: "On the Relationship of Intelligence to Size and Shape of Head, and to other Physical and Mental Characters," *Biometrika*, Vol. v. 1906-1907, pp. 105-146.

suggested in this paper, of the *analograph.* Pearson describes this last method as follows: "In the case of intelligence, I take a normal scale as my base line and plot up the *percentage* of the character for each grade of intelligence along the centroid vertical of the corresponding range, drawing a horizontal line to represent the mean percentage in the population at large. We thus obtain a diagram, which I will venture to call an *analograph.*

"If the percentage increases *or decreases* continually with intelligence (or with the base character, whatever it may be), I term the relationship *homoclinal*; if the percentage does not reach its maximum with the maximum or minimum of intelligence, I term the diagram *heteroclinal* *."

The mental characters investigated (which alone interest us here) were—temper, popularity, self-consciousness, shyness, conscientiousness, quiet habits, and handwriting, and in the case of the school children the correlations of intelligence with these, and also with age, were found to be as follows:

Character	Boys	Girls
Conscientiousness	·46	·43
Handwriting ...	·28	·30
Popularity ...	·22	·30
Temper	·19	·22
Shyness	·03	·18
Self-Consciousness	·10	·03
Age	·05	·08
Quiet Habits ...	·04	·09

The low value of the correlation of intelligence with *age* should be noted.

Space does not allow of further consideration of Karl Pearson's paper. It is undoubtedly the most important contribution that has hitherto been made to the subject of the relation of intelligence to other mental and physical characters in the human organism.

Other articles dealing with psychological classifications and measurements, which have recently appeared in *Biometrika*, are

* *Op. cit.* p. 129.

two accounts* of the results of applying the latest statistical methods to the working up of material collected by certain psychologists. In the case of the first of these two, viz., the research of Messrs Heymans and Wiersma into the inheritance of physical characters, the material had been obtained by the questionnaire method, 90 questions in all being sent out, divided into six classes according to the nature of the characters to which they referred. Examples of the questions are:

Class I. Movements and occupations (Q. 1—8), mobility or restfulness; steady or temporary industry or laziness; etc.

Class V. Inclinations (Q. 44—81), use of alcohol; vanity; ambition; love of truth; etc.

More than 400 sets of answers were returned.

In cases where the questions involved only two alternative answers, the four-fold table method was used in evaluating the correlation between parents and children; where four or five alternative answers were demanded, the contingency method was used, and the result tested by a four-fold division of the table. The results were corrected for assortative mating by means of the correlation coefficient between father and mother. (In the tables for coefficient of assortative mating, the parents were weighted with the number of their children.) The coefficients of assortative mating were found to be more frequently negative than positive. The account in *Biometrika* concludes as follows: "After allowing for the effect of assortative mating, the mean coefficients of resemblance between fathers and sons, and between mothers and daughters, calculated for one set of characters by the contingency method and for another set by the four-fold correlation method, come in each case to very

* (*a*) "The Inheritance of Psychical Characters," being further Statistical Treatment of Material Collected and Analysed by Messrs G. Heymans and E. Wiersma: by Edgar Schuster and E. M. Elderton, *Biometrika*, Vol. v. 1906–1907, pp. 460–469.

(*b*) "On the Association of Drawing with other Capacities in School Children," by E. M. Elderton; being a biometric evaluation of the results of E. Ivanhoff: "Recherches expérimentales sur le Dessin des Ecoliers de la Suisse Remande," *Archives de Psychologie*, III. VIII. 1908. *Biometrika*, Vol. VII. pp. 222–226.

nearly $\frac{1}{3}$, the value proposed originally for the parental inheritance coefficient by Sir F. Galton" (p. 469).

In the second case, that of Ivanhoff's data in his investigations into the relation of *drawing* to other mental abilities in school children, Miss Elderton evaluated the results by the contingency method (nine-fold grouping, divisions being "fort," "moyen," and "faible") and, in some cases, by the four-fold table method also. The conclusions she was able to draw were briefly as follows:

(1) Drawing correlates with other abilities more closely in girls than in boys.

(2) The correlations in all cases are extraordinarily *low*. Miss Elderton suggests that "this may arise possibly from capacity in drawing being a hereditary character, having small association with other school measures of fitness."

(3) General ability shows the highest of the correlations, in both sexes.

(4) The so-called "psycho-pedagogic characters"—attention, obedience, industry, cleanliness, and temper—correlate less closely than proficiency in the ordinary subjects of the school curriculum do. In the case of the girls, they are all the lowest on the list.

Accounts of other investigations into the relationship existing between the various subjects of the school and college curriculum are to be found in papers by W. P. Burris, A. G. Smith, Brinckerhoff, Morris and Thorndike*. The results are somewhat conflicting, and in every case surprisingly low. Thorndike and Fox† found low correlations between part-capacities of the same subject (arithmetic); and Burris obtained a coefficient for algebra and geometry which was nearly as low as that between mathematics and a non-mathematical subject. I have myself found‡ that, after correction for effects of irrelevant conditions, algebra and geometry show hardly

* All in Vol. XI. of *Columbia Contributions to Education.*

† *Op. cit.* pp. 138–143.

‡ William Brown: "An Objective Study of Mathematical Intelligence," *Biometrika*, Vol. VII. Part III. 1910.

any correlation at all,—in fact, the partial coefficient for "arithmetical ability constant" is zero.

An *experimental* investigation of the inter-relations of some of the higher mental processes in 50—90 normal school students was made by A. Peterson* in 1908. The marks used were good, medium and poor, and the following are some representative coefficients selected from those obtained:

Reasoning and generalizing	·95,	Generalizing and memory	·40
" " memory ...	·40,	" " accuracy	·28
" " accuracy ...	·45,	Memory and accuracy ...	·31

As one further illustration of the use to which the theory of correlation has been put in Educational Psychology, mention may be made of an investigation by Thorndike† into the correlation between entrance marks and later performances at college. For total marks the coefficients were:

Freshman year, ·62; Sophomore year, ·50; Junior year, ·47; Senior year, ·25.

Taking the various subjects separately, Thorndike found a slight superiority over the rest in the case of science, and a slight inferiority in the case of mathematics. The number of cases worked on was 56—84.

In a recent short paper‡, Thorndike brings forward correlation results which he considers to be "very strong evidence of the dependence of efficiency of memory upon content and of the specialisation of mental functions in general." On the other hand, he finds that "the relation between retention of the effects of an experience for one or two minutes and their retention for one or two days seems to be one of the closest yet measured in human nature." The "raw" correlation was $\cdot 5\frac{1}{2} \pm \cdot 1$, $(n = 40)$, which when corrected for mixture of sexes and inaccuracy of measurement gave a final value of $\cdot 8 \pm \cdot 1$.

* A. Peterson: "Correlation of Certain Mental Traits in Normal School Students," *Psychological Review*, Vol. xv. 1908, pp. 323–338.

† E. L. Thorndike: "An Empirical Study of College Entrance Examinations," *Science*, N.S. Vol. xxiii. June 1906, pp. 839–845.

‡ E. L. Thorndike: "The Relation between Memory for Words and Memory for Numbers, and the Relation between Memory over Short and Memory over Long Intervals," *American Journal of Psychology*, Vol. xxi. 1910, pp. 487, 488.

CHAPTER III

SOME EXPERIMENTAL RESULTS*

The research to be described in this chapter was devised for the purpose of determining to what extent correlation exists between certain very simple mental abilities in cases where the individuals experimented upon are, as near as may be, identically situated with respect to previous practice, general training, and environment; and how closely, if at all, these elementary abilities are related to general intellectual ability as measured by teachers' judgments, school marks, etc. Every effort was made to keep the groups of individuals tested as *homogeneous* as possible; and instead of measuring irrelevant factors and "correcting" for them in the later stages of the research, the influence of such irrelevant factors was excluded right from the beginning by a rigorous segregation of the material, and in other ways.

The groups of individuals to which the tests were applied, were as follows:

Group I, 66 boys of a London elementary school, all between the ages 11 and 12.

Group II, 39 girls of a London elementary school, all between the ages 11 and 12.

Group III, 40 boys of a London higher grade school, all between the ages 11 and 12.

Group IV, 56 training college students (women), of the same year and of approximately the same age.

Group V*a*, 35 university students (men).

[Group V*b*, 23 university students (women).]

* Published in the *British Journal of Psychology*, Vol. III. Pt. 3, Oct. 1910. A few unimportant alterations have been made, and a summary of the results has been added.

Little need be said as to the nature of the groups. Group III was as homogeneous as could possibly be expected or desired. The individuals were not only of the same age, but also belonged to the same form and had all worked for months past under exactly the same environment (same teacher, etc.). They were, however, a rigorously *selected* class, as might be expected from the character of the school.

Group IV was also thoroughly homogeneous. During an entire year previous to the application of the tests they had lived under exactly the same environment.

In Group II there was a slight mixing of "standards" which introduced a slight degree of heterogeneity, but the effect of this on the results must have been very small.

Group I was also slightly heterogeneous owing to mixture of standards, and the results show that the effect of this was somewhat greater than in the preceding case.

Group V*a* was fairly homogeneous, but was of course a "selected" group. The same remarks apply to Group V*b*, but owing to the smallness of the numbers (23) tested, the results were worked out by the method of ranks (ρ), which was considered good enough under such circumstances, and they are recorded avowedly as mere approximations.

Other groups of school children were also tested, but as the marking of the results is not yet complete, no further reference will be made to them here, except in the case of the "addition tests," for which see Appendix II, note at top of first page.

As regards the *tests* employed, they were chosen not so much for their novelty (though a few of them are new and the methods of applying the tests were determined in every case entirely by the requirements of the circumstances), nor so much for their *à priori* likelihood of showing inter-correlation, as for their convenience in admitting of application to an entire group of subjects simultaneously and *unobtrusively*. The following is a list of them:

1. Crossing through letters *e* and *r* in a page of print.
2. Crossing through letters *a*, *n*, *o*, and *s* in a page of print.

3. Crossing through every letter in a page of print.

4. Adding up single digits in groups of ten. Measurement of (*a*) speed, (*b*) accuracy.

5. Bisecting ten printed lines (80 mm. long), and putting in one of the points of trisection in each of ten other lines (90 mm. long).

6. Müller-Lyer Illusion. Measurement of (*a*) size, (*b*) mean variation.

7. Vertical-Horizontal Illusion. Measurement of (*a*) size (*b*) mean variation.

8. Mechanical Memory (permanent), tested by means of nonsense syllables.

9. Memory for poetry.

10. Combination test (Ebbinghaus).

In the case of Groups II and III, recourse was also had to:

11. Marks for Drawing.

12. Total School Marks.

13. Grading for General Intelligence (two independent measures).

Finally, with Groups V*a* and V*b*, the following test was also employed:

14. Association-time (uncontrolled). Measurement of rate of sequence of ideas called up by a stimulus-word.

The performances of the several groups in these tests admit of comparison in terms of the mean, standard deviation, and coefficient of variation, provided that the probable errors of these constants are also evaluated.

With the exception of test (9), and, in some cases, of test (8), every test was applied twice, the second test being given about a fortnight after the first, and at the same hour of the day. In the case of the school-children, I myself applied both tests in the presence of the form-master or mistress. The adults whose measurements are recorded and employed in the present research were also tested, with hardly an exception, by myself. It should be added that the research commenced with a very much larger number of adults (university students and others), amounting to over 100, but the much smaller numbers recorded as Groups V*a* and V*b* were alone used for the evaluation of coefficients, since they alone displayed sufficient reliability and homogeneity for the purpose.

(1) *ER Test.*

Pages of French words, arranged in irregular order so that they did not "make sense," and so chosen that the number of *e*'s and *r*'s was approximately constant from line to line, were employed. The page was given out face downwards, and at a given signal the subject turned it over and proceeded to cross through every *e* and every *r* that he came to, beginning with the first line and moving down line by line, until he received the signal to stop. The time allowed for the test was 5 minutes in the case of the children and 3 minutes in the case of the adults. The subject was urged to avoid passing over any of the stated letters but otherwise to work as quickly as possible. Before the commencement of the test, a full explanation of it was given to the group, illustrated by examples on the blackboard. This was done in the case of every test. A different set of words was employed in the second test.

System of marking:

1 mark for every letter crossed through correctly;

−1 mark for every letter passed over or crossed through incorrectly.

Results of ER Test.

Group	Mean	σ	Coefficient of Variation	Reliability coefficient (r_1) for each test	Rel. coeff. (r_2) for amalgamated pair of tests, $r_2 = \frac{2r_1}{1+r_1}$ †
I	377 ± 6	68 ± 4	18 ± 1·1	·60	·75
II	362 ± 8	71 ± 5	19·5 ± 1·5	·65	·79
III	417 ± 6	57 ± 4	13·7 ± 1·1	·75	·86
* V (*a*)	204 ± 4·5	41 ± 3·3	20 ± 1·7	·97	—

* One test only.

† r_2 measures the extent to which the amalgamated results of the two tests

(2) *ANOS Test.*

The method of procedure was identical with that described above for test (1), except that the letters to be crossed through were of four kinds instead of two, and that the time allowed was 5 minutes for children and also for adults.

Results of ANOS Test.

Group	Mean	σ	Coefficient of Variation	Reliability coefficient r_1	Reliability coefficient r_2
I	161 ± 5	$58 \pm 3{\cdot}5$	$36 \pm 2{\cdot}4$	·77	·87
II	$191 \pm 5{\cdot}6$	$51 \pm 3{\cdot}8$	$26{\cdot}5 \pm 2{\cdot}2$	·84	·91
III	228 ± 6	$56 \pm 4{\cdot}1$	$24{\cdot}4 \pm 2{\cdot}0$	·81	·89

(3) *Motor Test.*

In this test the subjects were asked to cross through *every* letter in a page of printed French words. Time allowed in all cases—three minutes. Method of procedure otherwise identical with that for (1) and (2).

---

would correlate with a similar amalgamated series of two other applications of the same test. If x_1, x_2, x_1', x_2' be two pairs of results (x denoting, as usual, deviation from the mean value), we may assume that

$$\sigma x_1 = \sigma x_2 = \sigma x_1' = \sigma x_2' = \sigma x \text{ (say)},$$

and that $$S(x_1 x_1') = S(x_1 x_2') = S(x_2 x_1') = S(x_2 x_2') = n\sigma_x^2 r_1.$$

Hence we get $$r_2 = \frac{S(x_1 + x_2)(x_1' + x_2')}{n\sigma x_1 + x_2 \, \sigma x_1' + x_2'}$$

$$= \frac{4n\,\sigma_x^2 r_1}{n(2\sigma_x^2 + 2r_1\sigma_x^2)}$$

$$= \frac{2r_1}{1 + r_1}. \quad \text{Q. E. D.}$$

It is easily seen that the amalgamation of 4 tests gives a reliability coefficient $= \frac{4r_1}{1 + 3r_1}$; and, in general, for n tests we have

$$r_n = \frac{nr_1}{1 + (n - 1)\, r_1}.$$

This last formula furnishes a ready means of determining, from the reliability coefficient of a single test, the number of applications of the test which would be necessary to give an amalgamated result of any desired degree of reliability.

Results of Motor Test.

Group	Mean	σ	Coefficient of Variation	Reliability coefficient r_1	Reliability coefficient r_2
I	718 ± 13	148 ± 9	20·6 ± 1·3	·91	·95
II	720 ± 16	148 ± 12	20·5 ± 1·7	·85	·92
III	813 ± 11	103 ± 7·7	12·7 ± 1·0	·76	·86

(4) *Addition Test.*

Duplicates of pages from one of Kraepelin's *Rechenhefte* were used, adapted to the purpose by the printing of short horizontal lines below each tenth figure, and by the omission of all figures below the thirtieth in each column. Time allowed for each test—five minutes. The speed of addition [4 (*a*)] was measured by the number of sums (groups of ten digits) worked in the given time, the accuracy of addition [4 (*b*)] by the percentage of correct answers. In the case of the children, 4 (*a*) was measured by the number of *digits* added, marks being allowed for the part of a sum with which they usually ended.

Results of Addition Test.

Group	Mean		σ	
	Speed	Accuracy	Speed	Accuracy
I	235 ± 6	163 ± 3	67 ± 4	31·3 ± 1·8
II	210 ± 6	149 ± 4	52 ± 4	33·4 ± 2·5
III	287 ± 9	171 ± 1·4	79 ± 6	13·3 ± 1
*IV	55 ± 1·5	173 ± 1·5	16 ± 1·0	16·8 ± 1·1
*V (*a*)	61 ± 2·1	184 ± 1·2	19 ± 1·5	11·3 ± ·9
V (*b*)	—	—	—	—

* One sum of 10 digits taken as unit; in other cases, the number of digits added was taken as the measure.

Results of Addition Test (cont.).

Group	Coefficient of Variation		Reliability coefficient r_1		Reliability coefficient r_2	
	Speed	Accuracy	Speed	Acc.	Speed	Acc.
I	28 ± 1·8	19 ± 1·2	·82	·33	·90	·50
II	24·6 ± 2·0	22·4 ± 1·8	·69	·46	·82	·63
III	27 ± 2·2	8 ± ·6	·68	[0]	·81	—
* IV	30 ± 2·9	9·6 ± ·86	·93	·29	·96	·45
* V (*a*)	32 ± 2·8	6 ± ·5	·95	·22	·97	·36
V (*b*)	—	—	·98	·59	·99	·74

(5) *Bisection and Trisection of Lines.*

Every test paper contained ten printed lines, each 8 cm. long, for bisection; and ten printed lines, each 9 cm. long, for trisection. The lines were printed three in a row, and those situated immediately under others were shifted a little to one side. It was very certain, however, that the bisection or trisection of any one of the lines was influenced by the positions of the neighbouring lines and of the edges of the paper. This fact diminishes the value of the test, but does not of itself deprive the test of all use as a measure of one form of sensory discrimination. A more serious drawback was found to be the very great individual variability displayed, which made the reliability coefficients very low. In order to get a more reliable measure for Group V the results for bisection and trisection in both tests (i.e. of the division of 40 lines in all by each individual) were thrown together and the total taken as a measure of sensory discrimination. *Only one* point of trisection was asked for, this being put in alternately towards the left and the right ends of the successive lines. Trisection was done very unsatisfactorily by the school-children, but, in the case of adults, gave a higher measure of reliability than did bisection.

* One sum of 10 digits taken as unit; in other cases, the number of digits added was taken as the measure.

The *average crude error* was taken as the measure of inaccuracy, since it was found to give more concordant results (higher reliability coefficients) than the other possible way of measuring the inaccuracy, viz. the mean variable error.

Results of Bisection and Trisection.

Group		Mean	σ	Coeff. of Variation	Rel. coeff. r_1	Rel. coeff. r_2
Bisection only	I	311 ± 11	129 ± 8	41 ± 2·8	·35	·52
Bisection & Trisection	V (*a*)	480 ± 14	129 ± 10	26 ± 2·2	B. ·36 T. ·35	B. ·53 T. ·52
	V (*b*)	—	—	—	B. ·44 T. ·87	B. ·93 T. ·82

(6) *Müller-Lyer Illusion.*

The adjustable apparatus, designed by Dr W. H. R. Rivers, was used to measure the size of this illusion. The length of the standard line was 75 mm., and results recorded below are also in mm. Each child of Groups II and III was tested individually by myself, being asked to make ten adjustments of the apparatus, alternately lengthening and shortening the variable line, and having the standard line alternately to the right and to the left. To obtain the reliability coefficient, the results were divided into two halves and correlated. The average deviation was taken as a size of the illusion, and the mean variation (M.V.) was also determined. Ten subjects of Group IV were tested, 10 times each, by one of the mistresses of the college. Groups V (*a*) and V (*b*) were tested by myself, but four times only.

The present test was the only one employed which involved the use of apparatus or the testing of the subjects separately.

Results of the Müller-Lyer Illusion Test.

Group	Mean		σ		Coeff. of Variation		Rel. coeff. for size of illusion only	
	Size	M. V.	Size	M. V.	Size	M. V.	r_1	r_2
II	17·3 ± ·31	3·3 ± ·14	2·8 ± ·22	1·3 ± ·10	16·2 ± 1·3	39 ± 3·4	·65	·79
III	16·8 ± 4	3·1 ± ·12	3·8 ± ·3	1·1 ± ·08	22·5 ± 1·8	36 ± 3·1	·86	·92
IV	13·7 ± 1·11	2·3 ± ·18	5·3 ± ·8	·8 ± ·13	38 ± 6·5	36 ± 6·1	·76	·86
* V (*a*)	16 ± ·67	—	3·7 ± ·48	—	23 ± 3·1	—	·57	—
* V (*b*)	—	—	—	—	—	—	·68	—

(7) *Vertical-Horizontal Illusion.*

The material for this test consisted of a set of 10 large L-shaped figures clearly printed on a very large sheet of paper, which could be folded in two. Each L had unequal arms, the shorter being 10 cms. in length, the longer 14 cms., and the vertical was alternately the shorter and the longer of the two. These papers having been distributed, it was explained by means of the black-board that the task to be performed was to mark off a part along the longer arm (estimated from the angle) such that it seemed equal to the shorter arm, the subject limiting his attention strictly to each figure in turn, and estimating *by eye only*. When all the ten figures had been marked in this way the subjects were asked *to go over them once more*, altering those which seemed too long or too short, the object of this being to make sure of the full effect of the illusion (of the existence of which, by the way, not one of the subjects tested—Groups I, II, III and IV—was aware). A fortnight later the test was repeated with other papers. The average size and the M.V. of the illusion were evaluated as in (6) above. Two objections may be made to this method of applying the test: (1) the presence of surrounding L's influenced the judgment; this was partly obviated by the way the figures were arranged on the page, and I believe the

* One test only.

influence was actually very small, each L being large enough to exclusively rivet the attention of the subject upon itself in its turn; (2) the *eyesight* of the subjects of the experiment was not previously tested. This objection is much more serious. Even if the illusion is not to be entirely explained as the effect of *astigmatism*, the latter must play an important part in determining the result. All we can say, then, is that the test measures the *balance of effect* of the various factors contributing towards the falsifying of judgments comparing horizontal and vertical distances. A somewhat remarkable result, which I do not remember to have heard or seen reported before, is that with as many as 20 measurements of each subject, quite a large proportion of the subjects show a *negative* illusion, i.e. they *under*estimate the vertical instead of overestimating it. One might retort that this is simply a case of *over-correction*, were it not for the still more remarkable fact that in the case of all the children measured, the proportion is exactly $\frac{1}{3}$, in Group I, 22 out of 66, in Group II, 13 out of 39, in Group III, 13 out of 40. In Group IV the second test has unfortunately not yet been marked; for the first test alone the proportion is $\frac{12}{56}$.

Results of Vertical-Horizontal Illusion Test.

Group	Mean		σ		Rel. coeff. r_1		Rel. coeff. r_2	
	Size	M. V.	Size	M. V.	Size	M.V.	Size	M.V.
I	25 ± 4·6	3·2 ± ·11	53 ± 3	1·3 ± ·08	·69	·43	·82	·60
II	29 ± 7·1	—	65 ± 5	—	·59	—	·74	—
III	31 ± 8·2	3·3 ± ·11	76 ± 6	1·0 ± ·08	·75	[0]	·86	—

(8) *Mechanical Memory Test.*

In this test a printed list of ten nonsense syllables was placed face downwards before each of the subjects, and at a given signal the subjects turned the papers over and applied themselves to the learning of the syllables as intently as possible. On a second signal, 2—3 minutes later, the papers were once more turned face downwards, and collected by the experimenter. The subjects were then asked to think no more about the syllables for the present. On the following day, at the same hour, blank slips of paper were distributed and the subjects were asked to write down the syllables they had learnt the previous day, so far as possible *in the right order*. As a system of marking which was found to be sufficiently satisfactory for the purpose, 2 marks were given for each syllable right and in the right order, and 1 mark for each right but in the wrong order. The time allowed for learning was 3 minutes in the case of Groups I and II, but this was found to be too long in the case of Groups III and IV, who were eventually given 2 minutes and 2½ minutes respectively. On account of the difficulty thus raised (and, in the case of Group II, for another reason) two series of results could unfortunately only be obtained from Groups I and IV.

Results of Mechanical Memory Test.

Group	Mean	σ	Coefficient of Variation	Reliability coefficient r_1	Reliability coefficient r_2
I	14·5 ± ·8	9·7 ± ·57	67 ± 5	·51	·68
*II	9·6 ± ·6	5·7 ± ·43	59 ± 6	—	—
*III	11·8 ± ·8	6·8 ± ·53	57 ± 6	—	—
IV	31·5 ± ·7	7·7 ± ·5	24 ± 2	·50	·67

* Results of one test only.

(9) *Memory for Poetry.*

This test was applied but once, and to Groups I and III only. Three verses of Hood's "Queen Mab," which apparently neither of the groups had seen or heard of before, were set to be learnt for 5 minutes, and the subjects were asked to attempt to reproduce them 24 hours later. The frequency constants were found to be as follows:

Group	Mean	σ	Coefficient of Variation
I	$20 \pm \cdot 85$	$9 \cdot 7 \pm \cdot 6$	$49 \pm 3 \cdot 7$
III	$28 \cdot 6 \pm 100$	$9 \cdot 0 \pm \cdot 75$	$31 \cdot 6 \pm 2 \cdot 9$

(10) *Combination Test.*

This was the well-known *Combinations-Methode* of Ebbinghaus, in which the subject is shown a passage of continuous prose with from one-third to one-quarter of the words replaced by blanks, and is asked to supply the missing words, or words of similar significance.

In applying this test, a thorough explanation, including black-board demonstrations and examples, was first given to the class and the papers were then distributed face downwards. At a given signal the class turned the papers over and proceeded to read the passage through carefully (*writing nothing*) with a view to grasping the general sense of the entire passage. On a second signal, three minutes later, they proceeded to fill in the blanks *in order* from the beginning, endeavouring to find in each case a word which would suit the sense both of the particular sentence in which it occurred and also of the entire passage. This second period lasted five minutes, at the end of which time the signal was given to stop. Such was the method of procedure in both applications of the test in the case of the school children, and both results were found to be quite satisfactory.

In the case of the adults the times allowed were different, being $1' + 10'$ for the first passage, and $1' + 3'$ for the second, and the reliability coefficient for Group V was found to be abnormally *low*. As this seemed to be due mainly to the unsatisfactory way in which the first test was performed (I had chosen the passage badly), the results of the second test were alone used for the purposes of correlation. In Group IV the reliability coefficient was higher, though still not very high, and the two series of results were therefore amalgamated in the usual way.

In marking the papers, words supplied by the subject were counted right if they made sense in their sentence and tolerable sense in the entire passage, and Ebbinghaus' system of values was adopted; viz.,

Each blank filled in correctly = 1 mark,
" " " " incorrectly = − 1 mark,
" " passed over = − ½ mark.

Results of Combination Test.

Group	Mean	σ	Coefficient of Variation	Reliability coefficient r_1	Reliability coefficient r_2
I	22 ± 1·3	16·3 ± ·95	74 ± 6·3	·74	·85
II	19·7 ± 1·2	11·4 ± ·87	58 ± 6	·56	·72
III	41·6 ± 1·8	17 ± 1·3	41 ± 3·5	·73	·84
IV	71·3 ± 2·1	24 ± 1·5	33 ± 3·3	·46	·63
* V (*a*)	18·5 ± ·78	7·14 ± ·56	38 ± 3·4	·22	—
V (*b*)	—	—	—	·69	·82

Measurements 11, 12, *and* 13.

Measurements 11 and 12 (marks for Drawing and Total School Marks) need no further explanation. The grading for General Intelligence was obtained from two of the schools—Groups II and III—from the former of which two separate and independent gradings, by different teachers, were provided. These independent gradings correlated with one another to the extent of ·90, which gave a reliability coefficient r_2 for the amalgamated grading = ·95 †.

* One test only (the 2nd).

† The high value of this coefficient contradicts the views of certain writers, but is borne out by a later research by Miss H. Gertrude Jones, whose results have been statistically evaluated by Karl Pearson, *Biometrika*, Vol. vii. 1910, pp. 542–548.

(14) *Association-time.*

This test was applied to certain individuals of Group V (*a* and *b*) only, and was of the following nature. A word of ordinary significance (a noun) was read out to the subjects and they were expected to write down as rapidly as possible, during the two minutes which followed, words representing the various "ideas" which passed through their mind in the time. After a short pause, another quite different word was called out and the writing repeated. Finally a third word was called out. The total number of words or phrases written down was taken as a measure of the rate of sequence of associated ideas in the subject's mind. The test was repeated a fortnight later.

This method gives fairly reliable results—for Group V*a*, $r_1 = \cdot 67$ and for Group V*b*, $r_1 = \cdot 87$—but is vitiated by the mechanical process of writing. The impurity could be eliminated by applying a simple writing test (speed) also, and then employing the formula for the "partial" correlation of three variables; it was not, however, done in the present research.

General Remarks upon the Tests.

The results tabulated in the last few pages, when tested by means of the formula for the P.E. of a difference $[\sigma_{x-y} = \sqrt{\sigma_x^2 + \sigma_y^2}$ and $\therefore$ $\text{P.E.}_{x-y} = \sqrt{\text{P.E.}_x^2 + \text{P.E.}_y^2}]$, show certain differences between group and group in respect of average ability, variability and reliability for correlation, which justify our plan of working correlation coefficients separately for the several groups, but do not seem otherwise to give many positive results of general significance and importance, such as, e.g., evidence as to the relative variability of the two sexes. A more careful and thorough examination of the tables may give cause for some qualification of the preceding statement. At any rate the individual figures are of considerable interest. The reliability coefficients, even for the single tests, are in most cases sufficiently high,—in fact much higher than I had dared to expect considering the circumstances that in all but one test the subjects were examined collectively. The less satisfactory tests, as applied in this research, seem to be those for accuracy of addition, bisection and trisection of lines, M.V. of the vertical-horizontal illusion, and, in a slighter degree, mechanical memory. The combination test, in the case of adults, was the most unsatisfactory, but in the case of the school children it gave fairly high results. The tests in which two applications gave very reliable results are the motor test, the *a n o s* test, speed of addition, the combination test (with children) and the *e r* test. The coefficients of variation are rather high, clustering about the values 20—30; in a few cases they are considerably higher.

The tests were applied during the course of the summer of 1909, and my sincere gratitude and thanks are due to the head masters, head mistresses and others, through whose kindness I was enabled to bring the research to a satisfactory conclusion.

As a fact of considerable importance, it should be added that the tests were so applied as to disturb the ordinary routine of the schools *as little as possible.*

Correlation Results.

The values of the frequency constants for the various groups of subjects show very clearly that any plan of throwing them together (as they stand) and working out coefficients from the combined series would produce a considerable amount of "spurious" correlation and make the results almost valueless. One exception, indeed, to this state of affairs was found in the case of Speed and Accuracy of Addition [4 (*a*) and 4 (*b*)] in Group I and in a group of 20 boys not otherwise included in the present research. The means and S.D.'s in these two sets of boys for these two characteristics were found to be the same, within the limits of probable error. These 86 boys were therefore taken together for this particular correlation and a correlation table was drawn up (see Appendix II and p. 59 above) whereby the value of η could be calculated as well as that of r, and the nature of the regression curve and regression line determined. To make the investigation into this particular problem of speed and accuracy in adding complete, small correlation tables were drawn up for all the other groups, separately from one another, and the value of η was calculated in each of these cases also (Appendix IV).

Apart from these cases the plan of grouping in correlation tables seemed quite unsuitable for such small numbers. The values were therefore taken as they stood, but the full product-moment formula, $\frac{S(xy)}{N\sigma_1\sigma_2}$, was employed throughout, with the single exception of Group V (*b*). Here the numbers were so very small (23) that the method of ranks $\left[\rho = 1 - \frac{6S(\nu_1 - \nu_2)^2}{N(N^2-1)},\ r = 2\sin\left(\frac{\pi}{6}\rho\right)\right]$ was considered sufficiently accurate, since nothing but a general impression of the nature of the correlation could be expected.

The following Tables give the values of the correlation coefficients between series formed by the amalgamation of the two measurements made in each test. The numbers immediately below the coefficients are the probable errors, and those in thick type are the reliability coefficients (r_2) for the amalgamated series, showing to what extent each amalgamated series would correlate with another quite similar series. In these tables, the tests are arranged according to order of magnitude of the *average* correlation of each with all the rest (within any particular group) and all coefficients smaller than their probable errors are put down as 0. Of the coefficients recorded, the total number of those > 2 P.E. is 139, and *of those* > 3 P.E. *the total is* 86. The first thing to be noticed in the groups of coefficients arranged in this way is that *not one of them shows the "hierarchical arrangement,"* and it is a very significant fact that the group which approaches it most nearly (Group I) is the group where "spurious correlation" due to heterogeneity of material was to be suspected (see p. 99). Now it will be apparent, on the slightest reflection, that any extraneous source of correlation (such as, e.g., difference of the state of *discipline* to which different members of the group had been accustomed immediately antecedent to the occasion of

GROUP I. 66 *boys* (*Elementary School*), *ages* 11—12.

	e r	a n o s	Combina-tion	Mechanical Memory	Memory for Poetry	Addition (speed)	Addition (acc.)	Motor (all letters)	M. V. of V. H. Ill.	Bisection	V. H. Ill.
e r	**·75**	·78 *·03*	·45 *·07*	·40 *·07*	·27 *·08*	·59 *·05*	·30 *·08*	·53 *·06*	– ·19 *·08*	0	0
a n o s	·78 *·03*	**·87**	·48 *·07*	·29 *·08*	·28 *·08*	·51 *·06*	·24 *·08*	·21 *·08*	– ·31 *·08*	0	·11 *·08*
Combination	·45 *·07*	·48 *·07*	**·85**	·52 *·06*	·52 *·06*	·40 *·07*	·38 *·07*	·13 *·09*	0	·15 *·08*	0
Mechanical Memory	·40 *·07*	·29 *·08*	·52 *·06*	**·68**	·49 *·07*	·27 *·08*	·31 *·08*	·14 *·08*	0	·10 *·09*	·24 *·08*
Memory for Poetry	·27 *·08*	·28 *·08*	·52 *·06*	·49 *·07*		·41 *·07*	·38 *·07*	·12 *·08*	0	·13 *·09*	·10 *·09*
Addition (speed)	·59 *·05*	·51 *·06*	·40 *·07*	·27 *·08*	·41 *·07*	**·90**	·13 *·08*	·25 *·08*	0	0	·12 *·09*
Addition (accuracy)	·30 *·08*	·24 *·08*	·38 *·07*	·31 *·08*	·38 *·07*	·13 *·08*	**·50**	0	– ·17 *·08*	·41 *·07*	·18 *·08*
Motor (all letters)	·53 *·06*	·21 *·08*	·13 *·09*	·14 *·08*	·12 *·08*	·25 *·08*	0	**·95**	·09 *·09*	0	0
M. V. of V. H. Ill.	– ·19 *·08*	– ·31 *·08*	0	0	0	0	– ·17 *·08*	·09 *·09*	**·60**	– ·22 *·08*	·21 *·08*
Bisection	0	0	·15 *·08*	·10 *·09*	·13 *·09*	0	·41 *·07*	0	– ·22 *·08*	**·52**	·12 *·09*
V. H. Ill.	0	·11 *·08*	0	·24 *·08*	·10 *·09*	·12 *·09*	·18 *·08*	0	·21 *·08*	·12 *·09*	**·82**

31 coefficients > 2 × P. E. 25 coefficients > 3 × P. E.

GROUP II. 39 *girls* (*Elementary School*), *ages* 11—12.

	School Marks	Combination	Mechanical Memory	a n o s	General Intelligence	e r	M. L. Ill.	Motor (all letters)	Addition (speed)	Drawing	V. H. Ill.	Addition (acc.)	M. V. of M. L. Ill.
School Marks		·54 *·08*	·59 *·07*	·27 *·10*	·64 *·06*	0	·16 *·11*	0	0	0	– ·15 *·11*	0	0
Combination	·54 *·08*	**·72**	·37 *·09*	0	·43 *·09*	– ·15 *·11*	0	0	– ·13 *·11*	·22 *·10*	·22 *·10*	– ·25 *·10*	0
Mechanical Memory	·59 *·07*	·37 *·09*		·20 *·10*	·55 *·08*	0	0	0	– ·13 *·11*	0	0	– ·23 *·10*	– ·16 *·11*
a n o s	·27 *·10*	0	·20 *·10*	**·91**	·13 *·11*	·80 *·04*	– ·21 *·10*	·21 *·10*	0	·13 *·11*	– ·11 *·11*	0	·15 *·11*
General Intelligence	·64 *·06*	·43 *·09*	·55 *·08*	·13 *·11*	**·95**	0	0	·13 *·11*	·10 *·11*	0	0	0	0
e r	0	– ·15 *·11*	0	·80 *·04*	0	**·79**	– ·20 *·10*	·49 *·08*	·13 *·11*	0	– ·18 *·10*	0	0
M. L. Ill.	·16 *·11*	0	0	– ·21 *·10*	0	– ·20 *·10*	**·79**	– ·21 *·10*	·12 *·11*	– ·44 *·09*	– ·21 *·10*	0	– ·21 *·10*
Motor (all letters)	0	0	0	·21 *·10*	·13 *·11*	·49 *·08*	– ·21 *·10*	**·92**	·33 *·10*	0	0	·30 *·10*	0
Addition (speed)	0	– ·13 *·11*	– ·13 *·11*	0	·10 *·11*	·13 *·11*	·12 *·11*	·33 *·10*	**·82**	– ·40 *·09*	0	·24 *·10*	0
Drawing	0	·22 *·10*	0	·13 *·11*	0	0	– ·44 *·09*	0	– ·40 *·09*		·27 *·10*	·11 *·11*	0
V. H. Ill.	– ·15 *·11*	·22 *·10*	0	– ·11 *·11*	0	– ·18 *·11*	– ·21 *·10*	0	0	·27 *·10*	**·74**	0	0
Addition (accuracy)	0	– ·25 *·10*	– ·23 *·10*	0	0	0	0	·30 *·10*	·24 *·10*	·11 *·11*	0	**·63**	0
M. V. of M. L. Ill.	0	0	– ·16 *·11*	·15 *·11*	0	0	– ·21 *·10*	0	0	0	0	0	

26 coefficients > 2 × P. E. 12 coefficients > 3 × P. E.

GROUP III. *Higher Grade School, 40 boys, ages 11—12 years.*

	School Marks	General Intelligence	Memory for Poetry	Combination	Drawing	e r	Addition (speed)	M. V. of M. L. Ill.	a n o s	M. V. of V. H. Ill.	Mech. Memory	M. L. Ill.	V. H. Ill.	Motor (all letters)	Addition (acc.)
School Marks		·78 *·04*	·60 *·07*	·60 *·07*	·51 *·08*	·30 *·10*	·28 *·10*	·28 *·10*	·17 *·10*	− ·20 *·10*	·40 *·09*	− ·20 *·10*	− ·30 *·10*	·23 *·10*	·11 *·11*
General Intelligence	·78 *·04*		·57 *·07*	·69 *·06*	·42 *·09*	·28 *·10*	·24 *·10*	·22 *·10*	·10 *·11*	− ·29 *·10*	·49 *·08*	0	− ·30 *·10*	·32 *·10*	0
Memory for Poetry	·60 *·07*	·57 *·07*		·44 *·09*	·44 *·09*	·23 *·10*	0	·52 *·08*	·14 *·10*	0	·38 *·09*	0	0	·19 *·10*	− ·11 *·11*
Combination	·60 *·07*	·69 *·06*	·44 *·09*	**·84**	·46 *·09*	0	·32 *·10*	0	·10 *·11*	− ·20 *·10*	·28 *·10*	0	− ·11 *·11*	·28 *·10*	0
Drawing	·51 *·08*	·42 *·09*	·44 *·09*	·46 *·09*		·11 *·11*	·14 *·10*	0	·19 *·10*	− ·24 *·10*	·39 *·09*	− ·19 *·10*	0	0	0
e r	·30 *·10*	·28 *·10*	·23 *·10*	0	·11 *·11*	**·86**	·35 *·10*	·26 *·10*	·74 *·05*	·26 *·10*	0	− ·31 *·10*	0	·25 *·10*	0
Addition (speed)	·28 *·10*	·24 *·10*	0	·32 *·10*	·14 *·10*	·35 *·10*	**·81**	0	·20 *·10*	− ·11 *·11*	0	− ·32 *·10*	− ·10 *·11*	·20 *·10*	·33 *·10*
M. V. of M. L. Ill.	·28 *·10*	·22 *·10*	·52 *·08*	0	0	·26 *·10*	0		·24 *·10*	·26 *·10*	·29 *·10*	0	− ·14 *·10*	− ·14 *·10*	0
a n o s	·17 *·10*	·10 *·11*	·14 *·10*	·10 *·11*	·19 *·10*	·74 *·05*	·20 *·10*	·24 *·10*	**·89**	− ·16 *·10*	0	− ·35 *·10*	0	0	− ·11 *·11*
M. V. of V. H. Ill.	− ·20 *·10*	− ·29 *·10*	0	− ·20 *·10*	− ·24 *·10*	·26 *·10*	− ·11 *·11*	·26 *·10*	− ·16 *·10*	$r_1 = 0$	0	·32 *·10*	·33 *·10*	0	·13 *·10*
Mechanical Memory	·40 *·09*	·49 *·08*	·38 *·09*	·28 *·10*	·39 *·09*	0	0	·29 *·10*	0	0		0	− ·16 *·10*	0	0
M. L. Ill.	− ·20 *·10*	0	0	0	− ·19 *·10*	− ·31 *·10*	− ·32 *·10*	0	− ·35 *·10*	·32 *·10*	0	**·92**	·29 *·10*	0	− ·11 *·11*
V. H. Ill.	− ·30 *·10*	− ·30 *·10*	0	− ·11 *·11*	0	0	− ·10 *·11*	− ·14 *·10*	0	·33 *·10*	− ·16 *·10*	·29 *·10*	**·86**	0	− ·14 *·10*
Motor (all letters)	·23 *·10*	·32 *·10*	·19 *·10*	·28 *·10*	0	·25 *·10*	·20 *·10*	− ·14 *·10*	0	0	0	0	0	**·86**	0
Addition (accuracy)	·11 *·11*	0	− ·11 *·11*	0	0	0	·33 *·10*	0	− ·11 *·11*	·13 *·11*	0	− ·11 *·11*	− ·14 *·10*	0	$r_1 = 0$

51 coefficients > 2 × P. E. 28 coefficients > 3 × P. E.

GROUP IV. (56 *women students.*) [*Provisional and incomplete table of coefficients.*]

	Combination	Addition (accuracy)	Addition (speed)	Mechanical Memory
Combination	**·63**	·53 *·06*	·34 *·08*	·31 *·08*
Addition (accuracy)	·53 *·06*	**·45**	·43 *·07*	·20 *·09*
Addition (speed)	·34 *·08*	·43 *·07*	**·96**	·18 *·09*
Mechanical Memory	·31 *·08*	·20 *·09*	·18 *·09*	**·67**

Variability coefficient for speed of addition:

$$r_{\substack{sp._1 + sp._2 \\ sp._1 \sim sp._2}} = 0{\cdot}33 \pm {\cdot}08.$$

Variability coefficient for accuracy of addition:

$$r_{\substack{acc._1 + acc._2 \\ acc._1 \sim acc._2}} = -\ 0{\cdot}66 \pm {\cdot}05.$$

[*Correlations of M. L. Ill.* (10 *cases only**.)

$r_{\substack{\text{M.L. Ill.} \\ \text{Comb.}}} = 0,$ $\qquad r_{\substack{\text{M.V. of M.L. Ill.} \\ \text{Comb.}}} = {\cdot}28 \pm {\cdot}19,$

$r_{\substack{\text{M.L. Ill.} \\ \text{Mech. Mem.}}} = {\cdot}35 \pm {\cdot}18,$ $\qquad r_{\substack{\text{M.V. of M.L. Ill.} \\ \text{Mech. Memory}}} = -\ {\cdot}21 \pm {\cdot}20,$

$r_{\substack{\text{M.L. Ill.} \\ \text{M.V. of M.L. Ill.}}} = -\ {\cdot}57 \pm {\cdot}14.$]

* Owing to the smallness of the sample, the P.E.'s (calculated by the ordinary formula) are too low, and the results cannot be regarded as anything more than very rough approximations.

GROUP V (*a*). (35 *men students.*)

	e r	Association time	Addition (acc.)	Combination	Addition (speed)	M. L. Ill.	Bisection + Trisection
e r	**·97**	− ·18 *·11*	− ·26 *·10*	·19 *·11*	0	·42 *·09*	− ·24 *·10*
Association time	− ·18 *·11*	**·87**	·39 *·09*	·33 *·10*	·37 *·09*	0	0
Addition (accuracy)	− ·26 *·10*	·39 *·09*	**·36**	− ·16 *·11*	·38 *·09*	0	0
Combination	·19 *·11*	·33 *·10*	− ·16 *·11*	**·22***	·19 *·11*	− ·24 *·10*	0
Addition (speed)	0	·37 *·09*	·38 *·09*	·19 *·11*	**·97**	0	·13 *·11*
M. L. Ill.	·42 *·09*	0	0	− ·24 *·10*	0	**·57**	− ·29 *·10*
Bisection + Trisection	− ·24 *·10*	0	0	0	·13 *·11*	·29 *·10*	**B = ·53** **T = ·52**

9 coefficients > 2 × P. E. 5 coefficients > 3 × P. E.

* The second test only was used in this case for correlation with other tests since the low correlation between the two was almost certainly due to the unsatisfactory nature of the first. The value ·22 is, then, r_1.

[GROUP V (*b*). (23 *women students**.)

{Method of ranks used $\rho = 1 - \frac{6S(d^2)}{N(N^2-1)}$,

$$r = 2\sin\left(\frac{\pi}{6}\rho\right).\}$$

	M. L. Ill.	Association time	Combination	Bisection + Trisection	Addition (acc.)	Addition (speed)	e r
M. L. Ill.	**·68**	– ·78 *·05*	– ·90 *·03*	– ·84 *·04*	– ·53 *·10*	– ·66 *·08*	– ·42 *·11*
Association time	– ·78 *·05*	**·67**	·58 *·09*	·53 *·10*	·58 *·09*	·35 *·12*	·43 *·11*
Combination	– ·90 *·03*	·58 *·09*	**·82**	0	·28 *·13*	·13 *·14*	·52 *·10*
Bisection + Trisection	– ·84 *·04*	·53 *·10*	0	B = **·61** T = **·93**	·29 *·13*	·40 *·12*	0
Addition (accuracy)	– ·53 *·10*	·58 *·09*	·28 *·13*	·29 *·13*	**·74**	·35 *·12*	0
Addition (speed)	– ·66 *·08*	·35 *·12*	·13 *·14*	·40 *·12*	·35 *·12*	**·99**	0
e r	– ·42 *·11*	·43 *·11*	·52 *·10*	0	0	0	**·58**]

16 coefficients > 2 × P. E. 12 coefficients > 3 × P. E.

* These results are recorded as avowedly rough approximations only, owing to the smallness of the sample.

applying the tests) the influence of which is in a *constant* direction but varies in amount from test to test according to the varying degrees to which the individual tests are susceptible to its influence, *must* tend to produce the hierarchical arrangement, and unless counteracted by other more potent tendencies, *will* do so. In fact, it will be the "central factor" supposed to be indicated by such a form of arrangement of coefficients. Spurious correlation of this nature might arise from the use of unfamiliar apparatus in the tests, or from the novelty of the tests, or in many other ways. The form of procedure adopted in the present research was specially devised to reduce such extraneous sources of correlation to a minimum, being assimilated as far as possible to the ordinary class-work of the school.

A definite solution of the question of the existence or non-existence of one central mental ability is yet to be sought. It can only be obtained by the use of much larger random samples than those hitherto employed, since the probable errors must be small compared with the coefficients, if precise inferences are to be drawn from the latter, and in the case of small samples, this condition is satisfied only for *large* correlation coefficients, which when obtained are often merely the result of selecting tests which measure closely similar mental abilities. In all results hitherto quoted in support of ultimate identity of general intelligence and general sensory discrimination, the correlations contributed by the latter are so small compared with their P.E.'s that nothing definite can be inferred from them. On the other hand, in such cases it is easy to propound hypotheses since the bounds of possibility are nowhere limited in any unambiguous way.

Certain sub-groups can be chosen from the tables so as to show a hierarchical arrangement, e.g. *Group III*, School Marks, General Intelligence, Mechanical Memory, and Combination. In fact the general law,—so far as the results allow of the confident formulation of any law at all—would seem to be that the tests fall into a number of such sub-groups, correlating highly among themselves, but not at all highly with members of other sub-groups, though an individual member of one sub-group may, exceptionally, correlate highly with an individual of another sub-group. In order to bring out these relations more clearly and also to show the relation of the main groups with one another, the table on pages 122 and 123 was drawn up. The results there to some extent explain themselves. Differences in correlation between the two sexes, though well marked, do not seem to follow any general law. On the whole, the correlations are lower* in the girls than in the boys, higher in the women than in the men. A striking feature is the fairly large number of instances of *negative* correlation in the girls corresponding to positive correlations in the boys.

Comparing relative *order* of tests in the tables for Groups I, II, and III, I get:

I	Rank	II	Rank	III	Rank
e r	1	Combination	3	Combination	3
a n o s	2	Mech. Memory	4	e r	1
Combination	3	a n o s	2	Addition (sp.)	5
Mech. Memory	4	e r	1	a n o s	2
Addition (sp.)	5	Motor	7	Mech. Memory	4
Addition (acc.)	6	Addition (sp.)	5	V. H. Ill.	8
Motor	7	V. H. Ill.	8	Motor	7
V. H. Ill.	8	Addition (acc.)	6	Addition (acc.)	6

The rank-correlations here are

$$\rho_{\text{I, II}} = \cdot 66, \quad \rho_{\text{I, III}} = \cdot 74, \quad \rho_{\text{II, III}} = \cdot 55.$$

As might be expected, the two groups of boys correspond more closely with one another than either does with the group of girls. (Compare value deduced from Mr Burt's results, p. 93 above.)

The method of multiple or "partial" correlation (see pp. 52, 63, 64 above) may be very advantageously employed to investigate the way in which the correlation coefficients are related to one another. Thus, taking the *three* variables, er, anos, and motor, and using the formula

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r^2_{13}}\sqrt{1 - r^2_{23}}},$$

I get the following values for the partial correlations in the first three groups:

	Group I	Group II	Group III
e r, a n o s	$\cdot 80 \pm \cdot 03$	$\cdot 82 \pm \cdot 04$	$\cdot 76 \pm \cdot 04$
e r, motor	$\cdot 59 \pm \cdot 05$	$\cdot 55 \pm \cdot 08$	$\cdot 37 \pm \cdot 09$
a n o s, motor	$- \cdot 38 \pm \cdot 07$	$- \cdot 35 \pm \cdot 09$	$- \cdot 28 \pm \cdot 10$

* This may be due to greater variability.

Coefficients < 3 P.E. *put in Square Brackets.*

Tests	Groups					
	I Boys (elementary)	II Girls (elementary)	III Boys (higher grade)	IV Women students	V (*a*) Men students	V (*b*) Women students
Combination Test and						
School Marks	—	·54 ± ·08	·60 ± ·07	—	—	—
General Intelligence	—	·43 ± ·09	·69 ± ·06	—	—	—
Drawing	—	[·22 ± ·10]	·46 ± ·09	—	—	—
Mechanical Memory	·52 ± ·06	·37 ± ·09	[·28 ± ·10]	·31 ± ·08	—	—
Memory for Poetry	·52 ± ·06	—	·44 ± ·09	—	—	—
Letters a n o s	·48 ± ·07	[0]	[·10 ± ·11]	—	—	—
Letters e r	·45 ± ·07	[– ·15 ± ·11]	[0]	—	[·19 ± ·11]	·52 ± ·10
Addition (speed)	·40 ± ·07	[– ·13 ± ·11]	·32 ± ·10	·34 ± ·08	[·19 ± ·11]	[·13 ± ·14]
Addition (accuracy)	·38 ± ·07	[– ·25 ± ·10]	[0]	·53 ± ·06	[– ·16 ± ·11]	[·28 ± ·13]
Association time	—	—	—	—	·33 ± ·10	·58 ± ·09
Mechanical Memory and						
School Marks	—	·59 ± ·07	·40 ± ·09	—	—	—
General Intelligence	—	·55 ± ·08	·40 ± ·08	—	—	—
Memory for Poetry	·49 ± ·07	—	·38 ± ·09	—	—	—
Letters a n o s	·29 ± ·08	[·20 ± ·10]	[0]	—	—	—
Letters e r	·40 ± ·07	[0]	[0]	—	—	—
Addition (speed)	·27 ± ·08	[– ·13 ± ·11]	[0]	[·18 ± ·09]	—	—
Addition (accuracy)	·31 ± ·08	[– ·23 ± ·10]	[0]	[·20 ± ·09]	—	—
Letters a n o s and						
Letters e r	·78 ± ·03	·80 ± ·04	·74 ± ·05	—	(·57)	(·56)
Addition (speed)	·51 ± ·06	[0]	[·20 ± ·10]	—	—	—
Addition (accuracy)	·24 ± ·08	[0]	[– ·11 ± ·11]	—	—	—
Letters e r and						
Addition (speed)	·59 ± ·05	[·13 ± ·11]	·35 ± ·10	—	[0]	[0]
Addition (accuracy)	·30 ± ·08	[0]	[0]	—	[– ·26 ± ·10]	[0]
Addition (speed) and						
Addition (accuracy)	[·13 ± ·08]	[·24 ± ·10]	·33 ± ·10	·43 ± ·07	·38 ± ·09	·35 ± ·12

Motor (*all letters*) and						
Letters e r	·53 ± ·06	·49 ± ·08	[·25 ± ·10]	—	—	—
[Letters a n o s	[·21 ± ·08]	[·21 ± ·10]	[0]	—	—	—]
Addition (speed)	·25 ± ·08	·33 ± ·10	[·20 ± ·10]	—	—	—
Addition (accuracy)	[0]	·30 ± ·10	[0]	—	—	—
Müller-Lyer Ill. and						
Drawing	—	– ·44 ± ·09	[– ·19 ± ·10]	—	—	—
V. H. Illusion	—	[– ·21 ± ·10]	[·29 ± ·10]	—	—	—
M. V. of M. L. Ill.	—	[– ·21 ± ·10]	[0]	(– ·57 ± ·14)	—	—
M. V. of V. H. Ill.	—	—	·32 ± ·10	—	—	—
Letters e r	—	[– ·20 ± – ·10]	– ·31 ± ·10	—	·42 ± ·09	– ·42 ± ·11
Letters a n o s	—	[– ·21 ± ·10]	– ·35 ± ·10	—	—	—
Addition (speed)	—	[·12 ± ·11]	– ·32 ± ·10	—	[0]	– ·66 ± ·08
Drawing and						
V. H. Illusion	—	[·27 ± ·10]	[0]	—	—	—
Addition (speed)	—	– ·40 ± ·09	[·14 ± ·10]	—	—	—
Vert. Hor. Ill. and						
M. V. of V. H. Ill.	[·21 ± ·08]	—	·33 ± ·10	—	—	—
School Marks	—	[– ·15 ± ·11]	– ·30 ± ·10	—	—	—
General Intelligence	—	[0]	– ·30 ± ·10	—	—	—
Combination	[0]	[·22 ± ·10]	[– ·11 ± ·11]	—	—	—
M. V. of V. H. Ill. and						
Letters a n o s	– ·31 ± ·08	—	[– ·16 ± ·10]	—	—	—
M. V. of M. L. Ill. and						
Memory for Poetry	—	—	·52 ± ·08	—	—	—
Mechanical Memory	—	[– ·16 ± ·11]	[·29 ± ·10]	[– ·21 ± ·20]	—	—
Drawing and						
School Marks	—	[0]	·51 ± ·08	—	—	—
General Intelligence	—	[0]	·42 ± ·09	—	—	—
Memory for Poetry	—	—	·44 ± ·09	—	—	—
Mechanical Memory	—	[0]	·39 ± ·09	—	—	—

Thus the original positive correlation between anos and motor is due entirely to the correlation of each with er. For "er constant," the correlation is large but negative, in all three cases.

Employing the four variables, School Marks, General Intelligence, Combination and Mechanical Memory (Group II) and using the formula

$$r_{12.34} = \frac{r_{12}(1 - r^2_{34}) - r_{13}(r_{23} - r_{24}r_{34}) - r_{14}(r_{24} - r_{23}r_{34})}{\sqrt{1 - r^2_{13} - r^2_{14} - r^2_{34} + 2r_{13}r_{14}r_{34}}\sqrt{1 - r^2_{23} - r^2_{34} - r^2_{24} + 2r_{23}r_{34}r_{24}}},$$

(see p. 64), I find the correlation between Combination and General Intelligence, assuming constant ability in Mechanical Memory, and as shown by School Marks

$$= 0{\cdot}11 \pm 0{\cdot}11.$$

The "entire" coefficients $= 0{\cdot}43 \pm {\cdot}09$. It is interesting to note that there is still correlation, though very slight, after the effect of memory and school industry and ability is eliminated.

A more thoroughgoing application of the method of partial correlation was made in the case of Group I, for the tests er, anos, Combination and Mechanical Memory. The full working, according to Mr G. U. Yule's method, is given in Appendix III. The results were:

Tests Correlated	Entire Coefficients	Partial Coefficients
er, anos	0·78 ± ·03	0·73 ± ·04
er, Combination	0·45 ± ·07	0·01 ± ·09
er, Mechanical Memory	0·40 ± ·07	0·26 ± ·08
anos, Combination	0·48 ± ·07	0·27 ± ·08
anos, Mechanical Memory	0·29 ± ·08	−0·15 ± ·08
Combination, Mechanical Memory	0·52 ± ·06	0·44 ± ·07

The *regression equation* for the calculation of ability in combination from abilities in the other three tests is

$$x_c = {\cdot}002x_{er} + {\cdot}099x_{anos} + {\cdot}703x_{mech.\ memory},$$

x in each case denoting deviation from mean ability.

For a larger number of variables than four, the arithmetic of partial correlation becomes extremely lengthy and rather fatiguing, but there can be no doubt whatever that this is the one sound method to adopt in investigating the relations between coefficients. Working with a large number of variables is only satisfactory when the original coefficients are *large* compared with the P.E.'s, since, as a rule (not universally *) the partial coefficient is smaller than the entire coefficient. The formula for the P.E. is the same in both cases.

It must also be remembered that the partial correlation formulae assume *linear regression*. In the case of the correlation of Addition (speed) and Addition (accuracy) throughout the Groups, and in the case of Combination and Mechanical Memory in one Group (Group I), the "correlation ratio" (η) was determined and compared with the corresponding value for r. The results are given in Appendix IV, together with the regression curves. It will be observed that although the criterion $\left(\frac{\sqrt{N}}{\cdot 67449} \cdot \frac{\sqrt{\eta^2 - r^2}}{2}\right)$ works out $< 2{\cdot}5$, there seems to be a small but well-defined deviation from linearity in all the cases of Addition (sp. and acc.), (see, too, the regression curve in Appendix II). η is in all cases fairly high, and the general form of curve is approximately identical in all. With increase of speed, the accuracy may at first decrease (Groups III and VI), but after that, it in all cases increases, then decreases, then increases once more, and in cases III and V (a) it finally decreases.

Variability coefficients $\left(r_{\substack{x_1+x_2 \\ x_1 \sim x_2}}\right)$ were obtained for Speed of Addition and Accuracy of Addition in Group IV, and were found to be

$$0{\cdot}33 \pm \cdot 08 \text{ and } -0{\cdot}66 \pm \cdot 05, \text{ respectively.}$$

A full analysis of the entire data in this and other additional ways must be reserved for a later occasion.

As mentioned above, pp. 84 and 85, tests were made of the applicability of Spearman's correction formula, with results which precluded the use of the formula. The question of correlation of errors of measurement with the true values of the variates and with one another has been made the subject of a separate piece of research by the present writer. The results of this investigation will be published shortly.

* See, e.g., p. 52 above.

In the case of the Vertical-Horizontal Illusion Test, it is perhaps of interest to note that if subjects showing a *negative* value of the illusion are excluded the value of the correlation between this Illusion Test and the Combination Test becomes positive and appreciable—in Group IV

$$r_{\substack{\text{Combination}\\ \text{V. H. Ill.}}} = 0{\cdot}24 \pm 0{\cdot}09,$$

(one test only of V.H. Ill.), in Groups I + III, it = 0·26 ± 0·08 (this latter value probably includes some "spurious" correlation: see tables pp. 107, 110). If negative values are included, we have the results:

$r_{\substack{\text{Combination}\\ \text{V. H. Ill.}}}$:

Group I, 0; Group II, 0·22 ± ·10; Group III, – ·11 ± ·11.

$r_{\substack{\text{Gen. Intelligence}\\ \text{V. H. Ill.}}}$:

Group I, 0; Group II, 0; Group III, – ·30 ± ·10.

Conclusions.

1. The correlation between different psychical abilities is not very close. Few coefficients are greater than ·6.

2. The size of the correlation coefficient varies greatly from one group of subjects to another. This shows how great is the danger of spurious correlation, due to heterogeneity of material, in psychical measurements.

3. The *Combinations-Methode* of Ebbinghaus is a good measure of intellectual ability. It correlates with "general intelligence" almost as closely as "scholastic intelligence" (school marks) does.

4. Mechanical memory correlates fairly closely with intelligence.

5. Drawing also correlates closely with intelligence in the case of higher grade school boys. On the other hand, for girls of the same age the correlation is nil.

6. The susceptibility to the Müller-Lyer Illusion is not at all closely correlated with that to the Vertical-Horizontal Illusion. This indicates that the factors involved are for the most part different in the two cases.

7. The "mean variation" of the Vertical-Horizontal Illusion is positively related to the size of the illusion (Group I, ·21 ± ·08, Group III, ·33 ± ·10).

8. In the case of the Müller-Lyer Illusion, there is a zero or slightly negative correlation between M.V. and size.

9. The two optical illusions correlate differently with drawing, the Müller-Lyer negatively (Group II, − ·44 ± ·09, Group III, − ·19 ± ·10), the Vertical-Horizontal positively or not at all (Group II, ·27 ± ·10, Group III, 0).

10. The mean variations of the illusions show a zero or very slight negative correlation with other measures of mental ability.

11. The correlation between speed and accuracy of mental performance is slightly non-linear.

12. Correlations may be very low even within a set of mental tests which appear to measure closely related mental abilities, and this when the reliability coefficients are high. Thus, the correlation between erasing the letters *a, n, o, s* and erasing all the letters, is less than three times the probable error in every group tested. For "erasing the letters *e, r*" constant, the partial correlation is negative and numerically greater than 3 P.E. in each group.

13. In *homogeneous* groups of subjects there is no positive evidence of the existence of one "central factor" to which the correlations between the individual mental abilities may be regarded as due.

14. There is in several cases a pronounced correlation between ability and variability. This may be either positive or negative. It is, e.g., positive for speed in adding figures, negative for accuracy.

15. There is definite evidence that variability, as measured by the coefficient of variation, is greater in boys than in girls. Differences in correlation, though sometimes appreciable, seem to follow no well-defined rule.

CHAPTER IV

THE SIGNIFICANCE OF CORRELATION IN PSYCHOLOGY

The full significance of correlation in Psychology is to be found in the general theory of multiple correlation, of which the correlation of two variables, measured by the correlation coefficient, is only a particular case. Karl Pearson's classic deduction of the product-moment formula* proceeds on these lines. We will therefore attempt to state the general problem, so that we may have something tangible on which to base our discussion.

Let there be n physical (or psychical) quantities

$$x_1, x_2, x_3 \ldots x_n,$$

which we observe and find correlated together; and let us assume that there are m causes

$$\epsilon_1, \epsilon_2, \epsilon_3 \ldots \epsilon_m,$$

which may be physical, biological, or psychological, but which are absolutely independent of one another. We may further assume that the x's are given by functions of the ϵ's, thus:

$$x_1 = f_1(\epsilon_1, \epsilon_2, \epsilon_3 \ldots \epsilon_m),$$

$$x_2 = f_2(\epsilon_1, \epsilon_2, \epsilon_3 \ldots \epsilon_m),$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$x_n = f_n(\epsilon_1, \epsilon_2, \epsilon_3 \ldots \epsilon_m).$$

All the ϵ's need not occur in each x-function.

* "Regression, Heredity and Panmixia," *Phil. Trans.* A. Vol. CXCVIII. pp. 443–459. The following proof is based partly upon a lecture demonstration given by Prof. Pearson.

We further assume that $m > n$, since otherwise the m ϵ's can be eliminated from these n equations.

Then our general problem is this: supposing the independent variables, ϵ, each follow a Gaussian distribution, what will be the frequency surface connecting the x's?

Assume that the variations of ϵ are small relative to their mean values. In this case we may expand the functions and retain only the first powers, in other words we may make the x's linear functions of the ϵ's.

Let ξ's be deviations of x's from mean values,

and η's „ „ „ ϵ's „ „ „

Then

$$\left.\begin{array}{l} \xi_1 = a_{11}\eta_1 + a_{12}\eta_2 + \ldots + a_{1m}\eta_m \\ \xi_2 = a_{21}\eta_1 + a_{22}\eta_2 + \ldots + a_{2m}\eta_m \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \xi_n = a_{n1}\eta_1 + a_{n2}\eta_2 + \ldots + a_{nm}\eta_m \end{array}\right\} \ldots\ldots\ldots\ldots \text{(i)}.$$

By hypothesis, the η's are independent and follow the Gaussian distribution. Hence

the chance of a complex of variables lying between

$$\eta_1, \eta_1 + \delta\eta_1 = c_1 \times e^{-\frac{1}{2}\frac{\eta_1^2}{s_1^2}} d\eta_1,$$

the chance of a complex of variables lying between

$$\eta_2, \eta_2 + \delta\eta_2 = c_2 \times e^{-\frac{1}{2}\frac{\eta_2^2}{s_2^2}} d\eta_2,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

the chance of a complex of variables lying between

$$\eta_m, \eta_m + \delta\eta_m = c_m \times e^{-\frac{1}{2}\frac{\eta_m^2}{s_m}} d\eta_m.$$

Whole chance = product of these.

We can now eliminate one of the η_1's in the following manner.

Let us solve the n equations (i), finding the value of

$$\eta_1, \eta_2 \ldots \eta_n$$

in terms of

$$\begin{array}{l} \xi_1 - a_{1n+1}\eta_{n+1} - a_{1n+2}\eta_{n+2} - \ldots - a_{1m}\eta_m, \\ \xi_2 - a_{2n+1}\eta_{n+1} - a_{2n+2}\eta_{n+2} - \ldots - a_{2m}\eta_m, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \xi_n - a_{nn+1}\eta_{n+1} - a_{nn+2}\eta_{n+2} - \ldots - a_{nm}\eta_m. \end{array}$$

Since these equations are linear, our chance of a complex of variables lying between

$$\xi_1 \text{ and } \xi_1+\delta\xi_1,\ \xi_2 \text{ and } \xi_2+\delta\xi_2,\ \dots\ \xi_n \text{ and } \xi_n+\delta\xi_n,$$
$$\eta_{n+1} \text{ and } \eta_{n+1}+\delta\eta_{n+1},\ \dots\ \eta_m \text{ and } \eta_m+\delta\eta_m$$

will be of the form

$$\text{Const.} \times e^{-\frac{1}{2}q_1-\frac{1}{2}q_2-\frac{1}{2}q_3} \times \delta x_1 \delta x_2 \dots \delta x_n \times \delta\eta_{n+1} \dots \delta\eta_m \dots \text{(ii)}.$$

In this expression q_1 is a function of the ξ's only, q_2 is a function of η_n to η_m only, and q_3 is a function of the products of ξ's into η's.

In order to find the chance that our complex of variables lies between ξ_1, $\xi_1+\delta\xi_1$ up to ξ_n, $\xi_n+\delta\xi_n$, we must integrate expression (ii) for all values of η_{n+1} up to η_m between the limits $+\infty$ and $-\infty$.

Integrating with regard to η_{n+1}, we find that the expression keeps its general form, only η_{n+1} has disappeared. If now the process be repeated with η_{n+2}, the integration will again leave the expression of the same form except that η_{n+2} will have disappeared. Continuing this for an indefinite number of η integrations, all the η's will at length disappear, and we shall be left with a quadratic function of ξ's only. In other words, the chance of a complex of variables lying between

$$\xi_1 \text{ and } \xi_1+\delta\xi_1,\ \xi_2 \text{ and } \xi_2+\delta\xi_2,\ \dots\ \xi_n \text{ and } \xi_n+\delta\xi_n$$
$$=\text{Const.} \times e^{-\frac{1}{2}(g_{11}\xi_1^2+g_{22}\xi_2^2+\dots+g_{nn}\xi_n^2+2g_{12}\xi_1\xi_2+2g_{13}\xi_1\xi_3+\dots)} d\xi_1 d\xi_2 \dots d\xi_n \dots \text{(iii)}.$$

This expression represents the general form of the frequency surface connecting the x's.

It may be noted that the frequency will be constant when

$$g_{11}\xi_1^2+g_{22}\xi_2^2+\dots+g_{nn}\xi_n^2+2g_{12}\xi_1\xi_2+2g_{13}\xi_1\xi_3+\dots=\text{const.}\dots\text{(iv)}.$$

This represents an ellipsoid in n-dimensional space.

Taking the simplest case of two variables, expression (iii) becomes

$$Ce^{-\frac{1}{2}(g_1\xi_1^2+g_2\xi_2^2+2g_{12}\xi_1\xi_2)} d\xi_1 d\xi_2 \dots\dots\dots\dots\dots \text{(v)}.$$

Integrating (v) for all values of ξ_1 from $-\infty$ to $+\infty$, we get the Gaussian curve of ξ_2 variation, so that

$$\frac{1}{2\sigma_1^2}=g_2\left(1-\frac{g_{12}^2}{g_1g_2}\right).$$

Similarly, integrating for all values of ξ_2,

$$\frac{1}{2\sigma_2^2} = g_1\left(1 - \frac{g_{12}^2}{g_1 g_2}\right).$$

If now we integrate for all values of ξ_1 and ξ_2, we get the total frequency, N:

$$N = \frac{C\pi}{\sqrt{g_1 g_2 - g_{12}^2}}.$$

Writing x for ξ_1, y for ξ_2, and r for $-\frac{g_{12}}{g_1 g_2}$, we have as the equation to the correlation surface

$$z = \frac{N}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} e^{-\frac{1}{2}\frac{1}{1-r^2}\left(\frac{x^2}{\sigma_1^2} - \frac{2rxy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right)} \quad \text{......(vi).}$$

r in expression (vi) is the coefficient of correlation. That it is a measure of correlation is apparent from the fact that with no correlation the compound probability for two variables, whose separate variations are represented by $e^{-\frac{1}{2}\frac{x^2}{\sigma_1^2}}$ and $e^{-\frac{1}{2}\frac{y^2}{\sigma_2^2}}$, would be proportional to $e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_1^2} + \frac{y^2}{\sigma_2^2}\right)}$. The term $\frac{2rxy}{\sigma_1\sigma_2}$ in the index marks the difference between the two expressions.

It can be shown that $S(xy)$, or $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} z\,xy\,dx\,dy$, is equal to $Nr\sigma_1\sigma_2$.

$$\therefore \quad r = \frac{S(xy)}{N\sigma_1\sigma_2} \quad \text{....................(vii).}$$

The correlation surface represented by equation (vi) is in three dimensions of space*, and is such that vertical sections parallel to the two axes of coordinates are all normal curves and horizontal sections (i.e. the "contour lines") are similar and similarly situated ellipses. With respect to these ellipses the two regression lines are conjugates to the axes of x and y respectively.

Other properties of the surface are:

(1) The means of parallel arrays all lie on a straight line, i.e. the regression is linear.

(2) The standard deviations of all parallel arrays are equal.

* See Fig. 29 in G. Udny Yule's *Theory of Statistics*, p. 166.

It should be noted that although normal distribution of the two variates involves linear regression, the converse is not always true. For this reason the proof of the product-moment formula in Chap. I is more general than that given in the present chapter; it assumes linear regression, but is quite independent of the form of distribution. Similarly, the formulae for partial regression follow from the proof of the present chapter, but they can also be obtained more generally on the assumption merely of linear regression, independent of form of distribution (see Chap. I, pp. 59, 63).

By means of the formulae for multiple or partial correlation we are able to deal with more than two variables, and to determine the relation between any two in independence* of the influence of all the rest. Hence in psychology, where the problem of *analysis* is one of the most important, the theory of partial correlation is of the utmost value. By its means the mental abilities of *a homogeneous group* of individuals can be analysed out and arranged in order of closeness of inter-relation to one another; in other words the abilities of the group may be exhibited as a system, quantitatively determined in its various parts. We must remember, however, that the formulae are only applicable in cases of *linear* regression. Evidence for this should therefore be forthcoming in all future researches aiming at completeness; r should be tested by η.

Logicians remind us that Mill's "Method of Concomitant Variations" is to be regarded not as a method of scientific *proof* but as a fruitful source of suggestions for scientific *hypothesis*. So it is with correlation. A correlation coefficient is a statement of probabilities. It does not prove anything. It merely suggests a hypothesis as regards causation within a particular sphere. The investigator must go beyond the coefficient to the facts themselves for his scientific hypothesis. In the light of this the coefficient gains new significance and, if carefully and accurately calculated, is now of absolute worth as a

* It is important to understand exactly in what sense this statement is correct. We do not *eliminate* the other variables when we employ the formula for partial correlation; we only make them constant, i.e. we eliminate the effect of their variation.

quantitative measure of a concomitance of variation *explained* by a reference to more ultimate factors. For reasons already set out in earlier chapters, a *true* correlation coefficient is not easy to obtain in psychology. Hence it is probable that much of the statistical work done by psychologists during the next few years will be limited to an effort to obtain more reliable values for their correlation coefficients. At the present time coefficients obtained by different people and even by the same people on different occasions and when working with different groups of subjects show a great lack of agreement with one another. So far as this is due to differences in the "irrelevant" conditions of the investigations a careful statement and analysis of these conditions and a comparison of those in one investigation with those in another are likely to explain the discrepancies and so turn them to good account. From the nature of the case, absolute agreement is not to be expected. If the discrepancies stimulate the investigator to go behind the coefficients to the facts themselves, the cause of science will be better served than if a closer agreement had been originally attained.

It is well to bear constantly in mind that a correlation between two mental capacities does not always and necessarily indicate identity or partial identity of certain part-capacities; it may be due to identity or partial identity of environment (taken in its widest sense). Special teaching may produce a correlation between two school-subjects which would not be found under ordinary conditions. Again, response to a common environment may have produced, in the course of evolution, a correlation between two mental functions introspectively quite dissimilar to one another.

Partial correlation will carry us far in the unravelling of the tangled complex of mental ability. It will also serve to give quantitative precision to results when classified in the light of an explanatory hypothesis. For the production of the hypothesis itself we must look elsewhere, viz. to psychological analysis and psychological insight.

APPENDIX I*

Fechner's Fundamental Table.

n	$t=h\delta$	n	$t=h\delta$	n	$t=h\delta$
0·50	0·0000	0·67	0·3111	0·84	0·7031*
0·51	0·0177	0·68	0·3307	0·85	0·7329
0·52	0·0355	0·69	0·3506	0·86	0·7639
0·53	0·0532	0·70	0·3708	0·87	0·7965
0·54	0·0710	0·71	0·3913	0·88	0·8308
0·55	0·0888*	0·72	0·4121	0·89	0·8673
0·56	0·1067*	0·73	0·4333	0·90	0·9062
0·57	0·1247	0·74	0·4549	0·91	0·9480*
0·58	0·1427*	0·75	0·4769	0·92	0·9935*
0·59	0·1609	0·76	0·4994	0·93	1·0435*
0·60	0·1792*	0·77	0·5224	0·94	1·0993*
0·61	0·1975	0·78	0·5460	0·95	1·1630*
0·62	0·2160	0·79	0·5702	0·96	1·2380*
0·63	0·2346*	0·80	0·5951	0·97	1·3300*
0·64	0·2535	0·81	0·6208	0·98	1·4520*
0·65	0·2725	0·82	0·6743	0·99	1·6450
0·66	0·2916*	0·83	0·6747	1·00	∞

If n is $< 0·5$, look in the Table not for n but for $1 - n$, and take t negative. Thus the t for $n = 0·25$ is $- 0·4769$.

* The first two tables in this appendix are quoted from E. B. Titchener, *Experimental Psychology*, Vol. II. *Student's Manual*, pp. 99 and 101. The "starred" values in the first table are the corrected values given by H. Bruns (*Wahrscheinlichkeitsrechnung und Kollektivmasslehre*, 1906), and quoted by F. M. Urban (*Psychological Review*, Vol. XVII. 1910, p. 251). Sheppard's Tables of the Probability Integral (*Biometrika*, Vol. II.) are more generally useful than this one of Fechner's.

Müller's Table of Coefficients of Weights.

n	w''	n	w''	n	w''
0·50	1·000	0·67	0·824	0·84	0·373
0·51	0·999	0·68	0·803	0·85	0·342
0·52	0·997	0·69	0·782	0·86	0·311
0·53	0·994	0·70	0·760	0·87	0·281
0·54	0·990	0·71	0·737	0·88	0·251
0·55	0·984	0·72	0·712	0·89	0·222
0·56	0·977	0·73	0·687	0·90	0·193
0·57	0·969	0·74	0·661	0·91	0·166
0·58	0·960	0·75	0·634	0·92	0·139
0·59	0·950	0·76	0·606	0·93	0·114
0·60	0·938	0·77	0·578	0·94	0·089
0·61	0·925	0·78	0·550	0·95	0·067
0·62	0·911	0·79	0·521	0·96	0·047
0·63	0·896	0·80	0·492	0·97	0·029
0·64	0·880	0·81	0·463	0·98	0·014
0·65	0·862	0·82	0·433	0·99	0·004
0·66	0·843	0·83	0·403		

The weight of an n which is $< 0·5$ is the same as the weight of an n which exceeds 0·5 by the same amount. Thus the weights of $n = 0·25$ and of $n = 0·75$ are both alike $= 0·634$.

*Improved Table of Weights**.

n	w	n	w	n	w
0·50	1·000	0·67	0·932	0·84	0·694
0·51	1·000	0·68	0·923	0·85	0·670
0·52	0·999	0·69	0·914	0·86	0·646
0·53	0·998	0·70	0·904	0·87	0·621
0·54	0·996	0·71	0·894	0·88	0·595
0·55	0·995	0·72	0·883	0·89	0·567
0·56	0·992	0·73	0·871	0·90	0·538
0·57	0·989	0·74	0·859	0·91	0·506
0·58	0·985	0·75	0·846	0·92	0·472
0·59	0·981	0·76	0·832	0·93	0·435
0·60	0·977	0·77	0·818	0·94	0·396
0·61	0·972	0·78	0·803	0·95	0·352
0·62	0·967	0·79	0·787	0·96	0·304
0·63	0·960	0·80	0·770	0·97	0·249
0·64	0·954	0·81	0·752	0·98	0·187
0·65	0·947	0·82	0·733	0·99	0·112
0·66	0·940	0·83	0·713	1·00	0·000

* The values in this table are given by the formula

$$w = \frac{1}{2\pi} \cdot \frac{Ne^{-2t^2}}{n(1-n)},$$

where N = no. of experiments made.

The table is quoted from F. M. Urban: "The Method of Constant Stimuli and its Generalizations," *Psychological Review*, Vol. XVII. 1910, p. 253. See also "Die psychophysischen Massmethoden als Grundlagen empirischer Messungen," by the same author, *Archiv f. d. ges. Psychologie*, Vols. XV. and XVI.

APPENDIX II

GROUPS I AND VII (*a small group of* 20 *elementary schoolboys aged* 11—12, *and showing in this case the same mean and* S.D. *as Group I*).

86 boys aged 11—12 years.

Correlation between speed and accuracy in the addition of groups of 10 single digits. Two tests, of 5 minutes' duration each.

Correlation Table.

⟶ Speed of Addition (x)

↓ Accuracy of Addition (y)

	100—140	*140—180*	*180—220*	*220—260*	*260—300*	*300—340*	*340—380*	*380—420*	Totals (n_y)	Means ($\bar{x}_y$)
50—110	—	3 *11*	0·5 *5·5*	0·5 *0*	1 *5·5*	—	—	1 *22*	6	− ·250
110—125	1 *9*	1 *6*	1 *3*	1 *0*	—	—	—	—	4	− 1·500
125—140	—	2 *4*	0·5 *2*	0·5 *0*	—	—	—	—	3	− 1·500
140—155	0·5 *3*	2·5 *2*	1 *1*	1·5 *0*	1·5 *1*	2 *2*	—	—	9	− ·222
155—170	—	3 *0*	2 *0*	4 *0*	5 *0*	3 *0*	—	1 *0*	18	+ ·389
170—185	1 *3*	4·5 *2*	4·5 *1*	5·5 *0*	5·5 *1*	3 *2*	0·5 *3*	0·5 *4*	25	− ·060
185—200	1 *6*	2·5 *4*	5 *2*	5 *0*	2·5 *2*	2·5 *4*	2 *6*	0·5 *8*	21	+ ·166
Totals (n_x)	3·5	18·5	14·5	18	15·5	10·5	2·5	3	$N=86$	$\bar{x}=-\cdot0698$ $\sigma_x^2=2\cdot796$
Means ($\bar{y}_x$)	− ·143	− 1·000	+ ·569	+ ·403	+ ·226	+ ·571	+ 1·800	− 1·333	$\bar{y}=\cdot081$ $\sigma_y^2=4\cdot034$	

Note. The figures in italics immediately beneath the frequency values within the correlation table are for the calculation of $S(xy)$. The row and column with zeros correspond to the arbitrary means from which the true means, S.D.'s and $S(xy)$ are calculated.

Frequency	x'	Frequency $\times x'$	Frequency $\times x'^2$	Frequency	y'	Frequency $\times y'$	Frequency $\times y'^2$
3·5	−3	10·5	31·5	6	−5·5	33	181·5
18·5	−2	37	74	4	−3	12	36
14·5	−1	14·5	14·5	3	−2	6	12
18	0	−62	—	9	−1	9	9
15·5	1	15·5	15·5	18	0	−60	—
10·5	2	21	42	25	1	25	25
2·5	3	7·5	22·5	21	2	42	84
3	4	12	48	86 = N		+67	347·5
86 = N		+56	248			+7	
		−6					

$$d_1 = -\frac{6}{86} = -0·06977, \qquad d_2 = \frac{7}{86} = 0·0814,$$

$$\sigma_1^2 = \frac{248}{86} - (d_1)^2 - \frac{1}{12}\,^*, \qquad \sigma_2^2 = \frac{347·5}{86} - d_2^2\,\dagger,$$

$$= 2·7955, \qquad = 4·0341,$$

$$\therefore\ \sigma_1 = 1·67. \qquad \therefore\ \sigma_2 = 2·013.$$

* Sheppard's correction. Remember that $\sigma^2 = \mu_2$, and $d = \nu'_1$; see above p. 39.

† Sheppard's correction cannot be used here, since the units of the sub-groups are not equal and there is not high contact at the ends of the frequencies.

Frequencies	$x'y'$ *	Total of frequencies (f)	$f \times x'y'$	
1+5·5−4·5−1·5	1	0·5	+0·5	
2·5+0·5+2·5+3−4·5−5−2	2	−3		−6
1+0·5+0·5−1	3	1	3	
2+2·5+0·5−2·5	4	2·5	10	
0·5−1	5·5	−0·5		2·75
1+2−1	6	2	12	
0·5	8	0·5	4	
1	9	1	9	
3	11	3	33	
−1	22	−1		22
			71·5 − 30·75	
			$S(x'y') = 40·75$	

$$S(xy) = S(x'y') - Nd_1d_2\dagger$$

$$= 40·75 + 86 \times ·0057$$

$$= 41·24,$$

$$r = \frac{S(xy)}{N\sigma_1\sigma_2}$$

$$= \frac{41·24}{86 \times 1·67 \times 2·013}$$

$$= 0·143,$$

$$\text{P.E.} = ·67449 \,.\, \frac{1 - r^2}{\sqrt{N}}$$

$$= 0·071.$$

$$\therefore\ r_{\substack{\text{speed of addition}\\ \text{acc. of addition}}} = \mathbf{0·14 \pm 0·07}.$$

* The figures in italics in the correlation table.

† $S(x'y') = S(x+d_1)(y+d_2)$

$= S(xy) + d_1 S(y) + d_2 S(x) + Nd_1d_2$

$= S(xy) + Nd_1d_2$, since $S(x) = S(y) = 0$.

$\therefore\ S(xy) = S(x'y') - Nd_1d_2$.

$\bar{y}_x - \bar{y}$	$(\bar{y}_x - \bar{y})^2 \times n_x$
$-\ \cdot224$	$\cdot050176 \times 3\cdot5 = \cdot175616$
$-1\cdot081$	$1\cdot168561 \times 18\cdot5 = 21\cdot618379$
$+\ \cdot488$	$\cdot238144 \times 14\cdot5 = 3\cdot453088$
$+\ \cdot322$	$\cdot103684 \times 18 = 1\cdot866312$
$+\ \cdot145$	$\cdot021025 \times 15\cdot5 = \cdot325888$
$+\ \cdot490$	$\cdot2401 \times 10\cdot5 = 2\cdot52105$
$+1\cdot719$	$2\cdot954961 \times 2\cdot5 = 7\cdot387403$
$-1\cdot414$	$1\cdot999396 \times 3 = 5\cdot998188$
	$S\{n_x(\bar{y}_x - \bar{y})^2\} = 43\cdot345924$

$$\eta^2 = \frac{S\{n_x(\bar{y}_x - \bar{y})^2\}}{N\sigma_y^2}$$

$$= \frac{43\cdot345924}{86 \times 4\cdot034}$$

$$= \cdot1249,$$

$$\therefore \eta = 0\cdot353.$$

$$\text{P.E.} = \cdot67449 \,.\, \frac{1-\eta^2}{\sqrt{N}}$$

$$= 0\cdot064,$$

$$\therefore \eta_{\substack{\text{speed of addition}\\ \text{acc. of addition}}} = 0\cdot35 \pm 0\cdot06.$$

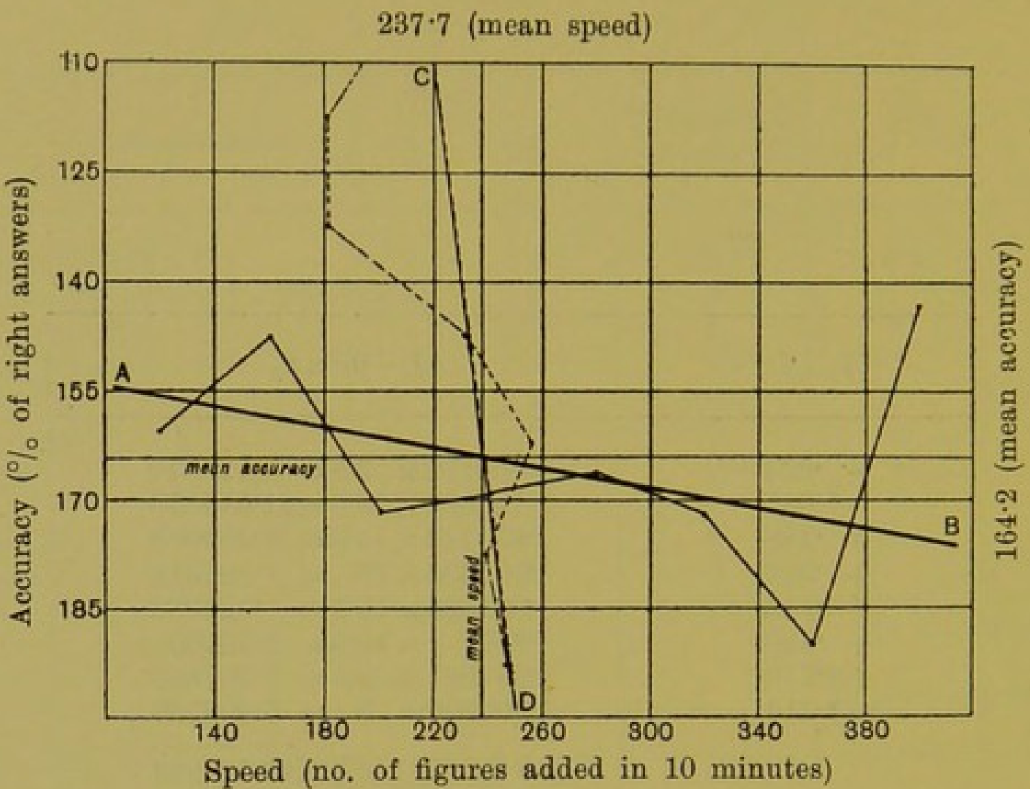


Fig. 16.

Calculating η from the means of the y-arrays, we have

$$\eta^2 = \frac{S\{n_y(\bar{x}_y - \bar{x})^2\}}{N\sigma_x^2}$$

$$= \frac{19{\cdot}68101}{86 \times 2{\cdot}796}$$

$$= {\cdot}0819.$$

$$\therefore\ \eta = 0{\cdot}29 \pm 0{\cdot}07.$$

To test the value of η obtained from the means of the x-arrays, for linear regression:

$$\frac{\sqrt{N}}{{\cdot}67449} \cdot \tfrac{1}{2}\sqrt{\eta^2 - r^2}\,{*} = \frac{{\cdot}323}{2 \times {\cdot}07273}$$

$$= 2{\cdot}22, \text{ i.e. } < 2{\cdot}5.$$

Hence regression may be considered to be linear.

Regression coefficients:

$$b_{12} = r\frac{\sigma_1}{\sigma_2} = {\cdot}118,$$

$$b_{21} = r\frac{\sigma_2}{\sigma_1} = {\cdot}172.$$

* See p. 59.

Equations to regression lines are

$$x - \bar{x} = b_{12}(y - \bar{y}),$$

and

$$y - \bar{y} = b_{21}(x - \bar{x}).$$

∴ equation to regression line AB is

$$y - 164{\cdot}2 = {\cdot}172\,(x - 237{\cdot}7),$$

i.e.

$$y = {\cdot}172x + 123{\cdot}316 \text{ ....(i).}$$

Similarly equation to line CD is

$$x - 237{\cdot}7 = {\cdot}118\,(y - 164{\cdot}2),$$

i.e.

$$x = {\cdot}118y + 218{\cdot}324 \text{ ....(ii).}$$

Equation (i) gives the most probable value of y associated with a given value of x, with a standard error

$$\sigma_2\sqrt{1 - r^2}, \text{ i.e. } 1{\cdot}991.$$

Similarly, *mutatis mutandis*, with equation (ii).

APPENDIX III

Example of Multiple Correlation.

GROUP I. (*Boys, ages* 11—12; $n = 66$.)

1. Crossing through two letters (e and r).
2. Crossing through four letters (a, n, o, s).
3. Combination test.
4. Mechanical memory test.

Formula for multiple correlation:

$$r_{12.34\ldots n} = \frac{r_{12.34\ldots n-1} - r_{1n.34\ldots n-1}\, r_{2n.34\ldots n-1}}{(1 - r^2_{1n.34\ldots n-1})^{\frac{1}{2}}\,(1 - r^2_{2n.34\ldots n-1})^{\frac{1}{2}}}.$$

For four variables this becomes:

$$r_{12.34} = \frac{r_{12.3} - r_{14.3}\, r_{24.3}}{(1 - r^2_{14.3})^{\frac{1}{2}}\,(1 - r^2_{24.3})^{\frac{1}{2}}},$$

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{(1 - r^2_{13})^{\frac{1}{2}}\,(1 - r^2_{23})^{\frac{1}{2}}}.$$

TABLE I.

Correlation coefficient		$\log (1 - r^2)$
12	0·78	$\bar{1}$·59284
13	0·45	$\bar{1}$·90173
14	0·40	$\bar{1}$·92428
23	0·48	$\bar{1}$·88627
24	0·29	$\bar{1}$·96185
34	0·52	$\bar{1}$·86308

TABLE II.

Correlation coefficient (zero order)		Product term of numerator	Numerator	Correlation coefficient (first order)		log $(1-r^2)$
12	0·78	0·2160	0·5640	12·3	0·7199	$\bar{1}$·68281
13	0·45	0·3744	0·0756	13·2	0·1377	$\bar{1}$·99169
23	0·48	0·3510	0·1290	23·1	0·2308	$\bar{1}$·97623
12	0·78	0·1160	0·6640	12·4	0·7570	$\bar{1}$·63038
14	0·40	0·2262	0·1738	14·2	0·2902	$\bar{1}$·96179
24	0·29	0·3120	−0·0220	24·1	−0·0386	$\bar{1}$·99348
13	0·45	0·2080	0·2420	13·4	0·3091	$\bar{1}$·95639
14	0·40	0·2340	0·1660	14·3	0·2176	$\bar{1}$·97893
34	0·52	0·1800	0·3400	34·1	0·4154	$\bar{1}$·91774
23	0·48	0·1508	0·3292	23·4	0·4027	$\bar{1}$·92316
24	0·29	0·2496	0·0404	24·3	0·0539	$\bar{1}$·99873
34	0·52	0·1392	0·3808	34·2	0·4536	$\bar{1}$·89996

TABLE III.

Correlation coefficient (first order)		Product term of numerator	Numerator	Correlation coefficient (second order)		log $(1-r^2)$
12·4	0·7570	0·1245	0·6325	12·34	0·727	$\bar{1}$·67345
13·4	0·3091	0·3048	0·0043	13·24	0·007	$\bar{1}$·99998
23·4	0·4027	0·2340	0·1687	23·14	0·272	$\bar{1}$·96662
12·3	0·7199	0·0117	0·7082	12·34	0·727	—
14·3	0·2176	0·0388	0·1788	14·23	0·258	$\bar{1}$·97009
24·3	0·0539	0·1567	−0·1028	24·13	−0·152	$\bar{1}$·98985
13·2	0·1377	0·1316	0·0061	13·24	0·007	—
14·2	0·2902	0·0625	0·2277	14·23	0·258	—
34·2	0·4536	0·0400	0·4136	34·12	0·436	$\bar{1}$·90843
23·1	C·2308	−0·0160	0·2468	23·14	0·272	—
24·1	−0·0386	0·0959	−0·1345	24·13	−0·152	—
34·1	0·4154	−0·0089	0·4243	34·12	0·436	—

The regression equation between changes in intelligence (as measured by the combination test) and changes in the other three variables is

$$x_3 = b_{31.24}\,x_1 + b_{32.14}\,x_2 + b_{34.12}\,x_4.$$

Calculation of regression coefficients.

$$\sigma_1 = 68{\cdot}19, \quad \sigma_2 = 58{\cdot}43, \quad \sigma_3 = 16{\cdot}34, \quad \sigma_4 = 9{\cdot}70,$$

$$\sigma^2_{1.23\ldots n} = \sigma_1^2 (1 - r^2_{12})(1 - r^2_{13.2})(1 - r^2_{14.23}) \ldots (1 - r^2_{1n.23\ldots n-1}).$$

$$\therefore \sigma_{3.24} = \sigma_3 (1 - r^2_{32})^{\frac{1}{2}} (1 - r^2_{34.2})^{\frac{1}{2}}$$
$$= 12{\cdot}77.$$

Similarly $\sigma_{1.24} = 40{\cdot}83$, $\sigma_{3.14} = 13{\cdot}27$, $\sigma_{2.14} = 36{\cdot}27$, $\sigma_{3.12} = 14{\cdot}23$, $\sigma_{4.12} = 8{\cdot}82$.

Again, $$b_{12.34\ldots n} = r_{12.34\ldots n} \frac{\sigma_{1.34\ldots n}}{\sigma_{2.34\ldots n}}.$$

$$\therefore b_{31.24} = r_{31.24} \frac{\sigma_{3.24}}{\sigma_{1.24}} = {\cdot}002.$$

Similarly, $b_{32.14} = {\cdot}099$, $b_{34.12} = {\cdot}703$.

Hence regression equation is

$$\mathbf{x_3 = {\cdot}002x_1 + {\cdot}099x_2 + {\cdot}703x_4}.$$

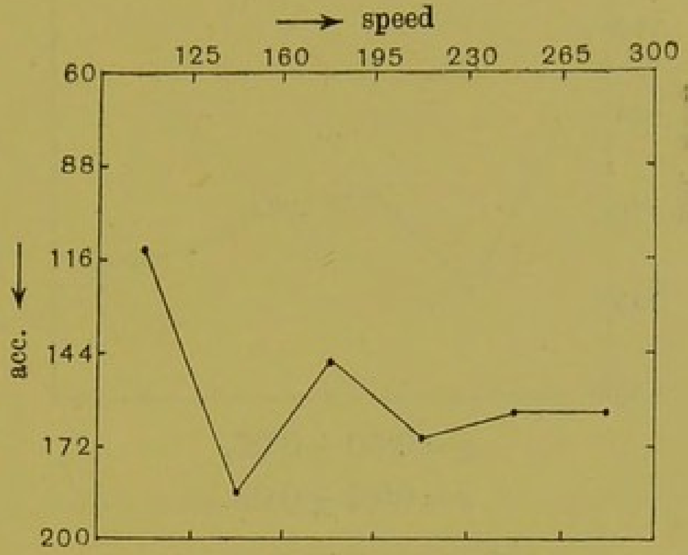
The standard error ($\sigma_{3.124}$) made in estimating x_3 from x_1, x_2, and x_4 by this equation

$$= \sigma_3 (1 - r^2_{31})^{\frac{1}{2}} (1 - r^2_{32.1})^{\frac{1}{2}} (1 - r^2_{34.12})^{\frac{1}{2}}$$
$$= \mathbf{12{\cdot}78}.$$

APPENDIX IV

Regression Curves for Speed and Accuracy of Addition.

GROUP II.

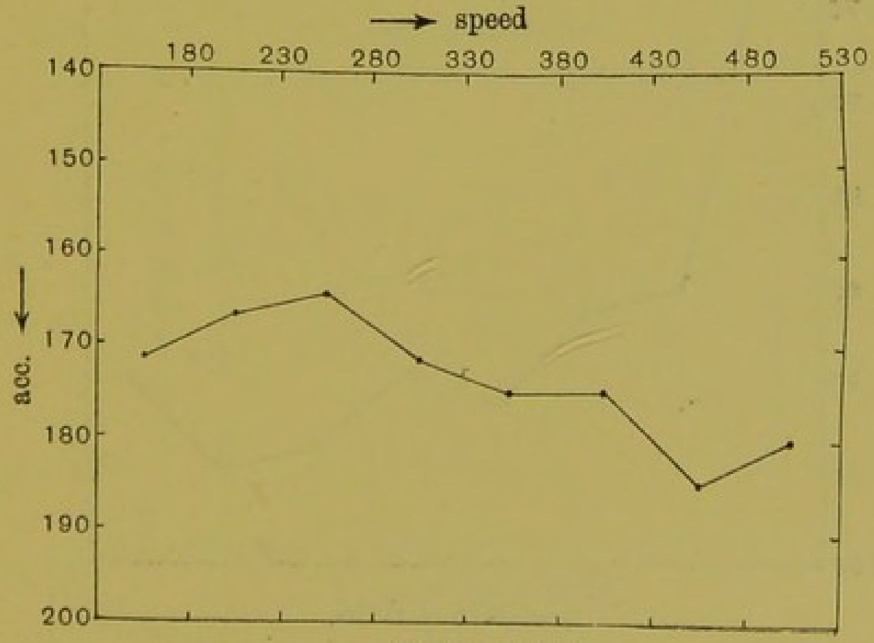


$$\eta = 0{\cdot}524 \pm 0{\cdot}079,$$

$$r = 0{\cdot}24 \pm 0{\cdot}101,$$

$$\frac{\sqrt{N}}{{\cdot}67449} \cdot \frac{\sqrt{\eta^2 - r^2}}{2} = 2{\cdot}16.$$

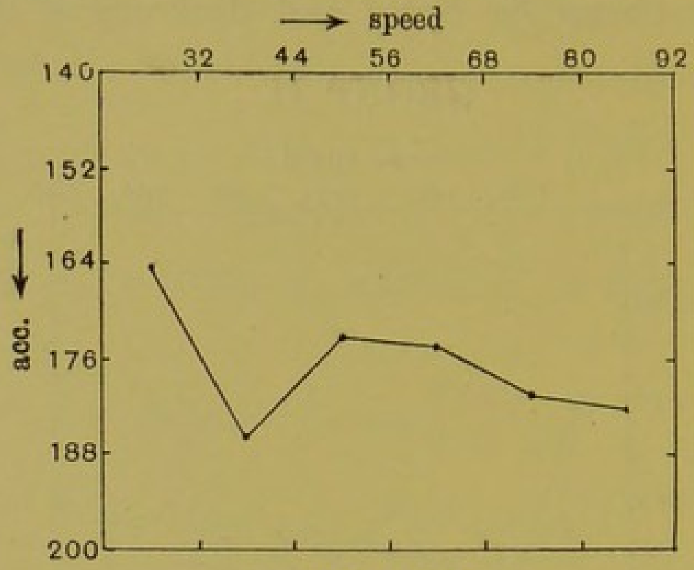
GROUP III.



$$\eta = 0{\cdot}38 \pm 0{\cdot}09,$$

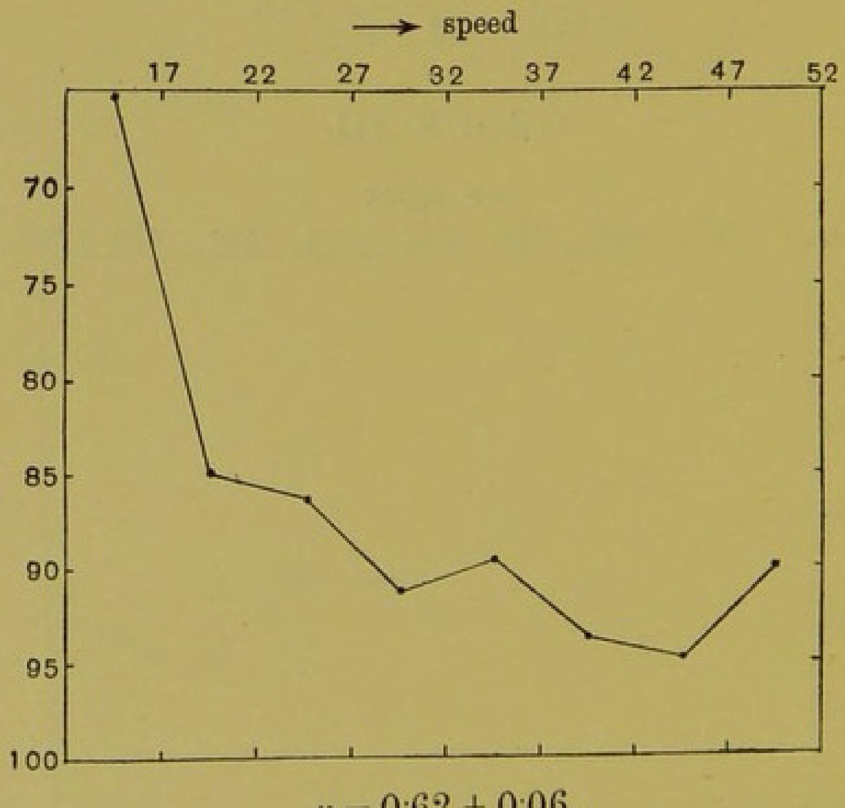
$$r = 0{\cdot}33 \pm 0{\cdot}09.$$

GROUP IV.



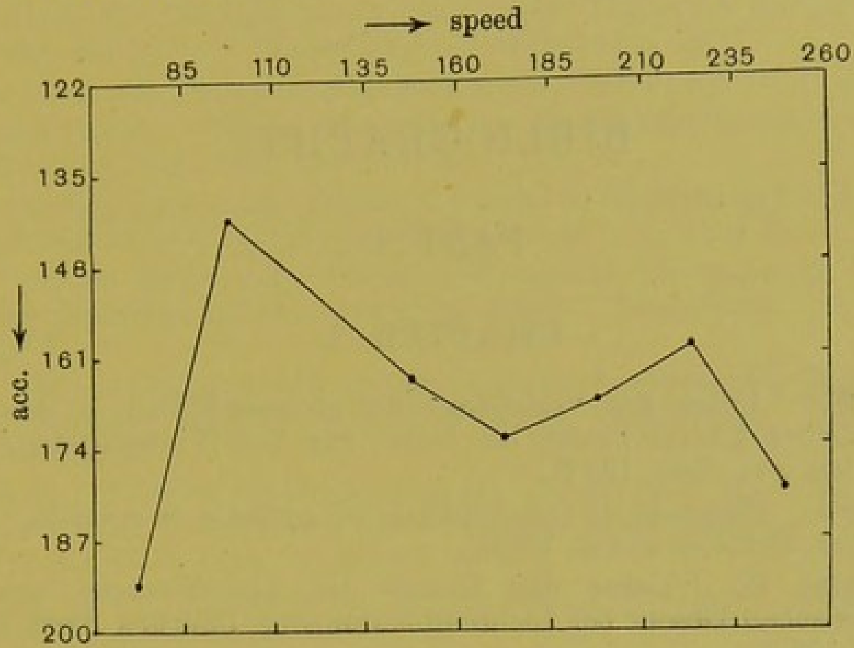
$$\eta = 0{\cdot}60 \pm 0{\cdot}06,$$
$$r = 0{\cdot}43 \pm 0{\cdot}08,$$
$$\frac{\sqrt{N}}{{\cdot}67449} \cdot \tfrac{1}{2}\sqrt{\eta^2 - r^2} = 2{\cdot}28.$$

GROUP V (*a*) (*enlarged*, $n = 51$).



$$\eta = 0{\cdot}62 \pm 0{\cdot}06,$$
$$r = 0{\cdot}38 \pm 0{\cdot}10,$$
$$\frac{\sqrt{N}}{{\cdot}67449} \cdot \tfrac{1}{2}\sqrt{\eta^2 - r^2} = 2{\cdot}26.$$

GROUP VI (*additional group*, 40 *boys*, *ages* 11—12).

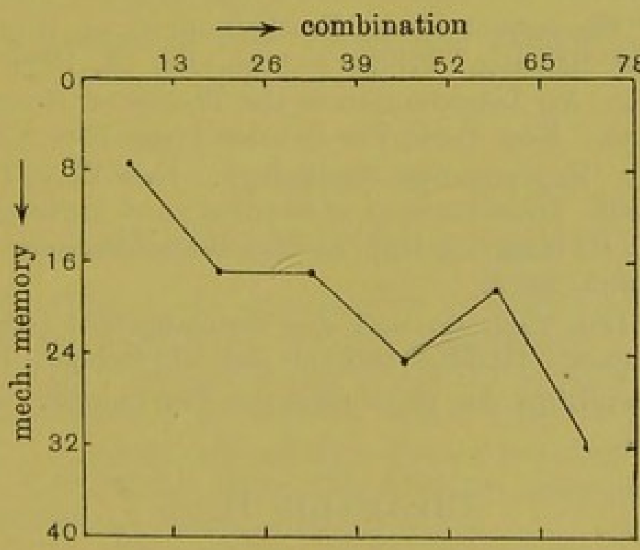


$$\eta = 0{\cdot}38 \pm 0{\cdot}09,$$

$$r = 0{\cdot}09 \pm 0{\cdot}10,$$

$$\frac{\sqrt{N}}{{\cdot}67449} \cdot \tfrac{1}{2}\sqrt{\eta^2 - r^2} = 1{\cdot}7.$$

Regression Curve for Combination and Mechanical Memory,
GROUP I.



$$\eta = 0{\cdot}63 \pm 0{\cdot}05,$$

$$r = 0{\cdot}52 \pm 0{\cdot}06,$$

$$\frac{\sqrt{N}}{{\cdot}67449} \cdot \tfrac{1}{2}\sqrt{\eta^2 - r^2} = 2{\cdot}13.$$

BIBLIOGRAPHY

PART I

CHAPTER I

AMENT, W. "Ueber das Verhältniss der ebenmerklichen zu den übermerklichen Unterschieden bei Licht- und Schall-intensitäten," Phil. Studien, XVI, 1900, 135 ff.

DELBŒUF, J. Éléments de psychophysique, générale et spéciale. Paris, Germer, Baillière et Cie, 1883.

EBBINGHAUS, H. "Ueber den Grund der Abweichungen von dem Weber'schen Gesetz bei Lichtempfindungen," Pflüger's Arch. Bd. 54, 1889, 113 ff.

—— Grundzüge der Psychologie. Leipzig, Veit and Co. 1905, §§ 6, 44–46.

FECHNER, G. T. Elemente der Psychophysik [1860]. 2 vols. Leipzig, Breitkopf and Härtel, 1889.

FRÖBES, J. "Ein Beitrag über die sogenannten Vergleichungen übermerklicher Empfindungsunterschiede," Zeitschr. für Psychol. XXXVI, 1904, 241 ff., 344 ff.

JAMES, W. Principles of Psychology. London, Macmillan and Co. 1901, Vol. I, Ch. XIII.

MYERS, C. S. A Textbook of Experimental Psychology. Cambridge University Press, 2nd edition, 1911, Ch. XIX.

STEINACH, E. "Electromotorische Erscheinungen an Hautsinnesnerven bei adäquater Reizung," Pflüger's Arch. Bd. 63, 1896, 495 ff.

THORNDIKE, E. L. An Introduction to the Theory of Mental and Social Measurements. New York, The Science Press, 1904, Chs. I and II.

TITCHENER, E. B. Experimental Psychology. New York, The Macmillan Company, 1905. Introductions of Student's and Instructor's Manuals.

WALLER, A. D. "Points relating to the Weber-Fechner Law," Brain, Vol. XVIII, 1895, 200 ff.

WEBER, E. H. Der Tastsinn und das Gemeingefühl [1846]. Offprint from R. Wagner's Handwörterbuch der Physiologie, 1851.

WUNDT, W. Grundzüge der physiologische Psychologie. Leipzig, 1902, Vol. I, 493 ff.

CHAPTER II

BRUNS, H. Wahrscheinlichkeitsrechnung und Kollektivmasslehre, Leipzig, 1906.

EBBINGHAUS, H. Grundzüge der Psychologie, 1905, § 6.

ELDERTON, W. PALIN. Frequency Curves and Correlation. London, C. and E. Layton, 1906, Chs. I–IV.

MERRIMAN, M. A Textbook of the Method of Least Squares. New York, John Wiley and Sons, 1907, 6 ff.

MÜLLER, G. E. Die Gesichtspunkte und die Tatsachen der psychophysischen Methodik. Wiesbaden, J. F. Bergmann, 1904.

MYERS, C. S. A Textbook of Experimental Psychology, 1911, Ch. XV.

PEARSON, KARL. "Skew Variation in Homogeneous Material," Phil. Trans. A, Vol. 186, 1895, 343 ff.

—— "On the Systematic Fitting of Curves to Observations and Measurements," Biometrika, Vol. I, 265 ff., Vol. II, 1 ff., 1901–1903.

—— "On an Elementary Proof of Sheppard's Formulae for correcting Raw Moments and on other allied Points," Biometrika, Vol. III, 1904, 308 ff.

—— "On the Curves which are most suitable for describing the frequency of Random Samples of a Population," Biometrika, Vol. V, 1906, 172–175.

RHIND, A. "Tables for Facilitating the Computation of Probable Errors of the Chief Constants of Skew Frequency-distributions," Biometrika, Vol. VII, 1909–1910, 127 ff. and 386 ff.

SANFORD, E. C. A Course in Experimental Psychology. Boston, D. C. Heath and Co. 1903, Ch. VIII.

SHEPPARD, W. F. "On the Calculation of the most Probable Values of Frequency Constants, for Data arranged according to Equidistant Divisions of a Scale," Proc. Lond. Math. Soc. Vol. XXIX, 353 ff.

SPEARMAN, C. "The Method of 'Right and Wrong Cases' ('Constant Stimuli') without Gauss's Formulae," Brit. Journ. of Psychology, Vol. II, 1908, 227 ff.

TITCHENER, E. B. Experimental Psychology, 1905, Vol. II.

URBAN, F. M. "Die psychophysischen Methoden als Grundlagen empirischer Messungen," Archiv f. d. ges. Psychologie, Vols. XV and XVI.

—— The Application of Statistical Methods to the Problems of Psychophysics, Philadelphia, 1908.

—— "On the Method of Just Perceptible Differences," Psychological Review, Vol. XIV, 1907, 244 ff.

PART II

INTRODUCTION AND CHAPTER I

BLAKEMAN, J. "On Tests for Linearity of Regression in Frequency-distributions," Biometrika, Vol. IV, 1905, 332 ff.

BLAKEMAN, J. AND PEARSON, KARL. "On the Probable Error of Mean-Square Contingency," Biometrika, Vol. V, 1906, 191 ff.

BRAVAIS, A. "Analyse mathématique sur les probabilités des erreurs de situation d'un point," Acad. des Sciences, Mémoires présentés par divers savants, IIe Série, IX, 1846, 255 ff.

DARBISHIRE, A. D. "Some Tables for Illustrating Statistical Correlation," Memoirs and Proceedings of the Manchester Literary and Philosophical Society, Vol. LI, Pt. III, 1906–1907.

EVERITT, P. F. "Tables of the Tetrachoric Functions for Fourfold Correlation Tables," Biometrika, Vol. VII, 1910, 437 ff.

GIBSON, WINIFRED. "Tables for Facilitating the Computation of Probable Errors," Biometrika, Vol. IV, 1905, 385 ff.

HERON, D. "An Abac to determine the Probable Errors of Correlation Coefficients," Biometrika, Vol. VII, 1910, 411.

MILL, J. S. A System of Logic, Bk. III, Ch. VIII, § 6.

PEARSON, KARL. "Regression, Heredity and Panmixia," Phil. Trans. Roy. Soc. Series A, Vol. 187, 1896, 253 ff.

—— "On a form of Spurious Correlation which may arise when Indices are used in the Measurement of Organs," Proc. Roy. Soc. Vol. LX, 1897, 489 ff.

—— "On the Theory of Contingency and its Relation to Association and Normal Correlation," Drapers' Company Research Memoirs, Biometric Series I, 1904, Dulau and Co., London.

—— "On the Theory of Skew-Correlation and Non-Linear Regression," Drapers' Company Research Memoirs, Biometric Series II, 1905.

—— "On Further Methods of Determining Correlation," Drapers' Company Research Memoirs, Biometric Series IV, 1907.

—— "On a New Method of Determining Correlation between a Measured Character *A* and a Character *B*, of which only the Percentage of Cases wherein *B* exceeds (or falls short of) a given Intensity is recorded for each grade of *A*," Biometrika, Vol. VII, 1909, 96 ff.

—— "On a New Method of Determining Correlation, when one Variable is given by Alternative and the other by Multiple Categories," Biometrika, Vol. VII, 1910, 248 ff.

SCHUSTER, E. AND ELDERTON, E. M. "The Inheritance of Psychical Characters," being further Statistical Treatment of Material Collected and Analysed by Messrs G. Heymans and E. Wiersma, Biometrika, Vol. V, 1906–1907, 460 ff.

SHEPPARD, W. F. "New Tables of the Probability Integral," Biometrika, Vol. II, 1903, 174 ff.

SPEARMAN, C. "Measurement of Association between Two Things," American Journal of Psychology, Vol. XV, 1904.

—— "'Footrule' for Measuring Correlation," Brit. Journ. of Psychology, Vol. II, 1906.

—— "Correlation Calculated from Faulty Data," Brit. Journ. of Psychology, Vol. III, 1910.

"STUDENT." "The Probable Error of a Coefficient of Correlation," Biometrika, Vol. VI, 1908, 302 ff.

WELDON, W. F. R. "Inheritance in Animals and Plants," in Lectures on the Method of Science, edited by T. B. Strong, Oxford, Clarendon Press, 1906, 100 ff.

YULE, G. U. "On the Theory of Correlation for any Number of Variables, treated by a New System of Notation," Proc. Roy. Soc. Vol. 79 A, 1907, 182 ff.

—— An Introduction to the Theory of Statistics. London, Charles Griffin and Co., Ltd., 1911.

CHAPTERS II AND III

AIKINS, H. A. AND THORNDIKE, E. L. "Correlations in the Perceptive and Associative Processes," Psychological Review, Vol. IX, 1902, 374 ff.

BETZ, W. "Ueber Korrelation," Beiheft III zur Zeitschrift für angewandte Psychologie und psychologische Sammelforschung, 1911 (in the press).

BONSER, F. G. "The Reasoning Ability of Children of the Fourth, Fifth, and Sixth School Grades." New York, Teachers' College, Columbia University, 1910.

BROWN, WILLIAM. "Some Experimental Results in Correlation," Comptes Rendus du VIme Congrès International de Psychologie, Genève, Août, 1909. Librairie Kündig, Genève, 1910, 571 ff.

—— "An Objective Study of Mathematical Intelligence," Biometrika, Vol. VII, April 1910, 352 ff.

—— "The Psychologist in the Secondary School," Journal of Phil., Psychol. and Sci. Methods, Vol. VII, Jan. 1910.

—— "Note on a Quantitative Analysis of Mathematical Intelligence," Op. cit. Vol. VII, Sept. 1910.

—— "Some Experimental Results in the Correlation of Mental Abilities," Brit. Journ. of Psychology, Vol. III, Oct. 1910.

BURT, CYRIL. "Experimental Tests of General Intelligence," Brit. Journ. of Psychology, Vol. III, Dec. 1909, 94 ff.

CLAPARÈDE, ED. "L'Unification et la Fixation de la Terminologie Psychologique," Comptes Rendus du VIme Congrès International de Psychologie, Genève, 1909. Librairie Kündig, Genève, 1910, 477.

ELDERTON, ETHEL M. "On the Association of Drawing with other Capacities in School Children," being a biometric evaluation of the results of E. Ivanhoff, "Recherches expérimentales sur le Dessin des Ecoliers de la Suisse Remande," Archives de Psychologie, VIII, 1908; Biometrika, Vol. VII, 1909–10, 222 ff.

KRUEGER, F. AND SPEARMAN, C. "Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten," Zeitschrift für Psychologie, Bd. XLIV, 1906, 50 ff.

MCDOUGALL, WM. "On a New Method for the Study of Concurrent Mental Operations and of Mental Fatigue," Brit. Journ. of Psychology, Vol. I, 1905, 435 ff.

MYERS, C. S. An Introduction to Experimental Psychology. Cambridge University Press, 1911, Chs. VI and VII.

PEARSON, KARL. "On the Mathematical Theory of Errors of Judgment, with special reference to the Personal Equation," Phil. Trans. A, Vol. 198, 235 ff.

—— The Grammar of Science, 2nd edition. London, A. and C. Black, 1900, Ch. XI, 3rd edition, Part I, 1911.

—— "Francis Galton's Difference Problem," Biometrika, Vol. I, 1902.

—— "On the Relationship of Intelligence to Size and Shape of Head and to other Physical and Mental Characters," Biometrika, Vol. V, 1906–1907, 105 ff.

PETERSON, A. "Correlation of Certain Mental Tests in Normal School Students," Psychological Review, Vol. xv, 1908, 323 ff.

SCHUSTER, E. AND ELDERTON, E. M. "The Inheritance of Psychical Characters," being further Statistical Treatment of Material Collected and Analysed by Messrs G. Heymans and E. Wiersma, Biometrika, Vol. v, 1906–1907, 460 ff.

SPEARMAN, C. "'General Intelligence' Objectively Determined and Measured," American Journal of Psychology, Vol. xv, 1904, 201 ff.

—— "Demonstration of Formulae for True Measurements of Correlation," American Journal of Psychology, Vol. XVII, 1906, 161 ff.

—— "Correlation Calculated from Faulty Data," Brit. Journ. of Psychology, Vol. III, Oct. 1910.

THORNDIKE, E. L. "An Empirical Study of College Entrance Examinations," Science, N.S. Vol. XXIII, 1906, 839 ff.

—— Educational Psychology, 2nd edition, New York, 1910.

—— "Empirical Studies in the Theory of Measurement," Archives of Psychology, New York, 1907.

—— "The Relation between Memory for Words and Memory for Numbers, and the Relation between Memory over Short and Memory over Long Intervals," American Journal of Psychology, Vol. XXI, 1910, 487, 488.

WHIPPLE, G. M. A Manual of Mental and Physical Tests. Baltimore, Warwick and York, 1910.

WINCH, W. H. "The Transfer of Improvement in Memory in School Children," Brit. Journ. of Psychology, Vols. II and III, 1908 and 1910.

WISSLER, CLARK. "The Correlation of Mental and Physical Tests," Psychological Review, Monograph Supplement, Vol. III, No. 16, June 1901.

—— "The Spearman Correlation Formula," Science, N.S. Vol. XXII, 1905. 309 ff.

CHAPTER IV

ELDERTON, W. P. Frequency Curves and Correlation, 1906, Ch. VI.

PEARSON, KARL. "Regression, Heredity and Panmixia," Phil. Trans. 1896.

YULE, G. U. An Introduction to the Theory of Statistics, 1911, Ch. XVI.

SUBJECT INDEX

CAMBRIDGE: PRINTED BY JOHN CLAY, M.A. AT THE UNIVERSITY PRESS.