THE USE AND INTERPRETATION OF COEFFICIENTS OF CORRELATION

WALTER S. MONROE, UNIVERSITY OF ILLINOIS

Within the last few years the statistical device known as the coefficient of correlation has become extensively used in various phases of educational research. As a result, the interpretation of the coefficient of correlation has become a matter of major inportance. In the brief time that is allotted to me I desire to call attention to two conditions which affect the meaning to be attached to a given value of the coefficient of correlation.

In the first place, the magnitude of the coefficient of correlation is affected by a selection of the population from which it was calculated. When studying general relationships it is obvious that we must resort to sampling. For example, if we wish to ascertain the relation between success in Latin and success in English, it is necessary for us to answer this question on the basis of the relation that exists between these two achievements in a limited population. If two populations are selected in a random or unbiased manner the resulting coefficients will not be identical, except by chance. The differences, however, will not generally be large, and if we are justified in assuming that the selection is a random one, the fluctuations of the resulting coefficients of correlation conform to a general law. For any given value of the coefficient of correlation calculated from a known number of cases we can determine the probable error due to sampling by means of the formula

$$P.\,E._{r} = .6745 \frac{1 - r_{12}^{2}}{\sqrt{n}}$$

In certain types of educational research the magnitude of the coefficient of correlation is affected by other aspects of the selection of the population group from which it is calculated. In the case of certain traits, the range of the magnitude of the traits is a potent factor in determining the magnitude of the coefficient of correlation. For example, if we are studying the relation between intelligence and achievement in silent reading, the coefficient of correlation will be much smaller when calculated from a random selection of the pupils belonging to a single half grade than when it is calculated from a random selection of pupils taken from grades three to eight, inclusive. In a recent study the coefficient of correlation between these two traits was calculated for each half grade from 3B to 8A, inclusive. The coefficients of correlation range from —.12 to .57. Since there were from 200 to 400 cases in each half grade the probable errors of these coefficients of correlation are relatively small. When the data from all grades were combined the coefficient of correlation for the same two traits was found to be .76. This is distinctly larger than the average of the coefficients of correlation for the separate half grades.

A more striking illustration is obtained from the correlation between chronological age and mental age. In the investigation referred to above, the coefficients of correlation were calculated between these two traits for each half grade. These range from —.12 up to —.44. In other words, the correlation between mental age and chronological age for the pupils belonging to a single half grade is inverse. In general, the pupils who are older chronologically tend to be the ones who are younger mentally, and vice versa. When the pupils from grades 3B to 8A, inclusive, are grouped together the correlation between chronological age and mental age was found to be .56. In this instance the coefficient is not only larger but it is positive, although when the pupils were taken by separate half grades all of the coefficients were negative. It, however, does not always happen that the coefficient of correlation is affected by the grade range from which the population is selected. For the same pupils the correlation was calculated between the intelligence quotient and the achievement quotient for arithmetic. When the pupils are taken by separate half grades, these coefficients range from —.19 to —.62. When all grades are taken together, the coefficient of correlation was found to be —.41, which is essentially the average of the coefficients of correlation for the separate half grades.

It is, therefore, necessary to take into account the type of selection that is made as well as its random character. It is not enough to make the selection of data in an unbiased way. In addition, one must interpret the coefficient of correlation that is obtained with reference to the particular type of selection that has been followed. If all of the data have been obtained from the pupils belonging to a given half grade, the interpretation will be generally on a distinctly different basis from that to be followed if the data were selected in a random way from the pupils belonging to a sequence of several grades.

My second thesis is that the meaning to be attached to the numerical value of a coefficient of correlation depends upon the type of relationship being studied. For example, if we were studying the relation between general intelligence and quality of handwriting, a coefficient of correlation between these two traits of .40 would be high. If we were to obtain this result for a representative student population drawn from the sequence of several grades, we would be justified in asserting that for this group the relationship between general intelligence and quality of handwriting was very high. On the other hand, if we were studying the relation between first trial and second trial scores on an intelligence test we would consider a coefficient of .40 exceedingly low, if it was based upon a representative group of pupils from a sequence of several grades. In order for a coefficient of correlation to be considered high in this case it would need to be .80 or larger.

When these two illustrations are analyzed with reference to the questions we are attempting to answer by means of the coefficient of correlation, we find that we are dealing with questions which are not identical. In the case of the relationship between general intelligence and handwriting we are primarily concerned with ascertaining whether or not any relationship exists. If any does exist, it is very small. If we secure a positive coefficient of correlation which is three or four times larger than its probable error, we have evidence of a positive relationship between these two trials. We are, of course, interested somewhat in the closeness of this relationship, but since it is slight we have answered the question in so far as we are able when we have determined that a positive relationship does exist. On the other hand, when we are dealing with the relation between first trial scores and second trial scores on the same test we are not concerned with ascertaining whether a positive relationship exists between these sets of facts. We know from our general experience with educational tests that a relatively high positive relationship exists between first and second trial scores. The question which we wish to answer in this case is how nearly perfect is this relationship, or perhaps to put it into a more effective form, how great are the departures from a perfect relationship. If we compare the coefficient of correlation with its probable error due to sampling we obtain little or no assistance in interpreting it in such case. For example, we might have a coefficient of correlation of .65 with a probable error of ± .03. In this case the coefficient of correlation is more than 20 times its probable error, but the relationship between the first trial scores and the second trial scores is far from perfect.

In order to secure a measure of departure from a perfect correlation the probable error of estimate has been proposed for use instead of the coefficient of correlation. It may be calculated by means of the formula,

$$P. E._{Est} = .6745 \ P \sqrt{1 - r_{12}}.$$

The calculation of the probable error of estimate, if the coefficient of correlation is known, involves some very simple arithmetic. It gives a measure of the magnitude of the departures from perfect correlation, or to put it in another way, it measures the amount of change which would be necessary in one set of data in order to bring them into perfect correlation with the other. For example, the correlation between the number of questions answered by eighty fourth grade children on forms 1 and 3 of the Courtis Silent Reading Test No. 2, was found to be .87 ± .02. In this case the coefficient of correlation is 43 times its probable error. This will be recognized as a very high correlation since it is based upon data of children from a single school grade. The probable error of estimate is 4.6. This means that in fifty percent of the cases one set of scores departed from perfect correlation with the other by more than 4.6 units. Since the average score is approximately 40 this amounts to about twelve per cent of the magnitude of the average scores.

In the formula for the probable error of estimate it will be noted that its magnitude depends upon the standard deviation of the distribution of the data from which the coefficient of correlation is computed as well as upon the coefficient of correlation. For this reason there is far from perfect correspondence between the magnitude of the probable error of estimate and the magnitude of the coefficient of correlation, even when the difference in the size of units is recognized by taking the ratio of the probable error of estimate to the average. For example, a certain collection of paired facts yielded a coefficient of correlation of .52. The probable error of estimate was 6.4 and the ratio of this to the average was .33. Another collection of data yielded a coefficient of correlation of .8, a probable error of estimate of 1.4 and the ratio of the probable error of estimate to the average of .14. A third collection of data yielded a coefficient of .51, a probable error of estimate of 17.6, and a ratio of .37. It will, therefore, be seen that widely different probable errors of estimate may be obtained for coefficients of correlation of approximately the same magnitude.