

and Rowe¹ give a list of seven formulae which by no means exhausts the logical possibilities.

In view of certain erroneous statements which have been made as to the nature and meaning of this coefficient, and of certain fallacious inferences drawn from its use, it is thought desirable to offer the following comments, based upon a preliminary examination of the problem from the biological point of view, in the hope that workers with the necessary statistical and mathematical equipment might be persuaded to give some attention to the development of this device which, it is thought, could be of considerable value in taxonomic practice.

The following points in connexion with this coefficient seem to have been overlooked:

(1) It is an index expressing one mean value (root mean square deviation) as a percentage of another mean value. Accordingly, the values normally calculated may be regarded as being possibly not true values since no correction is ever made for correlation effects, and notably no correction is made for the effect of spurious correlation which arises where an index is calculated from mean values, as shown by Pearson²; nor is correction ever made for the correlation which generally exists between a mean and its standard deviation.

(2) The numerator of the index is a quantity determined by many factors according to which it can be partitioned. The *CV* is customarily calculated only from such values of the standard deviation as may be to hand: at best these are sample values and it cannot be held that they accurately represent the variability of the particular group to which they refer. Accordingly, it is likely to be erroneous to compare the *CV*'s of two groups, unless the conditions of sampling, that is to say, the sources of variance, are identical. Even when sampling conditions are similar, such comparison of *CV*'s can be regarded as evidence of relative variability of the groups only in respect of the particular measurements to which they refer. Comparison on other measurements may reverse the relative positions.

(3) The correlation between a mean (of a particular measurement) and its variance is extremely variable and is itself a feature to be determined. Accordingly, values of the *CV* cannot usually be predicted on biometrical grounds alone. Thus, for the one measurement in a particular species the *CV* may or may not vary with sex, age, locality, season or other factor. Again, the coefficients of different measurements in the one species may or may not be the same and may or may not behave similarly in respect of such factors as sex, age and so on. Similar observations may be made in respect of the *CV* of a particular measurement made on different species, or genera. I have compiled tables of means, standard deviations and *CV*'s, and find it generally demonstrable that the characteristic value of the *CV* for any measurement in any group cannot be predicted, but must be separately determined. This does not deny that the behaviour of this coefficient might be according to some discernible law, particularly if some alteration were made in the manner of its calculation. However, this is a question of the relationship between two variables (the mean and the variance) under various conditions, and such relationship cannot be analysed by means of an index. Finally, as a matter of immediate practical importance, it is fallacious to attempt to set any general limits to the value of the *CV* or to draw any particular conclusions from departures from such limits.

(4) Since so many factors may contribute to the size of the variance, and since the manner of that contribution cannot be predicted, it is impossible to argue to the sample from the value of the *CV*. But departures in subsequent samples, from the value of the *CV* established for a particular specification of sampling, might serve to indicate that the conditions of sampling had been departed from, or that certain changes in the population had occurred.

However, despite these limitations to the *CV* it is desirable to have some measure of relative variability, and it is thought that with some modification, and with care in the specification of the conditions of use, the *CV* can serve this purpose. In the first place a *CV* should be cited, as a taxonomic feature, only where the material from which it was obtained can be precisely specified. Secondly, since the object is to permit comparisons, it would be desirable to effect some choice of conditions of sampling which can be generally reproduced; thus it might be wise to specify the *CV* for the sexes separately, to specify a single locality (say, the type locality) and only a few age groups. Thirdly, it would be a useful innovation to cite the *CV* partitioned according to the various sources of variance; coefficients could be quoted for each of the most important sources of variance and one for the residual variance. The latter might prove to be a fundamental characteristic of the species. The need for coefficients for 'interaction' would depend upon the magnitude of the effect. Finally, further refinements could be introduced by the adjustments possible through the covariance analysis. However, the *CV* should be an end-product of a detailed analysis: crude values should not anticipate such analysis.

While the *CV* in its present form is of very limited value it probably could be made a most useful adjunct to the usual set of statistical measures quoted in taxonomic works.

G. L. KESTEVEN

Fisheries Section,
Marine Biological Laboratory,
Cronulla,
New South Wales.
Sept. 3.

¹ Simpson and Rowe, "Quantitative Zoology" (McGraw-Hill, 1937).
² Pearson, K., *Proc. Mag. Soc.*, **60**, 489 (1897).

A General Class of Confidence Interval

AMONG types of statistical inference about unknown parameters statements are possible which have a statistical truth, that is, they are random variables such that within the statistical framework adopted the probability of their being in error is known. In these statements intervals, called by Neyman confidence intervals, are assigned to the value of an unknown parameter. On generalization to more than one unknown parameter these intervals become multi-dimensional regions; but I have pointed out¹ that the existence of

such regions does not of itself imply in Neyman's theory the corresponding existence of regions of lower order, equivalent to the elimination of irrelevant unknown parameters.

However, the logical statement concerning the simultaneous boundary of several parameters includes a statement about the maximum boundary of any selected set of these parameters, and consequently if the total statement is true with probability $1 - \epsilon$, the included statement is true with probability not less than $1 - \epsilon$. When the selected set consists of only one parameter, this fact gives rise to a general class of confidence interval for one parameter that includes all previously known 'exact solutions' with probability $1 - \epsilon$ and also new solutions with probability not less than $1 - \epsilon$. When optimum exact solutions do not exist, investigation of the optimum solution of the new type may still be possible.

In Fisher's most recent discussion² of his theory of fiducial probability, including the problem of testing the difference between two means, for which the Behrens-Fisher test does not constitute a solution in the above sense, he seems to throw out a challenge to critics of this test to provide an alternative 'tolerable solution'. In my original critical discussion³ the existence was noted of a two-parameter fiducial distribution for the true difference between the two means and the true ratio of variances. In addition to solutions of the confidence interval type previously noted, this two-parameter distribution implies a possible solution of the new type suggested above, for which the optimum (that is, shortest) confidence interval may be calculated. This particular solution, since it is based on an inequality, is not obviously more powerful than others based on exact solutions of a non-optimum type, but its statistical properties are open to investigation. But 'validity' and 'tolerability' should not be confused—the solution proposed here is valid in the sense defined, whether or not on more detailed examination it proves 'tolerable'.

Note added September 23. Since this letter was written, Dr. B. L. Welch has shown me the manuscript of a forthcoming paper in *Biometrika*, in which he puts forward a new solution of the confidence interval type for the 'difference between two means' problem. His solution appears to be exact, at least in the sense of allowing a series expansion for the true limits in terms of the initial large-sample normal approximation, and promises, much more than my own suggestion, to provide the so far missing 'tolerable' solution. I have also been interested to learn from recent conversation with Prof. A. Wald of some related unpublished work of his on the existence of such an exact solution.

Queens' College,
Cambridge.
Aug. 21.

M. S. BARTLETT

¹ *Ann. Math. Stat.*, **10**, 129 (1939).

² *Sankhya*, **7**, 129 (1945).

³ *Proc. Camb. Phil. Soc.*, **32**, 560 (1936).

Random Associations on a Lattice

GIVEN a lattice of $m \times n$ points, suppose that each may be 'black' or 'white' with probabilities p and $q = 1 - p$. The probability distribution of the number of 'black-white' joins is then of interest in several branches of science^{1,2,3}. The expected number is $2pq(2mn - m - n)$ and the second moment about the mean is $2pq(8mn - 7m - 7n + 4) + 4p^2q^2(13m + 13n - 14mn - 8)$.

As m and n increase, the distribution tends to normality, and this may be proved by methods similar to those used by Bernstein⁴ in his work on Markov chains. Similar results can be obtained for the number of 'black-black' joins and also the corresponding results in three dimensions. Levene⁵ has announced results dealing with a different but similar problem. A full account will appear later.

P. A. MORAN

Institute of Statistics,
University of Oxford.

¹ Mood, A. M., *Ann. Math. Stat.*, **11**, 367 (1940).

² Ising, E., *Z. Phys.*, **31**, 253 (1925).

³ Wishart, J., and Hirschfeld, H. O., *J. Lond. Math. Soc.*, **11**, 227 (1937).

⁴ Bernstein, S., *Math. Ann.*, **97**, 1 (1926).

⁵ Levene, H., *Bull. Amer. Math. Soc.*, **52**, 621 (1946).

Experiment and Theory in Statistics

MR. D. V. LINDLEY¹ distinguishes between the curve of best fit and the regression curve on the ground that the former gives the best estimate of the relation between the true values of the variables, while the latter gives the best estimate of the true value of one variable from the observed value of the other. He thus provides a convenient opportunity to raise a question that I have long wanted to raise, namely, whether there is any experimental evidence for this and similar deductions from statistical theory.

Direct experimental tests are not easily devised; for the meaning of the conceptions involved in such statements is derived largely from the theory on which they are based. But the first statement seems capable of test. The true values must surely be independent of the method of examining them. Accordingly, if many different sets of observations are made on a system by many different ways, the curves of best fit from the different sets, rightly calculated, should agree significantly better than curves calculated in any other way, for example, better than the regression curves. Is this true?

Nor are the tests easy to apply. A single worker seldom accumulates enough observational material to apply the test: he cannot use the observations of others, because sufficient details are scarcely ever published. The application of the tests requires organised co-operation. Until it is undertaken and the tests proved to be successful, all use of elaborate statistical theory is precarious.

NORMAN R. CAMPBELL

Sept. 2.

¹ *Nature*, **158**, 273 (1946).