Measures of the Amount of Ecologic Association Between Species
Author(s): Lee R. Dice
Reviewed work(s):
Source: *Ecology*, Vol. 26, No. 3 (Jul., 1945), pp. 297-302
Published by: Ecological Society of America
Stable URL: http://www.jstor.org/stable/1932409
Accessed: 26/05/2012 07:11

MEASURES OF THE AMOUNT OF ECOLOGIC ASSOCIATION BETWEEN SPECIES

LEE R. DICE
University of Michigan

In many ecologic studies there is need to express in a quantitative manner the degree to which two different species are associated in nature. Nevertheless, no such measure has up to the present time come into general use by ecologists.

Of the measures of association between species previously proposed, the "coefficient of association" of Forbes ('07) is useful, though no estimate of its statistical reliability has in the past been employed. The "coefficient of association" of Michael ('20) is too complicated for general use. The "association value" of Hacker ('21) varies with the total number of associations between all the species in a given locality and therefore is impracticable for most field studies. Van Deventer ('36) has calculated the percentage of association between certain species of birds, but his brief statement gives no details of the method of computation. Recently Edmondson ('44) applied the chi-square test as a measure of the deviation from chance expectation of the occurrences together of two given species of rotifers in a series of lakes, but he did not use any measure of the amount of association between the species.

The coefficient of association developed by Forbes ('07) considers the number of associated occurrences of two given species in a series of random samples compared to the number of such joint occurrences expected by chance. In order to calculate this coefficient we must therefore count the number of samples in which each of the two species occurs and also the number of samples in which both species occur together. Let us indicate by a the number of samples in which species A occurs either alone or together with species B, by b the number of samples in which species B occurs, by h the number of samples in which both species occur, and by n the total number of samples examined. The chance of species A occurring in any particular sample in the series is therefore a/n. Similarly, the chance of species B occurring in any sample is b/n. The chance that species A and B will both occur in any given sample is $a/n \times b/n$. The number of samples in the series in which species A and B would be expected to occur together is therefore $a/n \times b/n \times n$ or ab/n. To obtain the coefficient we now divide the number of samples (h) in which both species are observed to occur by the number of samples (ab/n) in which they would be expected to occur by chance:

$$\text{Coefficient of association} = \frac{h}{ab/n} = \frac{hn}{ab}.$$

This formula is exactly that proposed by Forbes, except that the symbols are different.

A coefficient of association of 1.0 shows that the two species under consideration occur together in exactly the number of sample units expected by chance. A coefficient smaller than 1.0 shows that they occur together in fewer samples than would be expected by chance. For example, a coefficient of 0.5 shows that they are found together only half as frequently as expected by chance. Similarly, a coefficient larger than 1.0 indicates that the species being compared are associated more frequently than would be expected by chance alone. The coefficient of association therefore gives a readily understandable indication of the degree to which the association between any two species conforms to expectation on the basis of chance.

The value of the coefficient of association between any two given species varies with the abundance of the forms in the area studied. If the two species compared are both abundant, both will appear in a high proportion of the samples taken. The chance of both species occurring together in any given sample is therefore high, and it will be impossible for any great deviation from expectation to occur. Under these conditions the coefficient of association can never be far from 1.0. On the other hand, if both species are rare in the situation, then the chance of their occurring together in any given sample is also low. Any deviation from expectation may result in a very high or very low value for the coefficient of association. A high coefficient therefore does not necessarily mean that the two forms are strongly associated, nor does a low coefficient always indicate that the forms are strongly repulsed. A coefficient of 1.0 shows that the two forms were observed to occur together in exactly the number of samples expected by chance, but it does not necessarily indicate that the forms do not tend to be associated. The coefficient of association is therefore not a measure of the amount of association between species, but only a measure of the amount of the deviation of the number of their occurrences together from the number expected by chance.

To provide a simple and direct measure of the amount of association between any given species (A), taken as the basis of comparison, and any other species (B) I here propose the *association index*. This index is obtained by dividing the number (a) of random samples of a given series in which species A occurs into the number (h) of samples in which species A and B occur together. Therefore:

$$\text{Association index B/A} = h/a.$$

The formula for the reciprocal index is:

$$\text{Association index A/B} = h/b.$$

For example, if species A should occur on a total of 40 sample plots ($a = 40$), and if species B should occur together with A on 30 of these plots ($h = 30$), then the association index of species B with species A (B/A) would be $30/40 = 0.75$. This index shows that on three-quarters of the plots on which species A occurred, it was associated with B. If we further assume that in the series of sample plots under consideration the number (b) of plots in which species B appeared was 30, then the association index of species A with species B (A/B) would be $30/30 = 1.0$. This index shows that on all the plots where species B appeared species A also was present.

It will be noted that the association index may differ depending on which species is used as the basis of comparison. Such a difference may point out an important ecologic relation between the two species. Frequently one species is dependent upon another, without there being any reciprocal dependency. It is believed that the association indices may sometimes be of value in discovering such unilateral dependencies.

Because the reciprocal association indices are different it will be important always to indicate which species is used as the basis of comparison. This may be done by placing the name of the base species second in the statement. Thus, for a theoretical example: "association index *gracilis/bairdii* = 0.01."

In some ecologic studies it will be desirable to use a measure of the amount of association between species that does not change depending on which species is used as a base. Such a measure, here called the *coincidence index*, has a value intermediate between the reciprocal association indices A/B and B/A. Thus:

$$\text{Coincidence index} = \frac{2h}{a+b}.$$

In the example previously given the coincidence index would be $(2 \times 30)/(30 + 40) = 60/70 = 0.86$. The coincidence index therefore is a measure of the amount of association between both the two species compared.

The values of the association indices and of the coincidence index range from 1.0, which indicates association of the two species in all the samples examined, to 0.0, which indicates complete failure of association under the conditions of observation. Intermediate values give the proportional amount of association. The indices are therefore easy to comprehend.

No indication of a possible deviation from chance expectation is given either by the association index or by the coincidence index. It will therefore often be desirable to give one or other of these indices and also the coefficient of association. All of these may be calculated with only slight effort from the same data.

No statistical test of the reliability of any of these measures of association has been devised. However, the chi-square test (Snedecor, '40: 16–19) will show whether the combinations of species in the samples of a particular series may possibly be due to chance errors in random sampling. This test should therefore always be applied to the data from which any association index, coincidence index, or coefficient of association is derived.

In applying the chi-square test it is necessary to determine the number of units on which each possible combination of the two species would be expected to occur by chance. The expected number of samples in which two species, A and B, would be expected to occur together has already been shown to be ab/n. Species A would be expected to occur alone in the number of samples from which it has been recorded (a) less the number of samples in which it is expected to occur together with B. Similarly, B is expected to occur alone in b samples minus the number of samples in which B is expected to occur along with A. The chance of species A not occurring in any given sample in the series is $(n - a)/n$. Similarly the chance of B not occurring in any sample is $(n - b)/n$. The chance that neither A nor B will occur in a given sample is therefore $(n - a)/n \times (n - b)/n$ and the total number of samples in which neither species is expected to be present will be $(n - a)/n \times (n - b)/n \times n$ or $(n - a)(n - b)/n$. From the deviations from these expected numbers chi-square is calculated.

As an example let us assume a series of 100 samples in which species A has been found alone in 5 samples, species B alone in 20 samples, both species A and B together in 20 samples, and neither species in 55 samples (table I). Then the value

TABLE I. *Theoretical example of the association between two species (A and B) shown by 100 samples*

Class	Number of samples		Deviation from expectation	$\frac{(\text{deviation})^2}{\text{expected}}$
	Observed	Expected		
A only	5	15	−10	6.667
A+*B*	20	10	+10	10.000
B only	20	30	−10	3.333
Neither	55	45	+10	2.222
Totals	100	100	0	22.222

$a = 25$; $b = 40$; $n = 100$; coefficient of association $= 2.0$; association index $A/B = 0.5$; association index $B/A = 0.8$; coincidence index $= 0.62$; degree of freedom $= 1$; $\chi^2 = 22.222$.

of a is 25, b is 40, and n is 100. The number of samples in which A and B are expected to occur together by chance (ab/n) is 10. The coefficient of association is accordingly 20/10 or 2.0, which shows that the two species in the example occur together twice as frequently as would be expected by chance. The association index of species A with species B (A/B) is 20/40 = 0.5, and of species B with species A (B/A) is 20/25 = 0.8. The coincidence index is $(2 \times 20)/(25 + 40) = 40/65 = 0.62$.

In calculating the chi-square deviations in our example we used data from two species. The number of degrees of freedom is therefore one. The total chi-square is 22.222 (table I). From a table of the values of chi-square (Fisher, '36, table III; or Snedecor, '40, table 9.1) it is evident that this amount of chi-square with one degree of freedom is far above the one per cent. level of significance. The association of species A and B on the sample plots must therefore be due to some factor other than chance errors in random sampling. We may conclude that there is a tendency in this situation for species A and B to be rather strongly associated in their distribution. From the difference in association indices, moreover, it is suggested that species A is somewhat more strongly associated with B than species B is with species A.

The chi-square test is not reliable when the number of expected units in any class falls below five. If possible the expected number in each class should be ten or more. Should the number of samples available for a particular study nevertheless be few and if there be only a single degree of freedom, Yates' "correction" (Snedecor, '40: 168–169) may be applied. It is better, however, in every field study of the amount of association between species to secure a sufficient number of samples so that no class will contain less than ten expected units.

The magnitude of chi-square varies to some degree with the abundance of the species compared. Two strongly associated species might occur together on a large number of sample plots and yet, if both were abundant, the deviation from expectation might be slight and chi-square therefore small and not statistically significant. Also, two species that have only a slight tendency toward association, if both were rare in the samples, might occur together frequently enough to have a large and very significant chi-square. Care should therefore be taken not to interpret the magnitude of chi-square as indicating the strength of association between the species under consideration. On the other hand, the degree of significance of chi-square may be an indirect but valid measure of association.

The amount of association between three or more species can be determined by means of the coincidence index and coefficient of association by an expansion of the formulas used for two species. Also the significance of the deviation from expectation of the observed amount of association between these species can be determined by the chi-square test. Attempts to measure the association of three or more species will, however, very rarely be practical. The number of samples needed to secure an adequate measure of the association of only two species is very considerable. A fully reliable measure of the amount of association between two species will seldom be secured with much less than 100 samples, and more than this number will nearly always be desirable. It will rarely be possible to secure the much greater number of comparable samples needed to measure in one step the amount of association between three or more species.

So far we have discussed only the association between two or more given species. We may also need to measure the amount of association between a species and some inanimate feature of its habitat. For this purpose we may use the indices of association or coincidence, the coefficient of association, and the chi-square test of significance. We may thus compare the occurrence of a given species with the simultaneous presence of rocks, logs, trees, clumps of bushes, or any other feature of the environment. It must not be expected, however, that the calculation of these statistics will alone solve all the problems of the relationships between organisms and their environments.

The abundance of the individuals of either of the species concerned on the sample units is not considered in the calculation of the coefficient of association or of the indices of association or coincidence. Such an omission from measures of association has been criticized by Calvert ('22). Furthermore, no account has been taken, in these computations, of the possible tendency of the individuals to be associated with one another in pairs, families, or larger social aggregations.

The abundance and also the social behavior of the individuals of any given species obviously affect the frequency with which the form appears in any kind of sample. Therefore the frequency with which any species appears in a given set of samples includes, to some degree, an indication both of its abundance and of its social behavior. So long as any given sample is as likely as any other sample to contain a representative of each of the species concerned we may use these sample counts as a basis for computing the association between the species. We will readily grant, however, that these measures of association apply only to the samples taken and that if the two species involved should occur in other densities or if one or both should have another type of social grouping, such as might be true in other seasons of the year, then different values for these measures of association would probably be obtained. No methods have yet been devised for predicting the values of these measures of association under other population densities or under other social groupings than those exhibited by a particular sample series.

The samples that are used as a basis for the measurement of association may be any of the usual sample units employed in ecologic field work. They may be quadrats, transects, or sample plots of any kind, or they may be sample catches obtained by sweeps of an insect net, secured by drags of a tow net or dredge, or collected in any other standardized manner. The essential feature is that the two species to be compared actually must have been in close association with one another if they are recorded in the same sample unit. Any sample unit that covers a large area or that includes several different types of ecologic communities, therefore, will not usually be a proper base for the calculations. Sample units based on time intervals will be admissible only if they also are space units each of which covers an area of not too great extent.

The size of the units from which the samples are obtained influences all these measures of association. This is true because the frequency with which any species appears in a series of samples is dependent not only upon its density and its social tendencies but also upon the size of the area from which the sample is drawn. If the sample plots are of large size, almost all the species in the local biota, except only the rarer ones, will appear on a large proportion of the plots. On sample plots of small size, on the contrary, even the most abundant species are likely to appear on only a small fraction of the plots. For this reason heterogeneous samples, such as the collections of fish compared by Forbes ('07), do not provide a sound basis for the calculation of the amount of association between species.

Inasmuch as the value of all these measures of association varies with the size of the sample units employed, it is evident that the sample units of the series on which any of these computations is based must all be of the same size. The size of sample unit to be employed must of course be adjusted to the abundance and also to the size and mobility of the species being compared. The most efficient measure of association seems generally to be obtained when the more abundant one of the species under consideration occurs on about half of the sample units in the series.

All sample units on which estimates of association are to be based should be taken at random, using this term in the statistical sense. If any selection is used in locating the sample units the method of selection should be fully described. Samples taken systematically, such as one collection in each of a series of stands or of communities, usually cannot be used properly as a basis for statistical calculations.

Any measure of association is of course true only for the particular place and time for which it is calculated. The ecologic relations between organisms vary from time to time, from community to community, and from place to place. It will always be important, therefore, to specify the particular conditions upon which any given measure is based. It will usually be impossible to combine data taken at different times, in different places, in different types of communities, in different manners, or from different sizes of samples to serve as a basis for the calculation of the degree of association between species.

The fact that two species may tend to occur together, as shown by a high association index, coincidence index, or coefficient of association, or by a significant chi-square deviation from expectation, does not in any way explain the reason for their association. It may be that the two forms are mutually attracted to one another or that one form is in some way dependent upon the other. The two species may, however, be associated only because of the presence in their habitat of certain features attractive or essential to both. In such a case the two associated species may have no relationship at all to one another, other than selection of the same habitat. These statistics at best are a measure only of the actual amount of association between two species rather than of the amount of attraction between them.

Summary

The coefficient of association of Forbes indicates the amount of association between two given species compared to the amount of association between them expected by chance. In order to provide a simple direct measure of the amount of association of one species with another the *association index* is proposed. If a is the number of random samples of a given series in which species A occurs and h is the number of samples in which another species B occurs together with A, then the association index B/A = h/a. Similarly, if b is the number of samples in which species B occurs, then the association index A/B = h/b. There is also proposed a *coincidence index*, $2h/(a + b)$, whose value is intermediate between the two reciprocal association indices. As a measure of the statistical reliability of the deviation shown by the samples of a given series from the amount of association expected by chance, the chi-square test may be used.

Literature Cited

Calvert, Philip P. 1922. Methods for expressing the associations of different species. Ecology **3**: 163–165.

Edmondson, W. T. 1944. Ecological studies of sessile Rotatoria. Part I. Factors affecting distribution. Ecol. Monog. **14**: 31–66, 5 figs.

Fisher, R. A. 1936. Statistical methods for research workers; Sixth edition. Edinburgh and London: Oliver and Boyd. xiii + 339 pp.

Forbes, S. A. 1907. On the local distribution of certain Illinois fishes: an essay in statistical ecology. Bull. Illinois State Lab. Nat. Hist. **7**: 273–303, 15 maps, 9 pls.

Hacker, H. P. 1921. On the distribution of anophelines in the Brickfields Road area of Kuala Lumpur, November, 1917. Federated Malay States Malaria Bureau Reports **2**: 3–28, 1 map.

Michael, Ellis L. 1920. Marine ecology and the coefficient of association. Jour. Ecol. **8**: 54–59.

Snedecor, George E. 1940. Statistical methods applied to experiments in agriculture and biology. Ames, Iowa: Iowa State College Press. xiii + 422 pp., Illus.

Van Deventer, William C. 1936. A winter bird community in western New York. Ecology **17**: 491–499, 3 figs.