

NOTES ON THE MEASUREMENT OF FAUNAL RESEMBLANCE

GEORGE GAYLORD SIMPSON

Museum of Comparative Zoology, Harvard University,
Cambridge, Massachusetts

ABSTRACT. Of various measures of taxonomic resemblance, the percentage in the smaller of two faunal samples of the number of taxa common to both is most useful. It tends to eliminate the effects of discrepancy in size between the two faunas or samples, and when that discrepancy is a factor in the problem being studied the percentage of common taxa may be used. When faunas closely similar taxonomically are compared, it may be desirable to take into account differences in the relative abundances of taxa in common. For that purpose, measures based on rank correlation are suggested, but no such measure seems full satisfactory. Several other measures used in current literature are discussed.

MEASURES BASED ON NUMBERS OF TAXA

Most studies of faunal resemblances are strictly zoogeographical or correlative (in a stratigraphic sense), although other and especially ecological factors cannot be wholly excluded and are sometimes of primary importance. The usual approach is concerned with the presence or absence of taxa and with degrees of phylogenetic relationships among them. Degrees of phylogenetic affinity are practically, even though somewhat crudely, quantified by the levels of hierarchic classification. Comparisons at the specific level involve close affinity, at the generic level broader affinity, at the family level still broader, and so on. Presence and absence of taxa, for purposes of measurement of faunal resemblances, are adequately quantified by the following symbols and concepts:

E_1 , Number of taxa (at a specified level) in the first, smaller (or equal) of the two faunas or samples compared, but absent in the second.

E_2 , Number of taxa in the second, larger (or equal) fauna or sample but absent in the first.

C , Number of taxa common to both.

$N_1 = E_1 + C$, Total taxa in first.

$N_2 = E_2 + C$, Total taxa in second.

$N_t = E_1 + E_2 + C = N_1 + N_2 - C$, Total taxa in both.

The most obvious, and apparently the most intuitively acceptable measure of faunal resemblance, is simply the percentage of taxa in common among the total taxa of the two faunas or samples in question:

$$\frac{C}{N_t} \times 100 \quad (1)$$

which for purposes of compilation and calculation is more conveniently represented as:

$$\frac{C}{N_1 + N_2 - C} \times 100 \quad (1a)$$

If both faunas are almost or quite completely known, and if they are of at least approximately equal size, index (1) is usefully indicative. The maximal value, 100, shows complete identity and the minimal, 0, complete difference at the chosen taxonomic level; and, in general, intermediate values

are clearly proportionate to degrees of resemblance. If, however, either fauna is incompletely known, or if the known numbers of taxa are distinctly unequal index (1) is difficult to interpret and may be quite misleading.

It is generally the case with fossil faunas and sometimes with recent faunas that the available samples, on which a measure must be based, are incomplete and do not include all the taxa of both faunas. If the two samples have the same degree of completeness relative to the populations¹ (e. g. each has half the taxa of the respective population) and neither is biased with respect to E/C, index (1) from the samples may give comparable estimates of the population values. The actual ratio of sample N to population N is, however, always unknown and can rarely be closely estimated. The best one can do is to assume, in the absence of indications to the contrary, that the two populations have about the same numbers of taxa. If that assumption is true, comparisons of samples of approximately equal size still give somewhat biased estimates of index (1), but estimates that are reasonably comparable and are adequate if the samples are fairly large.

Those conditions are, however, rarely met. Samples are frequently of quite unequal size. It is frequently probable that the smaller sample (smaller in terms of number of taxa included) also has a smaller proportion of the taxa originally present in the corresponding population. Then the value of index (1) will be especially strongly biased as an estimate of the population value, and the bias will be greater the greater the discrepancy in N_1 and N_2 in the samples. The mere fact that N_2 is larger makes index (1) smaller. A quick and intuitively clear way to check such bias, for this or any other index, is to consider samples drawn from indentially the same population. If bias is absent, they should of course tend to give the maximal value of the index. It will at once be seen that index (1) cannot have its maximal value, 100, if the samples are of unequal size and that it may give extremely low values for samples from the same population. Obviously, comparisons based on this index are usually unreliable for unequal faunas or samples.

The following index eliminates the worst disadvantages noted for (1):

$$\frac{C}{N_1} \times 100 \quad (2)$$

As an estimate of a population index from samples, index (2) minimizes effects of differences in size between N_1 and N_2 . When samples are small, both (1) and (2) of course have considerable sampling error; but (2) is also preferable in this respect, and indeed the larger the discrepancy between N_1 and N_2 (the better N_2 but not N_1 is sampled), the lower the sampling error without, as with (1), introduction of increasing bias. When the sample (and

¹ Here and elsewhere in this paper, the word *population* refers to the actual fauna under consideration as it exists in nature. It might, for instance, be all the native mammals now present in a defined area, or all the fossil pelecypods preserved (but not necessarily collected) in a given stratum. When taxonomic indices are being used, the pertinent population is all the *taxa* (at the level involved in the index) present in the defined fauna in nature. Sampling error is involved when any of those taxa are unknown. When indices take individual abundance into consideration (discussed later in this paper), the pertinent population is all the individuals in the defined fauna. Then study is virtually always on a sampling basis.

population) sizes are indeed equal, (2) is still at least as good an estimate as (1), so there is no reason to turn to (1) in this (unusual) situation.

Note that (1) and (2) both have scales from 0 to 100 but that they do not tend to be equal except when (1) is 100 or either one is 0. [(1) is not, however, necessarily 100 when (2) is 100.] Otherwise (2) is always somewhat larger than (1). These, like any two different indices, measure different things, and values are to be compared only when they represent the same index.

When both faunas are completely known taxonomically, the comparison is on the true population basis rather than being based on incomplete samples. This situation is common in comparing Recent faunas of higher vertebrates, at least, but rarely arises in paleontology. When the populations are completely known, the advantages of (2) may be less obvious but may still exist. If the numbers of taxa in the two populations are equal, both indices make consistent (not identical) comparisons and there is no particular reason to prefer one over the other. If difference in sizes is in itself a meaningful element in the desired comparison, then (1) is preferable. Other things being equal, index (1) but not (2) will be smaller the greater the discrepancy between N_1 and N_2 , and this is desirable if the discrepancy is important for the particular problem being investigated.

That is not, however, always or even usually the case. If, for instance, comparison is of the fauna of a smaller with that of a larger area, the zoogeographic relationships are more clearly indicated if it is possible to eliminate the tendency of larger areas to have larger faunas simply because the areas *are* larger, and then index (2) is clearly more indicative. Index (2) is even more obviously advantageous when the smaller fauna has been derived from the larger, or when a local fauna is compared with a regional fauna. The situation is comparable in stratigraphic correlation when, as is so often done, a particular faunule is compared with the whole known fauna of a given age. In these or other comparisons, it also frequently is true that the larger fauna is more varied ecologically. Then use of index (2) tends to minimize the merely ecological difference, although obviously this cannot be wholly eliminated. In effect, index (2) tends to stress the most nearly similar parts of the two faunas, which is usually an advantage in zoogeographic and especially in stratigraphic investigations. Finally, since index (2) is generally less biased for paleontological samples, it is advantageous to use (2) consistently and thus to permit more nearly valid comparisons of both fossil and recent faunas. (For further discussion and for extensive exemplification of index (2) see, e.g. Simpson, 1947).

Burt (1958) suggests that what is here called index (2) "should be applied in both directions to give the true picture," in other words that one should give not only

$$\frac{C}{N_1} \times 100 \quad (2)$$

but also

$$\frac{C}{N_2} \times 100 \quad (3)$$

This is not, in fact, the application of the same measure in two directions, but the addition of another measure with different properties. Unless the samples are equal in numbers of taxa, when (2) and (3) become identical, and except at the extreme values 100 and 0, (3) is always smaller than (2) for the same comparisons. If one thinks of these indices as truly analogous or as comparable with each other, it may appear that fauna 2 resembles fauna 1 less than fauna 1 resembles fauna 2, which is confusing. In fact (3) is a distinctly different measure that suffers to exaggerated degree from the disadvantages of index (1), already discussed. If the discrepancy in sizes of samples is considered significant and is wanted to influence the index, then (1) is available and is more readily understood and more widely used. Otherwise (2) is better. There seems little reason to use index (3).

Still another index that is a sort of average (but not the arithmetic mean) between (2) and (3) has been advocated especially by Pirlot (1956; see also Burt, 1958, and Crusafont and Truyols, 1958) :

$$\frac{2C}{N_1+N_2} \times 100 \quad (4)$$

This is the percentage of common taxa not in the total for both samples as in (1), or in one sample or the other as in (2) and (3), but in the mean number of taxa for the two, as is more obvious when the formula is written:

$$\frac{C}{(N_1+N_2)/2} \times 100 \quad (4a)$$

This is in effect a compromise between (1) and (2), for it reduces the effect of discrepancies in sample size, which is prominent in (1), but does not minimize that effect as does (2). It is hard to see a meaningful use for (4), because when discrepancy in size is pertinent to the problem (1) is better, and when it is not, (2) is better.

In contrasting (1) and (4), Burt (1958) notes that in (1) C appears in both numerator and denominator and in (4) only in the numerator. That is, however, misleading because C is really present in both numerator and denominator of all the indices (1)-(4), as is evident when they are rewritten as follows:

$$\frac{C}{E_1+E_2+C} \times 100 \quad (1b)$$

$$\frac{C}{E_1+C} \times 100 \quad (2a)$$

$$\frac{C}{E_2+C} \times 100 \quad (3a)$$

$$\frac{2C}{E_1+E_2+2C} \times 100 \quad (4b)$$

If it were desirable really to eliminate C from the denominator, this

could be done in some such form as:

$$\frac{C}{E_1 + E_2} \times 100 \quad (5)$$

but this is no longer a percentage with conveniently comparable values from 100 to 0, and it has no compensating advantages.

Finally, some such formula as this could be used:

$$\frac{E_1}{E_1 + C} \times 100 = \frac{E_1}{N_1} \times 100 \quad (6)$$

This is a percentage, but it is merely 100 minus index (2) and has the disadvantage of rating identity as 0 and complete difference as 100. There are, of course, comparable analogues for indices (1), (3), and (4).

TAXONOMIC LEVELS

An essential factor in all the indices of taxonomic resemblance is the hierarchic level of the taxa counted. It need hardly be emphasized that values of an index are comparable only if they are at the same taxonomic level and that the values will generally be higher the higher the level. Choice of too high a level for a given group of comparisons will give many or all values = 100, and too low a level will give many or all values = 0. It is almost always possible to find a level with most or all values less than 100 and more than 0, thus permitting meaningful comparisons. Determination of such a level is in itself a general indication of the degree of phylogenetic affinity involved in the comparisons.

As has been emphasized especially by Burt (1958), direct comparisons assume that the taxa really are comparable in the two faunas, that, for instance, genera are not notably more split in one than in the other. When taxa are evidently more split in one, that must be taken into account in interpreting the meanings of the indices. It may be added that index (2) has the additional advantage of tending to minimize effects of disproportionate splitting in the larger fauna.

Burt also finds it misleading that E (in symbols here used) includes forms that have relatives (at a higher taxonomic level) in both faunas as well as those that do not. For instance in comparing the Recent North American with the Asiatic mammalian fauna, *Odocoileus* has fairly close Asiatic relatives, *Antilocapra* does not, but both are counted in E. He proposes to include *Odocoileus* (and numerous other genera with Asiatic relatives) in C. If this adjustment is clearly understood and can be made consistently, it may indeed be useful. It should, however, be noted that the same purpose can be achieved more objectively by simply including analysis at an appropriate higher taxonomic level. In all of Burt's examples his objections would have been met more or less adequately by also giving indices at the subfamily or family level.

Another possible adjustment, still closer to Burt's purpose although he does not suggest it, would be to redefine C and E:

C_1' , Number of genera (or other appropriate specified lower taxa)

in the first fauna (or sample) belonging to families (or other appropriate higher taxa) common to both.

C_2' , Same for second fauna or sample.

E_1' , Number of genera (or as specified) in the first in families (etc.) absent in the other.

E_2' , Same for the second fauna or sample.

With these definitions, C' , unlike C , is not a single number the same for both faunas, but indices comparable to those already discussed can readily be devised, e. g.:

$$\frac{C_1 + C_2'}{2N_1 + 2N_2 - C_1 - C_2'} \times 100 \quad (7)$$

$$\frac{C_1}{N_1} \times 100 \quad (8)$$

Index (7) is analogous (but not directly comparable) to (1), and (8) to (2). The different form of the denominator of (7) is necessary to keep this index analogous with (1) by substituting the mean of C_1 and C_2' for C . A more logical but less convenient equivalent expression is:

$$\frac{(C_1 + C_2')/2}{N_1 + N_2 - [(C_1 + C_2')/2]} \times 100 \quad (7a)$$

All the taxonomic indices, (1)—(8), are based on existing classifications. They necessarily reflect such subjectivity and lack of consistency as may affect those classifications. They also necessarily neglect relationships of possible importance for zoogeography (e. g. probable direction of spread between two areas) that are not involved in or expressed by classification. Like any other statistics, they cannot automatically solve any problem. They can order and quantify the data that enter into them and thus they assist in sound comparisons. After being calculated, they must still be interpreted in the light of knowledge of exactly what it is that each of them measures and in accordance with zoological and stratigraphic principles.

MEASURES INVOLVING ABUNDANCE OF TAXA

The indices hitherto mentioned, which include those in most general use, directly measure resemblance in taxonomy only. They take no overt account of the abundance of individuals in the taxa or of various other factors. In another sense, it is evident that faunas will be more similar if the taxa in common are relatively abundant in individuals and less similar if the most abundant taxa are those peculiar to each. *Ecological* differences between *taxonomically* similar faunas may, indeed, be largely reflected in the relative abundance of their taxa. For discussion of this point an additional set of symbols is needed.

I_{E1} , I_{E2} , Number of individuals of taxa not in common in first and second samples, respectively.

I_{C1} , I_{C2} , Number of individuals of taxa in common in first and second samples, respectively.

$I_1 = I_{E1} + I_{C1}$, $I_2 = I_{E2} + I_{C2}$, Total number of individuals in first and second samples, respectively.

$I_C = I_{C1} + I_{C2}$. Total number of individuals in the taxa in common.

The new and interesting idea of establishing an index in which individual abundance figures is due to Crusafont and Truyols (1958). As a preliminary trial they proposed the following index:

$$\frac{I_C : 2C}{I_1 : N_1 + I_2 : N_2} \times 100 \quad (9)$$

The following algebraically identical formula may be somewhat easier to calculate and is more obviously indicative of the mathematical behavior of the index:

$$\frac{I_C N_1 N_2}{2C (I_1 N_2 + I_2 N_1)} \times 100 \quad (9a)$$

In the original publication Crusafont and Truyols noted some disadvantages of this formula, for instance that it is not scaled from 0 to 100 (not being a percentage) and may give values far above 100, with a seemingly erratic scale making comparisons uncertain. In further illustration of its peculiarity, it suffices to point out that for two identical samples the value of this index is 50 while for nonidentical samples the value may also be 50 or higher or lower. In view of this and other inconvenient properties of the index (some noted in the original publication and some in subsequent unpublished discussion) it is preferable to seek another form of index based on the same general principle.

Among possible alternatives is:

$$\frac{I_C}{I_1 + I_2} \times 100 \quad (10)$$

— the percentage of individuals in taxa in common among total individuals in both samples. This has a simple and rational scale, from 100 for complete identity to 0 for no resemblance, but it still has the drawback that differences in sizes of unbiased samples may result in different values of the index. That consideration suggests the use of:

$$\frac{1}{2} \left(\frac{I_{C1}}{I_1} + \frac{I_{C2}}{I_2} \right) \times 100 = \left(\frac{I_{C1}}{I_1} + \frac{I_{C2}}{I_2} \right) \times 50 \quad (11)$$

— which is the mean of the percentages of individuals in common taxa figured separately for each sample. This also has the logical percentage scale from 100 to 0 and (if sampling is unbiased) it is not biased by sample size. This simple index thus has none of the serious weaknesses of (9) and (10) and is, I think, the best to use when an index of this particular kind is wanted. It should, of course, still be borne in mind that different indices measure *different* things and that a value of, say, 50 for (11) does not necessarily (although it may) indicate the same degree of resemblance as one of 50 for any other index.

Indices (9), (10), and (11) all have in common another weakness: they will not reliably reflect differences in relative abundance of taxa in faunas that are closely similar taxonomically. If, for instance, two samples have all their taxa in common, (10) and (11) will always be 100 and (9) will always be 50 even if the faunas are ecologically different and have different dominant taxa. But as Crusafont and Truyols point out, it is exactly this situation that calls for considering relative abundance and not only taxonomic resemblance. If taxonomic resemblance is moderate or low, indices (9)—(11) will not be likely to give any reliable information not inherent in the purely taxonomic indices. Thus (9)—(11) are least indicative when most needed.

Of several possible solutions to this special problem, perhaps the simplest is to use rank correlation of the taxa in common. The taxa are listed with the abundance or frequency of each in the available samples. For each sample each taxon is assigned a rank, 1 for the most abundant, 2 for the next, etc. For each taxon, the rank for one sample (it does not matter which) is then subtracted from the rank of the other, giving a difference, d . The rank correlation coefficient is then:

$$r_r = 1 - \frac{\sum (d^2)}{C(C^2 - 1)} \quad (12)$$

This coefficient has a logical but somewhat peculiar scale, and its significance in this usage is not the same as in usual statistical correlation. Values range from +1 to -1. The value +1 reflects maximal resemblance, -1 minimal resemblance, and 0 indicates not absence of resemblance but an intermediate value. In order to make this formally similar to most of the previously discussed indices, with scales of 100 to 0, it may be transformed to:

$$50 + 50r_r \quad (13)$$

It must, however, again be remarked that a value for (13) does not indicate the same degree and sort of resemblance as the same value for any other index.

The coefficient (12) and index (13) leave entirely out of account the taxa that are not common to the two samples. These measures are therefore indicative only when most taxa are in fact common to the two. When that is not true, (11) may be used, or a purely taxonomic index, preferably (2), may be sufficient.

Measures (9)—(13) are all influenced both by purely taxonomic resemblance and by numbers of individuals in the common or noncommon taxa. Although (12) and (13) do not vary with numbers of taxa in either or both categories, they can only be used when in fact C is large, so that there is still a strong influence of numbers of common taxa. These are all compromise measures which either apply only under certain conditions as to taxonomic resemblance or seek to evaluate taxonomic and some sort of ecological resemblances simultaneously. They may correspondingly be ambiguous or difficult to interpret. They are, in particular, quite limited as measures of ecological resemblance, which frequently is largely or even almost completely indepen-

dent of taxonomic resemblance. None of these measures takes any account whatever of nontaxonomic ecological similarities or differences. Again, it is essential to know exactly what an index is actually measuring and to select an index that really is indicative for the particular problem in hand.

An important further point is that relative abundance of individuals in different taxa may in fact strongly influence the values of the apparently purely taxonomic indices (1)—(8). This is not the case if both faunas are completely known taxonomically and the question of sampling probability does not arise. If, however, either fauna is incompletely known, then the probability that a given taxon will appear in the corresponding sample will tend (other things being equal) to depend on the relative abundance of individuals in that taxon. Under these circumstances if the taxa in common are also more abundant, the taxonomic indices will tend to be larger than if these taxa were relatively rare. Thus in sampling situations such as are so common, especially with paleontological materials, all the indices, including those that are overtly taxonomic only, are likely also to be covertly affected by resemblances in individual abundances of taxa. This is not necessarily a disadvantage and may even be an advantage, but it is another factor to be taken into account in interpretation of the data.

Note on Sampling Effects

In the remarks just made, and also earlier in the paper, sampling effects have been mentioned in quite general terms. For fully reliable comparisons statistical confidence intervals should, of course, be established for any index used. For any of these indices, that is an exceptionally difficult problem that has not yet been worked out. The present notes are notes only, and not even by intention the last word on the problem. In the meantime use of the recommended indices, notably (2), is pragmatically justified because on the whole and with suitable caution they do give fairly consistent and meaningful results. To give them up and to rely on purely subjective comparisons until a complete statistical theory is available would be unjustified. The indices are objective data even when uncertainties remain in their statistical interpretation.

This problem does not arise for indices (1)—(8) when the faunas are completely known taxonomically. The population relevant here is a real and finite entity, quite distinct from any theoretical population of statistics. If the finite population, a list of taxa, is known, that is a direct and basic observation not affected by any considerations of probability. Then a value of, say, 78 for index (2) means exactly that. It is a measurement, not an estimate, of the true value and there is no probability at all that the true value is 79, 77, or any other figure than the one observed. If in a second such comparison a value 80 is obtained, then the difference between that and the first value of 78 is, so to speak, infinitely significant. There is no chance at all that the population values are the same. What this small difference means in terms of a particular problem is another matter; the difference, as such, exists.

When not only numbers of taxa but also numbers of individuals are taken directly into account [indices (9)—(13)], there is always a question of sta-

tistical estimation of probability. It is not likely ever to happen that *all* individuals in the populations can be counted, and there is bound to be some sample variance in proportions of the various taxa. The approach here is not too difficult, at least in theory, although the calculations would often be onerous. For any one taxon, confidence intervals for the relative frequencies of its individuals can be obtained by well-known statistical methods (e.g. Simpson, Roe, and Lewontin, 1960). Approximate confidence intervals for r_r [(12)] are also given in readily available texts.

The really serious problem is that of incomplete sampling of taxa, which directly or indirectly affects all the indices here mentioned. When all taxa are known, the confidence interval at any level is zero, as already mentioned. The intervals become larger, and apparently in a complex way, as the *proportion* of known taxa becomes smaller. But in this case the proportion itself is unknown, because the absolute number of taxa in the population is unknown. As already noted, the confidence intervals in terms of taxa are also indirectly influenced by the abundance of individuals and therefore by confidence intervals for relative abundance, which again greatly complicates the problem. There are, moreover, many other factors that may affect the probability of finding any particular taxon.

Finally, it is certainly true of some of these measures and perhaps of all of them that they are biased when based on an incomplete roster of taxa. That is, repeated samples of less than the full number of taxa will not tend to vary about, or to give as an average, the true value of the index in the population. Generally they will tend to give smaller values. A precise expression for this tendency would, again, be very complicated and has not been worked out.

All of this calls for further and better study. Nevertheless comparisons by the same index and with samples including about the same proportion of the populations' taxa are reasonably indicative. Indices based on a large proportion of population taxa also approach the population value nearly enough for most purposes.

AN EXAMPLE

Table 1 gives data for fairly good, representative fossil faunules from the same area, the relationships of which are reasonably well understood. Table 2 applies most of the suggested measures to comparisons of these faunules. The results exemplify in some detail the characteristics of these indices already discussed in a more theoretical way.

If differences in facies are minimized in favor of purely taxonomic or phylogenetic resemblances, comparison (a-b) should give a high value, (b-c) also high, (b-d) considerably lower, and (c-d) somewhat lower still. Index (2) most nearly agrees with these expectations, although the discrepancy between (b-d) and (c-d) is perhaps greater than might have been anticipated. This may be related to the probability that both c and d include only small fractions of the taxa of the original populations so that all the indices for these two comparisons are relatively less reliable.

TABLE 1

Comparative data for genera of fossil mammals from different localities and horizons in the Paleocene of the Crazy Mountain Field, Montana. (Based on Simpson, 1937).

Datum (defined in text)	Faunal Comparisons			
	a-b	b-c	b-d	c-d
E ₁	3	2	8	11
E ₂	16	21	29	13
N ₁	21	15	13	13
N ₂	34	34	34	15
C	18	13	5	2
E' ₁	0	0	1	6
E' ₂	7	13	12	3
C' ₁	21	13	12	7
C' ₂	27	21	22	12
I _{E1}	3	7	43	104
I _{E2}	83	156	349	59
I _{C1}	55	72	50	3
I _{C2}	299	226	33	20
I ₁	58	79	93	107
I ₂	382	382	382	79
I ₀	354	298	83	23

Samples from: a, Silberling quarry. b, Gidley quarry. c, Lebo surface localities. d, Scarritt quarry.

Comparisons: a-b, samples of unequal size similar in facies and age.

b-c, samples of unequal size and different facies but about the same age.

b-d, samples of unequal size and different (successive) ages with some resemblance in facies.

c-d, samples fairly close in size but distinctly different in both age and facies.

TABLE 2

Measures of faunal resemblance from the data of table 1

Measures (as in text)	Comparisons (defined in table 1)			
	a-b	b-c	b-d	c-d
(1a) $\frac{C}{N_1 + N_2 - C} \times 100$	49	36	12	8
(2) $\frac{C}{N_1} \times 100$	86	87	38	15
(3) $\frac{C}{N_2} \times 100$	53	38	15	13
(4) $\frac{2C}{N_1 + N_2} \times 100$	65	53	21	14
(7) $\frac{C_1 + C_2}{2N_1 + 2N_2 - C_1 - C_2} \times 100$	77	53	57	51
(8) $\frac{C_1}{N_1} \times 100$	100	87	92	54
(9a) $\frac{I_0 N_1 N_2}{2C (I_1 N_2 + I_2 N_1)} \times 100$	70	69	34	43
(10) $\frac{I_0}{I_1 + I_2} \times 100$	80	65	17	12
(11) $\left(\frac{I_{C1}}{I_1} + \frac{I_{C2}}{I_2} \right) \times 50$	85	75	31	14
(13) $50 + 50r_r$	86	54	— ^a	— ^a

^a For these two comparisons C is too small to permit meaningful use of this index.

For combined measures of taxonomic and ecological resemblances, (a-b) should be markedly high, and the other measures successively distinctly lower. Index (9) has this tendency but is still inadequate, while (10), (11), and, as far as it goes, (13) give reasonable results on this basis.

The whole table strikingly demonstrates the fact that comparisons cannot usefully be made from one index to another, although (3) does follow the same trends as (1). Indices (2), (3), and (4) tend to take the same value when N_1 is nearly equal to N_2 , as in the comparison (c-d); (7) and (8) at quite a different level also tend toward the same value under this condition. Discrepancies in sample size strongly affect (1), (3), (4), and (7) but have little influence on (2) and (8).

REFERENCES CITED

- Burt, W. H., 1958, The history and affinities of the Recent land mammals of western North America: Zoogeography, Am. Assoc. Adv. Sci., Pub. 51, p. 131-154.
- Crusafont Pairó, M., and Truyols Santonja, J., 1958, Ensayo sobre el establecimiento de una nueva fórmula de semejanza faunística: Barcelona Inst. Biología Aplicada Pub. 28, p. 87-94.
- Pirlot, P. L., 1956, Les formes européennes du genre *Hipparion*: Barcelona Inst. Geología Mem. y Comunicaciones, no. 14, p. 1-150.
- Simpson, G. G., 1937, The Fort Union of the Crazy Mountain Field, Montana, and its mammalian faunas: U. S. Natl. Mus. Bull. 169, 287 p.
- , 1947, Holarctic mammalian faunas and continental relationships during the Cenozoic: Geol. Soc. America Bull., v. 58, no. 7, p. 613-688.
- Simpson, G. G., Roe, A., and Lewontin, R. C., 1960, Quantitative zoology (revised ed.): New York, Harcourt, Brace, 440 p.