Empty Sites and the Analysis of Presence-Absence Data

S. Joseph Wright, Carl C. Biehl

American Naturalist, Volume 122, Issue 6 (Dec., 1983), 833-834.

EMPTY SITES AND THE ANALYSIS OF PRESENCE-ABSENCE DATA

The geographic distribution of a species may be affected by other species. For instance, competitors may not be able to coexist and mutualists may never be found apart. We analyzed presence-absence data to test the null hypothesis that species are distributed independently of one another (Wright and Biehl 1982). There are two alternative hypotheses: Species may be overdispersed among sites (e.g., competitors) or species may be aggregated among sites (e.g., mutualists). Reddingius (1983) recreates the analysis which we used to determine whether two species are over-dispersed, i.e., co-occurr on fewer islands than are expected by chance. He encounters two problems: a typographical error in our paper and a logical error of his own making.

Two probabilities which we both calculated differ (Reddingius 1983, table 2 cols. 2 and 4). One differs by .000009 or by rounding error. The other differs by precisely one order of magnitude, and Reddingius is correct. There are two, more serious, errors in our article. First, in the denominator of the expression on page 355 of Wright and Biehl (1982), the sum should be replaced by a product sum. Second, in our presentation of the competitive guild model on pages 347–348, we neglected to state that our equation (2) (for the probability that x species are shared between 2 sites) is an asymptotic probability and only applies when the number of sites is large. Fortunately, this does not change the conclusions we drew from this model.

Reddingius (1983) creates one additional problem. He adds six empty sites (the Dutch Wadden Isles, 53°, 30′ N in the Atlantic Ocean) to each of two of the original data sets (6°–2° S in the Pacific Ocean and 43°–45° N in the Adriatic Sea) and he finds that the statistic which we calculate changes. Reddingius concludes from this that our analysis is incorrect. He is wrong. A statistic (in this case the probability that x or fewer sites are co-occupied by 2 species) is a property of a sample taken from a study population, and a sample statistic estimates a population parameter. When a sample is altered by adding more items from the same population, the sample statistic may change even though the population parameter does not. Reddingius (1983) adds empty sites from a new population of sites. If this had not changed the sample statistic, there would be cause for concern. While we regret Reddingius's confusion with our paper, our Venn diagram representing the possible distribution of three species onto N islands does include the event that some islands are unoccupied (Wright and Biehl 1982, fig. 3).

Finally, empty sites could pose a problem for analysis of presence-absence data. To test the hypothesis that two species affect one another's distributions, we first census a sample of sites which we think could be colonized by both species. Nonetheless, both species could be absent from some of the sites. This could happen because the species are obligate mutualists, through the vagaries of chance extinction and colonization, or for any number of other reasons. In any

Am. Nat. Vol. 122, pp. 833–834.

event, the exclusion of empty sites after the sample is taken increases the probability of type I error. This is true because the probability that the observed number or fewer co-occurrences arose by chance decreases as the sample size, N, decreases while r, q, and the number of observed sites shared remain constant.

Note the implicit assumption of our analysis. Each species is assumed to be able to colonize each site. This assumption requires that species and sites are selected carefully. We do not add North Atlantic islands devoid of fruit pigeons and cuckoo-doves (and their habitat requirements) to our analysis of birds of equatorial Pacific islands. On the contrary, the population of study sites must be chosen carefully. For instance, if the null hypothesis is that aquatic species are distributed independently among sites, each site must include aquatic habitats. The inclusion of sites which lack aquatic habitats and therefore aquatic species will make it artificially difficult to reject the null hypothesis. Likewise, if the null hypothesis is that birds are distributed independently among sites, the population of species should be restricted to birds. On the other hand, if the null hypothesis concerns competitors, the population of species should be restricted to potential competitors. In sum, the population of sites and species must be tailored to match the null hypothesis and, to avoid circularity, this must be done without reference to the presence-absence data being analyzed.

The sample of sites must meet one further requirement. If too few sites are included, the null hypothesis that species co-occur at random can never be rejected (Biehl and Matthews 1984).

ACKNOWLEDGMENTS

James F. Quinn and an anonymous referee provided useful criticism of our paper. Daniel Simberloff kindly brought to our attention the omission we made regarding the competitive guild model.

LITERATURE CITED

Biehl, C. C., and W. J. Matthews. 1984. Small fish community structure in Ozark streams: improvements on statistical methods. Am. Midl. Nat. (in press).

Reddingius, J. 1983. On species sharing islands: comments on an article by S. J. Wright and C. C. Biehl. Am. Nat. 122:830–832.

Wright, S. J., and C. C. Biehl. 1982. Island biogeographic distributions: testing for random, regular, and aggregated patterns of species occurrence. Am. Nat. 119:345–357.

S. Joseph Wright

Smithsonian Tropical Research Institute
APDO 2072
Balboa, Republic of Panama

Carl C. Biehl

Department of Biology
University of California
Los Angeles, California 90024

Submitted February 17, 1983; Accepted June 29, 1983