Ann. Zool. Fennici 32: 435–437
Helsinki 8 December 1995
ISSN 0003-455X

Commentary

Analyzing insular distribution patterns: statistical approaches and biological inferences

Mark V. Lomolino, James H. Brown & Russell Davis

Lomolino, M. V., Department of Zoology and Oklahoma Biological Survey, University of Oklahoma, Norman, OK 73019 USA
Brown, J. H., Department of Biology, University of New Mexico, Albuquerque, NM, 87131 USA
Davis, R., Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721 USA

Received 4 May 1994, accepted 4 July 1994

In a recent paper by Rita & Ranta (1993) we (Lomolino et al. 1989) were criticized for not using logistic regression to investigate the factors associated with distributions of non-volant mammals on isolated, montane forests of the American Southwest. The purpose of this paper is to explain why we chose not to use logistic regression and to point out a number of oversights of Rita and Ranta.

First, it appears that Rita & Ranta (1993) were unaware of Adler and Wilson's paper (1985) that also proposed the use of logistic regression to describe patterns of insular occurrence. Since then, Schoener & Adler (1991) have developed more sophisticated techniques to study similar patterns given appropriately large data sets.

Second, even if they disagreed with our reasons for not using logistic regression, Rita and Ranta should have noted our expressed concerns with this method. We referred the reader to an earlier paper (Lomolino 1986:8) with the following explanation. "Logistic regression was not used because it requires quite large sample sizes; according to Harrell (1983:190)", with two independent variables (area and isolation) "the number of observations in the least frequent category should be 20... Therefore N should be >> 40". Harrell (1983:190) includes the following guidance for determining the adequacy of sample size for variable selection in logistic regression: "In general, if there are m observations for the least frequent category of a binary response variable, you should not examine more than $m/10$ variables in order to derive a model that is somewhat reliable". Applying this to our (Lomolino et al. 1989) data for mammals on 27 montane islands, m can not exceed 13 (i.e., < half of N), therefore $m/10$ (i.e., the number of independent variables) should not exceed 1. Other papers, including at least one of those cited by Rita & Ranta (1993), also caution on inappropriate sample sizes and the limitations of variable selection techniques (see Hosmer & Lemeshow 1989, James & McCulloch 1990). Perhaps Rita & Ranta (1993) disagree with the above cautions or feel that we were too conservative, but they still should have noted that we considered logistic regression as an alternative and we explained why we chose not to use it.

Rita & Ranta (1993) also failed to make clear that, for cases with binary dependent variables, multiple linear regression is equivalent to discri-

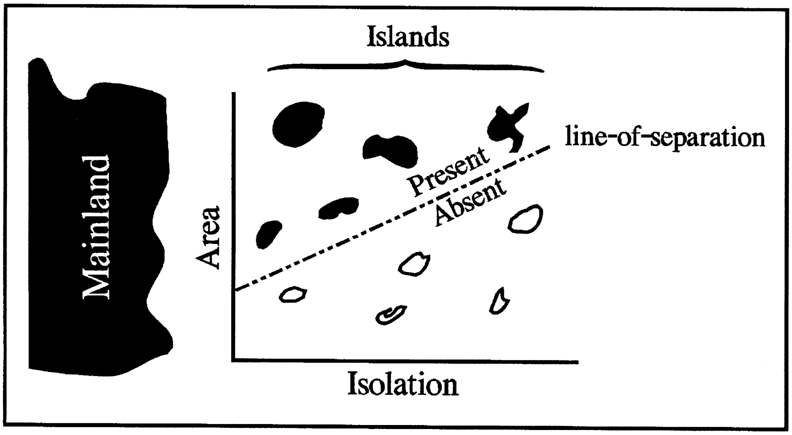

Fig. 1. Insular distribution function for a hypothetical species. Here we assume that insular distributions are influenced by recurrent immigrations and extinctions. Therefore, insular distributions should be biased toward the larger and less-isolated islands (presence or absence indicated by filled and open symbols, respectively). The line-of-separation indicates the combinations of area and isolation at which insular incidence should equal 0.5 (see Lomolino 1986, Hanski 1986 and Lomolino et al. 1989).

minant analysis and that both of these techniques are relatively robust against violations of multinormality and homoscedasticity (Lomolino et al. 1989, after Pimentel 1979). In addition, James & McCulloch (1990, after Effron 1976) hold that in comparison to logistic regression "if the data are multivariate normal, linear discriminant function analysis is a more efficient procedure". The search for factors associated with insular incidence (presence or absence) is clearly a problem of discrimination. Because the line separating absence from presence (Fig. 1) is easier to calculate from the results of linear regression analyses, we (Lomolino et al. 1989) preferred to use it in our analysis of distributions of montane mammals.

We also have a number of more technical problems with Rita & Ranta's (1993) analysis. Their selection criteria for including variables in the model are unclear, the fit of their model seems poor (based on high P-values in their Table 1), and their inclusion of lines-of-separation based on variables found not to be significant in their analysis seems misleading (see their Fig. 2 versus Fig. 6 in Lomolino et al. and Fig. 3 in Lomolino 1986). We agree with Rita & Ranta (1993) and Adler & Wilson (1985) that logistic regression is a potentially powerful tool for analyzing distribution patterns (see also James & McCulloch 1990 and Trexler & Davis 1993). Indeed, we would prefer logistic regression over alternative approaches if we had adequately larger sample sizes. Unfortunately, this is seldom the case in biogeographic studies (but see Schoener & Adler 1991 for a notable exception).

Finally, we feel Rita & Ranta (1993) should give more attention to the biological significance of their results, which they assume to be correct given the presumed primacy of their method. Most biologists now have an arsenal of diverse and powerful statistical tools at their finger tips ready to assist them in their search for patterns in nature. While easy access to these tools has improved our abilities to understand complex patterns, we should realize their limitations. Each statistical approach has assumptions that influence its appropriateness for a given biological problem. Decisions on which approach to employ and the inferences we draw from results of any statistical analysis must be guided by the biological characteristics of the problem under study. In one of the references on logistic regression cited by Rita & Ranta (1993), Hosmer & Lemeshow (1989:87) warn that "variable selection procedures can yield a biologically implausible model (…) the problem is not the fact that the computer can select such models, but rather that the analyst fails to carefully scrutinize the resulting model, and reports such results as the final, best model". We feel that, although they were well-intentioned, Rita & Ranta (1993) may have made this mistake — they failed to interpret and assess the biological relevance of their results. For example, their Figure 2 indicates that *Eutamias quadrivitattatus* and *Microtus mexicanus*, in addition to being restricted to the less-isolated islands, are also restricted to the smaller islands. Higher incidences on smaller islands seems counterintuitive but, if correct, would be a potentially insightful observation. The results of our study (Lomolino et al. 1986), however, seem more plausible. We found no significant bias for distributions on smaller islands. These two species of small mammals, being relatively limited in their abilities to disperse across dry woodlands and deserts, are limited to the less-isolated islands. The problem is not that Rita & Ranta's (1993) analysis suggests that distributions of these species were biased in favor of the smaller islands, but that they failed to comment on the biological significance of this apparent anomaly.

We conclude with a final quotation from Hosmer & Lemeshow's (1989:87–88) book on logistic regression. "The wide availability and ease with which stepwise methods can be used has undoubtedly reduced some analysts to a role where they are assisting the computer in model selection rather than the more appropriate alternative. It is only when the analyst understands the strengths, and especially the limitations of the methods that these methods can serve as a useful tool in the model-building process. The analyst, not the computer, is ultimately responsible for the review and evaluation of the model". Rita & Ranta (1993) are no more susceptible to this pitfall than any of us. Indeed, their note will probably have some positive impact on biogeography and statistical ecology. Hopefully, it will stimulate additional debate on the utility and limitations of discriminant analysis and logistic regression. We all, however, should be vigilant and take heed of the above caution.

References

Adler, G. H. & Wilson, M. L. 1985: Small mammals on Massachusetts islands: the use of probability functions in clarifying biogeographic relationships. — Oecologia 66: 178–186.

Effron, B. 1976: The efficiency of logistic regression compared to normal discriminant analysis. — J. Amer. Stat. Assoc. 70: 892–898.

Hanski, I. 1986: Population dynamics of shrews on small islands accord with the equilibrium theory. — Biol. J. Linn. Soc. 28: 23–36.

Harrell, F. E. 1983: The LOGIST procedure. — SAS Supplemental Library users guide: 181–202. Cary, NC: Institute SAS Institute Inc.

Hosmer, D. W. & Lemeshow, S. 1989: Applied logistic regression. — Wiley Interscience, New York.

James, F. C. & McCulloch, C. E. 1990: Multivariate analysis in ecology and systematics: panacea or Pandora's box? — Ann. Rev. Ecol. Syst. 21: 129–166.

Lomolino, M. V. 1986: Mammalian community structure on islands: the importance of immigration, extinction and interactive effects. — Biol. J. Linn. Soc. London 28: 1–21.

Lomolino, M. V., Brown, J. H., & Davis, R. 1989: Island biogeography of montane forest mammals in the American Southwest. — Ecology 70: 180–194.

Pimentel, R. A. 1979: Morphometrics. — Kendall/Hunt, Dubuque, Iowa, USA.

Rita, H. & Ranta, E. 1993: On analysing species incidence. — Ann. Zool. Fennici 30: 173–176.

Schoener, T. W. & Adler, G. H. 1991: Greater resolution of distributional complementarities by controlling for habitat affinities: a study with Bahamian lizards and birds. — Amer. Nat. 137: 669–692.

Trexler, J. C. & Travis, J. 1993: Nontraditional regression analyses. — Ecology 74: 1629–1637.