Small sample size problems in wildlife ecology: a contingent analytical approach

Author: Bissonette, John A.

ORIGINAL ARTICLES

Small sample size problems in wildlife ecology: a contingent analytical approach

John A. Bissonette

Administrative, political, and natural history constraints on the design of research studies in ecology often result in small data sets. In this paper, I identify some problems associated with small data sets and describe a contingent process for data analysis. I argue that exploring small data sets is heuristic and can be a valuable first step in the formulation of biologically interesting hypotheses.

John A. Bissonette, Utah Cooperative Fish and Wildlife Research Unit, U. S. Geological Survey - Biological Resources Division, College of Natural Resources & Department of Fisheries and Wildlife, Utah State University, Logan, UT 84322-5290, USA - e-mail: John.Bissonette@cnr.usu.edu

Wildlife studies by their nature and especially those involving species with sparse distributions often are characterised by small sample size, although notable exceptions exist, e.g. Berger 1986, Geist 1971, Sinclair 1977, McCullough 1979, Clutton-Brock & Ball 1987. Bears *Ursus* sp. and wolves *Canis lupus* in both Europe and North America are examples of sensitive, top level species for which political concerns and low population numbers appear to contribute to a limited data collection and often insufficiently replicated data sets. Similarly, mountain lions *Felis concolor* in North America are difficult to study because of their position as top carnivores and their wide-ranging habits. These species often operate over larger spatial and temporal scales, and often require longer-term studies over larger spatial extents to effectively capture the relevant dynamics. How-ever, studies with limited sample sizes can be extremely useful. Studies involving experimental techniques to breed endangered species held in captivity for subsequent release of adults and progeny into the wild, (red wolves *Canis niger* (Moore 1990, Jenks & Wayne 1992), condors *Gymnogyps californianus* (Kiff, Mesta & Wallace 1996), whooping cranes *Grus americana* (Longmire, Gee, Hardekopf & Mark 1992, Cannon 1996)) are valuable even though sample size is generally small. Unfortunately, in some studies there is a tendency to generalise small sample size results beyond the appropriate boundaries.

In addition to natural history constraints, political sensitivity, changing administrative priorities of funding sources, and funding constraints, viz., two-year master's level graduate studies, influence the length and intensity with which most research pro-

jects are conducted, often resulting in smaller data sets for a wide variety of studies. For example, Weatherhead (1986) found that the length of over 300 studies that he reviewed lasted an average of only 2.5 years. Further, Tilman (1989) reported that only 1.7% of field studies reported in the journal Ecology lasted 5 years or more, and only 7% of 180 papers that involved experimentation lasted 5 years or more; a large number (N = 72) lasted less than one year (see May 1994). Although during the past 10 years there has been an increased appreciation for studies of longer duration and larger spatial scale, the problem will still continue to plague ecology and the wildlife profession. In the two years (1988-1989) I served as Associate Editor of the Journal of Wildlife Management, the most common complaint from reviewers involved small sample size.

My purpose here is to present one approach to the small sample size problem. My objective was to develop a multiple regression model to predict barn owl *Tyto alba* reproductive success from several habitat characteristics; however, the use of automated multiple regression model selection techniques, e.g. stepwise selection, often hides important aspects of data. Additionally, and apparently not generally recognised, there is a general tendency for ecologists to "overemphasize the potential role of significance testing in.... scientific practice" (Yoccoz 1991) to the detriment of biological understanding. Yoccoz (1991) further stated that "most biologists and other users of statistical methods seem still to be unaware that significance testing *by itself* (italics mine) sheds little light on the questions they are posing". Thus, I adopted the philosophy of model selection recommended by Henderson & Velleman (1981) as an alternative approach. The approach is deliberately contingent: results at each step in the progressive analyses were evaluated, and the decisions about how to proceed to the next step were based on results from the preceding step and on biological insight (e.g. Myers 1990, Neter, Wasserman & Kutner 1985).

Analytical problems with small data sets

Small data sets pose structural problems for the investigator:
1) It is difficult to evaluate assumptions of the analyses, including the forms of the relationships between the response and explanatory variables.
2) Evaluation of any chosen model is ambiguous.

Characteristics of collinearity, outliers, and influential points that interfere in model selection for data sets of reasonable size are even more problematic with small data sets because they are more difficult to assess.
3) When the sample size is small or when the number of fitted parameters is a moderate to large fraction of the sample size, most model selection procedures will lead to models that appear to have high explanatory power and that select as significant, explanatory variables that are not truly related to the response (Freedman 1983, Freedman & Pee 1989, Hurvich & Tsai 1989).

There are additional concerns to be aware of when analysing small data sets. One risk involves the repeated analyses of the same set of data in a search for models that fit the data well. Consequently, a model may be fitting the random variation in the data set on which it is based, rather than the underlying biological relationship. As a result, the predictive ability of the model for a second, similar data set may be less than for the data upon which it was built (Neter et al. 1985, Maurer 1986). Prediction bias is increased when the number of observations is small with respect to the number of predictors (Magnusson 1983, Verbyla 1986). Maurer (1986), Rotenberry (1986), and more recently Anderson & Burnham (1998) have discussed in detail the problems associated with predictability of wildlife habitat models.

However, small data sets can be valuable for generating realistic hypotheses and testable models. All ecological models attempt to simplify the complexity of nature, using usually easily measured variables in equations that represent ecological relationships. But simplification of analytical models necessitates trade-offs between generality, precision, and realism (Levins 1968). In statistical correlation models, this trade-off may not apply. However, correlation points to pattern in ecology and to interesting questions for which we seek explanation.

No single criterion determines the 'best' model. Rather, model evaluation takes into account all criteria, as well as biological insight. Different criteria address different aspects of model goodness of fit:
1) One measure of goodness of fit is the adjusted coefficient of multiple determination, R^2, the proportion of the variation of the response variable that is jointly explained by the explanatory variables included in the model. Whereas R^2 always increases with the addition of an explanatory variable to the model, the adjusted R^2 increases

only if an added explanatory variable results in an improved fit of the model to the data (Zar 1996). This is especially important when sample size is small and the number of explanatory variables in the model relatively large because the adjustment is considerable (Sokal & Rohlf 1995). Among competing models, the one with the largest adjusted R^2 is favoured.

2) The error or residual mean square, s^2, another measure of goodness of fit, expresses variation in the residuals from the model (Myers 1990). Because the residual is the difference between observed value and the value predicted by the model, the model with the smallest s^2 is favoured. In selecting a 'best' model, balance between increased bias due to underfitting, i.e. failing to include important explanatory variables, and increased variance due to overfitting, i.e. incorporating unnecessary variables, must be achieved.

3) Mallow's C_p expresses variance plus bias, and thus is useful in discriminating between competing models (Myers 1990). The model with the smallest C_p is favoured. When the sample size is small, the data set cannot be split for validation reasons.

4) The PRESS (Prediction Sum of Squares) statistic is a criterion that can be used as a form of validation (Myers 1990). Each observation is set aside in turn, and a model is fit to the remaining observations in the sample. Using this model, the deleted response is estimated, and the PRESS residual is computed. The PRESS statistic is computed as the sum of squared residuals over all observations. Among competing models, the model with the smallest PRESS statistic is favoured.

Data set

Ault (1982) collected data on barn owl reproductive success from 11 nest sites over a 5-year period in Jackson County, near Eldorado, Oklahoma USA (Table 1). Six habitat variables were measured within a 1-km radius of each nest site. Nest sites were found in the bottom of empty cisterns (dry water wells). The variables he measured included: 1) road length (km); 2) edge length (measured as the linear distance (km) of contiguous habitats not including road edges); 3) area (ha) of wheat or sorghum (hereafter termed grain); 4) mesquite *Proposis glandula*; 5) herbland; and 6) an index of habitat interspersion (obtained by counting the number of discrete units of each cover type within the 1-km radius circle of each cistern nest site and then summing the number of units of each type across all cover types).

Contingent analyses

To avoid pseudoreplication (Hurlbert 1984), I used the cistern, not individual owls, as the sample unit. Regression analysis was performed using PROC REG in SAS and SYSTAT. Analysis proceeded as follows:

1) Reproductive success (# of young fledged) was plotted against each of the six explanatory variables to assess the nature (linear or non-linear) of the relationships and discover unusual points.

2) The set of explanatory variables was explored for collinearity.

3) All possible regressions were fit and screened as potential candidate models based on adjusted R^2, Mallow's C_p, and s^2 statistics.

Table 1. Observed and predicted reproductive success[a] of barn owls nesting in empty cisterns, and mean values for six habitat variables measured within a 1-km radius of each nest site in southwestern Oklahoma (USA), 1977-1981.

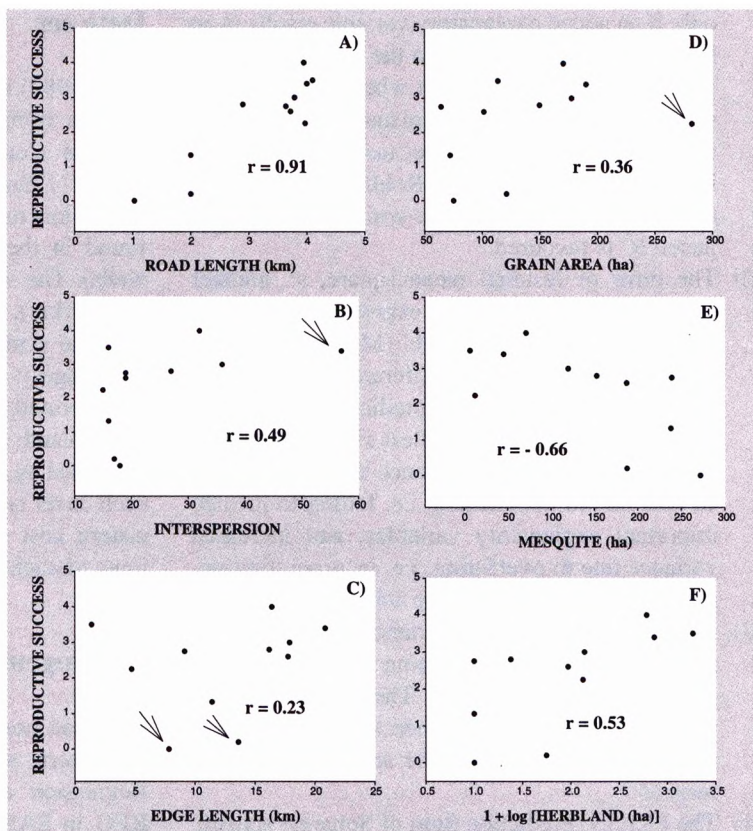| Nest site	Rs$_o$[b]	Rs$_p$[c]	Road length	I[d]	Edge	Mesquite	Herb	Grain
1	3.40	3.23	4.00	57	20.86	45.6	72.5	190.3
2	3.00	3.03	3.79	36	17.89	119.8	12.9	177.7
3	4.00	3.07	3.95	32	16.41	71.4	60.2	170.5
4	1.33	0.92	2.00	16	11.42	238.1	0.0	72.2
5	2.80	1.94	2.90	27	16.19	152.7	1.4	149.9
6	2.75	2.87	3.64	19	9.14	239.2	0.0	64.4
7	0.20	1.23	2.00	17	13.61	187.6	4.6	121.1
8	2.25	3.39	3.97	15	4.73	12.7	12.4	282.1
9	0.00	-0.17	1.03	18	7.83	272.0	0.0	75.1
10	3.50	3.35	4.10	16	1.39	6.9	183.3	113.7
11	2.60	2.99	3.72	19	17.76	187.2	8.4	101.5

[a] Mean number of young owls fledged over five years
[b] Observed reproductive success,
[c] Predicted reproductive success
[d] Interspersion index

4) Based on adjusted R^2, C_p, and s^2, a subset of models was selected for further evaluation, and the PRESS statistic was computed for each candidate model.

5) Residuals from each model were examined to evaluate the assumptions of linearity, normality, and homoscedasticity and to identify potential outliers. Partial regression plots and tests of significance of regression coefficients were used to assess the need for each variable. The susceptibility of any given model to influential data points was examined using various influence diagnostics.

6) One model was chosen to represent the data, based on the four criteria (adjusted R^2, C_p, s^2 and PRESS statistics), tests of significance, adherence to assumptions, susceptibility of each model to influential data points, and biological insight.

Figure 1. Barn owl reproductive success plotted against the six explanatory variables road length (A), interspersion (B), edge length (C), grain area (D), mesquite (E) and herbland (F). Arrows indicate data points whose presence or absence can alter the qualitative nature of the relationship, indicating the sensitivity of small data sets to single data points.

Results

The scatter plots (Fig. 1) illustrate one source of ambiguity in small data sets: individual observations may be highly influential. Snedecor & Cochran (1967: 175) state that "even the direction of inclination ... . may elude you if r is between -0.3 and + 0.3." This type of ambiguity may be most evident in Figure 1C. By omitting one or two arbitrarily chosen data points (see arrows in Fig. 1), the relationship between reproductive success and a given explanatory variable can appear positive or negative, linear or non-linear (e.g. Fig. 1 B, D), or even disappear (e.g.

Fig. 1C, D). For this analysis, all data points were considered valid.

No relationship between edge length and observed reproductive success (r = 0.23) was apparent in the scatterplot. Mesquite area exhibited a weak, negative linear relationship (r = -0.66); log-transformed herbland area (r = 0.53), grain area (r = 0.36), and interspersion (r = 0.49) exhibited weak, positive linear

Table 2. Pearson correlation coefficients for the six independent variables.

	Grain area	Mesquite area	Road length	Interspersion	Edge length	Log herbland area
Grain area	1.00	-0.77	0.53	0.36	0.09	0.51
Mesquite area		1.00	-0.76	-0.35	0.10	-0.89
Road length			1.00	0.38	0.11	0.71
Interspersion				1.00	0.72	0.44
Edge length					1.00	0.05
Log herbland area						1.00

Table 3. Candidate models and model selection criteria for barn owl reproductive success.

Model	Adjusted R^2	C_p	s^2	PRESS
Road	0.80	0.20	0.35	4.40
Road, interspersion	0.81	1.13	0.33	5.90
Road, grain	0.80	1.53	0.35	6.15
Road, interspersion, grain	0.82	2.02	0.31	7.09

relationships. Herbland area was logarithmically transformed to achieve a more linear relationship. Road length (see Fig. 1A) appeared most highly correlated with observed reproductive success (r = 0.91).

Simple correlation identified a potential collinearity problem (Table 2). A large negative correlation (-0.89) existed between mesquite area and log-transformed herbland area. Moderately large correlations were found for mesquite area and grain area (-0.77), mesquite area and road length (-0.76), road length and log-transformed herbland area (0.71), and edge length and interspersion (0.72).

A multiple linear regression with all six explanatory variables was fit to the data. Variance inflation factors, which represent the inflation of regression coefficients due to correlation among the explanatory variables, were large (>10) for mesquite area and log-transformed herbland area. Another measure of multicollinearity is the condition number. It is calculated as the ratio of the largest to the smallest eigenvalue of the correlation matrix for the explanatory variables. More formally, condition number is given as:

$$\emptyset = \lambda_{max} / \lambda_{min} \quad (1)$$

where λ is the eigenvalue. Numbers greater than 1,000 indicate serious multicollinearity problems (Myers 1990). Our condition number was 3,706. The large variance proportions indicated that collinearity appeared to involve the intercept, mesquite, grain, and to a lesser extent log-transformed herbland areas. Consequently, as a first step mesquite area was dropped as an explanatory variable. The multiple linear regression was re-fit without mesquite area; no further evidence of collinearity was found.

Four models were identified as potential candidates based on the first three criteria (Table 3). A larger adjusted R^2, a smaller C_p, a smaller s^2, or a smaller PRESS generally indicates a better model. No one model provided the best performance on all four criteria. The model with road length alone provided a reasonable fit to the data. Road length (x) was

positively associated with reproductive success (y) according to the following equation:

$$\hat{y} = -1.24 + 1.12 x \quad (2)$$

The standard errors for the intercept and slope estimates were 0.59 and 0.18, and the observed significance levels were 0.065 and 0.0001 (t-tests with df = 10), respectively; adjusted $R^2 = 0.80$.

Discussion

Multicollinearity causes regression coefficients to be unstable and exists when the explanatory variables are redundant (Myers 1990). Even small changes in explanatory variables produce large changes in coefficient estimates. However, note that lack of evidence does not confirm that collinearity problems no longer exist. Although other models, with additional variables, were better in terms of adjusted R^2, C_p, s^2, or PRESS statistics, support for the inclusion of additional variables was insufficient, as indicated by tests of significance for regression coefficients and partial regression plots. In addition, although the model with road length alone appeared to suffer from the influence of individual data points, more complex models were as susceptible or more so. The interpretation of the association between reproductive success and road length is biologically meaningful. Bar ditches (ditches along the roadsides) with relatively high densities of small mammals were associated with each road segment, and fence posts lined most roads, providing owl hunting perches. Accordingly, I chose the most parsimonious model.

The analysis suggests that reproductive success and road length are positively associated. However, this conclusion should not be accepted uncritically. Because the database is small, there is little guarantee that the linear association between road length and reproductive success would be evident in another data set. In addition, the power to detect associations of even moderate strength is low. Interspersion,

amount of area in grain, and amount of area in mesquite may be non-linear, however in this paper I focused on linear relationships. Possible non-linearities were not explored for the following reasons. There are hints of associations between reproductive success and other explanatory variables for which form and strength are determined by a single data point (e.g. see Fig. 1B, D). When sample size is small, the estimates of the likely size of chance error in the regression results are imprecise and there is scant basis for checking model assumptions like linearity, normality, and homoscedasticity. In addition, the nature of this data set is observational, not experimental, and cannot reliably determine the mechanisms involved. There may well be unobserved and unmeasured variables that influence barn owl reproductive success. Consequently, care should be taken not to interpret the results as a confirmation of associations without additional and corroborating data.

One of the most profitable uses of small data sets is to generate interesting questions and hypotheses for future studies. The patterns uncovered may suggest general conclusions that allow one to devise experimental field studies. Even further, a 'logical tree' (Platt 1964) can be employed where a hierarchy of groups of hypotheses can be sequentially tested in a mechanistic approach (Price 1986) to discover why the pattern occurs. It is my observation that this step-down approach is sometimes preceded by intriguing results from often-small data sets that have been subjected to pattern analyses of some sort. This, perhaps, is the strongest reason why it is fruitful to explore limited data sets. Exploratory data analysis methods can prove most helpful and may point to scale-sensitive effects that need to be addressed.

Small sample sizes will continue to plague ecologists. I argue that these kinds of data are important, but that extreme care must be taken in both the analyses employed and the interpretations made. I suggest that contingent data analyses procedures promote conservative interpretations and can be used heuristically to illuminate patterns and interesting questions in ecology. As Tukey (1980; cited in Yoccoz 1991) suggested: "finding the question is often more important than finding the answer". I have illustrated one possible approach here.

References

Anderson, D.A. & Burnham, K.P. 1998: Model selection and inference: a practical information theoretic approach. - Springer Verlag, New York, New York, USA, 353 pp.

Ault, J.W., III. 1982: A quantitative estimate of barn owl nesting habitat quality. - M.Sc. thesis, Oklahoma State University, Stillwater, Oklahoma, USA, 33 pp.

Berger, J. 1986: Wild horses of the Great Basin. - University of Chicago Press, Chicago, Illinois, USA, 326 pp.

Cannon, J.R. 1996: Whooping crane recovery: a case study in public and private cooperation in the conservation of an endangered species. - Conservation Biology 10: 813-821.

Clutton-Brock, T.H. & Ball, M.E. 1987: Rhum: the natural history of an island. - Edinburgh University Press, Edinburgh, Scotland, 159 pp.

Freedman, D.A. 1983: A note on screening regression equations. - American Statistician 37: 152-155.

Freedman, L.S. & Pee, D. 1989: Return to a note on screening regression equations. - American Statistician 43: 279-282.

Geist, V. 1971: Mountain sheep. - University of Chicago Press, Chicago, Illinois, USA, 400 pp.

Henderson, H.V. & Velleman, P.F. 1981: Building multiple regression models interactively. - Biometrics 37: 391-411.

Hurlbert, S.H. 1984: Pseudoreplication and the design of ecological field experiments. - Ecological Monographs 54: 187-211.

Hurvich, C.M. & Tsai, C.L. 1989: Regression and times series model selection in small samples. - Biometrika 76: 297-307.

Jenks, S.M. & Wayne, R.K. 1992: Problems and policy for species threatened by hybridization: the red wolf as a case study. - In: McCullough, D.R. & Barrett, R.H. (Eds.); Wildlife 2001: Populations. Elsevier Science Publications, Ltd., pp. 237-251.

Kiff, L.F., Mesta, R.I. & Wallace, M.P. 1996: Recovery plan for the California condor (Gymnogyps californianus). 3rd revision. - U. S. Fish and Wildlife Service, Portland, Oregon, USA, 76 pp.

Levins, R. 1968: Evolution in changing environments. - Princeton University Press, Princeton, New Jersey, USA, 120 pp.

Longmire, J.L., Gee, G.F., Hardekopf, C.L. & Mark, G.A. 1992: Establishing paternity in whooping cranes (Grus americana) by DNA analysis. - Auk 109: 552-529.

Magnusson, W.E. 1983: Use of discriminant function to characterize ruffed grouse drumming sites in Georgia: a critique. - Journal of Wildlife Management 47: 1151-1152.

Maurer, B.A. 1986: Predicting habitat quality for grassland birds using density habitat correlations. - Journal of Wildlife Management 50: 556-566.

May, R.M. 1994: The effects of spatial scale on ecological questions. - In: Edwards, P.J., May, R.M. & Webb, N.R. (Eds.); Large-scale ecology and conservation biology. Blackwell Scientific Publications, London, United Kingdom, pp. 1-17.

McCullough, D.R. 1979: The George Reserve deer herd. - University of Michigan Press, Ann Arbor, Michigan, USA, 271 pp.

Moore, D.E. 1990: The red wolf: extinction, captive propagation and reintroduction. - In: Mitchell, R.S., Sheviak, C.J. & Leopold, D.J. (Eds.); Proceedings 15th Annual Natural Areas Conference. New York State Museum Bulletin 471: 136-139.

Myers, R.H. 1990: Classical and modern regression with applications. 2nd edition. - PWS-Kent Publishing Company, Boston, Massachusetts, USA, 488 pp.

Neter, J., Wasserman, W. & Kutner, M.H. 1985: Applied linear statistical models: regression, analysis of variance, and experimental designs. 2nd edition. - R.D. Irwin, Homewood, Illinois, USA, 1127 pp.

Platt, J.R. 1964: Strong inference. - Science 146: 347-353.

Price, M.V. 1986: Structure of desert rodent communities: a critical review of questions and approaches. - American Zoologist 26: 39-49.

Rotenberry, J.T. 1986: Habitat relationships of shrubsteppe birds: even good models cannot predict the future. - In: Verner, J., Morrison, M.L. & Ralph, C.J. (Eds.); Wildlife 2000: modeling habitat relationships of terrestrial vertebrates. University of Wisconsin Press, Madison, Wisconsin, USA, pp. 217-221.

Sinclair, A.R.E. 1977: The African buffalo. - University of Chicago Press, Chicago, Illinois, USA, 368 pp.

Snedecor, G.W. & Cochran, W.G. 1967: Statistical methods. 6th edition. - The Iowa State University Press, Ames, Iowa, USA, 593 pp.

Sokal, R.R. & Rohlf, F.J. 1995: Biometry. 3rd edition. - W.H. Freeman and Company, New York, New York, USA, 887 pp.

Tilman, D. 1989: Ecological experimentation: strengths and conceptual problems. - In: Lichens, G.E. (Ed.); Longterm studies in ecology. Springer-Verlag, New York, New York, USA, pp. 136-157.

Tukey, J.W. 1980: We need both exploratory and confirmatory. - American Statistician 34: 23-25.

Weatherhead, P.J. 1986: How unusual are unusual events? - American Naturalist 128: 150-154.

Verbyla, D.L. 1986: Potential prediction bias in regression and discriminant analysis. - Canadian Journal of Forestry Research 16: 1255-1257.

Yoccoz, N.G. 1991: Use, overuse, and misuse of significance tests in evolutionary biology and ecology. - Bulletin of the Ecological Society of America 72: 106-111.

Zar, J.H. 1996: Biostatistical analysis. 3rd edition. - Prentice Hall, Upper Saddle River, New Jersey, USA, 662 pp.