Regression versus ANOVA

(Peer-reviewed letter)

In their recent article in *Frontiers*, Cottingham *et al.* (2005) argue that regression-based experimental designs (and analyses) are preferable to those based on ANOVA because of the greater inference gained from regression. We agree completely with the authors and commend them for addressing this important issue. Too often, ANOVA is used to analyze ecological experiments when regression would be more advantageous in terms of both inference and statistical power. Further-more, ecologists commonly rely on experiments with dichotomous treatment levels when multiple-treatment-level experiments (analyzed with regression) would provide stronger inference (Steury *et al.* 2002).

However, we contend that Cottingham *et al.* (2005) overlook the fact that the number and range of treatment levels can influence R^2 and thus power in regression (and ANOVA) and that, consequently, their recommendations for treatment-level selection in experimental design are misguided. When a treatment (independent variable) is continuous and has a proportional (linear) effect on the response (dependent variable), the dispersion in the treatment levels influences the model R^2, and thus the power of both ANOVA and regression. Specifically, R^2 can be expressed as:

$$R^2 = 1 - \frac{SSE}{TSS}$$

where *SSE* is the sum of squares due to error (the dispersion in the response variable that cannot be accounted for by dispersion in the treatment levels) and *TSS* is the total sum of squares (total dispersion) in the response variable. Increasing the dispersion in the treatment levels used in an experiment will also increase the dispersion in the measured response variable (TSS); however, SSE remains unchanged. Therefore, increasing the dispersion in the treatment levels improves R^2 and, consequently, power (note, however, that this conclusion may not hold for non-linear relationships). For example, an experiment with two treatment levels at the extremes in natural variation will have greater dispersion and thus higher R^2 and greater power than an experiment with two or more treatment levels at intermediate intensity (assuming same total sample size; Steury *et al.* 2002). Cottingham and colleagues suggest that the power of these two experimental designs should be equivalent and that power is not a function of the number of treatment levels in either regression or ANOVA; this conclusion is only true if R^2 is identical between experiments. However, the number and distribution of treatment levels (and samples among those levels) certainly affect their dispersion, and thus both R^2 and power. Ecologists should therefore carefully consider the relationship between the distribution of treatment levels and both precision (R^2) and power when designing experiments.

However, precision and power should not be the sole factors considered when selecting treatment levels. As Cottingham *et al.* note, one potential problem with regression is its assumption of linearity between dependent and independent variables. We agree that to address this limitation, experimenters should have replicates at each treatment level, so that lack-of-fit tests can be used to assess linearity. To perform a lack-of-fit test, an experimenter must have at least one more treatment level than the number of parameters in their model (three for linear, four for quadratic, etc; Draper and Smith 1998). Furthermore, the power of the lack-of-fit test is a function of the number of replicates at each treatment level, which influences the "within-treatment-level" variance (Draper and Smith 1998). Armed with this information and an appreciation of the importance of treatment levels to power, our suggestions for treatment-level selection conflict with those proposed by Cottingham *et al.* (2005). These authors suggest that if the assumption of linearity is likely to be upheld, experimenters should choose many treatment levels with few replicates. We argue that if the relationship between treatment and response variables is known to be linear, having many treatment levels is unnecessary, and one should put all replicates in two treatment levels at the extremes in natural variation. This design maximizes R^2 and power. Of course, rarely does one know a priori that a relationship will be linear. Alternatively, Cottingham and colleagues argue that when the assumption of linearity is likely to fail, the chosen experimental design should include few treatment levels, each with many replicates. However, while such a design may maximize power in ANOVA, it may also preclude fitting non-linear curves and conducting lack-of-fit tests. In general, to determine the best experimental design we recommend that: (1) the most parameterized model that may describe the data be determined a priori; (2) the number of treatment levels should be one greater than the number of parameters in that model in the experimental design; and (3) treatment levels should be distributed in a manner that maximizes dispersion, while maintaining the ability to reveal non-linear relationships (Draper and Smith 1998; Steury *et al.* 2002). Such designs should maximize power of both regression and lack-of-fit tests, and facilitate exploration of non-linear fits.

We agree with Cottingham *et al.* that when independent variables are continuous, regression-based experimental designs and analyses are preferable. However, we argue that the number of treatment levels and their distribution have greater importance in experimental design than the authors suggest.

Todd D Steury[1] and Dennis L Murray[2]
[1]*Dept of Biology*
Washington University
St Louis, MO 63130
[2]*Dept of Biology and Dept of Environmental Resource Studies*
Trent University
Peterborough, ON K9J 7B8 Canada

Cottingham KL, Lennon JT, and Brown BL. 2005. Knowing when to draw the line: designing more informative ecological experiments. *Front Ecol Environ* **3**: 145–52.

Draper NR and Smith H. 1998. Applied regression analysis. 3rd edn. New York, NY: Wiley-Interscience.

Steury TD, Wirsing AJ, and Murray DL. 2002. Using multiple treatment levels as a means of improving inference in wildlife research. *J Wildlife Manage* **66**: 292–99.

Cottingham *et al.*'s (*Front Ecol Environ* 2005; **3**: 145–52) recommendation in favor of regression over analysis of variance (ANOVA) left me with serious concerns, because these are two very different tools. Regression is an estimation procedure and is a wonderful tool if one wishes to describe the relationship between variables. Successful use of regression for model development depends on drawing a random sample of paired observations from the population of interest. Sokal and Rohlf (1995) provide a good summary of the uses of regression analysis; Burnham and Anderson (2002) give an excellent overview of how to select the best model.

Analysis of variance, in contrast, is a method for testing hypotheses. If one wants to test a specific hypothesis, then one should choose the number of treatment levels and replicates appropriate for that specific hypothesis (Sokal and Rohlf 1995; Petraitis 1998). Individuals, plots, etc, are randomly assigned to fixed treatment levels, which are controlled by the experimenter. Treatment levels used in an experiment are not randomly drawn from all possible levels, which underscores the distinction between estimation and hypothesis testing. Part of the problem is that Cottingham *et al.* make two common mistakes in their attempt to compare the merits of regression and ANOVA. First, they assume that data collected as part of an ANOVA can be used "as is" in a regression analysis. In a sense, they advocate pooling sources of variation to increase degrees of freedom, and thus power. This is not correct, and is a form of sacrificial pseudoreplication (Hurlbert 1984). A regression analysis can be done within an ANOVA, but only as a linear contrast that is nested within the ANOVA (Sokal and Rohlf 1995; Petraitis 1998). For example, a linear regression within an ANOVA with six treatment levels and 24 experimental units (as in Cottingham *et al.*'s Figure 1) has one and four degrees of freedom, not one and 22. The power of a linear regression done within an ANOVA will be similar to the power of a simple linear regression if done correctly and matched with the correct degrees of freedom. Second, Cottingham *et al.* incorrectly assume that power of different designs can be compared in a meaningful way. Petraitis (1998) provides several examples of how effect size, R^2, and power depend not only on the number of replicates, but also on the number of treatment levels.

More than 20 years ago, Hurlbert (1984) lamented the lack of statistical sophistication among ecologists and its effect on the field. Assuming Cottingham, her two co-authors, more than 12 acknowledged colleagues, at least two reviewers, and an editor represent a sample randomly drawn from the population of well-trained ecologists in the US, one might infer that not much has changed.

Peter S Petraitis
Dept of Biology
University of Pennsylvania
Philadelphia, PA

Burnham KP and Anderson DR. 2002. Model selection and multimodel inference: a practical information–theoretic approach. 2nd edn. New York, NY: Springer-Verlag.

Hurlbert SH. 1984. Pseudoreplication and the design of ecological field experiments. *Ecol Monogr* **54**: 187–211.

Petraitis PS. 1998. How can we compare the importance of ecological processes if we never ask, "compared to what?" In: Resetarits Jr WJ and Bernardo J (Eds). Experimental ecology: issues and perspectives. New York, NY: Oxford University Press.

Sokal RR and Rohlf FJ. 1995. Biometry. 3rd edn. New York, NY: WH Freeman and Company.

Cottingham *et al.* (*Front Ecol Environ* 2005; **3**: 145–52) consider the choice between discrete and quantitative versions of an explanatory variable in designing an experiment, and conclude that "regression [using a quantitative predictor] is generally a more powerful approach than ANOVA [using a discrete predictor]". Because of the way they choose to specify the alternative in their power calculations, however, their work amounts to showing that, given two models that "explain" the same amount of variability in the response, the one based on fewer parameters is preferred – not a very novel conclusion.

The point is that the two approaches will not, in general, have the same explanatory power. Depending on how linear the relationship is between predictor and response, the extra variability explained by the ANOVA model may or may not be enough to counterbalance the degrees of freedom it uses up, compared to the simpler regression model. One approach is not inherently more powerful than the other. These ideas are discussed in many statistics textbooks (eg Ramsey and Schafer 2002).

Paul Murtaugh
Dept of Statistics
Oregon State University
Corvallis, OR

Ramsey FL and Schafer DW. 2002. The statistical sleuth: a course in methods of data analysis. Pacific Grove, CA: Duxbury.

The authors reply

We are pleased that our review (*Front Ecol Environ* 2005; **3**: 145–52) on regression and ANOVA has generated such spirited discussion regarding the design of more effective ecological experiments. We agree with Murtaugh that our conclusions in favor of regression are "not … very novel" – they should come as no surprise to statistically savvy ecologists, particularly those weaned on newer textbooks, such as Ramsey and Schafer (2002) and Gotelli and Ellison (2004). Unfortunately, the major points raised in our review are not discussed in classic biostatistics texts (eg Sokal and Rohlf 1995), and it is clear that not all ecologists believe regression is appropriate for experimental data (see comment by Petraitis).

Petraitis argues that regression is an estimation procedure that cannot be used to test hypotheses. There is no theoretical or mathematical rationalization for this view. As explained in our paper, regression and ANOVA share the same underlying mathematical framework (see Web-only Appendix 1 of our paper) and differ only in how they are applied. Either approach can be used to test hypotheses, as long as the treatment levels are under the control of the investigator. Petraitis also suggests that using regression to analyze experimental data involves "sacrificial pseudoreplication". As defined by Hurlbert (1984), and clarified by Quinn and Keough (2002), pseudoreplication refers to the lack of independence caused by subsampling experimental units; we certainly do not advocate this. Petraitis specifically contends that we "[pool] sources of variation to increase degrees of freedom and thus power". We show that this is not the case in Table 2a of Web-only Appendix 2; the extra degrees of freedom gained from replicate samples at each level of X are appropriate in testing for a linear effect when there is no lack-of-fit (see also Draper and Smith [1998] and the comment from Steury and Murray above).

A common theme running through all three comments is the use of R^2 to generate power curves. Murtaugh notes that regression and ANOVA "will not in general have the same explanatory power". As explained in our Web-only Appendix 2, nonlinearity in the relationship between X and Y is captured by the lack-of-fit sums-of-squares, which become part of the error term in regression and part of the model term in ANOVA. R^2 will therefore always be bigger for ANOVA than for regression by an amount proportional to this lack-of-fit term (Table 2b). Of course, regression is not appropriate when the X–Y relationship is not linear, which is why regression is more powerful than ANOVA only in situations when the assumptions of both tests are met. Steury and Murray, and Petraitis, correctly critique our claim that R^2 for regression designs does not depend on the numbers of replicates and treatment levels. Importantly, Steury and Murray explain why this is the case and provide additional recommendations regarding the design of replicated regression experiments in different research situations. We encourage readers to study these recommendations carefully.

Clearly, the use of regression to analyze experimental data is a controversial topic for some ecologists. This controversy may stem from historical biases in the field of ecology, which have favored ANOVA in experimental studies. However, our review demonstrates that regression is often equally applicable, and in many cases superior to ANOVA. Because the printed text of our paper was written to be readily accessible to all readers, including those with little background in statistics, many of the statistical details supporting our recommendations appear online in the web-only materials. Our critics may have missed these. We therefore en-courage interested readers to read the web-only appendices carefully and, most importantly, to decide for themselves what statistical approach will be most appropriate for their research questions.

KL Cottingham[1], JT Lennon[2], and BL Brown[3]
[1]*Dept of Biological Sciences*
Dartmouth College
Hanover, NH 03755
[2]*Dept of Ecology and Evolutionary Biology*
Brown University
Providence, RI 02912
[3]*Integrative Biology*
University of Texas at Austin
Austin, TX 78712

Draper NR and Smith H. 1998. Applied regression analysis. New York, NY: Wiley-Interscience.
Gotelli NJ and Ellison AM. 2004. A primer of ecological statistics. Sunderland, MA: Sinauer Associates Inc.
Hurlbert SH. 1984. Pseudoreplication and the design of ecological field experiments. *Ecol Monogr* **54**: 187–211.
Quinn GP and Keough MJ. 2002. Experimental design and data analysis for biologists. New York, NY: Cambridge University Press.
Ramsey FL and Schafer DW. 2002. The statistical sleuth: a course in methods of data analysis. Pacific Grove, CA: Duxbury.
Sokal RR and Rohlf FJ. 1995. Biometry. New York, NY: WH Freeman and Company.