Knowing when to draw the line: designing more informative ecological experiments

Kathryn L Cottingham[1], Jay T Lennon[1,2], and Bryan L Brown[1,3]

Linear regression and analysis of variance (ANOVA) are two of the most widely used statistical techniques in ecology. Regression quantitatively describes the relationship between a response variable and one or more continuous independent variables, while ANOVA determines whether a response variable differs among discrete values of the independent variable(s). Designing experiments with discrete factors is straightforward because ANOVA is the only option, but what is the best way to design experiments involving continuous factors? Should ecologists prefer experiments with few treatments and many replicates analyzed with ANOVA, or experiments with many treatments and few replicates per treatment analyzed with regression? We recommend that ecologists choose regression, especially replicated regression, over ANOVA when dealing with continuous factors for two reasons: (1) regression is generally a more powerful approach than ANOVA and (2) regression provides quantitative output that can be incorporated into ecological models more effectively than ANOVA output.

Front Ecol Environ 2005; 3(3): 145–152

Designing informative ecological experiments can be a very challenging endeavor, particularly for researchers studying continuous independent variables for which either analysis of variance (ANOVA) or linear regression could be used to analyze the results. Although ecologists have relied heavily on ANOVA to design and analyze their experiments for most of the past century, there are many reasons to use regression-based experimental designs (cf Gotelli and Ellison 2004). The aim of this review is to demonstrate why ecologists should prefer experiments designed for analysis with regression, when appropriate.

ANOVA-based experiments are designed to answer qualitative questions, such as, "Does the response variable (Y) differ across different levels of the independent variable(s) (X)?" and "If there *are* differences in Y, which treatments are different?" Typically, X is either a discrete variable (eg type of disturbance; the presence/absence of predators, disease, or an invasive species) or a continuous variable that can be readily grouped into categories or classes (eg low, medium, and high levels of nitrogen; low and high population density).

In contrast, regression-based experiments are designed to answer the quantitative question, "How does the response variable change with the independent variable(s)?" by building a model that describes the shape of the relationship between X and Y using as few parameters as possible. Regression is therefore appropriate only for continuous independent variables, such as environmental characteristics that lie along gradients (eg light, pH, temperature, nutrient concentration, disturbance frequency) or continuous biological characteristics (eg species richness, organism size, disturbance magnitude).

For some ecological research scenarios, the choice between designing an experiment for analysis with ANOVA or regression is relatively straightforward. For example, ANOVA is the only appropriate approach for studying factors that cannot be made continuous (eg male versus female; genotypes that are sensitive to a pathogen versus those that are resistant), while regression is the most appropriate approach when the research question involves building a quantitative model to describe the relationship between X and Y. ANOVA is also a useful starting point for new empirical research, such as testing a specific hypothesis about the effect of X on Y derived from a theoretical model. More generally, simple ANOVA experiments to refute or accept a proposed effect allow a researcher to determine whether that factor is worthy of further investigation (see Case Study Panel 1).

There are, however, many ecological scenarios for which either ANOVA or regression would be appropriate. In such cases there is at least one independent vari-

In a nutshell:

- Analysis of variance (ANOVA) and linear regression are widely used by ecologists, but surprisingly little information is available regarding their relative merits
- As linear regression is more powerful than ANOVA and provides quantitative information that can be used to build ecological models, we suggest that ecologists use regression whenever possible
- In particular, replicated regression designs provide the flexibility to analyze data with regression when appropriate and with ANOVA otherwise

[1]*Department of Biological Sciences, Dartmouth College, Hanover, NH 03755 (cottingham@dartmouth.edu);* [2]*Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912;* [3]*Integrative Biology, University of Texas at Austin, Austin, TX 78712*

Case Study Panel 1. Julie's dilemma

Julie is a second-year graduate student trying to decide how to set up her next field experiment. Last summer, she conducted a number of preliminary studies; the most promising evaluated the effects of plant community diversity and light on ecosystem processes using a 2 x 2 factorial experiment in aquatic mesocosms. Although no effect of diversity was detected, the light treatment – shaded versus unshaded – had modest but interesting effects on several key response variables, including ecosystem respiration. Julie has therefore decided to evaluate the effect of light more thoroughly this year, but she is not sure how to design the experiment. Because she has a fixed number of mesocosms (24) available for her study, she faces an important decision about allocating experimental units to treatments (light levels) versus replicates. Should she repeat the design from last year, with just two levels of light, to maximize her power to detect an effect of light? Or should she create a gradient of light levels in order to map how her response variables vary with light? If she has more than two light levels, how many should she have, and how should they be selected? Moreover, because light is a continuous variable, should she plan to analyze her results with ANOVA, linear regression, or some combination of these approaches? This paper attempts to provide guidance to Julie and others who are faced with tough decisions about designing ecological experiments.

able that could be considered as either a continuous or discrete variable, depending on context, and the research question is flexible enough to be explored as either a regression or an ANOVA problem. For example, in Case Study Panel 1, Julie plans to evaluate the effect of light on ecosystem respiration, but she has not yet refined her research question to the point where the choice between regression and ANOVA is obvious. Julie can therefore define light as a discrete variable (by having shaded versus unshaded treatments) or a continuous variable (by using shade cloths that pass different percentages of the incident light). In a situation like this, what are the advantages and disadvantages of choosing a regression-based design versus an ANOVA-based design?

This review provides concrete suggestions for choosing between regression- and ANOVA-based experiments for research questions involving at least one continuous independent variable and for which the choice of approach is not dictated by the research question. We begin with an overview of the general linear model that underlies both techniques. We then make a head-to-head comparison between the power of regression and ANOVA models before introducing replicated regression, an approach that maximizes both power and flexibility. Throughout, we use the main text to make our major messages accessible to all readers, Case Study Panels to apply our findings (eg Case Study Panel 1), and Statistical Panels to provide details for interested readers (eg Statistical Panel 1).

Some key information about regression and ANOVA

Although most introductory statistics courses make a clear distinction between fitting a curve to data (regression) versus testing for differences between treatment means (ANOVA), few point out the underlying similarity between these techniques. ANOVA and linear regression share the same underlying mathematical model, the general linear model, which is expressed in matrix form as $Y = X\beta + \epsilon$ (Web-only Appendix 1; Neter *et al.* 1996). In this model, Y represents the response variable, X a matrix of the independent variable(s), β the parameters associated with each independent variable, and ϵ the errors. The matrix of independent variables X determines whether we are performing a regression or an ANOVA. In regression, the X matrix contains only continuous variables, while ANOVA uses only discrete variables (sometimes called "indicator" or "dummy" variables). The elements of the β matrix of a regression quantify the shape of the relationship between Y and X, while the elements of the β matrix of an ANOVA provide information about treatment means. Alternatively, the X matrix can contain a mix of discrete and continuous variables, allowing researchers to compare the shapes of relationships across different treatment groups (eg ANCOVA and indicator variables regression; Neter *et al.* 1996); we do not address these intermediate cases here.

Although they have the same underlying mathematical framework, regression and ANOVA are different in several fundamental ways. For example, because these techniques address different questions (or, alternatively, test different hypotheses), their underlying assumptions are subtly different (Statistical Panel 1). Most importantly, the general linear model assumes that the relationship between Y and X can be described using a linear equation (Neter *et al.* 1996), so that regression is inappropriate when the relationship cannot be made linear in the parameters (eg through transformations or polynomial terms). In contrast, ANOVA does not assume any particular relationship between Y and X, and so is appropriate even when the response to the independent variable(s) is highly nonlinear.

Another key difference between regression and ANOVA lies in the number of columns used to define the X matrix, which determines the number of parameters in the general linear model. Given a particular experimental design, the X matrix for ANOVA generally has more columns than the X matrix for regression

Statistical Panel 1. Assumptions and their violation

Here we highlight the major assumptions for regression and ANOVA. Violation of some of these assumptions can be a serious issue, leading to erroneous or biased conclusions, while violations of other assumptions may be less serious.

Response variable and residuals

Both regression and ANOVA assume that the response variable Y and residuals ϵ are independent, normally distributed random variables with the same variance (homoskedastic). Importantly, analysis of the residuals ϵ, not the response variable Y, is the best way to test these assumptions for both regression and ANOVA (Neter *et al.* 1996; Quinn and Keough 2002; Kéry and Hatfield 2003). If the residuals meet the assumptions of normality and equal variance, then the underlying rules of probability calculus imply that the response variable was also normally distributed and homoskedastic (Larsen and Marx 1986; Miller 1986).

Unequal variance (heteroskedasticity) can be extremely problematic in both regression and ANOVA. With regression, strong heteroskedasticity causes the variance around the estimated slope and intercept to be underestimated (Miller 1986), potentially leading to overestimates of statistical significance. In ANOVA, heteroskedasticity alters the assumptions underlying the F-test and may cause the P value to be over- or underestimated (Miller 1986). Most researchers cope with heterogeneous variances through transformations, commonly a logarithmic or root transformation for residuals that funnel out or a reciprocal transformation for residuals that funnel in (Neter *et al.* 1996). Importantly, moderate violation of homoskedasticity can be ignored in balanced ANOVA designs (those with equal numbers of replicates for each treatment), because the bias in the P value is small (Box 1954a,b). In regression designs, quantile regression can be a powerful tool for dealing with heteroskedasticity (Cade and Noon 2003).

Failure to meet the normality assumption is usually of minimal concern in both ANOVA and regression, unless the errors are highly non-normal (eg skewed). The F-tests used in ANOVA and regression tend to be robust to non-normal errors, except when an experiment is highly unbalanced, although power may be reduced by non-normality (Miller 1986; but see Wilson and Grenfell 1997). Moreover, parameter estimates from regression analyses are robust to non-normality, except when the non-normality is due to outliers (Miller 1986). Importantly, when errors are highly non-normal, generalized linear models need to be used instead of regression or ANOVA (eg McCullagh and Nelder 1997; Wilson and Grenfell 1997).

More generally, outliers that cause skew, unequal variance, or non-normality in the errors are extremely problematic and need to be dealt with carefully (Miller 1986).

Independent variable(s)

Unlike the rigid distributional assumptions for Y and ϵ, neither regression nor ANOVA make assumptions about the distribution(s) of the independent variable(s) X. Thus, X does not need to be normally distributed in order to proceed with regression. However, the independent variables need to be either controlled by the researcher or measured as accurately as possible.

Imprecise or inaccurate estimates of the independent variables are a particular concern for regression, which explicitly assumes that all predictors are measured without error, or at least with much less error than the response variable Y. Violation of this assumption leads to "errors in variables" (EIV) and biased parameter estimates. For example, in simple linear regression, EIV bias regression slopes towards zero (Sokal and Rohlf 1995; McArdle 2003), potentially altering biological conclusions and complicating the use of regression models in further research.

because ANOVA requires each treatment to be identified using a separate column of X (Web-only Appendix 1). To make this statement more concrete, consider our case study. Suppose that Julie set up her mesocosm experiment to quantify the effects of light on ecosystem respiration using five levels of light. A simple linear regression to account for light effects would have two columns in X, corresponding to the intercept and slope. On the other hand, a one-way ANOVA model for the same experiment would require five columns, each specifying the mean for a treatment. This difference in the number of parameters grows more extreme as the number of treatments increases. For example, suppose Julie added temperature as a second factor, such that she had two levels of temperature and five levels of light. A typical multiple regression model would have four parameters (intercept, main effects of light and temperature, and a light x temperature interaction), while the two-way ANOVA would require ten parameters (grand mean, four parameters for light effects, one parameter for temperature effects, and four light x temperature interactions).

The relative power of regression and ANOVA

This difference in the number of parameters leads us to one of the most important take-home messages from this review: because regression requires fewer parameters, it is generally a more powerful statistical approach than ANOVA. Statisticians define power as the probability of detecting an effect when that effect is present (ie the probability of rejecting the null hypothesis when the null hypothesis is false). In regression, the null hypothesis is that Y is not predicted by a specific linear function of X, while in ANOVA, the null hypothesis is that treatments do not differ. The power for the overall F-test is calculated in the same way for all general linear models (Statistical Panel 2); we used this procedure to generate power curves (graphs showing how the ability to detect an effect changes with effect size) for a variety of one- and two-way experimental designs (Figure 1). Several interesting features emerged from this analysis:

(1) The power curve for ANOVA is determined by the number of replicates per treatment, as power increases with increased replication (Figure 1). This should come as no surprise to anyone who has taken a course in experimental design. If the number of experimental units is fixed by logistical constraints, power increases when these units are allocated to fewer treatments with more replicates per treatment. Moreover, the power for the overall F-test is determined by the total number of treatment combinations, not the number of factors (independent vari-

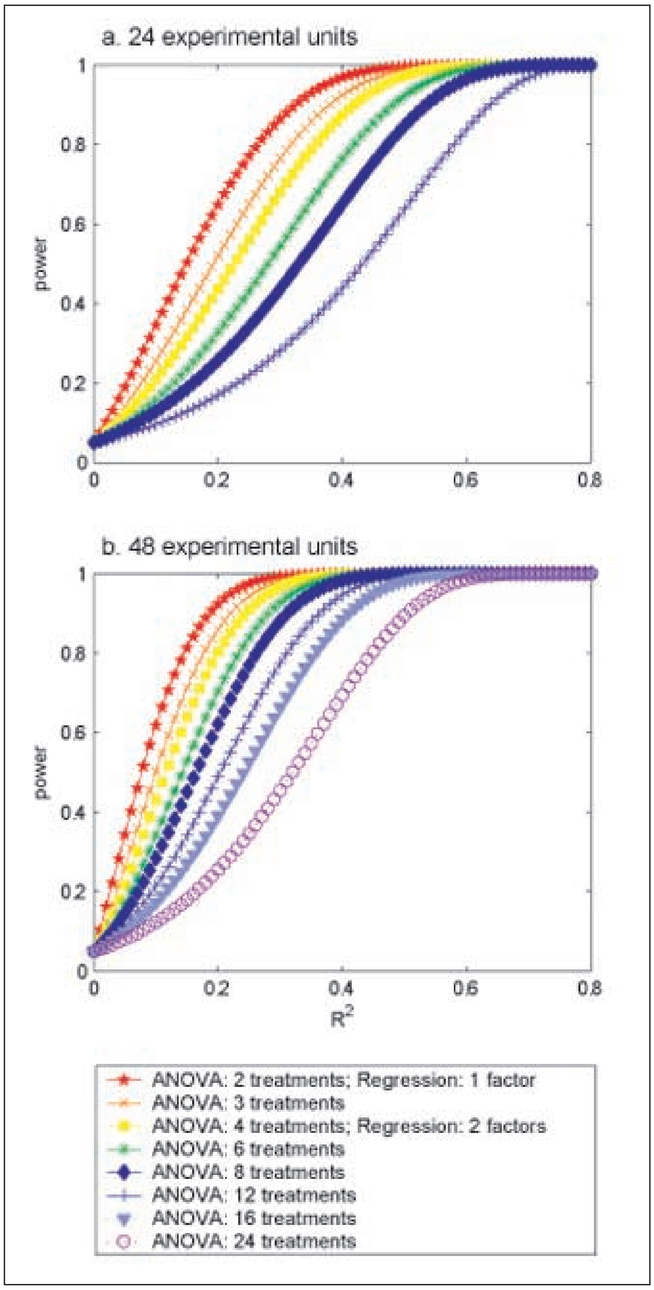

Figure 1. *Power curves for all possible balanced one- and two-way regression and ANOVA models when there are (a) 24 and (b) 48 experimental units. Identifying information for each curve is provided below the figure; the number of replicates per treatment can be determined by dividing the number of experimental units by the number of treatments. Note that regression generally has greater power than ANOVA, except in the special case where the ANOVA only involves two levels per factor.*

ables) or the number of levels of each factor (Statistical Panel 2). Thus, an experiment with eight levels of Factor A has the same power curve as an experiment with four levels of Factor A crossed with two levels of Factor B, or a three-way experiment with two levels of each factor.

(2) The power curves for regression are determined by the number of factors and the number of experimental units, but not the number of treatments or replicates (Figure 1). Given a fixed number of experimental units, the regression power curve is determined by the number of factors (compare the red and yellow lines in Figure 1). Given a fixed number of factors, power increases with the number of experimental units (compare lines with the same colors in the top and bottom panels of Figure 1). It is only in ANOVA that the allocation of experimental units to treatments versus replicates determines power.

(3) When there are only two levels per factor, the power of ANOVA is always equivalent to the power of regression because both have the same number of parameters. Thus, a one-way ANOVA with two levels of the independent variable has the same power as a simple linear regression (red lines in Figure 1), while a two-way ANOVA with two levels per factor has the same power as a multiple regression model with main effects and an interaction (yellow lines in Figure 1).

(4) For all other designs, regres-sion is more powerful than ANOVA. In designs with one factor, simple linear regression is more powerful than ANOVA, unless there are just two levels of the factor. Similarly, in designs

Statistical Panel 2. Details for power calculations

The power of the overall F-test for regression and ANOVA is calculated in the same way (Cohen 1988), as long as the ANOVA considers fixed effects and the regression is with little error in X. As with all power calculations, we begin by specifying the null (H_o) and alternative (H_a) hypotheses of interest and the significance level α for rejecting H_o. The null hypothesis in either case is that the variability in Y is due to chance rather than biological differences – that is, H_o: $R^2_{Y\bullet\beta} = 0$, where $R^2_{Y\bullet\beta}$ indicates the fraction of variation in Y explained by the model with parameters β. We express H_a as a function of the minimum variability explained by the model (a minimum $R^2_{Y\bullet\beta}$) thought to be of biological significance. We then translate the target $R^2_{Y\bullet\beta}$ into an effect size f^2 using the ratio of explained to unexplained variance:

$$f^2 = R^2_{Y\bullet\beta} / (1 - R^2_{Y\bullet\beta})$$

The critical value of the F-statistic (F_{crit}) that will cause us to reject H_o is determined from α and the numerator (u) and denominator (v) degrees of freedom (df) for the particular experimental design used. Because the total number of treatments determines u and v in the overall F-test (see table at the end of Web-only Appendix 1), there is no change in the power curves when there are multiple factors under investigation.

Given u, v, and a target f^2 (H_a), we calculate the non-centrality parameter λ of the non-central F-distribution with u,v df as

$$\lambda = f^2 (u + v + 1)$$

Finally, we calculate the power of the overall F-statistic as one minus the probability associated with the non-central F-distribution at the value specified by F_{crit}, u, v, and λ.

The power curves in Figure 1 were generated using this algorithm implemented in Matlab 6.5 (MathWorks, Natick, MA). For a particular experimental design, we calculated u, v, and F_{crit} for both the regression and ANOVA models. We then determined λ and power given these values for all effect sizes corresponding to R_2 from 0 to 1 at steps of 0.05. Our programs and data files with the power curves are available online (Web-only Appendix 4).

with two factors (ie at least four treatments), regression is more powerful than ANOVA unless the design is a 2 x 2 factorial.

Based on the above findings, we recommend that ecologists use regression-based experimental designs whenever possible. First, regression is generally more powerful than ANOVA for a given number of experimental units (Figure 1). Second, regression designs are more efficient than ANOVA designs, particularly for quantifying responses to multiple factors (Gotelli and Ellison 2004). Third, regression models have greater information content: regression results can be readily incorporated into theoretical ecological models (eg Aber *et al*. 1991) or used to make empirical predictions for new systems (eg Meeuwig and Peters 1996). Modelers frequently bemoan the lack of empirical data to develop equations and parameters for simulation studies (Canham *et al*. 2003), and a greater emphasis on regression-based designs may help to fill this gap (Gotelli and Ellison 2004).

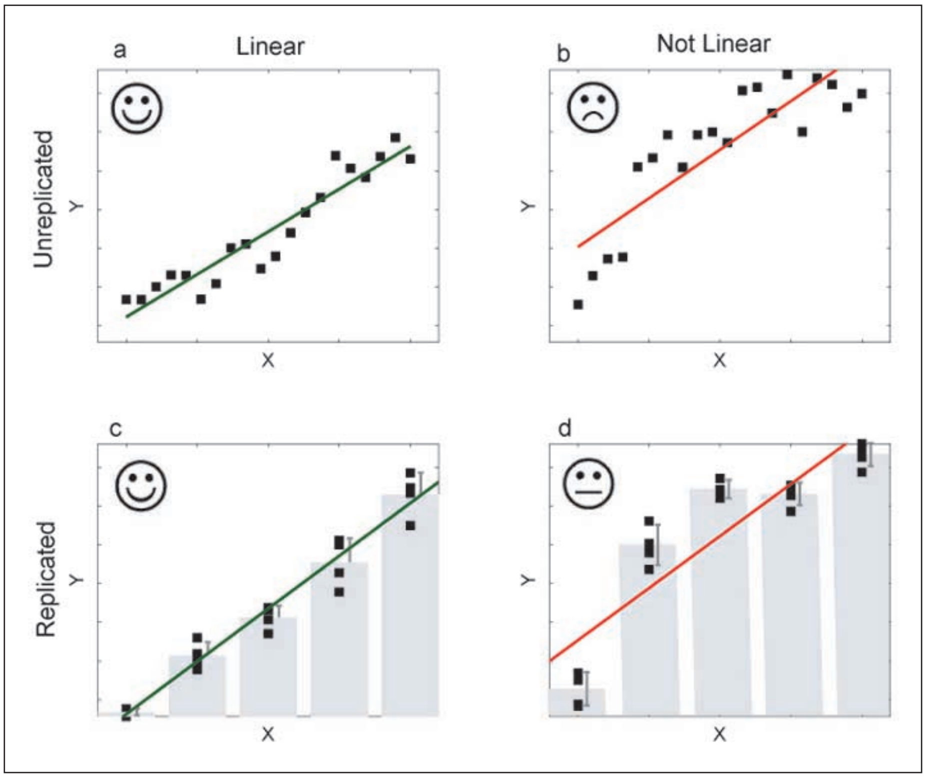

Figure 2. *Contrasting outcomes for replicated versus unreplicated regression-based experimental designs. In the left column, data were simulated using a linear relationship; in the right column, data were simulated using a sigmoidal relationship. In the top row, each level of X is unreplicated, so it is not possible to "fall back" to ANOVA when linear regression is not appropriate (b). However, in the bottom row, there are replicate observations at each level of X, allowing us to use ANOVA to test for differences in mean response across levels of X (grey bars ± 1 SE) – particularly when linear regression is not appropriate (d).*

It should be remembered, however, that regression is not appropriate in all situations. For example, standard linear regression is inappropriate when there are thresholds and non-linearities in the data that cannot be accommodated by a linear model or transformations (Figure 2; Web-only Appendix 2), or when there are measurement errors in one or more independent variables ("errors-in-variables"; Statistical Panel 1). Because these situations are not uncommon, a regression design that does not replicate treatments can be risky (Case Study Panel 2). This makes replicated regression experiments (Figures 2c and d), which provide the flexibility to analyze the resulting data with either regression or ANOVA, extremely attractive.

Case Study Panel 2. Choosing between regression and ANOVA

After seeing Figure 1, Julie becomes very enthusiastic about using a regression design for her field experiment. She decides that she should monitor changes in ecosystem respiration across 12 different levels of light (with two replicate mesocosms per level) and then analyze the results with linear regression. Pleased with herself, Julie goes to her advisor to explain her proposed design. A self-described "ANOVA-head", the advisor asks Julie to briefly justify why she has chosen this particular design. Julie argues that:

- By using more levels of light, she'll be able to better describe exactly *how* respiration changes with light.
- Her regression relating respiration and light could become part of a simulation model to evaluate how aquatic ecosystem respiration might respond to changes in cloud cover predicted by global warming.

Julie's advisor concedes that these are both worthy points, but then asks a single, pointed question: "What will you do if the relationship between ecosystem respiration and light cannot be described using a linear model?" At this point, Julie realizes that a regression-based experiment might be more complicated than she realized.

Replicated regression: a powerful hybrid

Replicated regression (RR) combines the pattern-distinguishing abilities and statistical power of regression with ANOVA-like replication of treatments (Figure 2). In RR designs, researchers make multiple independent observations of the response variable for at least some values of the independent variable(s). Here, we focus on the case where there are equal numbers of replicates for every treatment because balanced designs give unbiased results even with some heterogeneity in error variance (Statistical Panel 1). Because the regression power curve is determined by the number of experimental units, and not the number of replicates per treatment, allocating some experimental units to replication increases analytical flexibility without decreasing power.

RR designs make it possible to use lack-of-fit tests to evaluate the appropriateness of a regression model (Web-only Appendix 2) and/or use ANOVA as a "fall back" analysis when data violate the assumptions of standard linear regression (Figure 2). When there are thresholds and non-linearities in the response variable, nonlinear regression (eg Draper and Smith 1998), piecewise regression (eg Toms and Lesperance 2003), and quantile regression (eg Cade and Noon 2003) are often valid alternatives. However, many ecologists are unfamiliar or uncomfortable with these approaches. For these researchers, ANOVA is also a valid alternative, but only if the experiment included replicates at some levels of X.

"Falling back" to ANOVA almost always entails a reduction in statistical power (Figure 1), but it is possible to design experiments such that regression can be used to analyze the results when the resulting data are appropriate and ANOVA when they are not, without sacrificing too much statistical power (Case Study Panel 3). Designing such experiments requires balancing two competing needs: having enough levels of the independent variable(s) X to fit a meaningful quantitative model while at the same time protecting against the possibility of non-linearity or errors-in-variables by having more replicates at each level of X. Decisions about this tradeoff should be based on the following criteria:

Case Study Panel 3. Planning for a "fall back" ANOVA

After thinking about her advisor's comment, Julie realizes that a regression experiment with only two replicates per treatment might not be the best choice. She has 24 experimental units available for her experiment, so she decides to evaluate all of the options: she can have 12, 8, 6, 4, 3, or 2 treatments with 2, 3, 4, 6, 8, or 12 replicates, respectively.

Julie first decides that she would definitely like to know something about the shape of the response of respiration to light, and therefore needs at least four light treatments. Next, she admits that she knows little about how linear the response to light might be, since last year she only had two light levels. However, it seems reasonable to expect some sort of saturating function, based on plant physiology: at high light levels, physiological processes probably become limited by some other factor. She concedes that her advisor was right – she needs to plan for the contingency of a non-linear relationship. Moreover, because she wants her results to be publishable, regardless of the analysis used, Julie aims for a minimum ANOVA power of at least 0.8. She knows from her experiments last summer that the variability among replicate mesocosms (the sums-of-squares due to pure error, or SSPE; see Web-only Appendix 2) could be quite high, accounting for as much as 50–60% of the overall variability in the response variable (the total sums-of-squares, or SST).

Armed with this information, Julie consults Figure 3. She decides to ensure a power of 0.8 with 24 experimental units (blue curves) by using six treatments of four replicates each. This design will allow her the flexibility to "fall back" to ANOVA should she encounter a saturating response, but it also provides her with enough levels of the independent variable to reasonably map out a response, potentially using nonlinear regression. With six treatments, Julie needs an overall $R^2 > 0.4$ or a SSPE/SST ratio <0.6 to achieve a power of 0.8. She feels that these constraints are reasonable, given her results from last summer.

Julie returns to her advisor with this revised design, and they agree that a 6-level, 4-replicate design is an appropriate compromise between the potential power of the linear regression approach and the possible scenario requiring a "fall back" ANOVA.

The importance of building a quantitative model for the relationship between X and Y

When the primary research objective is to develop a predictive model for Y, then sampling as many levels of the independent variables as possible should be given the highest priority. In this situation, we recommend "falling back" to alternative regression models (eg nonlinear, piecewise, or quantile regression) instead of ANOVA, because ANOVA is unlikely to yield satisfactory conclusions.

The potential size of the experiment

The number of experimental units dictates the potential power of the regression analysis, as well as the list of potential RR designs. Generally speaking, the more experimental units there are, the more powerful the analysis will be, although logistical constraints usually provide an upper boundary on experiment size.

The probability of a regression model being inappropriate

If problems with regression are unlikely (see Statistical Panel 1), we suggest having more treatments and fewer replicates per treatment. However, when there may be problems with regression, we recommend adopting a design with fewer treatments and more replicates per treatment. The likelihood that regression will be inappropriate can be estimated by studying the literature, as well as by intuition and pilot experiments (see Case Study Panels 2 and 3 for an example).

The expected variability among replicates

As with all power analyses, an *a priori* estimate of variability within treatments is necessary (Quinn and Keough 2002). The greater the expected variability, the stronger the need for more replicates. In particular, a rough estimate of the expected ratio of the variability within treatments to the overall variability in the response variable can be used to choose between alternative RR designs (Case Study Panel 3; Web-only Appendices 2, 3).

The desired power of the "fall back" ANOVA

To ensure that a "fall-back" ANOVA has high power, a researcher should increase the number of replicates and

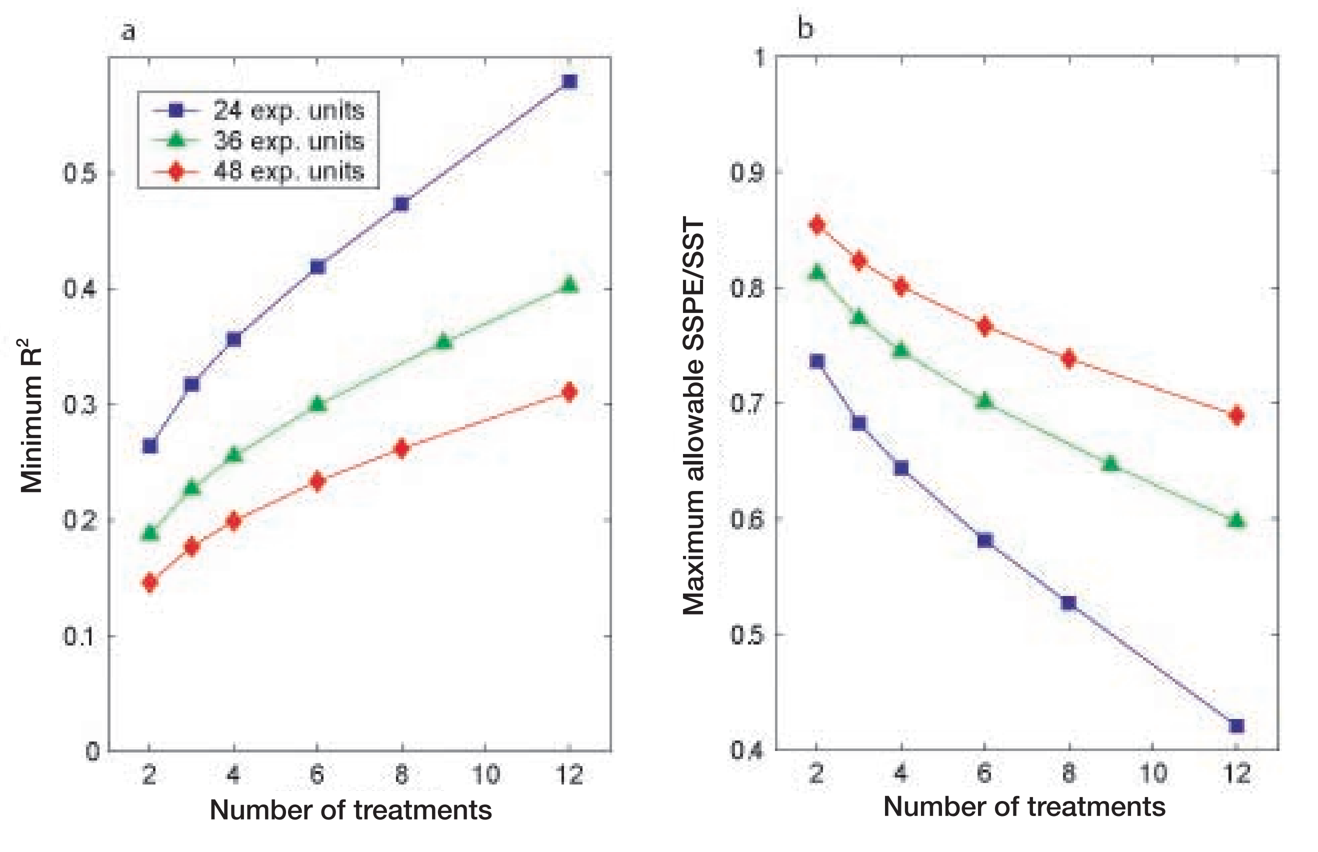

Figure 3. *Guidelines for choosing between possible replicated regression designs when 24, 36, or 48 experimental units are available. Lines show (a) the minimum required R^2 or (b) the largest possible allowable SSPE/SST for the target power level 0.8; line symbols and colors indicate the number of experimental units under consideration. To generate similar figures for other sample sizes or powers, see Web-only Appendix 3.*

decrease the number of treatments. The exact number of treatments and replicates required to meet a particular minimum power demand can be determined using power curves together with an estimate of the expected variability in the system (see Case Study Panel 3).

A cautionary note

Readers should be aware that there are situations for which the general linear model is inappropriate, prohibiting the use of either ANOVA or linear regression. For example, highly non-normal errors require generalized linear models, which allow for a diversity of error distributions (eg log-normal, Poisson, or negative binomial; McCullagh and Nelder 1997; Wilson and Grenfell 1997). It is currently impossible to state whether our conclusions regarding the relative power of regression and ANOVA also extend to generalized linear models, since calculations of power for such models are still in their infancy. However, we hypothesize that our conclusions will hold for this more general class of models, since regression models will include fewer parameters than ANOVA models for all but the simplest experiments. Testing this hypothesis is an important area for further research.

Conclusions

This review was motivated by a perceived shortage of information about the relative merits of regression- and ANOVA-based experiments when there is at least one continuous variable and the research question can be answered with either regression or ANOVA. Many current ecological questions fall into this category, including investigations of the relationships between species richness and ecosystem functioning (eg Loreau *et al*. 2001) and between metabolic rate and population/community parameters (eg Brown *et al*. 2004). To aid researchers working on these and other questions, we have shown that:

(1) Regression and ANOVA are more similar to one another than they are different. The key distinction is that regression builds a quantitative model to describe the shape of the relationship between X and Y, using as few parameters as possible.
(2) In testing the assumptions of regression and ANOVA, homogeneity of variance tends to be far more critical than normality for most ecological variables (Statistical Panel 1).
(3) Regression is generally more powerful than ANOVA, and also provides additional information that can be incorporated into ecological models quite effectively.
(4) Because unreplicated regression designs can be risky,

we recommend replicated regression designs that allow researchers to use either regression or ANOVA to analyze the resulting data.

(5) In replicated regression, how experimental units are allocated to treatments versus replicates has a major effect on the overall power of the "fall back" ANOVA. Decisions about the numbers of treatments should be based on the tradeoff between building a quantitative model and allowing for the possibility of falling back to ANOVA if necessary. To help ecologists choose among the alternatives, we have provided an example (Case Study Panel 3) and instructions for drawing Figure 3 for other design scenarios (Web-only Appendix 3).

Acknowledgments

Many people have provided constructive feedback on previous drafts of this manuscript, including R Thum, M McPeek, J Butzler, A Dawson, M Donahue, N Friedenberg, J Kellner, M Ayres, and J Pantel. J Aber, C Canham, J Pastor, and others who attended Cary Conference IX stimulated our thoughts about the role of regression-based designs in contributing to the development of ecological models. Our research is supported by NSF-DEB 0108474, NSF-DDIG 0206531, USGS/NIWR 2002NH1B, and the Cramer Fund at Dartmouth College.

References

Aber JD, Melillo JM, Nadelhoffer KJ, *et al*. 1991. Factors controlling nitrogen cycling and nitrogen saturation in northern temperate forest ecosystems. *Ecol Appl* **1**: 303–15.

Box GEP. 1954a. Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *Ann Math Stat* **25**: 290–302.

Box GEP. 1954b. Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and of correlation between errors in the two-way classification. *Ann Math Stat* **25**: 484–98.

Brown JH, Gillooly JF, Allen AP, *et al*. 2004. Toward a metabolic theory of ecology. *Ecology* **85**: 1771–89.

Cade BS and Noon BR. 2003. A gentle introduction to quantile regression for ecologists. *Front Ecol Environ* **1**: 412–20.

Canham CD, Lauenroth WK, and Cole JJ. 2003. Models in ecosystem science. Princeton, NJ: Princeton University Press.

Cohen J. 1988. The analysis of variance and covariance: statistical power analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Gotelli NJ and Ellison AM. 2004. A primer of ecological statistics. Sunderland, MA: Sinauer Associates, Inc.

Kéry M and Hatfield JS. 2003. Normality of raw data in general linear models: the most widespread myth in statistics. *Bull Ecol Soc Am* **84**: 92–94.

Larsen RJ and Marx MJ. 1986. An introduction to mathematical statistics and its applications. Englewood Cliffs, NJ: Prentice-Hall.

Loreau M, Naeem S, Inchausti P, *et al*. 2001. Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* **294**: 804–08.

McArdle BH. 2003. Lines, models, and errors: regression in the field. *Limnol Oceanogr* **48**: 1363–66.

McCullagh P and Nelder JA. 1997. Generalized linear models, 2nd edn. New York, NY: Chapman & Hall/CRC.

Meeuwig JJ and Peters RH. 1996. Circumventing phosphorus in lake management: a comparison of chlorophyll a predictions from land-use and phosphorus-loading models. *Can J Fish Aquat Sci* **53**: 1795–1806.

Miller RG. 1986. Beyond ANOVA: basics of applied statistics. New York, NY: Wiley Press.

Neter J, Kutner MH, Nachtsheim CJ, and Wasserman W. 1996. Applied linear statistical models. Chicago, IL: Richard D Irwin, Inc.

Quinn GP and Keough MJ. 2002. Experimental design and data analysis for biologists. New York, NY: Cambridge University Press.

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

Toms JD and Lesperance ML. 2003. Piecewise regression: a tool for identifying ecological thresholds. *Ecology* **84**: 2034–41.

Wilson K and B.T. Grenfell. 1997. Generalized linear modelling for parasitologists. *Parasitol Today* **13**: 33–38.

