Dartmouth College

Dartmouth Digital Commons

Open Dartmouth: Published works by Dartmouth faculty

Faculty Work

2005

Knowing When to Draw the Line: Designing More Informative Ecological Experiments

Kathryn L. Cottingham
Dartmouth College

Jay T. Lennon
Dartmouth College

Bryan L. Brown
Dartmouth College

Follow this and additional works at: https://digitalcommons.dartmouth.edu/facoa

Part of the Applied Statistics Commons, Biology Commons, and the Ecology and Evolutionary Biology Commons

---

Dartmouth Digital Commons Citation

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.

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.

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

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

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-
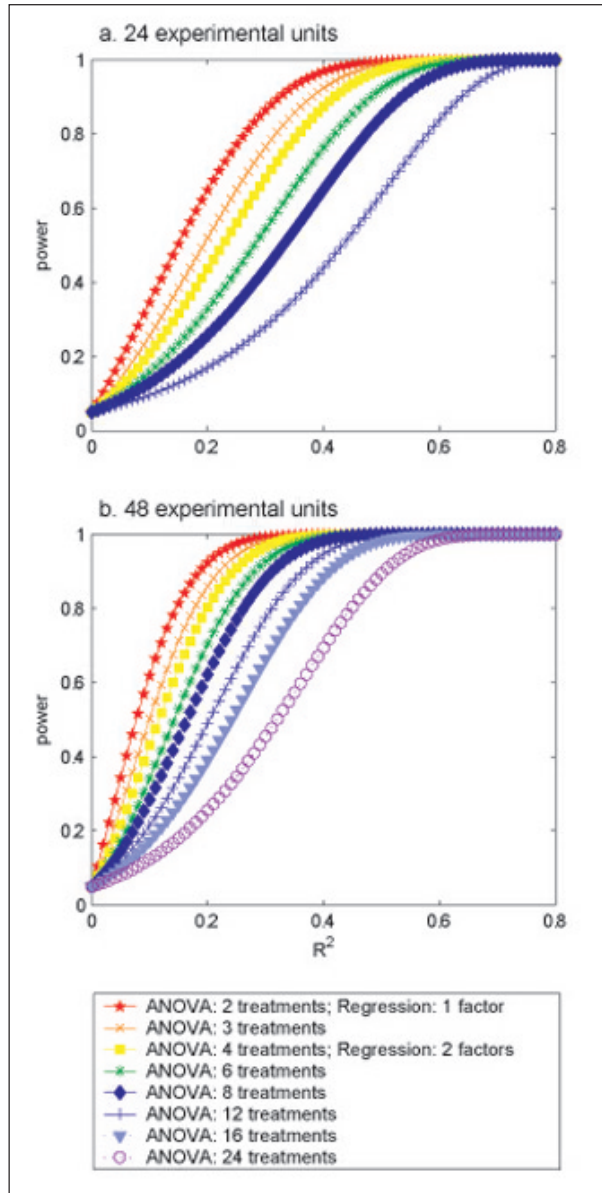
Figure 1. *Power curves for all possible balanced one- and two-way regression and ANOVA models when there are (a) 24 and (b) 48 experimental units. Identifying information for each curve is provided below the figure; the number of replicates per treatment can be determined by dividing the number of experimental 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 experimen-

tal 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, $H_o: R^2_{Y \cdot \beta} = 0$, where $R^2_{Y \cdot \beta}$ 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 $R^2_{Y \cdot \beta}$) thought to be of biological significance. We then translate the target $R^2_{Y \cdot \beta}$ into an effect size f^2 using the ratio of explained to unexplained variance:

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

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, 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).

It should be remembered, however, that regression is not appropriate in all situations. For example, standard linear regression is inappropriate
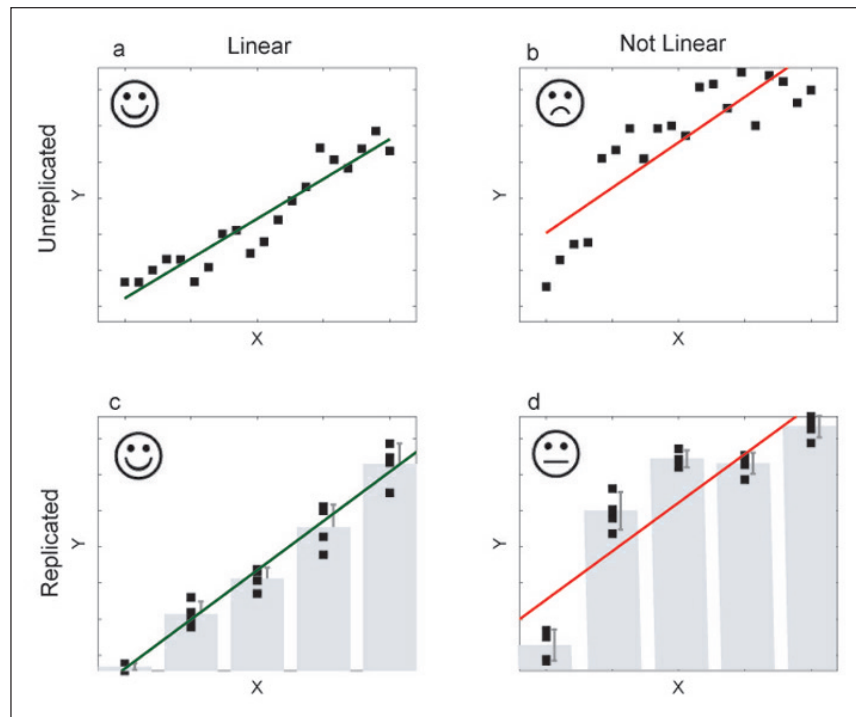

Figure 2. *Contrasting outcomes for replicated versus unreplicated regression-based experimental designs. In the left column, data were simulated using a linear relationship; in the right column, data were simulated using a sigmoidal relationship. In the top row, each level of X is unreplicated, so it 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).*

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.

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.

■ **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).

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.

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 trade-off should be based on the following criteria:

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
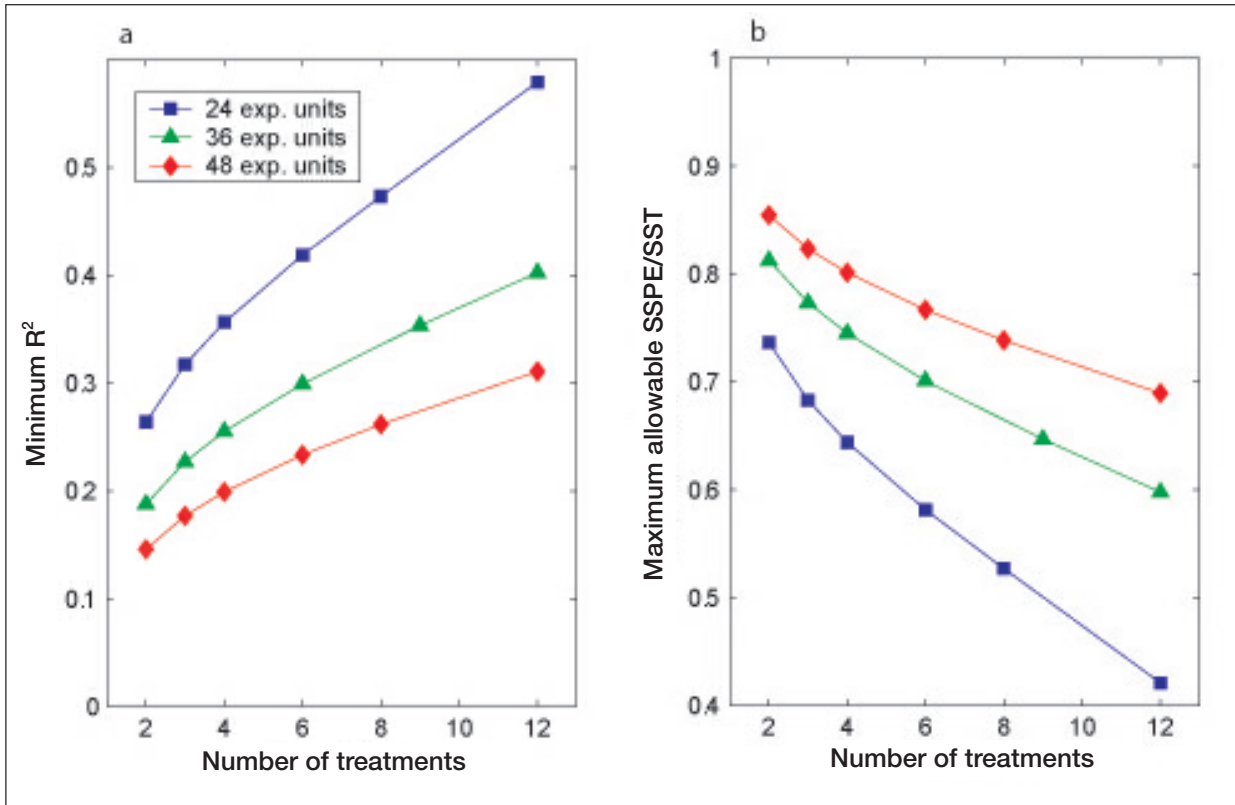
Figure 3. *Guidelines for choosing between possible replicated regression designs when 24, 36, or 48 experimental units are available. Lines show (a) the minimum required R^2 or (b) the largest possible allowable SSPE/SST for the target power level 0.8; line symbols and colors indicate the number of 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**

■ **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.

■ The matrix formulation for regression and ANOVA (Neter *et al.* 1996).

Both regression and ANOVA can be described using the general linear model $Y = X\beta + \epsilon$, where

> Y = an *n x 1* column vector of values of the response variable Y
> There are n observations.
> X = an *n x p* matrix with columns corresponding to the *p* predictor variables X_i
> β = an *p x 1* column vector of parameters, with row numbers corresponding to the column numbers in X
> ϵ = an *n x 1* column vector of errors

In regression, the columns in X are fairly straightforward. Most regression models contain an intercept (β_o), which is fit by setting the first column of X to a dummy variable X_o with value=1 for all observations. One column is added to the X matrix for each of the predictor variables, and if there are interaction terms or polynomial terms, the appropriate products or powers of the predictor variables are added as additional columns. For example, in simple linear regression, we use:

$$X = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ 1 & x_{31} \\ \vdots & \vdots \\ 1 & x_{n1} \end{bmatrix}$$

In multiple regression with two predictors and an interaction, we use:

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{11}*x_{12} \\ 1 & x_{21} & x_{22} & x_{21}*x_{22} \\ 1 & x_{31} & x_{32} & x_{31}*x_{32} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n1}*x_{n2} \end{bmatrix}$$

In ANOVA, the X matrix contains qualitative indicator variables indicating membership in treatment groups. If there are *m* groups, there are *m* columns in X. There are an infinite number of ways to define the qualitative variables, but one way is to calculate the overall mean for Y (using the same approach described for β_o, above) together with deviations of particular treatments from this overall mean.

This involves assigning a column in X to all but one treatment group; because the overall mean is already known, the deviation for the last group is determined from the sum of the other deviations. The indicator variables are set to 1 when an observation (row) is in the group that corresponds to that X variable, -1 if the observation is in the treatment group without its own column, and 0 otherwise. For example, suppose there were four treatment groups and three observations per group. The X matrix for the models described in Table 1 might look like:

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{bmatrix}$$

Regardless of how X is formulated, the equation $Y = X\beta + \epsilon$ is solved for β using the normal equations, giving the parameter estimates

$$\hat{b} = (X'X)^{-1} X'Y.$$

Once the parameters are estimated, we partition the overall variance in the data as follows, given that *p* is the number of columns in X (ie *p*=2 for a simple linear regression, *p*=4 for a two-factor regression, and *p*=m, the total number of treatments, for any ANOVA).

Source	SS	df	MS	F
Model (M)	$b' X' Y - n\bar{Y}^2$	p-1	MSM	MSM/MSE
Error (E)	$Y' Y - b' X' Y$	n-p	MSE	
Corrected Total (T)	$Y' Y - n\bar{Y}^2$	n-1		

We also determine the percent of variability explained by the model, R^2, as SSM/SST.

■ Reference

Neter J, Kutner MH, Nachtsheim CJ, and Wasserman W. 1996. Applied Linear Statistical Models. Chicago: Richard D Irwin, Inc.

Cottingham – Web-only Appendix 2. Replicated regression (RR) designs

■ Lack of fit tests

RR designs provide an underappreciated opportunity to test whether a particular regression model is appropriate for the data using lack-of-fit tests (Draper and Smith 1998). These tests are particularly good at diagnosing deviations from the linear model that may be difficult to detect by eye. Lack of fit tests work by partitioning residual variation around a regression line into two components: that due to variability among replicates within a treatment (the "pure error") and that due to deviations of the treatment means from the fitted curve (the "lack of fit"; Table 2A). The pure error is obtained from an ANOVA that uses the predictor variable(s) as a classification factor rather than as a quantitative one, and the lack-of-fit component is estimated from the difference between the error SS from the regression model and the error SS from the ANOVA model. There is significant lack-of-fit when the ratio of the mean squares lack-of-fit (MS_L) to mean squares pure error (s_e^2) exceeds a critical F-statistic. If there is no significant lack-of-fit, then the regression model is appropriate for the data and conclusions can be drawn accordingly. If, however, there is significant lack-of-fit, remedial action is required. In some cases, the regression model can be modified to be more appropriate for the data, for example by adding polynomial terms (which will reduce the power somewhat due to the additional parameters). However, in other cases, there is no appropriate linear model for the data. In this case, researchers can switch to non-linear regression or "fall back" to drawing conclusions using ANOVA.

We note briefly that lack-of-fit tests are also available for nonlinear regression, although we do not develop them here (see Draper and Smith 1998 instead), providing another argument for the use of RR designs in ecological research.

Table 2A. ANOVA table for a replicated regression
N indicates the total number of experimental units, p is the number of columns of X, m indicates the number of treatments with replicates, and n_j is the number of replicates for treatment j.

Source		SS	df	MS	
Model	Regression SS	SSR	p-1		
Error	Lack of fit	SSLOF	m-p	MS_L	s^2
	Pure error	SSPE	N-m	s_e^2	
Total (corrected)			N-1		

■ Using RR designs to see the parallels between regression and ANOVA

Replicated regression provides a currency for relating the model and residual sums-of-squares (SS) for regression and ANOVA models fit to the same data (Table 2B). The alternative partitioning of sums of squares and degrees of freedom has some interesting implications. Most importantly, the lack-of-fit SS (SSLOF) are part of the error SS in regression (SSE), but part of the model SS in ANOVA (SSA). As a result, we expect changes in R^2 (and thus effect size) between regression and ANOVA models applied to the same dataset. R^2 will always be bigger for ANOVA than for regression by the amount SSLOF/SST.

Table 2B. Partitioning variability in a RR dataset according to the regression, RR, and ANOVA models

	As a regression		As a replicated regression		As an ANOVA	
Source	**SS**	**df**	**SS**	**df**	**SS**	**df**
Model	SSR	1	SSR	1	SSA	m-1
Error	SSE	N-2	SSLOF	m-2		
			SSPE	N-m	SSPE	N-m
Total	SST	N-1	SST	N-1	SST	N-1

■ How Figure 3 was created

To create the scenarios in Figure 3, we started with power curves (as explained in Statistical Panel 2) for experimental designs with 24 (Figure 1a), 36 (not shown), and 48 (Figure 1b) experimental units. We then selected a minimum power for the ANOVA (0.8, following convention).

1. Left panel: Minimum R^2 vs the number of treatments. On the power curves for experiment size, we drew a line horizontally across the figure at the target power level. At each intersection of this "minimum power" line with a power curve, we dropped down to the X-axis and recorded R^2 at that point, which is the minimum R^2 required to produce that power for that experimental design. We then plotted this minimum R^2 versus the number of treatments in that design in Figure 3a.

2. Right panel: Maximum allowable ratio of SSPE/SST vs number of treatments
   Statistical Panel 3 introduces several abbreviations for the sum-of-squares terms in a replicated regression:
   • SSR = sums-of-squares due to regression
   • SSPE = sums-of-squares due to pure error, the variability around the mean for each level of the predictor variable(s)
   • SSLOF = sums-of-squares due to lack of fit, the deviation from the regression line not explained by the ANOVA (determined as SSR-SSPE).

   From Table 2B in Web-only Appendix 2, we also know that R^2_{anova} = (SSR+SSLOF) / SST.

   Therefore, we can define
   $1-R^2_{anova}$ = SST/SST – (SSR+SSLOF)/SST = SSPE/SST, which provided us with a formula to convert the minimum R^2 obtained in Step 1 to the fraction of the total variability that is explained by the pure error, or variability among replicates within a treatment.
   Estimates of SSPE/SST are closely related to those used to calculate power analyses in t-tests and straightforward ANOVA models, and so are frequently estimable from past experiments (eg Case Study Panel 3).

Cut and paste the code for use in Matlab. The raw data file used for the simulation is available from the authors.

```
% calculatepower.m
% determine power for a series of potential one- and two-way experimental designs specified by the user
% author KL Cottingham (cottingham@dartmouth.edu)
% created 19 Dec 03 from compareRvsA_vsf2.m;
% last modified 23 December 2004 for Frontiers website


% 888888888888888888888888888888888888888888888
% give the necessary info
% 888888888888888888888888888888888888888888888

clear;
lookpowerfigs=0; % toggle figures on and off
lookthresholds=0; % toggle evaluating thresholds on and off

% setups
output=[];
thresholds=[];

% specify the target p-value
alpha=0.05;

% prepare figures (if desired)
if lookpowerfigs,
        figure(1); clf; orient tall;
end;

% set constraints
minnr=2; % minimum number of replicates per treatment
maxnr=5; % maximum number of replicates per treatment


% 888888888888888888888888888888888888888888888888
% looping structure
% 888888888888888888888888888888888888888888888888

% specify number of levels of factor A
for Alevels=2:4, %input('Number of levels of factor A? ');

% specify number of levels of factor B
for Blevels=1:4, %input('Number of levels of factor B? ');

% specify number of replicates of each cell
for nreps=minnr:maxnr, %input('Number of replicates per cell? ');

        % 8888888888888888888888888888888888888888888888888
        % determine df for regression & for ANOVA
        % 8888888888888888888888888888888888888888888888888

        % calculate number of EU
        N=Alevels*Blevels*nreps;

        % assume we're fitting a regression with three parameters: effects of A & B
        % and their interaction
        if Blevels==1,
```

```
        DFM_reg=1;
          else DFM_reg=3;
        end;
          DFE_reg = N - DFM_reg - 1;

          % assume we're fitting an ANOVA with main effects and interactions
          DFM_anova=(Alevels-1) + (Blevels-1) + (Alevels-1)*(Blevels-1);
          DFE_anova=N - DFM_anova - 1;

          % 8888888888888888888888888888888888888888888888888
          % calculate the power of each design, based on case 0 of Cohen Ch 9
          % delta = (effect size)squared * (u+v+1)
          % 8888888888888888888888888888888888888888888888888

          % determine critical value of F needed to reject Ho: no difference for each design
          Fcrit_reg=finv(1-alpha,DFM_reg,DFE_reg);
          Fcrit_anova=finv(1-alpha,DFM_anova,DFE_anova);

          % list of R2 to compare
          R2 = (0 : 0.01 : 0.99)';

          % list of effect sizes that go with those R2 values
          % f2 = R2 / (1 - R2)
          ES = R2 ./ (1-R2);

          % calculate delta as f2 * (u+v+1)
          delta = N.*ES;

          %calculate the power for each design here following other program
          power_reg=1-ncfcdf(Fcrit_reg,DFM_reg,DFE_reg,delta);
          power_anova=1-ncfcdf(Fcrit_anova,DFM_anova,DFE_anova,delta);

          output=[output; ones(length(ES),1)*[Alevels Blevels nreps] ES power_reg power_anova];

% 8888888888888888888888888888888888888888888888888
          % plot power vs. effect size
          % 8888888888888888888888888888888888888888888888888

if lookpowerfigs,
              sb=sb+1;
              if sb>8, sb=1; figno=figno+1; figure(figno); clf; orient tall; end;
              subplot(4,2,sb);
              semilogx(ES,power_reg,'r-',ES,power_anova,'k:');

              if sb==1, legend('Regression','ANOVA',2); end;
              ylabel('power');
              xlabel('Effect Size');
              title([num2str(Alevels) ' x ' num2str(Blevels) ' x ' num2str(nreps) ' design']);
end;

          % 8888888888888888888888888888888888888888888888888
          % determine thresholds of interest
          % 8888888888888888888888888888888888888888888888888

          reggtpt8=min(ES(find(power_reg>=0.8)));
          anovagtpt8=min(ES(find(power_anova>=0.8)));
          reggtanova=min(ES(find(power_reg>=power_anova)));
```

```
reggtanovaandpt8=min(ES(find(power_reg>0.8 & power_reg>=power_anova)));

% 8888888888888888888888888888888888888888888
% use algebra to determine what SSR & SSPE need to be to exceed these
% thresholds
% 8888888888888888888888888888888888888888888

% regression power > 0.8
minpctSSR=reggtpt8./(reggtpt8+1);

% anova power > 0.8
maxpctSSPE=1./(anovagtpt8+1);

% (regression power > anova power) & (regr power > 0.8) -> works out to
minpctSSRforRtowin=reggtanovaandpt8./(reggtanovaandpt8+1);
maxpctSSPEforRtowin=1./(reggtanovaandpt8+1);


% 8888888888888888888888888888888888888888888
% collect these thresholds for particular designs: are there patterns?
% 8888888888888888888888888888888888888888888

    thresholds=[thresholds; Alevels  Blevels  nreps  minpctSSR  maxpctSSPE  minpctSSRforRtowin
maxpctSSPEforRtowin reggtanovaandpt8];

   end; % for nreps

 end; % for Blevels
 end; % for Alevels


 save powervsESinfo.dat output /ascii;
 save thresholds.dat thresholds /ascii;
```

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 lev-

els 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.

358

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.