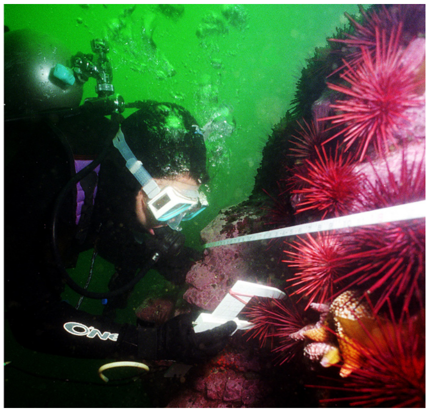

CONTRIBUTIONS

Commentary

Interpreting the Results from Multiple Regression and Structural Equation Models

The coefficients that are associated with pathways in multiple regression, as well as more advanced methods based on regression, such as structural equation models, are central to the interpretations made by researchers. The complex of factors that influence these coefficients make interpretations tricky and nonintuitive at times. Very often, inappropriate inferences are made for a variety of reasons. In this paper we discuss several important issues that relate to the interpretation of regression and path coefficients. We begin with a consideration of multiple regression. Here we discuss the different types of coefficients that can be obtained and their interpretations, with our focus on the contrast between unstandardized and standardized coefficients. Structural equation modeling is used to show how models that better match the theoretical relations among variables can enhance interpretability and lead to quite different conclusions. Here we again emphasize often-ignored aspects of the use of standardized coefficients. An alternative means of standardization based on the "relevant ranges" of variables is discussed as a means of standardization that can enhance interpretability.

Biologists have long used multiple regression in its various forms to examine relationships among explanatory and response variables. Over the past decade and a half, there has been a steady increase in the use of path analysis by biologists to serve the same purpose, but in the context of a more interpretive structure. Most recently, there has developed a considerable amount of interest in the more comprehensive capabilities of structural equation modeling (SEM) for understanding natural systems, again with the purpose of enhancing our interpretation of results. These methodologies have in common that they are based on the fundamental principles of regression and share many of the same issues when it comes to interpretation.

Researchers may not be aware that there has been a long history of discussion among quantitative social scientists and statisticians about the interpretation of results from both multiple regression and path analysis applications. The topic is sufficiently subtle and important that the central theme of Pedhazur's (1997) book on regression is the pitfalls of interpreting results. Among the many things he concludes is that results are frequently misinterpreted, particularly as they relate to the meaning of path coefficients. Many of these same issues apply to SEM. This discussion has involved a consideration of many topics, including the types of coefficients that can be calculated, the kinds of interpretations that can be supported using different coefficient types, and the importance of theory to interpretation. Here we illustrate some of these issues and discuss problems with the use of standardized coefficients, as well as a possible remedy.

An illustrative example

To illustrate the points being made in this paper we consider an example dealing with the response of shrublands to wildfire in Southern California (J. B. Grace and J. E. Keeley, *unpublished manuscript*). The data presented here represent a small subset of the variables in the complete study. In addition, the relationships among variables have been modified somewhat to meet the needs of the current paper. In this example, 90 sites were located in areas burned by a series of fires that occurred during a 2-week period in the fall of 1993 (Keeley et al., *in press*). Plots were established in all 90 sites and sampling began in spring of the first postfire year and continued for 4 more years, though only the data from the first sampling following fire are discussed here. At each site, the variables included (1) herbaceous cover (as a percentage of ground surface), (2) fire severity (based on skeletal remains of shrubs, specifically the average diameter of the smallest twigs remaining), (3) prefire stand age (in years), estimated from ring counts of stem samples, and (4) the elevation above sea level of the site. The data used in this analysis are summarized in Table 1. Again, the data presented are a subset of the original, and some relations in the data have been modified to make the example more applicable to our purposes.

Issues related to multiple regression

A multiple regression represents a particular model of relationships in which all potential explanatory variables (predictors) are treated as coequal and their interrelations are unanalyzed. As we shall see, the ability to obtain interpretable results from such models depends on the degree to which their structure matches the true relations among variables. Fig. 1 presents diagrammatic representations of a multiple regression model in which fire severity, stand age, and elevation are related to vegetation cover. Parameter estimates were obtained using the software Mplus (Muthén and Muthén 2005) under maximum likelihood estimation. Several types of coefficients were obtained from the analyses and are presented in Fig. 1, with each subfigure presenting a different view of the relations among variables.

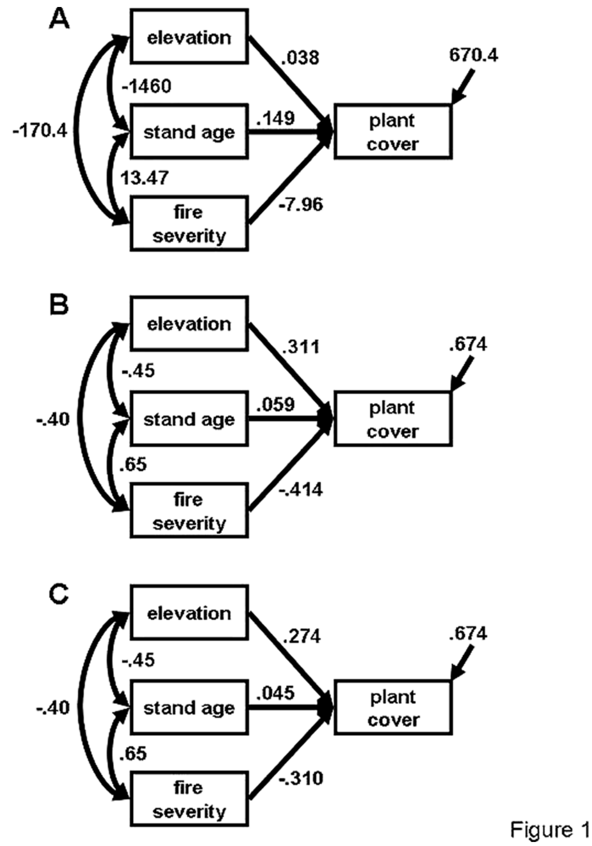

Fig. 1. Multiple regression results based on analysis of the data in Table 1. (A) Unstandardized parameters. (B) Standardized parameters. (C) Semipartial coefficients for the directional pathways.

Unstandardized coefficients

Fig. 1A presents the unstandardized path coefficients associated with the regression of plant cover on elevation, stand age, and fire severity. While the unstandardized coefficients are the most primary parameters obtained from a multiple regression, often they are not presented by investigators. In fact, typically the significance tests associated with regression are tests of the unstandardized parameters, and the standardized parameters are simply derived from the unstandardized coefficients and not directly tested. Characteristic of unstandardized parameters, they are expressed in the original units of the explanatory and dependent variables. With reference to a simple linear regression, unstandardized coefficients associated with directed paths represent the slope of the relationship. The same is true in multiple regression, although the slope is in *n*-dimensional space.

As we begin to interpret the results in Fig. 1A, note that the undirected relationships (double-headed arrows) represent the covariances among exogenous variables (predictors) in a model. In contrast, the coefficients associated with directed paths are partial regression coefficients. It is important for the discussion that follows to understand when the principles of partial regression apply. Simply put, partial regression represents a method of statistical control that removes the effect of correlated influences. Pathways that involve partial regression can be recognized by the following: (1) they involve a directed relationship (single-headed arrow), (2) the response variable (variable receiving the arrow) also receives other arrows, and (3) the multiple predictors affecting the response variable are correlated. As we can see from these criteria, all directed paths in multiple regression will involve partial regression as long as there are significant correlations among predictors. The question then is how are we to interpret such coefficients.

The literal definition of a partial regression coefficient is *the expected change in the dependent variable associated with a unit change in a given predictor while controlling for the correlated effects of other predictors*. There are actually several different ways we can look at partial regression coefficients. The most direct is to view them as parameters of an equation such as

$$\text{cover} = 0.038(\text{elevation}) + 0.149(\text{age}) - 7.96(\text{severity}) \qquad (1)$$

when variables are in their raw units. If we were able to plot a four-dimensional graph of cover against elevation, age, and severity, the unstandardized regression coefficients would be the slopes of the relationship in the plot. From this perspective, it should be clear that the coefficients estimate the mean influences of predictors on the response variable and the variation around the mean is ignored. Deviations from the mean in this case relate to the estimation of the probabilities that coefficients' values are zero. Thus, one interpretation of the unstandardized coefficients is that they are prediction coefficients. They also are descriptive coefficients in that they describe the association between cover and a one-unit change in the other variables. Hypothetically, these coefficients might also be viewed as explanatory. However, for such an interpretation to be valid, we must depend on the structure of the model to match the true dependencies among the predictors. As Pedhazur (1997:8) states, "Explanation implies, first and foremost, a theoretical formulation about the nature of the relationships among the variables under study." This point will be illustrated later in the paper when we discuss the structural equation model results for these data.

Referring back to our example, if we were to keep elevation constant for a set of plots, and the stands being burned were of a fixed age, a one-unit difference in the fire severity is associated with an average difference in cover of –7.96 cover units (i.e., the cover of the postfire community would differ by 7.96%). Similarly, if we were able to apply a fire of fixed severity while also holding stand age constant, a difference in elevation of 1000 m is associated with an expected difference of 38% in the postfire cover.

Standardized coefficients

Looking at Fig. 1A, we see that it is difficult to compare unstandardized coefficients among different pathways because the raw units are various. Cover varies in percentage points, elevation varies in meters, age varies in years, and fire severity varies in the units of an index based on the diameter of remaining twigs following fire. So, is a value of 0.038 (the coefficient for elevation effects on cover) large or small relative to the effect of another factor? The standardization of the coefficients based on the standard deviations of the variables is the approach typically used to make coefficients comparable. In essence, this puts variables in standard deviation units, and in that sense the expected impact of a standard deviation difference in one variable (say elevation) can be compared to a standard deviation difference in another variable (say fire severity). Though a convenient transformation, standardized regression coefficients are frequently misinterpreted, for reasons we will discuss next.

The most common misinterpretation of standardized coefficients is to interpret them as if they represent a partitioning of explained variance in the model. The fact that standardized coefficients are in standard deviation units contributes to the tendency to make this mistake. For example, the formula for standardized partial regression coefficients can be expressed in terms of the correlations among variables. In the case of two predictors, x_1 and x_2, and one response, y_1, this formula is

$$\gamma_{11} = \frac{r_{x1y1} - r_{x1x2} \times r_{x2y1}}{1 - r^2_{x1x2}} \tag{2}$$

where γ_{11} refers to the standardized partial regression coefficient representing the response of y_1 to x_1, and the r values represent the bivariate correlations among variables. This formula can be readily extrapolated to the case of more than two predictor variables (Pedhazur 1997).

Another relationship that applies to standardized coefficients is that the sum of all simple and compound associations between two variables equals the bivariate correlation between those two variables. For example, the bivariate correlation between elevation and cover is 0.45 (Table 1). With reference to Fig. 1B where standardized coefficients are presented, we find that the coefficients are those that satisfy the formula (allowing cover to be y_1, and elevation, stand age, and severity being $x_1 - x_3$)

$$r_{x1y1} = \gamma_{11} + r_{x1x2} \cdot \gamma_{12} + r_{x1x3} \cdot \gamma_{13}, \tag{3}$$

where γ_{11} is the response of y_1 to x_1, γ_{12} is the response of y_1 to x_2, γ_{13} is the response of y_1 to x_3, and r's refer to correlations.

A third property of standardized coefficients is that they can be related to the explained variance in our response variable using the equation

$$R^2 = r_{x1y1} \cdot \gamma_{11} + r_{x2y1} \cdot \gamma_{12} + r_{x3y1} \cdot \gamma_{12} \tag{4}$$

Pedhazur (1997). For our example presented in Fig. 1B, we find that the expression in Eq. 4 yields an R^2 of 0.326 (note the standardized error variance shown in Fig. 1B equals 1 minus the R^2).

Now, the properties of standardized coefficients give the impression that they solve a number of problems. Most obviously, they put all the coefficients in what seem to be the same units. However, they are only the "same" if we are willing to say that a standard deviation for one variable in one metric is *interpretationally equivalent* to a standard deviation of another variable that was measured in a different metric. This is an implicit assumption of using standardized coefficients and it is not obvious that this assumption is suitable other than in the fact that each is a standard deviation.

More seductive than that, however, is that standardized coefficients are expressed in terms of correlations, *which represent the variation associated with the relationships.* In the case of simple regression (involving one predictor variable), we know that the unstandardized coefficient represents the slope, while the standardized coefficient represents the square root of the variance explained in the response variable. Eq. 4 may give the false impression that this relationship between standardized coefficients and variance explained can be generalized to the case of multiple correlated predictors. However, it cannot be so generalized. To see why more readily, we now turn to the concept of semipartial coefficients and unique variance explanation.

Semipartial coefficients and the concept of shared variance explanation

The semipartial coefficient, when expressed in standardized form, represents a measure of the *unique ability of a predictor variable to explain variation in a response variable that cannot be explained by any other predictor variable in the model.* We can understand this in contrast to stepwise regression, which measures the sequential abilities of variables to explain residual variance. In sequential variance explanation, there is a pervading influence on the results by the logic used to determine the order of variables included. Here, the semipartials represent a measure of the minimum effect of a variable regardless of logical order. In the example in Fig. 1C, the coefficients associated with directed paths are semipartial coefficients, while the coefficients associated with undirected paths remain correlations. The unique variance explanation abilities of our three predictors (elevation, age, and severity) are 0.075, 0.002, and 0.096, the squares of the semipartial coefficients. Collectively, the three variables provide unique variance explanation of 0.173. Since the total variance explained by the full model is 0.326, we must conclude that 0.153 (roughly half) of the explained variance is shared among predictors.

The concept of *shared variance explanation* makes sense when we have predictor variables that are correlated for some unknown or unspecified reason. How are we to apportion the correlated explanatory power among predictors in a multiple regression? Since our relations among predictors are unanalyzed or not understood, we have no means to accomplish this. The implications of these relations can be seen if we compare the coefficients in Figs. 1B and C. It is to be expected that the partial regression coefficients are greater than the semipartial coefficients, with the degree of difference directly related to the strength of the correlations among predictors. It should be clear from the above discussion that as predictors become more highly correlated, their unique variance explanation ability decreases. It should also be clear from our presentation that the standardized partial regression coefficients (Fig. 1B) do NOT represent measures of variance explanation ability. Rather, *the standardized partial regression coefficients represent expected changes in* y *as a result of manipulations in* x *in standard deviation units while controlling for the correlated effects of other predictors.* The reason these coefficients cannot be used to represent variance explanation is simple; it is because we cannot guess how to apportion the variance explanation shared among predictors. In sum, the total variance explained in a multiple regression can only be attributed to the collection of predictors. The truth of this is most evident in nonlinear regression where individual predictors (e.g., x and x^2) may explain no variance by themselves, yet together they can explain substantial variance in some y.

Conclusions about the interpretability of multiple regression

While investigators commonly ask, "What is the relative importance of a set of causes controlling some observed phenomenon?" we must conclude that when predictor variables are correlated for unknown reasons, standardized partial regression coefficients do not provide an answer to this question. It is true that when correlations are not excessive, path coefficients can provide important insights. Multiple regression, which is inherently designed to ignore the causes behind the correlations among a set of predictors, makes for a particularly poor approach to understanding, however. This fundamental problem has been long recognized and is the central theme in Pedhazur's (1997) book on multiple regression. While Pedhazur discusses the problem from many different angles, his main conclusion is that without a theory to guide the analysis, a meaningful answer to the question of relative importance of factors is usually precluded in a multiple regression analysis. As we have seen, standardized regression coefficients do not equate to variance explanation. At the same time, measures of unique and shared variance explanation, which can be obtained using semipartial analysis, really don't address explanatory questions either, but instead, relate more to their unique roles as predictor variables.

Structural equation modeling

Since the interpretability of multiple regression results is typically limited by an insufficiently developed theoretical framework, we should consider what problems are solved using a theory-oriented method such as SEM. For those not familiar with SEM, it involves the use of a generalized multiequation framework that enables the analyst to represent a broad range of multivariate hypotheses about interdependencies (Bollen 1989). Path analysis, which is now familiar to most ecologists, is best known in analyses that only consider relations among observed variables. Modern SEM allows for the inclusion of unmeasured (latent) effects, as well as the specification of a wide range of model types. Importantly, SEM allows for evaluations of model fit that serve to permit overall testing of the model as a hypothesis. While SEM is most commonly based on maximum likelihood estimation, many model types can be solved using various least squares procedures. While we do not present latent variable examples in this paper, the issues discussed apply equally to such models.

We should begin by stating that SEM does not solve all problems associated with interpreting multivariate relations. Both inadequate data and insufficient theory can block substantial progress. Additionally, while SEM permits the implications of a causally structured theory to be expressed, the analysis itself does not contribute to the establishment of causality. This must come from other information. Nonetheless, the use of theory to guide our analysis within an SEM framework has the potential to remove many obstacles to interpretation. The example presented here is meant to illustrate that potential, but not to imply that the application of SEM automatically leads to a superior analysis.

Returning to the example of fire response by California shrublands, we now ask, "What do we know of the relations among our explanatory variables?" In this case, the authors of the original study felt they knew some important things, but we were unable to incorporate this information into the multiple regression performed in the previous section. First, substantial experience (Keeley 1991) indicates that postfire recovery by the plant community may be affected by fire severity because of impacts on seed survival. It is also possible that impacts to soil properties could contribute as well (Davis et al. 1989). The point is that fire severity is reasonably modeled as having a direct impact on plant cover. Stand age can be expected to have an effect on fire severity because older stands tend to have more fuel. A simple thought experiment illustrates the point. If we were to vary stand age (say, allow a stand to get older and accumulate more fuel), we might reasonably expect that it would burn hotter (though this would not be guaranteed). However, if we were to manipulate fire severity in a plot, that would certainly not affect the age of the stand. This logic and the experience upon which it is based encourages us to represent the relationship between stand age and fire severity as a directional one rather than a simple correlation. By a similar logic, we can see that the relationship between elevation and stand age should be represented as directional. If shrub stands tend to be younger as we go higher in elevation, which the data indicate, (e.g., if there were a reduced incidence of fire suppression at higher elevations), then picking a spot lower on the mountain will likely result in finding an older stand. On the other hand, if we were to allow a stand of shrubs to get older, we would not find that there was an associated change in elevation. Again, the use of thought experiments, which tap into our body of prior knowledge, suggest directional relationships among variables.

Some researchers may be uncomfortable with the logic used above to indicate directional relationships in causal models. This subject is beyond the scope of our discussion in this paper and we refer the reader to more in-depth treatments of the subject (e.g., Bollen 1989, Pearl 2000, Shipley 2000). For now, we accept such a procedure as reasonable and illustrate its consequences in Fig. 2. The path model represented in Fig. 2 illustrates the logic of the dependencies described above. In addition, it represents the possibility that there may be influences of elevation on cover that are unrelated to associated variations in stand age and fire severity. Because this model is not saturated (i.e., not all paths are specified), our model represents a testable hypothesis. Inherent in SEM practice is the evaluation of fit between model expectations and observed relations in the data. Our point here is not to elaborate on this point, but only to note this feature of SEM practice and then continue with our discussion of interpretation. The patterns of covariances specified in Table 1 in fact fit the model presented in Fig. 2 reasonably

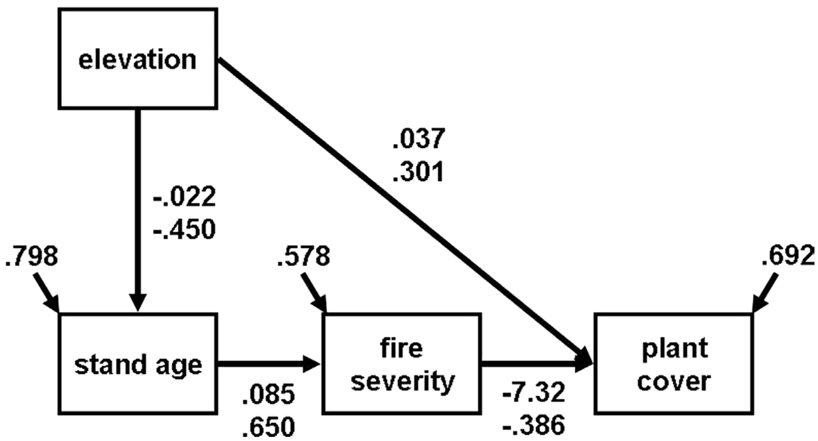

Fig. 2. Path model results based on analysis of data in Table 1. For path coefficients, upper numbers are the unstandardized coefficients, while lower numbers are standardized path coefficients. Error variables are standardized values. Model chi-square = 2.535 with 2 df, $P = 0.278$.

well (chi-square = 2.535 with 2 df and $P = 0.278$; note that a nonsignificant P value indicates the absence of significance deviations between data and model). This does not, of course, prove that the model is the correct one, only that it is consistent with the data.

The first thing we should do when interpreting the results in Fig. 2 is to consider which of our paths involve partial regression and which involve simple regression. Recall that response variables receiving two or more directed arrows will involve partial regression if the predictors involved are correlated. As stand age and fire severity only receive single directed arrows, their incoming pathways represent simple regression relations. We can see in fact that the correlations in Table 1 match the standardized path coefficients in Fig. 2 for these two pathways. Cover, on the other hand, has multiple influences and thus, the coefficients from elevation to cover and fire severity to cover are partial coefficients. What this means is that when we examine the relationship between elevation and stand age or between age and severity, there are no influences from other variables in the model to control for. On the other hand, the relationship between severity and cover is controlled for the covarying effects of elevation on cover. Similarly, the direct path from elevation to cover represents the effect once the influence of severity is removed.

Considering the unstandardized path coefficients in Fig. 2, we can see that the covariance between elevation and stand age can be understood as an expectation that age will decline on average by 2.2 years with an increase of 100 m. The covariance observed between stand age and fire severity can be understood as an expectation that severity will increase by 0.085 units with each year older a stand gets. Thus, we can understand the covariance between elevation and fire severity as the product of these two described relationships. Further, there is no indication of any other effect of elevation on fire severity except that mediated by stand age (because there is no direct path from elevation to severity to indicate some other effect).

The interpretation of unstandardized coefficients connecting severity and elevation to cover is somewhat different from those associated with a simple regression coefficient. We would draw the interpretation from Fig. 2 that increasing fire severity by one unit while holding all other conditions constant would cause a decrease in cover of 7.32%. The effect of elevation on cover is somewhat more interesting because of the presence of both direct and indirect effects on cover implied by the model. The direct path from elevation to cover predicts that if one were to choose a site 100 m higher than the mean and yet have an average severity fire, postfire cover would be 3.7% higher

Table 1. Covariances and correlations† among four variables relating vegetation regrowth in response to wildfire and the standard deviations of each variable (n = 90).‡ Matrix diagonals are the variances for the four variables.

Variables	Vegetation cover (% cover)	Fire severity (index values)	Prefire stand age (yr)	Elevation (m)
Cover	1,006.2	–26.2	–139.4	3686.3
Severity	–0.50	2.722	13.47	–170.4
Age	–0.35	0.65	157.8	–1459.6
Elevation	0.45	–0.40	–0.45	66,693
Standard deviations	31.72	1.65	12.56	258.25

†Note that the variance/covariance matrix can be reconstituted from the correlations and standard deviations presented. All analyses presented are based on the analysis of covariances.

‡The correlations among variables have been modified from the original to make the example more useful for the purposes of this paper. However, the standard deviations are as found by Keeley and Grace (submitted), thus the original scales for variables are preserved.

than the mean. On the other hand, the total effect of elevation on cover is 0.050, which indicates that if one moved upslope 100 m and allowed stand age and severity to vary as it naturally would (i.e., we are not holding them constant), there would be a net increase in cover of 5.0%. For the total effect of varying elevation, part of the increase in cover (1.3%) would result from the fact that stands would be younger (on average), 100 m higher, and associated fires would be expected to be less severe.

Consideration of standardized coefficients (Fig. 2) provides for an understanding of relationships expressed in terms of standard deviations. Such coefficients are both more easily compared (assuming different standard deviations can be thought of as equivalent) and somewhat more abstract. In these units, we see that if severity were increased by one standard deviation while elevation was held constant, cover would be expected to decrease by 0.386 standard deviations. On the other hand, if elevation was increased by one standard deviation, while holding severity constant, cover would increase by 0.301 standard deviations. Based on an estimated total effect of elevation on cover of 0.414, we can see that if elevation was increased one standard deviation without holding age and severity constant, then cover would increase 0.414 standard deviations. Thus, in terms of standardized units, the direct effect of elevation on cover is less (sign ignored) than the effect of severity (0.301 vs. 0.386), though the total effect of elevation on cover is greater (0.414).

So, how does all this relate to the question of the relative importance of different factors in affecting cover? If we accept standardization in terms of standard deviations as a reasonable basis for comparing coefficients (which is questioned below), it can be seen that the total influence of elevation on cover is greater than that of fire severity, with the total effect of stand age (–0.251) being least. The question we must now address is what it means to say that a pathway represents expected change in terms of standard deviation units.

Criticisms of standardization

While the above discussion appears to provide a suitable resolution of the question of how we may evaluate the importance of explanatory variables, we were forced to accept the caveat that standardization based on standard deviations was reasonable. Many metricians actually recommend that researchers avoid using standardized coefficients and focus on the unstandardized coefficients when seeking to draw conclusions from regression models (Darlington 1990, Luskin 1991). The reason for this is tied to the substantive meaning of unstandardized coefficients and the conditional nature of standardized coefficients. If we presume that our sample is fairly representative of some larger world, our unstandardized estimates represent the slopes of the relationships (i.e., the mean responses). When we use standardized coefficients, we interject additional variables into the problem, that of the sample variances. As Pedhazur (1997:319) so eloquently put it, "The size of a [standardized coefficient] reflects not only the presumed effect of the variable with which it is associated but also the variances and the covariances of the variables in the model (including the dependent variable), as well as the variances of the variables not in the model and subsumed under the error term. In contrast, [the unstandardized coefficient] remains fairly stable despite differences in the variances and the covariances of the variables in different settings or populations."

These criticisms of standardization appear rather powerful. In many ecological studies, we know that our samples often represent a tiny fraction of the total samples possible. Also, restrictions on randomization, for example because of accessibility problems or other sampling limitations, mean that sampling is often neither purely random nor fully representative; thus, variances can easily vary from sample to sample. Additionally, comparisons among populations based on standardized coefficients depend on the variances being constant across populations, which may frequently not be the case. Unstandardized coefficients are generally much more readily estimated with accuracy and less sensitive to differences in the variances of the variables across samples. Comparisons across populations (or between paths) in unstandardized coefficients do not depend on equal sample variances, and as a result, are more generalizable parameters than are those based on standardization. Altogether, there are assumptions that go into the interpretation of standardized coefficients and these are typically ignored, representing unknown influences.

A possible resolution using an alternative standardization procedure

Despite the criticisms of standardization, researchers generally would prefer a means of expressing coefficients in a way that would permit direct comparisons across paths. The debate over this issue goes back to Wright (1921), who originally developed path analysis using standardized variables. It was Tukey (1954) and Turner and Stevens (1959) who first criticized the interpretability of standardized values in regression and path models, and many others have since joined in that criticism. However, Wright (1960) argued in defense of standardized coefficients, saying that they provide an alternative method of interpretation that can yield a deeper understanding of the phenomena studied. Later, Hargens (1976) argued that when the theoretical basis for evaluating variables is based on their relative degrees of variation, standardized coefficients are appropriate bases for inference. Therefore, there are circumstances where standardized coefficients would be desirable. As Pedhazur's recent assessment of this problem concludes, ". . . the ultimate solution lies in the development of measures that have meaningful units so that the unstandardized coefficients . . . can be meaningfully interpreted."

So, how might we standardize using measures that have meaningful units? We must start by considering what it means to say that if x is varied by one standard deviation, y will respond by some fraction of a standard deviation? For normally distributed variables, there is a proportionality between the standard deviation and the range such that six standard deviations are expected to include 99% of the range of values.

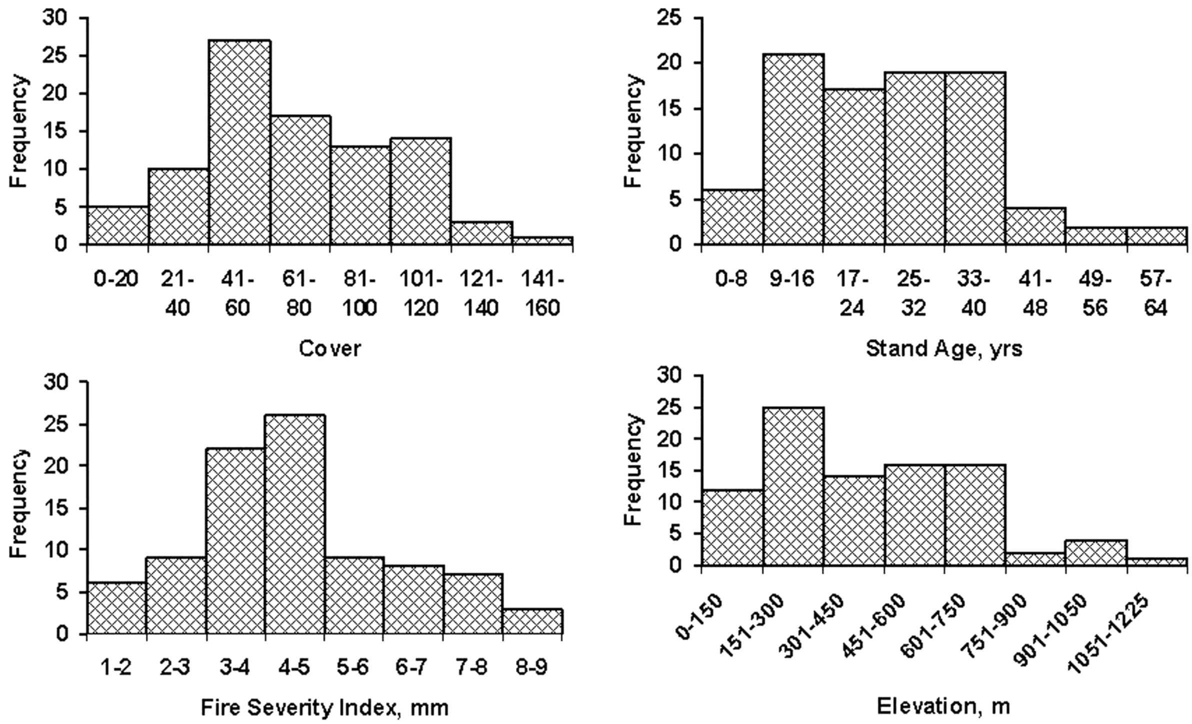

Fig. 3. Frequency diagrams for cover, stand age, fire severity, and elevation.

As discussed earlier, this may seem reasonable if (1) we have a large enough sample to estimate a consistent sample variance, (2) our variables are normally distributed, and (3) variances are equal across any samples we wish to compare. The reason why many metricians oppose standardized coefficients is because these three necessary conditions are not likely to hold generally. Of equal importance, rarely are these requirements explicitly considered in research publications and so we usually don't know how large violations of these requirements might be.

Fig. 3 presents frequency distributions for the four variables considered in our example. In the absence of further sampling, the repeatability of our sample variance estimate is unknown. This contributes to some uncertainty about the interpretability of coefficients standardized by the standard deviations. As for approximating a normal distribution, three of the four variables are truncated on the lower end of values. Cover can never be <0%, elevation likewise has a lower limit of expression relevant to terrestrial communities in this landscape, and stand age is also limited to a minimum value of between 0 and 1 year. None of these deviations are substantial enough to cause major problems with hypothesis tests (i.e., these variables are not wildly nonnormal); however, the deviations from idealized normality may very well impact the relationships between standard deviations and ranges. The observed range for cover was from 5% to 153% (overlapping canopies allow cover to exceed 100%), while six times the standard deviation yields an estimated range of 190%. The observed range for elevation was from 60 to 1225 m, while six times the standard deviation equals 1550 m. Stand age ranged from 3 to 60 years old, with six times the standard deviation equaling 75 years. Finally, fire severity index values ranged from 1.2 to 8.2 mm, while six times the standard deviation equals 9.9 mm. Thus, observed ranges are consistently less than would be estimated based

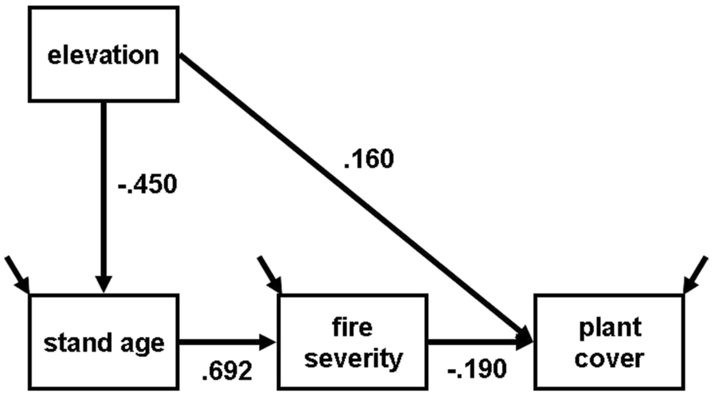

Fig. 4. Path analysis result showing coefficients standardized by the relevant ranges.

on standard deviations and the degree to which this is the case is slightly inconsistent (ratios of observed to predicted ranges for cover, elevation, age, and severity equal 0.78, 0.75, 0.76, and 0.71).

It is possible that in some cases information about the ranges of values likely to be encountered or of conceptual interest can provide a more meaningful basis for standardizing coefficients than can the sample standard deviations. We refer to such a range as the "relevant range." For example, if we have a variable whose values are constrained to fall between 0 and 100, it would not seem reasonable for the relevant range chosen by the researcher to exceed this value regardless of what six times the standard deviation equals. On the other hand, it may be that the researcher has no basis other than the observed data for selecting a relevant range. Even in such a case, we can choose to standardize samples that we wish to compare by some common range so as to clarify meaning across those samples. Whatever the basis for standardization, researchers should report both the unstandardized coefficients and the metrics used for standardization.

For the variables in our example, we specify the relevant range for cover to be from 0% to 270%. Obviously values cannot fall below 0%, but why chose an upper limit of 270%? Examination of cover values for all plots across the five years of the study show that values this high were observed in years 2 and 4 of the study. By using a relevant range of from 0% to 270%, we permit comparisons across years standardized on a common basis. Of course, this implies that the slopes measured will extrapolate to that full range, which is an assumption that should be evaluated closely. For elevation, the relevant range we choose is the observed range, from 60 to 1225 m. This span of 1165 m is chosen because we do not wish to extrapolate to lower or higher elevations, in case relationships to other variables are not robust at those elevations. For stand age, we specify the relevant range to be 60 years for basically the same reason. Finally, the fire index range chosen was also the observed range, which was 7.0 mm. It is clear that values could be obtained beyond this range in another fire. It is not known, however, whether the relationship between remaining twig diameter and herbaceous cover would remain linear outside the observed range.

Based on these determinations, we can generate path coefficients standardized on the relevant ranges. These coefficients are shown in Fig. 4. The biggest numeric differences between these values and those standardized using standard deviations (Fig. 2) is that the absolute values of the coefficients leading to cover are lower because of the large relevant range for this variable. The coefficient for the effect of age on severity is slightly higher, while that for the effect of elevation on age is unchanged. Using these coefficients now allows us to describe the importance of variables using their relevant ranges as the explicit context. These interpretations are *only* valid for relative comparisons within the *n*-dimensional parameter space defined by the relevant ranges. As fire severity increases across its relevant range, cover would be expected to decline by 19% of its relevant range. As elevation increases across its relevant range, the total change in cover from both direct and indirect causes would be an increase of 21.9% (the total effect). We now conclude from this analysis that the sensitivities of cover to fire severity and elevation (19% vs. 21.9%) are roughly equivalent in this study, though of opposing sign. It is possible to test whether these two estimates are reliable differences, which in this case, they are not.

Conclusions

It is important to recognize that the analysis of data has both an analytical element and a research element. By analytical element, we refer to the purely mathematical and statistical properties of the analytical methods. By research element, we refer to the fine art of applying analysis methods in the most meaningful ways. Formal training in statistics often emphasizes the analytical element and provides limited prescriptions for research applications that do not include a great deal of subjective judgment. What experienced statisticians have long known, however, is that for the application of statistical methods to be successful, strong guidance from the research perspective is required. Structural equation modeling is powerful specifically because it allows researchers to incorporate their accumulated knowledge into the analysis. Our advice regarding the interpretation of path coefficients is in that same vein. Rather than automatically allow sample standard deviations to represent the authoritative basis for standardizing coefficients, it is possible to insert our knowledge of the subject into the standardization process by explicitly considering the relevant ranges over which variables are to be considered. This procedure of standardizing based on substantive considerations acts to facilitate comparisons while avoiding problems associated with the sample-specific nature of standard deviations.

As with many new approaches, initial gains from defining and using the relevant range for standardization may be modest. Often the sample range will provide the best estimate available. However, as we accumulate additional information and focus on the ranges that are relevant to the inferences we wish to draw, much can be gained. Again, we recommend that unstandardized coefficients always be presented, regardless of the use of standardized coefficients of any sort. By also including either the sample standard deviations or the relevant ranges, which provide the bases for standardization, researchers can begin to compare both standardized and unstandardized values across studies. At present, there is a widespread and careless misapplication of standardized coefficients by researchers, both in the use of multiple regression and in the use of SEM/path analysis. Alternative means of comparing standardized coefficients may prove useful in drawing meaningful conclusions from analyses.

James B. Grace
US Geological Survey
and
Kenneth A. Bollen
University of North Carolina

Literature cited

Bollen, K.A. 1989. Structural equations with latent variables. John Wiley and Sons, New York, New York, USA.

Darlington, R. B. 1990. Regression and linear models. McGraw-Hill, New York, New York, USA..

Davis, F. W., M. I. Borchert, and D. C. Odion, 1989. Establishment of microscale vegetation pattern in maritime chaparral after fire. Vegetatio **84**:53–67.

Hargens, L. L. 1976. A note on standardized coefficients as structural parameters. Sociological Methods and Regression **5:**247–256.

Keeley, J. E. 1991. Seed germination and life history syndromes in the California chaparral. Botanical Review **57**:81–116.

Keeley, J. E., C. .J. Fotheringham, and M. B. Keeley. *In press.* Determinants of postfire recovery and succession in mediterranean-climate shrublands of

California. Ecological Applications.

Luskin, R. C. 1991. Abusus non tollit usum: standardized coefficients, correlations, and R^2s. American Journal of Political Science **35:**1032–1046.

Muthén, L. K., and B. O. Muthén. 2005. Mplus user's guide. Third edition. Muthén and Muthén, Los Angeles, California, USA.

Pearl, J. 2000. Causality: models, reasoning, and inference. Cambridge University Press, Cambridge, UK.

Pedhazur, E. J. 1997. Multiple regression in behavioral research. Third edition. Wadsworth, Belmont, California, USA.

Shipley, B. 2000. Cause and correlation in biology. Cambridge University Press, Cambridge, UK.

Tukey, J. W. 1954. Causation, regression and path analysis. Pages 35–66 *in* O. Kempthorne, T. A. Bancroft, J. W. Gowen, and J. L. Lush, editors. Statistics and mathematics in biology. Iowa State College Press, Ames, Iowa, USA.

Turner, M. E., and C. D. Stevens. 1959. The regression analysis of causal paths. Biometrics **15**:236–258.

Wright, S. 1968. Evolution and the genetics of populations. Volume 1. Genetic and biometric foundations. University of Chicago Press, Chicago, Illinois, USA.