7/28/05

Amy,

When we spoke of the formulas some time ago you mentioned that you weren't overly familiar with the math that went into them. I'll try to avoid getting too complex in my discussion below, but I'll need to talk about it to some extent so that you understand the difference between what the LSAC folks are telling you and what I'm suggesting concerning the use of the weights and the index. (My apologies if some of this is really basic.)

A main point of using an index is to get some idea of the potential for an applicant in a comparative context. From prior history we know that an applicant's LSAT and GPA are good predictors of the person's first year law school grades. Given the competition for spaces in a given entering class, it's not surprising that a formula was developed to allow comparisons between applicants. Eventually, someone tied the notion of an index with that of the predictors and put them into a formula where the index value is the dependent variable and the LSAT and GPA values are the independent variables.

To determine the "best" relationship for using the dependent variables as predictors for the dependent variable, we begin with data from current students. This allows us to know the values for the LSAT and GPA, but also for the index. (For applicants, this value will be unknown and the index value eventually calculated will be a prediction of their index score based on the actual values from prior students.)

The formula that any given school uses can vary, but let's stick to the simple model which can be written as:

$$I = \beta_1 \times \text{LSAT} + \beta_2 \times \text{GPA} + C$$

I, of course is the index value (which usually relates directly or indirectly to the student's first year law school grades). LSAT and GPA are the student's values for these measures. The "betas," the β_i, are called the coefficients of the equation and they are unknown at this point. The C is the "constant" and we don't need to worry about it for present purposes.

The equation above is in "linear form." That means we expect the dependent variable will vary directly with changes in the values of the independent variables. Thus, we would expect a one unit change in one of the independent variables to result in a corresponding change in the dependent variable. The relationship won't be one-to-one, for a couple of reasons, but that's where the betas come in.

Having constructed the equation we now want to "estimate" the model. That means we want to estimate the values for the coefficients (the betas). I'll avoid a discussion of "true" and "estimated" coefficients, but for our purposes we are looking for the estimates that give us the best "fit." In other words, the values of the coefficients for which our predicted index values come closest to the actual index values. To determine these

values we use a technique called regression analysis. (We call it "multiple regression" when there are two or more independent variables.)

We obtain coefficient estimates, at this point in the process usually designated as a lower case b (e.g., b_1, b_2), from the regression analysis. These tell us, as best we can determine within the confines of the model, the relationship between the independent variables and the dependent variable. The next step is to take these values and apply them to the data for the applicants. We have the LSAT and GPA scores for the applicants, but, as yet, no index value. The point, of course, is to determine some index value as an aid in the admission process. In the 8/27/04 letter from Lynne Norton, a copy of which you gave me, the results of the regression using Iowa's data for the relevant years was:

$$I = (0.409) \times \text{LSAT} + (3.559) \times \text{GPA} - 1.290$$

Again, the point of this equation is to give you the best idea of an applicant's future success in his or her first year law school grades based on these two variables. The coefficients can be thought of as the weights given to each variable given both their importance as a predictor and their size (based on units and such). Although Norton also refers to them as "constants," that is incorrect usage in this context

Following up on that point, the values for the coefficients will change from year to year given the specific data for the students included in the original estimation. Norton notes that the Iowa formula is based on three years worth of data (specifically 2000, 2001, and 2002 for these values), which helps to smooth out any yearly variations.

Because of the difference in the scales of the underlying variables, the coefficients for LSAT and GPA cannot be directly compared to determine their importance relative to each other. That the coefficient for GPA is much larger than for LSAT is largely a function of the much higher values for LSAT. To directly compare the coefficients one must first "standardize" them, which is what Norton does in the section of her letter labeled "Calculating relative weights."

Although I would prefer to see the original data on which her figures were based, her procedure for calculating the standardized coefficients is correct. There are, however, three caveats to then reaching the conclusion that, for the data used, the LSAT is weighted at 62.6% relative to the GPA. First, although the standardized coefficients are "dimensionless," there are still issues concerning the range of values for the underlying independent variables. In this case, the values for the LSAT variable range over 60 units while that of GPA only two (Norton mentions a high of 4.33, but I'll assume 4.00 for now and make a comment later about this.)

Second, the procedure also assumes that the two independent variables are actually independent from each other. In other words, that there is little or no correlation between LSAT and GPA. In fact, however, we know that the two are correlated, which requires one to question the meaningfulness of the final figures.

Third, even to the extent they are correct, they are based on the specific data for a set group of students. My concern was with how the structure of the formula treated the two variables that contributed to the index total. As an illustration, think of the index as a container into which you are going to pour some combination of two liquids. The container has a maximum volume. You also have a maximum amount of each liquid you can pour into the container, though the maximum for each liquid is less than the maximum for the container. If you pour the maximum amount for each liquid into the container it will be full. On the other hand, you can pour less than the maximum for either or both liquids. In my formula, I look at the maximums and determine how much each is weighted based on the maximum amount of each liquid. LSAC looks at the actual amount of the two liquids that each person (applicant) pours into the container and determines the average. Rather than rely on the container illustration to make my point, let me turn to some actual formulas.

The example I use to illustrate this in one of my classes comes from *Hopwood v. Texas*, 78 F 3d 932 (5th Cir. 1996). In footnote 1 the court notes that when the UT law school used the old two-digit LSAT scale (10 to 48) it wanted to place 60% of its index weight on the LSAT and 40% on the GPA. To do this it used the formula:

(1.25 x LSAT) + (10 x GPA) = Texas Index (TI)

With a maximum LSAT score of 48 and a maximum GPA of 4.0, this results in the following maximum TI of:

(1.25 x 48) + (10 x 4.0) = 60 + 40 = 100

It's pretty easy to see that the LSAT contribution to the maximum index value is 60% of the maximum of 100. Of course, the contribution of any individual's scores to his or her TI will vary. Thus, for example, a student with a 4.0 who scored 40 on the LSAT would have a TI of 50 + 40 = 90, and the LSAT portion would constitute 55.6% of that total. Even so, the formula is structured so that the LSAT in the abstract is weighted at 60%.

Although it's not directly related to the main point of this discussion, note that the court goes on to tell us that under the new, three-digits LSAT scale the formula was changed to

LSAT + (10 x GPA) = TI

The goal of the law school was to have the same 60-40 weighting, but this was not the case. Specifically,

180 + 40 = 220

and 180 is 81.8% of 220. Thus, during the transition period when some students had taken the LSAT on each of the scales, those who took the earlier version had their GPAs weighted more heavily than those who took the later LSAT version that used the three-digit scale.

I suppose I should also note here that UT artificially constructed its formulas. The goal was to apportion the weights in a particular way. They were successful with the first, but unsuccessful with the second. Although a few schools still use such a constructed formula, most others, like Iowa, now use an estimated formula that lets the data determine the relative weights of the two independent variables

Before moving to the Iowa formula, let me give another example from one of my classes. (This is part of the material I have posted on my website in case you've already seen it.)

The key to determining the relative weight in a scale situation such as this is to use the maximum values for each component to determine how much each contributes to the total. For example, in one of my courses I give two tests and a paper. The first test is worth 30 points, the paper is worth 30 points, and the second test is worth 40 points. As they total to 100 points, each constitutes the same percentage in the total as the number of points. What matters is the amount that each component contributes to the total possible value. In addition, and as opposed to the total, any particular applicant's index value may be made up of a different ratio of GPA and LSAT values depending on the person's score. Using my course example, a person could have gotten 50 points by scoring 30 on the first test, 20 on the paper, and 0 on the second test. A second student could have gotten 0, 30, and 20. Each student's score is made up of a different percentage of the three assignments, but the overall weight of the assignments stays the same (30, 30, 40). Put another way, the weight the school puts on the components is not necessarily the same as how much each component contributes to any given applicant's index score.

Returning to the Iowa formula we have the following:

$I = (.409 \times 180) + (3.599 \times 4.0) = 73.62 + 14.40 = 88.02$

With 88.02 as the maximum index value (when just considering the LSAT and GPA portions), the LSAT portion constitutes 83.6%, compared to 16.4% for the GPA.

By itself, this result is not necessarily good or bad. Given variations among the applicants' undergrad schools and majors, I can certainly see why the LSAT might be the better predictor of later law school grades—it is basically the great equalizer. My concern was more with the fact that some schools might not realize the relative weights they are giving the values (e.g., UT and the change in scales) or might be a little deceptive in their descriptions of their procedure (e.g., suggesting to applicants a 60-40-ish weighting when it's actually quite different).

I can see where law schools would not want to panic those preparing for law school by emphasizing that the LSAT is much more heavily weighted, and I've certainly encountered many students who get themselves in a tizzy about it, but I also think that it's better to be clear about the weighting so that a student can adjust his or her preparations accordingly. In addition, by the time a student is thinking about taking the LSAT it is often too late to do much about a low GPA. It comes as some comfort, therefore, to learn that all is not lost in terms of getting into a good law school.

Two final points The first concerns the maximum GPA. In the calculations above I used 4.0, but Norton used 4.33 in hers. The difference doesn't make much of a change in the numbers above, but it does raise some questions about how the additional value is handled. If Iowa uses a 4.33 maximum, is it for all students, including those who come from schools that only have a 4.0 maximum? That would seem to put such applicants at a disadvantage. I haven't looked at this closely in some time, so perhaps nearly all undergrad programs now have a 4.33 maximum. It used to be that some law schools would just truncate anything over a 4.0, but perhaps that has changed as well. Regardless, I note this in passing just in case it's not something that you've considered

Finally, despite all the above, Norton's point about standardized coefficients is probably worth mentioning, so I will likely post a version of all this on my website.