DR. ARMOR'S REGRESSIONS

JAN DE LEEUW
PROFESSOR AND CHAIR
UCLA DEPARTMENT OF STATISTICS

1. Gap Analysis

1.1. **Definition.** *Gap analysis* is defined by Dr. Armor as the following two-step procedure. It is not a standard or particularly well-known data analysis procedure, but there are many variations of it in the social and educational research literature. As we shall argue below, all these variations share the same fundamental flaw.

There is a binary variable `BLACK` in the regression analysis indicating if a student is black or not. We first do a regression analysis using only the `BLACK` variable (and a constant). This gives a regression coefficient for `BLACK`, say β_0, which is simply the difference in mean test score of blacks and non-blacks. This is called the *raw gap* (for this particular test and this particular group of students).

The next step in gap analysis is to attempt to "correct for background". We do a second regression analysis on the same data, where we now use the variable `BLACK` together with some other variables we want to "correct" for. These are often called *covariates* or *background*. In Dr. Armor's case these are `FREELUNCH`, `FIRSTGRADE`, and `GENDER` (and maybe `MAGNET`). Each of these variables gets a regression coefficient. Say the regression coefficient for `BLACK` in this second regression analysis is β_1. This is the *corrected gap*.

Finally, Armor argues that the background "explains"

$$\left\{\frac{\beta_0-\beta_1}{\beta_0}\right\}\times 100\%$$

of the raw gap. Thus if $\beta_0=-23$ and $\beta_1=-5$ (as we have for reading in grade 3), then the background "explains"

$$\left\{\frac{23-5}{23}\right\}\times 100\%=78\%$$

of the gap.

Thus if the background is social economic status (SES), as Dr. Armor seems to think, then SES explains the race gap almost completely, and there is very little room left for the influence of school-based variables.

1.2. **Criticism.** Anybody with any formal training in statistics at all can point out the obvious flaws in this argument. There is generally nothing wrong with the computations, but the interpretation of the results, and the discussion based on these interpretations has nothing to do with either statistics or science. Let us discuss some of the key points in detail.

1.2.1. *Explaining.* Regression analysis by definition does not "explain" anything. Scientific explanation requires embedding empirical results in a theoretical framework, which can then subsequently be tested empirically. Regression analysis merely establishes correlations between certain selected variables in certain selected groups of students.

1.2.2. *Causal Attribution.* It is *impossible*, on the basis of a regression analysis with correlated predictors, to state that a certain percentage of the variance of the outcome can be attributed to the effect of predictor A, and another percentage to the effect of predictor B. There is no consistent and logical way to do this, fundamentally because the Pythagorean Theorem only applies to rectangular triangles.

1.2.3. *Labeling.* Dr. Armor calls FREELUNCH a proxy for SES, in fact he seems to treat FIRSTGRADE and even GENDER as proxies for SES as well. Since nobody owns rights to the "correct" definition of SES, there is nothing inherently wrong with this. But it obviously is misleading. In subsequent discussion Dr. Armor argues that he could make the gap even smaller by including more SES proxies. Of course this is true, but it has nothing to do that these are SES variables. They merely have to be correlated with race.

1.2.4. *Weird Gaps.* There is nothing in the Gap Analysis procedure proposed by Dr. Armor that guarantees that β_0 is larger than β_1. In fact, it is easy to construct examples, where β_1 is smaller. In such cases, the background "explains" a negative percentage of the gap, whatever that means. Although such examples are perhaps somewhat artificial, they do show that the procedure is fundamentally unsound.

1.2.5. *Correcting Everything.* In removing the "effect" of the background from the outcome (test) in a regression analysis, we remove at the same time the "effect" of the background on the variable BLACK. Thus the Gap Analysis does **not** quantify the effect of race on test score corrected for background, but it quantifies the effect of *race corrected*

for background on *test score corrected for background*. If the background is correlated strongly with race, then the variable *race corrected for background* will be very different from the original race variable.

2. HLM versus OLS

In his discussion of `HLM`, Dr. Armor shows again that he is not trained as a statistician, and consequently he makes a number of obvious blunders.

2.1. **Meaning of HLM.** Dr. Armor confuses `HLM`, a computer program, with multilevel analysis, a statistical technique. Bryk and Raudenbush have produced the program `HLM`, and they have promoted its use in the educational community, but the technique has been around a very long time. Goldstein and Longford in England, statisticians such as Morris, Lindley, Novick, quantitatively oriented people in education such as Cronbach or Burstein, and sociologists such as Hauser or Robinson have been using these techniques for a very long time.

2.2. **Snowjob.** It is misleading to say that `OLS` and `HLM` both have their strengths and weaknesses. For hierarchically structured data sets, with variables both from the individual and school level, using `OLS` gives estimates of the regression coefficients which are unbiased but inefficient, and using `OLS` gives standard errors (and thus significance tests) which are often completely wrong. This has been discussed in enormous detail in the statistical, sociological, and educational literature for at least thirty years.

2.3. **In Praise of OLS.** Dr. Armor says that `OLS` is the most widely used and best understood statistical method in the social sciences. Both statements are wrong. The t-test is both more widely used and better understood, and so is the cross-table. In fact, multiple regression may be well understood from the computational and mathematical point of view, but its use is (after about 100 years of study) not at all well understood in non-experimental situations. Dr. Armor then says that OLS is quite robust. It is not that either. Regression coefficients in the case of multicollinearity are unstable, standard errors of regression coefficients vary widely depending on assumptions of the regression model. Dr. Armor goes on to say that OLS has the advantage of providing conservative estimates of school effects. If anything, this seems to be a political statement.

2.4. **Surpressors.** One of the main reasons why educational statisticians such as Cronbach and Burstein became interested in multilevel analysis was because they wanted to separate the effects of school-level variables from individual variables. In an analysis in which these variables are thrown together, and in which the fact that there is within-school correlation is ignored, the regression coefficients for the school variables are artificially suppressed. Multilevel analysis leaves room to capture both within-school and between-school variation.

3. Student and School Characteristics

In his regression analysis (on page 6/12 of the supplementary report) in which third grade MSPAP scores in reading and math are predicted from three individual and five school variables, Dr. Armor commits the usual sins. In the early sixties Lee Cronbach of Stanford wrote extensively on the fact that almost all educational analyses using both school and individual variables were misleading and basically flawed. In order to separate the between and within school variation, Cronbach and other have suggested to transform the individual level variables to deviations from the school mean. This makes them uncorrelated with the school variables, and thus we eliminate the most important source of confounding.

Dr. Armor ignores this discussion completely. Let me put this in perspective. The Cronbach type of regression, and other forms of ecological regression in sociology, were proposed in the sixties and seventies to replace the type of OLS analysis that Dr. Armor does. Burstein's "slopes-as-outcomes" analysis was proposed in the early eighties to refine the Cronbach-type of analysis. The method implemented in `HLM`, and in comparable programs such as `MLn` or `PROC MIXED` in `SAS` or `BMDP5V`, were suggested to make "slopes-as- outcomes" analysis statistically rigorous. Thirty years of research on this topic have passed Dr. Armor by. Either he is incompetent in the area of regression analysis, or he very much likes the fact that `OLS` squashes the school-level regression coefficients. Or both.

That the correlation between aggregated characteristics is higher than between individual characteristics is well known since the beginning of statistical time. In sociology, this is known as the "Robinson Effect", in general using aggregated and individual variables interchangeably in the same analysis is known as the "ecological fallacy" or the "aggregation fallacy".

It is of some interest that these regression analysis were done for black students only. There is nothing wrong with this, as long as conclusions are limited to that group as well. Obviously nothing at all

can be said about the impact on "black achievement relative to white" (Armor Comments on the Stringfield Report, page 2).

Dr. Armor argues that teacher resources which benefit all students cannot reduce the achievement gap. This is politics, not science. In the first place, it should be "teacher resources which benefit all students equally". There could easily be interaction with background, in fact "slopes-as-outcomes" theory and Cronbach's "aptitude-treatment-interaction" strongly suggests there is such an interaction.

Also, if we improve both black and white scores equally, their difference may remain the same, but their ratio becomes closer to unity. A white score of 15 is 1.5 times as large as a black score of 10, but 115 is only 1.05 times as large as 110. There is no inherent reasons why the gap should be measured by subtraction.