Assessment of Performance in Introductory Linear Models Courses

Earl Jennings - Austin, Texas, USA
Joe H Ward Jr - San Antonio, Texas, USA

1. Background

We have been teaching a General Linear Models approach (Ward and Jennings, 1973) to problem formulation and analysis to university students and applied researchers for over 25 years. The range of objectives and the heterogeneity of backgrounds create difficulties in the assessment of performance. Among these difficulties are defining the objectives of the instruction, developing instruments to evaluate student performance, and methods of awarding grades.

2. Defining objectives

The objectives of the course we teach are first identified at a conceptual level and then translated into specific exercises, activities, or examinations. The following are among the conceptual level objectives or goals.

(i) Imbed the mathematics required for understanding into the course. Courses that do not require a very extensive mathematics background frequently are taught at an intuitive level without rigor; or even worse, they are taught at an arithmetic level which puts the student in competition with a computer, a contest the student is sure to lose. In our judgement, a course that does not have a fairly rigorous mathematical base is likely to do the student a disservice. On the other hand, we are inclined to the view that in many applied areas, a heavy mathematical prerequisite is a sure way to guarantee that a course will disappear because of lack of enrolment. The problem is to identify those mathematical concepts that are essential and imbed them in the course to facilitate the development of mathematical skills necessary for statistical reasoning and argument.

(ii) Keep the technical vocabulary, special symbols, and computing formulae to a minimum. This is in keeping with our view that ideas are more important than the words used to describe the ideas. A technical vocabulary does more to speed communication than it does to enhance it. Mathematicians have long recognised the value of attempting to reduce or transform new problems into previously solved problems. For example, if one deals with the general notion of the standard deviation of the sampling distribution of a linear combination of the parameter estimates of a linear model, the formula for and the concept of standard errors of: means, differences between means, slopes, differences between slopes, adjusted means, differences between adjusted means, etc. are all just special cases of a more general formula and a more general concept.

(iii) Emphasise the notion that a model is a way of formalising an argument. Students in courses frequently know what kind of analysis they are expected to perform because they know what chapter in the book they are studying. Problems that arise in the real world are not accompanied with such information. We also attempt to discourage the notion that one can tell what to do by the nature of the data. We believe what one does should be dictated by the question one is trying to answer rather than by the data. In general, we believe that it is better to have an approximate answer to an interesting question than it is to have an exact answer to an uninteresting question, and that there may be a number of different models that will lead to useful conclusions.

(iv) Stress the conceptual importance of distinguishing between statistical inference and practical or substantive inference. We attempt to do this by giving the student many opportunities to state questions in a natural language or non-technical form (i.e. is Fertilizer A better than Fertilizer B?) and hypotheses in a parametric form (i.e. is β=0?).

(v) Provide opportunities to practice what we call "investigating the properties of models". Because, in general, our approach requires that students be able to create models that have specified properties, it is important that they develop the skill of verifying that a created model does in fact have the properties desired. In addition, such a skill also makes it possible to investigate the properties of models that have been created by others. This turns out to be a particularly useful skill in attempting to determine how to interpret the outputs from popular computer packages such as PROC GLM of SAS (1985). A demonstration of this can be found in Jennings and Ward (1982) in response to a problem suggested by Freund (1980).

The kinds of items and problems that we include on exercises and examinations may be inferred from the objectives. We have discovered that if an out-of-class exercise is called an out-of-class examination, students seem to work more diligently. Also, very short (perhaps 10 minutes) in-class examinations keep students on their toes. These examinations do not have to be "pop quizzes"; they can be announced, even scheduled.

3. Awarding grades

Some instructors feel that when they are placed in the role of an evaluator, an adversarial environment is created that does not conduce to learning. Thus a requirement to assign grades is viewed as an unattractive aspect of teaching. Many institutions now require that the process of awarding grades be made visible to the students. Such a requirement suggests that it might be desirable to make the final grade assessment formula-driven, for example, using a composite score that is a weighted sum of performance indicators such as examinations, papers, projects, attendance, etc. One of the major advantages of a formula-driven assessment process is that students may be encouraged to attempt exercises and activities that do not appear in the formula. If students are afraid that a revelation of ignorance will "count against them" they may attempt to minimise the number of opportunities they have to reveal ignorance. If the final grade is formula-driven and some features of the course are not in the formula, then presumably it will be clear that such features cannot lower the final grade. On the other hand, there are some students who will believe that if a particular activity is not in the formula, then it must not be important. For these and other reasons, conventional weighting schemes frequently fail to produce a grading policy that is acceptable to the instructor.

Consider, for example, an instructor who is teaching a course that is believed to have content that is hierarchically structured as we believe ours to be. In such a course, it is not unusual for a student to have a great deal of difficulty until "all the pieces fall into place" and the instructor might very well want to adopt a grading policy that results in a high final grade for superior performance on a final examination even if performance on other indicators has been relatively poor throughout the course. Not only does such a policy reward the student who finally does integrate the material, but it has the potential of encouraging effort throughout the course because a high final mark is never out of reach. Unfortunately, the common policy of weighting the final heavily and the other indicators little if anything may send the wrong signal. If students believe that activities other than the final "don't count" they may neglect the very behaviours that would increase the likelihood of superior performance on the final examination. What is needed is a policy that weights the final examination heavily when performance on other tasks is poor but less heavily when performance on other tasks is superior.

Figure 1 (see overleaf) contains a sketch of the desired relationship. The vertical axis is the composite score (C) and the horizontal axis is the final examination score (E). The line labelled "low value of P" describes the desired relationship between C and E for students with low performance (P) on tasks other than the final examination. The line labelled "high value of P" describes the desired relationship between C and E for students with high performance on P. The other two lines in Figure 1 demonstrate the desired relationship between C and E at two intermediate values of P.

It turns out that an equation with three weights will produce the grading policy sketched in Figure 1. As an example, suppose an instructor decides that C, E, and P are all to have a range of zero to 100. P might be a sum or average of other indicators. The form of the equation that will produce Figure 1 is $C = aE + bP + c(E*P)$ where a, b, and c are to be determined. We can find these by feeding in values of P and E for three hypothetical students. The following are suggested:

$$100 = a(100) + b(100) + c(100*100)$$
$$100 = a(100) + b(0) + c(100*0)$$
$$60 = a(0) + b(100) + c(0*100)$$

The first expresses the desire to give a high value of C when P and E are both high, the second the desire to give high C when E is high even if P is low, and the third the desire to give moderate C to a student with low E and high P. The solution for the three simultaneous linear equations is $a = 1$, $b = .6$ and $c = -.006$. Thus the equation that produces the desired result is:

$$C = 1*E + .6*P - .006*(E*P).$$

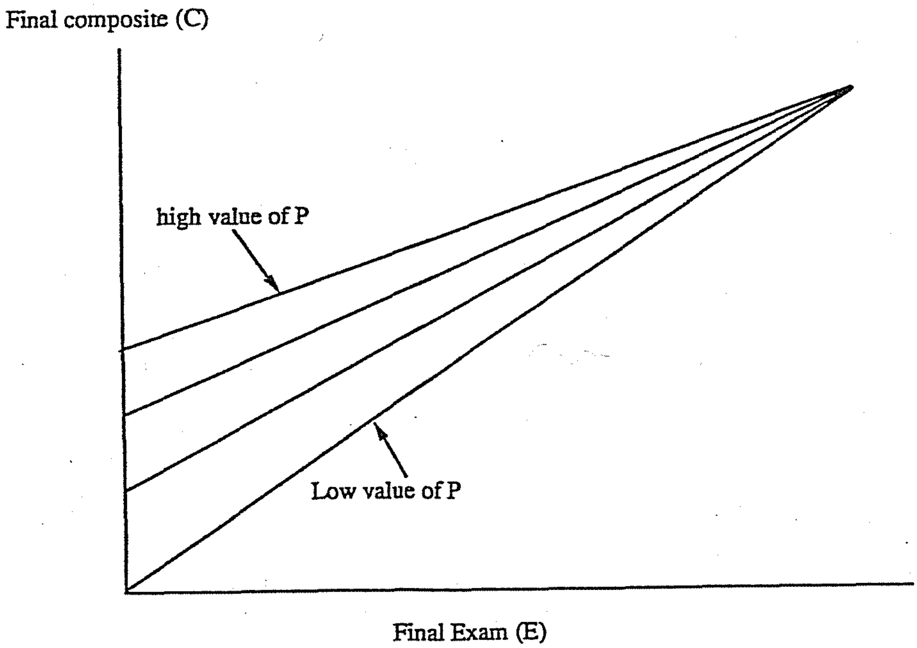

FIGURE 1
Relationship between C and E at various levels of P

This process of modelling human judgement is described in detail in Ward (1979). Other policies can be obtained by choosing other values for C, E, and P in the three equations. For example, some might feel that to assign a C value of 100 to a student with a P value of zero as in the second equation above is weighting E too heavily. If the second equation is changed to:

$$90 = a(100) + b(0) + c(100*0)$$

and the three linear equations solved simultaneously, the policy equation becomes

$$C = .9*E + .6*P - .005*(E*P).$$

In general, any policy that can be represented graphically can be reduced to a series of equations that can be solved and the result is a formula-driven grading policy.

References

Freund, R J (1980) The case of the missing cell. *The American Statistician* 34, 94-98.
Jennings, E and Ward, J H Jr (1982) Hypothesis identification in the case of the missing

cell. *The American Statistician* **36**, 25-27.

SAS Users Guide : Statistics, Version 5 Edition (1985) SAS Institute, Inc, Cary, NC.

Ward, J H Jr and Jennings, E (1973) *Introduction to Linear Models*. Prentice-Hall, Englewood Cliffs, New Jersey.

Ward, J H Jr (1979) Creating mathematical models of judgment processes : from policy capturing to policy specifying. *Journal of Experimental Education* **48**, 60-84.