BEST LINEAR PREDICTION OF COMPOSITE UNIVERSE SCORES

DAVID JARJOURA

AMERICAN COLLEGE TESTING PROGRAM

The problem of predicting universe scores for samples of examinees based on their responses to samples of items is treated. A general measurement procedure is described in which multiple test forms are developed from a table of specifications and each form is administered to a different sample of examinees. The measurement model categorizes items according to the cells of such a table, and the linear function derived for minimizing error variance in prediction uses responses to these categories. In addition, some distinctions are drawn between aspects of the approach taken here and the familiar regressed score estimates.

Key words: best linear prediction, test specifications, generalizability theory, measurement error variance, regressed score estimates.

Introduction

The problem of predicting universe scores for a sample of examinees from their observed responses to samples of items is described. Though the solution provided here is general, discussion of its application is restricted to a common situation in standardized testing. The situation is one in which many test forms have been or will be generated, each of which has been or will be administered to a different sample of examinees. The notion of an overall measurement procedure refers to the administration of potentially many test forms to different samples of examinees. The specific problem is to determine, for the overall measurement procedure, the linear combination of item responses which minimizes the error variance in predicting composite universe scores.

In the context of measurement consistency across the multiple forms, the focus here is on the categories or cells of a table of specifications used to generate these forms; i.e., the categories of items in such a table are viewed as fixed conditions for the overall measurement procedure. Though the categories remain constant across forms, the items in them change and thus constitute a random dimension in the measurement procedure. This perspective is discussed in Jarjoura and Brennan [1982] and is associated with multivariate generalizability theory [Cronbach, Gleser, Nanda, & Rajaratnam, 1972, chaps. 9 & 10]. The term composite universe score is adopted because the measurement model defines a universe score for each category of the table of specifications. Since a single score is of interest, a composite of the category universe scores is defined.

The general theoretical basis for the solution to the problem can be found in Searle [1973, 1974] as well as other sources. Because the model used here for the overall measurement procedure is a variance components model, Searle's [1974] focus on the prediction of random effects in such a model is of special relevance. The general problem discussed in Searle [1974] is that of predicting an unobservable random vector from an observable one. Essentially $u$ and $y$ are jointly distributed vectors of random variables such that those variables in $y$ are observable and those in $u$ are not; and $u$ is to be...
predicted from realized values of $y$. Call $\hat{u}$ some function of $y$ which predicts $u$ such that

$$E(\hat{u} - u)'Q(\hat{u} - u)$$

is minimized, where the expectation is over the joint distribution, and $Q$ is any positive definite symmetric matrix. Such a prediction function is often referred to as "best," and for any $Q$ the best function is simply the regression function, $E(u | y)$, where $y$, here, denotes a realization of the vector of variables defined above and denoted also as $y$. [Note that $y$ will be used to denote a vector of random variables in most contexts and a vector of realized values in a few others. A dual use will also be made of $u$, $\hat{u}$, $y^*$ (both symbols to be defined), as well as elements of these vectors. This will maintain the convention of uppercase for matrices and lowercase for vectors, and should not result in any confusion.]

The regression function requires knowledge about the joint distribution of $u$ and $y$, and unless some strong assumptions are made, such knowledge may be difficult to obtain. However, if the prediction function is restricted to one that is linear in $y$, then knowledge of the first and second moments of this distribution is all that is necessary to minimize (1) for this class of functions.

Searle [1973], among others, shows that the Best Linear Prediction (BLP) function of $u$ is

$$\hat{u} = E(u) + CV^{-1}(y - Ey),$$

where $C = \text{COV} (u, y) \equiv E(u - Eu)(y - Ey)'$, $V = \text{VAR} (y) = E(y - Ey)(y - Ey)'$. Note also that this is the regression (best) function under multivariate normality [Searle, 1974].

Besides minimizing (1) (for any $Q$) for the class of linear functions of $y$, two other well-known properties of such a function are:

$$E(\hat{u}) = E(u);$$

and for this class, $\rho(\hat{u}_j, u_j)$ is maximized which is to say the correlation between the $j$-th elements of $\hat{u}$ and $u$ is maximized.

Other results which will be used later are:

$$\text{VAR} (\hat{u} - u) = \text{VAR} (u) - CV^{-1}C';$$

$$\text{COV} (\hat{u}, u) = \text{VAR} (\hat{u}) = CV^{-1}C';$$

and

$$\rho(\hat{u}_j, u_j) = \left[ C_j'V^{-1}C_j \right]^{1/2} \left[ \text{VAR} (u_j) \right],$$

where $C_j$ is the $j$-th row of $C$.

Of course, these properties and results depend upon knowledge of the parameter values of the BLP function. For the type of application described here, it is argued that highly accurate estimates of these parameters are often readily available. The purpose then is to use these results in deriving a functional form for a measurement model that considers the general situation in which multiple test forms are constructed according to a table of specifications. First the measurement model is presented, then the BLP function is derived, and finally an example is provided.

Though not specifically suggested for the problem described, there is a clear precedent for minimizing measurement error variance through what has been referred to as "regressed score estimates." The approach taken here might even be viewed as one type of generalization of these familiar estimates. However, there are distinctions with some current perspectives in this area. For example, in their discussion on regressed score estimates in the context of generalizability theory, Cronbach et al. [1972, chaps. 5 & 10] emphasize problems in the practical implementation of such estimates without the classi-