DOCUMENT RESUME

ED 089 676 IR 000 388

AUTHOR	Harris, Chester W.; McArthur, David
TITLE	Studies of Two Sets of Variables.
INSTITUTION	California Univ., Santa Barbara. Bureau of Educational Research and Development.
REPORT NO	RB-1
PUB DATE	Apr 74
NOTE	9p.; Paper presented at the American Educational Research Association Annual Meeting (Chicago, Illinois, April 15 through 19, 1974)
EDRS PRICE	MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS	*Data Analysis; *Methodology; Methods; *Methods Research; Research; *Research Design; *Research Methodology; State of the Art Reviews; Techniques
IDENTIFIERS	*Two Battery Studies

ABSTRACT

A review is presented of several types of two-battery studies and of methods for analyzing the resultant data. Two-battery studies are defined as those in which observations on two sets of variables are available for the same sample of subjects. Several types are identified and described, including: 1) prediction studies, 2) studies of domain relationships, 3) studies of equivalence of samples of variables drawn from the same domain, 4) discriminant or dispersion studies, 5) battery reliability studies, 6) studies of change or development, and 7) studies of the relation of observed to theoretical variables. Methods of data analysis identified include regression analysis, canonical variate analysis, interbattery factor analysis, and multimode factor analysis. (Author)

28 02

Studies of Two Sets of Variables

Chester W. Harris

David McArthur

University of California, Santa Barbara

Presented at American Educational Research Association
Chicago, Illinois, April, 1974.

Research Bulletin No. 1
Bureau of Educational Research and Development
University of California, Santa Barbara

The purpose of this paper is to identify several types of two-battery studies and to comment on methods for analyzing such data. By a two-battery study is meant a study in which observations on two batteries or sets of variables are available for the same sample of subjects.* The primary purpose of such a study is to quantify relationships between the two batteries or sets of variables; the methods employed may be several. We first call to your attention two recent reviews which offer illustrations of studies of two sets of variables and comment primarily on canonical correlation and canonical variate analysis as methods of analysis. These reviews are by Weiss (1972) and by Darlington et al. (1973). We present here a slightly different analysis of two-battery studies than appears in either of these reviews, and we consider analogs of component and of factor analysis as applicable methods.

In considering types of studies it is helpful first to make a distinction between two-mode and three-mode data. Two-mode data may be conceptualized as existing in a matrix for which there are no linkages among the rows or among the columns. It is conventional to let the rows of such a matrix designate subjects; for this dimension lack of linkage simply means independence of the subjects, or members of the sample. The columns of this matrix designate the variables; the lack of linkage here means that although the variables may be segregated or classified into two sets, there is no linkage such as there would be if the first variable of set 1 were the first variable of set 2 administered at an earlier time. Thus two-mode data may have different numbers of variables in the two sets. If the numbers are

*We do not consider in this paper studies of the responses of linked pairs of subjects (such as siblings, husbands and wives, etc.) to the same set of variables.

equal, this represents a choice rather than a necessary feature of the data. In contrast, three-mode data may be arranged in a meaningful three-dimensional "matrix" as in the study of responses of the same sample of subjects to the same test battery on two different occasions. An analogy (which should not be pushed too far) with the difference between the one-way ANOVA and the two-way completely crossed ANOVA with attendant interactions suggests distinctions between studies involving two-mode and three-mode data.* For three-mode data the number of variables is necessarily the same in each of the two sets.

We now describe types of two-battery studies which involve two-mode data.

Type A: Study of a battery of predictor variables in relation to a battery of achievement or criterion variables. Alternatively, study of a battery of input variables in relation to a battery of output variables. There appear to be different definitions of predictor variables, including achievement variables at an earlier period in time employed as predictors. If identical tests are used as predictors and as criteria, then the study employs three-mode data and belongs with Type G, below. It helps to distinguish Type A from Types B and C below if we require that there be a necessary time lag between the gathering of the predictor (input) data and the gathering of the criterion (output) data for Type A.

Type B: Study of a battery of variables drawn from one domain (set of domains) in relation to a battery drawn from another domain (set of domains). Examples of two different domains (or sets of domains) would include: achievement in two different

*Sets of data involving four or more modes may also be conceptualized. See Horst (1963).

subject-matter areas, process variables vs. outcome variables in learning studies, personality variables vs. interest measures, affective measures vs. cognitive measures, factorial domain 1 vs. factorial domain 2, etc. In Type B studies there is not a <u>necessary</u> time lag between administrations of the two batteries; in fact, both batteries might be administered in a randomized order during the same gross time period.

Type C: Study of a sample of variables drawn from one domain (set of domains) in relation to a second sample of variables drawn from the same domain (set of domains). In Type B, the relations between different domains are being studied; in Type C, the equivalence of variables from the same domain (set of domains) is being studied.

Type D: Study of a battery of substantive variables in relation to a battery of dummy (classificatory) variables. This type of study handles the case of two or more categories or groups of subjects observed on the same set of substantive variables, and consequently is related to discriminant or dispersion analysis. As is well known, if there are only two categories of subjects (the Fisher two-group discriminant case) only one dummy variable exists, and the study is in effect a multiple regression study with the substantive variables playing the role of independent variables and the single dummy variable the role of the dependent variable.

Type E: Study of a battery of observed variables in relation to a battery of theoretical variables. An example is Rao's (1955) use of canonical correlation theory to relate observed scores to theoretical factor scores.

Studies of two sets of variables which employ three-mode data necessarily have the same number of variables in each of the two batteries. Such studies may be classified as follows:

Type F: Studies of the relations between two equivalent or parallel batteries: i.e., batteries made up of equivalent or parallel forms of the same tests. In such studies one generally would administer the two batteries within the same gross time period rather than on distinctly separate occasions as in Type G.

Type G: Studies of the relations between the "same" variables observed on two separate occasions. This type includes studies of "natural" or "normal" change or development. Alternatively these are studies of stability over time of a set of variables. Studies of the relations between the "same" set of variables administered before and after a particular treatment is applied to the subjects involve two separate occasions and may be classified here; however, in this latter situation the interest is in the effects of the intervention.

Type H: Studies of the relations between the "same" variables administered under two different conditions. An example would be a self-concept inventory administered under "your ideal self" and "your self as you are" conditions. For any individual subject, two different occasions are involved since he is unable to take the test under the two different conditions simultaneously;

randomizing over subjects with respect to the two occasions is possible for this type, whereas it is not possible to randomize over subjects with respect to the two occasions in Type G. A data gathering procedure that involves both scales and content in a completely crossed design usually constitutes three-mode data. Thus, applying several semantic differential scales to two different concepts (or applying two different scales to several concepts) belongs to Type H, as do "two-method, multi-trait" or "two-trait, multi-method" studies.

Type I: Study of the relations between observed variables and theoretical variables, such as underlying true scores, where each theoretical variable is linked to a specific observed variable. This specific linkage distinguishes Type I from Type E.

Since we have space for only a few comments on methods of analysis, we shall consider here only studies involving two-mode data (Types A through E). Studies involving three-mode data present problems that deserve a more extensive analysis.

In Types A, B, and C, one seldom is interested in comparing the means of the variables; instead, it is the relations among the variables that are of primary interest. With two sets of (non-linked) variables and one sample of subjects it is possible to build a supermatrix consisting of the variance-covariance matrices for each set of variables plus the cross-covariance matrix and its transpose. Such a supermatrix contains no information about the means of the variables. It is also possible to rescale the variables and thus alter this supermatrix so that its diagonal elements are each unity and its off-diagonal elements are correlation coefficients. Certain analytic procedures are scale independent and thus can yield similar findings for the two forms of the supermatrix. When this holds, we often prefer to make the analysis on the correlation matrix -- that is, the covariance matrix in standard form.

A number of matric equations describing regression analyses, canonical variate analyses, and interbattery factor analysis for the supermatrix in standard form are given in Harris and Harris (1973, pp. 171-176). If a matrix multiplication package is available at a computing center, these equations can be used to program the main analyses one is likely to want. Both the regression and the canonical variate analyses are scale independent; however, the Tucker (1958) interbattery analysis described by Harris and Harris is not. One could develop a scale independent interbattery analysis by following Kristof's (1967) lead and requiring that the initial analysis on which his procedure is based be a scale independent factor analysis of the complete supermatrix. It is important to recognize that a principal components analysis of the complete supermatrix would not provide the proper basis for the Kristof procedure.

Regression analyses for estimating any one variable (or any composite of variables) in one set from the variables in the other set or from the Tucker interbattery factors that are common to the two sets of variables can be performed. The various canonical correlation coefficients, the weights for forming each of the pairs of canonical variates, the correlations of the variables in each of the two sets with the canonical variates in each of the two sets, and analogs of interbattery factors can all be made available. Finally, the correlations of each variable with the interbattery factors, i.e., the factors common to the two sets, can also be produced routinely. A procedure for determining the number of statistically significant canonical correlations exists and can be found in a text such as

Tatsuoka's (1971). The redundancy index proposed by Stewart and Love (1968) is intended to define the proportion of overlapping variance between two batteries; Nicewander and Wood (1974) have criticized it severely, but - we believe - mistakenly.

Studies of Type D differ in an important respect. It is well-known that for the case of only two groups of subjects, and thus only one dummy variable, the canonical correlation analysis is related to the Fisher two-group discriminant analysis and it in turn to the Hotelling T^2 analysis, which provides a significance test for the two vectors of means (See Fisher, 1938). Thus, in a Type D study one is interested in the means of the variables for the various groups, and in general would like to test hypotheses about the separation, in the substantive variable space, of the group centroids. With more than two groups there is interest in the dimensionality (and configuration) of the group centroids. A Type D study can be analyzed by means of canonical variate analysis of the substantive and the dummy set of variables as well as by the discriminant-dispersion techniques. (See Tatsuoka, 1971, pp. 177-183).

Although we have listed Type E among this set of study types, it probably is true that Type E refers primarily to theoretical formulations rather than to empirical studies per se. As such Type E merely reminds us that a technique such as canonical correlation analysis may provide a basis for the solution of interesting theoretical problems.

REFERENCES

Darlington, R. B., et al. Canonical variate analysis and related techniques. Review of Educational Research, 1973, 43, 433-454.

Fisher, R. A. The statistical utilization of multiple measurements. Annals of Eugenics, 1938, 8, 376-386.

Harris, M. L. and Harris, C. W. A structure of concept attainment abilities. 1973, Wisconsin Research and Development Center.

Horst, P. Multivariate models for evaluating change. In Harris, C. W. (Ed.) Problems in measuring change, 1963, University of Wisconsin Press.

Kristof, W. Orthogonal interbattery factor analysis. Psychometrika, 1967, 1967, 32, 199-227.

Nicewander, W. A., and Wood, D. A. Comments on "a general canonical correlation index". Psychological Bulletin. 1974, 81, 92-94.

Rao, C. R. Estimation and tests of significance in factor analysis. Psychometrika. 1955, 20, 93-111.

Stewart, D. and Love, W. A general canonical correlation index. Psychological Bulletin. 1968, 70, 160-163.

Tatsuoka, M. M. Multivariate analysis. 1971, Wiley.

Tucker, L. R. An interbattery method of factor analysis. Psychometrika. 1958, 23, 111-136.

Weiss, D. J. Canonical correlation analysis in counseling psychology research. Journal of Counseling Psychology. 1972, 19, 241-252.