Proceedings of the Society Multivariate Experimental Psychology (SMEP), European Branch

Society of Multivariate Experimental Psychology (SMEP)
European Branch
Sixth Conference held at Bad Homburg/Ts., F.R. Germany
September 12 - 13, 1974

Research problems in multivariate experimental psychology 1974: a brief account

Viktor Sarris (Ed.)
Psychology Department,
University of Frankfurt,
F. R. Germany

Thirty-five SMEP members (out of more than 100 members) participated in the 6th conference of the Society of Multivariate Experimental Psychology (SMEP), European Branch, which holds its conferences annually in different European countries. In accordance with the SMEP's general policy the papers dealt mainly with methodological problems of application of multivariate experimental techniques, mostly in the fields of clinical and educational psychology. The papers and discussions concentrated, therefore, on the methodology of new trends in multivariate psychology.

In order to make the past and present SMEP's prolific scientific work more public, the participants decided to publish the conference paper abstracts jointly.
Where indicated, full reports can be obtained from the respective authors.

The SMEP's 1975 conference will be held in Switzerland (Organizers: Dr. U. Baumann, Psychiatrische Universitätsklinik, CH-8029 Zürich, and Dr. L. Hursch, Bern).

Special thanks are due the Werner-Reimers-Stiftung, Bad Homburg, for supporting the conference's location and for funding the printing costs.

Multivariate experiment as a catalyst of psychological theory

Raymond B. Cattell, Department of Psychology,
University of Illinois, USA

The design of experiment in which many variables are simultaneously recorded did not become truly possible until statistical advances of the first part of this century made analysis of multivariate models possible.
However, natural conservatism caused a further lag, so that only in the last twenty years has the proper relation to classical bivariate experiment been focussed and the characteristic virtues of multivariate experiment fully recognized.

After a reminder of the comprehensive parameters and the taxonomy of experimental designs, this paper turns to the potential and the actual historical impacts of multivariate designs upon psychological theory. The potential for a powerful "catalytic" action on theory arises in the first place from the more comprehensive determination which the multivariate approach gives both in regard to empirical constructs and theoretical concepts. Secondly, it gains access to new domains through not needing to depend upon manipulation. Thirdly, it is likely to lead to quicker recognition of higher order interaction and moderator effects. Fourthly, it permits deductions from a systematic theory to be more efficiently and broadly examined.

The extent to which these and other potentials of the method have yet been realized in the historical advance of theory is next examined. The effect on generation of concepts with a low infant mortality has been far greater than most psychologists yet realize, as is illustrated in personality trait concepts and in the definition of anxiety and other states. Its capacity to open doors on which classical experiment has beaten in vain is illustrated by the opening up of information on the number and nature of human drives. The higher order and moderator theoretical developments are still awaiting

greater precision of experiment and statistical concepts. The provocation and testing of broader deductions is illustrated in structured learning theory. The speed of testing and eradication of false models and deductions is illustrated by what has happened to the early cruder use of "types" in psychology.

The forms of theory which multivariate methods have especially catalysed are those initially expressed in mathematical models and sometime foreign to verbal and intuitive approaches. If successful these produce a more radical break away from the persistence of verbal stereotypes and start entirely new directions of advance, as may be claimed for the multivariate concept of role action and structured learning theory. But the multivariate methods also have their shortcomings, e.g. the linear limitation in factor analysis, and their complexity leads to far more instances of erroneous or inadequate usage than with simpler methods. It may be that the extensiveness of the theoretical models which is more readily developed by multivariate methods could also lead to rigidity.

All in all, there is reason to believe that a much wider teaching of multivariate methods would add substantially to the development of theory in almost every area of psychology while making the mathematical-statistical multivariate designs and methods better adapted.

REFERENCE

Cattell, R.B., et al. Handbook of Multivariate Experimental Psychology (Chaps 1-3), Chicago: Rand McNally, 1966.

Remission and deterioration after short term counselling in schools

T. Christie and Una Maguire, Department of Education, University of Manchester, England

Research on the effectiveness of counselling is characterized by the use of criteria which are either univariate, e.g. undifferentiated recovery rate or concerned with secondary or even tertiary gains rather than the primary objective, e.g. prevention of delinquency or most frequently both i.e. the grade point average.

The present study takes as its criteria amelioration of the presenting symptoms as measured by Berdie's (1957) Minnesota Counselling Inventory (MCI) and Cattell's (1966) High School Personality Questionnaire (HSPQ). It is hypothesized that (1) short term counselling will have no effect on mean levels of disturbance and (2) short term counselling will increase the variance of disturbance scores.

Sample

Complete test-retest data (15 week interval) on the HSPQ and the MCI were obtained from 648 fourteen year old pupils drawn from 35 schools. The mean IQ was 112.3 (s.d.-12.47): 103 children had IQs in the 85-99 range, but were functioning at an adequate scholastic and verbal level to meet Berdie's advice: "The inventory should not be used with children of less than eighth grade reading ability". From this group, 54 children who showed some disturbance on the MCI were given individual counselling for 6 weekly 35 minute periods by the counsellor attached to their school. Pupils were not volunteers. Ten counsellors, all having received the same training (essentially Rogerian) were involved, counselling between two and ten pupils each. Post-testing took place approximately four weeks after the end of therapy.

A further sample of 54 children, matched as a group on sex, level of disturbance and type of disturbance (initial MCI scores) but not on school acted as controls. CPI and HSPQ disturbations for the total subsample involved in the experi-

ment and for the main sample were gained; also, brief descriptions of the scales have been proposed. As the HSPQ profile (HSPQ was not used in the selection of the experimental sample) showed the two tests were not independent: the CPI probably measures states rather than the HSPQ's traits.

Results

Multivariate analyses of variance were computed for the post-test scores of each test separately. The full model in each case had treatment (intervention vs. no intervention) , sex, treatment by sex interaction as factors and the matrix of initial scores on the test as covariates, to give best estimates of the population means for post-test scores. There were found no significant differences between centroids.

A fuller report, including some tables and figures, can be obtained from:

Mr. T. Christie,
Department of Education,
University of Manchester,
Manchester,
England.

Visual representations of multivariate data

B.S. Everitt, Institute of Psychiatry,
University of London, England.

Visual presentation is valuable throughout a statistical analysis. Most people are aware of the methods useful for the visual representation of univariate data, for example, histograms, frequency polygons, pie charts etc. Methods for the visual representation of multivariate date are less common and less well known.

In this paper three relatively new techniques for representing multivariate data visually will be discussed, and their behavior illustrated by applying them to various sets of artificially generated data.

Hopefully such techniques will be most useful for what may be described as the pre-analysis stage of any investigation. Perhaps the major advantage to be derived from using these methods will be in the heightened qualitative awareness of which numerical calculations may be relevant.

The three techniques to be described are those given by Sammon (1969), Andrews (1972), and Chernoff (1973). The main purpose of this paper is to introduce people to these techniques, in the hope that this will encourage people to "look" at their data rather more than is done at present.

REFERENCES

Andrews, D.F. Plots of high dimensional data. Biometrics, 1972, 28, 125-136.

Chernoff, H. Using faces to represent points in k dimensional space graphically.

Sammon, J.W. A nonlinear mapping for data structure analysis. IEFE Trans Computers, 1969, C 18, 401-409.

A fuller report can be obtained from:

Mr. B.S. Everitt,
Biometric Unit,
Institute of Psychiatry,
University of London,
Denmark Hill,
London, S. E. 5,
England.

Some Models and Estimation Methods for Analysis of Longitudinal Data

K. G. Jöreskog and Dag Sörbom, Institute of Statistics, University of Uppsala, Sweden

This paper is concerned with statistical methodological problems arising in the analysis of data from large longitudinal studies, where several quantitative psychological or educational measurements are obtained at two or more occasions from large groups of people. Particularly, the paper focusses on the following problems:

A. The treatment of measurement error in observed variables
B. The scaling of latent variables (hypothetical constructs)
C. Design problems, i.e., the choice of measurements to adequately measure the latent variables and the relationships that the investigator wants to study.

We develop a series of models appropriate for various situations. These models may be broadly classified as follows:

1. Models for two occasions:
   (a) involving one latent variable at each occasion, each latent variable possibly measured by multiple measurements,
   (b) involving several latent variables at each occasion, these latent variables being measured by many observed variables.

2. Models for two occasions, as above, but with measures of background and environmental characteristics also available:
   (a) background and environmental variables are all qualitative or discrete,
   (b) some or all background and environmental variables are quantitative and measured without error,
   (c) some or all background and environmental variables are quantitative and fallible.

3. Two-wave, two variables models with measurement errors correlated over time.

4. Multi-wave, two variables models with various assumptions about the correlations of measurement error over time.

5. Multi-wave, multi-variables models with complex structures between and within occasions.

Each model is presented and discussed by means of a path diagram in which observed variables are marked as squares and latent variables as circles and where the direct causal influence of one variable on another is denoted by an arrow.

As a unifying approach to the handling of all the models, we develop a general model of which all the above models are special cases. Within the framework of this general model we consider the statistical problems of

(α) identification of parameters
(β) estimation of parameters
(γ) test of goodness of fit of the model.

In particular we consider the problems of detecting lack of fit and testing for fit of various parts of the model.

Some of the models discussed are illustrated on the basis of some data from a large nationwide longitudinal study conducted at Educational Testing Service in 1961 - 1971. In this

study a number of achievement and aptitude tests were administered to a large sample of fifth grade school children in 1961 and then again in 1963, 1965, and 1967 when they were in seventh, nineth and eleventh grade, respectively.

REFERENCES

Anderson, S. B. & Maier, M. H.: 34,000 pupils and how they grow. Journal of Teacher Education, 1963, 14, 212-216.

Joreskog, K. G. & Sörbom, D.: Some regression effects useful in the measurement of change. Research Report 74. Department of Statistics, University of Uppsala, 1974.

Sörbom, D. : A general method for studying differences in factor means and factor structure between groups. British Journal of Mathematical and Statistical Psychology, 1974, 27, (in press).

The stability of results: some examples of the effect of scale transformations

Bernt Larsson, Department of Educational and Psychological Research, School of Education, Malmö, Sweden

Scales used in educational research are, as a rule, loosely defined. The most common numerical coding of the possible outcomes of a measurement is successive integers. However, there is seldom anything in the educational measurement procedure which prescribes this rather than any other coding. Educational researchers will in most cases not have any fundamental objection to exchanging this coding for a monotonic transformation of it.

On the other hand, many statistical methods (or other mathematical models) used in educational research, are only invariant up to, e.g., linear transformations. The question is: how stable are results, described by these methods, when monotonic transformations constitute the class of acceptable codings? High stability admits conclusions with great generality. It may also be of interest to choose that scale which, under given restrictions, maximizes (or minimizes) a certain index of result.

The techniques used for investigating the stability are based on a general principle. By using binary coding, each many-valued variable can be expressed as a weighted sum of its binary variables, where the weights are the scale values. This implies that almost all analysis will be multivariate, e.g., a certain type of analysis of variance is transferred to the corresponding discriminant analysis, modified due to the restriction of monotonic transformations. The restriction may cause complications: in the above example an iterative optimization procedure is substituted for a common eigenvalue problem where difficulties with starts, convergence and locally extreme points have to be considered.

This report gives some simple examples of the stability of results for some statistical methods. The findings are very different: from superstability (no transformation - monotonic or not - can change the result) to almost total instability, where almost any value in the possible range of a certain index can be obtained by different monotonic transformations. It can be added that the basic optimization subroutines are not yet finally tested: many more data and simulations are needed to do this.
The article by Larsson (1973) containing information about the technique when no restriction to monotonic transformations is made can be obtained from:

Mrs. Barbro Bergstrom,
Department of Educational and
Psychological Research School of Education,
Fack,
200 45 Malmö 23,
Sweden.

REFERENCES

Bradley, R. A., Katti, S. K. , & Coons, I.J.: Optimal scaling for ordered categories. Psychometrika. 1962, 27, 355-374.

Larsson, B. Obtaining maximal correlations by the construction of binary variables. Didakometry, 1973, 38.

Configural frequency analysis of repeated measurement designs: A multivariate distribution - free approach.

G. A. Lienert and J. Krauth, Institute of Psychology, University of Dusseldorf, F. R. Germany

In most cases the aim of designs with one repetition is to record changes of an observable X from a first observation to a second one. In this case it is recommended to consider the difference $D_X = X' - X$ between the second and first observation and treat it like an original observable. Here it is assumed that the observables are interval scaled, as otherwise the differences need not be defined. In case of an ordinal scale it is asked only whether there is a change in the positive or negative direction or no change at all. Then we have for each observable at most three directions of change (+, O,-).

In case of an afactorial design with only one observable X, which is measured for a sample of N patients with (X') and without (X) treatment, we form the difference D_x and partition it into x intervals. The expectations of the frequencies are assumed to equal N/x and it is analyzed which frequencies are substantially greater than this value. This method can be used in a modified form for observables of change (+,O,-),too.

For an afactorial design with t observables $X_1,\ldots,X_t$, with one repetition for each, we form the differences $D_{x_1} = X'_1,\ldots,$ $D_{x_t} = X'_t - X_t$. These differences are partitioned into $x_1,\ldots,$ x_t intervals. In case of ordinally scaled observables we consider the observables of change (+,O,-). This gives $x_1 \ldots x_t$ or 3^t configurations, respectively. The corresponding frequencies are compared with their expectations under the hypothesis of independent observables by the method of simultaneous binomial tests. As a result we may formulate interaction types.

In case of an n-factorial design with factors $F_1,\ldots, F_n$ having numbers of levels $f_1,\ldots,f_n$ which affect the t observables $X_1,\ldots,X_t$, we again discreticize differences $D_{x_1},\ldots,D_{x_t}$. The corresponding frequencies are entered into a contingency table with $x_1 \ldots x_t$ rows and $f_1 \ldots f_n$ columns and are compared with their expectations under the hypothesis of indenepdence of factors and differences of observables. Using simultaneous fourfold table tests we may get interaction types which describe an interaction between certain factor levels and changes of observables.

In case of designs with m repeated measurements, where m may be one or greater than one, we may easily categorize the time series for small m with respect to shape if the observable is at least ordinally scaled. For example, in the case of m=2 repetitions we may distinguish between increasing, decreasing, u-shaped and reversed u-shaped time series. For large m this method no longer can be applied and in the case of interval scaled observables, approximating each of the N time series by a polynomial with a degree as small as possible is recommended. The coefficients of these polynomials are treated as pseudo observables much in the same way as the differences in case of one repetition, i.e. they are discreticized and we consider the frequencies corresponding to the respective factor levels and combinations of observables. The evaluation is quite the same as in the case of one repetition considered above.

REFERENCES

Krauth, J. & Lienert, G. Die Konfigurationsfrequenzanalyse. Freiburg Brsg. : Alber, 1973.

Lienert, G. & Krauth, J. Die Konfigurationsfrequenzanalyse. IX. Auswertung multivariater klinischer Untersuchungspläne (Part 1; Part 2). Zeitschrift fur Klinische Psychologie und Psychotherapie, 1974, 22, 3-17; (in press).

A fuller report can be obtained from:

Prof. Dr. G. Lienert,
Faculty of Education
Institute of Psychology
University of Erlangen-Nurnberg
Nurnberg
Regensburger Str. 160
F.R. Germany

Type analyses with U.I. factor scores in clinical psychology: some data and some problems

Lothar R. Schmidt, University Mental Hospital,
Homburg/Saar, F.R. Germany.

The importance of objective tests (T-data) and factors based on them (U.I. factors) in clinical psychology has been demonstrated in a number of studies. Cattell, Schmidt & Bjerstedt (1972) have shown that differential diagnoses of six psychiatric groups and a control group with U.I. factors are far better than expected. However, this approach is based on psychiatric a priori diagnoses with partially low validity and missing functional psychological basis. Therefore, it seems necessary to search for new clinical types using statistical methods without an a priori classification.

In this study, Bolz (1972) type analysis based on Cattell's r_p-coefficients has been used in order to find new types in the clinical setting. The profile elements were the U.I. factors of Cattell, Schmidt & Bjerstedt (1972). In a first analysis 17 U.I. factors in a sample of 114 Ss (six clinical groups and a control group) yielded seven types. The concordance between these types and a priori diagnoses was rather low. Ss from the control group were over-represented in most types.

Since Cattell, et al., (1972) found that only 8 to 10 of the 17 U.I. factors in their study had clinical importance, in a second study the 8 clinically most relevant U.I. factors were used: 16, 19, 21, 23, 24, 25, 30 and 32. The concordance between empirical typing (6 types) and a priori diagnoses was higher but only marked for three types. The meaning of the types can be inferred by going back to the factor score profiles of the central (ideal) Ss of the types. Type 3 appears for U.I. 21+ (''Exurbance''), U.I.23+ (''Mobilization of Energy''), U.I. 16+ (''Assertive ego'') and to a lesser degree U.I. 32+ (Extraversion). For the central S of type 5, the signs of the factor scores were almost all in the opposite direction of type 3. It seems to be a minor clinical counterpart of type 3, characterizing depression, especially endogenous depression. Fin-

ally, type 6 can be labeled as a psychotic type. No neurotic group or control group belong to it. It is characterized by salient factor scores of U.I. 21- (''Suppressibility''), U.I. 32- (''Introversion''), U.I. 23- (''Regression''), and U.I. 30- (''Dissociated frustration'').

If one evaluates all types together they certainly have clinical and functional psychological meaning. However, if the aim is to get new diagnostic and possibly therapeutic insights we need to know more about the members of each type. This might be additional information from case history, behavior analyses, or longitudinal observations. Unfortunately, it was not possible to get this information in our study. The only diagnostic step which could be taken with the present data was an attempt to classify Ss according to the similarity with central subjects of different types, in order to find out their general clinical importance.

The distribution of the r_p- coefficients of clinical and control Ss with the central S of type 3 shows that 68% of the r_p of clinical Ss are minus in contrast to only 13% of the control Ss. No clinical S has a significantly positive r_p and no control S a significantly negative r_p with the central S of type 3. The classification of Ss according to the distribution of these r_p coefficients was surprisingly efficient. One cutoff exceeded the base rate (.90 for control and .10 for clinical Ss), other cutoffs seemed to be important for different clinical problems. Multiple cutoffs with regard to the r_p- coefficients with the central persons of types 3, 5, and 6 yielded an almost perfect classification. With one cutoff a total valid rate of 99% could be reached. Different classification rules for different problems have been worked out.

Empirical typing in the clinical setting with U.I. factor scores has thus shown its significance. However, a number of problems have to be solved. After cross validation studies, additional information about the persons belonging to different

types is needed. The typing depends on the similarity coefficient, the method of type analysis, the number and selection of profile elements (especially a broad sampling of the "whole personality sphere" vs. clinically most important variables), and subject samples.

REFERENCES

Bolz, C.R. Types of personality. In: R.M. Dreger (Ed.) Multivariate personality research: Contributions to the understanding of personality in honor of Raymond B. Cattell. Baton Rouge, La.: Claitor, 1972, 161-260.

Cattell, R.B., Schmidt, L. R. & Bjerstedt, A. Clinical diagnosis by the objective-analytic personality batteries. Journal of Clinical Psychology, 1972, Monogr. Suppl., 34, 1-78.

Detection of correlated errors in longitudinal data

Dag Sorbom, Institute of Statistics,
University of Uppsala, Sweden.

A common objective in longitudinal studies is to examine the change in some ability between two occasions. Between the two occasions, different groups of subjects have undergone different treatments. By studying differences in change, one wants to get an idea of the influence of the treatment on the ability. In general, it is not possible to measure the ability directly; rather we have to use a number of tests, which, from the experience of earlier studies or by some rational arguments, are believed to measure the ability in question. For a good measure of change it is desirable to have comparable units of measurement at the two occasions. This can be achieved, at least approximately, by using the same battery of tests or equivalent forms of the same tests at both occasions. However, this procedure implies that there will be residuals in the test scores remaining after elimination of the influence of the true score, which is correlated between the two occasions. That is, there is a correlation between the errors in the regression of the test scores on true score, so the classical factor analytic model is not applicable.

There are other circumstances which may lead to correlated errors. In the paper, a study in which verbal ability is measured by three tests is described. Two of the tests are constructed to measure what we in a common sense way are inclined to call verbal ability. The third test, on the other hand, measures "vocabulary". No doubt, in the latter test we are measuring verbal ability too, but if we in some way could eliminate the influence of the true ability from the test score, there should be some residual left. Thus the true ability accounts for only a part of the correlation between the two occasions for the third test. This may cause our estimates of the measure of change to be distorted, and we should obtain better estimates if we take the correlation of the errors into consideration.

In other situations there may be correlations between the errors within occasions. Suppose, for example, we are measuring an ability by three tests and that two of these are very much alike. In this case some part of the correlation between the tests could originate from the mere construction of the tests and not from the common true score.

In the paper a computer program is described, which produces maximum likelihood estimates for the parameters in a factor analytic model in which the errors or unique variables may be correlated.

The main focus of the paper is on a procedure for detecting correlations between errors, since even if the number of observed variables is moderate the number of possible models is so large that a search strategy is needed. This problem has recently been considered by Costner and Schonberg (1973). The procedure proposed in the present paper is based on a simplified version of the Lagrangian multiplier test (see e.g. Silvey, 1958) and it is shown how it works in two simple cases.

REFERENCES

Costner, H.L. & Schonberg, R. Diagnosing indicator ills in multiple indicator models. In A.S. Goldgerger (Ed.) Structural equation models in the social sciences. New York: Seminar Press, 1973.

Silvey, S.D. The Lagrangian multiplier test. The Annals of Mathematical Statistics, 1958, 30, 389-407.

A fuller report can be obtained from:

Dr. Dag Sörbom,
Department of Statistics,
University of Uppsala,
P. O. Box 2300,
S-75002 Uppsala 2,
Sweden.

Inverted alphabet printing: reminiscence as a function of interpolated rest and extraversion.

P. O. White, Institute of Psychiatry,
University of London, England

Some 5,000 11-12 year old children performed an inverted alphabet printing test for 17 1-minute trials. Subjects are grouped in terms of the amount of rest interpolated between trials 12 and 13. Rest intervals utilized were 0 min. (no rest), 1 min., 5 min., 10 min., and 60 min. Thus the observations fall naturally into a 5 x 2 layout (Rest x Sex).

Pre-rest performance is analysed by Generalized Manova using orthogonal polynomial contrasts on the within subject model. While boys score consistently higher than girls, differences among rest groups both in terms of performance level and in terms of trends across trials are consistent across sexes.

A reminiscence score (T_{13}-T_{12}) relates very strongly to amount of interpolated rest but only very slightly to extraversion.

Principal component analyses of the variance-covariance matrices indicate a component common to all trials (individual differences in speed); a component which contrasts pre-rest performance with post-rest performance; and a component which contrasts early trials with later trials in both pre-rest and post-rest performance blocks.
These relationships are discussed in terms of a Hull-Kimble-Eysenck inhibition formulation.