University of Windsor

Scholarship at UWindsor

Electronic Theses and Dissertations | Theses, Dissertations, and Major Papers

1-1-1974

Issues involved in the empirical testing of a multivariate predictive system.

Robert T. Forsyth
University of Windsor

Follow this and additional works at: https://scholar.uwindsor.ca/etd

Recommended Citation

Forsyth, Robert T., "Issues involved in the empirical testing of a multivariate predictive system." (1974). *Electronic Theses and Dissertations*. 6715.
https://scholar.uwindsor.ca/etd/6715

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email (scholarship@uwindsor.ca) or by telephone at 519-253-3000ext. 3208.

NOTE TO USERS

This reproduction is the best copy available.

UMI®

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ISSUES INVOLVED IN THE EMPIRICAL TESTING OF A MULTIVARIATE PREDICTIVE SYSTEM

by
ROBERT T. FORSYTH
B.A., Brock University, 1973

A Thesis
Submitted to the Faculty of Graduate Studies through the
Department of Psychology in Partial Fulfillment
of the Requirements for the Degree of
Master of Arts at the
University of Windsor

Windsor, Ontario, Canada
1974

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

UMI Number: EC53126

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI®

UMI Microform EC53126
Copyright 2009 by ProQuest LLC.
All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 E. Eisenhower Parkway
PO Box 1346
Ann Arbor, MI 48106-1346

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A6A9053

©Robert S. Forsyth 1975

556541

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPROVED BY:

Dr. M. Morf Marvin Morf

Dr. H. Minton H. Minton

Dr. R. Daly R. Daly

Dr. S. Sadava

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ABSTRACT

This study utilized previously collected data from two samples of 187 and 184 college students to investigate problems inherent in multivariate data analysis as well as patterns of cannabis use. The multivariate issues considered were: the effects of the distributions of the scores of the variables; the use of factor scores as well as raw data in regression analysis to facilitate "conceptual cross-validation"; effects of sequential orthogonalization on regression equations; and the utility of criteria exhibited by factor analyses and canonical correlations. It was found that in the field of marijuana research, the usual criterion - frequency of use, was not suitable both for measuremeant and inferential reasons. Using canonical analysis two patterns of use were found -- moderate and abnormal, which were related to social and personality predictors respectively.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ACKNOWLEDGEMENTS

I acknowledge the help of the members of my committee: M. Morf, H. Minton, R. Daly, and S. Sadava. Appreciation is also extended to Marybeth Heilbronn and Larry M. Starr.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE OF CONTENTS

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF TABLES

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

List of Figures

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter I

Introduction

One of the major aims, if not the only aim of psychology, is useful and accurate prediction of human behaviour. The approaches used vary with the particular behaviours studied, ranging from the study of minute reflexes, to overall life patterns, carrying with them constructs from conditioned stimuli to self-concept. For behaviour of medium complexity, the social learning approach, developed by Rotter and his colleagues (eg. Rotter, 1954; Rotter, Chance & Phares, 1972) or a similar approach may be most useful, provided that constructs are made sufficiently clear. It is the intent of this thesis to operationalize some of these constructs, to investigate methods of testing a predictive system, and to consider specific issues that pertain to this system and generally to complex multivariate data sets.

In this study the behaviour of interest is marijuana use, specifically in college students. Studies on the drug have proliferated, perhaps in keeping with greater public interest in its use and abuse in the last few years. Although much of the literature deals with the physiological and psycho-pharmacological properties of cannabis, the psychological viewpoint demands an exploration of social and personality correlates and the consequent predictions of who uses the drug, and with what effects. Recent studies by Jessor (Jessor, Jessor & Finney; 1973) and Sadava (1974b) have had some success in this task, by viewing marijuana use as a functional behaviour, "caused" by variables both of an interpersonal (environmental) and of an intrapersonal (personality) nature, and as an ongoing process

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

changing over time. This conception of the problem may be traced back to Rotter's social learning framework, and a short description of this approach (based mainly on Applications of a social learning theory of personality, chapter 1, 1972) would make the later work more meaningful.

Rotter's Social Learning Theory

In general, Rotter's Social Learning Theory utilizes the ideas of expectancy and the empirical law of effect, as underlying constructs emphasizing the interaction of the individual and his meaningful environment. Thus, the approach is much more cognitive in nature than most other types of learning theory. Specific efforts are directed to determining the subject's perceptions of his goals and his subjective expectancy that they will be fulfilled according to his own past experiences. Needs and goals are social in nature since they are initially fulfilled by others. Behaviour becomes available if it has led to reinforcement, either directly, or through observation and modelling. Generalization takes place, with functionally related behavious/reinforcers leading to the same goals/satisfactions. Cognitions are more important in this theory than others since the generalization may be of symbols and their referents. The general formula used to predict behaviour is: $BP=f(E \& RV)$ where BP is the probability of the behaviour, relative to other behaviours in the specific (psychological or perceived) situations; E is the expectancy or subjective probability of a particular reinforcer occurring as a result of a certain behaviour; and RV, the reinforcement value, is the relative preference for a given reward with expectancies for all rewards kept constant.

So, for instance, the potential for marijuana smoking to occur, in

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

college, in relation to peer approval, is a function of the expectancy of the occurance of peer approval, following marijuana smoking in college and the value of peer approval. This is a specific example, and may be expanded by including other reinforcements, both positive and negative, other expectancies etc.

Jessor's Approach

Jessor has taken Rotter's system, developed in a clinical setting, and applied it to studies on alcohol use (eg. Jessor, Carnan, & Grossman, 1968; Jessor & Finney, 1973). Marijuana use is thought of as "problem behaviour... considered to be purposeful, goal oriented or functional (pl, 1973)". Variables have been classified into structures and systems by Jessor: a personality system consisting of motivational instigation, belief and personal control structures; a perceived environment system with these and other behaviours ranging from closely related to "normal". His methodology consists of administering questionaires at two or more points in time, using multiple-regression techniques, and gain scores (over time), on social and personal variables to increase "accounted-for" variance. He was able to assign users and non-users to their respective groups with 73% accuracy and to achieve multiple R's of up to .39 (Jessor, et al, 1973).

Sadava's Approach

Sadava's approach has focused in on marijuana use specifically. One major problem in marijuana research is that while ultimate clinical interest may lie in determining what leads to abuse, research has been defined in terms of use versus non-use. Even this distinction is made in different ways by various researchers. For instance the category of "light use" ranges from one to twenty experiences, and there are at least twenty-eight

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

labels for use patterns (Sadava, 1974a). To cut down this confusion, Sadava designed a short group of questions, with four stages of use (Sadava, 1972) with the additional stages of non-user and "Have stopped" added (Sadava 1974b). For the non-user and the four user stages this scale has Guttman-scale type properties. Continuing to refine criteria, Sadava uses the following measures; frequency of use, time span as a user, contexts of use and adverse consequences of use in a sample comprised of drug users. These measures have average intercorrelations of .23 with a maximum of .54 (Sadava, 1974b), indicating relationships amongst criteria, and the possibility of patterns of drug use. The complex relationships between behaviour and predictors may be obscured by a poor choice of criterion. Meaningful correlate may be hidden from our investigation if the criterion is poorly defined, or if its scale properties are ignored.

Sadava's schema of variable systems may be seen in Figure 1. The major difference between the systems of Sadava and Jessor is that in the former cognitive functions are conceptualized as a set of predictors of drug use (by Sadava), and as being predicted by other systems of variables. Using his system Sadava (1974b) has found multiple R's of up to .62 for variables predicting criteria measured at the same time, and multiple R's up to .58 for longitudinal analyses with samples of drug users. Both Jessor and Sadava use longitudinal analyses, and often utilize both the scores on variables at the earlier point in time, and change scores in these variables, to predict behaviour at the second point in time.

Social Learning Theory: Applied to Complex Operations

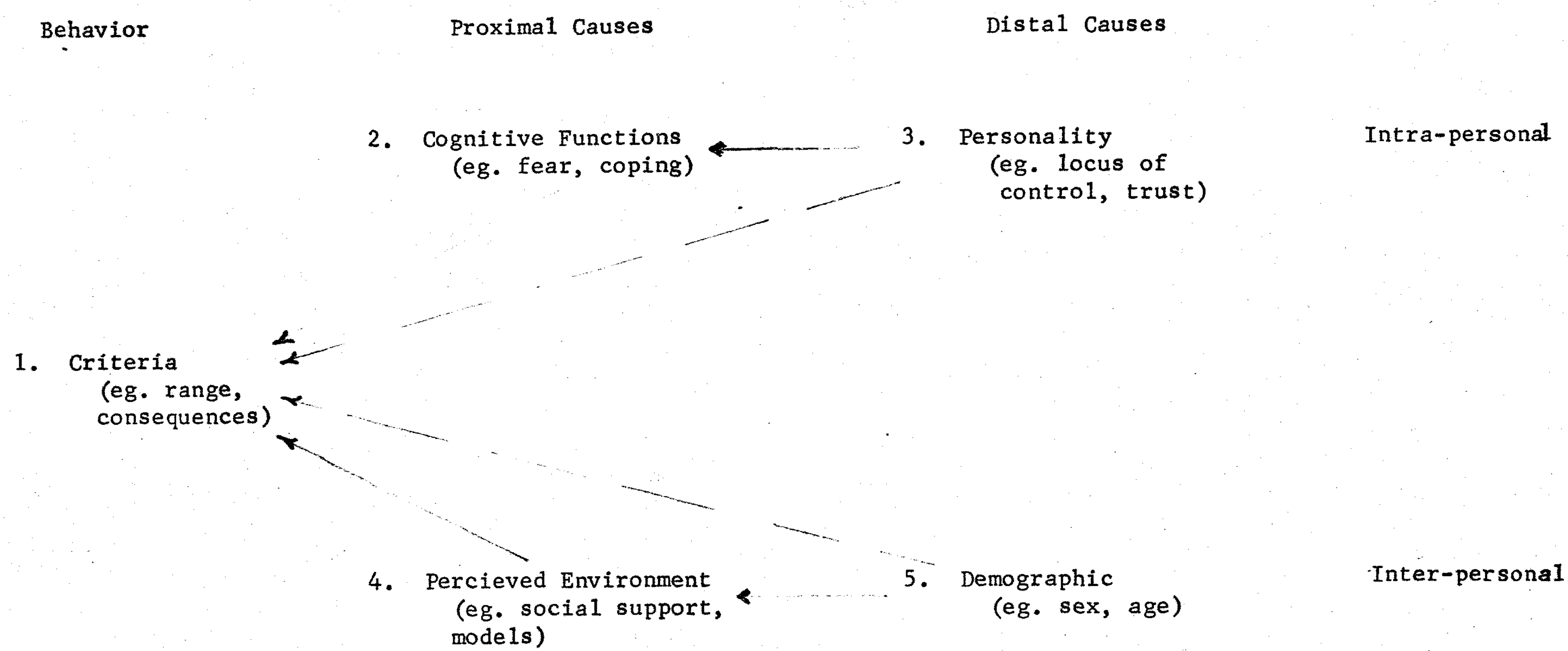
In the analyses of both these researchers, the variables have been analysed bit by bit, using sub groups of predictor variables to predict

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

FIGURE 1

Simplified Non-recursive

Social Learning Model



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

each criterion, then another sub group to predict each criterion and so on. Perhaps this is due to a lack of a statistical framework, or of a notational system that would allow easy manipulation of the large number of variables and extensive data. Thus the first stage of this thesis is the expression of Sadava's and Jessor's systems in a notation that allows manipulations of variables in order to deduce and to subject to verification, hypotheses about relationships between predictors and criteria. This is a necessary step since some of the terms used (eg. 'related structures', 'covary with other kinds of problem behaviours') and also relationships between variables (eg. BP=F(E & RV)) are hard to operationalize. Furthermore, although the underlying principles assume an integrated predictive system, the tendency is to examine concepts and variables in isolation, and with less than complete thoroughness.

Notation

Let B_{qtn} be a criterion observation [section 1 of Figure 1] for individual n (n=1,...,N) at time t (t=1,...,T) on criterion q (q=1,...,Q) which after appropriate transformations(s) is in Z-score form (Exp $(B_{qtn})=0$, $Var(B_{qtn})=1$). Similarly, let C_{1tn}(l=1,...,L) be a normalized variable for cognitive functions [section 2 of Figure 1]; P_{mtn}(m=1,...,M) be a normalized variable for personality characteristics [section 3 Figure 1]; E_{ktn}(k=1,...,K) be a normalized variable for perceived environmental data [section 4 Figure 1]; and D_{stn}(s=1,...,S) be a normalized variable for demographic data [section 5 Figure 1].

Let $\Delta(B_{qtn})$ be a normalized variable corresponding to the unpredicted part (i.e., change over time-delta) of B_{qtn} from B_{q1n}. Similarly, $\Delta(C_{1tn})$, $\Delta(P_{mtn})$, $\Delta(E_{ktn})$ and $\Delta(D_{stn})$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Let $F_{rn}(B)$ be the factor score for the r'th factor $(r=1,\ldots,R)$. for individual n, taken from the $B_{qtn}(D)$.

Let $\Gamma^{P}_{mtn}(E+C)$ denote a normalized variable consisting of that covariance-free part (gamma) of P_{mtn} remaining after the variance due to the E and C variable systems is partialled out, and so on for combinations of P, E, C and D.

Let $R_{Bq.ECP}$ be the multiple correlation between the qth variable of B and the variables of E, C and P.

Thus using the above notation for measurement times 1 and 2, it is hypothesized that an equation of the following general form leads to better estimations of criteria at time 2 and higher multiple R's than obtained previously:

$$\hat{B}_{q2n} = \sum_{s=1}^{S} r_{qs} F_{sn}\left(D_{s1n}\right) + \sum_{k=1}^{K} r_{qk} F_{kn}\left(\Gamma^{E}_{k1n}(D) + \Gamma^{\Delta E}_{k2n}(D)\right) + \sum_{m=1}^{M} r_{qm} F_{mn}\left(\Gamma^{P}_{m1n}(D+E) + \Gamma^{\Delta P}_{m2n}(D+E)\right) + \sum_{l=1}^{L} r_{ql} F_{lm}\left(\Gamma^{C}_{l1n}(D+E+P) + \Gamma^{C}_{l2n}(D+E+P)\right)$$

$$R_{B_q \cdot ECPD} = R_{\hat{B}_{q2n} B_{q2n}} .$$

The score for the n'th person, on the q'th criterion, at time 2, may be estimated from a linear composite comprised of the sum of the variables in each system weighted such that the predictors are orthogonal. Thus in this case, the estimated criterion score is the sum of the factor scores for demographic data weighted by first order correlation coefficients with

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the criterion, plus the sum of the personality factor scores, free of covariance with demographic data, and weighted by the correlations with the criterion, plus the percieved environment factor scores similarly partialled (of both demographic and personality scores) and weighted, plus cognitive function factor scores treated in the same manner.

Problems with this Predictive System

In the present study several important limitations exist and should be noted beforehand. Firstly, the model calls for all variables to be normally distributed, zero mean, and unit variance. This condition may not be met because of several considerations in the scales involved; measures may be nominal or at best polychotomous with very discrete values found for what may or may not be underlying continuous variables; some variables may be extremely skewed, and even with transformations may not really resemble a normal distribution.

A second deviation from the model relates to the fact that the various systems of variables should include all non redundant measures in the domain of interest. This condition is not fulfilled as: (a) those measures may not at this point in time exist and (b) only a limited subset (hopefully representative) can be administered in the time permitted.

Another problem exists in that the model is linear in nature whereas either the real or theoretical variables may combine multiplicatively in the form of moderators to other variables, or in other non-linear combinations (eg. curvilinear). This must be kept as a consideration in the research, but as the non-linear models tend to be limited operationally more or less to small numbers of variables, the present work will (hopefully) serve as a good first approximation.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Other problems exist in the nature of the instruments used and the type of behaviour studied. In the latter case, marijuana use is, of course, illegal and therefore responses to questions about its use can be seen as incriminating. King (1972) reports that anonymous and identifiable questionnaires on this subject do not significantly differ on reported data, however this does not mean there will be no difference between actual behaviour and reported behaviour due to some fear of exposure. With respect to questionnaire studies in general, we must consider the possibilities of response styles and sets (eg. Jackson, 1967), and also false content through misconceptions on the part of the subject. The latter is more easily dealt with, in that according to the assumptions of the Rotter, Jessor and Sadava models behaviour is determined by subjective expectancies, values, etc. not merely objective facts, and distortion of this kind do not render the information invalid.

To compare the possible distortion of responses on the questionnaire, it is best to consider that the data have two different logical sequences. The first sequence is that of cognitive functions followed by perceived environment, personality and demographic data, ordered in terms of decreasing proximity to the behaviour concerned. The subjects' own perceptions will have the most weight and be least subject to cumulative response biases.

The second sequence will be that of demographic data, personality, perceived environment, and cognitive function, in decreasing order of objective verifiability, as opposed to proximal subjectivity.

Specific Areas of Investigation

There are four specific issues that must be examined in dealing with

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the model. These are: distributional properties of predictor and criterion variables; the use of raw data vs factor scores; the use of two sequences of orthoganalization, ie. starting with cognitive functions, and starting with demographic variables; and the patterns of behaviour criteria.

Distributions of the Variables

It is extremely important that we examine the distributions of scores on the variables we use before interpreting results from statistical analyses, since we may be misled by our results if assumptions necessary to the analysis are not met, or if more useful approaches to data analysis are ignored due to a lack of information. This warning takes on more importance when the variables haven't been used extensively in other research, and/or several "non-robust" (requiring stringent assumptions for use and interpretation) statistics such as stepwise regression or canonical correlation are to be used. For instance, we might have a distribution with a few extreme values at each end, being correlated with another variable that also has extreme values. If the extreme values "match", there will be a very high correlation that is not reflected at all by the other variable pairs. Another possibility is that the two variables have a parabolic relationship, or that one variable is related to the logarithm of the other. In both cases, the correlations between the variables would be very low, and the true relationships found only by a theoretical or, more likely, an empirical investigation of the univariate and bivariate frequency distributions of the variables (see Carroll, 1967). As the first part of the exploration of this model, these distributions will be examined.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Raw Data vs Factor Scores

Since the scales for this investigation were chosen on the basis of availability, or constructed to measure a specific area, there may be both a large overlap in the domain of measurement, and a great deal of variance "superfluous" to the model. If the variables are of this nature, the use of a multiple regression technique may lead to different weights for the variable in each sample, even though the particular underlying factors don't change from sample to sample. Cooley and Lohnes (1971) point out that "chance" is an important consideration in multivariate analysis, and since we may have several variables contributing almost the same variance, chance could determine which variable is chosen (the covariance with similar variables being partialled out and consequently not appearing as significant). This process may have more importance than seems readily apparent since we will be influenced in our model building and generalization by the labels or names attached to the "significant" variables, and may draw the wrong inference about the meaning of a variable's loading on the regression equation.

One solution to this problem is to assure ourselves, prior to the dependence analysis (eg. multiple regression), that we have predictors, whose domains, empirically, do not overlap greatly. This may be achieved by orthogonalization of our predictors. We may also collapse the number of variables into a smaller number of factors. Using this method, we may substantiate the findings of analyses of raw data, and more easily infer from the measured variable the underlying trait or cause.

Sequence of Data Entry

When orthogonalizing the data sequentially, the first variables may

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

retain greater meaningful variance than later variables because of the method itself. The last variables will, most likely, have little or no correlation with the criterion, whereas the corresponding original variable may, in reality, be highly related to the criteria. Ideally every variable should have equal chance to covary with criteria, but this leads to an enormous number of analyses to include each sequential combination. Given these circumstances, together with the theoretical reasons discussed previously for the different types of response sets and kinds of verifiability two sequences of orthogonal predictors will be used: Functions- in which the function variables are entered first (the sequence of proximity); and Demo.- in which the demographic variables are entered first (the objective verifiability order). These sets will be examined to determine what differences exist due to ordering effects.

Criteria Patterns

Considering Sadava's findings (1974b) concerning the relationships between criteria, and the different predictors useful in each case, it is important to look at the criteria carefully, to examine specifically their inter-relationships (eg. by factor analysis), and to find out whether different sets of predictors predict different patterns of criteria (using canonical correlation). Using this information we can evaluate the criteria as to their reliability and usefulness. In addition, a methodology becomes available to ascertain predictive systems for genuine, multidimensiona patterns of use.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter II

Method

Subjects

The questionnaires were filled out by first and second year students enrolled full-time at Brock University, St. Catharines, Ont. in the 1972-73 academic year. Four hundred and eighty Ss completed questionaires in Nov. 1972; and of these 371 filled out the questionnaire in Mar. 1973. The ages of the Ss ranged from 16 to 74 but most were under 26. Fifty-five per cent of those responding twice were female.

The group (371) was split into two samples randomly, Sample 1 was 187 Ss and Sample 2 was 184. Scores were standardized within samples, but all other estimated parameters (e.g. correlations, factor loadings, etc.) were obtained from Sample 1 and applied to Sample 2.

Questionnaire

Table 1 lists the scales used, number of items, format and source, if other than Sadava (1972, 1973, 1974b). This table is organized by variable systems, i.e. behaviour, cognitive functions, perceived environment, personality, and demographic data, and excludes some scales not used in the present study.

Major Statistics

Univariate and Bivariate Distributions

Descriptive statistics and frequency distributions were found for each scale, using data before standardization. Scattergrams were constructed for ten predictor variables (standardized) with each of the four criterion variables, to give an indication of the types of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE 1

Variables Used

System/ Variable	Number of Items	Item Format	Source**
Criteria System	....	....	......
Consequences	17	3 point scale	
Range	19	yes/no	
Time Span	1	"how long ago"	
Frequency	1	# of occasions	
Cognitive Functions	....	....	......
Total Positive	14	3 point scale	
Instruemental	6	3 point scale	
Coping	6	3 point scale	
Total Negative	10	3 point scale	
Ideological	4	3 point scale	
Fear	6	3 point scale	
Percieved Environment	....	....	......
Social Support	10	4 point scale	
Social Sanct'n	6	3 point scale	
Availability	2	5 point scale	
Parental Model	16	4 point scale	
Sibling Model	5	4 point scale	
Personality	....	....	......
Peer Confor'y	10	3 point scale	Lin(1972)
Peer Indep.	10	3 point scale	Lin(1972)
Peer Anticon'y	10	3 point scale	Lin(1972)
Family Con'y	10	3 point scale	Lin(1972)
Family Indep.	10	3 point scale	Lin(1972)
Family Anti.	10	3 point scale	Lin(1972)
Tol. Drug Use*	3	10 point scale	
Att. To Devian	15	10 point scale	
Total Risk	8	9 point scale	Jackson(1972)
Physical Risk	2	9 point scale	Jackson(1972)
Financial Risk	2	9 point scale	Jackson(1972)
Social Risk	2	9 point scale	Jackson(1972)
Ethical Risk	2	9 point scale	Jackson(1972)

continued

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2 continued

System/ Variable	Number of Items	Item Format	Source**
IE Locus of Con	23	Forced Choice	Rotter(1966)
IE personal	9	Forced Choice	Mirels(196)
IE political	4	Forced Choice	Mirels(196)
Trust	25	5 point scale	Rotter(1966)
Personal Trust	6	5 point scale	
Political Trust	8	5 point scale	
Interper'l Al'n	10	5 point scale	Keniston(unp)
Social Alien'n	5	5 point scale	Keniston(unp)
Moral Judge.*	35	mixed	
Religiousity*	3	mixed	
Delay Grat'n	4	5 point scale	Stumphauser(1972)
Time Pers've	5	number of months	Shybut(1968)
Demographic	....	....	......
Sex *	1		
Age *	1		
Social Econ.*	1		
Yr in College*	1		
Poli. Orien'n*	1		
Residence*	1		
Reference Gr.*	1		
Height *	1		
Weight *	1		
Self Des. Obs.*	1		
Exp. GPA *	1		
GPA Difference*	1		

*Not repeated Spring, 1973
** If other than Sadava

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

bivariate distributions present in the data.

Orthogonalization of Data

Delta and gamma scores. In order to arrive at a set of mutually independent predictors several types of transformation are needed. First, we must get an estimate of change-over-time scores. Given two sets of scores on the same variable, from the same subjects, we can partial out of the later score, the variance accounted for by the score taken at a point earlier in time (e.g. Jessor et al, 1973). Thus, starting with two related scores, we obtain a base score, and an independent estimate of change, a delta (Δ) score. Delta is estimated by Ferguson's formula $\Delta = (Z_1 - r_{12}Z_2)/\sqrt{1-r_{12}^2}$ (1971, p. 387) where Z_1 is the score at the first point in time, Z_2 is the same subjects' score at second point in time, and r12 is the correlation between Z_1 and Z_2 over all subjects.

A generalization of the above formula estimates a score that is independent of more than one variable (this may be done across time, as for , or with other scores from a larger test battery). If we wish an estimate of this covariance-free score, a gamma (Γ) score, we may use the following formula: $\Gamma_{1.23\ldots n} = (Z_1 - \sum_{m=2}^{n} r_{1m}Z_m)/\sqrt{1-\sum r_{1m}^2}$ [see appendix A for proof], where Z_1 is our variable of interest, $Z_2,\ldots,Z_n$ are the covariates, r_{1m} is the correlation between the Z_1 scores for $Z_m(m=2,\ldots,n)$ over all subjects. These scores will have $E(\Gamma)=0, VAR(\Gamma)=1$ and will tend to a normal distribution.

Principal components. The third type of transformation needed is a principal components factor analysis, producing appropriate

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

factor scores. Since we are interested in linear composites with the characteristics of mutual independence, interpretability and completeness, the proper procedures for this factor analysis are: the use of unities in the main diagonal; an adaptation of the scree test (see Cattell, 1952, for a discussion of this criterion) to determine the number of factors retained; and a varimax rotation of those factors.

We must use unities in the diagonal instead of estimates of communality since our data are not being used to infer underlying structure, but rather, are being described by the factor scores. The question of establishing the number of variables to be rotated is more complex. It may be resolved by considering the nature of the factor structure and the communality of a variable. We may regard communality as the square of a variable's multiple correlation with all the extracted factors. Thus, if each variable has a high communality after a certain number of factors has been extracted, we may assume that each of the common and unique factors has been obtained. For example, if after five factors are extracted, the communality of a variable is .78, there remains to be accounted for, only 22% of the original variable's variance in all the remaining factors.

We can always (with unities in the diagonal) account for all of the variance by taking as many factors as variables, but we most likely will have "error" factors, due to faulty measurement, discrete scale values etcetera. A procedure that tends to eliminate these factors is Cattell's scree test (Cattell, 1952). A sharp drop in eigenvalues after relatively high eigenvalues means the following eigenvalues may be neglected. The original argument by Cattell

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

recommended both a scree test, and a minium eigenvalue over 1.0, however, since we are using principal components, the later requirement may be dropped, and the requirement of high communalities added.

The varimax method of rotation is used so that each factor can be most easily identified as related to a few (and if possible only one) variables. Varimax simplifies factors, as opposed to variables (Mulaik, p.259, 1972), and this means that the "output" of our transformation will tend to be more easily interpretable than if we used either a variable-simplifying or both factor and variable simplifying rotation.

Stepwise Multiple Regression

This technique (SMR) is really a combination of regression and factor analysis (Mulaik, p.412, 1972). The first step is to correlate the criterion with the criterion as the variable in the regression equation. The factor collinear with this variable is extracted from the predictor matrix, and a residual correlation matrix is obtained. Then the variables (or factors since this factor analysis method-that of Cholasky [Mulaik p.412] maintains a correspondence between factors and variables),remaining are correlated with the predictor, the corresponding factor extracted, leaving a residual matrix, and so on. At each step a variable is added to (in occasional circumstances removed from) the regression equation, and the multiple correlation is increased.

This process of adding (and/or removing) variables can continue until all the variables are in the equation, however, as with other factor analytic procedures, later factors (i.e. variables in Cholesky method) may be unique or error terms. Since we are dealing also with

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

regression, the unique factors that do not correlate with the criterion, may be, for practical purposes, considered error. Either error may increase the multiple correlation, but to such a small extent that, given the probability of measurement error, the predictive utility of the equation is diminished. Cooley and Lohnes (p.56-57, 1971) point out that SMR can capitalize on chance to a large extent, and care must be taken to replicate this procedure on another sample(s). The replicated R may shrink appreciably if too many variables are allowed in the original equation. For our purposes, we may use an $\underline{F}$ ratio (or an equivalent $\underline{t}$ test) to determine whether a variable should be added to the equation. This $\underline{F}$ ratio is computed as the square of the ratio of regression coefficient and its standard error (SPSS, 1970).

Orthogonal Predictor Variables

If the correlation matrix of predictor variables is an identity matrix (i.e. the variables are correlated), the regression equation weights may be obtained directly from simple correlations with the criterion (Mulaik, p.404, 1972). Furthermore, the square of the multiple correlation is the sum of the squares of the simple correlation with the criterion. If we have orthogonal predictor variables, the order of entry of these variables into a stepwise MR is the same as the rank order of the absolute values of the correlation coefficients, and the SMR is not necessary. In practice, a SMR can be useful in these circumstances since the correlation with real data probably will not be an exact identity matrix. Even if our regression weights are from a non-orthogonal group of predictors, Cooley and Lohnes (p.56, 1971) point out that sample predictor-criterion correlation

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

coefficients of the first order may yield better predictive utility than regression weights from more "sophisticated" processes.

Canonical Correlation

Stepwise Multiple Regression may be considered a special case of canonical correlation. The problem of canonical correlation is to find relationships between a set of predictor variables and a set of criterion variables (in SMR we have a set of one criterion variable). In general, we ask if there is a combination of predictors X (a pattern) that has a high correlation with a combination or pattern of criterion Y variables. To do this we find a set of weights for the predictors such that the composite variable (W_1) is maximally correlated to a composite variable Y_1 made up of a weighted combination of the Y variables. A factor corresponding to W_1 is extracted from X which leaves the residuals X_r unrelated to W_1, in the same manner as the residual matrix is found in SMR. The composite V_2 is treated in the same way to produce a residual matrix Y_r. This process is repeated producing (usually) m set of weights, where m is the number of variables in the smaller of the predictor and criteria groups (see Van de Geer, 1971). We must then evaluate the sets of weights called canonical variates to find which are significant. The canonical correlation coefficient R_c can easily be misinterpreted. It is not the correlation or overlap between X and Y but between the linear composites W and V. To evaluate the "overlap" or shared variance of the two sets, we may use a statistic R_d, a redundancy coefficient discussed by Stewart and Love (1968). For each canonical correlation there are two redundancy coefficients, one for the X variables given the composite V from the Y variables (i.e. R_{dx}) and one for the Y

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

variables given the composite W from the X variables (i.e.R_{dy}). R_{dx} is calculated as the proportion of variance extracted by the factor (W_n) times the proportion of shared variance (R_{cn}) between the factor and corresponding canonical factor of the other battery (Cooley and Lohnes, p. 170, 1972). Thus we square the weights of the X variables, divide by the number of X variables, and multiply by the canonical correl ation

The canonical correlation analysis can be useful in exploring criterion patterns however it should be used in conjunction with other measures such as the multiple correlation of each criterion (Cooley and Lohnes, p.176, 1972), chiefly because of the complexity of the procedure, and the possible misinterpretation of results.

<u>Summary of Programmes and Formulae</u>

1. Univariate Distributions - SPSS (1970), subroutine CODEBOOK

2. Bivariate Distributions - BASIS (1971), subroutine Plot.

3. Delta Score - $_{Z1Z2} = (Z_1 - r_{12}Z_2)/ \; 1 - r_{12}^{\;2}$. A Fortran programme was written and an example given in Appendix B.

4. Gamma Scores: $= (Z_1 - \; (r_{1m}Z_m)) \; / \; 1 - r_{1m}^{\;2}$. This was done by a Fortran programme in Appendix B.

5. Principal Components Analysis: Both SPSS (1970), subroutine FACTOR, and SSP factor analysis (1970: p.429) were used, with several sets of data run on both to ensure accuracy (identical results to fifth signific- ant digit).

7. Correlation Coefficients. SPSS (1970), subroutine PEARSON CORR was used and compared to SSP subroutine CORRE.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

8. Stepwise Mutiple Regression: SPSS (1970), subroutine REGRESSION, and SSP programme for stepwise multiple regression (p. 419) were used and compared.

9. Canonical Correlation: SPSS (1970), subroutine CAN CORR was used, there were no suitable programmes for comparison.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter III

Results

Univariate Distributions

The raw data may be grouped in sets of variables with common forms of frequency distribution. These distributions seem to approximate: the normal, truncated normal, superimposed two population normal, the Poisson, the rectangular and the dichotomous or binomial distributions. Table 2 gives examples of variables from each group, descriptive statistics and type of distribution.

Bivariate Distributions and Correlations

Selected pairs of standardized variables were plotted to check for usual bivariate distributions. None of the distributions seemed curvilinear (e.g. U shaped). The scattergrams including Range seemed to indicate that extreme values (for Range) were depressing the correlations, and suggested that correlation with Log (Range) might be considerably higher. Scattergrams are included as Figure 2.

Correlations matrices between the criteria and each of the predictor variables may be found in Appendix C. There is a matrix for each of the Raw, Functions and Demo. modes of analysis for Sample 1 and for Sample 2.

Transforms

The resulting variables, processed in either direction, i.e. partialling out covariance of functions from social, of functions and social from personality etc; and partialling out covariance of demographic from personality variables etc., were relatively easily identified, having

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE

Univariate Distributions

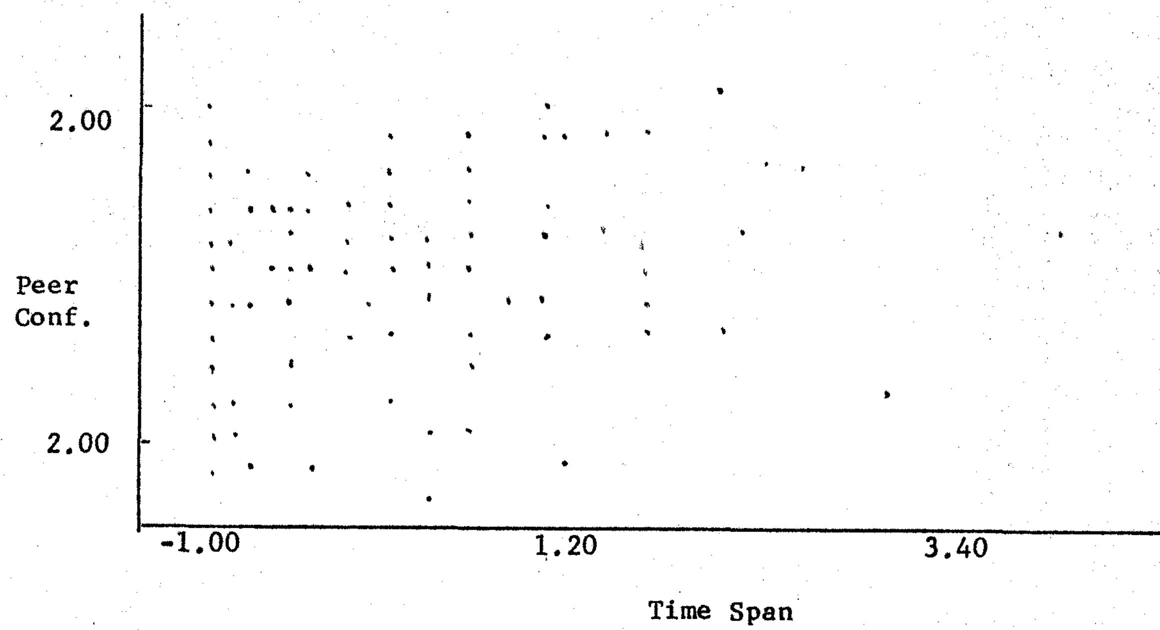
Variable**	Central Moments *				Range	Quartiles	Type of Distribution
	Mean	Variance	Skewness	Kurtosis			
Frequency	17.976	6095.527	10.940	133.106	999	0/0/10	Poisson/ Dichotomous
Time Span	19.505	648.855	1.365	1.860	144	0/4/36	Poisson
Range	3.585	13.238	2.725	12.793	33	1/2/5	Poisson
Consequen.	31.118	116.599	0.400	-1.095	34	22/30/41	Rectangular
Social Supp	19.865	36.436	0.406	-0.572	28	15/20/24	Bimodal Nor.
Avail.	8.435	2.436	-1.382	2.536	8	8/9/10	Trunc. Nor.
Sib. Model	7.557	5.433	1.675	4.733	15	6/7/9	Trunc. Nor.
Total Pos.	22.862	27.451	0.491	0.156	28	19/22/26	Trunc. Nor.
Ideo. Neg.	6.470	3.432	0.799	0.159	8	5/6/8	Trunc. Nor.
Fam. Conf.	22.849	6.501	0.053	0.079	15	21/23/25	Normal
Att.To Dev.	36.891	81.661	0.069	0.472	58	31/37/43	Normal
Total Risk	38.192	89.906	0.052	0.010	54	32/38/44	Normal
I.E.	10.593	24.697	-0.014	-0.585	22	7/11/14	Normal
Exp. GPA	73.843	42.218	0.511	1.014	41	70/75/80	Disc. Nor.
S.D.Obes.	3.114	0.530	-0.047	0.370	4	3/3/4	Normal
Sex	1.445	0.248	0.220	-1.952	1	1/1/2	Dichotomous

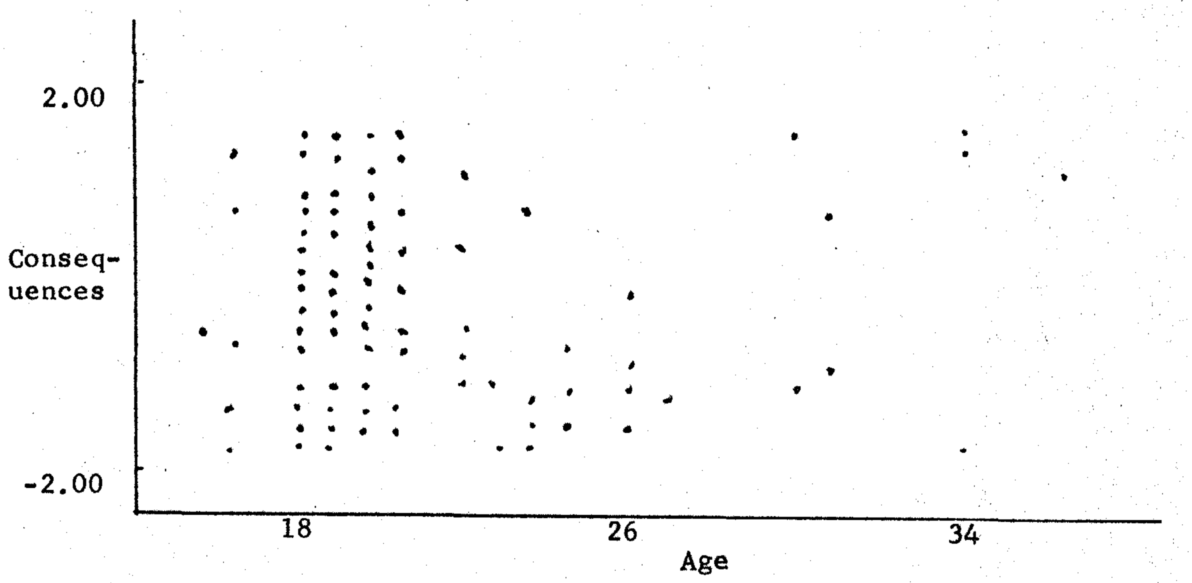
* Based on 371 observations

** Spring variables except Att. to Dev.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

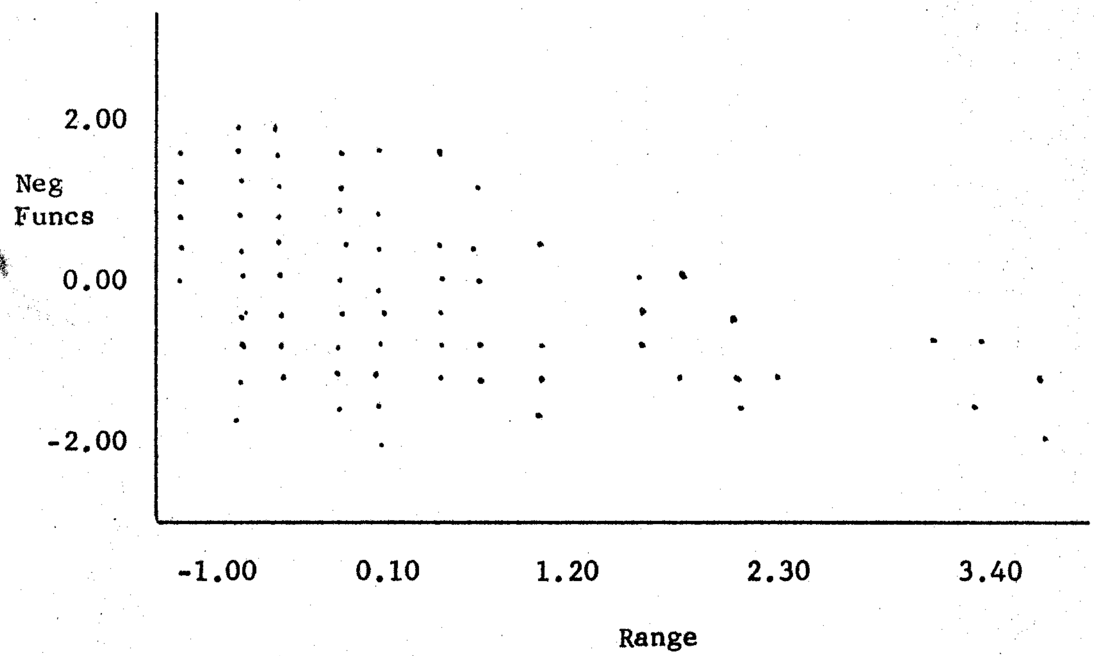
FIGURE 2

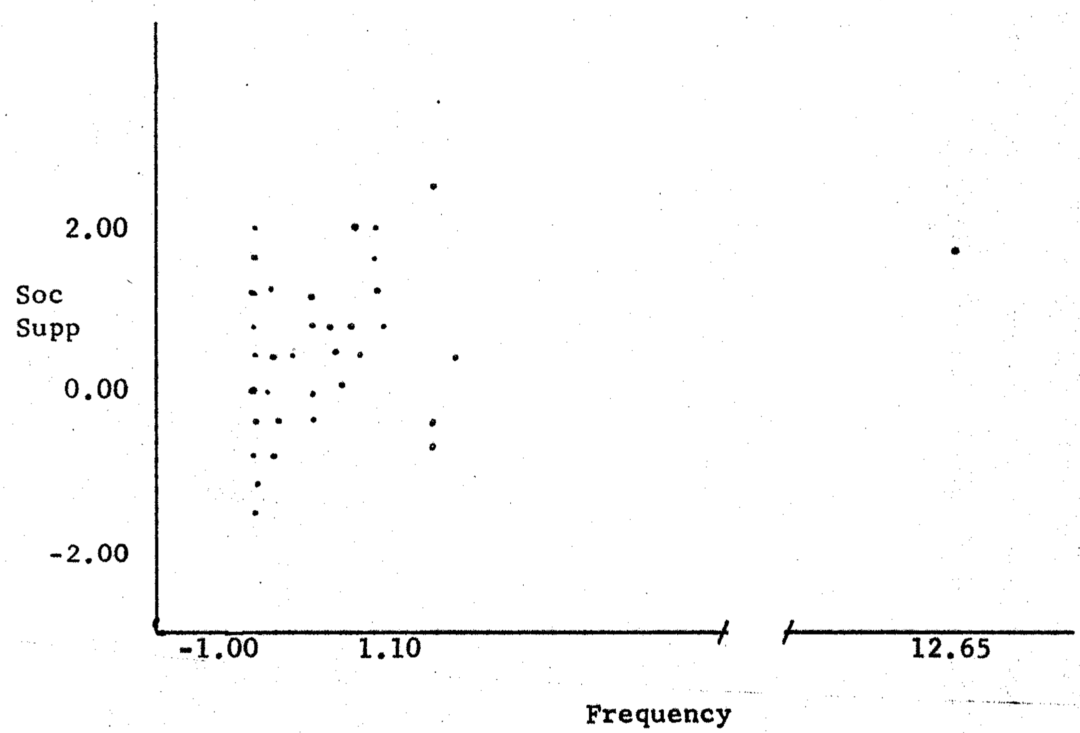




continued...

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.





Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

high loadings on principal components analyses. Table 3 includes rotated factor loadings, communalities, and lists new variable names for the two modes of factoring.

Factor Analysis of Behaviour

In both Samples 1 and 2, frequency is unrelated to the other criteria, with time and range loading together. Table 4 presents rotated loadings and eigenvalues .

Stepwise Multiple Regression Analysis

A total of 24 SMR's were calculated. Tables 5 through 10 present the variables with significant Beta weights, the cumulative variance in thousandths and the final multiple correlation coefficient. Table 11 presents a 2 x 3 x 4 breakdown of these Multiple correlations, and average correlations (using Fisher's Z transformation) for each classification.

Canonical Correlations

Five canonical correlations were performed, results for canonical correlation with raw data were not obtained due to the size of the r matrix and the programmes available. One of the canonical correlations was meaningless due to a singular matrix resulting in a canonical correlation greater than 1.0. Table 12 gives the significant canonical correlations, loadings variables with loadings for each of the four meaningful analyses.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE 3

Functions and Demo. Factors

Original Variable		Functions Factors				Demo. Factors			
Name	#	Name	#	Loading	Comm.	Name	#	Loading	Comm.
Tot Pos Func	5	Pos Func	7	.99	.99	Pos Func	47	.99	.99
Inst Pos Func	6	Pos Func	7	.89	.80	Pos Func	47	.88	.78
Cope Pos Func	7	Pos Func	7	.87	.80	Pos Func	47	.88	.78
Tot neg Func	8	Neg Func	5	.99	.99	Neg Func	44	.99	.99
Ideo Neg Func	9	Neg Func	5	.89	.77	Neg Func	44	.85	.74
Fear Neg Func	10	Neg Func	5	.94	.77	Neg Func	44	.91	.84
ΔTot Pos	11	ΔPos	8	.99	.99	ΔPos	45	.99	.99
ΔInst Pos	12	ΔPos	8	.87	.80	ΔPos	45	.86	.76
ΔCope Pos	13	ΔPos	8	.86	.88	ΔPos	45	.87	.78
ΔTot Neg	14	ΔNeg	6	.98	.99	ΔNeg	46	.98	.99
ΔIdeo Neg	15	ΔNeg	6	.86	.76	ΔNeg	46	.83	.70
ΔFear Neg	16	ΔNeg	6	.89	.83	ΔNeg	46	.86	.81
Soc supp	17	Soc Sanc	9	.58	.83	Soc Clim	37	.84	.78
Soc Sanc	18	Soc Sanc	9	.94	.92	Soc Clim	37	.82	.70
Avail'y	19	Avail	15	.95	.95	ΔSoc Supp	38	.64	.74
Par Model	20	Par Mod	13	.95	.92	Par Mod	41	.95	.93
Sib Model	21	Sib Mod	10	.81	.87	Sib Mod	39	.78	.81
ΔSoc Supp	22	ΔSoc Supp	14	.96	.96	ΔSoc Supp	38	.85	.84
ΔSoc Sanc	23	ΔSoc Sanc	12	.98	.98	ΔSoc Sanc	40	.97	.95
ΔAvail'y	24	ΔAvail	16	.99	.99	ΔAvail	42	.97	.96
ΔPar Mod	25	ΔPar Mod	11	.97	.97	ΔPar Mod	43	.97	.99
ΔSib Mod	26	Sib Mod	10	-.77	.86	Sib Mod	39	-.74	.79

Continued...

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Cont'd

Name	#	Name	#	Loading	Comm.	Name	#	Loading	Comm.
Peer Conf	27					Peer Conf	27	.95	.94
Peer Indep	28	Indep	22	.87	.88				
Peer Anti	29	Anticonf	21	.85	.80				
Fam Conf	30	Fam Conf	30	.93	.92				
Fam Indep	31	Indep	22	.87	.87	Fam Indep	29	.98	.99
Fam Anti	32	Anticonf	21	.87	.86	Anticonf	21	.97	.98
Att Tow Dev	33								
Tol Dr Use	34	Tol Dr Use	27	.96	.95				
Tot Risk	35	Tot Risk	20	.90	.87				
Phys Risk	36	Tot Risk	20	.91	.87				
Fin Risk	37								
Soc Risk	38					Soc Risk	20	.98	.99
Eth Risk	39								
Tot I-E	40	I-E	17	.94	.95				
I-E Pers	41	I-E	17	.85	.89				
I-E Poli	42	I-E	17	.70	.82	I-E Poli	36	.93	.96
Tot Tr	43	Poli Tr	26	.91	.90				
Pers Tr	44					Pers Tr	22	.90	.91
Poli Tr	45	Poli Tr	26	.90	.89	Poli Tr	31	.97	.99
IP Alien	46								
Soc Alien	47					Soc Alien	35	.96	.98
Moral Judg	48								
Relig	49					Relig	33	.96	.98
Del Grat'n	50	Del Grat	24	.93	.91	Del Grat	32	.99	.99
Time Pers	51					Time Pers	23	.99	.99

Continued...

Cont'd

Name	#	Name	#	Loading	Comm.	Name	#	Loading	Comm.
ΔPeer Conf	52	ΔConf	23	.83	.82				
ΔPeer Indep	53	ΔIndep	28	.68	.84				
ΔPeer Anti	54					ΔFam Indep	24	.51	.89
ΔFam Conf	55	ΔConf	23	.76	.82	ΔConf	17	.94	.93
ΔFam Indep	56	ΔIndep	28	.92	.90	ΔFam Indep	24	.90	.91
ΔFam Anti	57	ΔNon-conf	33	.87	.88				
ΔAtt Tow Dev	58	ΔAtt Tow Dev	35	.98	.99	ΔAtt Tow Dev	16	.94	.95
ΔTot Risk	59								
ΔPhys Risk	60	ΔPhys Risk	29	.96	.95				
ΔFin Risk	61	ΔFin Risk	32	.94	.94				
ΔSoc Risk	62					ΔSoc Risk	28	.96	.98
ΔEth Risk	63	ΔSoc Risk	34	.86	.88	ΔEth Risk	34	.94	.96
ΔTotal I-E	64	ΔI-E Pers	19	.93	.91				
ΔI-E Pers	65	ΔI-E Pers	19	.94	.92	ΔI-E	15	.65	.91
ΔI-E Poli	66					ΔI-E	15	.89	.93
ΔTotal Tr	67	ΔPoli Tr	18	.92	.91				
ΔTr Pers	68					ΔPers Tr	26	.95	.96
ΔTr Poli	69	ΔPoli Tr	18	.92	.90	ΔPoli Tr	30	.98	.99
ΔIP Alien	70	ΔAlien	25	.74	.76				
ΔSoc Alien	71	ΔAlien	25	.87	.83	ΔSoc Alien	25	.98	.98
ΔDel Grat	72					ΔDel Grat	19	.97	.97
ΔTime Pers	73	ΔTime Pers	31	.97	.96	ΔTime Pers	18	.96	.97

Continued...

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Cont'd

Name	#	Name	#	Loading	Comm.	Name	#	Loading	Comm.
Sex	74	Size	36	.87	.82	Size	5	.87	.82
Age	75	Age	42	.99	.99	Age	10	.99	.99
Soc Ec St	76	SES	44	.99	.99	SES	11	.99	.99
Yr in Coll	77	Yr in Coll	43	.99	.99	Yr in Coll	7	.98	.99
Poli Orien	78	Poli Orien	41	.98	.99	Poli Aff	14	.99	.99
Residence	79	Res	39	.99	.99	Res	9	.99	.99
Ref Group	80	Ref Gr	40	.99	.99	Ref Gr	12	.99	.99
Height	81	Size	36	.81	.91	Size	5	.92	.88
Weight	82	Size	36	.91	.86	Size	5	.88	.93
Self D Obs	83	SDO	38	.98	.96	SDO	6	.98	.97
Exp GPA	84	Exp GPA	45	.98	.99	Exp GPA	8	.98	.99
Grade Diff	85	Gr Diff	37	.97	.99	Gr Diff	13	.97	.99

Note : All loadings after varimax rotation and based on Sample 1

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE 4

Factor Loadings of Criteria

(Varimax Rotation)

Sample	Factor	Criterion				Initial Eigenvalue
		Conseq.	Range	Time	Freq.	
One	1	-.214	.398	.900	.090	2.284
	2	-.016	.293	.101	.975	0.968
	3	.966	-.192	-.258	-.014	0.522
	4	-.147	.848	.336	.202	0.226
Two	1	-.106	.936	.347	.109	2.076
	2	-.060	.124	.130	.987	0.870
	3	.974	-.113	-.232	-.058	0.727
	4	-.189	.309	.900	.105	0.327
One *	1	-.165	.915	.391	not included	2.089
	2	.964	-.180	-.252		0.643
	3	.209	-.360	-.885		0.268

* Freq. excluded from analysis

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE 5

Stepwise Multiple Regression

Raw Data Sample 1

	Criterion Variable			
	Frequency	Time Span	Range	Consequences
Predictor Variable and Cumulative Variance ** (in thousandths)	ΔTotal Risk(110) Del Grat'n(133)* Pers Trust(152)* Poli Trust(181)	Social Supp(450) Tol Dr Use(486)* Age (508) ΔPoli Trust(528)* Pr Anti (545) ΔTime Pers(561) ΔSoc Al'n(575) ΔPers Trust(587)* ΔNeg Cope (596)* ΔSocial Supp(606)	Social Supp(449) Tol Dr Use(491)* ΔTime Pers(528) ΔFin Risk(549) Social Sanc(562) ΔTrust (578)* Pr Anti(589)* ΔAtt Tow Dev(598)* Sib Model (610)	Social Supp(247)* ΔFear (334) Fear (406) ΔPers trust(457) ΔSocial Sanc(477)* Avail (494)* ΔPhys Risk(509) ΔFam Anti (523)* ΔTime Pers(532) ΔNeg Cope (543) ΔFam Indep(554)
Multiple Correlation	.426	.778	.781	.774

*Variables with negative loadings
**All variables load significantly ($p<.05$)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE 6

Stepwise Multiple Regression

Raw Data Sample 2

	Criterion Variable			
	Freq.	Time Span	Range	Conseq.
Predictor Variable and Cumulative Variance** (in Thousandths)	Tol.Dr.Use (133)*	Social Supp (458)	Social Supp (357)	Social Supp (259)*
	Grades (132)*	ΔSocial Supp (512)	Tol.Dr.Use (407)	ΔSocial Sanc. (307)*
	ΔDel. Grat. (177)*	Age (538)	ΔIP.Al'n (440)*	Soc. Sanc. (339)*
	Time Persp. (209)	Sex (562)	Sex (458)	ΔInst. Fn (374)
	Att. Tow. Dev (231)	Tol. Dr. Use (583)	Att. Tow. Dev (473)	Fin Rk (393)*
	ΔEth. Rk (250)	ΔSib. Mod. (599)	ΔPeer Conf. (484)	Yr. in coll. (411)*
	Weight (272)	ΔPar. Mod. (609)	ΔSoc. Supp. (494)	ΔAvail. (427)
			ΔFin Rk. (506)	Self Des. Obs (442)
				ΔEth. Rk (455)
Multiple Correlation	.521	.780	.711	.674

* Variables with negative loadings
** All variables load significantly

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE 8

Stepwise Multiple Regression

Functions Sample 2

	Criterion Variable			
	Frequency	Time Span	Range	Consequences
Predictor	Neg Func (022)*	Neg Func (177)*	Soc Sanc (120)	Neg Func (139)
Variable**	ΔNeg Func (046)*	Soc Sanc (293)	Neg Func (230)*	ΔSoc Sanc (207)*
in order of	Avail (061)	Avail (386)	Avail (315)	Soc Sanc (292)*
entry with		ΔNeg Func (440)*	ΔAlien'n (365)*	Avail (331)*
Cumulative		Sib Mod (463)	ΔNeg Func (398)*	ΔPos Func (360)
Variance		ΔIE (494)*	ΔNon-Conf (412)*	ΔNeg Func (388)
		Poli Orien (518)	ΔSoc Supp (427)	Del Grat (408)*
		ΔSoc Sanc (544)		Poli Orien (425)*
		Age (557)		
		ΔEth Rk (567)		
		IE (577)*		
		ΔSoc Supp (591)		
Multiple Correlation	.246	.769	.653	.652

** All variables load significantly ($p < .05$)
* Variables with negative Loading

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE 9

Stepwise Multiple Regression

Demo Sample 1

	Criterion			
	Frequency	Time Span	Range	Consequences
Predictor Variable** in Order of entry with Cumulative Variance	ΔDel Grat (046)*	Soc Clim (139)	Soc Clim (163)	Soc Clim (119 *
	ΔAtt Tow Dev (077)*	Size (259)	Size (269)	Size (190)*
	Self D.Ob (106)*	ΔPoli Tr (324)*	ΔEth Rk (316)	Poli Tr (261)
	ΔTime Persp (128)*	Poli Aff'n (387)	ΔDel Grat (354)*	ΔPoli Aff'n (289)*
	Time Persp (153)*	Del Grat (425)	Grade Diff (383)	Neg Func (317)
	Pol Tri (178)*	Grade Diff (453)	Self D Ob (410)*	Age (339)
	Soc Clim (200)	ΔSoc Sanc (480)	Poli Aff'n (432)	Residence (361)*
		Residence (498)	Sib Mod (452)	Grade Diff (382)*
		Soc Alien (511)*	ΔPoli Tr (470)*	ΔPos Func (402)
		Sib Mod (524)	Poli Tr (488)*	ΔAtt Tow Dev (418)
		Age (537)	ΔSoc Supp (502)	Relig'y (444)*
		Peer Confor (550)*	ΔConform (515)	ΔSoc supp (460)*
			Soc Al'n (527)*	ΔSoc Sanc (473)*
				Yr in Coll (486)*
Multiple Correlation	.448	.741	.726	.697

**All Variables load significantly ($p < .05$)
*Variables with negative loading

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE 10

Stepwise Multiple Regression

Demo. Sample 2

	Criterion Variable			
	Frequency	Time Span	Range	Consequences
Predictor Variable** in order of entry with Cumulative Variance	ΔTime Persp (067)*	Soc Clim (145)	Soc Clim (144)	Poli Aff'n (121)*
	Soc Clim (094)	Poli Aff'n (304)	Pol Aff'n (253)	Soc Clim (170)*
		Size (381)	Size (295)	Residence (209)*
		ΔSoc Alien (407)	ΔDel Grat (326)*	Poli Tr (250)
		Del Grat (434)	ΔPoli Tr (355)*	ΔI E (275)*
		ΔSoc Supp (461)	Relig'y (379)	Size (309)*
		Yr in Coll (482)	Ref Grp (395)	ΔSoc Sanc (355)*
		Sib Mod (505)	ΔEth Rk (409)	Relig (357)*
		ΔI.E (518)		Time Persp ((380)*
				Neg Func (401)
				Soc Al'n (421)*
				Yr in Coll (437)*
				Self D Ob (451)
Multiple Correlation	.307	.720	.640	.672

** All variables load significantly ($p < .05$)
 * Variables with negative loading

TABLE 11

Summary of Multiple Correlations

		Mode of analysis			
Sample	Criterion	Raw	Functions	Demo.	Total **
One	Conseq.	.744 (11)*	.673 (8)	.697 (14)	.705
	Range	.781 (9)	.743 (10)	.726 (13)	.750
	Time Span	.778 (10)	.771 (11)	.741 (12)	.765
	Freq.	.426 (4)	.272 (2)	.448 (7)	.384
Two	Conseq.	.674 (9)	.652 (8)	.672 (13)	.666
	Range	.711 (8)	.653 (7)	.640 (8)	.669
	Time Span	.780 (7)	.769 (12)	.720 (9)	.758
	Freq.	.521 (7)	.214 (2)	.307 (2)	.354
Total **	Conseq.	.708	.664	.713	.686
	Range	.747	.701	.684	.712
	Time Span	.779	.770	.758	.761
	Freq.	.457	.244	.380	.370

* Averages over modes/samples (using Fisher's Z)
** Indicates number of variables with significant loadings

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE 12

Canonical Correlations

	Mode of Analysis							
	Functions 1		Demo 1		Demo 2	Demo 1		
	1	2	1	2	1	1	2	3
Canonical** Correlation	.876	.591	.860	.656	.855	.854	.641	.546
Coefficient of Redundancy (Rdx)	.088	.171	.094	.197	.096	.101	.237	.151
Conseq. Range Time Span Frequency	-0.382 0.364 0.419 0.080	1.010 0.895 0.138 -0.366	-0.560 0.282 0.290 0.173	0.965 0.403 0.777 -0.360	-0.475 0.281 0.475 0.042	-0.522 0.398 0.292 not	1.024 0.213 0.802 included	0.082 1.378 1.266
Predictors	Neg Func ΔNeg * Soc Sanc Avail	ΔNeg Func ΔTime Per	Size ΔSoc Rk* ΔPol Tr* Soc Clim	Age ΔA.T.D. ΔTimePer* Anticon* ΔSoc Al* Soc Rk Fam Ind Pol Tr* ΔEth Rk* IE Pol* Relig* Soc Clim	Size Pol Or Pol Tr* Soc Clim	Size Pol Or Soc Rk* ΔPol Tr* Soc Clim	Age ΔTimePer* ΔDel Gr* Anticon* Time Per* ΔSoc Al* Soc Rk Fam Ind Pol Tr* Del Gr Relig* ΔEth Rk* IE Pol	ΔDel Gr* Soc Rk ΔPer Tr Fam Ind Pol Tr* Del Gr* ΔEth Rk

*Variables with negative loadings ** p<.05

Chapter IV

Discussion

Several importamt issues must be considered in testing a predictive system. The first point is the investigation of the nature of the distributions of scores for the variables, the effects of these distributions on results, and the possibilities of reformulating measures or transforming distributions to eliminate undesirable characteristics. Then, the effect of error components, and covariance of chosen predictors is examined by means of raw score-factor score comparisons with reference to stability of results and interpretation of regression weights. The importance of different sequences of data entry should be considered. And finally, consideration will be given to patterns existing within the criteria, and in the relationships between predictors and criteria patterns. Discussion of each of these issues follows, concluding with a general overview and discussion of implications for multivariate research.

Distributional Properties of the Variables

The universal and bivariate frequencies of the data seem to deviate from normal, and bivariate normal distributions. Descriptive statistics for 15 of the variables appeared in Table 2. This sample represents the kinds of distributions found, and includes each of the variable systems. An examination of these fifteen varialbes represents a consideration of the problems in the complete set of 85 variables. First the variables will be considered individually, with the empirical distribution arising from both scaling properties, and the properties

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of the underlying "real" distributions.

Criteria

Frequency. This one item scale requires the subject to state the number of times he (she) has used marijuana in the last six months. The answers range from '0' (marijuana not used) to '999', and the distribution has very high skewness (a measure of symmetry, high positive meaning a long tail to the right, high negative a long tail to the left), and a very high kurtosis (a positve kurtosis means the distribution is more peaked than the normal curve). An examination of the histogram for frequency shows that approximately 55% of the respondents gave an answer of '0'. Three possibilities exist as to the nature of the distribution: a dichotomy with unequal intervals; a poisson distribution; or a log-normal distribution. The log-normal is properly a distribution for a continuous variable, but may be approximated for a discrete variable. It is a distribution in which the logarithm of the variable is distributed normally, and results from either a particular process occurring to the "true score" or as a result of error variance being related to the value of the true score (the error variance of a normal variable is unrelated to the true score). The distribution may not take on the true value of '0' (log (0) is undefined), and since the probability of '0' is so high, the addition of a constant in this case may be more distortion than transformation.

A poisson distribution occurs when we have "rare occurrences events in a fixed time interval (Tsokos, p.113, 1972)". Since it seems strange that an event that has low probability could occur 100, 200 or 999 times, as with frequency we must consider what these numbers mean.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

First of all over the six month period, the respondent answering '100' has smoked every second day, the respondant answering '400' has smoked twice a day, etc. Also we must consider the implications of the number given as to probable error involved. Most likely those responding '0' will give an exact (true) report, but those responding with '5' may be including the interval 3,7 , and with '200' may be including 180, 220 etc. This is similar to psycho-physical problems (e.g. JND's) where the error is proportional to the magnitude of the true score. The differences between answers of '1' and '100' and between '101' and '200' cannot be considered as equal intervals. Perhaps the most important interval is that between '0' and '1'. In other research (see Sadava, 1974b), the investigators often treat frequency as a dichotomous criterion 0 and 1 , all those answering '1' or more being grouped together and predictors analyzed by point-biserial correlation or analysis of variance. Upon reanalyzin the present data, classifying frequency as 0 or 1, the multiple correlation with the predictors rose to .767 (as opposed to a high of .521 with raw data, for sample 1.; where a continuous distribution was assumed).

One approach to the frequency criterion would be to first analyze assuming a dichotomy and then reanalyze assuming a continuous distribution. An investigation of the portion of the curve including all those answering '1' or more may allow us to transform answers by taking log(x) or to rescale by intervals and a poisson approximation, resulting in either case in a more normal distribution.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Time Span. Although this criterion does not have the extremities of Frequency (Pr['0']<.5 lower kurtosis and skewness), it is by no means normal. In theory, this measure, i.e. how many persons started smoking marijuana in a given month, should take on a poisson distribution, however in practice the person's score is complicated by errors that most likely are proportional to the 'true' answer. The most direct way of transforming this variable would be to add a very small constant (e.g. .001) to avoid zeros and then take log(x), this at least would reduce skewedness.

Range. This variable is scored by adding up the number of drugs taken by the individual. The responses are not independent, as in a personality scale, but rather strongly interrelated. In a personality scale, each item is assumed to have constant probability of being endorsed, due to the underlying trait measured. With the Range scale, it is likely that endorsement of one drug in a group (e.g. opium, heroin, morphine) implies a large probability of endorsing the other drugs outside the group (e.g. tranquilizers). Instead of 19 separate questions, we may be asking about 5 or 6 categories of drugs, and assigning different weights to these categories rather arbitrarily (as above-a score of 3 for opium derivatives vs a score of 1 for tranquilizers). It we refine the scale to give equal weights to drug categories, we would most likely find a different distribution. This new variable might have the properties of a Guttman scale (if the theories about "progression to harder drugs" were true), or that of a poisson distribution. In either case, the addition of a small constant, and the use of a log(x) transform would probably make the Range variable approximate a normal distribution.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Consequences. This variable has a much different form than those above, being relatively symetric (skewness = 0.400) and flat rather than peaked (the kurtosis is negative). In general however this distribution (labelled rectangular) is a close enough approximation of the Gaussian (or normal) to be used in most statistics. In Carroll's (1961) review of the effects of marginal distributions on correlations, the quality that most often gives rise to problems is that of skewness combined with peakness. The rectangular (as opposed to tapering tails as in the normal) nature of this distribution may be due to extreme scores of a response set. There are 17 three point items all keyed, and worded, in the same direction. The subjects may be responding habitually to a particular key ('often' or 'never') and more discrimination may be afforded by a reversal of keying or wording, or an extension from three point items to five or seven (see below for comments on the personality scales).

Perceived Environment

Social support. This scale more closely approximates normal than any of the preceding, but appears to have two modes. This may be a result of chance, or perhaps there are two reponse patterns-that of users and that of non-users. A discriminant analysis would provide more evidence on this point. For our purpose this distribution may be considered normal.

Availability. This scale and those of sibling model, total positive functions and ideological negative functions all appear to be normal, with one tail truncated. The loss of symmetry caused by

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

this truncation is a more serious problem than the apparent bi-modality of social support, but the trunctation can be corrected. The score distribution of Availability is negatively skewed and probably results from the common milieu of the subjects. In a less resrticted sample, we most likely would find more subjects at the lower level of the scale. A second approach to approximating nomality could be the extension of this scale from the 3 items to 10 or 15, adding questions of a higher "difficulty" level (e.g. have you had a pond of marijuana (or more) in your possession at any given time).

<u>Sibling model</u>. As with Availability more items (and more difficult items) should be added.

<u>Cognitive Functions</u>

<u>Total positive functions and ideological negative functions</u>. These two scales are similar in construction except that the former has 14 items and the latter 6 items. The Ideological scale has a higher skewness (.799 vs .491) and this may be a direct result of there being a fewer items. This comparison suggests to us that the number (and difficulty level) be increased, at least for the ideological scale.

<u>Personality Scales</u>

<u>Value family conformity, attitude towards deviance, total risk and locus of control (I-E)</u>. Of all the scales used in this study the personality scales (not just the four cited), in general, may be most closely described as normal. This would seem to result from both the theoretical underlying distributions involved (as opposed to time span for example), and the scaling strategies involved (as opposed to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Idêological negative functions). The personality traits are usually assumed to be normally distributed in the general population, and the underlying traits are continuous (not dichotomous or discrete). This follows from a psychological assumption that there are many contributing causes for the trait, and the mathematical principle that the sum of a number of independent, small, random variables has a distribution that is approximately Gaussian (normal). This latter principle is used again in the scale construction, by utilizing either a large number of questions, or a wide range of valuse for each question (or both). The four scales in Table 2 all follow this strategy: Value Family Conformity has 10 items, each a three point scale; Attitude Towards Deviance had 15 ten point items; Total Risk has 8 nine point items; and Locus of Control (IE) has 23 two point items.

Demographic Variables

Expected grade point average. This scale has normal properties with a minor perperbation. Numbers ending in '0' or '5' (e.g. 65, 80, 85) seem to include most of the marks otherwise found in the surrounding interval. We may assume that '65' implies the range from '63' to '67'. (The practice may result from prior experiences with rounded marks). In our interpretation of regression weights etc., connected with Expected GPA, we may have to consider this, but there should be no major problem.

Self description: obesity. Although only a single five point item, this scale appears to be normal. This may result from connection to physical characteristics (weight, height) which are normally distributed, or may be due to chance.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Sex. This variable, at least for the purpose of this study has a dichotomous distribution. We must be careful when interpreting the meaning of any regression weights or other statistics. Furthermore, the combination of this scale with another whose distribution is markedly non-normal may result in strange values.

Bivariate Distributions

Although only a few scattergrams were produced (out of thousands possible), some of the results were supportive of the consideration above. Range consistently showed that a transformation could have increased correlation. The plot of consequences and age showed that several extremely high values of age combined with high values for consequences to change the correlation. Because of the vast majority of the subjects were 26 or under, the older subjects may have had disproportionate effects on findings. (This may also have occured with Frequency, Time and Range.) The use of scattergrams should be extended (both in this study and mutlivariate analyses in general), and used to substantiate hypotheses about univariate distribution.

Raw Data Versus Factor Scores

The issue we are concerned with is cross-validation, both in the classic statistical sense, across samples, and in a more abstract conceptual sense, i.e. what the label of a reliable variable means. First of all, if the stepwise mutliple regressions performed on frequency are excluded (as previously mentioned, its underlying scale may be unsuitable for this analysis, and as will be pointed out below, it is unrelated to other criteria), the range of multiple R's over two three criteria and three modes of analysis, is remarkably small-the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

lowest is .640 and the highest .781 (variance-accounted for is between 40% and 60%).

The parameters for Sample 2 were calculated on Sample 1, yet the differences between R's is not great (see Table 11). Without statistical tests (for which we assume, rather tenuously, a multivariate normal distribution), it is safe to say that we have not captilized on chance (we have used large samples and demonstrated a reasonable replication, lowest variance accounted for is per cent) even though the correlation for Sample 1 are slight;y higher than those of Sample 2. That is, multivariate cross validations (Cooley and Lohnes, 1971), have been demonstrated.

The multiple correlations for the raw data are higher than for either of the factor data sets, however the difference is not great (the largest difference is .070), and the functions (and demo) have slightly greater stability in individual variable entry into the regression equation. On the average, three variables are replicated from sample one to two, but with raw data, but 5.3 variables are replicated with the factor scores. This is to be expected since there are about half as many with Functions and Demo as there are with the raw data, and these variables are "boiled down" through a process that tends to eliminate unique or error variance.

Both the raw score and the factor score approaches have advantages, but the greatest advantage is gained in using both of them. The analyses for consequences should serve as an example. If we look for raw score variables that are significantly loaded for both Samples 1 and 2,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

we find only social support and Δ Social sanctions. Social support appears in the Functions and Demo. analyses (as the social sanctions and social climate factors respectively). Social Sanctions also appear on all of the analyses. We have gained more confidence in these variables since they might have been eliminated by the errors of proximity and verifiability discussed earlier. Availability appears in in Raw data, Sample 1, both Function analyses, and as social climate in Demo analyses; social sanctions appear in five of the analyses. Now it is clear that the variable complex: social support, social sanctions, availability and social support is of great importance to the consequences criterion. The labels may merge in the factor score analyses, but it is clear that this complex represents a major correlate. Further investigation would seem worthwhile. The variables negative Functions and Δ Negative Functions (or their subscale-fear negative function) appear four times each in the six analyses and similarly indicate that this complex of variables is of importance. In contrast, Δ personal trust is the fourth variable to enter the first raw data sample equation, but

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

is not replicated in raw or factor score equations. It would appear that this may have been as a result of covariance with those variables, rather than being a significant predictor.

The above discussion is brief, and could be continued for the other criteria, but should be sufficient to show the importance of a simultaneous raw score/factor score approach.

Sequence of Data Entry

It is now appropriate to focus on the comparison of the Functions (sequence of proximity) and Demo (objective verifiability order) analyses. The averages (over samples) multiple R's of the two methods for each of the criteria are very close, and differences are more parsimoneously attributable to chance than one method's being more effective in capturing variance than the other. The two modes of analysis have different sets of individual factors included as having significant beta weights in each analysis. Whereas the function factors, contribute greatly to the multiple R's in Function mode of analysis (e.g. 42% of the variance accounted for, Function Data, Sample 1,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Time Span), they contribute almost nothing in the Demo. analyses. Only with consequences as the criterion do these factors appear at all. The reverse is true of demographic factors in Functions analysis (age appears with time span but may be redundant with the criterion rather than an independent correlate). However social factors appear prominently in all analyses (e.g. over 40% of the accounted-for-variance in each of Functions and Demo data, Sample 1, Range), and although not as powerful, the personality factors also enter significantly in each analysis.

Although these results need more exploration, the signs are clear that sequential data entry may influence our interpretation of results, if not the results themselves.

Criteria Patterns

Of the criteria, Frequency was the least predictable by Stepwise Multiple R's (average R is .370 for frequency and ranges from .686 to .761 for the other criteria). Also its distributional properties deviate much farther from normality than those of Time Span, Range, and Consequences. In the Factor analyses of the criteria, each of the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

loads highly on a separate component. Finally, Frequency has no high beta weights on any of the canonical variates. It would seem that, in general, Frequency, as measured in this study at least, is not a constituent of the overall criteria set. This may be because of the dual nature of the Frequency score distribution (dichotomous-Poisson), rendering it unsuitable for these kinds of analyses. This could be investigated by transforming it into a dichotomy and analysing; then, excluding non-users, transforming the Frequency score distribution into an approximately normal distribution and running the analysis again. The results from this might indicate a five criteria set (e.g. Consequences, Range, Time Span, Use vs. Non-use and User Frequency). Alternatively, Frequency may just be a poor criterion; but, as indicated by Sadava (1974a), this is usually the sole continuous criterion in marijuana research. Further, since the practical aim of much of the research is to treat abusers (i.e. those with high consequences) of marijuana, or conceptualize marijuana as an example of the broader range of drugs, the absence of investigation into the criterion domain may be particularly unfortunate. Given the results of the present criteria factor analysis,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and the canonical correlation analysis, the inferential links from predictors (of marijuana smoking) to frequency, to the "real" interest may be very weak.

Canonical correlation analysis

In considering the results of this analysis, first a further explanation of the coefficient of redundancy (R_{dx}). As Stewart and Love (1968) point out, canonical correlations are not correlations between two sets of variables, but between two linear functions of those variables. The square of the canonical correlation cannot be interpreted as r^2 for simple correlation, or R^2 in multiple correlation: R_{dx} (or R_{dy}) is the quantity used to indicate the relationship between the sets of variables. This tells us the proportion of the variance of the criteria shared with the predictors. Thus the second canonical variate of Demo data, Sample 1 has twice the predictive power of the first variate, even though the first canonical correlation is higher.

The loadings of the criteria indicate two patterns of marijuana use (The second canonical variate of Demo., sample two, corresponds to the other second variates, but was not statistically significant).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

These might be characterized as the "normal" (i.e. low consequences associated with higher range and time span) and "abuser" (high perceived consequences associated with a long time span and low frequency of use) patterns. The abuser pattern is associated with personality variables, while the normal pattern seems to be associated with social and environmental predictors. In the case of the second variate, it would be very difficult to draw inferences on the causality sequence. To do this would require some form of path analysis, or a set of canonical correlation analyses utilizing Δ predictor scores and Δ criterion scores (see Blalock, 1971, for a set of readings on problems with causal models). At this point in the research, it seems clear that patterns in the criteria exist, and that these may result from or result in different variable complexes.

Multivariate Prediction with Conceptual Systems

A consistent finding in this investigation has been the system of variable sets at work. The original model (Figure 1) included five variable systems: behaviour, cognitive function, personality, perceived

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

environment, and demographic variables. The last is the least coherant (due perhaps to its addition, as an afterthought, to the questionnaire, without a definite conceptual basis). The other three predictive systems seem to have a degree of unity. Even if the individual variables enter differently in the SMR equations, the systems or factors were consistent (e.g. the social support variable, social sanctions factor, and social climate factor loaded significantly on all analyses). These systems also seem to predict different patterns in the canonical correlations. Although the domains of individual variables may be uncertain, by considering the predictors as sets, more reliance may be given to the results. Individual variable instability is not unique to this project, for instance, in the work of Jessor, Jessor & Finney (1973) significant differences between means did not always appear for individual variables when distinguishing between categories of marijuana users. However across samples, the same groups or sets of variables included significant differences. The large number of variables used, certainly included those with unknown domains, and a reconstruction of the variables as

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

orthogonal factors would probably improve reliability (Jesorr, Jessor & Finney, 1973, pp. 6-8).

The empirical testing of a multivariate system must be very carefully done. The consideration of score distributions is often never made, although this inattention may alone invalidate or compromise later findings. Similarly multiple regression replication must be more thoroughly done since useful predictors may be excluded or unreliable ones included. Finally, criteria patterns and the links between the measurement of a behaviour and its "real life" correspondant(s) must be examined conceptually and empirically.

The above work did not attempt to examine the results in context of Sadava's system or to relate them to the literature on marijuana research specifically, or deviant behaviour in general. The focus was on issues common to multivariate systems, and to problems often ignored for some reason or another. It is the consideration of these problems, and their solution, that will allow the construction of useful predictive systems, and their meaningful empirical testing.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CONCLUSIONS

The results of this thesis may be divided in two parts: those findings with regards to the content matter, cannabis use; and those related to data analysis and methodology.

The most intriguing and suggestive of the content findings is the type of predictor and criterion patterns found in the canonical analysis. Several implications point to the potential in replication with different measures and samples. First, consider the two criteria patterns that seem to emerge: that of a moderate or "social smoker" which may parrellel the social drinker in alcohol use; and the "abusive" pattern which includes high adverse consequences of use. If these are "true" patterns they are important in considering the treatment of marijuana problems users and for the attitudes towards marijuana itself. Further, comparisons between patterns for marijuana and alcohol abuse, may clarify the origins of and meaning of "drug abuse".

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

On the other side of the canonical analysis, the predictor patterns may not only help us to understand drug abuse but may help us visualize the importance of (and the utility of) the differences between situational and intrapersonal factors in psychology. It would seem that there are at least two components in the predictor set: one comprised of social-environmental variables (e.g. peer and family relationships, political orientation size etc.) which is related to the variance in normal behaviour, and a component of personal variables (e.g. time perspective, conformity issues, etc.) related more to the "extremes" of behaviour. A thorough investigation of these kinds of relationships can point out what each approach can do and how to implement a situational-personality system in psychology.

A second finding in the content area is much more specific but, if typical of marijuana research, of some importance. As pointed out by Sadava (1974a) the criterion used most often in marijuana research is frequency of use. This variable is used in a variety of forms: dichotomous (use - not use); ordinal group

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

scale (e.g. 0, 1 - 6, 7 - 15, etc.) and continuous interval (implicit in the use of Pearson product moment correlation coefficient or variants thereof). This study found that there were two major drawbacks to the "frequency of use" criterion. First, the scale and underlying distribution ssverely limit available statistics and restrict interpretations of findings with respect to this variable. Secondly, this variable seems to have little empirical relationship to other criteria of drug use, and more seriously may not be suitable as a criterion to be linked with drug abuse or problem behaviour - the aim of much drug research. These findings cast doubt on the results of previous research and on the generalization of findings to the problem area.

The methodology and data analysis issues raised by this thesis are important when considering the task of analyzing and interpreting the large data arrays generated in multivariate studies. Four problem areas were pointed out: frequency distri-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

butions, patterns of variables, raw data vs. factor scores and sequential variable entry. Warnings about the effects of the first two problems are always given when introducing the student to statistics, however, judging from the apparent lack of attention to these problems in non-medical drug use research these warnings may not be needed in practice. Although this thesis has not covered any new points with regards to the topics of frequency distributions and variable patterns, the results at least give a good example of the effects of focused consideration on the two areas.

The issues of type of data (raw vs. factor scores) and orthogonalization have received much less attention in the literature and the present thesis attempted to compare procedures and exemplify their use. Raw data scores alone, and factor scores alone, have been used in data analysis previously but this study shows the utility of considering both and cross-checking results. Quite often the name assigned to a factor, variable, or "cause" with influence

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

greatly interpretation and future research directions. If a result is obtained by "chance" (and as noted by Cooley & Lohnes, 1971, multivariate analysis often capitalizes on "chance"), succeeding research may be ambiguous, or interpretations forced by the historical "momentum". Although the statistical results may be verified by cross-validation techniques, interpretations are still based upon a single source. However, the concurrent use of factor and raw scores may provide "conceptual cross-validation". This means that one set of variables (factors) is mapped onto a set of factors (variables) formed in a different manner, and both are checked against the criterion (criteria). Hopefully this will establish more clearly just what it is that predicts (explains) the variance of the criteria.

Finally the utility of partitioning the predictor variables so that orthogonalization takes place in different sequences was demonstrated. Almost all techniques for orthogonalization (factor analysis) must extract components sequentially and it is important to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

know whether the results are relatively invariant or whether the technique determines the results. Since we used two opposite sequences of entry, the common results may be more trustworthy, and our subsequent interpretations may be more valid.

With increased use of multivariate analyses in the social sciences, the potential pitfalls must be avoided, and guidelines for meaningful interpretation drawn up. This thesis has attempted to explore some problem areas in multivariate analyses and to highlight potentially useful procedures.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPENDIX A

Proof For Gamma (Γ_{yi})

Let Z_i $(i=1,\ldots,m)$ be a series of independent, normally distributed variables with $Exp(Z_i)=0, Var(Z_i)=1$. Let Z_y be another normally distributed variable $Exp(Z_y)=0, Var(Z_y)=1$. Then Z_y, the predictable part of Z_y^y from Z_i equals $\sum_{i=1}^{m} \beta_i Z_i$.

But $\beta_i=r_i$ for orthogonal predictors (Mulaik, 1972, p.404-405), then $Z_y=\sum_{i=1}^{m} r_{yi} Z_i$. Let Γ_{yi} $(Exp(\Gamma_{yi})=0, Var(\Gamma_{yi})=1)$ be some variable, and α some constant such that :

$$Z_y = \alpha \Gamma_{yi} + \sum_{i=1}^{m} r_{yi} Z_i ,$$

Γ_{yi} indepent of Z_i. Then $Var(Z_y) = Var(\alpha \Gamma_{yi}) + \sum_{i=1}^{m} (Var(r_{yi} Z_i))$,

$$1 = \alpha^2 + \sum_{i=1}^{m} r_{yi}^2 ,$$

$$\alpha = \sqrt{1 - \sum_{i=1}^{m} r_{yi}^2} ,$$

Therefore: $$\Gamma_{yi} = \frac{Z_y - \sum r_{yi} Z_i}{\sqrt{1 - \sum r_{yi}^2}} .$$

(see also Anderson, 1958, Chapter 2)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPENDIX B

Delta Score Programme

```
      DIMENSION XDAT(35),YDAT(35),RXY(35),XYDEN(35),DELTA(35)
   50 FORMAT(8F10.4)
  500 FORMAT(5F7.3/28X,6F7.3,/8F7.3,7X,F7.3/10F7.3/F7.3,14X,2F7.3,21X,
     22F7.3/3F7.3,28X,3F7.3/2(10F7.3/),5F7.3,40X,I3)
  700 FORMAT(5F7.3,40X,I3,'91'/6F7.3,33X,I3,'92'/10F7.3,5X,I3,'93'/
     110F7.3,5X,I3,'94'/2F7.3,50X,I3,8X,I3,'95')
      READ(5,50)(RXY(K),K=1,33)
      DO 300 N=1,33
      XYDEN(N)=1./SQRT(1.-RXY(N)*RXY(N))
  300 CONTINUE
      DO 400 M=1,187
      READ(5,500)(XDAT(N),N=1,33),(YDAT(N),N=1,33),NUM
      DO 100 N=1,33
      IF(XDAT(N).LT.-20..OR.YDAT(N).LT.-20.) GO TO 99
      DELTA(N)=(YDAT(N)-RXY(N)*XDAT(N))*XYDEN(N)
      GO TO 100
   99 DELTA(N)=-99.
  100 CONTINUE
      WRITE(7,700)(DELTA(N),N=1,5),NUM,(DELTA(N),N=6,11),NUM,(DELTA(N),
     2N=12,21),NUM,(DELTA(N),N=22,31),NUM,(DELTA(N),N=32,33),M,NUM
  400 CONTINUE
      STOP
      END
```

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Gamma Score Programme

```
      DIMENSION ALPHA(10,12),NCOF(12),XDAT(12),FACSC(10)
     1,YDAT(10),SIGMA(10),XMEAN(12),XSTD(12)
      NFAC=10
      NVAR=12
      NSS=187
  500 FORMAT(2F3.0,3(6X,F3.0),F3.0,4F4.0,4X,2F4.0,11X,I3)
  700 FORMAT(7F10.5,2X,2I3/3F10.5,42X,2I3)

  550 FORMAT(20I4)
  555 FORMAT(8F10.6)
   10 READ(5,555) ALPHA
      READ(5,555) XMEAN
      READ(5,555) XSTD
   20 READ(5,550) NCOF
      DO 1000 NUM=1,NSS
      DO 155 N=1,NFAC
  155 FACSC(N)=0.
      READ(5,500) XDAT,NSUB

      TEST=0.
      DO 150 M=1,NVAR
      TEST=TEST+XDAT(M)
      IF(XDAT(M).GT.-95.) GO TO 150
      HOLD=NCOF(M)
      FACSC(HOLD)=-99.
      XDAT(M)=XMEAN(M)
  150 XDAT(M)=XDAT(M)*XSTD(M)
      IF(TEST.LE.-250.) GO TO 999
      DO 250 N=1,NFAC
      IF(FACSC(N).LT.-20.) GO TO 250
      DO 350 M=1,NVAR
      FACSC(N)=FACSC(N)+XDAT(M)*ALPHA(N,M)
  350 CONTINUE
  250 CONTINUE
      WRITE(7,700) (FACSC(J),J=1,7),NUM,NSUB,(FACSC(J),J=8,10),NUM,NSUB
  999 CONTINUE
 1000 CONTINUE
      STOP
      END
```

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix C

CORRELATIONS — Raw Data , Sample 1

	Freq. 1	Time 2	Range 3	Cons. 4
1	1.000000	0.249884	0.492633	-0.078437
2	0.249884	1.000000	0.719662	-0.488306
3	0.492633	0.719662	1.000000	-0.400560
4	-0.078437	-0.488306	-0.400560	1.000000
5	0.158142	0.100957	0.141028	-0.048416
6	0.134254	0.072574	0.099419	0.000323
7	0.073119	-0.023714	0.026709	0.035238
8	-0.219039	-0.455573	-0.429082	0.452180
9	-0.150383	-0.395357	-0.371124	0.369365
10	-0.232579	-0.435367	-0.411263	0.445978
11	-0.061778	-0.104573	-0.103255	0.155118
12	-0.019670	-0.112870	-0.084368	0.160672
13	-0.113771	-0.137686	-0.164088	0.197381
14	-0.062933	-0.229010	-0.177363	0.388277
15	-0.049222	-0.240948	-0.143274	0.330141
16	-0.076242	-0.221222	-0.204102	0.392063
17	0.310186	0.670891	0.670350	-0.497251
18	0.271774	0.519601	0.567349	-0.457342
19	0.181094	0.439777	0.451085	-0.437401
20	0.132019	0.123747	0.200803	-0.202592
21	0.161394	0.375104	0.348916	-0.267977
22	0.078461	0.140631	0.178301	-0.243349
23	0.082504	0.229872	0.165228	-0.265691
24	0.038802	0.037785	0.062818	-0.169117
25	-0.067883	-0.017869	-0.058009	0.074304
26	0.103474	0.050599	0.052732	-0.171120
27	-0.049509	0.129755	0.013475	-0.097367
28	-0.043021	-0.086985	-0.135877	0.040042
29	0.025137	-0.116460	-0.037855	0.049935
30	0.140115	0.266597	0.208295	-0.206468
31	-0.141523	-0.124588	-0.209398	0.038877
32	-0.030782	-0.092753	-0.052023	0.086532
33	-0.225432	-0.306802	-0.380525	0.356702
34	-0.285306	-0.561495	-0.573743	0.488240
35	0.289792	0.363108	0.321359	-0.286597
36	0.199915	0.190613	0.155593	-0.199181
37	0.136935	0.150606	0.107153	-0.050313
38	0.216896	0.214331	0.168295	-0.103007
39	0.239060	0.419471	0.426420	-0.397905

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Raw Data, Sample [redacted]

	Freq.	Time	Range	Cons.
40	-0.035849	-0.104491	-0.033839	-0.016754
41	-0.018426	-0.063958	-0.002862	-0.106413
42	-0.061578	-0.090256	-0.023628	0.098447
43	-0.171932	-0.056381	-0.061400	0.006823
44	-0.135311	-0.029986	-0.053166	-0.028680
45	-0.113455	-0.071120	-0.120675	-0.008519
46	-0.138392	-0.045218	-0.041850	0.031295
47	-0.074792	-0.114172	-0.126073	0.038616
48	-0.144323	-0.273848	-0.290843	0.174677
49	-0.156106	-0.209104	-0.113080	0.184917
50	-0.048777	-0.113761	-0.132248	0.034387
51	0.131824	0.013334	-0.059344	0.112991
52	0.182495	0.240187	0.267980	-0.117212
53	-0.089362	-0.078355	-0.035014	0.020376
54	-0.035042	-0.165129	-0.054121	0.126191
55	-0.013633	0.112736	0.109212	-0.166621
56	-0.097393	-0.042180	0.004833	-0.016089
57	-0.047837	-0.027029	-0.063839	0.174374
58	-0.126339	-0.285267	-0.327007	0.194890
59	0.232978	0.511072	0.440134	-0.454710
60	0.015394	0.069188	0.032714	-0.039310
61	0.000581	0.056794	0.007861	-0.149611
62	0.015042	0.201076	0.080043	-0.191414
63	0.197205	0.182411	0.108366	-0.244914
64	-0.023296	-0.020255	-0.012791	-0.019369
65	-0.002607	-0.008903	0.003936	0.017468
66	-0.024335	0.017548	0.035711	-0.023232
67	-0.061793	-0.023959	-0.002371	-0.063665
68	-0.081001	0.060187	0.047309	-0.029800
69	-0.016494	-0.119390	-0.071170	0.043318
70	-0.013367	-0.057711	-0.225459	0.103692
71	0.026921	-0.063157	-0.186523	0.031341
72	-0.239104	-0.086315	-0.069771	0.055263
73	0.048265	-0.098938	-0.151697	0.002221
74	0.117004	0.345497	0.284135	-0.198318
75	0.011005	0.086855	0.010076	0.083211
76	-0.072025	-0.013370	-0.079895	-0.127454
77	-0.041047	0.112238	0.106158	-0.170981
78	0.135379	0.344109	0.275445	-0.281755
79	0.040934	0.100866	0.115661	-0.169396
80	0.071905	0.117117	0.106189	-0.064011
81	0.047718	0.093208	0.120916	0.020491
82	0.087363	0.191803	0.141475	-0.039770
83	0.000726	-0.039886	0.012252	0.178589
84	-0.250550	-0.020636	-0.009452	0.067499
85	-0.045720	0.012354	0.066343	0.077519

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Raw Data, Sample 2

CORRELATIONS

	Freq 1	Time 2	Range 3	Cons. 4
1	1.000000	0.272934	0.261781	-0.147041
2	0.272934	1.000000	0.641323	-0.437990
3	0.261781	0.641323	1.000000	-0.273362
4	-0.147041	-0.437990	-0.273362	1.000000
5	-0.001012	-0.010568	0.027101	-0.055540
6	0.056412	-0.050184	-0.042733	0.033977
7	-0.085382	-0.115064	-0.024031	0.006418
8	-0.142843	-0.427808	-0.332615	0.392143
9	-0.118898	-0.336735	-0.259334	0.305632
10	-0.139025	-0.426925	-0.332146	0.389409
11	0.109492	0.043309	0.113640	0.159899
12	0.120411	-0.011775	0.066758	0.174971
13	0.056682	0.007871	0.053717	0.177641
14	-0.113028	-0.174229	-0.164380	0.151694
15	-0.114864	-0.204603	-0.126781	0.137321
16	-0.107660	-0.167179	-0.183853	0.166812
17	0.250197	0.676824	0.597502	-0.508666
18	0.188474	0.499672	0.472565	-0.440783
19	0.166267	0.412208	0.366035	-0.349605
20	-0.065232	0.167089	0.154589	-0.172229
21	0.093312	0.409423	0.321522	-0.268596
22	0.005936	0.172757	0.128030	-0.037533
23	0.129239	0.241958	0.158071	-0.303939
24	0.059736	0.079430	0.120643	0.041306
25	-0.051722	0.078845	0.048154	-0.016038
26	0.024920	-0.070630	0.030931	-0.052613
27	0.043093	0.107286	0.159868	-0.156139
28	-0.067160	-0.186069	-0.052360	0.097508
29	-0.071325	-0.044168	-0.066821	0.181855
30	0.140121	0.110642	0.140035	-0.204669
31	-0.062250	-0.194843	-0.065344	0.134471
32	-0.062950	-0.134916	-0.123562	0.202409
33	0.068655	-0.169081	-0.122211	0.251279
34	-0.271097	-0.555370	-0.531158	0.421321
35	0.115777	0.301799	0.294776	-0.270209
36	0.023191	0.087635	0.151912	-0.029822
37	0.124173	0.217327	0.165384	-0.231446
38	0.094855	0.266992	0.195055	-0.228340
39	0.075716	0.246914	0.279778	-0.250647

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Raw Data, Sample 2

	Freq.	Time	Range	Cons.
40	-0.020568	0.038497	0.020274	-0.011100
41	0.008844	0.072383	0.052494	-0.039822
42	0.052197	0.015363	0.120204	-0.006288
43	-0.048970	0.013875	0.025232	-0.008370
44	-0.162864	-0.000937	-0.062504	-0.001168
45	0.124506	0.007584	0.034163	0.006196
46	-0.028716	-0.018076	-0.055226	0.055302
47	-0.077674	-0.159395	-0.163141	0.034627
48	-0.030834	-0.308852	-0.223155	0.195306
49	-0.146138	-0.200646	-0.168780	0.211513
50	-0.245119	-0.239201	-0.289156	0.131079
51	-0.099408	-0.015915	-0.001178	-0.018500
52	0.012603	0.012488	-0.032954	-0.098488
53	-0.093575	-0.110700	-0.177571	0.055059
54	-0.028048	0.003952	0.076260	0.078838
55	0.069455	-0.045244	0.025421	-0.122811
56	0.009641	0.020003	-0.068163	0.080879
57	-0.085119	-0.023139	-0.009815	0.032512
58	-0.097746	-0.408900	-0.284830	0.352052
59	0.331460	0.548586	0.561500	-0.443214
60	0.047789	0.066854	0.128677	0.075123
61	0.154537	0.254332	0.312611	-0.169877
62	0.045282	-0.038266	0.008338	-0.061103
63	0.054731	0.170527	0.158124	-0.152805
64	-0.030481	0.086975	0.005601	-0.169014
65	-0.091244	0.085863	-0.012662	-0.153076
66	0.045931	-0.077083	-0.013259	0.015372
67	-0.052744	-0.153395	-0.130068	0.080998
68	-0.079875	-0.061894	-0.047070	0.156342
69	-0.003931	-0.187378	-0.098102	-0.021325
70	-0.087573	-0.069588	-0.009512	0.054919
71	0.053732	0.081023	-0.052584	-0.011282
72	-0.015106	-0.116641	-0.114434	0.000725
73	0.006564	0.115753	0.187922	0.102412
74	0.137960	0.288669	0.271514	-0.234173
75	-0.051669	0.035965	-0.059823	0.141460
76	0.067147	0.015717	0.071866	-0.116554
77	-0.056981	-0.058117	-0.107382	-0.111862
78	0.090251	0.254256	0.159333	-0.127882
79	0.029995	0.166005	0.039428	-0.154525
80	-0.005772	0.034903	-0.012962	-0.042216
81	0.050283	0.044637	0.047505	0.030474
82	0.039144	0.170525	0.110728	-0.056558
83	-0.128563	-0.101911	-0.173433	0.123835
84	0.001264	-0.036330	-0.041599	0.057133
85	0.040913	0.216409	0.201962	-0.145882

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

	Freq. VAR001	Time VAR002	Range VAR003	Cons. VAR004
VAR001	1.00000	0.25101	0.49528	-0.07856
VAR002	0.25101	1.00000	0.72244	-0.49253
VAR003	0.49528	0.72244	1.00000	-0.40006
VAR004	-0.07856	-0.49253	-0.40006	1.00000
VAR005	-0.21435	-0.44835	-0.42161	0.42877
VAR006	-0.05171	-0.22701	-0.16393	0.37785
VAR007	0.13509	0.04931	0.10094	0.00974
VAR008	-0.04955	-0.07579	-0.08291	0.11958
VAR009	0.16667	0.29956	0.36304	-0.13292
VAR010	0.02770	0.17736	0.16651	-0.00430
VAR011	-0.05854	-0.02171	-0.05587	0.07107
VAR012	0.05762	0.17872	0.10808	-0.16467
VAR013	0.08807	0.01995	0.07888	-0.08225
VAR014	0.03662	-0.00937	0.04658	-0.10043
VAR015	0.09826	0.27825	0.28676	-0.22326
VAR016	0.00724	-0.01086	0.01743	-0.08805
VAR017	-0.03808	0.00425	0.02693	-0.00877
VAR018	-0.04329	-0.19911	-0.11839	0.03403
VAR019	-0.10857	0.01369	-0.09022	-0.06671
VAR020	0.13414	-0.01822	-0.07398	0.01395
VAR021	0.07300	0.05325	0.09611	-0.03723
VAR022	-0.02941	0.02978	-0.04163	-0.04552
VAR023	0.05353	0.01926	0.02988	-0.09652
VAR024	-0.12187	-0.01676	-0.06263	-0.02203
VAR025	-0.00471	0.09258	0.02976	-0.01677
VAR026	0.05526	0.03931	0.08445	-0.09145
VAR027	-0.10191	-0.13558	-0.14851	0.04269
VAR028	-0.03114	0.00402	-0.07849	0.08930
VAR029	-0.00399	-0.05501	0.03445	0.13780
VAR030	0.04798	0.09903	0.01934	-0.02552
VAR031	0.01190	0.09667	0.20238	0.09403
VAR032	0.06036	0.07933	0.13121	-0.00054
VAR033	-0.08585	0.03752	-0.01100	-0.01169
VAR034	-0.03815	-0.03472	-0.05247	0.05466
VAR035	0.08965	-0.05752	0.07155	-0.02675
VAR036	-0.01619	0.07966	0.01409	-0.01115
VAR037	-0.06647	0.02146	0.01804	-0.09292
VAR038	-0.08315	0.04900	-0.02020	-0.07906
VAR039	0.07040	0.06357	0.02114	-0.07984
VAR040	-0.10212	-0.10584	-0.05078	0.05275
VAR041	-0.03143	0.04471	-0.05826	0.02242
VAR042	-0.01091	0.19284	0.10156	0.01987
VAR043	0.04680	-0.04128	0.01626	-0.04740
VAR044	0.02025	-0.07698	-0.01365	-0.00892
VAR045	0.00742	-0.00335	0.05270	-0.06249

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Functions Factor, Sample 2

	Freq. VAR001	Time VAR002	Range VAR003	Cons. VAR004
VAR001	1.00000	0.28003	0.26321	-0.15418
VAR002	0.28003	1.00000	0.62854	-0.44206
VAR003	0.26321	0.62854	1.00000	-0.26488
VAR004	-0.15418	-0.44206	-0.26488	1.00000
VAR005	-0.14855	-0.41287	-0.31759	0.38669
VAR006	-0.13211	-0.18294	-0.16946	0.13041
VAR007	-0.01255	-0.07325	-0.02633	0.03133
VAR008	0.14063	0.09244	0.14057	0.12987
VAR009	0.13678	0.33229	0.34221	-0.23222
VAR010	0.01523	0.27899	0.15847	-0.04740
VAR011	-0.05134	0.08402	0.06888	-0.05310
VAR012	0.11156	0.18417	0.10105	-0.31537
VAR013	-0.11007	0.05270	0.07966	-0.09076
VAR014	-0.05119	0.04021	-0.00366	0.01937
VAR015	0.09523	0.19352	0.18137	-0.16562
VAR016	0.03840	0.04023	0.09353	0.08169
VAR017	-0.06583	-0.13788	-0.06242	0.02876
VAR018	-0.09332	-0.12187	-0.03149	0.04863
VAR019	-0.04642	-0.13880	0.02818	0.05059
VAR020	-0.04831	-0.09354	-0.08934	0.06698
VAR021	-0.08240	0.00880	0.04351	-0.02558
VAR022	-0.02743	-0.15206	-0.03686	0.09202
VAR023	-0.12569	0.03850	0.01765	-0.08624
VAR024	0.11070	0.01322	-0.03318	-0.10762
VAR025	0.00097	-0.00796	-0.20405	0.02198
VAR026	-0.04971	-0.04823	-0.04885	0.11334
VAR027	0.11205	-0.03745	-0.02757	0.00152
VAR028	0.02045	0.13063	0.05201	-0.10017
VAR029	0.05722	0.09226	0.05753	0.00564
VAR030	0.02096	0.01374	-0.08174	-0.04740
VAR031	0.11752	-0.01713	0.03400	-0.06674
VAR032	0.02045	0.13063	0.05201	-0.10017
VAR033	-0.06999	-0.12608	-0.13512	-0.05682
VAR034	0.05406	0.16017	0.11397	0.08375
VAR035	0.01245	0.15330	0.17705	-0.20506
VAR036	-0.02196	0.07990	0.06207	-0.09379
VAR037	0.01242	0.02223	0.05453	-0.02149
VAR038	0.02580	0.01837	-0.02426	0.09587
VAR039	0.02534	-0.05329	0.01509	-0.10864
VAR040	0.01266	0.01448	0.05838	0.10423
VAR041	0.03239	0.11118	0.02314	-0.09154
VAR042	0.06922	0.12605	0.05772	0.05320
VAR043	-0.03796	-0.02856	-0.01374	-0.09330
VAR044	-0.07238	-0.00104	-0.06642	-0.07239
VAR045	-0.04677	0.06581	0.03402	-0.00439

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Demo Factors, Sample 1

	Freq. VAR001	Time VAR002	Range VAR003	Cons. VAR004
VAR001	1.00000	0.25101	0.49528	-0.07856
VAR002	0.25101	1.00000	0.72244	-0.49253
VAR003	0.49528	0.72244	1.00000	-0.40006
VAR004	-0.07856	-0.49253	-0.40006	1.00000
VAR005	0.11901	0.33803	0.31697	-0.25865
VAR006	-0.16892	-0.08259	-0.16370	0.05065
VAR007	-0.05574	-0.06503	-0.09372	-0.11056
VAR008	-0.04512	-0.07553	-0.04068	-0.01765
VAR009	0.01763	0.13557	0.02049	-0.14814
VAR010	-0.02876	0.11128	0.01499	0.15168
VAR011	0.05941	-0.02158	0.04421	-0.07589
VAR012	-0.00570	0.04712	0.07421	-0.08513
VAR013	0.01438	0.16786	0.17100	-0.14671
VAR014	0.07608	0.25154	0.14654	-0.15730
VAR015	0.03832	0.02403	0.05105	-0.10468
VAR016	-0.15273	-0.02861	-0.01092	0.15867
VAR017	0.02889	0.03269	0.10392	-0.03310
VAR018	-0.07911	-0.04150	-0.09929	-0.02491
VAR019	-0.21542	-0.06165	-0.14827	-0.03173
VAR020	-0.04361	-0.01691	0.01484	-0.02660
VAR021	0.12596	-0.06750	-0.01295	-0.12461
VAR022	0.01526	-0.01701	0.01810	0.10838
VAR023	-0.10105	0.05704	-0.01164	-0.05339
VAR024	-0.01881	-0.05005	-0.01668	-0.02495
VAR025	0.06843	0.03001	-0.05863	-0.08198
VAR026	0.03268	0.19337	0.03611	-0.12843
VAR027	-0.07857	-0.11560	0.00237	-0.01435
VAR028	-0.20230	0.04911	-0.06264	-0.08895
VAR029	0.01905	0.01177	0.00345	0.03030
VAR030	-0.02002	-0.20857	-0.09392	0.22341
VAR031	-0.11916	-0.05502	-0.13339	-0.04261
VAR032	0.05446	0.18165	0.04966	-0.07219
VAR033	-0.11641	0.11990	-0.03014	-0.12002
VAR034	0.03717	0.08620	0.16841	-0.04209
VAR035	-0.06064	-0.08521	-0.10878	-0.08693
VAR036	0.02576	0.03168	-0.02331	-0.05981
VAR037	0.12560	0.37265	0.40352	-0.34459
VAR038	0.05837	0.06953	0.14099	-0.08770
VAR039	0.01882	0.09296	0.12362	-0.03008
VAR040	0.16276	0.17022	0.14819	-0.10369
VAR041	0.08277	-0.00362	0.10297	-0.03161
VAR042	0.07360	-0.01301	0.04887	-0.04424
VAR043	0.00162	-0.04875	-0.03160	0.12225
VAR044	-0.05203	-0.03576	0.00287	0.10126
VAR045	-0.08808	-0.08095	-0.12994	0.15654
VAR046	0.09747	0.00574	0.05784	-0.02270
VAR047	-0.03885	-0.01059	-0.07536	0.14703

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Demo. Factors, Sample 2

	Freq. VAR001	Time VAR002	Range VAR003	Cons. VAR004
VAR001	1.00000	0.27443	0.26321	-0.14785
VAR002	0.27443	1.00000	0.64483	-0.44039
VAR003	0.26321	0.64483	1.00000	-0.27486
VAR004	-0.14785	-0.44039	-0.27486	1.00000
VAR005	0.11122	0.28282	0.20975	-0.15423
VAR006	0.00234	-0.03956	-0.02685	0.13053
VAR007	-0.05253	0.14347	0.13516	-0.20280
VAR008	-0.00426	0.02284	0.04226	0.04883
VAR009	0.02486	0.04275	0.08938	-0.17852
VAR010	0.00763	0.06850	-0.02984	0.08625
VAR011	-0.08946	-0.02490	-0.11290	-0.12572
VAR012	0.07318	0.07703	0.06678	0.00574
VAR013	0.04274	0.00865	0.08465	0.07730
VAR014	0.14153	0.38063	0.31208	-0.34728
VAR015	0.12358	0.04976	0.03508	-0.15380
VAR016	-0.02727	-0.03072	-0.00006	0.04871
VAR017	-0.01400	-0.07811	-0.05805	-0.12879
VAR018	-0.25966	-0.07731	-0.02014	0.07123
VAR019	-0.00792	-0.02702	-0.11947	0.06287
VAR020	0.13887	0.01995	0.01563	-0.04799
VAR021	-0.02608	0.01210	-0.06541	-0.03336
VAR022	-0.04647	-0.05557	-0.01223	0.06490
VAR023	-0.11565	-0.01106	-0.10785	-0.08834
VAR024	0.07522	0.07396	0.02463	0.04612
VAR025	0.02097	0.10395	-0.01369	-0.12568
VAR026	-0.02388	0.02046	-0.05411	-0.03545
VAR027	-0.05991	-0.04219	0.01728	0.03901
VAR028	-0.13344	-0.09201	-0.09793	0.06431
VAR029	-0.10034	-0.02049	-0.03929	-0.04745
VAR030	-0.03461	-0.09861	-0.18079	0.00678
VAR031	-0.00221	-0.08691	0.00460	0.12752
VAR032	0.00938	0.20090	0.20715	-0.17465
VAR033	-0.00794	0.08124	0.13933	-0.12503
VAR034	-0.02833	-0.01079	0.04332	0.00827
VAR035	-0.05627	-0.10089	-0.10386	-0.01482
VAR036	0.01694	0.07796	0.08253	-0.00531
VAR037	0.15675	0.38100	0.38002	-0.20629
VAR038	-0.05547	0.04151	-0.02657	0.02091
VAR039	0.06213	0.21437	0.15566	0.04784
VAR040	0.08805	0.07861	0.07557	-0.11753
VAR041	-0.03145	0.06484	0.10331	0.01184
VAR042	-0.00904	-0.00934	-0.04122	0.06873
VAR043	-0.06197	0.08787	0.07726	-0.03944
VAR044	0.00351	-0.05133	-0.04724	0.12675
VAR045	0.00321	-0.01064	0.03339	0.05894
VAR046	0.05590	0.01127	0.00020	0.04778
VAR047	0.01358	0.07259	0.03016	0.04284

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

REFERENCES

BASIS: Burroughs Advanced Statistical Inquiry System, as described in Basis at Brock. (ed. F.R.Skilton) Brock University, 1971.

Blalock, H.M., Causal models in the social sciences. Chicago, Ill.: Aldine, Atherton, Inc, 1971.

Carroll, J.B., The nature of the data, or how to choose a correlation coeffient. Psychometrika, 1961, 26, 4, 347-372.

Cattell, R.B., Factor Analysis. New York: Harper and Row., 1952.

Cooley, W.W., & Lohnes, P.R., Multivariate data analysis. Toronto: John Wiley and Sons, Inc, 1971.

Ferguson, G.A., Statistical analysis in psychology and education (3rd. Ed). Toronto: McGraw-Hill Book Co., 1971.

Jackson D.N., Aquiescence response styles: Problems of identification and control. In I.A.Berg (Ed), Response set in personality assessment. Chicago: Aldine. 1967.

Jessor, R., Collins, M.I., & Jessor, S.L. On becoming a drinker: Social psychological aspects of an adolescent transition. In R.E.Seixas (Ed), Nature and nurture in alcoholism. New York: Annals of N.Y.Academy of Sciences, 1972, 197, 199-213.

Jessor, R., Carman, R.S., & Grossman, P.H. Expectations of need satisfaction and drinking patterns of college students. Quarterly Journal of Studies on Alcohol. 1968, 29, 101-116.

Jessor, R., Jessor, S.L., & Finney, J. Asocial psychology of marijuana use: Longitudinal studies of high school and college youth. JPSP, 1973, 26, 1-15.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

King, F.W. Anonymous versus identifiable questionnaires in drug usage surveys. American Psychologist, 1970, 25 982-985.

Mulaik, S.A. The foundations of factor analysis. Toronto: McGraw-Hill Book Company, 1972.

Rotter, J.B. Social learning and clinical psychology. Englewood Cliffs, N.J.: Prentice-Hall, 1954.

Rotter, J.B., Chance, J.E., and Phares, E.J. Applications of a social learning theory of personality. Toronto: Holt, Rinehart and Windston, Inc., 1972.

Sadava, S.W. Stages of college student drug use: A methodological contribution to cross-sectional study. Journal of Consulting and Clinical Psychology, 1972, 38, 298 (and extended report, mimeo.).

Sadava, S.W. Patterns of college student drug use: A longitudinal social learning study. Psychological Reports, 1973, 33, 75-86.

Sadava, S.W. Research on illicit drug use: A critical analysis. Genetic Psychology Monographs, 1974, in press (a).

Sadava, S.W. Patterns and meanings of college student drug use: A social learning approach. Canadian Journal of Behavioural Science, 1974, in press (b).

SPSS: Statistical package for the social sciences as described in Statistical package for the social sciences. Nie, N., Bent, D.H., and Hull, C.H., Toronto: McGraw-Hill, 1970.

SSP: Scientific Subroutine Package, Version III, IBM release GH20-0205-4, 1970.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Stewart, D., and Love, W., A general canonical correlation index. Psychology Bulletin, 1968, 70, 3(1).

Tsokos, C.P., Probability distributions: An introduction to probability theory with applications. Belmont, California: Duxbury Press, 1972.

Van de Geer, J.P., Introduction to multivariate analysis for the social sciences. San Francisco: W.H.Freeman and Co., 1971.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

VITA AUCTORIS

1947	Born in St. Catharines. Ont., to Alexander Simson and Edythe Jean Forsyth.
1953-66	Educated at L.G.Lorriman and Thorold High Schools, Thorold Ont.
1973	Graduated with B.A., from Brock University, St. Catharines.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.