Psychological Bulletin
1968, Vol. 70, No. 6, 426–443

MULTIPLE REGRESSION AS A GENERAL DATA-ANALYTIC SYSTEM [1]

JACOB COHEN

New York University

Techniques for using multiple regression (MR) as a general variance-accounting procedure of great flexibility, power, and fidelity to research aims in both manipulative and observational psychological research are presented. As a prelude, the identity of MR and fixed-model analysis of variance/covariance (AV/ACV) is sketched. This requires an exposition of means of expressing nominal scale (qualitative) data as independent variables in MR. Attention is given to methods for handling interactions, curvilinearity, missing data, and covariates, for either uncorrelated or correlated independent variables in MR. Finally, the relative roles of AV/ACV and MR in data analysis are described, and the practical advantages of the latter are set forth.

If you should say to a mathematical statistician that you have discovered that linear multiple regression analysis and the analysis of variance (and covariance) are identical systems, he would mutter something like, "Of course—general linear model," and you might have trouble maintaining his attention. If you should say this to a typical psychologist, you would be met with incredulity, or worse. Yet it is true, and in its truth lie possibilities for more relevant and therefore more powerful exploitation of research data.

That psychologists would find strange the claimed equivalence of multiple regression (MR) and the fixed-model analysis of variance (AV) and covariance (ACV) is readily understandable. The textbooks in "psychological" statistics treat these matters quite separately, with wholly different algorithms, nomenclature, output, and examples.

MR is generally illustrated by examples drawn from the psychotechnology of educational or personnel selection, usually the prediction of some criterion (e.g., freshman grade point average) from predictors (e.g., verbal and quantitative score, high school rank). The yield is a multiple correlation (R) and a regression equation with weights which can be used for optimal prediction. The multiple R and the weights are subjected to significance testing, and conclusions are drawn about the effectiveness of the prediction, and which predictors do and do not contribute significantly to the prediction.

By way of contrast, AV and ACV are generally illustrated by pure research, manipulative experiments with groups subjected to different treatments or treatment combinations. Means and variances are found and main effect, interaction, and error mean squares computed and compared. Conclusions are drawn in terms of the significance of differences in sets or pairs of means or mean differences. More analytic yield of one or both of these systems is sometimes presented, but the above is a fair description of the respective thrusts of the two methods, and they are clearly different.

The differences are quite understandable, but the basis for this understanding comes primarily from the history and sociology of behavioral science research method and not from the essential mathematics. MR began to be exploited in the biological and behavioral sciences around the turn of the century in the course of the study of *natural* variation (Galton, Pearson, Yule). A couple of decades later, AV and ACV came out of the structure of (agronomic) experimentation, that is, of *artificial* or experimentally manipulated variation, where the treatments were carefully varied over the experimental material in efficient and logically esthetic experimental designs. The

[1] This work was supported by Grant No. MH 06137 from the National Institute of Mental Health of the United States Public Health Service, and by an open computing grant from Abacus Associates, Inc., New York, N. Y., to whom grateful acknowledgement is accorded. The author is also grateful to the members of the Society of Multivariate Experimental Psychology for their constructive response when this material was presented at their annual meeting in Atlanta, Georgia, November 1966. This work profited greatly from detailed critiques supplied by Robert A. Bottenberg and Joe H. Ward, Jr., but since not all their suggestions were followed, they share no responsibility for any defects in the result.

chief architect here was R. A. Fisher. These historical differences resulted in differences in tradition associated with substantively different areas and value systems in the psychological spectrum (cf. Cattell, 1966).

Yet the systems are, in the most meaningful sense, the same.

One of the purposes of this article is to sketch the equivalence of the two systems. In order to do so, it is necessary to show how nominal scales ("treatment," religion) can be used as "independent" variables in MR; the same is shown for "interactions." It is also necessary to demonstrate how multiple R^2 (and related statistics) can be computed from fixed-model AV and ACV output. Once the case is made for the *theoretical* equivalence of the two systems, the *practical* advantages of MR will be presented, which, given the foregoing, will be seen to constitute a very flexible general system for the analysis of data in the most frequently arising circumstance, namely, where an interval scaled or dichotomous (dependent) variable is to be "understood" in terms of other (independent) variables, however scaled.

A word about originality. Most of the material which follows was "discovered" by the author, only to find, after some painstaking library research, that much of it had been anticipated in published but not widely known works (chiefly Bottenberg & Ward, 1963; Li, 1964). Thus, no large claim for originality is being made, except for some of the heuristic concepts and their synthesis in a general data-analytic system realized by means of MR.

The Equivalence of the Systems: Nominal Scales as Independent Variables in MR

Some of the apparent differences in MR and AC/ACV lie in their respective terminologies. The variable being analyzed (from AV and ACV) and the criterion variable (from MR) are the same, and will be called the dependent variable and symbolized as Y. The variables bearing on Y, variously called main effect, interaction, or covariate in AV and ACV (depending on their definition and design function), and predictor variables in MR will be called independent variables, and symbolized as X_i $(i = 1, 2, \cdots k)$. Each X_i consumes *one* degree of freedom (*df*). In complex problems (e.g., factorial design, curvilinear analysis), it is convenient to define sets of the X_i, each such set representing a single research variable or factor.

In the conventional use of MR, the X_i are ordered quantitative variables, treated as equal interval scales. Thus, in a study of the prediction of freshman grade point average (Y), one might have X_1 = verbal aptitude score, X_2 = quantitative aptitude score, X_3 = percentile rank in high school graduating class, and X_4 = Hollingshead socio-economic status index. Thus, $k = 4$, and the question of sets need not arise (or, they may be thought of as four sets, each of a single variable). But what if one wanted to include *religion* among the X_i? Or alternatively, if the entering class were to be assigned randomly to four different experimental teaching systems, how would experimental group assignment be represented? More generally, how does one accommodate a purely nominal or qualitative variable as an independent variable in MR?

Imagine a simple situation in which a dependent variable Y is to be studied as a function of a nominal scale variable G, which has four "levels": groups G_1, G_2, G_3, and G_4. For concreteness, Y and G may be taken as having the following alternative meanings:

Research Area	Y	The G Set: G_1, G_2, G_3, G_4	
Social Psychology	Attitude toward United Nations	Religion:	Protestant Catholic Jewish Other
Clinical Psychology	Suggestibility	Diagnosis:	Paranoid Schizophrenia Nonparanoid Schizophrenia Compulsive Neurosis Hysterical Neurosis

Physiological Psychology	Retention	Treatment: Drug and Frontal Lesion Drug and Control Lesion No Drug and Frontal Lesion No Drug and Control Lesion

Formally, what is being posited is the assignment, not necessarily equally, of each of n cases into (four) mutually exclusive and exhaustive groups, no matter whether G is an organismic, naturally occurring variable or one created by the experimenter's manipulative efforts on randomly assigned subjects.

The expression of group membership as independent variables in MR can be accomplished in several ways, all equivalent in a sense to be later described. The intuitively simplest of these is "dummy" variable coding (Bottenberg & Ward, 1963; Suits, 1957).

Dummy Variable Coding

Table 1 presents various coding alternatives for the rendition of membership in one of four groups. Columns 1, 2, and 3 represent a dummy variable coding scheme. It involves merely successively dichotomizing so that each of $3(= g - 1)$ of the $4(= g)$ groups is distinguished from the remainder as one aspect of G. For example, on X_1 all subjects in G_1 are scored 1 and all others, without differentiation, are scored 0. Thus, this variable by itself carries only some of the information in the G variable as a whole, for example, Protestant versus all other, or Paranoid Schizophrenia versus all other. However, the three variables coded as in Columns 1, 2, and 3 together exhaust the information of the G variable. One might think that a fourth independent variable, one which distinguishes G_4 from all others, would be necessary, but such a variable would be redundant. In the usual MR system which uses a constant term in the regression equation, it requires no more than $g - 1$ independent variables (no matter how coded) to represent g groups of a G nominal scale. A fourth X_i here is not only unnecessary, but its inclusion would result in indeterminacy in the computation of the MR constants. This is an instance of a more general demand on the set of independent variables in any MR system: no independent variable in the set may yield a multiple R with the remaining independent variables of 1.00. This constraint on the independent variables (in matrix algebraic terms, the demand that their data matrix be nonsingular or of full rank) would be violated if we introduced a fourth variable, since, in that case, any of the four X_i would yield $R = 1.00$ when treated as a dependent variable regressed on the other three. In terms that are intuitively compelling, one can see that members of G_4 are identified uniquely on the X_1, X_2, X_3 vector as 0, 0, 0, that is, as not G_1, not G_2, and not G_3, thus not requiring a fourth dichotomous X_i. G_4 is not being slighted; on the contrary, as will be shown below, it serves as a reference group. Any group may be designated for this role, but if one is functionally a control or reference group, so much the better.[2]

Before we turn to a consideration of X_1, X_2 and X_3 as a set of variables, let us consider them separately. Each can be correlated with the dependent variable Y. A set of artificial data was constructed to provide a concrete illustration. For $n = 36$ cases, a set of three-digit Y scores was written, the cases assigned to four groups and coded for X_i as described. The resulting product moment r's (point-biserial) were $r_{Y1} = -.5863$, $r_{Y2} = .0391$, and $r_{Y3} = .4965$. When squared, the resulting values indicate the proportion of the Y variance each distinction accounts for: $r^2_{Y1} = .3437$, $r^2_{Y2} = .0015$, and $r^2_{Y3} = .2465$. Thus, for example, the Protestant versus non-Protestant variable accounts for .3437 of the vari-

[2] It is of interest to note that information about the "omitted" group, here G_4 (more generally, G_0), is readily recovered. The value for the correlation of the dichotomy for that group with *any* variable Z (r_{Z0}) is a simple function of the r's of the other variables with Z (r_{Zi}) and the standard deviations of the X_i, namely

$$r_{Z0} = (-\sum_{i=1}^{g-1} r_{Zi}\sigma_i)/\sigma_0$$

where

$$\sigma_i = [n_i(n - n_i)/n^2]^{\frac{1}{2}}, \quad \text{similarly for } \sigma_0.$$

When all groups are of the same size, this simplifies to

$$r_{Z0} = -\sum_{i=1}^{g-1} r_{Zi}$$

This relationship will hold whatever the nature of Z; it need not even be a real variable,—it will hold if Z is a factor in the factor-analytic sense, unrotated or rotated, with the r_{Zi} being factor loadings.

TABLE 1

ILLUSTRATIVE CODING FOR A NOMINAL SCALE

Nominal scale variable	Columns																	
	1	2	3		4	5	6		7	8	9		10	11	12		13	14	15
	X_1[a]	X_2	X_3		X_1	X_2	X_3		X_1	X_2	X_3		X_1	X_2	X_3		X_1	X_2	X_3
G_1	1	0	0		1	1	0		1	1	1		5	25	125		1	−7	0
G_2	0	1	0		1	−1	0		1	−1	−1		0	0	0		−1	−1	0
G_3	0	0	1		−1	0	1		−1	1	−1		−4	16	−64		−4	$\frac{1}{2}$	24
G_4	0	0	0		−1	0	−1		−1	−1	1		6	36	216		1	6	−1

[a] Independent variable.

ance in Attitude toward the United Nations dependent variable, as represented in the sample.

Whether the .3437 value can be used as an estimator of the proportion of variance which G_1 versus remainder accounts for *in the population* of naturally occurring G depends on the way G was sampled. If the n cases of the sample were obtained by randomly sampling from the population as a whole so that the proportion of G_1 cases in the sample, n_1/n reflects their population predominance, .3437 estimates the proportion of variance in the natural population. However, if G was sampled to yield equal n_i in the g groups (or some other nonrepresentative numbers), the .3437 value is projectible to a similarly distributed artificial population. The statistical purist would abjure the use of r or r^2 (and R or R^2) in such instances, but if one understands that the parameters being estimated are for populations whose X_i characteristics are those of the sample, no inappropriate errors in inference need be made, and a useful analytic tool becomes available.

Although the separate r^2_{Yi} are analytically useful, our purpose is to understand the operation of X_1, X_2, X_3 as a set, since it is as a set that they represent G as the four-level nominal scale. The r^2_{Yi} cannot simply be added up to determine how much Y variance G accounts for, since dummy variables are inevitably correlated with each other. Mutually exclusive assignment means that membership in one group G_i necessarily means nonmembership in any other, G_j, hence a negative relationship. The product moment r (i.e., the phi coefficient) between such dichotomies, that is, between G_i and G_j or X_i and X_j when expressed in dummy variable form, is

$$r_{ij} = -\sqrt{\frac{n_i n_j}{(n - n_i)(n - n_j)}} \qquad [1]$$

where n_i, n_j are the sample sizes of each group, and n is the total sample size over all g groups. When sample sizes are all equal, the formula simplifies to

$$r_{ij} = -\frac{1}{g - 1}, \qquad [2]$$

that is, the negative reciprocal of one less than the number of groups; thus, in our running artificial example, if we assume the four groups equal in size, the phi coefficients among the X_i dichotomies are all $-\frac{1}{3}$.

The fact that the independent variables representing group membership are correlated with each other poses no special problem for MR, which is designed to allow for this in whatever guise it appears. But it does alert us to the fact that the proportions of Y variance given by the r^2_{Yi} are overlapping. If we now compute the multiple R^2 using X_1, X_2, and X_3 as independent variables, the value we find in the artificial data is $R^2_{Y.123} = .4458$. This is interpreted as meaning that G (religion, diagnosis, or treatment group membership) accounts for .4458 of the variance in the dependent variable Y, and in the exact sense ordinarily understood.

Identity with Analysis of Variance

Consider the more familiar AV analysis of these data. The Y scores can be assembled into the four G groups and a one-way AV performed. This yields the usual sums of squares for between groups (B SS), for within groups (W

SS), and their total (*T SS*). If we determine the proportion of *T SS* which *B SS* constitutes, we have η^2 (eta square), the squared correlation ratio. This statistic has, as its most general interpretation, the proportion of variance of the dependent variable accounted for by *G*-group membership, or, equivalently, accounted for by the group *Y* means. (Unfortunately, tradition in applied statistics textbooks and courses has focused on a narrow, special-case interpretation of η as an index of curvilinear correlation. For a broader view, see Cohen, 1965, pp. 104–105 and Peters & Van Voorhis, 1940, pp. 312–325 and, particularly, 353–357).

If we compute $\eta^2_{Y \cdot G}$ for the artificial data, we find

$$\eta^2_{Y \cdot G} = \frac{B\,SS}{T\,SS} = \frac{12127.0}{27205.6} = .4458 \quad [3]$$

Thus, our MR coding procedure yields an $R^2_{Y \cdot 123}$ exactly equal to $\eta^2_{Y \cdot G}$, interpretable as the proportion of *Y* variance for which *G* accounts. The parallel goes further. It is demonstrable that the "shrunken" or *df*-corrected R^2 (McNemar, 1962, pp. 184–185) is identically the same as Kelley's "unbiased" squared correlation ratio, epsilon-square (Cohen, 1965, p. 105; Cureton, 1966; Peters & Van Voorhis, 1940, pp. 319–322).

Furthermore, if one tests either of these results for significance, one obtains identically the same *F* ratio, for identically the same *df*:

For the $R^2_{Y \cdot 123}$, using the standard formula (e.g., McNemar, 1962, p. 283)

$$F = \frac{R^2_{Y \cdot 123\ldots k}/k}{(1 - R^2_{Y \cdot 123\ldots k})/(n - k - 1)}$$

$$= \frac{R^2_{Y \cdot 123\ldots k}/(g - 1)}{(1 - R^2_{Y \cdot 123\ldots k})/(n - g)}$$

$$= \frac{.4458/(4 - 1)}{(1 - .4458)/(36 - 4)} = 8.580, \quad [4]$$

for numerator (regression) $df = k = g - 1 = 3$ and denominator (residual or error) $df = n - k - 1 = n - g = 32$.

The significance of η^2 is, of course, the significance of the separation of the *G* groups' *Y* means, that is, the usual AV *F* test of the between-groups mean square (*MS*):

$$F = \frac{\text{between } G \text{ groups } MS}{\text{within } G \text{ groups } MS} = \frac{(B\,SS)/(g-1)}{(W\,SS)/(n-g)}$$

$$= \frac{(12127.0)/(4-1)}{(15078.6)/(36-4)} = \frac{4042.33}{471.21} = 8.580, \quad [5]$$

for numerator (between *G* groups) $df = g - 1 = 4 - 1 = 3$, and denominator (within *G* groups, or error) $df = n - g = 36 - 4 = 32$.

These *F* ratios must be identical, since $B\,SS = (R^2_{Y \cdot 123\ldots k})$ (total *SS*), and $W\,SS = (1 - R^2_{Y \cdot 123\ldots k})$ (total *SS*). Formula 4 differs from Formula 5 only in that the total *SS* has been cancelled out from numerator and denominator.

The formulas help clarify the identity of the two procedures. We obtain another perspective on why $3(= g - 1)$ independent variables carry all the group membership information for $4(= g)$ groups, —there are only 3 *df* "associated with" *G* group membership. By either the MR or AV route, the total *SS* (or variance) of *Y* has been partitioned into a portion accounted for by *G* group membership (or by *G* group *Y* means), and a portion not so accounted for (i.e., within group, residual, or "error"), the latter, by either route, based on $n - g$ *df*.

Conceptually, the *F* ratios can be understood to be the same because they are testing null hypotheses which are mathematically equivalent, even though they are traditionally differently stated:

MR: H_0: Population $R^2_{Y \cdot 123} = 0$
AV: H_0: Population $m_1 = m_2 = m_3 = m_4 = m$

If the AV H_0 is true, then knowledge of group membership and the use of group means leads to the same least squares prediction of the *Y* value of a given case as no knowledge, namely, the grand mean, thus one can account for none of the variance in *Y* by such knowledge, hence $R^2_{Y \cdot 123} = 0$, and conversely.

A full MR analysis also yields the regression coefficients and constant for the regression equation:

$$\hat{Y} = B_1X_1 + B_2X_2 + \cdots + B_kX_k + A \quad [6]$$

where $\hat{Y}$ is the least-squares estimated ("predicted") value of *Y*, the B_1 are raw score partial regression coefficients attached to each X_i, and *A* is the regression constant or *Y*-intercept,

that is, the estimated value of Y when all X_i are set at zero. (Its computation is accomplished by including a "unit vector" with the X_i; see Draper & Smith, 1967.)

In any MR problem, a B_i coefficient gives the amount of the effect in Y expressed in Y units which is yielded by a unit increase in X_i. But since as dummy variables the X_i are coded 0 — 1, a unit increase means 1, membership in the group, rather than 0, nonmembership in the group. Solving for the values of the general regression Equation 6 for the artificial data, and using dummy variables, we obtain:

$$\hat{Y} = -30.34X_1 - .56X_2 + 21.22X_3 + 84.12$$

Since group membership is all-or-none, the B_i values give the *net* consequence of membership in G_i relative to G_4 for groups G_1, G_2, and G_3. Thus,

$$\hat{Y}_1 = \bar{Y}_1 = -30.34(1) - .56(0) + 21.22(0) + 84.12 = 53.78$$

$$\hat{Y}_2 = \bar{Y}_2 = -30.34(0) - .56(1) + 21.22(0) + 84.12 = 83.56$$

$$\hat{Y}_3 = \bar{Y}_3 = -30.34(0) - .56(0) + 21.22(1) + 84.12 = 105.34$$

And G_4 has not been slighted, since, substituting its scores on X_1, X_2, and X_3, we find:

$$\hat{Y}_4 = \bar{Y}_4 = -30.34(0) - .56(0) + 21.22(0) + 84.12 = 84.12.$$

Thus, one can understand that "B_4," the "missing" reference group's weight, is always zero, and that therefore $\hat{Y}_4 = A$. The exact values of the B_i will vary, depending on which group is taken as the reference group (i.e., is coded $0, \cdots, 0$), but the differences among the B_i's will always be the same, since they are the same as the differences between the group Y means. That is, whichever the reference group, the separation of the B_i's in the example will be the same as that among the values −30.34, −.56, +21.22 and 0. (For example, if G_1 is taken as the reference group, the new B_i are 0, 29.78, 51.56, and 30.33, and the regression constant $A = \bar{Y}_1 = 53.78$.)

Not only are the B_i meaningful, but also the multiple-partial correlations with the criterion, that is, the correlation of Y with X_i, partialing out or holding constant all the other independent variables, which for the sake of notational simplicity, we designate p_i. With dummy variable coded X_i, p_i can be more specifically interpreted as the correlation between Y and the dichotomy made up of membership in G_i versus membership in G_0, the reference group. The p_i thus give, in correlational terms, the relevance to Y of the distinction between each G_i and the reference group.

Furthermore, the p_i, B_i, and β_i (the standardized partial regression coefficient) can be tested for significance by means of t (or equivalently, F with numerator $df = 1$). Indeed, the null hypothesis is the same for all three,—the respective population parameter equals zero. But for a given X_i, if any one of the three is zero, all are zero, and the value of t is identical for all three tests. For the artificial data, the results are

	X_1	X_2	X_3
B_i	−30.34	−.56	+21.22
β_i	− .478	−.009	.334
p_i	− .464	−.010	.344
t_i	− 2.96	−.05	2.07

Thus, the G_1-G_4 distinction and also the G_3-G_4 distinction with regard to Y are significant (two tailed .01 and .05, with 32 df) while the G_2-G_4 is not. These are identically the results one would obtain for t tests between the respective Y means, using the within-group mean square (with 32 df) as the variance estimate.

The reader, having been shown the MR-AV identities, may nevertheless react, "O.K., that's interesting, but so what?" Other than the provision of correlational (or regression) values, no advantage of MR over AV is claimed for this problem. But if there were other independent variables of interest (main effects, either nominal, ordinal, or interval; interactions; covariates; nonlinear components; etc., whether or not correlated with G or each other), their addition to the G variable could proceed easily by means of MR, and not at all easily in an AV/ACV framework. This possibility is the single most important advantage of the MR procedure, and will receive further attention below.

To summarize, dummy variable coding of nominal scale data yields the multiple R^2 and

F test (proportion of variance accounted for by group membership and an overall significance test) and the group Y means, but also information on the degree of relevance to Y of membership in any given group, G_i, relative to the remainder (r_{Yi}), and to a reference group in terms of either regression weights (B_i or β_i) or correlation (p_i), as well as specific significance tests on the relevant null hypotheses. The importance of dummy variable (or other nominal scale) coding lies not so much in its use when only a single nominal scale constitutes the independent variables, but rather in its ready inclusion with other independent variables in MR.

Contrast Coding

Another system for representing nominal data can be thought of as contrast or "issues" coding. Here, each independent variable carries a contrast (in the AV/ACV sense) among group means. Each subject is characterized for each contrast according to the role he plays in it, which depends upon his group membership. With all contrasts so represented, the MR analysis can proceed.

As an example, reconsider the representation of the G variable. We can contrast membership neither G_1 or G_2 versus membership in either G_3 or G_4. This could be substantively interpreted as, for example, majority versus minority religions, schizophrenic versus neurotic, or drug versus no-drug treatment condition. The coding or scoring of this issue may be rendered as in Column 4 in Table 1: the value 1 is assigned the subjects in G_1 and G_2 and the value -1 to those in G_3 and G_4, as is done in the computation of orthogonal contrasts in AV (e.g., Edwards, 1960). Actually, any two different numbers can be used to render this issue by itself, but there are advantages for some purposes in using values which sum to zero. The simple correlation between the dependent variable and this X_1 is a point-biserial correlation (as were the dummy variable correlations) whose square gives directly the proportion of Y variance attributable to the G_1, G_2 versus G_3, G_4 distinction. For the artificial data, the $r^2_{Y1} = .2246$ ($r_{Y1} = -.4739$). This is a meaningful value which gives the size of the relationship in the sample. This r_{Y1} can be tested for significance, and confidence limits for it (or for r^2_{Y1}) can be computed by conventional procedures.

Other issues or contrasts can be rendered as independent variables. For example, a second issue which may be rendered is the effect on Y of the G_1 versus G_2 distinction, ignoring G_3 and G_4. A third issue may be the analogous G_3 versus G_4 distinction, ignoring G_1 and G_2. These are rendered, respectively, in Columns 5 and 6 in Table 1. Each yields an r and r^2 with the criterion which is interpretable, testable for significance, and confidence boundable.

Beyond the separate correlations of these three contrast variables, there is the further question of what their *combined* effect is on Y. We compute the $R^2_{Y.123}$ and F and obtain *exactly* the same values as when the arbitrary or dummy variable coding was used, .4458 and 8.580 (for the artificial data). This follows from the fact that the three independent variables satisfy the nonsingularity condition, that is, no one of them gives a multiple R with the other two of unity. This is a necessary *and sufficient* condition for *any* coding of $g - 1$ independent variables to represent G (see next section).

As before, the partial statistics, that is, the p_i, B_i and β_i and the common t test of their significance are also meaningful. If the independent variables all correlate zero with each other, the β_i will equal their respective r_{Yi}. That this must be the case can be seen from the fact that each r^2_{Yi} represents a *different* portion of the Y variance whose sum is the multiple $R^2_{Y.123}$ and thus the relationship $R^2_{Y.123} = \Sigma r_{Yi}\beta_i = \Sigma r^2_{Yi}$ must hold. The X_i as presented in Columns 4, 5, and 6 will be mutually uncorrelated if and only if the group sample sizes are equal. If they are not equal, the correlations among the X_i will be nonzero, which means that the contrasts or issues posed to the data are not independent. Such would be the case, in general, in the example if it were religion or diagnosis which formed the basis for group membership, and the actual natural population randomly sampled. Given unequal n_i for the four samples, although it is possible to make the three contrasts described above mutually uncorrelated, the coding of Columns 4, 5, and 6 does not do so. The scope of this article precludes discussion of the procedures whereby contrasts are coded so as to be uncorrelated. We note here merely that although it is

always possible to do so, it is not necessarily desirable (see below).

Since, in AV terms, the between-groups *SS* can be (orthogonally) partitioned in various ways, there are sets of contrasts other than the set above which can be represented in the coding. A particularly popular set is that automatically provided by the AV factorial design. If the four groups of this example are looked upon as occupying the cells of a 2×2 design (an interpretation to which the physiological example of drug versus no drug, frontal lesion versus control lesion particularly lends itself), each of the usual AV effects can be represented as X_i by the proper coding. The first is the same as before, and contrasts G_1 and G_2 with G_3 and G_4, for example, the drug–no-drug main effect, reproduced as Column 7 of Table 1. The second main effect, for example, frontal-control lesion, contrasts G_1 and G_3 with G_2 and G_4 and is given by the coding in Column 8. This latter X_2 gives r_{Y2}, the (point-biserial) r for (e.g.) site of lesion with the dependent variable (e.g.) retention, and r^2_{Y2} is the proportion of Y variance accounted for by this variable.

The remaining *df* is, as the AV has taught us, the interaction of the two main effects, for example, Drug–No-Drug $\times$ Frontal-Control Lesion. It can always be rendered as a multiplicative function of the two single *df* aspects of the main effects. Here, it is simply coded as the product of each group's "scores" on X_1 and X_2 (given as Column 9 in Table 1): $1 \times 1 = 1$, $1 \times -1 = -1$, $-1 \times 1 = -1$, and $-1 \times -1 = 1$. Rendering the interaction as X_3, one can interpret it as carrying the information of that aspect of group membership which represents the *joint* (note, *not* additive) effect of the drug and frontal lesion conditions. Its (point-biserial) r_{Y3} is an expression in correlational terms of the degree of relationship between Y and the *joint* operation of drug and lesion site. r^2_{Y3} gives the proportion of Y variance accounted for by this joint effect.

In the example, these three issues are *conceptually* independent, thus it would be desirable that the X_i be uncorrelated, that is, $r_{12} = r_{13} = r_{23} = 0$. The coding values given in Columns 7, 8, and 9 of Table 1 will satisfy this condition if (and only if) the sample sizes of the four cells are equal. (If not, other coding, not discussed here, would be necessary.)

The conceptual independence of the issues arises from the consideration that they are both manipulated variables. When this is the case, it is clearly desirable for them to be represented as mutually uncorrelated, since then the $\beta_{Yi} = r_{Yi}$ and the $R^2_{Y \cdot 123}$ is simply a sum of the separate r^2_{Yi}. Thus, the total variance of Y accounted for by group membership is unambiguously partitioned into the three separate sources. Further, the factorial AV F test values of each of the separate (one *df*) effects is *identical* with the t^2 of the analogous MR partial coefficients (β_i, B_i, or p_i).

However, whether one wishes to represent the issues as uncorrelated depends on whether they are conceptually independent and the differing n_i are a consequence of animals randomly dying or test tubes being randomly dropped on the one hand, or whether they carry valid sampling information about a natural population state of affairs. Assume Y is a measure of liberalism-conservatism and reconsider the problem with the groups reinterpreted as G_1: low education, low income ($n_1 = 160$), G_2: low education, high income ($n_2 = 20$), G_3: high education, low income ($n_3 = 80$), and G_4: high education, high income ($n_4 = 100$). These unequal and disproportional n_i carry valid sampling information about the univariate and bivariate distributions of education and income as defined here, the product moment r_{12} (phi) between them (coded as in Columns 7 and 8) equalling .4714. They may also be correlated with their interaction. One would ordinarily not wish to render these effects as uncorrelated, since the resulting X_i would be quite artificial, but rather by the coding given in Columns 7, 8, and 9, where, again, X_3 is simply the X_1X_2 product.

Note that whether the X_i are correlated or uncorrelated, or whether the n_i are equal or unequal, *all* of these coding systems yield the same $R^2_{Y \cdot 123}$ and associated F.

Two systems of rendering nominal scale (group membership) information into independent variables have been described: dummy variable coding and contrast coding. They result in identically the same multiple R^2 (and associated F) but different per independent variable partial statistics which are differently interpreted. Either involves expressing the nominal scale of g levels (groups) into $g - 1$

independent variables, each carrying a distinct aspect of group membership whose degree of association and statistical significance can be determined.

Nonsense Coding

It turns out, quite contraintuitively, that if one's purpose is merely to represent G so that its R^2_Y and/or its associated F test value can be determined, it hardly matters how one codes $X_1, X_2, \cdots, X_{g-1}$. *Any* real numbers, positive or negative, whole or fractional, can be used in the coding subject only to the nonsingularity constraint, that is, no X_i may have a multiple R of 1.00 with the other independent variables.

Consider, for example, the values of Columns 10–12 of Table 1. The numbers for X_1 in Column 10 were obtained by random entry into a random number table and their signs by coin flipping. Column 11 for X_2 was constructed by squaring the entries in Column 10, and Column 12 for X_3 by cubing them. Powering the X_1 values assures the satisfaction of the nonsingularity constraint. Now, using these nonsense "scores" to code G and the same Y values of the artificial example, we find the *same* $R^2_{Y.123}$ of .4458 with associated $F = 8.580$!

Or, alternatively, the coding values of Columns 13, 14, and 15 were obtained by haphazard free association with a quick eyeball check to assure nonsingularity. They, too, yield $R^2_{Y.123} = .4458$ and $F = 8.580$.

Why these, or any other values satisfying nonsingularity will "work" would require too much space to explain nontechnically. Ultimately, it is a generalization of the same principle which makes it possible to score a dichotomy with *any* two different values (not only the conventional 0 and 1) and obtain the same point-biserial r^2 against another variable.

Of course, the statistics per X_i, that is, r_{Yi}, p_i, B_i, β_i, are as nonsensical as the X_i. But the regression equation will yield the correct group means on Y, and, as noted, R^2 and its F remain invariant. Thus, with the aid of an MR computer program and a table of random numbers (or a nonsingular imagination), one can duplicate the yield of an AV.

Apart from its status as a statistical curiosity, of what value is the demonstration that one can simulate an AV by means of an arbitrarily coded MR analysis? Not much, taken by itself. However, despite this disclaimer, it should be pointed out that for most investigators, the yield sought from the AV of such data is the significance status of the F test on the means, which the MR provides; the latter also "naturally" yields, in R^2, a statement of proportion of variance accounted for. True, this is identically available from the AV in η^2, but this is not generally understood and computed. The MR approach has the virtue of calling to the attention of the investigator the existence of a rho (relationship) value and its distinction from a tau (significance test) value (Cohen, 1965, pp. 101–106), an issue usually lost sight of in AV contexts (but, see Hays, 1963, pp. 325–333).

But if it hardly matters how we score G and still get the same $R^2_{Y.123}$ and F ratio, we can score it in some meaningful way, one which provides analytically useful intermediate results, that is, by dummy variable or contrast coding. For other approaches to nominal scale coding, see Bottenberg and Ward (1963) and Jennings (1967).

Aspects of Quantitative Scales as Independent Variables

As noted in the introduction, psychologists are familiar with the use of quantitative variables as independent variables in MR. This, indeed, is the only use of MR illustrated in the standard textbooks. Thus, given duration of first psychiatric hospitalization as the dependent variable Y, and as independent variables: age (X_1), Hollingshead SES Index (X_2), and MMPI Schizophrenia (*Sc*) score (X_3), the psychologist knows how to proceed. But MR provides opportunities for the analysis of quantitative independent variables which transcend this very limited approach.

Curvilinear Regression

From the enlarged conceptual framework of the present treatment of MR, we would say that this analysis is concerned with the *linear* aspects of age, SES, and *Sc*. There are other functions or aspects of these variables which can be represented as independent variables.

It has long been recognized that curvilinear relationships can be represented in linear MR by means of a polynomial form in powered terms. The standard Equation 6

$$\hat{Y} = B_1X_1 + B_2X_2 + \cdots B_kX_k + A$$

is linear in the X_i. If the X_i are $X_1 = Z$, $X_2 = Z^2$, $X_3 = Z^3, \cdots X_k = Z^k$, the equation is *still* linear in the X_i, even though not linear in Z. The result of this strategem is that nonlinear regression of Y on Z can nevertheless be represented within the linear *multiple* regression framework, the "multiplicity" being used to represent various aspects of nonlinearity, the quadratic, cubic, etc. The provision of any given power u of Z, that is, Z^u allows for $u - 1$ bends in the regression curve of Y or Z. Thus Z^1 or Z provides for $1 - 1 = 0$ bends, hence a straight line, Z^2 provides for $2 - 1 = 1$ bend, Z^3 for 2 bends, etc. In most psychological research, provision for more than one or two bends will rarely be necessary.

It is the same strategem of polynomial representation further refined to make these aspects orthogonal to each other, which is utilized in the AV, also a linear model, in trend analysis designs.

A note of caution must be injected here. Such variables as Z, Z^2, and Z^3 are in general correlated, indeed, for score-like data, usually highly so. Table 2 presents some illustrative data. In this example, the correlations are .9479, .8840, and .9846. For reasons of ordinary scientific parsimony, unless one is working with a strong hypothesis, we normally think of them as a hierarchy: how much Y variance does Z account for? (.5834) If Z^2 is added to Z as a second variable, how much do both together account for? (.5949) The difference represents the increment in variance accounted for by making allowance for quadratic (parabolic) curvature. In the example, it is a very small amount,—.0115. If to Z and Z^2 we add Z^3, the multiple $R^2_{Y.123}$ becomes .5956, an increment over $R^2_{Y.12}$ of only .0007. Each of these separate increments, or the two combined can be tested

TABLE 2

ILLUSTRATIVE DATA ON POLYNOMIAL MULTIPLE REGRESSION

Variable	Correlations (r)			Cumulative R^2	Increment	p_i
	Y	Z	Z^2			
Z (= X_1)	.7638			.5834	.5834	.1399
Z^2 (= X_2)	.7582	.9479		.5949	.0115	−.0116
Z^3 (= X_3)	.7268	.8840	.9846	.5956	.0007	.0419

for significance. In general, *any* increment to an $R^2_{Y.A}$ due to the addition of B can be tested by the F ratio:

$$F = \frac{(R^2_{Y.A,B} - R^2_{Y.A})/b}{(1 - R^2_{Y.A,B})/(n - a - b - 1)} \quad [7]$$

with $df = b$ and $(n - a - b - 1)$, where

$R^2_{Y.A,B}$ is the incremented R^2 based on $a + b$ independent variables, that is, predicted from the combined sets of A and B variables,

$R^2_{Y.A}$ is the smaller R^2 based on only a independent variables, that is, predicted from only the A set,

a and b are the number of original (a) and added (b) independent variables, hence the number of df each "takes up."

This F test of an increment to R^2 is much more general in its applicability than the present narrow context, and its symbols have been accordingly given quite general interpretation. It is used several times later in the exposition, in other circumstances where, because of correlation among X_i, it provides a basis for judging how much a set of independent variables contributes *additionally* to Y variance accounting. Since what is added is independent of what is already provided for, this is a general device for partitioning R^2 into orthogonal portions. Since the size of such portions depends on the *order* in which sets are included, the hierarchy of sets is an important part of the investigator's hypothesis statement. The generality of Formula 7 is further seen in that Formula 4 is actually a special case of Formula 7, where $R^2_{Y.A}$ is zero because no X_i are used (hence $a = 0$) and $R^2_{Y.B}$ is the R^2 based on b $(= k)$ df which is being tested, that is, an increment of R^2 from zero.

Either set may have one or more independent variables. Thus, to test the increment of Z^2 to Z alone, assuming total $n = 36$,

$$F = \frac{(.5949 - .5384)/1}{(1 - .5949)/(36 - 1 - 1 - 1)} = \frac{.0115}{.4051/33} = .934$$

with $df = 1$ and 33 (a chance departure).
To test the pooled addition of both Z^2 and Z^3 to

Z,

$$F = \frac{(.5956 - .5834)/2}{(1 - .5956)/(36 - 1 - 2 - 1)}$$

$$= \frac{.0122/2}{.4044/32} = .483$$

with $df = 2$ and 32 (also a chance result).

The need for caution arises in that if one studies the results of the regression analysis which uses Z, Z^2 and Z^3, where the solution of the partial (regression or correlation) coefficients is simultaneous, not successive, the three variables are treated quite democratically. Each is partialed from the others without favor or hierarchy. Since such variables are highly correlated, when one partials Z^2 and Z^3 from Z, one is robbing Z of Y variance which we think of as rightfully belonging to it. Table 2 gives the p_i of the three predictors when one treats them as a set. The values are smaller (reflecting the mutual partialing), and may be negative (reflecting "suppression" effects). Because the p_i are so small, they may well be nonsignificant (as they are here), even though r_{YZ} is significant and any of the other variables may yield a significant increment. Thus, the significance interpretation of the regression of a set of polynomial terms simultaneously may be quite misleading when the usual hierarchical notions prevail.

On the other hand, if the analyst's purpose to portray a polynomial regression fit to an observed set of data, he can solve for the set simultaneously and use the resulting MR equation. For the data used for Table 2, the regression Equation 6 is:

$$Y = 11.70X_1 - .50X_2 + .25X_3 + 55.90$$

the values being the B_i regression coefficients and constant, and the X_i successively Z, Z^2, and Z^3. One can substitute over the range of interest of Z and obtain fitted values of Y for purposes of prediction or of graphing of the function.

There are other means whereby curvilinear relationships can be handled in an MR framework. Briefly, one can organize an independent variable Z into g class intervals (ordinarily, but not necessarily equal in range) and treat the resulting classes as groups, coding them by the dummy variable technique described above. This results in $g - 1$ independent variables, each a segment of the Z range. The resulting $R^2_{Y \cdot G}$ is the amount of Y variance accounted for by Z (curvilinearly, if such is the case) and the Y means for the g intervals, computable from the resulting raw score regression equation, can be plotted graphically against the midpoints of the class intervals of Z to portray the function.

A more elegant method is the transformation (coding) of the Z values to *orthogonal* polynomials. This has the advantages in that the resulting X_i terms representing linear, quadratic, cubic, etc., components of the polynomial regression are uncorrelated with each other; thus each contributes a separate portion of the Y variance capable of being tested for significance. Unfortunately, this method becomes computationally quite cumbersome unless the Z values are equally spaced and with equal n_i per interval. The latter is the usual case when Z is an experimentally manipulated variable, where the standard trend analysis designs of the AV can be used (Edwards, 1960).

Finally, although the first few powers of a polynomial is a good *general* fitting function, in some circumstances, such transformations of Z as $\log Z$, $1/Z$, or $Z^{\frac{1}{2}}$ may provide a better fit. Draper and Smith (1967) provide a useful general reference for handling curvilinearity (and other MR problems).

Joint Aspects of Interactions

Given two independent variables, $X_1 = Z$ and $X_2 = W$, one may be interested in not only their separate effects on Y, but also on their joint effect, over and above their separate effects. As noted above (Contrast Coding), where this was discussed in the narrow context of a 2×2 design, this joint effect is carried by a third independent variable, a score defined for each subject by the product of his Z and W scores, that is, $X_3 = ZW$. This variable contains this joint effect, which is identically the (first-order) interaction effect of AV, or the "moderator" effect of Saunders (1956). This identity is quite general, so that a triple interaction is carried by a triple product, say ZWV, etc. Furthermore, the above are all interactions or joint effects of *linear* aspects of the variables. The more complex interactions of nonlinear aspects, such as the linear by quad-

ratic, or quadratic by cubic, made familiar by advanced treatments of AV trend analysis (Winer, 1962, pp. 273–278), would be represented by products of powered variables, for example, ZW^2, Z^2W^3, each a single independent variable.

The presentation of joint effects as simple products in MR requires the same caution as in the polynomial representation of a single variable. (Indeed, a powered variable can be properly understood as a special case of an interaction, for example, Z^2 contains the Z by Z interaction.) If one uses simultaneously as independent variables $X_1 = Z$, $X_2 = W$, $X_3 = ZW$, the correlations of Z with ZW, and W with ZW will ordinarily not be zero, may indeed be large, and the partial coefficients for Z and W (β, B, p) will have lost to ZW some Y variance which properly is theirs (just as Z would be robbed of some of its Y variance by Z^2 and Z^3). The problem is solved as in the polynomial regression analysis: Find $R^2_{Y\cdot123}$, the variance proportion accounted for by all three variables; then find $R^2_{Y\cdot12}$, the amount accounted for *without* the interaction. The increment is tested for significance by the F ratio of Formula 7.

This, too, generalizes. In more complex systems, involving either more variables and higher order interactions or interactions among polynomial aspects (or both), one forms a hierarchy of sets of independent variables and tests for the significance of increments to R^2 by means of the same F ratio (Formula 7). For example, if one has three variables Z, W, and V, represented both linearly and quadratically with all their interactions, one possible way of organizing the variables is by means of the following sets:

A: Z, W, V

B: ZW, ZV, WV

C: ZWV

D: Z^2, W^2, V^2

E: Z^2W, ZW^2, Z^2V, ZV^2, W^2V, WV^2

F: Z^2W^2, Z^2V^2, W^2V^2

G: $Z^2W^2V^2$

One would then test $R^2_{Y\cdot AB} - R^2_{Y\cdot A}$, $R^2_{Y\cdot ABC} - R^2_{Y\cdot AB}$, etc., each by the F ratio for increments. When a set containing more than one variable is significant, one can "break out" each variable in it and test its increment for significance by the same procedure. Of course, one can elect to make all sets contain only one variable, but the number of resulting tests (in the example there would be 20) brings with it an increased risk of spuriously significant results over the complete analysis. This strategy parallels that of the AV, where the avoidance of this risk is implicit. In a 4×5 factorial design AV, for example, the interaction involves a single mean square based on $3 \times 4 = 12$ df which is tested by a single F test. One ordinarily does not test each of these 12 effects separately unless the set as a whole is significant. The principle, of course, obtains even for the main effects, involving sets of 3 and 4 df, where each set normally is tested "wholesale."

Other combinations and priorities of the X, Y, and Z variables are, of course, possible. This operation involves formulating hypotheses about what constitutes a relevant class of independent variables and the priorities of these classes. It depends not only on mechanical variance-stealing considerations, but also on substantive issues in the research and the judgment of the investigator.

Although the discussion in this section has been concerned with interactions among quantitative variables, the principles of forming interaction variables hold also for nominal variables, and for mixtures of variables. Let an "aspect" of a research variable such as religion or IQ be one of the X_i of the set which represent it. Then, for example, if the interaction of u aspects of one variable U and v aspects of another variable V are desired, one may form a total of uv interaction X_i, by multiplying each of the u aspects by each of the v aspects. Each of the resulting uv independent variables is a single (one df) variable which represents a specific aspect by aspect joint or interaction effect. Either U or V may be nominal or quantitative. Where nominal, their aspects may be dummy variables or contrasts; where quantitative, the aspects may be powered polynomial terms or missing data dichotomies (see below). One can thus generate such single interaction X_i as "majority-minority religious group by authoritarianism," "experimental group D versus control group by

quadratic of stimulus intensity," etc. It is both convenient and enlightening to have each such joint aspect separately and unambiguously (but not necessarily orthogonally) represented in the set of independent variables. Their individual increment to R^2 and significance can then be determined.

Perhaps as important as being able to represent the interaction X_i in specific detail is the availability of the option *not* to represent some or all of them. The textbook paradigms for factorial design AV lead data analysts to dutifully harvest all possible interactions of all possible orders up to the highest, whether or not they are meaningful or interpretable or, if interpretable, communicable. There emanate from psychology departments many silent prayers to the spirit of R. A. Fisher that high-order interactions will not prove significant! Obviously, one need not (indeed cannot) analyze for all possible aspects including joint aspects of variables if for no other reason than the rapid loss of *df* for estimating error. The need to "specify the model," that is, the set of X_i to be studied in MR has the salutary effect of requiring an incisive prior conceptual analysis of the research problem. This goes hand in hand with the flexibility of the MR system, which makes readily possible the representation of the research issues posed by the investigator (i.e., multiple regression in the service of the ego!), rather than the canned issues mandated by AV computational routines.

Missing Data

In nonexperimental, particularly survey, research, it frequently occurs that some subjects are missing data on one or more (but not all) of the independent variables under study. Typically, the data are not missing randomly, but for reasons frequently related to values for other independent variables, and particularly to values for the dependent variable under study. For example, in a study of factors associated with the rehabilitation of drug addicts, reported weekly wages on last job is used as an independent variable, among others. Some respondents claim they do not recall or refuse to respond. As another example, consider a retrospective study of the school records of adult mental retardates where the recorded IQ is abstracted for use as an independent variable but found missing in some cases. In neither of these cases can one prudently assume that the mean of these cases on the X_i in question, other X_i, and, particularly, Y is the same as that for the cases with data present. The practice of excluding cases lacking some of the data has the undesirable properties of analyzing a residual sample which is unrepresentative to an unknown degree of the population originally sampled, as well as the loss of information (viz., the *fact* of data being missing) which may be criterion relevant.

MR provides a simple method for coping with this problem. Each such variable has *two* aspects, its value (where present) and whether or not the value *is* present. Accordingly, two independent variables are constructed: X_1 is the value itself, with the mean of X_1 for those cases where it is present entered for the cases where it is missing, and X_2 is the missing data aspect, a dummy variable dichotomy coded 0–1 for absent-present. These two aspects contain all the information available in the variable. Moreover, as scored, $r_{12} = 0$, hence X_1 and X_2 are each contributing an independent portion of the Y variance.

Actually, *any* value entered for the missing data in X_1 will "work" in the sense of accounting for Y variance, that is, the $R^2_{Y.12}$ will be the same. The use of the mean will uniquely result in $r_{12} = 0$, which may be advantageous interpretively. For some purposes, this advantage may be offset by using some (or any) other value, obviating the necessity of a prior computation of the mean.

The researcher, normally sensitive about tampering with data, may find the prospect of "plugging" empty spaces in his data sheet with means singularly unappealing. He may even correctly point out that this will have the effect of reducing r_{Y1} from what r_{YX} is for the subsample having X values present. In rebuttal, it must be pointed out that the subsample is not representative of the originally defined population, and the method proposed can be thought of as reflecting the fact that the population studied contains missing data, and fully incorporates this fact as positive information.

Analysis of Covariance

Viewed from the perspective of the MR system, the fixed-model ACV turns out to be a

rather minor wrinkle, and not the imposing parallel edifice it constitutes in the AV/ACV framework. A covariate is, after all, nothing but an independent variable, which, because of the logic dictated by the substantive issues of the research, assumes priority among the set of independent variables as a basis for accounting for Y variance. Consider a research in educational psychology in which the Y variable is some performance measure in children, X_1 is midparental education, X_2 is family income, and G, carried by the set X_3, X_4, X_5 represents some differential learning experience in four intact classes. This situation is a "natural" for ACV (assuming its assumptions are reasonably well met). One would think of it as studying the effect of learning experience or class membership on Y, using X_1 and X_2 as covariates. Thus considered, we are asking how much variance in Y (and its significance) the variables X_3, X_4, and X_5 account for, *after* the variance due to X_1 and X_2 is allowed for, or held constant, or "partialed out" (the terms being equivalent). The form of the MR analysis to accomplish this purpose is directly suggested. Find $R^2_{Y \cdot 12345}$, the proportion of Y variance all independent variables account for. Then find $R^2_{Y \cdot 12}$, the proportion of Y variance attributable to the covariates education and income. Their difference is the increment due to group membership, which is tested for significance by the F test of Formula 7 used in a different design context above. Note that no problem arises if the four groups are defined by a 2×2 factorial. If X_3, X_4, X_5 are coded as in Columns 7, 8, 9 in Table 1 to represent the two main effects and their interaction, the respective ACV significance tests are performed by (Formula 7) F ratio tests of the increments $R^2_{Y \cdot 12345} - R^2_{Y \cdot 1245}$ (for the main effect represented by X_3), $R^2_{Y \cdot 12345} - R^2_{Y \cdot 1235}$ (for the main effect represented by X_4) and $r^2_{Y \cdot 12345} - r^2_{Y \cdot 1234}$ (for the interaction or joint effect). Note that X_1 and X_2 are always included in the debited R^2, because of their priority in the issues as defined. This principle is readily generalized to designs of greater complexity.

That a covariate is nothing but another independent variable except for priority due to substantive considerations is evident when one considers a study formally almost identical to the above, now, however, done by a social psychologist. Since there are four different classes and four different teachers, the classes ipso facto have had different learning experiences. But this research is concerned with the effects of parental education and income on the performance criterion, with group membership now the contaminant which must be removed, hence the covariate. Using the same set up and data, he would find $R^2_{Y \cdot 12345} - R^2_{Y \cdot 345}$ as the combined effect of education and income, $R^2_{Y \cdot 12345} - R^2_{Y \cdot 2345}$ as the net effect of education (i.e., over and above that of income as well as the covariates of class membership), and $R^2_{Y \cdot 12345} - R^2_{Y \cdot 1345}$ for the net effect of income, each F-testable as before. Thus, one man's main effect is another man's covariate.

The MR approach to ACV-like problems opens up possibilities for statistical control not dreamed of in ACV. We have just seen how purely nominal or qualitative variables (class membership) can serve as covariates. Beyond this, we can apply other principles which have been adduced above: (*a*) Any aspects of data can, by appropriate means, be represented as independent variables. (*b*) Any (sets of) independent variables can serve as covariates by priority assignment in variance accounting. Thus, for example, one can make provision for a covariate being nonlinearly related to Y (and/or to other independent variables) by writing a polynomial set of independent variables and giving the set priority; or, one can carry two variables *and* their interaction as a covariate set; or, one can even carry as a covariate a variable for which there are missing data by representing the two aspects of such a datum as two independent variables and giving them priority. Finally, one can combine the priority principle with those of contrast coding to achieve analytic modes of high fidelity to substantive research aims.

The ACV assumption that the regression lines (more generally, surfaces) of the covariate (U) on Y have the same slopes (more generally, regression parameters) between groups (V) is equivalent to the hypothesis of no significance for the set of uv interaction independent variables. This hypothesis can be F-tested as a Set B following the inclusion of U as Set A, using Formula 7.

Discussion

In the introduction it was argued that MR and AV/ACV are essentially identical systems, and so they are, at least in their theory. In the actual practice of the data-analytic art, many differences emerge, differences which generally favor the MR system as outlined above.

Before turning to these differences, a closer look at their similarity in regard to statistical assumptions is warranted. This article has concerned itself only with the fixed-model AV/ACV, wherein it is assumed that inference to the population about the independent variables is for just those variables represented (and not those variables considered as samples) and that values on these variables are measured without error. This means that in a MR whose set of X_i include quantitative variates (e.g., scores), the population to which one generalizes, strictly speaking, is made up of cases having just those X_i values, only the Y values for any given combinations of values for the X_i varying; moreover, the Y distribution (and only this distribution) is assumed normal and of equal variance for all the observed combinations of X_i values. These seem, indeed, to be a constraining set of assumptions. However, the practical effect on the validity of the generalizations which one might wish to draw is likely to be vanishingly small. It seems likely that the substantive generalizations made strictly for the particular vectors of X_i values in the rows of the basic data matrix of the sample would hold for the slightly differing values which the population would contain if the sampling is random. As for the normality and variance homogeneity assumptions for Y, the robustness of the F test under conditions of such assumption failure is well attested to (for a summary, see Cohen, 1965, pp. 114–116). Particularly when reasonably large samples are used, itself desirable to assure adequate statistical power, no special inhibition need surround the drawing of inferences from the usual hypothesis testing, certainly no more so than in AV.

A discussion of the practical differences between MR and AV is best begun with a consideration of the nature of classical fixed-model AV. Its natural use is in the analysis of data generated by experimental manipulations along one or more dimensions (main effects), resulting in subgroups of observations in multifactor cells, treatment combinations. Each main effect is paradigmatically a set of qualitative distinctions along some dimension. These dimensions are conceptually independent of each other, and since they are under the control of the designer of the experiment, the data can, in principle, be gathered in such a way that the dimensions are actually mutually orthogonal in their representation in the data. (This condition is met by the proportionality of cell frequencies in all two dimensional subtables.) This also results in interactions being orthogonal to each other and to the main effects. Thus, the paradigm is of a set of batches (one batch per AV main effect or interaction) of qualitative independent variables, all batches mutually orthogonal.

Now, under such conditions, one *can*, as illustrated above, analyze the data by MR, but there is no advantage in so doing. The AV can be seen as a computational shortcut to an analysis by the linear model which analyzes by batches and capitalizes on the fact that batches are orthogonal. Thus, the classical fixed factorial AV is a special simplified case of MR analysis particularly suited to neat experimental layouts, where qualitative treatments are manipulated in appropriate orthogonal relationships. Later refinements allows for quantitative independent variables being exploited by trend analysis designs, but these, too, demand manipulative control in the form of equally spaced intervals in the dimension and equal sized samples per level if the computational simplicity is to be retained.

These designs are quite attractive, not only in their efficiency and relative computational simplicity, but also in the conceptual power they introduced to the data analyst, for example, interactions, trend components. They were presented in excellent applied statistics textbooks. Inevitably, they attracted investigators working in quite different modes, who proceeded to a Procrustean imposition of such designs on their research.

A simple example (not too much of a caricature) may help illustrate the point. Dr. Doe is investigating the effects of Authoritarianism (California F scale) and IQ on a cognitive style score (Y), using high school students as subjects. He is particularly interested in the F × IQ interaction, that is, in the possibility

that r_{YF} differs as a function of IQ level. He gives the three tests, and proceeds to set up the data for analysis. He dichotomizes the F and IQ distributions as closely as possible to their medians into high-low groups and proceeds to assign the Y scores into the four cells of the resulting 2×2 fixed factorial design. He then discovers that the number of cases in the high-low and low-high cells distinctly exceed those in the other two, an expression of the fact that F and IQ are correlated. He must somehow cope with this disproportionality (nonorthogonality). He may (*a*) throw out cases randomly to achieve proportionality or equality; (*b*) use an "unweighted means" or other approximate solution (Snedecor, 1956, pp. 385–387); or (*c*) "fit constants by least squares" (Snedecor, 1956, pp. 388–391; Winer, 1962, pp. 224–227), which is, incidentally, an MR procedure.

Clearly, this is a far cry from experimentally manipulated qualitative variables. These are, in fact, naturally varying correlated quantitative variables. This analysis does violence to the problem in one or both of the following ways:

1. By reducing the F scale and IQ to dichotomies, it has taken reliable variables which provide graduated distinctions between subjects over a wide range, and reduced them to two-point (high-low) scales, squandering much information in the process. For example, assuming bivariate normality, when a variable is so dichotomized, there is a reduction in r^2_{YX}, the criterion variance it accounts for, and hence in the value of F in the test of its significance, of 36%. This wilful degradation of available measurement information has a direct consequence in the loss of statistical power (Cohen, 1965, pp. 95–101, 118).

2. The throwing out of cases to achieve proportionality clearly reduces power, but, even worse, distorts the situation by analyzing as if IQ and F scale score were independent, when they are not. Other approximations suffer from these and/or other statistical deficiencies or distortions.

If Dr. Doe uses the MR-equivalent exact-fitting constants procedure, he has still given up computational simplicity, and, of course, the measurement information due to dichotomization. If he seeks to reduce the latter and also allow for the possibility of nonlinearity of Y on X_i regressions by breakdown of IQ and/or F scale into smaller segments, say quartiles, his needs for equality of intervals and cases will be frustrated, and he will not be able to find a computational paradigm, which, in any case, would be very complicated. It seems quite clear that, however considered, the conventional AV mode is the wrong way to analyze the data.

On the other hand, the data can be completely, powerfully, and relevantly analyzed by MR. A simple analysis would involve setting $X_1 =$ IQ, $X_2 =$ F scale score, $X_3 =$ (IQ)(F). By finding $R^2_{Y.123} - R^2_{Y.12}$ and testing it for significance (or equivalently, by testing the significance of p_3), he learns how much the interaction contributes to Y variance accounting and its significance. Determinations of the values of r^2_{Y1}, r^2_{Y2}, $R^2_{Y.12}$, $R^2_{Y.12} - r^2_{Y1}$, and $R^2_{Y.12} - r^2_{Y2}$ and testing each for significance fully exploits the information in the data at this level. If he believes it warranted, he can add polynomial terms for IQ and F score and their interaction in order to provide for nonlinearity of any of the relationships involved.

Another practical difference between MR and AC/ACV is with regard to computation. The MR procedure, in general, requires the computation and inversion of the matrix of correlations (or sums of squares and products) among the independent variables, a considerable amount of computation for even a modest number of independent variables. It is true that *classical* AV, whose main effects, interactions, polynomial trend components, etc., are mutually orthogonal, capitalizes on this orthogonality to substantially reduce the computation required. Whatever computational reduction there is in AV or MR depends directly on the orthogonality of the independent variables, which we have seen is restricted to manipulative experiments, and is by no means an invariant feature even of such experiments.

However, given the widespread availability of electronic computer facilities, the issue of the *amount* of computation required in the analysis of data from psychological research dwindles to the vanishing point, and is replaced by problems of programming. The typical statistical user of a typical computer facility

requires that a computer program which will analyze his data be available in the program library. Such programs will have been either prepared or adapted for the particular computer configuration of that facility. Unfortunately, it is frequently the case that the available AV program or programs will not analyze the particular fixed AV design which the investigator brings. Some AV programs are wanting in capacity in number of factors or levels per factor, some will handle only orthogonal designs, some will handle only equal cases per cell, some will do AV but not ACV, some of those that do handle ACV can handle only one or two covariates. Many will not handle special forms of AV, for example, Latin squares.

On the other hand, even the most poorly programmed scientific computer facility will have at least one good MR program, if for no other reason than its wide use in various technologies, particularly engineering. All the standard statistical program packages contain at least one MR program. Although these vary in convenience, efficiency, and degree of informativeness of output, all of them can be used to accomplish the analyses discussed in this article. In contrast to the constraints of AV programs, the very general MR program can be particularized for any given design by representing (coding) those aspects of the independent variables of interest to the investigator according to the principles which have been described.

A note of caution: as we have seen, given even a few factors (main effects of nominal variables or linear aspects of quantitative variables), one can generate very large numbers of distinct independent variables (interactions of any order, polynomials, interactions of polynomials, etc.). The temptation to represent many such features of the data in an analysis must be resisted for sound research-philosophical and statistical reasons. Even in researches using a relatively large number of subjects (n), a small number of factors (nominal and quantitative scales) can generate a number of independent variables which exceed n. Each esoteric issue posed to the data costs a df which is lost from the error estimate, thus enfeebling the statistical power of the analysis.

This, ultimately, is the reason that it is desirable in research that is to lead to *conclusions* to state hypotheses which are relatively few in number. This formulation is not intended to indict exploratory studies, which may be invaluable, but by definition, such studies do not result in conclusions, but in hypotheses, which then need to be tested (or, depending on the research context, cross-validated). If one analyzes the data of a research involving 100 subjects by means of MR, and utilizes 40 independent variables, what does one conclude about the 4 or 5 of them which prove to have partial regression weights "significant" at the .05 level? Certainly not that *all* of them are real effects, when one realizes that an overall null hypothesis leads to an expectation that 5% of 40, or 2 are expected by change. But which two?

A reasonable strategy depends upon organizing a hierarchy of sets of independent variables, ordered, by sets, according to a priori judgments. Set A represents the independent variables which the investigator most expects to be relevant to Y (perhaps all or some of the main effects and/or linear aspects of continuous variables). These may be thought of as *the* hypotheses of the research, and the fewer the better. Set B consists of next order possibilities (perhaps lower order interactions and/or some quadratic aspects). These are variables which are to be viewed less as hypotheses and more as exploratory issues. If there is a Set C (perhaps some higher order interactions and/or higher degree polynomials), it should be thought of as unqualifiedly exploratory. (If there are covariates in the design, they, of course, take precedence over all these sets, and would enter first.) The "perhaps" in the parenthetical phrases in this paragraph are included because it is *not* a mechanical ordering that is intended. In any given research, a central issue may be carried by an interaction or polynomial aspect while some main effect may be quite secondary. In most research, however, it is the simplest aspects of factors which are most likely to occupy the focus of the investigator's attention. However, the decision as to what constitutes an appropriate set depends on both research-strategic issues that go to the heart of the substantive nature of the research, and subtle statistical issues beyond the scope of

this article. The latter are discussed by Miller (1966, pp. 30–35).

The independent variables so organized, one first does an MR analysis for Set A, then Sets A + B, then Sets A + B + C. Each additional set is tested for the increment to R^2 by means of the F test of Formula 7. A prudent procedure would then be to test for significance the contribution of any *single* independent variable in a set only if the set yields a significant increment to R^2. A riskier procedure would be to dispense with the latter condition, but then the results would clearly require cross-validation.

REFERENCES

BOTTENBERG, R. A., & WARD, J. H., JR. *Applied multiple linear regression.* (PRL-TDR-63-6) Lackland AF Base, Texas, 1963.

CATTELL, R. B. Psychological theory and scientific method. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology.* Chicago: Rand McNally, 1966.

COHEN, J. Some statistical issues in psychological research. In B. B. Wolman (Ed.), *Handbook of clinical psychology.* New York: McGraw-Hill, 1965.

CURETON, E. E. On correlation coefficients. *Psychometrika,* 1966, **31,** 605–607.

DRAPER, N., & SMITH, H. *Applied regression analysis.* New York: Wiley, 1967.

EDWARDS, A. E. *Experimental design in psychological research.* (Rev. ed.) New York: Rinehart, 1960.

HAYS, W. L. *Statistics for psychologists.* New York: Holt, Rinehart & Winston, 1963.

JENNINGS, E. Fixed effects analysis of variance by regression analysis. *Multivariate Behavioral Research,* 1967, **2,** 95–108.

LI, J. C. R. *Statistical inference.* Vol. 2. *The multiple regression and its ramifications.* Ann Arbor, Mich.: Edwards Bros., 1964.

MCNEMAR, Q. *Psychological statistics.* (3rd ed.) New York: Wiley, 1962.

MILLER, R. G., JR. *Simultaneous statistical inference.* New York: McGraw-Hill, 1966.

PETERS, C. C., & VAN VOORHIS, W. R. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill, 1940.

SAUNDERS, D. R. Moderator variables in prediction. *Educational and Psychological Measurement,* 1956, **16,** 209–222.

SNEDECOR, G. W. *Statistical methods.* (5th ed.) Ames: Iowa State College Press, 1956.

SUITS, D. B. Use of dummy variables in regression equations. *Journal of the American Statistical Association,* 1957, **52,** 548–551.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received November 13, 1967)