DOCUMENT RESUME

ED 367 687 TM 021 143

AUTHOR Thompson, Bruce
TITLE Why Multivariate Methods Are Usually Vital in Research: Some Basic Concepts.
PUB DATE Feb 94
NOTE 35p.; Paper presented at the Biennial Meeting of the Southwestern Society for Research in Human Development (Austin, TX, February 1994).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Analysis of Covariance; *Behavioral Science Research; Causal Models; Correlation; Educational Research; Error of Measurement; Error Patterns; Inferences; *Measurement Techniques; *Multivariate Analysis; *Research Methodology; *Statistical Significance; Statistical Studies
IDENTIFIERS *Research Replication; *Type I Errors

ABSTRACT

The present paper suggests that multivariate methods ought to be used more frequently in behavioral research and explores the potential consequences of failing to use multivariate methods when these methods are appropriate. The paper explores in detail two reasons why multivariate methods are usually vital. The first is that they limit the inflation of Type I error rates, and the second is that multivariate methods best honor the reality to which the researcher is purportedly trying to generalize. Three general analytic premises provide a framework for the discussion: (1) all statistical analyses are correlational; (2) the OVA (analysis of variance and covariance) analyses do not in and of themselves yield the capacity to make causal inferences; and (3) statistical significance tests do not evaluate the probability (p) that results will replicate. Eleven tables illustrate some analyses. (Contains 51 references.) (Author/SLD)

swsrhd.wp1 2/15/94

Why Multivariate Methods Are Usually Vital in Research: Some Basic Concepts

Bruce Thompson

Texas A&M University 77843-4225
and
Baylor College of Medicine

Paper presented as a Featured Speaker at the biennial meeting of the Southwestern Society for Research in Human Development, Austin, TX, February 25, 1994.

ABSTRACT

The present paper suggests that multivariate methods ought to be used more frequently in behavioral research, and explores the potential consequences of failing to use multivariate methods when these methods are appropriate. The paper explores in detail two reasons why multivariate methods are usually vital. Three general analytic premises provide a framework for the discussion.

Hinkle, Wiersma and Jurs (1979, p. 415) noted that "it is becoming increasingly important for behavioral scientists to understand multivariate procedures even if they do not use them in their own research." And recent empirical studies of research practice do confirm that multivariate methods are employed with some regularity in behavioral research (Elmore & Woehlke, 1988). However, empirical studies of practice indicate that univariate analyses, and particularly ANOVAs, still remain the dominant analytic choices in published research (Edgington, 1964, 1974; Elmore & Woehlke, 1988; Goodwin & Goodwin, 1985; Willson, 1980).

The purpose of the present paper is to suggest that multivariate methods ought to be used more frequently in behavioral research, and to explore the potential consequences of failing to use multivariate methods when these methods are appropriate. The paper explores in detail two reasons why multivariate methods are usually vital.

But, before these two issues are considered, it may be helpful to review three general premises regarding statistical analyses. The implications of these premises for multivariate analyses will then briefly be considered.

Three Premises and Their Implications for Multivariate Analyses

Premise #1: *All* Statistical Analyses are *Correlational*

All classical parametric methods are least squares procedures that implicitly or explicitly (a) use least squares weights to optimize explained variance and minimize model error variance, (b) focus on latent synthetic variables, and (c) yield variance-accounted-for effect sizes analogous to r^2. Put more parsimoniously, all classical analytic methods are correlational (Knapp, 1978; Thompson, 1988a). Designs may be experimental or correlational, but *all analyses are correlational*. In fact, all parametric analyses can be represented as bivariate analyses invoking Pearson product-moment correlation coefficients (Thompson, 1984, 1991).

The data (n=25) in Table 2, involving variables Y, X_1 and X_2 from Edwards (1985, p. 41), can be used as a heuristic to partially illustrate (a) the importance of weights in all analyses, (b) the obsession in all analyses with latent or synthetic variables, as against observed variables, and (c) that all analyses yield variance-accounted-for effect sizes analogous to r^2. The tabled data were analyzed using the SPSS program reported in Appendix A.

INSERT TABLE 1 ABOUT HERE.

When variables X_1 and X_2 are used to predict Y, the SPSS analysis yields the results presented in Table 2, reported in exactly the format output by SPSS. The Table 2 results indicate that the variance-accounted-for effect size for this analysis, R^2, was 15.470%, and thus R was .39332.

INSERT TABLE 2 ABOUT HERE.

The table also reports the unstandardized B weights that can be applied to the observed variables, X_1 and X_2, to optimize the prediction of Y_i for each of the i=25 subjects. Applying optimal weights to observed variables yields a score for each subject on the *synthetic* variable, $\hat{Y}_i$, a variable that is *latent* since the variable is not directly observable. The equation here takes the form:

$$Y_i \text{ <---- } \hat{Y}_i = a + B_1 (X_{1i}) + B_2 (X_{2i})$$

For the Table 1 data, as noted in Table 2, the unstandardized B weights are:

$$Y_i \text{ <---- } \hat{Y}_i = +9.930727 + +.113561 (X_{1i}) + +.098024 (X_{2i})$$

If the variables are standardized, the optimal weights are designated β weights as against B weights. Because the regression surface always contains that coordinate defined by the means of all the variables involved in the analysis, and since the mean of all the standardized variables is 0, the additive, a, weight is always 0 for standardized data. Thus, the regression equation for standardized data simplifies to:

$$Z_{Yi} \text{ <---- } \hat{Y}_{Zi} = \beta_1 (Z_{X1i}) + \beta_1 (Z_{X2i})$$

For the Table 1 data, also as noted in Table 2, the standardized β weights are:

$$Z_{Yi} \text{ <---- } \hat{Y}_{Zi} = +.124569 (Z_{X1i}) + +.332448 (Z_{X2i})$$

Table 2 also reports scores on both versions of the synthetic variable, predicted Y ($\hat{Y}_i$), for all i=25 subjects in the example data set. However, a second set of synthetic variables is present in the data set too. This set represents the deviation scores between the related actual and predicted variable scores for the i=25 subjects. These deviations take the form, respectively:

$$e_i = Y_i - \hat{Y}_i,$$

and

$$e_{Zi} = Z_{Yi} - \hat{Y}_{Zi}$$

Table 2 also reports these "error" scores for all i=25 subjects.

As implied by this discussion, all analyses involve two classes of variables: *observed* variables, and *latent* or *synthetic* variables. In the present example, for both the unstandardized and the standardized forms of the data, there are 3 observed variables, and 2 latent variables. Thus, the data set involves a total of 5 variables.

These data can also be used to make three conceptual points:

1. Conceptually, the $\hat{Y}_i$ (or the $\hat{Y}_{Zi}$) scores can be thought of as *the predicted portion* of Y_i. Therefore, logically $R_{Y \text{ with } X1,X2}$ should always exactly equal $r_{Y \text{ with } YHAT}$.
2. Conceptually, the $\hat{Y}_i$ (or the $\hat{Y}_{Zi}$) scores can be thought of as *the predicted portion* of Y_i, and the e_i scores represent the *unexplained portion* of Y_i. Therefore, by definition, the scores on the two synthetic variables, $\hat{Y}_i$ (or the $\hat{Y}_{Zi}$) and e_i (or e_{Zi}), will always be correlated 0 with each other.
3. Conceptually, the $\hat{Y}_i$ (or the $\hat{Y}_{Zi}$) scores can alternatively be thought of as *the predictively useful portion* of the predictor variables, discarding the predictively useless part of the predictor variables. Since the e_i scores represent the

unexplained portion of the Y_i scores, therefore, logically each predictor will always be correlated 0 with the synthetic variable, e_i.

These expectations are confirmed within the results reported in Table 3.

INSERT TABLE 3 ABOUT HERE.

The various tabled results illustrate the (a) use of least squares weights to optimize explained variance and minimize model error variance. Changing any of the weights reported in Table 2 in any way would, for the data in hand, (1) increase the discrepancy between the actual Y_i scores and the predicted scores $\hat{Y}_i$, i.e., result on the average in larger absolute e_i scores (or e_i^2 scores), and (2) result in a smaller $r_{Y \text{ with } YHAT}^2$ or variance-accounted-for effect size. These are the optimal weights for these data.

All classical parametric methods use such least squares weights to optimize prediction, in just this way. However, many researchers do not recognize that such is the case, because most computer packages do not print the least squares weights that are actually invoked in ANOVA, for example, or when t-tests are conducted. Thus, some researchers unconsciously presume that such methods do not invoke optimal weighting systems.

The tabled results also illustrate the (b) focus on latent synthetic variables in all analyses. The fact that the multiple correlation, $R_{Y \text{ with } X1,X2}$, equals $r_{Y \text{ with } THAT}$ indicates that the synthetic variable is implicitly the focus of multiple regression analysis. In fact, the synthetic variables scores are the focus in all classical parametric analyses (cf. Thompson, 1991).

Finally, the tabled results also illustrate that all classical parametric analyses (c) yield variance-accounted-for effect sizes analogous to r^2. These statistics can equivalently be conceptualized as (a) the ratio of explained to total variance (c.f. the sum-of-squares between divided by the total sum-of-squares) or as (b) the squared correlation between synthetic variable scores and other scores. Again, many computer packages do not automatically print eta^2 or other variance-accounted-for r^2 effect sizes present in analyses such as ANOVA or t-tests, so many researchers fail to appreciate that *all analyses are correlational*, but these effect sizes are always present (Snyder & Lawson, 1993).

Premise #2: OVA Analyses Do Not in and of Themselves Yield the Capacity to Make Causal Inferences

In a seminal article, Cohen (1968, p. 426) noted that ANOVA and ANCOVA are special cases of multiple regression analysis, and argued that in this realization "lie possibilities for more relevant and therefore more powerful exploitation of research data." Since that time researchers have increasingly recognized that conventional multiple regression analysis of data as they were initially collected (no conversion of intervally scaled independent variables into dichotomies or trichotomies) does not discard information or distort reality, and that the "general linear model"

...can be used equally well in experimental or non-experimental research. It can handle continuous and categorical variables. It can handle two, three, four, or more independent variables... Finally, as we will abundantly show, multiple regression analysis can do anything the analysis of variance does--sums of squares, mean squares, F ratios--and more. (Kerlinger & Pedhazur, 1973, p. 3)

Discarding variance is not generally good research practice (Thompson, 1988b). As Kerlinger (1986, p. 558) explains,

...partitioning a continuous variable into a dichotomy or trichotomy throws information away... To reduce a set of values with a relatively wide range to a dichotomy is to reduce its variance and thus its possible correlation with other variables. A good rule of research data analysis, therefore, is: <u>Do not reduce continuous variables to partitioned variables</u> (dichotomies, trichotomies, etc.) unless compelled to do so by circumstances or the nature of the data (seriously skewed, bimodal, etc.).

Kerlinger (1986, p. 558) notes that variance is the "stuff" on which all analysis is based. Discarding variance by categorizing intervally-scaled variables amounts to the "squandering of information" (Cohen, 1968, p. 441). As Pedhazur (1982, pp. 452-453) notes,

Categorization of attribute variables is all too frequently resorted to in the social sciences... It is possible that some of the conflicting evidence in the research literature of a given area may be attributed to the practice of categorization of continuous variables... Categorization leads to a loss of information, and consequently to a less sensitive analysis.

One reason why researchers may be prone (a) to categorizing continuous variables and also (b) to overuse of ANOVA is that some researchers <u>unconsciously</u> and <u>erroneously</u> associate ANOVA with the power of experimental designs. As Thompson (1993) noted,

Even most experimental studies invoke intervally scaled "aptitude" variables (e.g., IQ scores in a study with academic achievement as a dependent variable), to conduct the aptitude-treatment interaction (ATI) analyses recommended so persuasively by Cronbach (1957, 1975) in his 1957 APA Presidential address. (pp. 7-8)

Thus, many researchers employ interval predictor variables, even in experimental designs, but these same researchers too often convert their interval predictor variables to nominal scale merely to conduct OVA analyses.

It is *true* that experimental designs allow causal inferences and that ANOVA is appropriate for many experimental designs. However, it is *not* therefore *true* that doing an ANOVA makes the design experimental and thus allows causal inferences.

Humphreys (1978, p. 873, emphasis added) notes that:

> The basic fact is that a measure of individual differences is not an independent variable [in a experimental design], and it *does not become one* by categorizing the scores and treating the categories as if they defined a variable under experimental control in a factorially designed analysis of variance.

Similarly, Humphreys and Fleishman (1974, p. 468) note that categorizing variables in a nonexperimental design using an ANOVA analysis "not infrequently produces in both the investigator and his audience the illusion that he has experimental control over the independent variable. Nothing could be more wrong." Since all analyses are correlational, and it is the design and not the analysis that yields the capacity to make causal inferences, the practice of converting intervally-scaled predictor variables to nominal scale so that ANOVA and other OVAs (i.e., ANCOVA, MANOVA, MANCOVA) can be conducted is inexcusable, at least in most cases.

As Cliff (1987, p. 130, emphasis added) notes, the practice of discarding variance on intervally scaled predictor variables to perform OVA analyses creates problems in almost all cases:

> Such divisions are not infallible; think of the persons near the borders. Some who should be highs are actually classified as lows, and vice versa. In addition, the "barely highs" are classified the same as the "very highs," even though they are different. Therefore, reducing a reliable variable to a dichotomy makes the variable more unreliable, not less.

These various realizations have led to less frequent use of OVA methods, and to more frequent use of general linear model approaches such as regression (Elmore & Woehlke, 1988). However, canonical correlation analysis, and not regression analysis, is the most general case of the general linear model (Baggaley, 1981, p. 129; Bagozzi, 1981; Fornell, 1978, p. 168). In an important article, Knapp (1978, p. 410) demonstrated this in some mathematical detail and concluded that "virtually all of the commonly encountered tests of significance can be treated as special cases of canonical correlation analysis." Thompson (1988a, 1991) illustrates how canonical correlation analysis can be employed to implement all the parametric tests that canonical methods subsume as special cases.

Premise #3: Statistical Significance Tests Do Not Evaluate the p(robability) that Results Will Replicate

Science is about the business of identifying relationships that recur under stated conditions. Unfortunately, too many researchers, consciously or unconsciously, incorrectly assume that the *p* values calculated in statistical significance tests evaluate the probability that results will replicate (Carver, 1978, 1993). Such researchers often explain what *p* calculated is by invoking vague amorphisms such as, *p* calculated (or statistical significance testing) evaluates whether results "were due to chance".

It is true that statistical significance tests do focus on the null hypothesis. It is also true that such tests evaluate sample statistics (e.g., *sample* means, standard deviations, correlations coefficients) in relation to unknowable population parameters (e.g., *population* means, standard deviations, correlations coefficients).

But far too many resea··:hers incorrectly interpret statistical significance tests as evaluating *the probability that the null is true in the population, given the sample statistics for the data in hand*. This would, in fact, be a very interesting issue to evaluate.

If p calculated informed the researcher about the truth of the null in the population, then this information would directly test the replicability of results. Assuming the population itself remained stable, future samples from the population, if representative, should yield similar results. In this case, results for which the null was found to not be true in the population would therefore be likely to be replicated in future samples from the same population where the null would also likely be rejected. Unfortunately, this is not what statistical significance tests, and not what the associated p calculated values evaluate.

It is true that the p(robability) values calculated in statistical significance testing, which range from 0 to 1 (or 0% to 100%), do require that a "given" regarding the population parameters must be postulated. The characteristics of the population(s) directly affect what the calculated p values will be, and are considered as part of the calculations of p.

For example, if we draw two random samples from two populations, both with equal means, then the single most likely sample statistics (i.e., the sample statistics with the *largest p calculated value*) will be two equal sample means. These sample results are the most likely for these populations. But these exact same sample statistics would be less likely (i.e., would yield a *smaller p calculated value*) if the two populations had parameter means that differed by one unit. And the sample statistics involving exactly equal sample means would be still less likely (i.e., would yield a still *smaller p calculated value*) if the two population means differed by two units.

Indeed, specific population parameters must unavoidably be assumed even to determine what the p calculated is for the sample statistics. Given that population parameters directly affect the calculated p(robability) of the sample statistics, one must assume particular population parameters associated with the null hypothesis being tested (e.g., specific means, medians, standard deviations, correlation coefficients), because there are infinitely many possibilities of what these parameters may be in the population(s).

Only by assuming specific population parameters can a single answer be given to the question, "what is the p(robability) of the sample statistics, assuming the population has certain parameters?"

Without the assumption of specific population parameters being true in the population, there are infinitely many plausible estimates of p, and the answers to the question actually posed by statistical significance testing become mathematically indeterminate.

Classically, to get a single estimate of the p(robability) of the sample statistics, the null hypothesis is posited to be exactly true in the population. Thus, statistical significance testing evaluates *the probability of the sample statistics for the data in hand, given that null hypothesis is presumed to be exactly true as regards the related parameters in the population.*

Of course, this p is a very different animal than one which evaluates the probability of the population parameters themselves, and the statistical significance testing logic itself means that p evaluates something considerably less interesting than result replicability. As Shaver (1993) recently argued so emphatically:

> [A] test of statistical significance is not an indication of the probability that a result would be obtained upon replication of the study. A test of statistical significance yields the probability of a result occurring under [an assumption of the truth of] the null hypothesis [in the population], not the probability that the result will occur again if the study is replicated. Carver's (1978) treatment should have dealt a death blow to this fallacy.... (p. 304)

Furthermore, the requirement that statistical significance testing presumes an assumption that the null hypothesis is true in the population is a requirement that an untruth be posited. As Meehl (1978, p. 822) notes, "As I believe is generally recognized by statisticians today and by thoughtful social scientists, the null hypothesis, taken literally, is always false." Similarly, Hays point out that "[t]here is surely nothing on earth that is completely independent of anything else [in the population]. The strength of association may approach zero, but it should seldom or never be exactly zero."

One logic explaining why the null cannot be true in the population is mathematical. There are infinitely many possible parameters (e.g., means, standard deviations) in the population(s). Probability is the frequency of occurrence of an event divided by the total number of possible events. Therefore, the "point probability" of any single event (e.g., two populations with exactly equal means, a population with the parameter correlation coefficient exactly equal to zero) in the population is infinitely small. Thus, the probability of the null hypothesis being exactly or literally true in the population is infinitely small.

There is a very important implication of the realization that the null is not literally true in the population. The most likely sample statistics for samples drawn from populations in which the null is not literally true are sample statistics which do not correspond to the null hypothesis, e.g., there are some differences in sample means, or r in the sample is not exactly 0. Whenever the null is not exactly true in the sample(s), then the null hypothesis will always be rejected at some sample size. As Hays (1981, p. 293) emphasizes, "virtually any study can be made to show significant results if one uses enough subjects."

Although statistical significance is a function of at least seven interrelated features of a study (Schneider & Darcy, 1984), sample size is a basic influence on significance. Thus, some researchers (Thompson, 1989, 1993) have advocated interpreting statistical significance tests only within the context of sample size. In any case, all this means that:

> Statistical significance testing can involve a tautological logic in which tired researchers, having collected data from hundreds of subjects, then conduct a statistical test to evaluate whether there were a lot of subjects, which the researchers already know, because they collected the data and know they're tired. This tautology has created considerable damage as regards the cumulation of knowledge... (Thompson, 1992b, p. 436)

Thus, statistical significance testing can be a circuitous logic requiring us to invest energy to determine that which we already know, i.e., our sample size. And this energy is not invested in judging the noteworthiness of our effect sizes or the replicability of our effect sizes, since statistical significance testing does not evaluate these considerably more important issues. The recent Summer, 1993, special issue (Vol. 61, No. 4) of the Journal of Experimental Education provides a lucid and thorough treatment of these and related matters. Decades of effort "to exorcise the null hypothesis" (Cronbach, 1975, p. 124) continue.

Implications of the Three Premises for Multivariate Practice

These three premises have implications for correctly conducting multivariate analyses. First, *because all classical parametric analyses are positively biased in the effect sizes they yield, we should estimate the magnitude of this bias as part of the interpretation of all analyses.* It is true that all parametric methods yield positively biased estimates of effects, because all these analyses use least squares weights to optimize prediction, and "one tends to take advantage of chance [both sampling and measurement errors] in any situation where something is optimized from the data at hand" (Nunnally, 1978, p. 298).

Traditionally, many researchers recognized that so-called "correlational analyses" (e.g., *r*, multiple regression, canonical correlation analysis) yielded positively-biased effects, and therefore eschewed these analyses. However, all analyses are correlational and invoke least squares weights, so exactly the same dynamics occur in so-called, incorrectly-called, "non-correlational analyses" (e.g., the OVA methods, *t*-tests).

We can take this positive bias into account in several ways. For example, we can apply statistical corrections to our variance-accounted-for effect size estimates (see Snyder & Lawson, 1993; Thompson, 1990). Alternatively, we can employ replicability analyses, such as cross-validation, the jackknife, or the bootstrap, to empirically estimate the degree of positive bias (cf.

Thompson, 1993, in press). But however we do so, we should consider the positive bias in our detected effects as part of result interpretation.

Second, *because statistical significance does not evaluate the importance of results, effect sizes should initially be subjectively evaluated during interpretation*, to determine if the results are noteworthy and worth more detailed interpretation. As Thompson (1993) argued,

> Statistics can be employed to evaluate the probability of an event. But importance is a question of human values, and math cannot be employed as an atavistic escape (a la Fromme's Escape from Freedom) from the existential human responsibility for making value judgments. If the computer package did not ask you your values prior to its analysis, it could not have considered your value system in calculating p's, and so p's cannot be blithely used to infer the value of research results. (p. 365)

Third, *because all classical parametric analyses create synthetic variables that become the focus of the analysis, the nature of the synthetic variable should be explored as part of the analysis*, once it is decided that the origins of detected effects are worth exploring. There are two ways to explore the nature of the latent or synthetic variables in our analyses.

One way to explore the character of the synthetic variables is to evaluate the standardized weights used to create the synthetic variables. All parametric analyses involve standardized weights similar to the beta weights generated in regression. As Thompson (1992a) noted,

> These weights are all analogous, but are given different names in different analyses (e.g., beta weights in regression, pattern coefficients in factor analysis, discriminant function coefficients in discriminant analysis, and canonical function coefficients in canonical correlation analysis), mainly to obfuscate the commonalities of [all] parametric methods, and to confuse graduate students. (pp. 906-907)

If all standardized weights across analytic methods were called by the same name (e.g., beta weights), then researchers might (correctly) conclude that all analyses are part of the same general linear model.

A variable given a standardized weight of zero is being obliterated by the multiplicative weighting process, indicating either that (a) the variable has zero capacity to explain relationships among the variables or that (b) the variable has some explanatory capacity, but one or more other variables yield the same explanatory information and are arbitrarily (not wrongly, just arbitrarily) receiving all the credit for the variable's predictive power. On the other hand, as the standardized weights for variables deviate more from zero, these variables have more power to explain relationships among the variables.

Because a variable may be assigned a standardized multiplicative weight of zero when (b) the variable has some explanatory capacity, but one or more other variables yield the same explanatory information and are arbitrarily (not wrongly, just arbitrarily) given all the credit for the variable's predictive power, it is *essential* to evaluate other coefficients in addition to standardized weights during interpretation, to determine the specific basis for the weighting. Just as it would be incorrect to evaluate predictor variables in a regression analysis only by consulting beta weights, it would be inappropriate in multivariate analyses to only consult standardized weights during result interpretation (Borgen & Seling, 1978, p. 692; Kerlinger & Pedhazur, 1973, p. 344; Levine, 1977, p. 20; Meredith, 1964, p. 55).

One candidate for this second way of evaluating variable importance involves calculating the bivariate correlation coefficients between the observed variables and the predicted latent or synthetic variable(s). These correlation coefficients are called *structure coefficients* (Thompson & Borrello, 1985).

Two Reasons Why Multivariate Methods Are Usually Essential

There are two reasons why multivariate methods are so important in behavioral research. These are elaborated by Fish (1988), and explored in more detail here.

Controlling "Experimentwise" Type I Error Rates

First, multivariate methods limit the inflation of Type I "experimentwise" error rates. The seriousness of "experimentwise" error inflation, and what to do about it, are both matters prompting some disagreement (e.g., Bray & Maxwell, 1982, p. 343, 1985, p. 10; Hummel & Johnston, 1986). But it is clear that, "Whenever multiple statistical tests are carried out in inferential data analysis, there is a potential problem of 'probability pyramiding'" (Huberty & Morris, 1989, p. 306). And as Morrow and Frankiewicz (1979) emphasize, it is also clear that in some cases inflation of experimentwise error rates can be quite serious.

Most researchers are familiar with "testwise" alpha. But while "testwise" alpha refers to the probability of making a Type I error for a given hypothesis test, "experimentwise" error rate refers to the probability of having made a Type I error anywhere within the study. When only one hypothesis is tested for a given group of people in a study, "experimentwise" error rate will exactly equal the "testwise" error rate. But when more than one hypothesis is tested in a given study with only one sample, the two error rates may not be equal.

Given the presence of multiple hypothesis tests (e.g., two or more dependent variables) in a single study with a single sample, the testwise and the experimentwise error rates will still be equal only if the hypotheses (or the dependent variables) are perfectly correlated. Logically, the correlation of the dependent variables will impact the experimentwise error rate, because, for example, when one has perfectly correlated hypotheses, in actuality, one is still only testing a single hypothesis. Thus, two factors impact the inflation of experimentwise Type I error: (a) the number of hypotheses tested using a single sample of data, and (b) the degree of correlation among the dependent variables or the hypotheses being tested.

When the dependent variables or hypotheses tested using a single sample of data are perfectly <u>un</u>correlated, the experimentwise error rate (α_{EW}) can be calculated. This is done using the Bonferroni inequality:

$$\alpha_{EW} = 1 - (1 - \alpha_{TW})^K,$$

where *k* is the number of perfectly uncorrelated hypotheses being tested at a given testwise alpha level (α_{TW}).

For example, if three perfectly uncorrelated hypotheses (or dependent variables) are tested using data from a single sample, each at the α_{TW}=.05 level of statistical significance, the experimentwise Type I error rate will be:

$$\begin{aligned} \alpha_{EW} &= 1 - (1 - \alpha_{TW})^K \\ &= 1 - (1 - .05)^3 \\ &= 1 - (.95)^3 \\ &= 1 - (.95(.95)(.95)) \\ &= 1 - (.9025 (.95)) \\ &= 1 - .857375 \\ \alpha_{EW} &= .142625 \end{aligned}$$

Thus, for a study testing three perfectly uncorrelated dependent variables, each at the α_{TW}=.05 level of statistical significance, the probability is .142625 (or 14.2625%) that one or more null hypotheses will be incorrectly rejected within the study. Most unfortunately, knowing this will not inform the researcher as to which one or more of the statistically significant hypotheses is, in fact, a Type I error. Table 4 presents these calculations for several conventional α_{TW} levels and for various numbers of perfectly uncorrelated dependent variables or hypotheses.

INSERT TABLE 4 ABOUT HERE.

But these concepts are too abstract to be readily grasped. Happily, Witte (1985, p. 236) explains the two error rates using an intuitively appealing example involving a coin toss. If the toss of heads is equated with a Type I error, and if a coin is tossed only once, then the probability of a head on the one toss (α_{TW}), and of at least one head within the set (α_{EW}) of one toss, will both equal 50%.

If the coin is tossed three times, rather than only once, the "testwise" probability of a head on each toss is still exactly 50%, i.e., α_{TW}=.50 (not .05). Now the Bonferroni inequality is a literal fit to this example situation (i.e., is a literal analogy rather than a figurative analogy), because the coin's behavior on each flip is literally uncorrelated with the coin's behavior on previous flips. That is, a coin is not aware of its behavior on previous flips and does not alter its behavior on any single flip given some awareness of its previous behavior.

Thus, the "experimentwise" probability (α_{EW}) that there will be at least one head in the whole set of three flips will be exactly:

$$\alpha_{EW} = 1 - (1 - \alpha_{TW})^K$$

$$= 1 - (1 - .50)^3$$
$$= 1 - (.50)^3$$
$$= 1 - (.50(.50)(.50))$$
$$= 1 - (.2500 (.50))$$
$$= 1 - .125000$$
$$\alpha_{EW} = .875000$$

Table 5 illustrates these concepts in a more concrete fashion. There are eight equally likely outcomes for sets of three coin flips. These are listed in the table. Seven of the eight equally likely sets of three flips involves one or more Type I error, defined in this example as a heads. And 7/8 equals .875000, as expected, according to the Bonferroni inequality.

INSERT TABLE 5 ABOUT HERE.

Researchers control "testwise" error rates by picking small values, usually 0.05, for the "testwise" alpha. "Experimentwise" error rates can be limited by employing multivariate statistics to test omnibus hypotheses as against lots of more discrete univariate hypotheses.

Paradoxically, although the use of several univariate tests in a single study can lead to too many null hypotheses being spuriously rejected, as reflected in inflation of the "experimentwise" error rate, it is also possible that the failure to employ multivariate methods can lead to a failure to identify statistically significant results which actually exist. Fish (1988) and Maxwell (1992) both provide data sets illustrating this equally disturbing possibility. This means that the so-called "Bonferroni correction" is not a satisfactory solution to this problem.

The "Bonferroni correction" involves using a new testwise alpha level, $\alpha_{TW}*$, computed, for example, by dividing α_{TW} by the number of *k* hypotheses in the study. This approach attempts to control the experimentwise Type I error rate by reducing the testwise error rate level. However, the use of the "Bonferroni correction" does not address the second (and more important) reason why multivariate methods are so often vital, and so even with this correction univariate methods usually still remain unsatisfactory.

Multivariate Methods Honor the Nature of Reality

Multivariate methods are also often vital in behavioral research because multivariate methods best honor the reality to which the researcher is purportedly trying to generalize. As noted previously, since statistical significance testing and error rates may not be the most important aspect of research practice (Thompson, 1989, 1993), this second reason for employing multivariate statistics is actually the more important of the two grounds for using these methods.

Implicit within all analyses is an analytic model. Each researcher also has a presumptive model of what reality is believed to be like. It is critical that our analytic models and our models of reality match, otherwise our conclusions will be invalid. It is generally best to consciously reflect on the fit of these two models whenever we do research. Of course, researchers with different models of reality may make different analytic choices, but this is not disturbing since analytic choices are philosophically driven anyway (Cliff, 1987, p. 349).

But Thompson (1986, p. 9) notes that the reality about which most researchers wish to generalize is usually one "in which the researcher cares about multiple outcomes, in which most outcomes have multiple causes, and in which most causes have multiple effects." Given such a model of reality, it is critical that the full network of all possible relationships be considered *simultaneously* within the analysis.

Most researchers recognize that independent variables can interact with each other to create important and independent effects on the dependent variable. Table 7 illustrates these effects for the heuristic example data presented in Table 6. Here the analyses involved scores on the variables, Z, A and B. The analyses were performed with the SPSS command file contained in Appendix B.

INSERT TABLES 6 AND 7 ABOUT HERE.

In these analyses there is a zero effect size for the main effect, A, in the one-way ANOVA for only that independent variable. The effect size for the main effect, B, for the one-way ANOVA for only that independent variable is also zero.

However, when all three variables are simultaneously considered in a single two-way factorial ANOVA, the eta^2 effect size of 100% is detected for the two-way interaction effect, because the full network of variable relationships is now evaluated within this single analysis. It is exactly for this reason that most researchers conduct factorial analyses when they use OVA methods.

But since all analyses are correlational, the designation of which variable sets are on which side of the analytic equation is arbitrary, because the relationship between variables sets is unaltered by exchanging the two variable sets. Thus, interactions can also occur among dependent variables, and these relationships are equally important to consider.

Table 8 illustrates this possibility using variables X and Y and group membership variable B from Table 6. Neither of the two ANOVAs yield statistically significant effects. And the eta^2 effect sizes ($eta^2 = SOS_{BETWEEN}/SOS_{TOTAL}$) associated with these analyses, respectively 7.9% and 1.0%, might be deemed not noteworthy by some researchers.

INSERT TABLE 8 ABOUT HERE.

But when these same data are analyzed using a one-way MANOVA, as reported in Table 8, the result is statistically significant, even at α_{TW}=.001. Furthermore, a multivariate variance-accounted-for effect size here can be computed as 1 - lambda, and for this analysis the estimate would be .39888 (1-.60112 = .39888 or 39.888%).

Clearly, the multivariate analysis of the Table 6 data yields radically different results than the univariate analysis of the same data. The reason is that the multivariate analysis *simultaneously* considers all the relationships among all the variables, and for these data honoring the relationship between the two dependent variables makes a big difference.

Conceptually, dependent variables can interact in a multivariate analysis, just as independent variables can interact in a multi-way ANOVA. Thus, Tatsuoka's (1973, p. 273) previous remarks remain telling:

> The often-heard argument, "I'm more interested in seeing how each variable, in its own right, affects the outcome" overlooks the fact that any variable taken in isolation may affect the criterion differently from the way it will act in the company of other variables. It also overlooks the fact that multivariate analysis--precisely by considering all the variables simultaneously--can throw light on how each one contributes to the relation.

Put differently, the latent variable that is actually analyzed in the multivariate analysis is more than the conceptual sum of the parts of the two variables taken separately, because the latent variable created in the multivariate analysis takes into account the relationships among all the variables.

The results reported in Table 9 can be used both (a) to reinforce the notion that all parametric analyses are interrelated and (b) to make explicit the synthetic variable that is actually analyzed in the multivariate analyses. A one-way MANOVA yields identical results to those in a discriminant analysis. In fact, the discriminant analysis usually yields more interpretive information that a classical MANOVA (Borgen & Seling, 1978), and therefore one-way MANOVAs are best conducted in the form of discriminant analysis.

INSERT TABLE 9 ABOUT HERE.

Table 10 presents the latent variables actually analyzed in the MANOVA/discriminant analysis. These latent variables were computed by the applying the standardized discriminant function coefficients, which are just like regression beta weights, to the data in a z-score form.

INSERT TABLE 10 ABOUT HERE.

Table 11 presents the ANOVA on these discriminant scores conducted within the SPSS command file presented in Appendix B. The eta^2 effect size for this analysis (39.88758%) matches the multivariate effect size for the Table 8 results.

INSERT TABLE 11 ABOUT HERE.

The degrees of freedom (1/30) in the Table 11 analysis are correct for an ANOVA, i.e., for a two-group one-way ANOVA involving a single dependent variable. But the latent variables actually represent two dependent variables for the two-group problem, so the degrees of freedom actually should be 2/29. When the ANOVA is recalculated using the correct degrees of freedom, the Table 11 ANOVA *F*-calculated for the multivariate latent variable also matches the MANOVA results reported in the bottom section of Table

8. Thus, the focus in all parametric analyses on a latent or synthetic variable is once again illustrated. The fact that multivariate analyses investigate relationships involving a multivariate synthetic variable has a very important implication for analytic practice.

In classical ANOVA, post hoc comparisons are necessary to determine which groups differ if (a) a statistically significant omnibus test is isolated and (b) there are more than two groups involved in the effect. But in multivariate analyses, such as classical MANOVA, when there is a statistically significant omnibus effect post hoc tests will be necessary to address either or both of two questions: (1) which groups differ?, and (2) on which dependent variables do groups differ?. Thus, even when there are only two groups in a multivariate analysis, a statistically significant omnibus result will still require post hoc exploration to address the second question, (2) on which dependent variables do groups differ?.

Too often researchers use MANOVA to test the full network of variable relationships, and if they obtain statistically significant results then employ univariate ANOVAs or *t*-tests to do the post hoc work. This is the so-called "protected *F*-test" analytic approach.

The "protected *F*-test" analytic approach is inappropriate and wrong-headed. The multivariate analysis evaluates multivariate synthetic variables, while the univariate analysis only considers univariate latent variables. Thus, *univariate post hoc tests do not inform the researcher about the differences in the multivariate latent variables actually analyzed in the multivariate analysis.*

Understandably, Borgen and Seling (1978) argue:

> When data truly are multivariate, as implied by the application of MANOVA, a multivariate follow-up technique seems necessary to "discover" the complexity of the data. Discriminant analysis is multivariate; univariate ANOVA is not. (p. 696)

It is illogical to first declare interest in a multivariate omnibus system of variables, and to then explore detected effects in this multivariate world by conducting non-multivariate tests!

References

Baggaley, A. R. (1981). Multivariate analysis: An introduction for consumers of behavioral research. Evaluation Review, 5, 123-131.

Bagozzi, R.P. (1981). Canonical correlation analysis as a special case of a structural relations model. Multivariate Behavioral Research, 16, 437-454.

Borgen, F.H., & Seling, M.J. (1978). Uses of discriminant analysis following MANOVA: Multivariate statistics for multivariate purposes. Journal of Applied Psychology, 63(6), 689-697.

Bray, J.H., & Maxwell, S.E. (1982). Analyzing and interpreting significant MANOVAs. Review of Educational Research, 52, 340-367.

Carver, R. (1978). The case against statistical significance testing. Harvard Educational Review, 48, 378-399.

Carver, R. (1993). The case against statistical significance testing, revisited. Journal of Experimental Education, 61(4), 287-292.

Cliff, N. (1987). Analyzing multivariate data. San Diego: Harcourt Brace Jovanovich.

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Cronbach, L.J. (1957). The two disciplines of psychology. American Psychologist, 12, 671-684.

Cronbach, L.J. (1975). Beyond the two disciplines of psychology. American Psychologist, 30, 116-127.

Edgington, E.S. (1964). A tabulation of inferential statistics used in psychology journals. American Psychologist, 19, 202-203.

Edgington, E.S. (1974). A new tabulation of statistical procedures in APA journals. American Psychologist, 29(1), 25-26.

Edwards, A.L. (1985). Multiple regression and the analysis of variance and covariance (2nd ed.). New York: Freeman.

Elmore, P.B., & Woehlke, P.L. (1988). Statistical methods employed in American Educational Research Journal, Educational Researcher, and Review of Educational Research from 1978 to 1987. Educational Researcher, 17(9), 19-20.

Fish, L.J. (1988). Why multivariate methods are usually vital. Measurement and Evaluation in Counseling and Development, 21, 130-137.

Fornell, C. (1978). Three approaches to canonical analysis. Journal of the Market Research Society, 20, 166-181.

Goodwin, L.D., & Goodwin, W.L. (1985). Statistical techniques in AERJ articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Researcher, 14(2), 5-11.

Hays, W. L. (1981). Statistics (3rd ed.). New York: Holt, Rinehart and Winston.

Hinkle, D.E., Wiersma, W., & Jurs, S.G. (1979). Applied statistics for the behavioral sciences. Chicago: Rand McNally.

Huberty, C.J, & Morris, J.D. (1989). Multivariate analysis versus multiple univariate analysis. Psychological Bulletin, 105, 302-308.

Hummel, T.J., & Johnston, C.B. (1986, April). An empirical comparison of size and power of seven methods for analyzing multivariate data in the two-sample case. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 269 480)

Humphreys, L.G. (1978). Doing research the hard way: Substituting analysis of variance for a problem in correlational analysis. Journal of Educational Psychology, 70, 873-876.

Humphreys, L.G., & Fleishman, A. (1974). Pseudo-orthogonal and other analysis of variance designs involving individual-differences variables. Journal of Educational Psychology, 66, 464-472.

Kerlinger, F. N. (1986). Foundations of behavioral research (3rd ed.). New York: Holt, Rinehart and Winston.

Kerlinger, F. N., & Pedhazur, E. J. (1973). Multiple regression in behavioral research. New York: Holt, Rinehart and Winston.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. Psychological Bulletin, 85, 410-416.

Levine, M. S. (1977). Canonical analysis and factor comparison. Newbury Park, CA: SAGE.

Maxwell, S. (1992). Recent developments in MANOVA applications. In B. Thompson (Ed.), Advances in social science methodology (Vol. 2, pp. 137-168). Greenwich, CT: JAI Press.

Meehl, P.E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. Journal of Consulting and Clinical Psychology, 46, 806-834.

Meredith, W. (1964). Canonical correlations with fallible data. Psychometrika, 29, 55-65.

Morrow, J.R., Jr., & Frankiewicz, R.G. (1979). Strategies for the analysis of repeated and multiple measures designs. Research Quarterly, 50, 297-304.

Nunnally, J.C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.

Pedhazur, E. J. (1982). Multiple regression in behavioral research: Explanation and prediction (2nd ed.). New York: Holt, Rinehart and Winston.

Schneider, A. L., & Darcy, R. E. (1984). Policy implications of using significance tests in evaluation research. Evaluation Review, 8, 573-582.

Shaver, J. (1993). What statistical significance testing is, and what it is not. Journal of Experimental Education, 61(4), 293-316.

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. Journal of Experimental Education, 61(4), 334-349.

Tatsuoka, M. M. (1973). Multivariate analysis in educational research. In F. N. Kerlinger (Ed.), Review of research in education (pp. 273-319). Itasca, IL: Peacock.

Thompson, B. (1984). Canonical correlation analysis: Uses and interpretation. Newbury Park, CA: SAGE.

Thompson, B. (1986, November). Two reasons why multivariate methods are usually vital. Paper presented at the annual meeting of the Mid-South Educational Research Association, Memphis.

Thompson, B. (1988a, April). Canonical correlation analysis: An explanation with comments on correct practice. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 295 957)

Thompson, B. (1988b). Discarding variance: A cardinal sin in research. Measurement and Evaluation in Counseling and Development, 21, 3-4.

Thompson, B. (1989). Statistical significance, result impo tance, and result generalizability: Three noteworthy but somewhat different issues. Measurement and Evaluation in Counseling and Development, 22, 2-6.

Thompson, B. (1990). Finding a correction for the sampling error in multivariate measures of relationship: A Monte Carlo study. Educational and Psychological Measurement, 50, 15-31.

Thompson, B. (1991). A primer on the logic and use of canonical correlation analysis. Measurement and Evaluation in Counseling and Development, 24(2), 80-95.

Thompson, B. (1992a). DISCSTRA: A computer program that computes bootstrap resampling estimates of descriptive discriminant analysis function and structure coefficients and group centroids. Educational and Psychological Measurement, 52, 905-911.

Thompson, B. (1992b). Two and one-half decades of leadership in measurement and evaluation. Journal of Counseling and Development, 70, 434-438.

Thompson, B. (1993). The use of statistical significance tests in research: Bootstrap and other alternatives. Journal of Experimental Education, 61(4), 361-377.

Thompson, B. (in press). The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results. Journal of Personality.

Thompson, B., & Borrello, G. M. (1985). The importance of structure coefficients in regression research. Educational and Psychological Measurement, 45, 203-209.

Willson, V.L. (1980). Research techniques in AERJ articles: 1969 to 1978. Educational Researcher, 9(6), 5-10.

Witte, R.S. (1985). Statistics (2nd ed.). New York: Holt, Rinehart and Winston.

Table 1
Edwards (1985, pp. 41-45) Data and Related Latent Variables

Id	X1	X2	Y	ZX1	ZX2	ZY	YHAT	E	YHATZ	EZ
1	11	38	10	-.05317	1.14354	-1.41426	14.90	-4.90	.37	-1.79
2	7	42	16	-1.38232	1.57344	.77274	14.84	1.16	.35	.42
3	12	38	18	.27912	1.14354	1.50174	15.02	2.98	.41	1.09
4	13	36	15	.61141	.92859	.40824	14.94	.06	.38	.02
5	14	40	15	.94370	1.35849	.40824	15.44	-.44	.57	-.16
6	15	32	11	1.27599	.49868	-1.04976	14.77	-3.77	.32	-1.37
7	5	20	13	-2.04690	-.79102	-.32076	12.46	.54	-.52	.20
8	14	44	18	.94370	1.78839	1.50174	15.83	2.17	.71	.79
9	14	34	12	.94370	.71363	-.68526	14.85	-2.85	.35	-1.04
10	10	28	16	-.38546	.06878	.77274	13.81	2.19	-.03	.80
11	8	24	10	-1.05003	-.36112	-1.41426	13.19	-3.19	-.25	-1.16
12	16	30	16	1.60828	.28373	.77274	14.69	1.31	.29	.48
13	15	26	15	1.27599	-.14617	.40824	14.18	.82	.11	.30
14	14	24	12	.94370	-.36112	-.68526	13.87	-1.87	.00	-.68
15	10	26	12	-.38546	-.14617	-.68526	13.61	-1.61	-.10	-.59
16	9	18	14	-.71774	-1.00597	.04374	12.72	1.28	-.42	.47
17	11	30	16	-.05317	.28373	.77274	14.12	1.88	.09	.69
18	9	26	13	-.71774	-.14617	-.32076	13.50	-.50	-.14	-.18
19	7	18	11	-1.38232	-1.00597	-1.04976	12.49	-1.49	-.51	-.54
20	10	10	17	-.38546	-1.86577	1.13724	12.05	4.95	-.67	1.81
21	9	12	8	-.71774	-1.65082	-2.14327	12.13	-4.13	-.64	-1.51
22	10	32	18	-.38546	.49868	1.50174	14.20	3.80	.12	1.38
23	10	18	14	-.38546	-1.00597	.04374	12.83	1.17	-.38	.43
24	16	20	15	1.60828	-.79102	.40824	13.71	1.29	-.06	.47
25	10	18	12	-.38546	-1.00597	-.68526	12.83	-.83	-.38	-.30

Note. Variables X_1, X_2, and Y are from Edwards (1985, p. 41). The latent/synthetic variables, YHAT and E, and YHATZ and EZ, were computed by applying the weights from Table 2 to observed/manifest variables, X_1 and X_2.

Table 2
SPSS Regression Output for the Table 1 Data

Multiple R	.39332	Analysis of Variance			
R Square	.15470		DF	Sum of Squares	Mean Square
Adjusted R Square	.07786	Regression	2	27.94522	13.97261
Standard Error	2.63452	Residual	22	152.69478	6.94067
		F = 2.01315		Signif F = .1574	

------------------ Variables in the Equation ------------------

Variable	B	SE B	Beta	T	Sig T
X2	.098024	.061603	.332448	1.591	.1258
X1	.113561	.190462	.124569	.596	.5571
(Constant)	9.930727	2.270875		4.373	.0002

Note. This table presents the SPSS output exactly as it was produced by the package.

Table 3
SPSS Output Correlating 3 Manifest/Observed and 2 Latent Variables

	X1	X2	Y	YHAT	YHATZ	E	EZ
X1	1.0000	.3461	.2396	.6092[a]	.6092[a]	.0000	.0000
X2	.3461	1.0000	.3756	.9548[a]	.9548[a]	.0000	.0000
Y	.2396	.3756	1.0000	.3933	.3933	.9194	.9194
YHAT	.6092[a]	.9548[a]	.3933	1.0000	1.0000	.0000	.0000
YHATZ	.6092[a]	.9548[a]	.3933	1.0000	1.0000	.0000	.0000
E	.0000	.0000	.9194	.0000	.0000	1.0000	1.0000
EZ	.0000	.0000	.9194	.0000	.0000	1.0000	1.0000

Note. As noted within the narrative, underlined values are always 0, by definition. The bivariate correlations between predicted and actual dependent variable scores equal *R*, also as explained within the narrative. This table presents the SPSS output exactly as it was produced by the package.

[a]Correlation coefficients between observed predictor variable scores and synthetic predictor variable scores are structure coefficients, as explained by Thompson and Borrello (1985).

Table 4
Formula for Estimating Experimentwise Type I Error Inflation When Hypotheses are Perfectly Uncorrelated

	TW alpha		Tests		Experimentwise alpha
1 - (1 -	0.05	) **	1	=	
1 - (	0.95	) **	1	=	[a]
1 -	0.95			=	0.05000
Range Over Testwise (TW) alpha = .01					
1 - (1 -	0.01	) **	5	=	0.04901
1 - (1 -	0.01	) **	10	=	0.09562
1 - (1 -	0.01	) **	20	=	0.18209
Range Over Testwise (TW) alpha = .05					
1 - (1 -	0.05	) **	5	=	0.22622
1 - (1 -	0.05	) **	10	=	0.40126
1 - (1 -	0.05	) **	20	=	0.64151
Range Over Testwise (TW) alpha = .10					
1 - (1 -	0.10	) **	5	=	0.40951
1 - (1 -	0.10	) **	10	=	0.65132
1 - (1 -	0.10	) **	20	=	0.87842

Note. "**" = "raise to the power of".

[a]These calculations are presented (a) to illustrate the implementation of the formula step by step and (b) to demonstrate that when only one test is conducted, the experimentwise error rate equals the testwise error rate, as should be expected if the formula behaves properly.

Table 5
All Possible Families of Outcomes for a Fair Coin Flipped Three Times

	Flip # 1	2	3
1.	T	T	T
2.	H	T	T
3.	T	H	T
4.	T	T	H
5.	H	H	T
6.	H	T	H
7.	T	H	H
8.	H	H	H
p of H on each Flip	50%	50%	50%

p of 1 or more H's (TW error analog) in set of 3 Flips = 7/8 = 87.5%

or

where TW error analog = .50,

$$EW\ p = 1 - (1 - .5)^3 = 1 - (.5)^3 = 1 - .125 = .875$$

Note. The probability of one or more occurrences of a given outcome in a set of events is $1 - (1-p)^k$, where *p* is the probability of the given occurrence on each trial and *k* is the number of trials in a set of perfectly independent events.

Table 6
Data Illustrating Interaction Effects
Can Occur in Both Independent and Dependent Variable Sets

Id	A	B	Y	X	Z
1	1	1	4	2	5
2	1	1	5	3	5
3	1	1	4	4	5
4	1	1	4	5	5
5	1	1	3	4	5
6	1	1	6	5	5
7	1	1	5	6	5
8	1	1	7	5	5
9	2	1	6	6	15
10	2	1	8	6	15
11	2	1	7	6	15
12	2	1	9	7	15
13	2	1	8	7	15
14	2	1	8	8	15
15	2	1	9	8	15
16	2	1	9	9	15
17	1	2	1	2	15
18	1	2	3	3	15
19	1	2	3	5	15
20	1	2	3	5	15
21	1	2	2	5	15
22	1	2	4	6	15
23	1	2	4	5	15
24	1	2	5	6	15
25	2	2	6	6	5
26	2	2	6	6	5
27	2	2	6	7	5
28	2	2	7	7	5
29	2	2	7	7	5
30	2	2	8	9	5
31	2	2	8	9	5
32	2	2	9	9	5

Note. Data for variables *B*, *X*, and *Y* are from Fish (1988).

Table 7
Illustration that Independent Variables A and B from Table 6 Can Create Interaction Effects that Influence Z

One-way ANOVA Predicting Z With Group Membership Variable A

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Main Effects	.000	1	.000	.000	1.00
B	.000	1	.000	.000	1.00
Explained	.000	1	.000	.000	1.00
Residual	800.000	30	26.667		
Total	800.000	31	25.806		

One-way ANOVA Predicting Z With Group Membership Variable B

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Main Effects	.000	1	.000	.000	1.00
A	.000	1	.000	.000	1.00
Explained	.000	1	.000	.000	1.00
Residual	800.000	30	26.667		
Total	800.000	31	25.806		

Two-way ANOVA Predicting Z With Group Membership Variables A and B

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Main Effects	.000	2	.000		
A	.000	1	.000		
B	.000	1	.000		
2-Way Interactions	800.000	1	800.000		
A B	800.000	1	800.000		
Explained	800.000	3	266.667		
Residual	.000	28	.000		
Total	800.000	31	25.806		

Note. This table presents the SPSS output exactly as it was produced by the package.

Table 8
Illustration that Dependent Variables (X and Y from Table 6) Can Also Interact As Regards Their Relationships with Each Other and A Group Membership Variable (B)

One-way ANOVA Predicting X With Group Membership Variable B

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Main Effects	1.125	1	1.125	.300	.588
B	1.125	1	1.125	.300	.588
Explained	1.125	1	1.125	.300	.588
Residual	112.375	30	3.746		
Total	113.500	31	3.661		

One-way ANOVA Predicting Y With Group Membership Variable B

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Main Effects	12.500	1	12.500	2.577	.119
B	12.500	1	12.500	2.577	.119
Explained	12.500	1	12.500	2.577	.119
Residual	145.500	30	4.850		
Total	158.000	31	5.097		

One-way MANOVA Predicting X and Y With Group Membership Variable B
Multivariate Tests of Significance (S = 1, M = 0, N = 13 1/2)

Test Name	Value	Exact F	Hypoth. DF	Error DF	Sig. of F
Pillais	.39888	9.62150	2.00	29.00	.001
Hotellings	.66355	9.62150	2.00	29.00	.001
Wilks	.60112	9.62150	2.00	29.00	.001
Roys	.39888				

Note. This table presents the SPSS output exactly as it was produced by the package.

Table 9

Illustration that a One-Way MANOVA on Dependent Variables X and Y and Group Membership Variable B from Table 6 Is the Same as a Discriminant Analysis on the Same Variables

CANONICAL DISCRIMINANT FUNCTIONS

FUNCTION	EIGENVALUE	PERCENT OF VARIANCE	CUMULATIVE PERCENT	CANONICAL CORRELATION	:	AFTER FUNCTION	WILKS' LAMBDA	CHI-SQUARED	D.F.	SIGN.
					:	0	0.6011236	14.760	2	0.0006
1*	0.66355	100.00	100.00	0.6315666	:					

* MARKS THE 1 CANONICAL DISCRIMINANT FUNCTIONS REMAINING IN THE ANALYSIS.

STANDARDIZED CANONICAL DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1
X	-1.97800
Y	2.10394

STRUCTURE MATRIX:
POOLED WITHIN-GROUPS CORRELATIONS BETWEEN DISCRIMINATING VARIABLES
AND CANONICAL DISCRIMINANT FUNCTIONS
(VARIABLES ORDERED BY SIZE OF CORRELATION WITHIN FUNCTION)

	FUNC 1
Y	0.35982
X	-0.12283

CANONICAL DISCRIMINANT FUNCTIONS EVALUATED AT GROUP MEANS (GROUP CENTROIDS)

GROUP	FUNC 1
1	0.78872
2	-0.78872

Note. This table presents the SPSS output exactly as it was produced by the package. The *F*-test from the MANOVA analysis reported in Table 8 is equivalent to the χ^2 test SPSS uses to test the same hypothesis in discriminant analysis. Thus, the likelihood ratios in the two tables differ only as to the arbitrary number of decimals to which these values are reported. Similarly, the *p* calculated values are identical. The $p=.0006$ for the discriminant results does round to $p=.001$, given the arbitrary decision by the programmers to report to fewer decimal for the MANOVA results.

Table 10
The Latent Variables Scores and Other Results

CASE SEQNUM	ACTUAL GROUP		HIGHEST PROBABILITY GROUP	P(D/G)	P(G/D)	2ND HIGHEST GROUP	P(G/D)	DISCRIMINANT SCORES
1	1		1	0.1337	0.9737	2	0.0263	2.2884
2	1		1	0.1518	0.9708	2	0.0292	2.2217
3	1		1	0.5862	0.5952	2	0.4048	0.2444
4	1	**	2	0.9911	0.7732	1	0.2268	-0.7776
5	1	**	2	0.9380	0.7543	1	0.2457	-0.7110
6	1		1	0.7306	0.8566	2	0.1434	1.1331
7	1	**	2	0.9557	0.7911	1	0.2089	-0.8443
8	1		1	0.1937	0.9642	2	0.0358	2.0884
9	1		1	0.4980	0.5437	2	0.4563	0.1111
10	1		1	0.2175	0.9604	2	0.0396	2.0218
11	1		1	0.7812	0.8432	2	0.1568	1.0664
12	1		1	0.2434	0.9562	2	0.0438	1.9551
13	1		1	0.8328	0.8288	2	0.1712	0.9998
14	1	**	2	0.4434	0.5088	1	0.4912	-0.0222
15	1		1	0.8852	0.8134	2	0.1866	0.9331
16	1	**	2	0.4840	0.5350	1	0.4650	-0.0889
17	2		2	0.8328	0.7132	1	0.2868	-0.5777
18	2	**	1	0.6329	0.6203	2	0.3797	0.3110
19	2		2	0.3450	0.9390	1	0.0610	-1.7330
20	2		2	0.3450	0.9390	1	0.0610	-1.7330
21	2		2	0.0575	0.9858	1	0.0142	-2.6883
22	2		2	0.3121	0.9447	1	0.0553	-1.7996
23	2		2	0.9911	0.7732	1	0.2268	-0.7776
24	2		2	0.9557	0.7911	1	0.2089	-0.8443
25	2	**	1	0.4980	0.5437	2	0.4563	0.1111
26	2	**	1	0.4980	0.5437	2	0.4563	0.1111
27	2		2	0.9027	0.8080	1	0.1920	-0.9109
28	2	**	1	0.4567	0.5175	2	0.4825	0.0444
29	2	**	1	0.4567	0.5175	2	0.4825	0.0444
30	2		2	0.7983	0.8385	1	0.1615	-1.0442
31	2		2	0.7983	0.8385	1	0.1615	-1.0442
32	2		2	0.4840	0.5350	1	0.4650	-0.0889

Note. This table presents the SPSS output exactly as it was produced by the package.

Table 11
ANOVA on Discriminant Latent/Synthetic Variable
Using Both 1/30 and 2/29 Degrees of Freedom

One-way ANOVA Predicting Table 10 DSCORE1 With Group Membership Variable B

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	1	19.9065	19.9065	19.9065	.0001
WITHIN GROUPS	30	30.0000	1.0000		
TOTAL	31	49.9065			

Note. eta^2 = (19.9065 / 49.9065) = .3988758. eta = $.3988758^{.5}$ = .6315661.

Recalculated ANOVA Using Actual Degrees of Freedom (2/29)

SOS	df	MS	Fcalc
19.9065	2	9.9533	9.6215
30.0000	29	1.0345	
49.9065	31		

Note. This upper ANOVA table presents the SPSS output exactly as it was produced by the package.

Appendix A
SPSS Command File for the Edwards (1985, pp. 41-45) Data

```
TITLE 'Edwards (1985, pp. 41-45) Example   EDWARDS.SPS #####'
DATA LIST FILE=ABC RECORDS=1/1 X1 1-2 X2 4-5 Y 7-8
LIST VARIABLES=ALL/CASES=25/FORMAT=NUMBERED
SUBTITLE '1.  Multiple Regression $$$$$$$$$$'
REGRESSION VARIABLES=X1 TO Y/DESCRIPTIVES=DEFAULTS/DEPENDENT=Y/
  ENTER X1 X2
subtitle '2.  get z score version of the 3 observed variables'
descriptives variables=x1 to y/save
compute yhat=9.930726782+(.113561*x1)+(.098024*x2)
compute e=y-yhat
compute yhatz=(.124569*zx1)+(.332448*zx2)
compute ez=zy-yhatz
list variables=all/cases=25/format=numbered
subtitle '3.  show R Y with X1,X2 = r Y with YHAT, etc.'
correlations variables=x1 to y yhat yhatz e ez/
  statistics=descriptives
```

Note. Commands in upper case letters were run first, to obtain the information needed for the lower case commands, which were then inserted, and the job was re-executed.

Appendix B
SPSS Command File For Fish (1988) and Other Data

```
TITLE 'Demo Multivariate Important  See Fish, MECD, Vol 21, 130-137'
DATA LIST FILE=ABC RECORDS=1/1
  A 1 B 2 Y 3 X 4 Z 5-6
LIST VARIABLES=ALL/CASES=50/FORMAT=NUMBERED
SUBTITLE '1a  Example Interaction of I.V. on Dependent @@@@@'
ANOVA Z BY B(1,2)/STATISTICS=MEAN
SUBTITLE '1b  Example Interaction of I.V. on Dependent @@@@@'
ANOVA Z BY A(1,2)/STATISTICS=MEAN
SUBTITLE '1c  Example Interaction of I.V. on Dependent @@@@@'
ANOVA Z BY A(1,2) B(1,2)/STATISTICS=MEAN
SUBTITLE '2a  Show Can Have Interactions in D.V. Too #######'
ANOVA Y BY B(1,2)/STATISTICS=MEAN
SUBTITLE '2b  Show Can Have Interactions in D.V. Too #######'
ANOVA X BY B(1,2)/STATISTICS=MEAN
SUBTITLE '2c  Show Can Have Interactions in D.V. Too #######'
MANOVA X,Y BY B(1,2)/
  PRINT CELLINFO(MEANS,COV,COR) HOMOGENEITY(BOXM)
  SIGNIF(MULTIV EIGEN DIMENR) DISCRIM(RAW,STAN,COR,ALPHA(.99))/
  DESIGN=B
SUBTITLE '3   Show One-way MANOVA is Discriminant $$$$$$$$'
DISCRIMINANT GROUPS=B(1,2)/VARIABLES=X,Y/ANALYSIS=X,Y/METHOD=DIRECT/
  SAVE = SCORES=DSCORE
STATISTICS 1,2,3,4,6,7,8,9,11,13,14,15
SUBTITLE '4a  Show Confidence Intervals About Centroids %%%%%%'
ONEWAY DSCORE1 BY B(1,2)/STATISTICS=ALL
SUBTITLE '4b Show Effect Size for Synthetic Variable Same %%%'
IF (B EQ 1)BCONTRAS=-1
IF (B EQ 2)BCONTRAS=1
REGRESSION VARIABLES=DSCORE1 BCONTRAS/DESCRIPTIVES=DEFAULTS/
  DEPENDENT=DSCORE1/ENTER BCONTRAS
SUBTITLE '5  Show Calculation of Discriminant Scores !!!!!!!!'
LIST VARIABLES=ALL/CASES=5000/FORMAT=NUMBERED
COMPUTE PZX=(X-5.875)/(3.745833**.5)
COMPUTE PZY=(Y-5.750)/(4.850000**.5)
COMPUTE DS1=(-1.97800*PZX)+(2.10394*PZY)
LIST VARIABLES=ALL/CASES=5000/FORMAT=NUMBERED
SUBTITLE '6  Show Coefficient Relationships ****************'
CORRELATIONS VARIABLES=X Y PZX PZY B BCONTRAS DS1 DSCORE1/
  STATISTICS=DESCRIPTIVE
SUBTITLE '7a  Graphically Show How Multivariate Can Matter^^^^'
TEMPORARY
SELECT IF (B EQ 1)
REGRESSION VARIABLES=X Y/DESCRIPTIVES=DEFAULTS/DEPENDENT=Y/ENTER X
TEMPORARY
SELECT IF (B EQ 2)
REGRESSION VARIABLES=X Y/DESCRIPTIVES=DEFAULTS/DEPENDENT=Y/ENTER X
REGRESSION VARIABLES=X B/DESCRIPTIVES=DEFAULTS/DEPENDENT=X/ENTER B
REGRESSION VARIABLES=Y B/DESCRIPTIVES=DEFAULTS/DEPENDENT=Y/ENTER B
SUBTITLE '7b  Plot X Across B(1,2) *******************'
PLOT
  VERTICAL='Variable X' MIN(1) MAX(9)/
  HORIZONTAL='Variable B Group' MIN(.5) MAX(2.5)/
  PLOT=X WITH B
SUBTITLE '7c  Plot Y Across B(1,2) *******************'
PLOT
  VERTICAL='Variable Y' MIN(1) MAX(9)/
  HORIZONTAL='Variable B Group' MIN(.5) MAX(2.5)/
  PLOT=Y WITH B
SUBTITLE '7d  Plot DSCORE Across B(1,2) *******************'
PLOT
  VERTICAL='Variable D' MIN(-4) MAX(4)/
```

```
  HORIZONTAL='Variable B Group' MIN(.5) MAX(2.5)/
  PLOT=DSCORE1 WITH B
SUBTITLE '7e  Plot Y and X Across B(1,2) ^^^^^^^^^^^^^^^^^^'
VALUE LABELS B 1 '1' 2 '2'
PLOT
  VERTICAL='Variable X' /
  HORIZONTAL='Variable Y ("1" = B Group 1, "2" = B Group 2)'/
  PLOT=Y WITH X BY B
```