9 Canonical Correlation Analysis

Bruce Thompson

Canonical correlation analysis (CCA) is an analytic method that can be used to investigate relationships among two or more variable sets. Each variable set usually consists of at least two variables (otherwise the canonical analysis is typically called something else, such as a *t* test or a regression analysis). Although in theory the canonical logic can be generalized to more than two variable sets (Horst, 1961), in practice most researchers use CCA in situations involving only two variable sets.

Canonical analysis was originally conceptualized by Hotelling (1935). Notwithstanding its long history, as Krus, Reynolds, and Krus (1976) noted, "dormant for nearly half a century, Hotelling's (1935) canonical variate analysis has come of age. The principal reason behind its resurrection was its computerization and inclusion in major statistical packages" (p. 725). Of course, having sophisticated statistical packages available does not in and of itself justify the use of CCA or any other analysis.

For two reasons, however, multivariate methods are being used with increasing frequency. First, multivariate methods control the inflation of experimentwise (Type I) error rates ($\alpha_{experimentwise}$) that can occur when several univariate tests are conducted with a single sample's data, even when the testwise error rate ($\alpha_{testwise}$) is very small. Thompson (1994d) provided further explanation of what experimentwise error is and how this error rate can be estimated.

Second, multivariate methods, such as CCA, best honor the nature of the reality that most researchers want to study because many of us believe that we live in a reality where most effects have multiple causes and most causes have multiple effects. As Tatsuoka (1973) emphasized,

> 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. (p. 273)

That is, univariate and multivariate analyses of the same data can yield results that differ like night and day with regard to both statistical significance and effect sizes (R^2, eta^2, etc.), and the multivariate picture in such cases is the accurate portrayal. Fish (1988) provided an empirical example of how univariate and multivariate analyses of the same data can yield contradictory results.

Thompson's (1999a) example was even more dramatic. For his data, two univariate analyses of variance yielded statistically nonsignificant results (both p values were .774) with eta^2 variance-accounted-for effect sizes of both 0.5%. For the same data analyzed by multivariate analysis of variance (MANOVA), $p_{calculated}$ was .000239, and the multivariate eta^2 was 62.5%.

This second reason for the more frequent use of multivariate methods is the more noteworthy of the two. Some researchers avoid inflated experimentwise error rates by making so-called Bonferroni corrections (i.e., downward adjustments in $\alpha_{testwise}$ so as to moderate increases in $\alpha_{experimentwise}$), but this second reason still applies even when such adjustments are invoked. Furthermore, this second reason for using multivariate analyses is more noteworthy because the first reason involves statistical significance testing, and social scientists have been placing less emphasis on statistical significance testing (cf. Cohen, 1994; Thompson, 1996; Thompson & Snyder, 1997, 1998). Indeed, the APA Task Force on Statistical Inference has issued a report that greatly emphasizes the importance of focusing interpretations on effect sizes, particularly in relation to the previous effects found in related prior studies (Wilkinson & APA Task Force on Statistical Inference, 1999).

It is also important to emphasize that although some researchers incorrectly believe that they can appropriately first conduct multivariate tests and then conduct so-called "protected" univariate tests (Maxwell, 1992, pp. 138–140), again the second rationale for conducting multivariate analyses still exists (Thompson, 1994d). That is, univariate tests cannot reasonably be used to investigate and understand the patterns first isolated in multivariate analyses; only a multivariate analysis can explore a multivariate effect.

Thus, for these two reasons, CCA has been used in a variety of published research. Wood and Erskine (1976) and Thompson (1989) provided extensive bibliographies of applications of CCA. Example applications include those reported by Chastain and Joe (1987); Dunst and Trivette (1988); Estabrook (1984); Fowler and Macciocchi (1986); Fuqua, Seaworth, and Newman (1987); Pitts and Thompson (1984); and Zakaahi and Duran (1982). One particularly interesting application involves studies of multivariate test–retest score reliability or of multivariate criterion-related score validity (cf. Sexton, McLean, Boyd, Thompson, & McCormick, 1988).

The purpose of this chapter is to provide a primer on CCA. A longer and more technical treatment is provided by Thompson (1984). The chapter (a) explains the basic logic of CCA using a heuristic data set, (b) provides a brief explanation of how CCA is related to other commonly used univariate and multivariate parametric analyses, (c) illustrates the steps in interpreting canonical results, and (d) details some common errors to avoid in interpreting canonical analyses.

Basic Logic of Canonical Calculations

Problems With the Nonmultivariate Alternative

Imagine that the director of personnel for a national chain of department stores wishes to determine the relationship between characteristics of sales staff and indices of job performance, using a random selection of salespeople from various stores. The first set of variables is obtained from personnel files and is comprised of scores on three questionnaires: leadership potential, need for achievement, and empathy; and of scores on a fourth variable, years of previous sales experience. Call this variable set "employee attributes." A conceptually discrete second set of variables, called "job performance," might be comprised of scores on three measures: past year's sales, absenteeism, and numbers of suggestions previously submitted for improving store operations. How can the director of personnel make sense of these data?

The investigator could first separately examine the *intradomain* matrix of bivariate correlations between the variables in each set. For example, an examination of only the employee attributes variables could provide insight regarding the nature of their interrelatedness. If these measures were highly correlated, this would suggest that these scores measure one underlying dimension or construct; such an outcome might be unlikely for these variables. However, if only some of the measures correlated highly with each other, this would suggest that the use of only one dimension to represent all the employee attributes variables is inappropriate.

The investigator could also examine the bivariate relationships between scores on the measures across the two variable sets (i.e., the *interdomain* correlation coefficients). The director of personnel may wonder whether all of the measures of employee attributes correlate strongly with all of the measures of job performance. Of course, for these variables, this outcome is also unlikely.

The problem with examining only the 21 ([7 × (7 − 1)]/2 = (7 × 6)/2 = 42/2) unique bivariate correlation coefficients defined by the scores on the seven variables is that even for this relatively simple problem, the dynamics represented within the data are beyond the perceptual ability of nearly anyone unless the patterns within the data are ridiculously obvious (e.g., all 21 correlation coefficients are nearly +1). Furthermore, several different patterns of relationships may simultaneously underlie the system of relationships. These can be discovered only by using a multivariate analysis. Of course, even if the patterns within the data are blatantly obvious, the multivariate analysis will still detect them.

The same conclusion about why the bivariate approach does not work can be seen from the perspective of individual scores. In the simplest of all worlds, all of the employee attributes variables measure one underlying construct, and the same is true for the job performance measures. In this case, one single score (rather than four different scores) could be used to describe the employee attributes data, and one single score could be used to describe the job performance data. Assuming that the variables are measured using a common metric (e.g., *z* scores), a simple approach would then involve adding together the four employee attributes scores to obtain one overall "employee" score for each salesperson and adding together the three job performance scores to obtain one overall "performance" score for each person. The single bivariate Pearson correlation between these two scores could then reasonably be interpreted as the sole basis for understanding the relationship between the two variable sets.

If the variable sets cannot each be appropriately considered purely unidimensional, however, then computing this simple sum of scores is akin to adding apples and oranges. In addition, regardless of the dimensionality of each variable set, a simple sum of scores does not reflect the possibility that some variables are more important and thus need to be weighted more strongly than others when exploring relationships between the two variable sets. Also such a simple summation weighting system would not intentionally maximize the relationship between the variable sets. CCA is the method of choice in such situations.

In this example, CCA would determine the exact weighting scheme for computing one or more employee scores and for computing one or more performance scores. The specific combination of weights is called a *canonical function* (conceptually, a *function* consists of weights [multiplicative constants] similar to the beta [β] weights that constitute a regression equation), and the score obtained by applying the canonical function coefficients to a set of actual measured scores for a given person is known as a *synthetic score.* CCA determines weighting schemes that create synthetic–latent scores that are maximally correlated. That is, no other possible combination of weighting schemes can be devised that would ultimately lead to a higher correlation between the resulting two *synthetic–latent variables* on a given CCA function.

Heuristic Data Set

An actual data set is used to make this discussion concrete with regard to the basic logic of CCA. The illustration uses scores on five measured–observed variables (i.e., two in one set and three in the other set) from 301 cases from the Holzinger and Swineford (1939, pp. 81–91) data. These scores on ability batteries have classically been used as examples in both popular textbooks (Gorsuch, 1983, *passim*) and computer program manuals (Jöreskog & Sörbom, 1989, pp. 97–104) and thus are familiar to many readers.

The illustrative data involve five variables, each of which is intervally scaled. Other levels of scale can be used in canonical analyses, however, if data are still somewhat normally distributed (e.g., Cooley & Lohnes, 1976, p. 209). As Maxwell (1961) noted, "the theory of canonical variate analysis, widely used with continuous variables, can be employed when the variables are dichotomous" (p. 271; for more technical detail, see Thompson, 1984, pp. 16–18).

One set of measured variables involves scores on three measures of "perceptual abilities": visual perception (VP), lozenges flipped shapes (LFS), and counting groups of dots (CGD). The second, conceptually discrete set of variables measured "academic achievement": word meaning (WM) and mixed math fundamentals (MMF). The syntax commands to conduct the canonical analysis of these variables within the popular SPSS programs, including the Windows versions, are a bit tricky because within SPSS, CCA is run within the MANOVA procedure. The commands for this example are

```
MANOVA VP LFS CGD WITH WM MMF/
PRINT = SIGNIF(MULTIV EIGEN DIMENR)/
DISCRIM = STAN CORR ALPHA (.999)/
```

Table 9.1 presents the matrix of Pearson product–moment correlation coefficients for these data. For example, the interdomain bivariate correlation between VP (a member of the first variable set) and MMF (a member of the second variable set) is .2826. The correlation matrix is "symmetric" about the diagonal. The diagonal contains all *1*s, reflecting the correlation of each measured–observed variable with itself. The correlation between VP and MMF in the first row and last column is reproduced exactly as the correlation between MMF and VP in the first column and last row.

Of course, one could subject the Table 9.1 correlation matrix to a

Table 9.1

Bivariate Correlation Matrix

	Variable				
Variable	**VP**	**LFS**	**CGD**	**WM**	**MMF**
VP	1.0000	0.4407	0.2239	0.3568	0.2826
LFS	0.4407	1.0000	0.1860	0.1977	0.1668
		R11		**R12**	
CGD	0.2239	0.1860	1.000	0.1496	0.3111
WM	0.3568	0.1977	0.1496	1.000	0.4401
		R21		**R22**	
MMF	0.2826	0.1668	0.3111	0.4401	1.0000

Note. VP = visual perception; LFS = lozenges flipped shapes; CGD = counting groups of dots; WM = word meaning; MMF = mixed math fundamentals. The quadrant names are designated in bold; **R11**, for example, is an intradomain quadrant in that both subscripts are the same.

factor analysis to evaluate relationships among the variables. If the measured variables consist of theoretically discrete variable sets, however, or if the variables were measured at chronologically discrete times, such a factor analysis would not honor the view that the measured variables exist within meaningful sets.

Indeed, it is only when the researcher believes the measured–observed variables exist within meaningful variable sets that CCA would be an appropriate analysis. In the present example, the conceptualization of the two variable sets as being discrete seems reasonable. Of course, such classifications are matters of researcher judgment, and even reasonable researchers differ regarding such judgments (just as researchers may reasonably disagree about most aspects of the research endeavor).

CCA computer programs first partition the correlation matrix into quadrants associated with the variable sets, as illustrated in Table 9.1. Note that each quadrant is identified by a boldface **R** with two subscripts. **R11** includes the intradomain correlations between variables that are measures of perceptual abilities (variable set 1), and **R22** includes correlations between variables that are measures of academic achievement (variable set 2). Both **R12** and **R21** include the interdomain correlations between variables from across the two variable sets.

After the correlation matrix is computed, a quadruple-product matrix is then computed from the four quadrants, using the following matrix algebra formula:

$$\mathbf{R22}^{-1}_{2\times2}\ \mathbf{R21}_{2\times3}\ \mathbf{R11}^{-1}_{3\times3}\ \mathbf{R12}_{3\times2} = \mathbf{A}_{2\times2}.$$

This matrix, $\mathbf{A}_{2\times2}$, is then subjected to a principal components analysis, and the results are expressed as standardized weights (Thompson, 1984, pp. 11–14 provides more detail) called standardized canonical function coefficients.

Although further discussion of the mathematical underpinnings of CCA is beyond the scope of this chapter, note that these function coefficients are directly akin to beta (β) weights in regression or the pattern coefficients from exploratory factor analysis. As noted shortly, these function coefficients are one important element within the process of result interpretation.

It may occur to the reader that many statistical analyses invoke weights but that different names are used for the same concepts across techniques (e.g., beta weights vs. pattern coefficients vs. function coefficients and equation vs. factor vs. function). Sometimes it appears that the sole purpose of such misnomers is to confuse graduate students into thinking that these analyses are unrelated to each other rather than all being part of one general linear model (GLM; Fan, 1996, 1997; Knapp, 1978; Thompson, 1991b).

Table 9.2 presents the standardized function coefficients for the current example. The number of functions (sets of weights) in a CCA is always equal to the number of variables in the smaller of the two variable sets (i.e., in this example, two). In Table 9.2, the first canonical function is labeled Function I and the second is labeled Function II. As is always the case, these functions are perfectly uncorrelated with each other and so are the scores on the latent or synthetic variables computed by applying the weights to the observed or measured variables (i.e., here five measured variables). Thus, synthetic–latent variables are never directly measured. Synthetic variables are obtained by applying weights to the measured variables. The synthetic variables are estimates of the latent constructs of interest and are the actual focus of all statistical analyses.

Computing synthetic variable scores is actually a simple matter. For example, for the variable set consisting of three variables, the scores of the first person in the data set were VP, 20; LFS, 3; and CGD, 115. For the second variable set, this person's scores were WM, 9, and MMF, 24. To compute the synthetic variables, raw scores must first be transformed to *z* scores (i.e., scores having a mean of 0 and a standard deviation of 1). The *z*-score equivalents of the first person's five scores were −1.373, −1.658, +.220, −.821, and −.056, respectively.

For the first participant, based on the first canonical function, two synthetic scores are computed. The first synthetic score, PRED1, is for the perceptual abilities variable set. The second synthetic score, CRIT1, is for the cognitive academic achievement variable set. In principal components analysis, this would be analogous to computing factor scores on Factor I; however, because in CCA there are two variable sets, one distinguishes synthetic variables by using two names, namely, PRED1 and CRIT1. If two canonical functions were present, then synthetic variables on the second factor would be labeled PRED2 and CRIT2.

For the first participant, the synthetic PRED1 score can be obtained by applying the standardized function coefficients to the *z* scores for the measured variables. For example, the synthetic score for first participant equals the *z* score for VP (−1.373) multiplied by the standardized function coefficient (0.733), plus the *z* score for LFS (−1.658) multiplied by the standardized function coefficient for LFS (0.091), plus

Table 9.2

Canonical Coefficients in the Format Recommended for CCA Reports

Variable–statistic	Function I: Function	Function I: r_s	Function I: r_s^2	Function II: Function	Function II: r_s	Function II: r_s^2	h^2
VP	0.733	0.879	77.26%	−0.622	−0.462	21.34%	98.61%
LFS	0.091	0.502	25.20%	−0.101	−0.206	4.24%	29.44%
CGD	0.473	0.654	42.77%	0.915	0.756	57.15%	99.93%
Adequacy			48.41%			27.58%	
Rd			8.71%			0.97%	
Rc^2			18.00%			3.50%	
Rd			12.94%			0.98%	
Adequacy			71.88%			28.06%	
WM	0.548	0.825	68.05%	−0.969	−0.565	31.92%	99.98%
MMF	0.629	0.870	75.69%	0.919	0.492	24.21%	99.90%

Note. CCA = canonical correlation analysis; VP = visual perception; LFS = lozenges flipped shapes; CGD = counting groups of dots; Rd = redundancy coefficient for a given variable set; Rc^2 = squared canonical correlation coefficient; WM = word meaning; MMF = mixed math fundamentals.

the z score for CGD (0.220) multiplied by the standardized function coefficient for CGD (0.473). The result of the computation for the first participant yields PRED1 = −1.053.

Similarly, the synthetic–latent variable score CRIT1 for the first participant would be computed as the z score on WM (−0.821) multiplied by the standardized function coefficient for WM (0.548), plus the z score for MMF (−0.056) multiplied by the standardized function coefficient for MMF (0.621). The result of this computation yields CRIT1 = −0.485. Note that all these synthetic scores are themselves in z-score form. That is, PRED1, PRED2, CRIT1, and CRIT2 are all z scores, with a mean of 0 and a standard deviation of 1.

Figure 9.1 presents a scatterplot of the 301 scores on PRED1 and CRIT1. For variables in z-score form, the best-fitting regression line has a slope, ß, that is equal to the correlation of the two synthetic variables (+.4246). This bivariate product–moment correlation coefficient is nothing more or less than the multivariate canonical correlation between the weighted variables in the two variable sets, Rc. Remember, no other possible weighting scheme can produce two synthetic variables that have a higher correlation between them: CCA maximizes Rc.

In Table 9.3, note that PRED and CRIT synthetic scores are uncorrelated with each other except when they are from the same function. As shown, PRED1 and CRIT1 are correlated with each other (the first Rc) and so are PRED2 and CRIT2 (the second Rc), but the remaining correlations among these different synthetic scores are all zero. This establishes that all the synthetic variable scores, except PRED1 with CRIT1 and PRED2 with CRIT2, are "bi-orthogonal." Note that the last row of Table 9.3 presents the bivariate product–moment correlation coefficients involving the four synthetic variables and, for illustrative purposes, one arbitrarily selected observed or measured variable, WM. This facilitates the understanding of a second coefficient (in addition to the standardized function coefficient), called a *structure coefficient,* which is important in all multivariate analyses and in univariate analysis (Thompson, 1997; Thompson & Borrello, 1985).

A structure coefficient (indicated as r_s in Table 9.2) is the bivariate product–moment correlation between scores on an observed or measured variable and scores on a synthetic or latent variable for that measured variable's variable set. Thus, because structure coefficients are correlation coefficients, they range from −1 to +1, inclusively; standardized function coefficients, however, are not usually correlation coefficients and have no definitive boundaries (see Thompson, 1984,

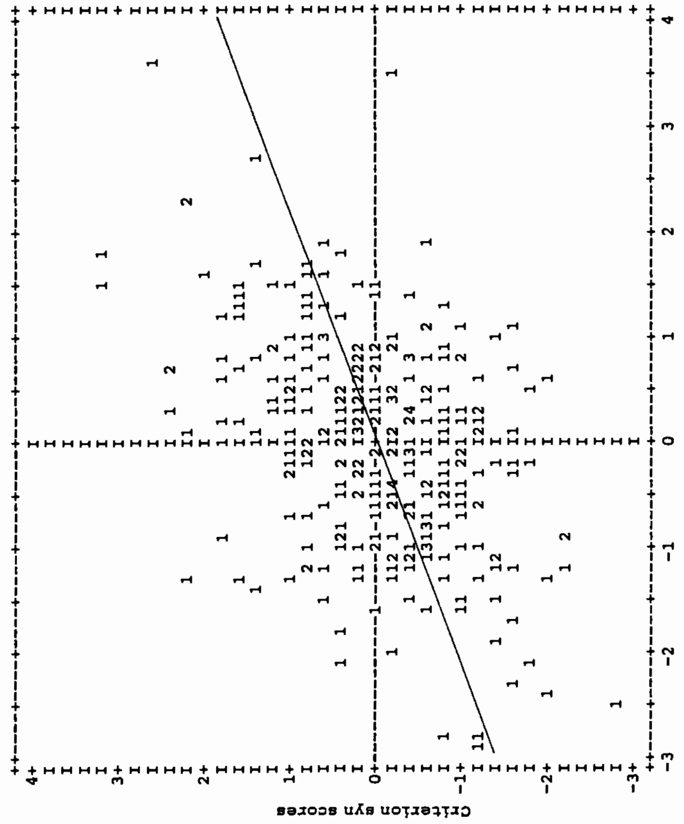

Figure 9.1

Scatterplot of CRIT1 and PRED1 scores.

Table 9.3

Product-Moment Correlation Coefficients Among the Four Synthetic Variables Scores and One Measured Variable

	Variable				
Variable	CRIT1	CRIT2	PRED1	PRED2	WM
CRIT1	1.0000				
CRIT2	0.0000	1.0000			
PRED1	0.4246[a]	0.0000	1.0000		
PRED2	0.0000	0.1863[b]	0.0000	1.0000	
WM	0.8253[c]	−0.5647[d]	0.3504[e]	−0.1052[e]	1.0000

Note. WM = word meaning. [a]RC_I, as reported in Table 9.2. [b]Rc_{II}, as reported in Table 9.2. [c]The structure coefficient for measured variable WM on Function I, as reported in Table 9.2. [d]The structure coefficient for measured variable WM on Function II, as reported in Table 9.2. [e]The index coefficients for measured variable WM (see Thompson, 1984, pp. 30–31).

pp. 21–24). The use of function and structure coefficients are further explained when rubrics for result interpretation are presented.

Table 9.2 also presents other indices that (unlike function and structure coefficients, which are almost always essential to interpret) are at least in some instances important in evaluating and interpreting CCA results. The special circumstances when the following two coefficients are important are detailed in the subsequent section on interpreting results. A common error by less-experienced CCA researchers involves interpreting the following two coefficients when they are in fact irrelevant and do not to be interpreted.

A *canonical adequacy coefficient* indicates how adequately a given function, on average, reproduces the variance of a given set of measured variables. For example, consider the first canonical function and the academic achievement variable set. For WM, the structure coefficient is $r_s = .825$. Here, the squared structure coefficient is $r_s^2 = .6806$. This means that 68.06% of the observed WM variable is useful within this canonical function. Also here, the mean of the squared structure coefficients ([.6806 + .7569]/2 = .7188) is the canonical adequacy coefficient for the first function. Thus, on average, this canonical function reproduces approximately 72% of the variance of the WM and MMF variables. The largest this number can be is 100%, indicating that all the variance of the measured variables in the given set has been reproduced within the function and the associated synthetic variable.

Table 9.2 also presents a *redundancy coefficient* (Rd) for each canonical function and variable set (i.e., four Rds, two for each function). Rd is defined as the product of the canonical adequacy coefficient for a variable set multiplied by the squared canonical correlation, Rc^2 (for further discussion, see Thompson, 1991b). For example, for the academic achievement variable set, the redundancy coefficient for the first canonical function is Rd = .7188 × .1800, or Rd = .1294.

Finally, Table 9.2 also presents *canonical communality coefficients* (h^2), which unlike the previous two coefficients are often important to evaluate when interpreting results. Canonical commonality coefficients are equal to the sum of the squared structure coefficients for a given variable across the canonical functions. For example, for VP, $h^2 = .7726$ (i.e., r_s^2 for Function I) + .2134 (i.e., r_s^2 for Function II) = 0.9861. This indicates that the two synthetic variables, considered together, can reproduce 98.61% of the variance of the VP variable, or conversely, that 98.61% of this measured variable was useful in defining the function.

Note that h^2 was greater than 98% for all measured variables in the example, except for LFS, for which h^2 was less than 30%. This suggests that compared with the other measured variables, the synthetic variables obtained in this CCA had relatively less to do with scores on the LFS variable. Sometimes measured variables with anomalously low communality coefficients may be deleted from the analysis to obtain a more parsimonious solution (see Thompson, 1984, pp. 47–51).

Canonical Correlation Analysis in the General Linear Model

CCA is the most general case of the parametric general linear model (Baggaley, 1981; Fan, 1996; Fornell, 1978; Thompson, 1991b), unless one wishes to delineate an even broader general linear model that also directly takes into account measurement error (Bagozzi, Fornell, & Larcker, 1981; Fan, 1997; Thompson, 1999b). Knapp (1978) demonstrated these views 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" (p. 410). Thus, Knapp's work was cited in a compilation of the seminal methodology publications produced during the past several decades (Thompson & Daniel, 1996).

Saying that CCA constitutes the parametric general linear model, subsuming all other parametric univariate and multivariate analyses, means that the other analyses are special cases of canonical analysis, and that the other analyses (e.g., *t* tests, analysis of variance [ANOVA], regression, MANOVA, descriptive discriminant analysis) can actually be conducted using the logic of CCA (Campbell & Taylor, 1996). This realization is of immense heuristic value for people trying to understand how various analytic methods are related to each other.

Three Realizations

Three important realizations can be extracted from the view of CCA as the general linear model. First, the GLM view forces researchers to understand that all analyses are correlational. Some designs are experimental, but all analyses are correlational, and it is the design (not the analysis) that enables the making of causal inferences.

Too many researchers use OVA methods (ANOVA, ANCOVA [analysis of covariance], MANOVA, and MANCOVA [multivariate analysis of covariance]) because they have come to associate making causal inferences with OVA methods; these erroneous associations are all the more pernicious because the associations tend to be made unconsciously. As Humphreys (1978) emphasized,

> the basic fact is that a measure of individual differences is not an independent variable, 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. (p. 873)

Similarly, Humphreys and Fleishman (1974) noted that categorizing variables in a nonexperimental design using an ANOVA or other OVA 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" (p. 468).

So the use of OVA methods to analyze nonexperimental data can be unnecessary and is usually harmful if researchers discard variance on intervally scaled predictor variables, so that they can use OVA methods (Thompson, 1994c). As Cliff (1987) explained,

> 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. (p. 130)

Second, the GLM view of canonical analysis correctly indicates that all parametric analyses either explicitly or implicitly invoke systems of weights applied to measured variables to produce synthetic variables. Again, these weight systems are often arbitrarily (and confusingly) given different names across different analyses (e.g., beta weights vs. pattern coefficients vs. function coefficients and equation vs. factor vs. function).

The fact that these weights are used to define synthetic–latent variables correctly suggests that it is the synthetic variables that are actually the focus in all analyses. That the synthetic variables are the analytic focus can be seen in the realization that the Rc equals the product–moment *r* between the synthetic variables for a given canonical function.

Third, the fact that all analyses are correlational implies that effect sizes ought to be reported and interpreted in all research studies. No knowledgeable researcher reporting bivariate or multiple correlation coefficients fails to comment on the magnitude of the squared correlation coefficients. Because all analyses are correlational, all researchers reporting *t* test, ANOVA, descriptive discriminant analysis, or any other analysis should always interpret uncorrected (e.g., eta^2) or corrected (e.g., $omega^2$) effect sizes or some other measures of effect size (Snyder & Lawson, 1993). The necessity of reporting such effect sizes is reinforced by the 1994 *Publication Manual of the American Psychological Association*, which notes that "neither of the two types of probability values reflects the importance or magnitude of an effect because both depend on sample size. . . . You are *encouraged* to provide effect-size information" (p. 18, emphasis added).

Unhappily, 11 empirical studies, each of either one or two volumes of 23 different journals, demonstrate that this "encouragement" has been ineffective (Vacha-Haase, Nilsson, Reetz, Lance, & Thompson, 2000). Thompson (1999c) observed that only encouraging effect-size reporting "presents a self-canceling mixed message. To present an 'encouragement' in the context of strict absolute standards regarding the esoterics of author note placement, pagination, and margins is to send the message, 'these myriad requirements count, this encouragement doesn't'" (p. 162). Given (a) a rationale as to why an encouragement is doomed to continued impotence and (b) empirical evidence that the encouragement indeed has been ineffective, editors of several journals have now come to require effect-size reporting (cf. Heldref Foundation, 1997; Murphy, 1997; Thompson, 1994a).

In fact, the recent report of the APA Task Force on Statistical Inference emphasized, "*always* provide some effect-size estimate when reporting a *p* value" (Wilkinson & APA Task Force on Statistical Inference, 1999, p. 599, emphasis added). Later the task force wrote that researchers should

> *always* present effect sizes for primary outcomes. . . . It helps to add brief comments that place these effect sizes in a practical and theoretical context. . . . We must stress again that reporting and interpreting effect sizes in the context of previously reported effects is *essential* to good research. (p. 599, emphasis added)

A Demonstration of the General Linear Model Concept

An illustration is provided to clarify how CCA subsumes other parametric analyses as special cases. The reader can consult other resources (e.g., Fan, 1996; Thompson, 1991b) for more comprehensive demonstrations of these various linkages, including SPSS or SAS syntax programs that conduct the related analytic proofs. Space precludes a complete treatment here, except for the following illustration. Specifically, the fact that CCA subsumes regression as a special case is demonstrated. This relationship is relatively straightforward because both analyses are explicitly correlational.

Figure 9.2 presents the SPSS for Windows regression printout for an analysis in which WM was treated as the dependent variable, whereas scores on VP, LFS, and CGD were used as predictor variables. Figure 9.3 presents the printout from the same analysis conducted using the logic of CCA.

The correlation coefficients from the analysis are the same (multiple R = +.36586 and Rc = +.366), except that the computer programmers arbitrarily elected to report results to different numbers of decimal places. However, the standardized weights (i.e., β and functional coefficients) from the two analyses seem to be different.

Actually, the weights are merely scaled differently, and such differences are not meaningful. Table 9.4 demonstrates how β weights can readily be converted into standardized canonical function coefficients and vice versa. Other resources show how CCA subsumes and can therefore be used to perform *t* tests, ANOVA, ANCOVA, MANOVA, and descriptive discriminant analyses (cf. Thompson, 1991b). In short, all analyses (a) are correlational, (b) invoke weights being applied to measured variables to estimate synthetic variables, and (c) yield variance-accounted-for effect sizes analogous to r^2.

Figure 9.2

```
Equation Number 1    Dependent Variable..   T9    WORD MEANING TEST

Block Number  1.  Method:  Enter      T1       T4       T12

Variable(s) Entered on Step Number
   1..     T12        SPEEDED COUNTING OF DOTS IN SHAPE
   2..     T4         LOZENGES FROM THORNDIKE--SHAPES FLIPPED
   3..     T1         VISUAL PERCEPTION TEST FROM SPEARMAN VPT

Multiple R            .36586
R Square              .13385
Adjusted R Square     .12511
Standard Error       7.17347

Analysis of Variance
                      DF       Sum of Squares       Mean Square
Regression             3         2361.87366          787.29122
Residual             297        15283.21604           51.45864

F =       15.29950        Signif F =  .0000

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

Variable                B         SE B        Beta          T  Sig T

T1               .353005      .066737     .322413      5.289  .0000
T4               .036160      .051249     .042660       .706  .4810
T12              .026311      .021088     .069479      1.248  .2131
(Constant)      1.285475     2.632565                   .488  .6257
```

Abridged SPSS printout showing regression results.

Interpreting Canonical Results

Because all classical parametric analyses are special cases of CCA, the same interpretation strategy can be used for any parametric analyses. The interpretation is approached as a hierarchical, two-stage contingency model. Not to put the matter too technically, two questions are addressed:

1. Do I have anything?
2. Where does what I have originate?

One reaches the second question only if the answer to the first question is *yes*.

In addressing these two questions, the interpretation should be framed within the context of sample size. CCA is a large-sample method, potentially requiring 15–20 participants per measured variable (cf. Barcikowski & Stevens, 1975). When one is interpreting the results, more

Figure 9.3

```
- - - - - - - Eigenvalues and Canonical Correlations - - - - - - -

Root No.    Eigenvalue    Pct.      Cum. Pct.    Canon Cor.    Sq. Cor

   1           .155      100.000    100.000        .366          .134

- - - - - - Standardized canonical coefficients for DEPENDENT variables
            Function No.

Variable         1

T1            .881
T4            .117
T12           .190
```

Abridged SPSS printout showing regression performed as a canonical correlation analysis. Pct. = percent; Cum. Pct. = cumulative percent; Canon Cor. = canonical correlation; Sq. Cor = squared canonical correlation.

Table 9.4

Regression and Canonical Results

	Coefficients						
Variable	β	/	Rc	=	Function	×	R
VP	.322413	/	.366	=	.881	×	.36586
LFS	.042660	/	.366	=	.117	×	.36586
CGD	.069479	/	.366	=	.190	×	.36586

Note. VP = Visual perception; LFS = lozenges flipped shapes; CGD = counting groups of dots. The beta (β) and *R* coefficients came from the Figure 9.2 regression as regression results. The Rc and function coefficients came from the Figure 9.3 regression as canonical results.

confidence can be vested in result stability as sample size is larger. Furthermore, interpretations must be framed in the knowledge that paradoxically one can typically be more confident in the stability of the overall effect size than in conclusions regarding the specific origins of the effect (cf. Thompson, 1990, 1991a).

1. Do I Have Anything?

A researcher can select any combination of three sorts of evidence to address the question, "Do I have anything?" First, the researcher can interpret the statistical significance of the canonical correlation coefficients. Of course, because the calculated probability (p) for a given set of results is highly dependent on sample size, consideration of statistical significance provides limited information. Furthermore, notwithstanding common misconceptions to the contrary, statistical significance tests do *not* evaluate whether the sample results occur in the population or are likely to be replicated in future samples (cf. Cohen, 1994; Thompson, 1996; Snyder & Thompson, 1998; Thompson & Snyder, 1997, 1998).

Second, the researcher can interpret some measure of effect size (see Kirk, 1996). There are many choices, as partially enumerated by Snyder and Lawson (1993). In the canonical case, the squared canonical correlation coefficients can be interpreted, or so-called "adjusted" values of these coefficients (similar to the adjusted R^2 in multiple regression) can be interpreted (Thompson, 1990). The interpretation of effect sizes invokes the researcher's subjective judgment of the noteworthiness of results. This intimidates some researchers, who become afraid that they will say or write judgments with which others will disagree.

Some such researchers seek an atavistic escape from these fears by trying to use statistical significance tests as an objective standard for whether results are noteworthy. As Thompson (1993) explained, however, "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). Clearly, research is in part an inherently subjective business, and researchers must inescapably make the necessary judgments.

There are no clear boundaries regarding what effect sizes are noteworthy. The judgment is made at the nexus of the researcher's value system and the substantive focus of the research. For example, the variance-accounted-for effect size associated with the effects of smoking on longevity is reportedly about 2% (cf. Gage, 1978, p. 21). This is a small number, but most people take any definitive decrease in their potential days on earth seriously and thus deem the effect noteworthy. Others, of course, may feel that the pleasures of smoking outweigh the possible costs, especially because most effects occur "on the average" (i.e., the relation between smoking and longevity does not apply equally to each individual who smokes).

Third, researchers can evaluate whether their effects are replicable. If science is about the business of identifying relationships that recur under specified conditions, so that knowledge is cumulated, then the replicability issue is a critical one. Because statistical significance tests do *not* evaluate result replicability (cf. Thompson, 1996; Snyder & Thompson, 1998; Thompson & Snyder, 1997, 1998), two classes of methods can be invoked to evaluate this important issue.

Ultimately, the best way to evaluate result replicability is to replicate a given study. This represents a so-called "external" replicability analysis because a completely new sample is drawn. For pragmatic reasons, however, most researchers feel unable to replicate all their studies (e.g., tenure or promotion or merit-raise decisions are approaching, one's spouse has threatened to abandon the doctoral student if the dissertation is not completed by date certain). In such cases, researchers can instead invoke so-called "internal" replicability analyses to try to evaluate replicability. These internal analyses are not as accurate as true replication efforts but are certainly better than no attempt to evaluate result replicability.

There are three basic logics for such empirical internal replicability analyses: (a) cross-validation, (b) the jackknife, and (c) the bootstrap. As Thompson (1996) explained, "basically, the methods combine the subjects in hand in different ways to determine whether results are stable across sample variations, i.e., across the idiosyncrasies of individuals which make generalization in social science so challenging" (p. 29). An expanded discussion of these three methods is beyond the scope of this treatment. Thompson (1994b) provided an overview. Crossman (1996) described canonical cross-validation. Thompson (1995) described a computer program, CANSTRAP, that implements a bootstrap CCA.

2. Where Does What I Have Originate?

Given that a decision has been made that the canonical results are noteworthy, the question then arises, "Where does what I have originate?" Many canonical coefficients can be interpreted to address this question (Thompson, 1984), but the two coefficients of primary importance are the standardized canonical function coefficients and the structure coefficients. Both must be interpreted (for a contrary view, see Harris, 1989).

Variables with function coefficients of zero on a given function clearly have no effect on defining the synthetic variables associated with the function. That is, multiplying the scores on an observed variable by the multiplicative constant of zero statistically kills that observed variable.

But it is critical to remember that a given observed variable can get a function coefficient of zero for either of two reasons: (a) The measured variable has nothing to contribute with regard to the relationship between the variable sets; or (b) whatever the measured variable has to contribute with regard to the relationship between the variable sets (which may even be quite a lot), one or more other variables also contain this variance or information and the given measured variable is arbitrarily denied any credit for providing this information. Obviously, the same thing can occur in regression, but certainly no reasonable researcher believes that in regression measured variables with ß weights of zero are inherently useless (Thompson & Borrello, 1985).

So measured variables with near-zero function coefficients may or may not be useless in creating the detected effects, and this ambiguity can only be resolved by consulting structure coefficients. As Meredith (1964) suggested, "if the variables within each set are moderately intercorrelated the possibility of interpreting the canonical variates by inspection of the appropriate regression weights [function coefficients] is practically nil" (p. 55). Similarly, Kerlinger and Pedhazur (1973) argued that, "a canonical correlation analysis also yields weights, which, theoretically at least, are interpreted as regression β weights. These weights [function coefficients] appear to be the weak link in the canonical correlation analysis chain" (p. 344). Levine (1977) was even more emphatic:

> I specifically say that one *has* to do this [interpret structure coefficients] since I firmly believe as long as one wants information about the nature of the canonical correlation relationship, not merely the computation of the [synthetic function] scores, one must have the structure matrix. (p. 20, emphasis in original)

One can determine that a measured variable is arbitrarily being denied credit for providing predictive information on a given function when the variable has a near-zero function coefficient but has a structure coefficient that is large in absolute value (i.e., approaching −1 or +1). For example, Sexton et al. (1988) presented a canonical analysis in which one variable had a function coefficient of +.02 on Function I but a structure coefficient of +.89.

Measured variables can also have near-zero structure coefficients but function coefficients that are large in absolute magnitude. Such results indicate the presence of so-called "suppression" effects explained in most regression textbooks. Horst (1966) provided the classic example. An accessible explanation is presented by Lancaster (1999).

Only measured–observed variables that have function and structure coefficients that are both near zero contribute nothing to defining a given function and its associated effects. Only measured variables with all function coefficients near zero and canonical communality coefficients near zero contribute nothing to the canonical solution as a whole.

A hybrid case, however, in which only one set of coefficients could be correctly interpreted should be noted (Thompson, 1984, pp. 22–23). The function and structure coefficient matrices for a given variable set are identical when the measured variables in the set are perfectly uncorrelated, as would be the case, for example, if the variables in a set consisted of factor scores on orthogonally rotated principal components.

A Brief Illustrative Interpretation

The Table 9.2 results can be interpreted to provide a model of the recommended interpretive procedure. First, the noteworthiness of detected effects is evaluated. Here, the squared canonical correlation coefficients (Rc^2) are 18.00% and 3.50%. One can presume that the researcher deemed only the first canonical function noteworthy, given personal values and the substantive context of the study.

Although it is important to judge which, if any, of the Rc^2 values (and the associated canonical functions) are deemed noteworthy to be retained for interpretation, a thorough researcher also assesses whether a similar outcome would result were the analysis repeated with an independent random sample. There are various ways to accomplish this (see Crossman, 1996; and Thompson, 1990, 1995). In reading applications of CCA in journals, look for whether the researcher has attempted to estimate replicability.

In the present context, for example, presume that the researcher first evaluated the replicability of the detected effect initially by applying the commonly used Ezekiel (1930) regression R^2 correction formula to compute an adjusted Rc^2 (Thompson, 1990) for Function I and considered the "shrunken" Rc^2 still noteworthy. Next, presume the researcher conducted either a canonical cross-validation (Crossman, 1996) or a canonical bootstrap analysis (Thompson, 1995) and empirically determined from this "internal" replicability analysis that the detected effect was reasonably stable across variations in the sample.

At this juncture, the researcher addresses the issue of where the effects originate. All five measured variables have function and structure coefficients on Function I that are all positive, as reported in Table 9.2. Thus, all the observed variables are positively related with the underlying synthetic variables on this function.

On the one hand, the LFS variable has a function coefficient of +.091. This variable has a structure coefficient of +.502 on the function, however, and this indicates that the other observed variables on this function are arbitrarily getting credit for the predictive power that LFS brings to the table.

On the other hand, LFS has a disproportionately low squared structure coefficient (25.20%) as against the remaining coefficients on this function, which range from 42.77% to 77.26%. Thus, the overarching pattern is that higher scores on the two perceptual tasks, VP and CGD, are predictive of higher scores on both achievement measures, WM and MMF scores.

These results can also be viewed from another intriguing perspective. Note that both the WM and MMF variables have function and structure coefficients that are both positive on Function I. The WM and MMF variables have function and structure coefficients with opposite signs on Function II. Function I can therefore be viewed as explaining aspects of positive covariation between the two measured achievement variables, which is most noteworthy (Rc^2 = 18.00%). In contrast, Function II evaluates aspects of differences between the two achievement variables, although here such differences are not very noteworthy (Rc^2 = 3.50%). This contrast illustrates the richness of canonical analysis, which, in this example, can be used to explore prediction of both the commonalities among and the differences between given measured variables.

Two Analytic Pitfalls to Avoid

For researchers interested in conducting CCA and for readers who must interpret the work of others who use this method, two potential analytic pitfalls can lead to erroneous conclusions and should be avoided. First, there is the problem of incorrect interpretation of the statistical tests performed in CCA. Most statistical packages that perform CCA produce multiple functions and multiple test statistics, but only the last test statistic is a test of the effect size associated with a single function.

For example, imagine a hypothetical problem in which there are three canonical correlations. In this instance, most programs provide three sets of test statistics, and this may lead one to believe that each test statistic can be used to evaluate each correlation coefficient independent of the others; however, this is not true. In this case, the first test statistic is used to evaluate all three canonical correlations (and their squared values as well); the second test statistic is appropriate for evaluating the second and third coefficients as a set; and only the third test statistic is a test of a single correlation coefficient, that is, the third and final canonical correlation. Most computer programs do not test each single canonical correlation, except the last one, and conducting tests of individual canonical correlations, other than the last one, is not a straightforward matter (see Stevens, 1992, pp. 411–412).

In the context of the illustrative problem, for example, that involved the two Rc^2 values reported in Table 9.2, most computer programs would report two test statistics (not shown in Table 9.2). For this example, the computer output reported an *F* statistic of 12.25 (degrees of freedom [*df*] = 6,592, $p < .001$) for "roots 1 to 2" and an *F* of 5.34 (*df* = 2,297, $p < .005$) for "roots 2 to 2." Only the last test is an appropriate test of a single canonical correlation (Rc = .1863, Rc^2 = 3.50%).

Second, researchers should generally avoid interpreting canonical redundancy coefficients (Rd; see Table 9.2). Stewart and Love (1968) conceptualized these statistics, and Miller (1975) developed a partial test distribution to test the statistical significance of redundancy coefficients.

Although Cooley and Lohnes (1971, p. 170) suggested that redundancy coefficients have great value, more recent thinking suggests that the interpretation of redundancy coefficients does not make much sense in a conventional canonical analysis. As Cramer and Nicewander (1979) clearly established, redundancy coefficients are not truly multivariate (see also Thompson, 1988). This is important to note because univariate results are generally not useful in interpreting multivariate effects.

Furthermore, canonical analyses optimize Rc (not Rd); it seems contradictory to emphasize statistics not optimized in a given analysis. If one were interested in redundancy coefficients, then a redundancy analysis should be performed rather than a CCA (see Thompson, 1984).

Indeed, a redundancy coefficient can only equal one when (a) the synthetic variables for the function represent all the variance of every variable in the set (i.e., all squared structure coefficients are one), and (b) the squared Rc also equals one. Such an outcome would be rare. In short, redundancy coefficients are useful only to test outcomes that rarely occur and that may even be unexpected (Thompson, 1984). Nevertheless, there can be exceptions to this rule, such as perhaps in multivariate test–retest reliability or multivariate concurrent validity studies, where one might expect redundancy coefficients to approach one. Sexton et al. (1988) reported just such an exception.

Conclusion

CCA can be a useful analytic tool, as noted previously, but interpreting canonical results may challenge even seasoned analysts. As Thompson (1980) noted,

> [one] reason why the technique is [too] rarely used involves the difficulties which can be encountered in trying to interpret canon-

> ical results. . . . The neophyte student of [CCA] may be overwhelmed by the myriad coefficients which the procedure produces. . . . [But CCA] produces results which can be theoretically rich, and if properly implemented the procedure can adequately capture some of the complex dynamics involved in educational reality. (pp. 1, 16–17)

Such difficulties can be mitigated by following the admonitions suggested in this chapter. As with most analytic methods, real understanding is best facilitated by practice in the context of actual analytic problems of intrinsic interest to a given researcher.

Suggestions for Further Reading

A good starting point for further reading would be Thompson (1991b), followed by Crossman (1996). Stevens (1992; or other editions of his book) provided more comprehensive treatments of canonical analyses. Next, the comprehensive and more technical treatment in Thompson (1984) would be useful. With regard to statistical testing in this and in other contexts, Cohen (1994), Kirk (1996), and Thompson (1996) are all strongly recommended.

Glossary

CANONICAL CORRELATION COEFFICIENT (Rc) The Pearson product–moment correlation between the two sets of *synthetic variable* scores computed for a given canonical function.

COMMUANLITY COEFFICIENT (h^2) The proportion or percentage of variance in a measured variable that is useful in defining the canonical solution; conversely, the proportion or percentage of variance in a *measured variable* that the CCA solution can reproduce.

EFFECT SIZE The measures of magnitudes of effect or relationship that can and should be calculated in all studies, which generally fall into two major classes: (a) variance-accounted-for effect sizes analogous to r^2 (e.g., Rc^2) and (b) standardized mean differences (see Kirk, 1996; Snyder & Lawson, 1993).

EXPERIMENTWISE ERROR RATE ($\alpha_{experimentwise}$) The probability of making one or more Type I errors in a set of hypothesis tests conducted in a single study, ranging from a minimum of ($\alpha_{testwise}$) to a maximum of $1 - (1 - \alpha_{testwise})^k$, where k is the number of hypotheses tested (see Thompson, 1994d).

EXTERNAL REPLICABILITY ANALYSIS An analysis evaluating result replicability in which new data are collected to determine the degree to which (a) the same effect sizes occur and (b) the effects originate with the same *measured variables.*

FUNCTION The set (in some analyses called "equation" or "factor") of weights (e.g., regression β weights, factor pattern coefficients, canonical function coefficients) applied to the *measured variables* to yield scores on synthetic variables (e.g., regression predicted Y scores, factor scores, canonical or discriminant function scores).

FUNCTION COEFFICIENT The multiplicative constant or *weight* applied to a given *measured variable* as part of the calculation of scores on *synthetic variables*; the weights are standardized if the measured variables to which they are applied are in *z*-score form.

GENERAL LINEAR MODEL (GLM) The concept that CCA subsumes all classical parametric methods (from *t* tests through MANOVA and descriptive discriminant analysis) as special cases and that therefore all analyses (a) are correlational, (b) invoke *weights* being applied to *measured variables* to estimate *synthetic variables,* and (c) yield variance-accounted-for effect sizes analogous to r^2 (see Knapp, 1978; Thompson, 1991b).

INTERDOMAIN CORRELATION The bivariate correlation between scores on two variables, both of which are members of two different conceptually discrete variable sets.

INTERNAL REPLICABILITY ANALYSIS An analysis (e.g., cross-validation, jackknife, or bootstrap) attempting to evaluate result replicability using the data in hand, without a true replication, thus resulting in a somewhat positively biased estimate of replicability (see Thompson, 1993, 1994b, 1995).

INTRADOMAIN CORRELATION The bivariate correlation between scores on two variables, both of which are members of a single conceptually discrete variable set.

MEASURED–OBSERVED VARIABLE A variable for which scores are derived by direct measurement by the researcher, as opposed to by applying *weights* to other variables.

REDUDANCY COEFFICIENT (Rd) A canonical coefficient in a squared metric that is not multivariate and that is useful in CCA only in unusual cases in which a "g" (general) function with a perfect effect size ($Rc^2 = 100\%$) is expected.

STRUCTURE COEFFICIENT (r_S) The Pearson product–moment correlation, which should be reported and interpreted in all CCA analyses, between the scores on a given *measured variable* and the *synthetic variable* scores on a given function for the variable set to which the measured variable belongs.

SYNTHETIC–LATENT VARIABLE Estimates of latent constructs, and the actual focus of all statistical analyses, computed by applying *weights* to the *measured variables* (e.g., regression predicted Y scores, factor scores, discriminant or canonical function scores).

TESTWISE ERROR RATE ($\alpha_{testwise}$) The probability of making a Type I error in the test of a single hypothesis.

WEIGHT The multiplicative constants (e.g., regression β weights, factor pattern coefficients, canonical *function coefficients*) applied to the *measured variables* to yield scores on *synthetic variables* (e.g., regression predicted Y scores, factor scores, discriminant or canonical function scores).

References

American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.

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

Bagozzi, R. P., Fornell, C., & Larcker, D. F. (1981). Canonical correlation analysis as a special case of a structural relations model. *Multivariate Behavioral Research, 16,* 437–454.

Barcikowski, R. S., & Stevens, J. P. (1975). A Monte Carlo study of the stability of canonical correlations, canonical weights and canonical variate–variable correlations. *Multivariate Behavioral Research, 10,* 353–364.

Campbell, K. T., & Taylor, D. L. (1996). Canonical correlation analysis as a general linear model: A heuristic lesson for teachers and students. *Journal of Experimental Education, 64,* 157–171.

Chastain, R. L., & Joe, G. W. (1987). Multidimensional relationships between intellectual abilities and demographic variables. *Journal of Educational Psychology, 79,* 323–325.

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

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49,* 997–1003.

Cooley, W. W., & Lohnes, P. R. (1971). *Multivariate data analysis.* New York: Wiley.

Cooley, W. W., & Lohnes, P. R. (1976). *Evaluation research in education.* New York: Irvington.

Cramer, E. M., & Nicewander, W. A. (1979). Some symmetric, invariant measures of multivariate association. *Psychometrika, 44,* 43–54.

Crossman, L. L. (1996). Cross-validation analysis for the canonical case. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 4, pp. 95–106). Greenwich, CT: JAI Press.

Dunst, C. J., & Trivette, C. M. (1988). A family systems model of early intervention with handicapped and developmentally at-risk children. In D. R. Powell (Ed.), *Parent education as early childhood intervention: Emerging directions in theory, research, and practice* (pp. 131–179). Norwood, NJ: Ablex.

Estabrook, G. E. (1984). A canonical correlation analysis of the Wechsler Intelligence Scale for Children–Revised and the Woodcock–Johnson Tests of Cognitive Ability in a sample referred for suspected learning disabilities. *Journal of Educational Psychology, 76,* 1170–1177.

Ezekiel, M. (1930). *Methods of correlational analysis.* New York: Wiley.

Fan, X. (1996). Canonical correlation analysis as a general analytic model. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 4, pp. 71–94). Greenwich, CT: JAI Press.

Fan, X. (1997). Canonical correlation analysis and structural equation modeling: What do they have in common? *Structural Equation Modeling, 4,* 65–79.

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.

Fowler, P. C., & Macciocchi, S. N. (1986). WAIS-R factors and performance on the Lauria-Nebraska's Intelligence, Memory, and Motor Scales: A canonical model of relationships. *Journal of Clinical Psychology, 42,* 626–635.

Fuqua, D. R., Seaworth, T. B., & Newman, J. L. (1987). The relationship of career indecision and anxiety: A multivariate examination. *Journal of Vocational Behavior, 30,* 175–186.

Gage, N. L. (1978). *The scientific basis of the art of teaching.* New York: Teachers College Press.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Harris, R. J. (1989). A canonical cautionary. *Multivariate Behavioral Research, 24,* 17–39.

Heldref Foundation. (1997). Guidelines for contributors. *Journal of Experimental Education, 65,* 95–96.

Holzinger, K. J., & Swineford, F. (1939). *A study in factor analysis: The stability of a bi-factor solution* (No. 48). Chicago: University of Chicago.

Horst, P. (1961). Generalized canonical correlations and their applications to experimental data. *Journal of Clinical Psychology, 26,* 331–347.

Hotelling, H. (1935). The most predictable criterion. *Journal of Experimental Psychology, 26,* 139–142.

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.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications* (2nd ed.). Chicago: SPSS.

Kerlinger, F. N.. & Pedhazur, E. J. (1973). *Multiple regression in behavioral research.* New York: Holt, Rinehart, & Winston.
Kirk, R. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56,* 746–759.
Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. *Psychological Bulletin, 85,* 410–416.
Krus, D. J., Reynolds, T. S., & Krus, P. H. (1976). Rotation in canonical variate analysis. *Educational and Psychological Measurement, 36,* 725–730.
Lancaster, B. P. (1999). Defining and interpreting suppressor effects: Advantages and limitations. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5, pp. 139–148). Stamford, CT: JAI Press.
Levine, M. S. (1977). *Canonical analysis and factor comparison.* Beverly Hills, CA: Sage.
Maxwell, A. E. (1961). Canonical variate analysis when the variables are dichotomous. *Educational and Psychological Measurement, 21,* 259–271.
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.
Meredith, W. (1964). Canonical correlations with fallible data. *Psychometrika, 29,* 55–65.
Miller, J. K. (1975). The sampling distribution and a test for the significance of the bimultivariate redundancy statistic: A Monte Carlo study. *Multivariate Behavior Research, 10,* 233–244.
Murphy, K. R. (1997). Editorial. *Journal of Applied Psychology, 82,* 3–5.
Pitts, M. C., & Thompson, B. (1984). Cognitive styles as mediating variables in inferential comprehension. *Reading Research Quarterly, 19,* 426–435.
Sexton, J. D., McLean, M., Boyd, R. D., Thompson, B., & McCormick, K. (1988). Criterion-related validity of a new standardized developmental measure for use with infants who are handicapped. *Measurement and Evaluation in Counseling and Development, 21,* 16–24.
Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *Journal of Experimental Education, 61,* 334–349.
Snyder, P. A., & Thompson, B. (1998). Use of tests of statistical significance and other analytic choices in a school psychology journal: Review of practices and suggested alternatives. *School Psychology Quarterly, 13,* 335–348.
Stevens, J. (1992). *Applied multivariate statistics for the social sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
Stewart, D. K., & Love, W. A. (1968). A general canonical correlation index. *Psychological Bulletin, 70,* 160–163.
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. (1980, April). *Canonical correlation: Recent extensions for modelling educational processes.* Paper presented at the annual meeting of the American Educational Research Association, Boston, MA. (ERIC Document Reproduction Service No. ED 199 269)
Thompson, B. (1984). *Canonical correlation analysis: Uses and interpretation.* Newbury Park, CA: Sage.
Thompson, B. (1988, 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, LA. (ERIC Document Reproduction Service No. ED 295 957)
Thompson, B. (1989, August). *Applications of multivariate statistics: A bibliography of canonical correlation analysis studies.* Paper presented that the annual meeting of the American Psychological Association, New Orleans, LA. (ERIC Reproduction Service No. ED 311 070)
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. (1991a). Invariance of multivariate results. *Journal of Experimental Education, 59,* 367–382.
Thompson, B. (1991b). A primer on the logic and use of canonical correlation analysis. *Measurement and Evaluation in Counseling and Development, 24*(2), 80–95.
Thompson, B. (1993). The use of statistical significance tests in research: Bootstrap and other alternatives. *Journal of Experimental Education, 61,* 361–377.
Thompson, B. (1994a). Guidelines for authors. *Educational and Psychological Measurement, 54,* 837–847.
Thompson, B. (1994b). The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results. *Journal of Personality, 62*(2), 157–176.
Thompson, B. (1994c). Planned versus unplanned and orthogonal versus nonorthogonal contrasts: The neo-classical perspective. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 3, pp. 3–27). Greenwich, CT: JAI Press.
Thompson, B. (1994d, February). *Why multivariate methods are usually vital in research: Some basic concepts.* Paper presented as a featured speaker at the biennial meeting of the Southwestern Society for Research in Human Development (SWSRHD), Austin, TX. (ERIC Document Reproduction Service No. ED 367 687)
Thompson, B. (1995). Exploring the replicability of a study's results: Bootstrap statistics for the multivariate case. *Educational and Psychological Measurement, 55,* 84–94.
Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher, 25*(2), 26–30.
Thompson, B. (1997). The importance of structure coefficients in structural equation modeling confirmatory factor analysis. *Educational and Psychological Measurement, 57,* 5–19.
Thompson, B. (1999a, April). *Common methodology mistakes in educational research, revisited, along with a primer on both effect sizes and the bootstrap.* Invited address presented at the annual meeting of the American Educational Research Association, Montreal, Canada. (ERIC Document Reproduction Service No. ED 429 110)
Thompson, B. (1999b). Five methodology errors in educational research: A pantheon of statistical significance and other faux pas. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5, pp. 23–86). Stamford, CT: JAI Press.
Thompson, B. (1999c). Journal editorial policies regarding statistical significance tests: Heat is to fire as p is to importance. *Educational Psychology Review, 11,* 157–169.
Thompson, B., & Borrello, G. M. (1985). The importance of structure coefficients in regression research. *Educational and Psychological Measurement, 45,* 203–209.
Thompson, B., & Daniel, L. G. (1996). Seminal readings on reliability and validity: A "hit parade" bibliography. *Educational and Psychological Measurement, 56,* 741–745.
Thompson, B., & Snyder, P. A. (1997). Statistical significance testing practices in the *Journal of Experimental Education. Journal of Experimental Education, 66,* 75–83.
Thompson, B., & Snyder, P. A. (1998). Statistical significance and reliability analyses in recent *JCD* research articles. *Journal of Counseling and Development, 76,* 436–441.
Vacha-Haase, T., Nilsson, J. E., Reetz, D. R., Lance, T. S., & Thompson, B. (2000). Reporting practices and APA editorial policies regarding statistical significance and effect size. *Theory & Psychology, 10,* 413–425.

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594–604. (Available online http://www.apa.org/journals/amp/amp548594.html)

Wood, D. A., & Erskine, J. A. (1976). Strategies in canonical correlation with application to behavioral data. *Educational and Psychological Measurement, 36,* 861–878.

Zakaahi, W. R., & Duran, R. L. (1982). All the lonely people: The relationships among loneliness, communication competence, and communication anxiety. *Communication Quarterly, 30,* 203–209.

10 Repeated Measures Analyses: ANOVA, MANOVA, and HLM

Kevin P. Weinfurt

Consider the following designs:

- *Example 1. Cognitive Psychology.* A psychologist was interested in how people's abilities to recognize stimuli are impaired when their attention is diverted. Each research participant indicated as soon as he or she identified a word flashed quickly on a computer screen; the experimenter measured the time between the presentation of the word and the participant's reaction. For each participant, this was conducted under three conditions. In the "full attention" condition, the participant performed the word identification task in silence. In the "distraction" condition, the participant performed the task while an audio recording of a man reading a story was played. In the "divided attention" condition, the participants performed the task while listening to another story. This time, however, the participants were required to count the number of times the reader used the word *but.* For each participant, the experimenter calculated the mean reaction time for each of the three conditions.
- *Example 2. Clinical Psychology.* A researcher wanted to examine the efficacy of cognitive-behavioral therapy (CBT) compared with treatment as usual for patients with major depression who come to primary care clinics. Patients screening positive for depression were randomized to either the CBT or control group. A baseline assessment of depression was made using the Beck Depression

I am grateful to Paul Yarnold and Jean Lennon Weinfurt for their helpful comments on an earlier draft of this chapter.

READING AND UNDERSTANDING MORE MULTIVARIATE STATISTICS

EDITED BY
LAURENCE G. GRIMM
AND PAUL R. YARNOLD

AMERICAN PSYCHOLOGICAL ASSOCIATION
WASHINGTON, DC