AD-A079 367 TEXAS CHRISTIAN UNIV FORT WORTH INST OF BEHAVIORAL R--ETC F/G 5/10
TESTS OF PARALLELISM IN SUBGROUPING MODERATOR ANALYSIS AND SEQU--ETC(U)
NOV 79 L R JAMES, G W JOE, D M IRONS N00014-77-C-0123

UNCLASSIFIED IBR-79-15 NL

1 OF 1
AD
A079367

END
DATE
FILMED
2-80
DDC

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	IBR Report No. 79-15
2. GOVT ACCESSION NO.	
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle)	Tests of Parallelism in Subgrouping Moderator Analysis and Sequential Moderation
5. TYPE OF REPORT & PERIOD COVERED	Technical Report
6. PERFORMING ORG. REPORT NUMBER	IBR - 79-15
7. AUTHOR(s)	Lawrence R. James, George W. Joe, and Dennis M. Irons
8. CONTRACT OR GRANT NUMBER(s)	N00014-77-C-0123
9. PERFORMING ORGANIZATION NAME AND ADDRESS	Institute of Behavioral Research Texas Christian University Fort Worth, Texas 76129
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	NR 170-840
11. CONTROLLING OFFICE NAME AND ADDRESS	Organizational Effectiveness Research Programs Office of Naval Research (Code 452) Arlington, Virginia 22217
12. REPORT DATE	Nov [redacted] 79
13. NUMBER OF PAGES	42
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	
15. SECURITY CLASS. (of this report)	Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Dynamic Criteria
Moderator
Parallelism
Predictor
Regression Slopes
Repeated Measurements
Sequential Moderation
Subgrouping

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Analytic procedures are presented which cast subgrouping moderator analysis and sequential moderator analysis in the roles of tests of parallelism of regression slopes, planes, and hyperplanes. With respect to subgrouping moderator analysis, the procedures include tests for comparing multiple independent subgroups, multiple predictors, and one or more criteria in one overall analysis. With respect to sequential moderation, the procedure addresses a test for comparing predictor-criterion relationships for one set of measurements on multiple predictors and repeated measurements on a criterion.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The application of the sequential moderation test to issues associated with dynamic criteria is discussed.

Tests of Parallelism in Subgrouping Moderator Analysis and Sequential Moderation

Subgrouping moderator analysis has been employed extensively in Industrial/Organizational psychology. This procedure involves typically (a) the sorting of a heterogeneous sample into more homogeneous subgroups, where the variable used for sorting purposes is referred to as a moderator (e.g., male versus female; high, middle, low socioeconomic status), and (b) comparisons of relationships, based on variables other than the moderator, among the subgroups (Zedeck, 1971). A significant difference in relationships among the subgroups suggests that the magnitude of any particular relationship is dependent on the subgroup in which it was calculated. Given significant differences, it is often said that the moderator "moderated" the relationships.

Another form of moderation is sequential moderation (cf. Weitz, 1966; Zedeck, 1971). An illustration of sequential moderation is the comparison of predictor-criterion relationships for the same subjects, given the same predictor scores and repeated measurements on a criterion. For example, it might be found that the selection tests that predict job performance six months after hire are not the same tests as those that predict job performance twelve months after hire. Results such as these suggest that the magnitude of a predictor-criterion relationship for a particular predictor is dependent on the time of measurement of the criterion. The moderator in this illustration is time, although the true moderator may be changes in organizational demands or abilities used (Smith, 1976), and the analytic question is whether predictor-criterion relationships change (differentially)

A

as a function of time. Sequential moderation has not been employed frequently, although its potential contribution to longitudinal studies of psychological processes is clearly evident.

The objective of this article is to present analytic procedures for subgrouping moderator analysis and sequential moderation based on tests of "parallelism of regressions." With respect to subgrouping moderator analysis, the question addressed is whether unstandardized regression weights (*b*-weights), determined by separate regressions of one or more criteria on one or more predictors in *each* subgroup, differ significantly among the subgroups. Procedures are discussed also for multiple comparisons (when more than two subgroups are involved) and for ascertaining the relative contribution of a predictor to the overall differences in *b*-weights among the subgroups. With respect to sequential moderation, a test is presented for assessing the equality of *b*-weights over time, given the same subjects, same predictor scores, and repeated measurements on a criterion variable. Analytic procedures are discussed separately for subgrouping moderator analysis and sequential moderation. Empirical illustrations are presented for each procedure.

Subgrouping Moderator Analysis

Need for New Approaches

A review of recent articles from the job characteristic and role perception literatures illustrates problems associated with subgrouping moderator analytic procedures presently in use.[1] Figure 1 displays an analytic design not unlike many found in these studies/reviews, where *A* is the moderator variable that defines the subgroups (e.g., urban versus rural background). In Figure 1, the typical statistical procedure is to compare, separately, pairwise correlation coefficient (e.g., r_{111} vs. r_{112}, r_{121} vs. r_{122}, and so forth), which in this illustration would result in a series of six univariate *t*-tests (three for each criterion). Results of the six tests are then usually discussed with respect to whether *A* is a moderator for (a) some predictor-criterion relationships but not other predictor-criterion relationships in regard to *each* criterion, and (b) overall predictor-criterion relationships for some criteria but not others (e.g., *A* moderates *X*, *Y* relationships for Y_1 but not Y_2). In cases where more than two subgroups are involved, the approaches have been either to contrast only two subgroups, typically extreme subgroups (e.g., top vs. bottom quartiles), or to contrast all possible pairs of subgroups. In either case, the analysis is again based on a series of univariate *t*-tests.[2]

Insert Figure 1 about here

Problems with the analytic procedures described above include, but are not necessarily limited to, the following.

(1) It has been well-documented that comparisons of correlation coefficients across independent subgroups is a potentially hazardous procedure if standard deviations on the variables differ among subgroups (cf. Blalock, 1964; Gulliksen & Wilks, 1950; Linn, 1978; Tukey, 1964). The preferred approach is to employ *b*-weights (cf. Linn, 1978; Tukey, 1964). However, when multiple, correlated predictors are involved, a simple conversion of *r*-coefficients to bivariate *b*-weights will not suffice because the predictors do not operate independently. The same point applies to criteria if the criteria are correlated. Failure to consider relationships among predictors and criteria not only may lead to overestimation/underestimation of the extent of moderator

effects, but also fails to capitalize on the increased statistical power that is potentially achievable by employing multivariate procedures (Winer, 1971).

(2) It is straightforward that a series of univariate *t*-tests compounds the probability of making a Type I error, especially given that the same criterion data are employed in more than one test.

(3) Comparisons based on extreme subgroups (e.g., high vs. low quartiles on *A*) may suffer not only from the two problems above, but also from a drop in statistical power resulting from using only part of the overall sample (cf. Schmidt & Hunter, 1978), as well as the inability to generalize findings to all the subpopulations.

(4) Comparisons of all possible combinations of subgroups may be vulnerable to problems discussed in points (1) through (3) and may further compound the Type I error as a result of the increased number of *t*-tests conducted.

In summary, since the problems cited are representative of common practice in the job characteristic and role perception literatures [as well as in other literatures such as differential validity (cf. Katzell & Dyer, 1977)], it would not be unreasonable to expect that these literatures contain both Type I and Type II errors. If anything, we have probably understated the case inasmuch as conclusions are often drawn regarding the efficacy of different moderators, again without the benefit of significance tests.

Without belaboring these issues further, corrective actions for the four problems discussed above are proposed. The corrective actions consist of analytic procedures in which controls are effected for Type I errors, correlations among predictors (and criteria) are taken into account, statistical power is believed to be increased, and a basis is provided for contrasting two or more subgroups simultaneously.

Proposed Analytic Procedures

Univariate tests. An illustration of the first model to be addressed is presented in Figure 2. In this figure, there are K independent subgroups (i.e., A defines K independent subgroups, which is represented statistically by $k = 1, \ldots, K$), J predictors ($j = 1, \ldots, J$), and one criterion (Y_1). The b_{1jk} are are b-weights, determined by separate multiple regressions of Y_1 on the J predictors in each of the subgroups. (Note that the use of multiple regression takes into account the covariances among the X_j in each subgroup). The question to be asked is whether a significant difference exists among the b-weights across the K subgroups. The null hypothesis is: $H_0 : \Gamma_1 = \Gamma_2 = \Gamma_k = \ldots = \Gamma_K$, where each Γ_k is a column vector of subpopulation b-weights resulting from the regression of Y_1 on the X_j in subpopulation k.

Insert Figure 2 about here

The logic of the test is that if H_0 cannot be rejected, then a common set of b-weights can be applied to all K subgroups without affecting significantly the pooled residual sum of squares for predicting Y_1, as compared to the pooled residual sum of squares obtained by determining the b-weights uniquely for each subgroup (Timm, 1975). Consequently, the significance test can be viewed as (a) a test of differences between B_c and uniquely determined B_k, where B_c is an (estimated) vector of b-weights common to all subgroups (Γ_c) and the B_k are estimates of the Γ_k; or (b) the degree to which the pooled residual sum of squares is increased by using a common regression equation to predict Y_1 in all subgroups versus employing unique regression equations for each subgroup. If the significance test leads to rejection of the null hypothesis, then a common B_c should not be applied in all K subgroups. This implies directly that not all of the Γ_k are equal, which in turn suggests that A may be regarded as a moderator, where moderation is indicated by unequal b-weight vectors (the question of intercepts is addressed later in this article).

The required significance test is presented in many advanced statistical texts as a test of parallelism of regression slopes, planes, or hyperplanes; the term "heterogeneous regression" is also employed [cf. Finn (1974, pp.377-378); Rao (1965, pp.237-240); Timm (1975, pp.331-347); Velicer (1972); Williams (1959, chapter 8)] [3]. That is, the null hypothesis, $\Gamma_1 = \Gamma_2 = \Gamma_k = \ldots = \Gamma_K$ is tested by ascertaining whether the sample estimates of the Γ_k, the B_k, are parallel (across subgroups). The test is predicated on determining whether one B_c can be used in place of the separate B_ks, as discussed above. Given the rather extensive treatments of the test in several of the references (see especially Timm), as well as the fact that the test is the same as that employed in analysis of covariance to test the assumption of common regression slopes (cf. Tatsuoka, 1971, pp.40-60), only the significance test itself is presented. The form of the test we have employed is presented in Timm (1975, p.335), and, for sample data, is as follows:

$$F = \frac{\sum_k (B_k - B_c)' \, S_{xy_k} \big/ J\,(K-1)}{\sum_k (S_{yy_k} - B_k' \, SS_{xx_k} \, B_k) \big/ (N - J\,K - K)} \qquad (1)$$

where:

B_k is a (column) vector of unique b-weights determined by multiple regression in subgroup k (primes connote the transpose);

S_{xy_k} is a column vector of sum of cross-products between y_1 and the x_j in subgroup k (lower-case y and x represent deviation scores);

S_{yy_k} is the sum of squares of y_1 in subgroup k;

SS_{xx_k} is the sum of squares and cross-products (SSCP) matrix of the x_j predictors in subgroup k;

B_c is a (column) vector of common b-weights, obtained by $(\sum_k SS_{xx_k})^{-1} (\sum_k S_{xy_k})$;

and

N is the total number of subjects across all subgroups; K is the total number of subgroups, and J is the number of predictors.

Prior to division by degrees of freedom, the numerator of Equation 1 indicates the extent to which the pooled residual sum of squares is increased by employing B_c for all subgroups rather than the unique B_k for each subgroup. The denominator indicates the pooled residual sum of squares resulting from using the unique B_k in each subgroup. The question, therefore, is whether error in the prediction of Y_1 based on the X_j is increased significantly by employing B_c rather than the separate B_k, in relation to the error made by using only the B_k. The F-test has $J(K-1)$ and $(N-JK-K)$ degrees of freedom, and, to employ the test, it is assumed that the residuals (errors) are normally distributed in the subpopulations. As noted above, a significant F-test indicates that not all of the Γ_k are equal, and further suggests that A moderates the X_j, Y_1 relationships.

If the F-test presented in Equation 1 is significant, then it may be of interest to ascertain whether a particular predictor contributed significantly to overall moderation among the subgroups. For example, review of the b-weights might indicate that predictor X_2 in Figure 2 was a major contributor to overall moderation. Consequently, the investigator may wish to compare the b-weights for X_2 separately. The null hypothesis is: H_0 : $\gamma_{121} = \gamma_{122} = \gamma_{12k} = \ldots = \gamma_{12K}$, where the γ_{12k} are subpopulation parameters, estimated by the b_{12k} subgroup weights. A significance test for a "general" predictor, X_j , presented by Williams (1959, p.132), is as follows:

$$F = \frac{\sum_k [(b_{ijk} - \bar{b}_{ij})^2 / S^{jj}_{xx_k}] / (K-1)}{g} \qquad (2)$$

where:

the b_{ijk} are the separate b-weights for criterion Y_i and predictor X_j across the subgroups (based on the regressions of Y_i on all X_j in each subgroup).

$\bar{b}_{ij}$ is the weighted mean of the b_{ijk} weights, determined by

$$\sum_k (b_{ijk} / S^{jj}_{xx_k}) / (\sum_k 1/S^{jj}_{xx_k});$$

$S^{jj}_{xx_k}$ is the jj diagonal element in $SS^{-1}_{xx_k}$; and

g is the residual mean square presented in the denominator of Equation 1.

The F-test has (K-1) and (N - JK - K) degrees of freedom.

Of additional interest might be post hoc, multiple comparisons of specific subgroups. For example, suppose $K = 4$, and following a significant F-test based on Equation 1 and review of the b-weights in each subgroup, it is decided that differences in the b-weights for the A_1 and A_4 subgroups were primarily responsible for the overall differences among the four subgroups. An F-test for the difference between the B_1 and B_4 sample b-weight vectors follows, in part, the procedures underlying the development of

Equation 1. That is, the numerator has the following form:

$$\left[(\underline{B}_1 - \underline{B}_{c_{14}})' \underline{S}_{xy_1} + (\underline{B}_4 - \underline{B}_{c_{14}})' \underline{S}_{xy_4} \right] / \underline{J} \qquad (3)$$

which has $\underline{J}$ degrees of freedom. $\underline{B}_{c_{14}}$ in Equation 3 is a vector of common $\underline{b}$-weights for subgroups $\underline{A}_1$ and $\underline{A}_4$, and is obtained by the equation given for $\underline{B}_c$; only here $\underline{k}$ assumes values of 1 and 4 only. Fo''owing the general logic of multiple comparison tests, the denominator fo testing the hypothesis that Γ_1 and Γ_4 are equal is the same as that for Equation 1 (with the same degrees of freedom).

Of concern at this point is that three tests of significance have been described, two of which may involve more than one test (i.e., Equations 2 and 3). Corrections for the possibility of compounding Type I errors are required. In this respect, rationale presented by Specht and Warren (1976, p.57, fn.2) regarding comparisons of structural equations across subgroups is recommended. This rationale consists of a sequential reduction in the alpha levels for different types of significance tests. Applying this logic to the present discussion suggests that the alpha level for the overall test of parallelism might be set at .05 (or whatever the investigation deems appropriate), the alpha level for the test of a general predictor (Equation 2) set at .025, and the alpha level for the multiple comparison test(s) set at.01(the order of the latter two tests is arbitrary). In addition, with respect to the multiple comparison tests, it is advisable to adjust further the alpha level for <u>each</u> test in order to insure that the alpha level for the set of tests does not exceed a predetermined level (e.g., $\underline{p} < .01$). In this condition, the investigator may wish to employ the logic of error rates (e.g., Winer, 1971) to set the alpha level for each multiple comparison test. Similar logic applies to more than one test based on

Equation 2 inasmuch as the identification of one or more general predictors requires an a posteriori review of results.

Multivariate test. The presentation thus far has addressed only one criterion. It is possible to include more than one criterion in the test of parallelism for independent data sets, in which case the hypothesis to be tested is that matrices of unstandardized regression weights are the same for all K subgroups. That is, the null hypothesis if $\Gamma_1 = \Gamma_2 = \Gamma_k = \ldots = \Gamma_K$, where the Γ_k are J by I subpopulation matrices of unstandardized regression weights. In the sample estimates of the Γ_k, the B_k, each column of each matrix represents the b-weights for the regression of a particular Y_i ($i = 1, \ldots, I$ criteria) on the X_j predictors in the kth subgroup (the X_j are the same for each criterion). The analog of Equation 1 for the overall test of parallelism for I criteria across K subgroups is as follows (cf. Timm, 1975, p.343).

$$\frac{|Q_E|}{|Q_C|} = \frac{\left|\sum_k (SS_{yy_k} - B_k' SS_{xx_k} B_k)\right|}{\left|\sum_k (SS_{yy_k} - B_c' SS_{xx_k} B_c)\right|} \qquad (4)$$

where:

SS_{yy_k} is the SSCP matrix of the y_i criteria in subgroup k;

B_k is a $J \times I$ matrix of b-weights;

SS_{xx_k} is the SSCP matrix of the x_j predictors in subgroup k; and

B_c is a matrix of common b-weights (one column for each criterion), obtained by $(\sum_k SS_{xx_k})^{-1} (\sum_k SS_{xy_k})$, were SS_{xy_k} is the x, y sum of cross-products matrix for subgroup k.

The numerator in Equation 4 is the determinant of an error SSCP matrix, designated Q_E. Q_E is an estimate of the pooled residuals for the criteria after each Y_i is regressed on the X_j in each of K subgroups. The denominator of Equation 4 is the determinant of the sum of Q_E and a hypothesis SSCP matrix. The sum is designated Q_C , and is equivalent to an estimate of the pooled residuals in the criteria after a common set of b-weights are determined for each Y_i across the K subgroups (the b-weights for a particular Y_i are the same across the K subgroups, but the b-weights for different Y_i may be different). Division of $|Q_E|$ by $|Q_C|$ provides Wilks' lamba criterion (Λ), which can be tested by the U distribution with (I, J (K-1), N - JK - K) degrees of freedom (other multivariate test criteria may also be employed). The required assumption is that the criterion variables have an underlying multivariate normal distribution (in the subpopulations).

If a significant Λ is obtained, the investigator may wish to conduct tests on one or more of the separate Y_i, using the procedures discussed in Equations 1 through 3. This requires additional concern for Type I errors. Following prior logic, a sequential decrease in alpha levels may be desirable, beginning with .05 for the overall multivariate test, and progressively smaller alpha levels for each succeeding test or set of tests.

Empirical illustration. Data were created and used to provide examples of the tests in Equations 1 through 4. Created data were employed to provide investigators an opportunity to check computer programs against a full set of empirical data.[4] The created data are presented in Table 1, and consist of three subgroups, two criteria, and two predictors, with an n of 15 in each subgroup ($N = \Sigma\, n = 45$). Descriptive statistics are presented at the bottom of Table 1, where a review of the correlation coefficients and multiple correlations displays the patterns used to create the data. With respect to the first criterion (Y_1), the salient features are (a) relatively large predictor-criterion relationships in subgroups 1 and 3, and comparatively lower predictor-criterion relationships in subgroup 2; and (b) a disordinal interaction between subgroups 1 and 3, where the predictor-criterion relationships are positive in subgroup 1 and negative in subgroup 3. In regard to the second criterion ($\underline{Y}_2$), the pattern of relationships is the same as that for $\underline{Y}_1$; the difference occurs in that the predictor-criterion relationships are smaller in magnitude for $\underline{Y}_2$. These patterns were chosen because they ensured (a) at least moderate, although not necessarily significant, correlations among the predictors and among the criteria in most cases, from which multivariate analysis draws its power; and (b) rejection of the null hypothesis of parallelism for tests involving at least the $\underline{Y}_1$ criterion, which is desirable for an illustration. Moreover, correlations of the magnitudes presented in Table 1 are not unrealistic given ns of 15, although the disordinal interactions (subgroup 1 versus subgroup 3) are admittedly somewhat illusional, especially for $\underline{Y}_1$.

- -

Insert Table 1 about here

- -

Statistical information required to conduct the tests in Equations 1 through 4 is shown in Table 2. In effect, the information in this table is required for some aspect of the multivariate test of parallelism (Equation 4). However, examination of selected portions of the matrices also provides most of the data for the univariate tests. For example, the first column of the SS_{xy_k}, B_k, and B_c matrices contains data required to conduct the univariate test of parallelism for the regression of $\underline{Y}_1$ on $\underline{X}_1$ and $\underline{X}_2$. The only other information required is the SS_{xx_k} matrices and the upper-left portions of the SS_{yy_k} matrices (i.e., the S_{yy_k} s), which are provided in the table, and the degrees of freedom, which are based on $\underline{J} = 2$, $\underline{K} = 3$, and $\underline{N} = 45$ ($\underline{I} = 2$ for the multivariate test). Information not included in this table can be computed directly from the data presented (e.g., pooled within-group SSCP matrices). Finally, it is important to note that B_c in Table 2 is identical for each subgroup, which connotes that the same sets of weights are applied in each subgroup.

Insert Table 2 about here

Illustrative results of the tests presented in Equations 1 through 4 are reported in Table 3. Section I of this table displays the results of the multivariate test of parallelism for the two criteria, based on the rationale discussed in the development of Equation 4. The null hypothesis is that the three subpopulation (unique) b-weight matrices, whose estimates are seen by the three B_k in Table 2, are equal; that is $\Gamma_1 = \Gamma_2 = \Gamma_3$. Equation 4 was employed to estimate the two SSCP matrices, Q_E and Q_C, the determinant values of which are shown in Table 3. Division of $|Q_E|$ by $|Q_C|$ resulted in a Λ equal to .63, significant at the .05 level as tested by the U distribution with (2, 4, 36) degrees of freedom. The null hypothesis of equal b-weight matrices was rejected; however, more in-depth analyses are needed to ascertain the basis for rejection.

Insert Table 3 about here

Separate tests of parallelism for each criterion, across the three subgroups, are exhibited in section II of Table 3. The null hypothesis for each test was $\Gamma_{i1} = \Gamma_{i2} = \Gamma_{i3}$ (the second subscript indicates subgroup). For example, the test for criterion Y_1 (i=1) was based on a test of the equality of the first column of the B_k matrices in Table 2, which are the sample estimates of the Γ_{1k}. The actual test employed is given in Equation 1, where each of the columns above was compared to column 1 of the B_c matrices in Table 2. As shown in Table 3, the null hypothesis was rejected for Y_1 but not Y_2. This indicated that the b-weight vectors were nonparallel across the three subgroups for criterion Y_1 only. Based on discussions of the need to protect against Type I error, the alpha level for the two F-tests was set at .025.

Review of the first column of each B_k matrix in Table 2 suggested that the X_1 predictor contributed to the rejection of the hypothesis of parallelism for criterion Y_1. To test this hypothesis, Equation 2 was employed, the null hypothesis being $\gamma_{ij1} = \gamma_{ij2} = \gamma_{ij3}$ ($i = Y_1$, $j = X_1$). The null hypothesis was rejected, as shown in section III of Table 3. The alpha level for this test was set at .01 based on the logic of sequential reduction in alpha levels. No further correction for error rates was initiated, based on Fisher's least significant difference approach (Winer, 1971) and two possible tests for Y_1.

The final question addressed here was a post hoc comparision between subgroup 1 and subgroup 3 based on the unique b-weight vectors for Y_1. Equation 3 was employed to conduct the test, and, as presented in section IV of Table 3, the null hypothesis that the b-weight vectors were equal was rejected. The alpha level employed for this test was .0017; a per experiment alpha level of .005 was protected for three possible comparisons using Fisher's least significant difference approach (less conservative tests may be desired).

Sequential Moderation

The issue addressed in this section concerns a test of the equality of b-weights when the predictors are measured at a base time period (T_0) and repeated measurements are taken later on the same criterion variable at times T_1 through T_S, where s is the subscript for the time of measurement on the criterion ($s = 1 \ldots S$). The possibility that the b-weights might change differentially over time is illustrated by studies of the dynamic nature of criteria (Fleishman & Fruchter, 1960; Fleishman & Hempel, 1954; Ghiselli, 1956; Ghiselli & Haire, 1960; Inn, Hulin, & Tucker, 1972; MacKinney, 1967; Smith, 1976). As reviewed by MacKinney (1967), the nature of the requirements for job performance might change over time. These changes may be a function of changes in environmental presses, the skills required to perform the job, types of training received, and so forth. It is reasonable to expect that if the requirements for job performance change, then the rank-order of individuals on a job performance criterion, such as quality of work, will change over repeated measurements (at different times) on this criterion. It is also reasonable to expect that predictors of job performance, such as selection tests, will exhibit different relationships with job performance as a function of when job performance is measured.

A model for the type of problem discussed above is shown in Figure 3, which presents b-weight vectors (columns) for the separate regressions of one criterion variable on the same J predictors (e.g., selection tests) on S different occasions. The J predictors are measured at a base period only (T_0), repeated measurements are taken on the criterion at times 1 through S, thus providing S separate Y on X multiple regressions. The same sample of subjects is employed in each of the S regressions. The question is whether a significant difference exists among the b-weight vectors across the $\underline{S}$ different time periods. A significant difference in the $\underline{b}$-weight vectors suggests that predictor-criterion relationships are moderated by time, which, as noted earlier, is indicative of changes in rank-order on the criterion, and may connote changes in such things as skill requirements.

Insert Figure 3 about here

The null hypothesis is $\Gamma_{\underline{1}} = \Gamma_{\underline{2}} = \Gamma_{\underline{s}} = \ldots = \Gamma_{\underline{S}}$, where it is understood that each $\Gamma_{\underline{s}}$ refers to an unstandardized regression weight vector for the same criterion variable (i.e., same scale of measurement) assessed at different points in time. The approaches presented for tests of independent sets of data will not suffice here inasmuch as the data, and thus the $\underline{b}$-weight vectors, are likely correlated. Specifically, the predictor data are identical for each criterion, and the criteria are presumably correlated over time, although these correlations need not be of large magnitudes. The correlations among the $\underline{b}$-weight vectors must be taken into account in the same sense that correlation among means must be considered in correlated $\underline{t}$-tests. In fact, we employed the logic of Hotelling's $\underline{T}^2$ for correlated data and repeated measures ANOVA to develop the test presented below. To our knowledge, this test has not been presented previously, although we benefitted from treatments of tests for correlated data presented by Carter (1949), Kullback and Rosenblatt (1957), and Williams (1959), and from derivations pertaining to the analysis of multiple-cue judgment tasks using the "Lens Model" approach (Castellan, 1973).[5]

The discussion of determining an analytic procedure begins with Hotelling's $\underline{T}^2$ statistic for correlated data, which may be viewed as comprising a hypothesis matrix, $\underset{\sim}{Q}_H$, which takes the form:

$$Q_H = (\bar{Y}_1 - \bar{Y}_2)(\bar{Y}_1 - \bar{Y}_2)' \quad (5)$$

and an error matrix, Q_E, which takes the form:

$$Q_E = \left(V_{\bar{Y}_1} + V_{\bar{Y}_2} - 2C_{\bar{Y}_1\bar{Y}_2} \right) \quad (6)$$

$\bar{Y}_1$ and $\bar{Y}_2$ are vectors of means based on repeated measurements on a set of Y variables for the same subjects, V indicates variance, and C indicates covariance.

If $\bar{Y}_1$ and $\bar{Y}_2$ are replaced by b-weight vectors resulting from the regressions of the Y_s on the same X_j for the same subjects, at times 1 and 2 (s = 1, 2), Q_H in Equation 5 is transformed into

$$Q_H = (B_1 - B_2)(B_1 - B_2)' \quad (7)$$

while Q_E in Equation 6 is

$$Q_E = \left(V_{B_1} + V_{B_2} - 2C_{B_1B_2} \right) \quad (8)$$

Note that the Y_s must be based on the same scale in Equations 7 and 8 in order for the comparison of b-weights to be meaningful. It is not necessary, however, for the Y_s to have the same means or same standard deviations across the S time periods.

We wish now to employ the approach of using common regression coefficients as the basis for testing differences among the correlated b-weight vectors. This provides the opportunity to extend the analytic design to accommodate more than two b-weight vectors (i.e., more than two measurements on the criterion). The multivariate design employs Q_H to represent a hypothesis SSCP matrix, and Q_E to represent a residual SSCP matrix.

Equation 7 was employed as a starting point to design Q_H, but using the difference between each b-weight vector and a common b-weight vector rather than calculating the difference between two b-weight vectors directly. With two b-weight vectors, Q_H is as follows:

$$Q_H = (B_1 - B_c)(B_1 - B_c)' + (B_2 - B_c)(B_2 - B_c)'$$

$$= \sum_s (B_s - B_c)(B_s - B_c)' \qquad (9)$$

where B_s is a b-weight vector for criterion Y measured at time s (s = 1,2), B_c is a vector of b-weights common to Y_1 and Y_2, and the order of the Q_H matrix is J by J. Using logic similar to that presented for the tests of independent data, if a common B_c can be applied for predicting the separate criterion measurements without significantly increasing the residuals, then the uniquely determined b-weight vectors for each Y_s are not significantly different.

A straightforward extension of Equation 9 allows one to include more than two measurements on the criterion in the development of Q_H. That is, s may assume values beyond 2. The key question now is how to determine B_c for two or more repeated measurements on Y. For illustrative purposes, S is set equal to 3 to demonstrate the calculation of B_c.

The determination of B_c is predicated on the inverse of the pooled (across time) X_j variance-covariance (VC) matrices, postmultiplied by the pooled X_j, X_s covariance vectors. In matrix form, this is

$$B_c = \left(\sum_s VC_{xx_s}\right)^{-1}\left(\sum_s C_{xy_s}\right) \qquad (10)$$

where VC_{xx_s} is the X_j variance-covariance matrix associated with the criterion measured at time s, and C_{xy_s} is the J by 1, X_j, Y_s covariance vector for the criterion measured at time s. However, the VC_{xx_s} matrices are identical, which, with minor manipulation and S = 3, provides (the subscript for VC_{xx_s} is deleted):

$$\underline{B}_c = 1/3\left(VC_{xx}^{-1}\, \underline{C}_{xy_1} + VC_{xx}^{-1}\, \underline{C}_{xy_2} + VC_{xx}^{-1}\, \underline{C}_{xy_3} \right)$$

$$= 1/3\ (\underline{B}_1 + \underline{B}_2 + \underline{B}_3)$$

or, in general terms

$$\underline{B}_c = 1/\underline{S}\ (\Sigma_{\underline{s}}\ \underline{B}_s) \qquad (11)$$

In other words, the common coefficient for each $\underline{X}_j$ is simply the average of the separate regression coefficients associated with that $\underline{X}_j$ for each time period.

Equation 8 provided the base for calculating the residual SSCP matrix Q_E . In general form, where $\underline{s}$ may assume values greater than 2, Equation 8 generalizes to the following:

$$\underline{Q}_E = \Sigma_{\underline{s}}\ \underline{V}_{B_s} - 2 \Sigma_{\underline{s}<\underline{p}}\ \underline{C}_{B_s B_p} \qquad \text{(for } \underline{s} < \underline{p}) \qquad (12)$$

The equation above is expanded below, where use is made of derivations presented by Finn (1974) and Castellan (1973). The expansion reflects simplifications resulting from the fact that the predictor variance-covariance matrices (VC_{xx_s}) and SSCP matrices (SS_{xx_s}) are identical at each time of measurement, which also implies, for example, that $VC_{xx_s} = VC_{x_s x_p}$, where $\underline{s} < \underline{p}$ (i.e., $\underline{p} = \underline{s} + 1, \ldots, \underline{S}$, which applies only to the second term on the right-side of Equation 12). Moreover, these identities allowed deletion of the subscripts for the predictor variance-covariance and SSCP matrices. The expansion is as follows.

$$\underline{V}_{B_s} = (1 - \underline{R}_{y_s}^{\ 2})\ \underline{V}_{y_s}\ SS_{xx}^{-1} \qquad (13)$$

where $\underline{R}_{y_s}^{\ 2}$ is the squared multiple correlation resulting from the regression of $\underline{Y}$ on the $\underline{X}_j$ at time $\underline{s}$, and $\underline{V}_{y_s}$ represents the variance of the $\underline{Y}$ criterion measured at time $\underline{s}$.

$$C_{B_s B_p} = \left(C_{y_s y_p} - B_s' VC_{xx} B_p \right) SS_{xx}^{-1} \qquad (s < p) \qquad (14)$$

where $C_{y_s y_p}$ represents the covariance between repeated measures on the criterion, with $s < p$, and B_s and B_p are b-weight vectors for the regressions of Y on the X_j at times s and p.

Based on Equations 13 and 14, Q_E may now be viewed as follows.

$$Q_E = \sum_s \left[(1 - R_{y_s}^2) V_{y_s} SS_{xx}^{-1} \right] - 2 \sum_{s<p} \left[(C_{y_s y_p} - B_s' VC_{xx} B_p) SS_{xx}^{-1} \right]$$

$$= \left[\sum_s (1 - R_{y_s}^2) V_{y_s} - 2 \sum_{s<p} \left(C_{y_s y_p} - B_s' VC_{xx} B_p \right) \right] SS_{xx}^{-1} \qquad (15)$$

Q_E in Equation 15 is of order J by J.

A multivariate significance test has the following form:

$$\Lambda = \frac{|Q_E|}{|Q_H + Q_E|} \qquad (16)$$

which follows the U distribution with $[S, J(S-1), (n-1)(S-1) - SJ]$ degrees of freedom, given that the values on the variables were sampled from an underlying, joint multivariate normal distribution.

A significant Λ indicates that the b-weight vectors for the X_j differ significantly for repeated measurements on Y over S time periods. Given a significant Λ, inspection of each b-weight vector may show that certain predictors have significant relationships with the criterion at one time period, while other predictors have significant relationships with Y at a different time period. These results would be consistent with the logic of dynamic criteria and suggest sequential moderation. However, a significant Λ might also reflect, in part, various statistical inadequacies and artifacts. For example, the predictors may have differential rates of stability over time, where the more stable predictors would have a greater likelihood of being related to the criterion as the time interval between predictor and criterion measurements increased. Moreover, criterion measurements at each point in time may not be equally reliable, which could lead to spurious differences in the *b*-weight vectors. Thus, one must be careful when interpreting the results provided by Equation 16, which is to say that consideration should be given to both substantive issues and potential statistical bias.

Empirical illustration. Created data were employed to simulate a sequential moderation design and to illustrate the use of Equation 16. The design was similar to Figure 3, although more austere, and included two predictors (X_1, X_2) measured at a base period (T_0 -- e.g., two selection tests), and one criterion (Y_s), on which repeated measurements were taken at T_1 and T_2 (i.e., $s = 1,2$). The analytic question is whether the *b*-weights for the Y_1 on X_1 and X_2 regression differ significantly from the *b*-weights for the Y_2 on X_1 and X_2 regression; the null hypothesis is $\Gamma_1 = \Gamma_2$.

The created data ($n = 30$) are presented in Table 4. Means, standard deviations, and correlations among the variables are presented in section I of Table 5. The correlations among the variables reflect the simulated design, where (a) X_1 and X_2 were not correlated significantly, (b) Y_1 and Y_2 had a significant but moderate correlation (which allowed for changes in rank-order on the criterion over time), and (c) X_1 had a high positive relationship with Y_1 and a comparatively lower relationship with Y_2, whereas X_2 had a high positive relationship with Y_2 and a comparatively lower relationship with Y_1. These data reflect a sequential moderation design wherein both the criterion and the salience of a predictor change as a function of time, which in turn suggest that the Y on X regression planes for T_1 and T_2 will be unequal (i.e., nonparallel).

- -

Insert Tables 4 and 5 about here

- -

Statistical information required to conduct some aspects of the test of parallelism presented in Equation 16 is displayed in section II of Table 5. It is important to note that the VC_{xx} and SS_{xx} matrices are the same for T_1 and T_2 (i.e., the same predictor scores are employed in both time periods), the (unique) b-weight vectors (B_1 and B_2) imply nonparallelism, and the vector of common b-weights (B_c) is based on the average of the B_1 and B_2 vectors (see Equation 11). (Exact replication of some values is not possible due to round-off error.)

The results of the test of parallelism are shown in section III of Table 5. Equation 15 was used to estimate Q_E and Equation 9 was employed to estimate Q_H. Division of $|Q_E|$ by $|Q_E + Q_H|$ provided a Λ of .04 ($p < .01$), which indicated rejection of the null hypothesis of parallelism. Thus, the results implied that the relationships between the criterion and the two predictors were a function of the time of measurement on the criterion; that is, the Y on X regressions were sequentially moderated.

Discussion

Analytic procedures have been presented which cast subgrouping moderator analysis and sequential moderation in the roles of tests of parallelism of regression slopes, planes, and hyperplanes. With respect to subgrouping moderator analysis, when (separate) regression equations are to be compared across independent subgroups for one or more criteria, the proposed methods provide a basis for (a) comparing multiple subgroups and multiple predictors simultaneously, and (b) increasing statistical power, in relation to separate comparisons of each predictor-criterion relationship, by taking into consideration covariances among the variables, although this statement is qualified below. Procedures for more in-depth analyses for a particular criterion included a test to identify a general or salient predictor (i.e., a predictor which led to the rejection of parallelism) and a post hoc multiple comparison test for comparing the regression equations of two subgroups. Methods for protecting against Type I errors were also presented, and involved sequential reduction in alpha levels and the use of error rates. It is noteworthy that these methods might result in potentially severe losses of power for tests conducted in the later phases of an analysis (e.g., multiple comparison tests), especially in multivariate designs involving three or more subgroups. In general, this is the price one must pay for "data-snooping", although the methods employed in this paper were conservative and other investigators may wish to use alternative procedures.

It is important to note that the usual caveats associated with subgrouping moderator analysis apply also to the analyses discussed here. The moderator should be uncorrelated, or at least have very low correlations, with the predictors and criteria; the reliabilities of all variables should be within acceptable limits and highly similar across different subgroups; relationships within subgroups should be linear; differences among regressions should be of practical as well as statistical significance, especially if large samples are employed, although the subgroup sample sizes should in any case be of sufficient size to provide meaningful power for the tests; and cross-validation of results is strongly recommended (Abrahams & Alf, 1972; Guion, 1976; Linn, 1978; Schmidt & Hunter, 1977, 1978; Schmidt et al., 1976; Zedeck, 1971). Problems associated with differential range restriction across subgroups is generally not a problem inasmuch as unstandardized regression weights are employed.

It should also be mentioned that the usual caveats apply regarding the use of multiple regression (cf. Cohen & Cohen, 1975). Of particular concern is avoidance of problems resulting from multicollinearity, namely the "bouncing" regression weight problem. Moreover, although the subgroup sample sizes may vary, large differences in sample sizes are likely to result in the same problem associated with comparisons of means, namely a greater probability of committing a Type II error.

The caveats above generally apply to the test for sequential moderation, although the same sample of subjects is required to conduct this test. The sequential moderation test was designed to ascertain whether the regressions of a criterion on a set of predictors could be considered nonparallel for repeated measurements on the criterion. It is noteworthy that this test does not provide a basis for causal inference; other methods are required for this purpose. Nevertheless, descriptive studies of changes in predictor-criterion relationships over time are important, especially in relation to the dynamic nature of criteria and long-range test validity. It is also important to reiterate the point that differences in relationships over time may in part be due to statistical inadequacies and artifacts, such as differential stabilities among the predictors or differences in criterion reliabilities at different times of measurement. Careful consideration should be given to both short-term reliabilities and long-term stabilities in tests of sequential moderation.

In conclusion, the analytic procedures presented here do not exhaust the tests that might be conducted in subgrouping or sequential moderator analysis. For example, tests for differences in intercepts were not addressed.

Tests for differences in intercepts are available in Timm (1975) for the overall tests on independent subgroups (analogues of Equations 1 and 4), and with minor extension the logic can be applied to the multiple comparison test (Equation 3). Other tests that might be of interest include (a) planned comparisons for subgrouping moderator analysis, (b) tests of intercepts, post hoc multiple comparisons, and planned comparisons for the sequential moderation designs, (c) tests involving multiple criteria in sequential moderation designs, and (d) tests which contrast the efficacy of different moderators in both subgrouping and sequential moderator analyses. Additional work is required to develop these tests, although James, Hater, and Jones (Note 1) have proposed a planned comparison test for subgrouping moderator analysis.

References

Abrahams, N.M., & Alf, E., Jr. Pratfalls in moderator research. Journal of Applied Psychology, 1972, 56, 245-251.

Beehr, T.A. Perceived situational moderators of the relationship between subjective role ambiguity and role strain. Journal of Applied Psychology 1976, 61, 35-40.

Beehr, T.A., Walsh, J.T., & Taber, T.D. Relationship of stress to individually and organizationally valued states: Higher order needs as a moderator. Journal of Applied Psychology, 1976, 61, 41-47.

Blalock, H.M., Jr. (Ed.) Causal inferences in nonexperimental research. Chapel Hill: University of North Carolina Press, 1964.

Brief, A.P., & Aldag, R.J. Employee reactions to job characteristics: A constructive replication. Journal of Applied Psychology, 1975, 60 182-186.

Brief, A.P., & Aldag, R.J. Correlates of role indices. Journal of Applied Psychology, 1976, 61, 468-472.

Carter, A.H. The estimation and comparison of residual regressions where there are two or more related sets of observations. Biometrika, 1949, 36, 26-46.

Castellan, N.J., Jr. Comments on the "lens model" equation and the analysis of multiple-cue judgment tasks. Psychometrika, 1973, 38, 87-100.

Cohen, J., & Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. New York: Wiley, 1975.

Dunham, R.B. Reactions to job characteristics: Moderating effects of the organization. Academy of Management Journal, 1977, 20, 42-65.

Finn, J.D. A general model for multivariate analysis. New York: Holt, Rinehart & Winston, 1974.

Fleishman, E., & Fruchter, B. Factor structure and predictability of successive stages of learning Morse code. Journal of Applied Psychology, 1960, 44, 97-101.

Fleishman, E., & Hempel, W. Changes in factor structure of a complex psychomotor test as a function of practice. Psychometrika, 1954, 19, 239-252.

Ghiselli, E.E. Differentiation of individuals in terms of their predictability. Journal of Applied Psychology, 1956, 40, 374-377.

Ghiselli, E.E., & Haire, M. The validation of selection tests in the light of the dynamic character of criteria. Personnel Psychology, 1960, 13, 225-231.

Guion, R.M. Recruiting, selection, and job placement. In M.D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally, 1976.

Gulliksen, H., & Wilks, S.S. Regression tests for several samples. Psychometrika, 1950, 15, 91-114.

Hackman, J.R., & Lawler, E.E., III. Employee reactions to job characteristics. Journal of Applied Psychology Monograph, 1971, 55, 259-286.

Inn, A., Hulin, C.L., & Tucker, L. Three sources of criterion variance: Static dimensionality, dynamic dimensionality, and individual dimensionality. Organizational Behavior and Human Performance, 1972, 8, 58-83.

Johnson, T.W., & Stinson, J.E. Role ambiguity, conflict, and satisfaction: Moderating effects of individual differences. Journal of Applied Psychology, 1975, 60, 329-333.

Katzell, R.A., & Dyer, F.J. Differential validity revived. Journal of Applied Psychology, 1977, 62, 137-145.

Kullback, S., & Rosenblatt, H.M. On the analysis of multiple regression in k categories. Biometrika, 1957, 44, 67-83.

LaRocco, J.M., & Jones, A.P. Co-worker and leader support as moderators of stress-strain relationships in work situations. Journal of Applied Psychology, 1978, 63, 629-634.

Linn, R.L. Single-group validity, differential validity, and differential prediction. Journal of Applied Psychology, 1978, 63, 507-512.

Lyons, T.F. Role clarity, need for clarity, satisfaction, tension, and withdrawal. Organizational Behavior and Human Performance, 1971, 6, 99-110.

MacKinney, A.C. The assessment of performance change: An inductive example. Organizational Behavior and Human Performance, 1967, 2, 56-72.

Malgady, R.G., & Huck, S.W. Predicting multiple criteria: Comparison of two vectors of regression coefficients. Educational and Psychological Measurement, 1978, 38, 869-872.

McLaughlin, D.H. A test for homogeneity of regression without homogeneity of variance. Educational and Psychological Measurement, 1975, 35, 79-86.

Oldham, G.R., Hackman, J.R., & Pearce, J.L. Conditions under which employees respond positively to enriched work. Journal of Applied Psychology, 1976, 61, 395-403.

Rao, C.R. Linear statistical inference and its applications. New York: Wiley, 1965.

Schmidt, F.L., & Hunter, J.E. Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 1977, 62, 529-540.

Schmidt, F.L., & Hunter, J.E. Moderator research and the law of small numbers. Personnel Psychology, 1978, 31, 215-232.

Schmidt, F.L., Hunter, J.E., & Urry, V.W. Statistical power in criterion-related validity studies. Journal of Applied Psychology, 1976, 61, 473-485.

Schuler, R.S. Role perceptions, satisfaction, and performance: A partial reconciliation. Journal of Applied Psychology, 1975, 60, 683-687.

Sims, H.P., Jr., & Szilagyi, A.D. Job characteristic relationships: Individual and structural moderators. Organizational Behavior and Human Performance, 1976, 17, 211-230.

Smith, P.C. Behaviors, results, and organizational effectiveness: The problem of criteria. In M.D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally, 1976.

Specht, D.A., & Warren, R.D. Comparing causal models. In D.R. Heise (Ed.), Sociological Methodology 1976. San Francisco: Jossey-Bass, 1976.

Steers, R.M., & Spencer, D.G. The role of achievement motivation in job design. Journal of Applied Psychology, 1977, 62, 472-479.

Stone, E.F., Mowday, R.T., & Porter, L.W. Higher order need strengths as moderators of the job scope-job satisfaction relationship. Journal of Applied Psychology, 1977, 62, 466-471.

Tatsuoka, M.M. Multivariate analysis. New York: Wiley, 1971.

Timm, N.H. Multivariate analysis with applications in education and psychology. Monterey , CA: Brooks/Cole, 1975.

Tosi, H. Organization stress as a moderator of the relationship between influence and role responses. Academy of Management Journal. 1971, 14, 7-22.

Tukey, J.W. Causation, regression, and path analysis. In O. Kempthorne, T.A. Bancroft, J.W. Gowen, & J.L. Lush (Eds.), Statistics and mathematics in biology. New York: Hofner, 1964.

Velicer, W.F. The moderator variable viewed as heterogeneous regression. Journal of Applied Psychology, 1972, 56, 266-269.

Wanous, J.P. Individual differences and reactions to job characteristics. Journal of Applied Psychology, 1974, 59, 616-622.

Weitz, J. Criteria and transfer of training. Psychological Reports, 1966, 19, 195-210.

Werts, C.E., Rock, D.A., Linn, R.L., & Joreskog, K.G. Comparison of correlations, variances, covariances, and regression weights with or without measurement error. Psychological Bulletin, 1976, 83, 1007-1013.

Williams, E.J. Regression analysis. New York: Wiley, 1959.

Winer, B.J. Statistical principles in experimental design. New York: McGraw-Hill, 1971.

Zedeck, S. Problems with the use of "moderator" variables. Psychological Bulletin, 1971, 76, 295-310.

Footnotes

Support for this research was provided under Office of Naval Research Contract N00014-77-C-123, Office of Naval Research Project RR042-08-10 N170-743, and by the National Institute on Drug Abuse, Department of Health, Education, and Welfare, Grant No. H-81-DA-01931-01. Opinions expressed in this report are those of the authors and are not to be construed as necessarily reflecting the official view or endorsement of the Department of the Navy or the National Institute on Drug Abuse.

The authors would like to thank Robert G. Demaree, Christopher W. Hornick, Michael K. Lindell, and S.B. Sells for their helpful suggestions and advice.

Requests for reprints should be sent to Lawrence R. James, Institute of Behavioral Research, Texas Christian University, Fort Worth, Texas 76129.

[1] The articles reviewed were as follows (a) job characteristics (Brief & Aldag, 1975; Dunham, 1977; Hackman & Lawler, 1971; Hackman & Oldham, 1976; Oldham, Hackman, & Pearce, 1976; Sims & Szilagyi, 1976; Steers & Spencer, 1977; Wanous, 1974), and (b) role perceptions (Beehr, 1976; Beehr, Walsh, & Taber, 1976; Brief & Aldag, 1976; Johnston & Stinson, 1975; Larocco & Jones, 1978; Lyons, 1971; Schuler, 1975; Tosi, 1971).

[2] An exception is a study by Stone, Mowday, and Porter (1977), which employed a chi-square test to assess the equality of three correlation coefficients provided by three subgroups.

[3] See also articles by Malgady and Huck (1978), McLaughlin (1975), and Werts, Rock, Linn, and Jöreskog (1976).

4 Our programs, which are written in APL and available from the authors, were checked against data presented in Timm (1975). These data were not used here because only two subgroups were employed, and it was our desire to illustrate the procedures with more than two subgroups.

5 The authors are indebted to Michael K. Lindell for pointing out the relevance of the "Lens Model" approach.

Table 1

Illustrative Data and Descriptive Statistics for Three Subgroups

Two Criteria and Two Predictors

	Subgroups, Data, and Statistics											
	Subgroup 1				Subgroup 2				Subgroup 3			
Subjects	X_1	X_2	Y_1	Y_2	X_1	X_2	Y_1	Y_2	X_1	X_2	Y_1	Y_2
1	1	2	1	1	1	3	1	1	5	4	2	1
2	1	4	2	4	1	4	4	4	5	2	2	4
3	2	4	1	1	3	4	1	1	4	2	1	1
4	2	3	2	3	2	3	5	5	4	3	2	3
5	2	2	3	3	2	2	3	4	4	4	3	4
6	3	2	1	4	3	2	4	5	3	4	1	4
7	3	1	2	1	3	1	1	1	3	5	2	1
8	3	3	3	3	3	3	1	5	3	3	3	3
9	3	3	4	5	3	3	5	5	3	3	4	5
10	4	2	3	5	4	2	5	5	2	4	3	5
11	4	4	4	4	4	4	2	2	3	2	4	4
12	4	4	5	5	4	4	5	5	2	2	5	5
13	5	3	4	5	5	3	5	5	1	2	4	5
14	5	5	4	4	5	5	4	4	1	1	4	4
15	5	4	5	4	5	4	4	4	1	2	5	4
Statistics												
M	3.13	3.07	2.93	3.47	3.20	3.13	3.33	3.73	2.93	2.87	3.00	3.53
SD	1.36	1.10	1.39	1.46	1.32	1.06	1.68	1.62	1.33	1.13	1.31	1.46
$r_{x_1x_2}$	.33				.29				.37			
$r_{y_1y_2}$			.69*				.80**				.64*	
$r_{x_jy_1}$	.76*	.47			.32	.13			-.69*	-.48		
$r_{x_jy_2}$	.58*	.34			.26	.02			-.53*	-.35		
R_{y_i}[a]			.80*	.60			.33	.27			.74*	.55

[a] Multiple correlation based on the regression of each criterion on two predictors in each subgroup.

* $p < .05$

Table 2

Statistical Information Required for Multivariate and Univariate

Tests of Parallelism on Three Independent Subgroups

		Subgroups		
Statistics	Order	Subgroup 1	Subgroup 2	Subgroup 3
SS_{xx_k}	JxJ	$\begin{bmatrix} 25.73 & 6.87 \\ 6.87 & 16.93 \end{bmatrix}$	$\begin{bmatrix} 24.40 & 5.60 \\ 5.60 & 15.73 \end{bmatrix}$	$\begin{bmatrix} 24.93 & 7.86 \\ 7.87 & 17.73 \end{bmatrix}$
$SS_{xx_k}^{-1}$	JxJ	$\begin{bmatrix} .04 & -.02 \\ -.02 & .07 \end{bmatrix}$	$\begin{bmatrix} .04 & -.01 \\ -.02 & .07 \end{bmatrix}$	$\begin{bmatrix} .05 & -.02 \\ -.02 & .07 \end{bmatrix}$
SS_{yy_k}	IxI	$\begin{bmatrix} 26.93 & 19.47 \\ 19.47 & 29.73 \end{bmatrix}$	$\begin{bmatrix} 39.33 & 30.33 \\ 30.33 & 36.93 \end{bmatrix}$	$\begin{bmatrix} 24.00 & 17.00 \\ 17.00 & 29.73 \end{bmatrix}$
SS_{xy_k}	JxI	$\begin{bmatrix} 20.13 & 16.07 \\ 10.07 & 7.53 \end{bmatrix}$	$\begin{bmatrix} 10.00 & 7.80 \\ 3.33 & .53 \end{bmatrix}$	$\begin{bmatrix} -17.00 & -14.47 \\ -10.00 & -7.93 \end{bmatrix}$
B_k	JxI	$\begin{bmatrix} .70 & .57 \\ .31 & .21 \end{bmatrix}$	$\begin{bmatrix} .39 & .34 \\ .07 & -.08 \end{bmatrix}$	$\begin{bmatrix} -.59 & -.51 \\ -.30 & -.22 \end{bmatrix}$
B_c	JxI	$\begin{bmatrix} .18 & .14 \\ .00 & -.05 \end{bmatrix}$	$\begin{bmatrix} .18 & .14 \\ .00 & -.05 \end{bmatrix}$	$\begin{bmatrix} .18 & .14 \\ .00 & -.05 \end{bmatrix}$

Table 3

Results of Tests of Parallelism, a General Predictor, and a Multiple Comparison for Independent Subgroups

I. Test of Parallelism across Three Subgroups Based on Two Criteria

Source-Residual	Determinant Value	Λ	df (U-test)
Q_E	2592.66	.63*	(2,4, 36)
Q_C	4121.07		

II. Tests of Parallelism for Each of Two Criteria, Across Three Subgroups

	Criterion	
Degrees of Freedom	Y_1	Y_2
Common Weights	4	4
Unique Weights	36	36
Mean Square-Residual		
Common Weights	8.02	5.29
Unique Weights	1.55	2.05
F	5.17**	2.58

III. Test to Ascertain if predictor X_1 Contributed to Nonparallelism for the Y_2 Criterion

Source-Residual	df	MS	F
Weighted Mean	2	14.57	9.66***
Unique Weights	36	1.51	

IV. Comparison of Subgroup 1 and Subgroup 3 Based on b-weights for the Y_1 Criterion

Source-Residual	df	MS	F
Common Weights Based on Two Groups	2	15.00	9.66****
Unique Weights	36	1.55	

Note. The p-values are as follows: * < .05, ** < .025, *** < .01, **** < .0017.

Table 4

Illustrative Data for One Sample, Two Predictors,

and Two Repeated Measurements on One Criterion

Subjects	X_1[a]	X_2[a]	Y_1[b]	Y_2[c]
1	1	1	1	1
2	6	2	4	2
3	1	4	2	6
4	2	2	3	2
5	2	3	2	4
6	2	6	1	6
7	3	8	2	7
8	3	5	6	5
9	3	7	2	6
10	4	7	3	8
11	4	5	4	7
12	4	9	3	7
13	5	4	6	6
14	5	8	5	7
15	5	9	6	8
16	4	5	5	7
17	3	8	2	8
18	6	10	6	9
19	5	6	6	8
20	7	8	7	9
21	7	4	8	9
22	6	7	7	9
23	8	8	8	10
24	8	7	9	5
25	9	3	8	4
26	7	9	9	8
27	9	6	4	7
28	10	9	7	8
29	10	7	9	7
30	10	6	10	7

[a] Measured at time T_0.

[b] Y measured at time T_1.

[c] Y measured at time T_2.

Table 5

Descriptive Statistics, Statistical Information Required for Test of Parallelism, and Results of Test of Parallelism for Sequential Moderation

I. Means, Standard Deviations, and Correlations

	X_1	X_2	Y_1	M	SD
X_1	-			5.30	2.69
X_2	.33	-		6.10	2.36
Y_1	.84**	.28	-	5.17	2.63
Y_2	.38*	.78**	.40*	6.57	2.17

II. Statistical Information Required for Test of Parallelism

$$VC_{xx}^{a} = \begin{bmatrix} 7.21 & 2.07 \\ 2.07 & 5.56 \end{bmatrix} \qquad VC_{xx}^{-1\,a} = \begin{bmatrix} .16 & -.06 \\ -.06 & .20 \end{bmatrix}$$

$$V_{y_1} = \begin{bmatrix} 6.94 \end{bmatrix} \qquad V_{y_2} = \begin{bmatrix} 4.71 \end{bmatrix} \qquad C_{y_1 y_2} = \begin{bmatrix} 2.27 \end{bmatrix}$$

$$C_{xy_1} = \begin{bmatrix} 5.95 \\ 1.72 \end{bmatrix} \qquad C_{xy_2} = \begin{bmatrix} 2.23 \\ 4.01 \end{bmatrix}$$

$$B_1 = \begin{bmatrix} .82 \\ .00 \end{bmatrix} \qquad B_2 = \begin{bmatrix} .11 \\ .68 \end{bmatrix} \qquad B_c^{b} = \begin{bmatrix} .47 \\ .34 \end{bmatrix}$$

III. Test of Parallelism for Two Predictor and Two Measurements on One Criterion

Source	Determinant Value	Λ_1	df
Q_E	.0003	.04**	(2, 2, 25)
$Q_E + Q_H$	.0060		

[a] Identical for $s = 1, 2$; $SS_{xx} = n\, VC_{xx}$ and $SS_{xx}^{-1} = \frac{1}{n} VC_{xx}^{-1}$.

[b] Based on average of B_1 and B_2.

* $p < .05$

** $p < .01$

	Subgroup	
	A_1	A_2
Criterion Y_1		
Predictor X_1	r_{111}	r_{112}
Predictor X_2	r_{121}	r_{122}
Predictor X_3	r_{131}	r_{132}
Criterion Y_2		
Predictor X_1	r_{211}	r_{212}
Predictor X_2	r_{221}	r_{222}
Predictor X_3	r_{231}	r_{232}

Criterion Y_1	Subgroup					
	A_1	A_2	$\cdots$	A_k	$\cdots$	A_K
Predictor X_1	b_{111}	b_{112}	$\cdots$	b_{11k}	$\cdots$	b_{11K}
Predictor X_2	b_{121}	b_{122}	$\cdots$	b_{12k}	$\cdots$	b_{12K}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
Predictor X_j	b_{1j1}	b_{1j2}	$\cdots$	b_{1jk}	$\cdots$	b_{1jK}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
Predictor X_J	b_{1J1}	b_{1J2}	$\cdots$	b_{1Jk}	$\cdots$	b_{1JK}

Predictors at T_0	Time						
	T_1	T_2	...	T_s	...	T_S
X_1	b_{11}	b_{21}	...	b_{s1}	...	b_{S1}
X_2	b_{12}	b_{22}	...	b_{s2}	...	b_{S2}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
X_j	b_{1j}	b_{2j}	...	b_{sj}	...	b_{Sj}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
X_J	b_{1J}	b_{2J}	...	b_{sJ}	...	b_{SJ}

Figure 1. An illustration of a typical analytic model for subgrouping moderator analysis. Subscripts for the correlation coefficients, r_{ijk}, are: i = 1,2 criteria, j = 1,2,3 predictors, k = 1,2 subgroups.

Figure 2. An illustration of an analytic model for tests of independent subgroups.

Figure 3. An illustration of an analytic model for a test of sequential moderation. The *b*-weights refer to predictor-criterion relationships for repeated measurements on the same criterion.