

Covariance Analyses with Heterogeneity of Slopes in Fixed Models

Leland J. Hendrix, Melvin W. Carter and Del T. Scott

Department of Statistics, Brigham Young University, Provo, Utah 84602, U.S.A.

SUMMARY

Techniques that describe the use of covariance when heterogeneity of slopes exists are severely limited. Although a few procedures for model selection have been recommended, none, except the hierarchical approach, is straightforward and usable with present computer programs. The hierarchical subset selection procedure presented in this paper is based on the proposition that heterogeneity may be present only for certain terms in the model. After hierarchical selection, those terms which do not involve heterogeneity are interpreted as in the usual analysis for covariance. The interpretations of those terms which do involve heterogeneity are modified with respect to significance tests performed at various values of the covariate. The hierarchical subset selection method allows one to investigate heterogeneity of slopes in covariance models as functions of the classification variables present in the design.

1. Introduction

Often, in the analysis of experimental and observational studies, both discrete and continuous independent variables are present. One approach to the analysis of this type of data is to assume that the slopes on the continuous independent variables are homogeneous for all combinations of the discrete variables. The validity of this assumption permits the traditional analysis of covariance (ANCOVA) suggested in most introductory texts (Ostle and Mensing, 1975). However, this assumption is often not satisfied. Tests of whether the slopes are homogeneous can be found in Gujarati (1970), Searle (1971) and Seber (1977). Traditionally, a separate intercept and slope have been fitted for each cell whenever the hypothesis of homogeneity of slopes is rejected. Until recently, little additional help has been available to the researcher wishing to proceed with the analysis.

Potthoff (1964) and Rogosa (1980), extending results given by Johnson and Neyman (1936), presented a method for the comparison of populations that have different slopes and/or intercepts. The method is based on the construction of simultaneous confidence regions about regression lines or selected points.

Searle (1979) presented some models in which the heterogeneity of slopes took on specific forms; for example, in a two-way experimental design layout, heterogeneity may be present for the row effects but not for the interaction or column effects. M. W. Reeder, in a master's thesis from Brigham Young University (1973), provided theoretical results concerning which tests of the means or intercepts are unaffected by different values of the covariate. For the example cited by Searle (1979), Reeder's results suggested that tests of column and interaction effects are unaffected by different choices of covariate values. However, row effects need to be interpreted in light of the heterogeneity. The investigation of the heterogeneity of the row effects with respect to the covariate resembles the analysis and interpretation of multiway layouts when interaction is present

Key words: Covariance; Heterogeneity; Homogeneity.

This paper provides a methodology for model determination, analysis and interpretation where the assumption of homogeneity is not met, currently available software being used. The procedure is illustrated with data taken from a PhD dissertation by L. J. Hendrix (Brigham Young University, 1967).

2. Notation

Consider an experiment with treatments in a completely randomized design. A covariate measure is taken on the experimental units prior to the assignment of treatments, the response measurement being taken after imposition of the treatments.

The general model takes the form

$$y_{ij} = \beta_{0i} + \beta_{1i}x_{ij} + \varepsilon_{ij}, \quad (1)$$

with

y_{ij} , the observed response for the j th observation of the i th treatment, where $i = 1, 2, \dots, t$, $j = 1, 2, \dots, n_i$ where $n_i \geq 2$ for all t and $n_i \geq 3$ for some t ;

β_{0i} , the y-intercept for the i th treatment $= E(y_{ij} | x = 0)$;

$x_{ij} = (X_{ij} - \bar{X})$;

X_{ij} , the covariate value for the j th observation of the i th treatment;

β_{1i} , the slope for the i th treatment;

$\bar{X} = n^{-1} \sum_{ij} X_{ij}$;

ε_{ij} , the error for the j th observation of the i th treatment.

The ε_{ij} values are assumed to be normally and independently distributed with constant variance.

Although β_{0i} is the average response for the population when Treatment i is given, one may be interested in estimating the treatment means and differences for those units in the population with a particular value for the covariate, say $x_c = X_c - \bar{X}$. In this case,

$$\mu_i | x_c = \beta_{0i} + \beta_{1i}x_c. \quad (2)$$

When heterogeneity of slopes is present it is apparent from (2) that differences in means depend on the value of x_c .

The method shown in this paper uses a full-rank reparameterization for covariance models given by Reeder (1973) and the theory and notation given by Bryce, Scott and Carter (1980). In matrix form the 'cell means' and 'cell slopes' model, (1), is

$$\mathbf{y} = (\mathbf{W} | \mathbf{W}_x) \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \boldsymbol{\varepsilon}, \quad (3)$$

where $\mathbf{W}$ is an incidence matrix linking an observation to its cell mean, and $\mathbf{W}_x$ is a matrix with the same structure as $\mathbf{W}$, linking an observation to its slope. For example, for an observation y_{ij} , the j th row of $\mathbf{W}$ is $(00 \dots 010 \dots 0)$, while the j th row of $\mathbf{W}_x$ is $(00 \dots 0X_{ij}00 \dots)$. The nonzero quantities are in the i th column. The β_0 in (3), i.e. the cell intercepts or the cell means at $x_c = 0$, are reparameterized to correspond to the contrasts and terms of interest to the researcher. The contrasts used to reparameterize β_0 are also used to reparameterize β_1 .

Thus, after reparameterization, (3) becomes

$$\mathbf{y} = (\mathbf{W} | \mathbf{W}_x) \begin{pmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \boldsymbol{\varepsilon}. \quad (4)$$

The first row of $\mathbf{A}$ is $\mathbf{a}'_1 = (1/t)\mathbf{j}$, and the remaining rows of $\mathbf{A}$ define the $t-1$ contrasts of interest. These contrasts must be chosen so that $\mathbf{A}$ is nonsingular. Equation (4) can be restructured as

$$\mathbf{y} = (\mathbf{Z} \mid \mathbf{Z}_x) \begin{pmatrix} \boldsymbol{\delta}_0 \\ \boldsymbol{\delta}_1 \end{pmatrix} + \boldsymbol{\epsilon}, \quad (5)$$

where $\mathbf{Z} = \mathbf{W}\mathbf{A}^{-1}$, $\mathbf{Z}_x = \mathbf{W}_x\mathbf{A}^{-1}$, $\boldsymbol{\delta}_0 = \mathbf{A}\boldsymbol{\beta}_0$ and $\boldsymbol{\delta}_1 = \mathbf{A}\boldsymbol{\beta}_1$.

If the t 'treatments' of (1) represent treatments arising from two factors, B and C, then the $t-1$ contrasts are chosen in the usual manner so that the rows of matrix $\mathbf{A}$ reflect the main effect contrasts of B and C followed by their interaction. After $\mathbf{Z}$ and $\mathbf{Z}_x$ have been partitioned into the columns associated with the terms in the two-way model, Model (5) becomes

$$\mathbf{y} = \mathbf{z}_\mu \boldsymbol{\delta}_{0(\mu)} + \mathbf{Z}_B \boldsymbol{\delta}_{0(B)} + \mathbf{Z}_C \boldsymbol{\delta}_{0(C)} + \mathbf{Z}_{BC} \boldsymbol{\delta}_{0(BC)} + \mathbf{z}_{x\mu} \boldsymbol{\delta}_{1(\mu)} + \mathbf{Z}_{xB} \boldsymbol{\delta}_{1(B)} + \mathbf{Z}_{xC} \boldsymbol{\delta}_{1(C)} + \mathbf{Z}_{xBC} \boldsymbol{\delta}_{1(BC)} + \boldsymbol{\epsilon}, \quad (6)$$

where $\mathbf{z}_\mu$ is a column vector of ones and $\mathbf{z}_{x\mu}$ is a column vector of the covariates.

As shown in Reeder's thesis, the last three terms in Model (6) can be expressed as

$$\mathbf{Z}_{xQ} = \mathbf{z}_{x\mu} * \mathbf{Z}_Q,$$

where $*$ indicates elementwise multiplication of the elements in the column vector $\mathbf{z}_x$ by each of the corresponding elements of each of the columns of $\mathbf{Z}_Q$, where Q is used to represent each of the terms in the model, i.e. B, C and BC. The terms $\mathbf{Z}_{xQ}$ refer to the heterogeneity of slopes.

For the purposes of this paper, the following simplified form for (6) is written term-for-term:

$$\mathbf{y} = \mu + B + C + BC + X + XB + XC + XBC + \boldsymbol{\epsilon}. \quad (7)$$

It should be understood that this does not represent the usual overparameterized model.

3. Procedures for Selecting the Model

The first step is to determine whether or not the regression coefficients that represent the relationship between the covariate and the dependent measure (response) are heterogeneous across the cells of the experimental design. This can be done by forming the heterogeneity terms or products of the covariate and the analysis of variance (ANOVA) terms and including them, along with the covariate and ANOVA terms, in the full model as indicated in (7). It may be noted from (7) that the heterogeneity terms are just 'interactions' with the ANOVA terms. The sum of squares associated with the error from this model, $SSE(F)$, is then compared with the sum of squares associated with the error from the reduced model, $SSE(R)$, which contains only the covariate and independent variable(s). Testing whether or not one can use a model with a common slope, i.e. testing for homogeneity, is similar to the problem of testing incompletely specified models. Bozivich, Bancroft and Hartley (1956) proposed the use of an α value greater than .05, for example $\alpha = .20$ or .25, for such tests in order to help control Type II error. Here we use $\alpha = .25$. Thus, if the F test

$$F = \frac{\{SSE(R) - SSE(F)\} / \{df_{(e)R} - df_{(e)F}\}}{MSE(F)}$$

is nonsignificant at $\alpha = .25$, homogeneity among the regression coefficients is assumed to

exist. The reduced model with only the covariate and the ANOVA terms is then used in the traditional ANCOVA. If the F test for heterogeneity is significant at $\alpha = .25$, several different approaches can be used to continue the analyses. These include analysis using the full model described above or using one of several subset selection procedures to arrive at a modified reduced model.

3.1 *Use of the Full Model*

When the preliminary test indicates that the full model should be used, an F test should be performed on each of the heterogeneity terms adjusted for the remainder of the model. Interpretations of the ANOVA terms will vary from one model term to another according to the pattern of heterogeneity present. Whenever homogeneity exists for a particular item, i.e. the terms are nonsignificant at $\alpha = .25$, the corresponding ANOVA term is interpreted as in the traditional ANCOVA, since the results are invariant to changes in the covariate (Reeder's thesis). However, where heterogeneity of slopes exists, the numerical value of a comparison depends on the value of the covariate, and the ANOVA terms have different interpretations at various values of the covariate. This is analogous to ANOVA when interpretation of a main effect is modified by the presence of a significant interaction.

3.2 *Use of Subset Selection Procedures*

Three different subset selection procedures are suggested here as possible ways to proceed when the original F test indicates heterogeneity. The first two procedures use traditional regression techniques, but may be difficult to implement due to the lack of computer software. The third is a technique usually used in ANOVA for factorial or polynomial models. In all three procedures, only the heterogeneity terms in the model are subject to the subset selection procedures. The covariate and the terms which would be included in an ordinary ANOVA are constrained to be present in the model.

For the full model, Reeder's thesis showed that those ANOVA terms, for which the corresponding heterogeneity terms are not significant, can be interpreted as in the traditional ANCOVA; he also showed that those ANOVA terms, for which heterogeneity terms were selected and included, require different interpretations at various values of the covariate. Reeder's results also hold for a reduced model when it is 'correct.'

3.2.1 *Stepwise subset selection procedure by terms.* The stepwise procedure by terms determines which subset of the heterogeneity terms is to be included in the model. Here this would involve the selection of terms as a whole rather than the individual independent variables (contrasts) which compose a term. This procedure usually results in a considerable decrease in the number of variables involved in the selection process. However, the available stepwise computer programs cannot be used at present since they select on individual variables.

3.2.2 *Subset selection procedure by terms, using algorithms suggested by La Motte and Hocking (1970) and Furnival and Wilson (1974).* The stepwise subset selection procedure by terms does not necessarily yield the 'optimal' subset of a particular size which explains the maximum amount of variance. The examination of all possible subsets can be an expensive procedure. The algorithms suggested by the above authors guarantee the optimal solution without actually analyzing all possible subsets of various sizes. Such a procedure helps minimize computer time and provides a feasible alternative to the

inexpensive stepwise procedure by terms. However, as in the stepwise subset selection by terms, present computer programs that employ this procedure select on individual independent variables rather than on terms. Modifications would be required to make the programs applicable to the covariance process suggested. At present all the possible subsets would actually have to be generated in order to use this approach.

3.2.3 Hierarchical approach to subset selection. In ANOVA, it is often the case that the higher the order of a term the less likely it is to explain a significant proportion of the variance. It is a common practice to pool third- or higher-degree interactions with error, on the assumption that they explain only random variation (Ostle and Mensing, 1975). Heterogeneity terms involving the covariate can be treated in a similar manner. All the covariate heterogeneity terms are one degree higher than the corresponding terms in the original ANOVA model. The hierarchical procedure suggests the testing of sequential sets of heterogeneity terms. First, the highest-order heterogeneity term is tested. If it is nonsignificant, it is pooled with error and the test is repeated for the set of heterogeneity terms of next highest order. These steps are performed until terminated by the finding of a significant lower-order set of heterogeneity terms. All heterogeneity terms within the remaining set of terms are then tested individually for significance (each term being adjusted for all the others) to see if further model reduction is possible. Those which are nonsignificant can be pooled together and tested as a set for nonsignificance. Once all of the nonsignificant heterogeneity terms have been deleted, the ANOVA terms that have no significant heterogeneity term associated with the covariate are interpreted as in ordinary ANCOVA. Those with significant heterogeneity terms have their interpretation modified by the presence of the heterogeneity. The approach described can be performed with any computer package that uses general linear model procedures.

4. An Illustration of Hierarchical Model Selection and Interpretation of Fixed Effects in the Presence of Heterogeneity

Hendrix (1967) conducted a study to determine the difference in the ability to discriminate aurally between environmental sounds, with respect to sex (S), cultural status (C) and a treatment program (T). Subjects within sex and cultural strata were randomly assigned to the treatment programs (10 within each three-way cell). Pre-tests and post-tests in auditory discrimination were administered. These data are summarized in Table 1.

The gain scores were analyzed as the dependent variable with the pre-test score used as a covariate. This is equivalent to analyzing the post-test score as the dependent variable with the pre-test score as the covariate (Hendrix, Carter and Hintze, 1978). A discussion of other covariance assumptions is to be found in the paper by Hendrix *et al.* (1978) and so is not presented here. The primary purpose of the present paper is to illustrate the model selection procedures that should be followed when heterogeneity is encountered.

The covariate (pre-test) was standardized by subtraction of its mean and then division by its standard deviation. The full model contains culture, sex, treatment, their interactions, the standardized covariate and the heterogeneity terms, i.e.

$$\begin{aligned} y = & \mu + C + S + T + CS + CT + ST + CST \\ & + X + XC + XS + XT + XCS + XCT + XST + XCST + \epsilon. \end{aligned}$$

The sequential analysis resulting from the full model is summarized in Table 2.

Table 1
Raw data used in the example for the analyses of covariance

Culturally-nondeprived								Culturally-deprived							
Males				Females				Males				Females			
Experimental subject		Control		Experimental subject		Control		Experimental subject		Control		Experimental subject		Control	
Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
58	71	35	49	64	71	68	70	53	67	60	61	52	54	41	45
57	69	31	69	39	71	52	64	49	58	24	42	50	70	41	51
63	71	54	69	69	71	53	67	44	64	59	63	59	65	49	60
66	70	65	65	56	76	43	63	60	62	57	64	42	62	29	35
45	65	54	63	67	71	54	63	39	56	55	44	62	63	42	56
51	69	37	55	39	65	35	53	63	68	55	65	35	44	41	52
62	69	64	66	32	66	62	65	62	66	40	47	41	47	45	34
58	66	69	69	62	70	67	69	60	60	51	69	36	44	62	66
52	61	70	69	64	68	51	68	35	32	32	43	37	45	45	33
59	63	39	57	66	68	42	61	63	64	56	65	64	70	64	68

The reduced model contains all of the above terms with the exception of the covariate heterogeneity terms, i.e.

$$y = \mu + C + S + T + CS + CT + ST + CST + X + \epsilon.$$

The sums of squares associated with the error terms of the two models were compared to determine whether or not heterogeneity exists. The test for heterogeneity gives a

Table 2
Sequential analysis of covariance on auditory discrimination gains, calculated at the mean of the standardized pre-test score and including all heterogeneity terms

Source	df	Mean square
Mean	1	6498.0125
Culture, C	1	409.5125
Sex, S	1	0.1125
Treatment, T	1	13.6125
CS	1	15.3125
CT	1	27.6125
ST	1	49.6125
CST	1	17.1125
Pre-test score, X	1	2588.3647
XC	1	644.3927
XS	1	2.8271
XT	1	32.1984
XCS	1	28.2504
XCT	1	18.3854
XST	1	24.1407
XCST	1	1.0736
Error	64	40.5073

significant result ($P < .25$):

$$F_{7,64} = \frac{(3342.73532 - 2592.46710)/(71 - 64)}{2592.46710/64} = 2.65$$

The next step was to test the third-order heterogeneity term involving culture, sex, treatment and the covariate:

$$F_{1,64} = \frac{1.07364/1}{2592.46710/64} = 1.0736/40.5073 = 0.027.$$

Since the result of this test is nonsignificant the ANOVA term CST can be interpreted in the usual covariance manner.

The three second-order heterogeneity terms involving the covariate were then tested as a set:

$$F_{3,65} = \frac{70.77638/3}{2593.54074/65} = \frac{23.5921}{39.9006} = 0.59.$$

This test result is also nonsignificant, so all first-order ANOVA terms can be interpreted as in the usual covariance analysis.

Finally, the set of first-order heterogeneity terms was tested:

$$F_{3,68} = \frac{679.41817/3}{2664.31713/68} = 5.78.$$

This test result is significant, indicating that one or more of the first-order heterogeneity terms are significant. Individual F tests were then performed on the heterogeneity terms involving culture ($F_{1,68} = 17.09$), sex ($F_{1,68} = 0.29$) and treatment ($F_{1,68} = 0.82$). Only the culture heterogeneity term is significant. An F test combining the sex and treatment heterogeneity terms ($F_{2,68} = 0.45$) is nonsignificant. Thus the sex and treatment variables may be interpreted as in ordinary ANCOVA, but the interpretation of the culture variable needs to reflect the covariate values. Table 3 summarizes the analysis obtained from the adjusted or partial sums of squares that resulted from use of the hierarchical model selection procedure.

Table 3
*Adjusted analysis of covariance on auditory discrimination gains
calculated at the mean of the standardized pre-test score*

Source	df	Mean square	F	P
Mean	1	6498.0125	168.51	.000
Culture, C	1	994.8492	25.80	.000
Sex, S	1	4.5562	0.12	.732
Treatment, T	1	149.0909	3.87	.053
Pre-test score, X	1	2395.1014	62.11	.000
XC	1	644.3927	16.71	.000
CS	1	32.1855	0.83	.364
CT	1	0.0931	0.0024	.961
ST	1	21.4874	0.56	.458
CST	1	29.7949	0.77	.382
Error	70	38.5620		

The significant F test for the mean ($F = 168.51$) shows that an overall gain was made between the pre-test and the post-test. The test on the pre-test scores ($F = 62.11$) shows that the amount of the gain is related to the initial performance as measured by the pre-test. This relationship is negative, as indicated by the regression coefficient associated with the pre-test (-5.8121). This is an artifact of the test ceiling, because the higher a person scores on the pre-test, the less the possible gain.

The result of the F test for culture is also significant ($F = 25.80$). Those in the higher socio-economic group have an adjusted mean gain of 13.3606, compared to an adjusted gain of 6.0890 for the lower socio-economic group. These mean gains were calculated at the mean of the standardized covariate, 0.000. However, since the heterogeneity involving the pre-test and culture is significant, the difference between these means is not constant for all values of the pre-test score. Figure 1 is a plot of the gains made by the two cultural groups against scores from the standardized pre-test. The nonparallel regression lines for the two cultural groups depict the heterogeneity between the pre-test and culture.

To shed further light on this significant heterogeneity, the pre-test scores were redefined by the addition of one and by the subtraction of one to give results at, respectively, one standard deviation below and one standard deviation above the mean of the covariate. Since the results of these two analyses are identical to the results summarized in Table 3, except for the analysis of the culture variable, only the culture results are given in Table 4.

Table 4

Adjusted analysis of covariance on auditory discrimination gains associated with culture, at one standard deviation below the mean, at the mean, and at one standard deviation above the mean of the covariate

Location	Source	df	Mean square	F	P
-1 SD	Culture, C	1	1643.1755	42.61	.000
Mean	Culture, C	1	994.8492	25.80	.000
+1 SD	Culture, C	1	13.8428	0.36	.551

The difference between cultures is significant ($F = 42.61$) at one standard deviation below the mean, as suggested by Fig. 1. The adjusted mean gains for the higher and lower socio-economic groups are 22.1874 and 8.8864, respectively. At one standard deviation above the mean the difference between cultures is no longer significant ($F = 0.36$). Figure 1 also suggests this, as do the adjusted means of 4.5338 and 3.2917. Thus, for low and medium values of the pre-test score, the higher socio-economic group makes significantly greater gains than the lower socio-economic group; but for high values of the covariate, there is no significant difference in gains.

The treatment variable is the only other term in Table 3 that approaches significance ($F = 3.87$). Since treatments have no significant heterogeneity with the pre-test, the data indicate that the difference between the two treatments remains constant for all values of the covariate. Figure 2 shows the near parallelism of the regression lines related to the treatment variable under the full or unconstrained model. The F ratio for testing this lack of parallelism is 0.59. Therefore the constraint of parallelism was used to obtain the adjusted means. The adjusted means at one standard deviation below the mean of the covariate (16.9200, 14.1538), at the mean of the covariate (11.1079, 8.3417) and at one standard deviation above the mean of the covariate (5.2958, 2.5297) show the constant difference between the treated and control groups.

A summary of the analyses indicates that the difference between cultures cannot be

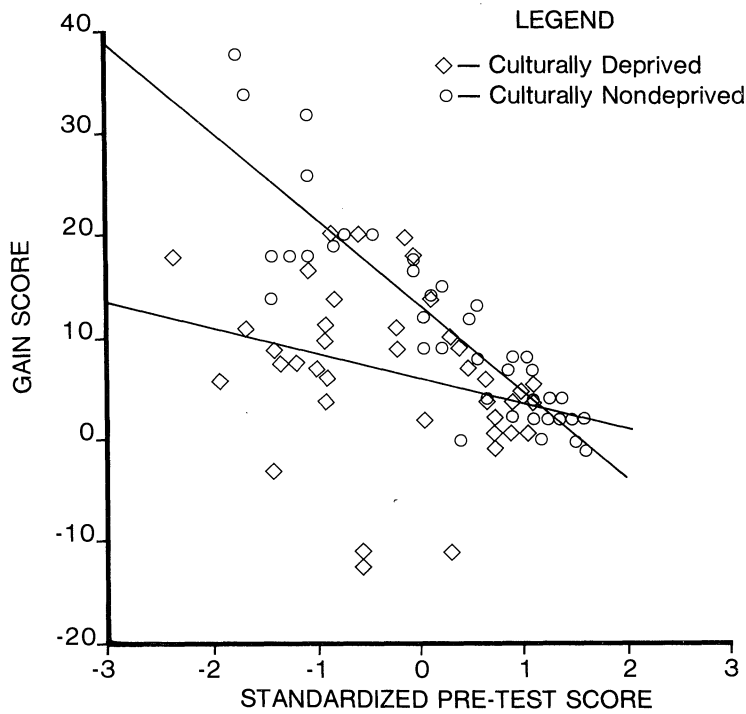


Figure 1. The linear relationship between gain scores and standardized pre-test scores for culturally-deprived and nondeprived subjects; (heterogeneity of slopes present).

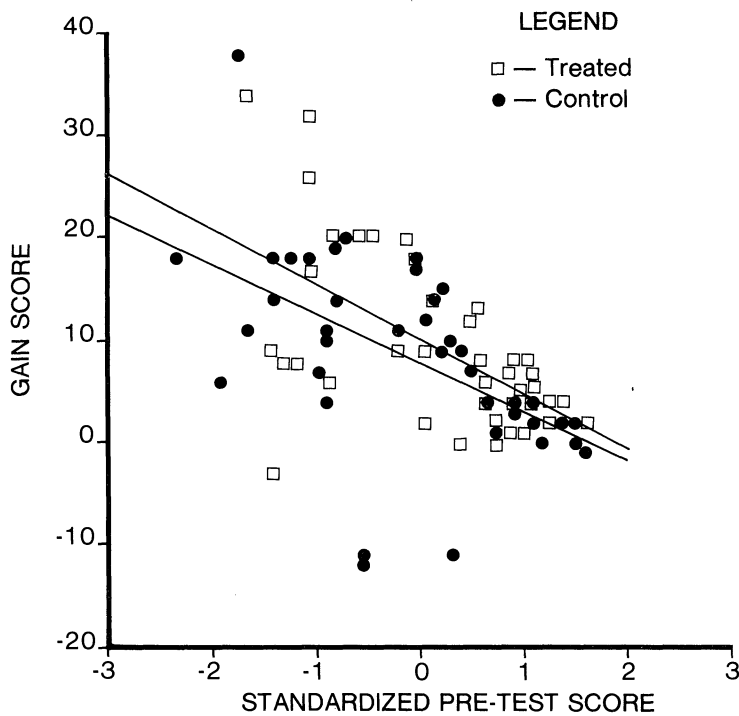


Figure 2. The linear relationship between gain scores and standardized pre-test scores for experimental and control subjects; (homogeneity of slopes indicated).

interpreted by the traditional ANCOVA, which compares the adjusted means only at the mean of the covariate. The heterogeneity or nonparallel regression lines preclude such a comparison. The adjusted means are conditional on the values of the covariate. However, since treatments are homogeneous with the pre-test, the difference between the adjusted treatment means can be interpreted at the mean of the covariate or at any other covariate value since the difference remains constant.

ACKNOWLEDGEMENTS

We very much appreciate the suggestions of the Guest Editor and the referees in the revision of this manuscript.

RÉSUMÉ

Il existe des techniques qui décrivent l'utilisation de la covariance, quand les pentes sont différentes, mais elles sont très limitées. Bien que peu de procédures de sélection de modèle aient été recommandées, aucune, sauf l'approche hiérarchique, n'est facile et n'est utilisable avec les programmes actuellement disponibles sur ordinateurs. La procédure de sélection d'un sous-ensemble hiérarchique que nous présentons ici est fondée sur la proposition que les différences de pente n'existent que pour certains termes du modèle. Après sélection hiérarchique, les termes qui n'impliquent pas de différence sont interprétés comme dans une analyse de covariance classique. L'interprétation des termes qui impliquent des différences est faite en rapport avec les tests de signification réalisés pour différentes valeurs de la covariable. La méthode de sélection d'un sous-ensemble hiérarchique permet d'analyser la différence des pentes dans des modèles de covariance, comme des fonctions des variables de classification présentées dans le dispositif expérimental.

REFERENCES

- Bozovich, H., Bancroft, T. A. and Hartley, H. O. (1956). Power of analysis of variance test procedures for certain incompletely specified models. *Annals of Mathematical Statistics* **27**, 1017–1043.
- Bryce, G. R., Scott, D. T. and Carter, M. W. (1980). Estimation and hypothesis testing in linear models—a reparameterization approach to the cell means model. *Communications in Statistics, Series A* **9**, 131–150.
- Furnival, G. M. and Wilson, R. W., Jr (1974). Regressions by leaps and bounds. *Technometrics* **8**, 28–51.
- Gujarati, D. (1970). Use of dummy variables in testing for equality between sets of coefficients in two linear regressions. *American Statistician* **24**, 50–52.
- Hendrix, L. J., Carter M. W. and Hintze, J. L. (1978). A comparison of five statistical methods for analyzing pretest-post designs. *Journal of Experimental Education* **47**, 96–102.
- Johnson, P. O. and Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs* **1**, 57–93.
- La Motte, L. R. and Hocking, R. R. (1970). Computational efficiency in the selection of regression variables. *Technometrics* **12**, 83–93.
- Ostle, B. and Mensing, R. W. (1975). *Statistics in Research*. Ames, Iowa: Iowa State University Press.
- Potthoff, R. F. (1964). On the Johnson–Neyman technique and some extensions thereof. *Psychometrika* **29**, 241–255.
- Rogosa, D. (1980). Comparing nonparallel regression lines. *Psychological Bulletin* **83**, 307–321.
- Searle, S. R. (1971). *Linear Models*. New York: Wiley.
- Searle, S. R. (1979). Alternative covariance models for the 2-way crossed classification. *Communications in Statistics, Series A* **8**, 799–818.
- Seber, G. A. F. (1977). *Linear Regression Analysis*. New York: Wiley.

Received September 1981; revised April 1982