HARKINS, Ralph Douglas, 1939-
METHODS FOR TESTING THE ASSUMPTION OF HOMOGENEITY OF REGRESSION COEFFICIENTS FOR THE RANDOMIZED COMPLETE BLOCK DESIGN ANALYSIS OF COVARIANCE.

The University of Oklahoma, Ph. D., 1967
Statistics

University Microfilms, Inc., Ann Arbor, Michigan

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

METHODS FOR TESTING THE ASSUMPTION OF HOMOGENEITY OF REGRESSION COEFFICIENTS FOR THE RANDOMIZED COMPLETE BLOCK DESIGN ANALYSIS OF COVARIANCE

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

RALPH DOUGLAS HARKINS

Oklahoma City, Oklahoma

1967

METHODS FOR TESTING THE ASSUMPTION OF HOMOGENEITY OF REGRESSION COEFFICIENTS FOR THE RANDOMIZED COMPLETE BLOCK DESIGN ANALYSIS OF COVARIANCE

APPROVED BY

M Clinton Miller III

Wilson S. Stearns

John C. Brixey

Thomas J. Lynn

Robert C. Duncan

DISSERTATION COMMITTEE

ACKNOWLEDGMENT

I should like to express my appreciation and thanks to Dr. M. Clinton Miller, III for suggesting this topic and his encouragement and aid in the preparation of the dissertation. I am also grateful for the helpful comments and suggestions given by Dr. Edward N. Brandt, Jr., Dr. Robert C. Duncan and Dr. John C. Brixey. I should like, also, to thank Dr. Wilson D. Steen and Dr. Thomas N. Lynn for their careful reading of this dissertation. Finally, I should like to express my sincere thanks for the care shown by Mrs. Rose Titsworth in the typing of the manuscript.

TABLE OF CONTENTS

LIST OF TABLES

METHODS FOR TESTING THE ASSUMPTION OF HOMOGENEITY OF REGRESSION COEFFICIENTS FOR THE RANDOMIZED COMPLETE BLOCK DESIGN ANALYSIS OF COVARIANCE

CHAPTER I

INTRODUCTION

A problem often encountered in statistical practice is that of controlling variation due to sources other than the treatments being evaluated. Control of these extraneous sources of variation may be obtained by blocking on them, by using them as covariates, or by a suitable combination of these two techniques. The necessity for this control is a direct result of the desire on the part of the investigator to make realistic tests for differences of treatment means. That is to say, the investigator wants the comparisons among the treatment means to be independent of those effects due to sources other than the treatments.

In general, there are two methods by which sets of concomitant variables (X_1, X_2, ..., X_p) and the corresponding dependent variable Y can be obtained. They are:

(1) The set of random observation vectors (X_1, X_2, ..., X_p, Y) can be obtained as a random sample from the general population, or

(2) The set of vectors (X_1, X_2, ..., X_p, Y) can be obtained

by first selecting the X values and then sampling the corresponding distribution for the dependent variable, Y.

In either case the assumption from linear regression that the covariates be fixed and measured without error must be satisfied.

Method (1) is of great use in studies of a nature such that it is impossible or impractical to preselect the values for the concomitant variables. For example, in a water quality study designed to evaluate the effect of some pollutant in a stream, it is impractical to preselect flow rate and dissolved oxygen concentration values and then obtain the value of a dependent variable. Method (2) is of value in controlled laboratory experiments where it is possible for the investigator to preselect the values for the concomitant variables and hold them at a steady state during the course of the experiment.

The analysis of covariance randomized complete block design is an experimental design which embodies both blocking and covariates in the statistical control of variation due to extraneous variables. The model for this design, for two covariates, is

$$Y_{ij} = \mu + \tau_i + \rho_j + \beta_{y.x} (X_{ij} - \bar{X}_{..}) + \beta_{y.z} (Z_{ij} - \bar{Z}_{..}) + \epsilon_{ij}, \qquad (1.1)$$

where Y_{ij} is the observed response recorded for the i^{th} treatment in the j^{th} block, X_{ij} and Z_{ij} are concomitant variables on which Y_{ij} has a linear regression with coefficients $\beta_{y.x}$ and $\beta_{y.z}$ respectively. It is further assumed that μ is the true overall mean response, τ_i is the true effect for the i^{th} treatment, $i = 1, 2, \ldots, t$; and ρ_j is the true block effect for the j^{th} block; $j = 1, 2, \ldots, r$. The ϵ_{ij} are random variates assumed to be normally distributed [1, p. 272] with mean zero and variance σ^2. Clearly this model is a combination of the models for the randomized complete block design and that for linear regression.

This design can be of use in the analysis of clinical trial data where the experiment is designed to evaluate a particular set of drugs or a therapeutic regimen. Variables such as sex or age could be controlled by using them as blocking factors. Variables such as time of administration of the drug or initial condition of health could be controlled by using them as covariates. In water quality analyses designed to evaluate the effects of various concentrations of a pollutant, location could be used as a blocking factor, with flow and dissolved oxygen concentration being used as concomitant variables in order to control the variation due to them.

Many investigators hesitate to use the analysis of covariance for a variety of reasons. Experimentally, it is often difficult to measure a variable without error. This may be a function of the measuring device which is itself incapable of giving error free measurements, or it may be a function of the inability of the experimenter to preselect a value of a variable, that is to fix a variable. It is many times quite easy to define a variable as fixed and then assume that corresponding to each of these preselected values there is a corresponding population of dependent variables to be sampled. It is quite another matter to define away or discount errors of measurement. For example, in clinical trial experiments a condition of health can be described, thus fixing the variable "condition of health." However, if in the subsequent investigation the measuring device is primarily human judgment, which most usually is subject to varying degrees of bias, imprecision and the inaccuracies due to inadequate knowledge or individual physiological disturbances, the variable has not been measured without error. If, on the other hand, a series of physiological tests which are capable of giving reproducible results are used, it usually can be said that the variable was measured without error. If human judgment is the measuring device, this could easily lead to the collection of a body of data which may not be reproducible in a later study. Furthermore, since such measurements are on an ordinal scale at best, parametric statistical techniques are not truly applicable. The desired control over variation due to concomitant variables measured on an ordinal scale or less is many times obtained by blocking on them.

It is important that the concomitant variable be measured without error since a measurement error introduces bias into the estimates of the regression coefficients. This in turn introduces bias into the adjusted dependent variable. Since the purpose for using covariates is the removal of bias from the dependent variable, an error of measurement defeats the whole purpose of the analysis. The presence of bias gives rise to a faulty test for differences of adjusted means.

Finally, one additional caution should be made in utilizing the analysis of covariance. If any covariate is affected by the treatment being evaluated, extreme caution should be used in interpreting the results. In fact, it is not generally advisable to use such a variable as a covariate.

The concept of the analysis of variance in the randomized complete block design is as follows: given r blocks and t treatments, such that each treatment appears an equal number of times in each block, it is possible to compute a total sum of squares. Subsequently, this total sum of squares is partitioned into sums of squares due to blocks, treatments, and error. If the mean square for treatment means is significantly large, the hypothesis of equality of treatment means is rejected. In the analysis of covariance randomized complete block design, the total sum of squares is partitioned in the same manner, but now the test for equality of treatment means is performed on a residual mean squares. These residuals are the differences between the actual observed dependent variables and a regression quantity based on the concomitant variable or variables. This adjustment is made after the variation due to blocks has been removed by subtraction.

It is stated above that in the analysis of covariance randomized complete block design, a test for equality of residual or "adjusted" treatment means is made. A test of the assumption of homogeneity of regression coefficients would aid in determining whether this test is valid. The failure of this assumption indicates that the treatments under study do not react at the same rate for all levels of the independent factors. To ignore this potential source of bias in the tests for differences among adjusted treatment means will lead to invalid interpretations of results since, if the underlying assumptions are not met, any inferences made are of dubious value. This is due to the fact that the numbers read from the percentile tables of the F statistic will not give the true level of significance in repeated testing of hypotheses based on the invalid model. There is now way to determine the effects of slight deviations from homogeneity among the regression coefficients upon the level of significance. However, since the F test is not usually greatly affected by minor disturbances of its underlying assumptions, the tabular percentile values are probably close enough for practical purposes. The F statistic is often called "robust" due to this property.

If an investigator desires realistic tests for differences of treatment means, he should also be interested in determining whether the assumptions upon which the test statistic is based are satisfied.

Steel and Torrie state [7, p. 319]

> Where the design is a completely random one, the regression of Y on X can be computed for each treatment. In this case, the usual assumption of homogeneity of regression coefficients can be posed as a null hypothesis and tested by an appropriate F test in an analysis of covariance. (For other designs, we are not aware of methods presently available for testing homogeneity of regression coefficients.)

Furthermore, a search of recent literature has not revealed a method for testing the assumption of homogeneity of regression coefficients when blocking is included in the model.

Therefore, it is the purpose of this paper to develop an easily applied method for testing the assumption of homogeneity of regression coefficients when the model is as given in equation (1.1), that is the analysis of covariance, randomized complete block design.

CHAPTER II

NOTATION AND THEORETICAL DEVELOPMENT

Two problems exist in testing the assumption of homogeneity of regression coefficients. The first problem is that of testing whether the slopes of the regression planes within the t treatment groups are the same. The second problem is determining the source of heterogeneity if it exists.

The first problem suggests a test of equality of the sum of the residual sum of squares about each treatment plane with the residual error sum of squares. This is the same test as is made in the analysis of covariance completely random design for homogeneity of regression coefficients. The second problem suggests a need for testing the equality of a specified set of the p X 1 vectors of regression coefficients, i.e. a null hypothesis that a linear combination of the vectors

$$(\beta_{11},\beta_{12},\cdots,\beta_{1p})^T,\ (\beta_{21},\beta_{22},\cdots,\beta_{2p})^T\ ,\ \ldots,\ (\beta_{t1},\beta_{t2},\cdots,\beta_{tp})^T \quad (2.1)$$

Solutions to both of these problems will be developed in this paper for the analysis of covariance randomized complete block design.

The Test of Homogeneity

The model for the analysis of covariance randomized complete block design was given in equation (1.1) and is repeated here for convenience.

$$Y_{ij} = \mu + \tau_i + \rho_j + \beta_{y.x} (X_{ij} - \overline{X}..) + \beta_{y.z} (Z_{ij} - \overline{Z}..) + \epsilon_{ij} \quad . \qquad (2.2)$$

Blocks serve as a means of increasing the precision of the experiment by grouping comparable experimental units. It will be assumed that each treatment appears only once in each block. Under such circumstances the j^{th} block should contribute some mean amount to each of the observations recorded on the t treatments in that block. In fact, in the analysis of variance for this design a sum of squares due to blocks is partitioned from the total sum of squares and generally is not considered in the subsequent test for differences of treatment means. The same holds true for the analysis of covariance randomized complete block design, but in this case a regression coefficient for Y on each covariate is computed after the variation due to blocks is removed. The idea is to remove the variation due to blocks before computing the regression coefficients. These block effects shall be removed in the following paragraphs such that the assumption of homogeneity of regression coefficients can be tested in a manner similar to that for the completely random design. Subsequently a test for homogeneity of regression coefficients will be developed using the sums of squares and cross products derived below in conjunction with the standard sums of squares and cross products for this model. The following procedure will be utilized to remove the block effects.

From each observation pair Y_{ij} and X_{ij}, subtract the mean of the j^{th} block, that is compute the following difference $(Y_{ij} - \sum_i Y_{ij}/_t)$, $(X_{ij} - \sum_i X_{ij/t})$ etc. for each observation pair in the j^{th} block. The model reflects the effects of this procedure in the following way, as it is now

$$Y_{ij} - \frac{1}{t}\sum_{i=1}^{t} Y_{ij} = (\mu+\tau_i+\rho_j+\beta_{y.x}(X_{ij} - \overline{X}..)+\beta_{y.z}(Z_{ij} - \overline{Z}..)+\epsilon_{ij})$$

$$- \frac{1}{t}\{t\mu + \tau. + t_{pj} + \beta_{y.x}X_{.j} - t\beta_{y.x}\overline{X}.. + \beta_{y.z}Z_{.j}$$

$$- t\beta_{y.z}\overline{Z}.. + \epsilon_{.j}\}$$

$$= \tau_i - \overline{\tau} + \beta_{y.x}(X_{ij} - \overline{X}_{.j}) + \beta_{y.z}(Z_{ij} - \overline{Z}_{.j}) + \epsilon_{ij} - \overline{\epsilon}_{.j} \quad . \qquad (2.3)$$

For simplicity this can be rewritten as

$$Y_{ij}{*} = \tau_i{*} + \beta_{y.x} X_{ij}{*} + \beta_{y.z} Z_{ij}{*} + \epsilon_{ij}{*} \quad . \qquad (2.4)$$

This result is free of the effects due to blocks and is identical to the model for the analy:is of covariance completely random design with μ equal to zero.

Using the notation from [4, p. 330], the likelihood equation for the i^{th} treatment plane is

$$L = f(i*) = \frac{1}{2\pi\sigma^2}^{\frac{r}{2}} \exp\ \{- \frac{1}{2\sigma^2}\sum_{j=1}^{r}(Y_{ij}{*}-\tau_i{*}-\beta_{y.x}X_{ij}{*}-\beta_{y.z}Z_{ij}{*})^2\} \quad (2.5)$$

and

$$\text{Log } L = - \frac{r}{2}\log 2\pi - \frac{r}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{j=1}^{r}(Y_{ij}{*}-\tau_i{*}-\beta_{y.x}X_{ij}{*}-\beta_{y.z}Z_{ij}{*})^2. \quad (2.6)$$

Setting the partial derivatives of (2.6) equal to zero for $\tau_i{*}$, $\beta_{y.x}$, $\beta_{y.z}$, and σ^2 gives

$$\frac{\partial \log L}{\partial \tau_i{*}} = \frac{1}{\hat{\sigma}^2}\sum_j (Y_{ij}{*}-\hat{\tau}_i{*}-\hat{\beta}_{y.x} X_{ij}{*}-\hat{\beta}_{y.z} Z_{ij}{*}) = 0,\ i=1,\ 2,\ \ldots,\ t$$

$$\frac{\partial \log L}{\partial \beta_{y.x}} = \frac{1}{\hat{\sigma}^2}\sum_j (Y_{ij}{*} - \hat{\tau}_i{*} - \hat{\beta}_{y.x} X_{ij}{*} - \hat{\beta}_{y.z} Z_{ij}{*})\, X_{ij}{*} = 0 \qquad (2.7)$$

$$\frac{\partial \log L}{\partial \beta_{y.z}} = \frac{1}{\hat{\sigma}^2} \sum_j (Y_{ij}* - \hat{\tau}_i* - \hat{\beta}_{y.x} X_{ij}* - \hat{\beta}_{y.z} Z_{ij}*) Z_{ij}* = 0$$

$$\frac{\partial \log L}{\partial \sigma^2} = \frac{r}{2\hat{\sigma}^2} \quad \frac{1}{2\hat{\sigma}^4} \sum_j (Y_{ij}* - \hat{\tau}_i* - \hat{\beta}_{y.x} X_{ij}* - \hat{\beta}_{y.z} Z_{ij}*)^2 = 0$$

Solving these equations, first for $\hat{\tau}_i*$, then for the regression coefficients estimators gives

$$\hat{\tau}_i* = \frac{1}{r} (Y_{i.}* - \beta_{y.x} X_{i.}* - \beta_{y.z} Z_{i.}*),$$

$$\hat{\beta}_{y.x} = \frac{G_{zzi}\, G_{xyi} - G_{xzi}\, G_{yzi}}{G_{xxi}\, G_{zzi} - (G_{xzi})^2}$$

and (2.8)

$$\hat{\beta}_{y.z} = \frac{G_{xxi}\, G_{yzi} - G_{xzi}\, G_{xyi}}{G_{xxi}\, G_{yzi} - (G_{xzi})^2},$$ where

$$G_{xxi} = \sum_{j=1}^{r} (X_{ij}* - \overline{X}_{i.}*)^2$$

$$G_{yyi} = \sum_{j=1}^{r} (Y_{ij}* - \overline{Y}_{i.}*)^2$$

$$G_{zzi} = \sum_{j=1}^{r} (Z_{ij}* - \overline{Z}_{i.}*)^2$$

$$G_{xyi} = \sum_j (X_{ij}* - \overline{X}_{i.}*)(Y_{ij}* - \overline{Y}_{i.}*)$$

$$G_{xzi} = \sum_j (X_{ij}* - \overline{X}_{i.}*)(Z_{ij}* - \overline{Z}_{i.}*) \qquad (2.9)$$

$$G_{yzi} = \sum_j (Y_{ij}* - \overline{Y}_{i.}*)(Z_{ij}* - \overline{Z}_{i.}*)$$

and the adjusted treatment group sum of squares is

$$G'_{yyi} = G_{yyi} - \hat{\beta}_{y.xi}\, G_{xyi} - \hat{\beta}_{y.zi}\, G_{zyi} \quad .$$

Equations (2.9) are the sums of cross products within each treatment group plane, thus i = 1, 2, ... t.

The standard equations for the error sums of squares and cross products, treatment sums of squares and cross products, block sums of squares and cross products, and total sums of squares and cross products are now introduced.

Cochran [1, p. 271] gives the following equations for the error sums of products

$$\begin{aligned}
E_{xx} &= \sum_{i,j} (X_{ij} - \overline{X}_{.j} - \overline{X}_{i.} - \overline{X}_{..})^2 \\
E_{yy} &= \sum_{i,j} (Y_{ij} - \overline{Y}_{.j} - \overline{Y}_{i.} - \overline{Y}_{..})^2 \\
E_{zz} &= \sum_{i,j} (Z_{ij} - \overline{Z}_{.j} - \overline{Z}_{i.} - \overline{Z}_{..})^2 \qquad\qquad (2.10) \\
E_{xy} &= \sum_{i,j} (X_{ij} - \overline{X}_{.j} - \overline{X}_{i.} - \overline{X}_{..})\,(Y_{ij} - \overline{Y}_{.j} - \overline{Y}_{i.} - \overline{Y}_{..}) \\
E_{xz} &= \sum_{i,j} (X_{ij} - \overline{X}_{.j} - \overline{X}_{i.} - \overline{X}_{..})\,(Z_{ij} - \overline{Z}_{.j} - \overline{Z}_{i.} - \overline{Z}_{..}) \\
E_{yz} &= \sum_{i,j} (Y_{ij} - \overline{Y}_{.j} - \overline{Y}_{i.} - \overline{Y}_{..})\,(Z_{ij} - \overline{Z}_{.j} - \overline{Z}_{i.} - \overline{Z}_{..})
\end{aligned}$$

where $\overline{X}_{..} = \sum_{i,j} X_{ij}/rt$, $\overline{Y}_{..} = \sum_{i,j} Y_{ij}/rt$, $\overline{Z}_{..} = \sum_{i,j} Z_{ij}/rt$

$\overline{X}_{.j} = \sum_{i} X_{ij}/t$, $\overline{Y}_{.j} = \sum_{i} Y_{ij}/t$, $\overline{Z}_{.j} = \sum_{i} Z_{ij}/t$

$\overline{X}_{i.} = \sum_{j} X_{ij}/r$, $\overline{Y}_{i.} = \sum_{j} Y_{ij}/r$, and $\overline{Z}_{i.} = \sum_{j} Z_{ij}/r$.

The treatment sums of squares and cross products are given by the following equations:

$$T_{xx} = \frac{1}{r} \sum_i X_{i.}^{\ 2} - \frac{X_{..}^{\ 2}}{r\ t}$$

$$T_{xy} = \frac{1}{r} \sum_i X_{i.}\ Y_{i.} - \frac{X_{..}Y_{..}}{r\ t} \tag{2.11}$$

$$T_{yy} = \frac{1}{r} \sum_i Y_{i.}^{\ 2} - \frac{Y_{..}^{\ 2}}{r\ t}$$

The sums of squares for T_{xz}, T_{yz}, and T_{zz} are obtained in the same manner.

Those for blocks are obtained from

$$R_{xx} = \frac{1}{t} \sum_j X_{.j}^{\ 2} - \frac{X_{..}^{\ 2}}{r\ t}$$

$$R_{xy} = \frac{1}{t} \sum_j X_{.j}Y_{.j} - \frac{X_{..}Y_{..}}{r\ t} \tag{2.12}$$

$$R_{yy} = \frac{1}{t} \sum_j Y_{.j}^{\ 2} - \frac{Y_{..}^{\ 2}}{r\ t}$$

with R_{xz}, R_{yz}, and R_{zz} being computed in the same manner.

Those for the total are given by

$$\sum_{i,j} x_{ij}^{\ 2} = \sum_{i,j} X_{ij}^{\ 2} - \frac{X_{..}^{\ 2}}{r\ t}$$

$$\sum_{i,j} x_{ij}y_{ij} = \sum_{i,j} X_{ij}Y_{ij} - \frac{Y_{..}X_{..}}{r\ t} \tag{2.13}$$

$$\sum_{i,j} y_{ij}^{\ 2} = \sum_{i,j} Y_{ij}^{\ 2} - \frac{Y_{..}^{\ 2}}{r\ t}$$

with $\Sigma x_{ij}z_{iz}$, $\Sigma y_{ij}z_{ij}$ and Σz_{ij}^2 being obtained in the same manner.

The error sums of squares and cross products can also be obtained by summing over i the sums of squares for the treatment groups. That is,

$$\sum_i G_{xxi} = E_{xx}$$

$$\sum_i G_{xyi} = E_{xy} \tag{2.14}$$

and so forth for the other sums of products. The truth of this becomes evident upon making the following algebraic manipulations. First, consider the quantity $\sum_i G_{xxi}$. This is

$$\sum_i G_{xxi} = \sum_i [\sum_j (X_{ij}^* - \overline{X}_{i.}^*)^2], \tag{2.15}$$

but $X_{ij}^* = X_{ij} - \sum_j X_{ij}/r$ and

$$\overline{X}_{i.}^* = \frac{1}{t} \sum_i (X_{ij} - \sum_j X_{ij}/r).$$

Therefore, making these substitutions, equation (2.15) is

$$\sum_i G_{xxi} = \sum_i [\sum_j ((X_{ij} - \overline{X}_{i.}) - \overline{X}_{.j} + \overline{X}_{..})^2]$$

$$= \sum_{i,j} (X_{ij} - \overline{X}_{i.} - \overline{X}_{.j} + \overline{X}_{..})^2 = E_{xx} \quad . \tag{2.16}$$

Similarly

$$\sum_i G_{xyi} = \sum_i [\sum_j (X_{ij} - \overline{X}_{i.} - \overline{X}_{.j} + \overline{X}_{..})(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})]$$

$$= \sum_{i,j} (X_{ij} - \overline{X}_{i.} - \overline{X}_{.j} + \overline{X}_{..})(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}) = E_{xy} \; . \tag{2.17}$$

Making similar substitutions the other identities for the error sums of squares and cross products are obtained.

The foregoing results are tabulated in Table 1 which gives all the quantities needed to test the assumption of homogeneity of regression coefficients. This table, with the exception of the sums of squares due to blocks, is identical to that for the analysis of covariance, completely random design when the test of homogeneity of regression coefficients is

TABLE 1

ANALYSIS OF COVARIANCE TABLE

Source	df	SS:XX	SS:XY	SS:XZ	SS:ZZ	SS:ZY	SS:YY	df	SS:YY'
Total	rt-1	x^2_{ij}	$x_{ij}y_{ij}$	$x_{ij}z_{ij}$	z^2_{ij}	$z_{ij}y_{ij}$	y^2_{ij}		$y_{ij}'^2$
Blocks	r-1	R_{xx}	R_{xy}	R_{xz}	R_{zz}	R_{zy}	R_{yy}		
Treatments	t-1	T_{xx}	T_{xy}	T_{xz}	T_{zz}	T_{zy}	T_{yy}	(t-1)	T'_{yy}
Trt. Plane (1)	r-1	G_{xx1}	G_{xy1}	G_{xz1}	G_{zz1}	G_{zy1}	G_{yy1}	(r-1-p)	G'_{yy1}
Trt. Plane (2)	r-1	G_{xx2}	G_{xy2}	G_{xz2}	G_{zz2}	G_{zy2}	G_{yy2}	(r-1-p)	G'_{yy2}
⋮									
Trt. Plane (t)	r-1	G_{xxt}	G_{xyt}	G_{xzt}	G_{zzt}	G_{zyt}	G_{yyt}	(r-1-p)	G'_{yyt}
Error	(r-1)(t-1)	E_{xx}	E_{xy}	E_{xz}	E_{zz}	E_{zy}	E_{yy}	(r-1)(t-1)-p	E'_{yy}

included.

To test the assumption of homogeneity of regression coefficients use

$$F = \frac{(E'_{yy} - \sum_i G'_{yyi}) \;/\; [(r-1)(t-1)-p]-[t(r-1-p)]}{\sum_i G'_{yyi} \;/\; t(r-1-p)}$$

$$= \frac{(E'_{yy} - \sum_i G'_{yyi}) \;/\; p(t-1)-(r-1)}{\sum_i G'_{yyi} \;/\; t(r-1-p)} \qquad (2.18)$$

This is a ratio of the variation among the regression coefficients of the different treatment groups to the sum of that about each group.

Two items should be noted. First, the number of covariates cannot exceed r-2 although as a maximum p is equal to r-2. Second, the inequality $t(r-1-p) < (r-1)\ (t-1)-p$ must hold.

Test Statistic for a Linear Combination of the Vectors of Regression Coefficients

In this section a test of significance for a linear combination of the vectors of regression coefficients,

$$(\beta_{11},\beta_{12},\cdots,\beta_{1p})^T\ ,\ (\beta_{21},\beta_{22},\cdots,\beta_{2p})^T,\ \ldots,\ (\beta_{t1},\beta_{t2},\cdots,\beta_{tp})^T, \qquad (2.19)$$

will be developed. Equation (2.4) is rewritten for convenience of presentation.

$$Y^*_{ij} = \tau_i^* + \beta_{y.x_{(k)}} \sum_{i=1}^{p} X^*_{ij_{(k)}} + \epsilon_{ij}, \qquad (2.20)$$

The problem is to devise a method for testing the equality of linear combinations of specified sets of the above vectors in order to determine the source and ultimately the nature of the heterogeneity among the regression coefficients. Each of the above vectors is pX1 and the form of (2.19) suggests that it is possible to construct a specified linear combination of these vectors of regression coefficients such that its expected value is zero. This vector of the linear combination will also be pX1. The elements of this new vector will be a linear combination of the individual regression coefficients. Recalling the rule for multiplying a matrix by a scalar and making use of the following definition, a linear combination of vectors can indeed be formed.

Definition: The vector Z is a linear combination [5, p. 370] of the vectors $\lambda_1, \lambda_2, \ldots, \lambda_t$ with coefficients $d_1, d_2, \ldots, d_t$ (scalars) if $Z = d_1\lambda_1 + d_2\lambda_2 + \ldots + d_t\lambda_t$. (2.21)

Choosing these coefficients, the d_i, such that their sum is zero and making use of the null hypothesis that the vectors of regression coefficients are equal, the following theorem is true.

Theorem I: The expected value of any linear combination of the vectors of regression coefficients, under the null hypothesis that the regression coefficients are equal and the restriction that the sum of the scalar coefficients be zero, is zero.

Proof: Let the vector $Z_{(pX1)}$ be a linear combination of the vectors of regression coefficients where the scalar coefficients d_k; $k = 1, 2, \ldots, t$; are chosen such that their sum is zero. That is, let

$$Z = \begin{bmatrix} d_1\hat{\beta}_{11} + d_2\hat{\beta}_{21} + \ldots + d_t\hat{\beta}_{t1} \\ d_1\hat{\beta}_{12} + d_2\hat{\beta}_{22} + \ldots + d_t\hat{\beta}_{t2} \\ \vdots \\ d_1\hat{\beta}_{1p} + d_2\hat{\beta}_{2p} + \ldots + d_t\hat{\beta}_{tp} \end{bmatrix}, \text{ where } \sum_{k=1}^{t} d_{(k)} = 0. \tag{2.22}$$

The expected value of Z is

$$E(Z) = \left(E\left(\sum_{k=1}^{t} d_k \hat{\beta}_{k1} \right), E\left(\sum_{k=1}^{t} d_k \hat{\beta}_{k2} \right), \ldots, E\left(\sum_{k=1}^{t} d_k \hat{\beta}_{kp} \right) \right)^T \quad (2.23)$$

but under H_o, $E(\hat{\beta}_{ki}) = E(\hat{\beta}_{k'i}) = \beta_i$; i=1, 2, ..., p; k = 1, 2, ..., t; k' = 1, 2, ..., t; $k \neq k$. Therefore

$$E(Z) = \left(\beta_1 E\left(\sum_{k=1}^{t} d_k \right), \beta_2 E\left(\sum_{k=1}^{t} d_k \right), \ldots, \beta_p E\left(\sum_{k=1}^{t} d_k \right)\right)^T$$

$$= (0, 0, \ldots, 0)^T , \quad (2.24)$$

since the scalar coefficients were chosen such that their sum is zero. Hence, the expected value of the Z vector, a linear combination of the regression coefficients, is zero.

Distribution of the Test Statistic

In equation (2.8) the estimators for $\hat{\beta}_{ki}$ were given. Using the notation of equation (2.20) and adopting matrix notation [2, p. 113] these estimators take the form

$$\hat{\beta}_k = (S^{-1} X^{*\prime} Y^*)_k, \; k = 1, 2, \ldots, t$$

where $$S^{-1} = [X^{*\prime} X^*]^{-1} \; . \quad (2.25)$$

This matrix notation and the following theorems will be helpful in ascertaining the distribution of a test statistic for the hypothesis of equality of regression coefficients.

Theorem II [2, p. 68]: If the pX1 vector Y is distributed normally with mean μ and variance V, and if A is a qXp matrix ($q \leq p$) of rank q, the vector W = AY is distributed as a q-variate normal distribution with mean $A\mu$ and covariance AVA'.

Theorem III [2, p. 56]: A linear combination of p-normally distributed variables, u_{ik}, is distributed normally with mean

$$\sum_{i=1}^{p} \alpha_i \mu_i$$

and variance

$$\sum_{i=1}^{p} \alpha_i^2 \sigma_i^2 + \sum_{\substack{i=1 \\ i \neq j}}^{p} \sum_{j=1}^{p} \alpha_i \alpha_j \sigma_{ij}$$

Theorem IV [4, p. 348]: The vectors of regression coefficients, $(\hat{\beta}_{k1}, \hat{\beta}_{k2}, \ldots, \hat{\beta}_{kp})^T$ are distributed as the p-variate normal with mean vector β_k and variance $\sigma_k^2 \, S^{-1}{}_k$; $k = 1, 2, \ldots, t$.

Recall that the Z vector given in equation (2.22) is pX1. By Theorems I, III, and IV, each element has mean zero and by Theorems III and IV each element has variance

$$\sum_{k=1}^{t} d^2{}_k \sigma^2{}_k + \sum_{\substack{k=1 \\ k \neq k'}}^{t} \sum_{k'=1}^{t} d_{kk'} \sigma_{kk'} \quad . \tag{2.26}$$

However, since $\hat{\beta}_k$ and $\hat{\beta}_{k'}$ were computed from independent sets of data, their covariance is zero and the variance reduces to $\sum_{k=1}^{t} d^2{}_k \sigma^2{}_k$.
Now let A be a pX1 vector of ones having rank 1, then by Theorem II, the 1 X 1 vector u = AZ is a univariate normal with mean and variance given

$$\mu = \sum_{k=1}^{t} d_k \left[\sum_{i=1}^{p} a_i \beta_{ki} \right] \qquad \text{and}$$

$$V = \sum_{k=1}^{t} d^2{}_k \left[\sum_{i=1}^{p} a^2{}_i C_{iik} \sigma^2{}_k + \sum_{\substack{i=1 \\ i \neq j}}^{p} \sum_{j=1}^{p} a_i j_i (-C_{ijk}) \sigma_k^2 \right] \tag{2.27}$$

respectively, where C_{ijk} is the $(i,j)^{th}$ element of $S^{-1}{}_k$.
Since,

Theorem V [2, p. 116]: The best linear unbiased estimate of any linear combination of the β_{ki} is the same linear combination of the best linear unbiased estimate of the β_{ki}. Then the best linear unbiased estimate of u is obtained by replacing β_{ki} by $\hat{\beta}_{ki}$, that is, the best linear unbiased estimate of u = AZ is

$$\mu = \sum_{k=1}^{t} d_k \left[\sum_{i=1}^{p} a_i \hat{\beta}_{ki} \right] \quad . \tag{2.28}$$

Significance Test for the Test Statistic

The sample statistic to be used for testing the assumption of homogeneity of regression coefficients was given in equation (2.28). This section shall be devoted to the development of a suitable test of significance for this linear contrast of regression coefficients. In this development the following theorems will be used.

Theorem VI [2, p. 113]: If $Y_k = X_k\beta_k + \epsilon_k$ is a general linear hypothesis of full rank and if ϵ_k is distributed normally with mean zero and variance $\sigma_k{}^2 I$ the estimators

$$\hat{\beta}_k = (S^{-1}X\ Y)_k \text{ and } \hat{\sigma}_k{}^2 = \left[\frac{Y'(I-X\ S^{-1}X')Y}{n-p} \right]_k \tag{2.29}$$

have the following properties:

(1) Consistent, (2) Efficient, (3) Unbiased, (4) Sufficient, (5) $\hat{\beta}_k$ is distributed $N(\beta_k, \sigma^2 S^{-1}{}_k)$, (6) Complete, (7) Minimum variance unbiased, (8) $\frac{(n-p)\ \hat{\sigma}^2 k}{\sigma^2}$ $X^2 n-p$, (9) $\hat{\beta}_k$ and $\hat{\sigma}_k{}^2$ are independent.

Theorem VII [2, p. 77]: If $\theta_1, \theta_2, \ldots, \theta_k$ are independently distributed chi square variates with $n_1, n_2, \ldots, n_k$ degrees of freedom, respectively, then $w = \sum_{i=1}^{k} \theta_i$ is distributed as chi square having-degrees of freedom $N = \overset{k}{\Sigma}\, n_i$.

Theorem VIII [4, p. 233]: Let X be a normal variate with mean zero and variance one. Let w be a chi square variate with k degrees of freedom and let X and w be independent. Then the random variable

$$t = X/\sqrt{w/k} \tag{2.30}$$

is distributed as student's t with k degrees of freedom.

Recall now that $\hat{\sigma}_k{}^2$ and $\hat{\sigma}_{k'}{}^2$; $k \neq k'$, $k = 1, 2, \ldots, t$, $k' = 1, 2, \ldots, t$; were computed from different sets of data and are therefore independent. Thus, the variables $\dfrac{(n-p)_k\, \hat{\sigma}_k{}^2}{\sigma^2}$ and $\dfrac{(n-p)_{k'}\, \hat{\sigma}_{k'}{}^2}{\sigma^2}$ are also independent. Furthermore, by property (8) of Theorem VI, these two variables are distributed as chi square variables. Applying Theorem VII to these results, the chi square variate

$$w' = \sum_{k=1}^{t} \frac{(n-p)_k\, \hat{\sigma}_k{}^2}{\sigma^2} \tag{2.31}$$

can be written having degrees of freedom

$$N = \sum_{k=1}^{t} (n-p)_k = t(n-p) .$$

Note also that $\hat{\beta}_k$ and $\hat{\beta}_{k'}$, $k \neq k'$, $k = 1, 2, \ldots, t$; $k' = 1, 2, \ldots, t$; are independent, being computed from different sets of data. Also, by property (9) of Theorem VI, $\hat{\beta}_k$ and $\hat{\sigma}_k{}^2$ are independent. Furthermore, $\hat{\beta}_k$ and $\hat{\sigma}_k{}^2$ are independent as they were computed from different sets of data, and $\hat{u}$, a linear combination of the $\hat{\beta}_k$, is independent of w', a linear combination of the $\hat{\sigma}_k{}^2$.

Let

$$v = \frac{\hat{u} - E(\hat{u})}{\sigma\sqrt{\{\sum_{k=1}^{t} d^2_k\,[\sum_{i=1}^{p} a_i^2\ C_{iik} + \sum_{\substack{i=1\\ 1\neq j}}^{p} \sum_{j=1}^{p} a_i a_j(-C_{ijk})]\}}}, \tag{2.32}$$

then v is distributed normally with mean zero and variance one, and by Theorem VIII the random variable

$$w = \frac{v}{w' \,/\, \sum_{k=1}^{t} (n-p)_k}$$

$$= \frac{\hat{u}}{\sigma_p\sqrt{\{\sum_{k} d^2_k\ [\sum_{i}^{p} a_i^2\, C_{iik} + \sum_{\substack{i=1\\ i\neq j}}^{p} \sum_{j=1}^{p} a_i a_j(-C_{ijk})]\}}}\quad, \tag{2.33}$$

where σ_p^2 is the pooled estimate of the error variance obtained from the analysis of covariance table, is distributed as student's t with $N = t(n-p)$ degrees of freedom.

This w ratio is to be compared to tabular student t values to test the hypothesis of homogeneity of regression coefficients in the analysis of covariance randomized complete block design.

CHAPTER III

APPLICATION OF THE METHODS

The following set of hypothetical data, Table 2 has been devised to fit the model for the analysis of covariance, randomized complete block design. Furthermore, this data has been arranged such that the null hypothesis of homogeneity of regression coefficients is false.

Test of the Null Hypothesis by the F-test

In addition to the usual sums of squares computed for this design, the sums of squares for treatment will be partitioned into portions due to each treatment group and a remainder or residual sum of squares. Equations (2.9) are used to partition the treatment sum of squares. These sums of squares will then be used to test the null hypothesis of homogeneity of regression coefficients.

The residual values used in computing the contributions of the individual treatment groups are given in Table 3. The values are obtained by subtracting the mean for each variable within a block from its row block value. For example, the value 3.3 for T_1 and Y^* in block 1 is 13.3 - 10.0.

In Table 4 the sums of squares and cross products for these data are given. The total, block, treatment and error sums of cross products were computed using the standard equation. The values for treatment

TABLE 2

DATA

	BLOCK 1			BLOCK 2			BLOCK 3			BLOCK 4			TOTAL		
	Y	X	Z	Y	X	Z	Y	X	Z	Y	X	Z	Y	X	Z
t_1	13.3	5.5	07.8	14.5	5.8	9.1	16.9	7.4	9.8	20.2	10.1	11.8	64.9	28.8	38.5
t_2	7.9	4.4	08.3	12.4	6.6	8.8	12.2	6.6	10.1	13.9	8.2	12.8	46.4	25.8	40.0
t_3	8.8	5.1	07.9	12.1	5.9	9.1	12.9	7.0	10.1	15.1	9.6	12.0	48.9	27.6	39.1
Sum	30.0	15.0	24.0	39.0	18.3	27.0	42.0	21.0	30.0	49.2	27.9	36.6	160.2	82.2	117.6
Mean	10.0	5.0	8.0	13.0	6.1	9.0	14.0	7.0	10.0	16.4	9.3	12.2			

TABLE 3

DIFFERENCES

	BLOCK 1			BLOCK 2			BLOCK 3			BLOCK 4			TOTAL		
	Y*	X*	Z*	Y*	X*	Z*	Y*	X*	Z*	Y*	X*	Z*	Y*	X*	Z*
t_1	3.3	.5	-.2	1.5	-.3	.1	2.9	.4	-.2	3.8	.8	-.4	11.5	1.4	-.7
t_2	-2.1	-.6	.3	- .6	.5	-.2	-1.8	-.4	.1	-2.5	-1.1	.6	- 7.0	-1.6	.8
t_3	-1.2	.1	-.1	- .9	-.2	.1	-1.1	0	.1	-1.3	.3	-.2	- 4.5	.2	-.1

TABLE 4

SUMS OF CROSS PRODUCTS

Source	df	xx	xy	xz	zz	zy	yy
TOTAL	11	33.29	27.92	51.59	29.86	36.66	118.61
BLOCKS	3	30.03	29.52	42.09	29.04	41.28	63.21
T_1	1	.65	1.375	- .285	.127	- .597	2.928
T_2	1	1.34	1.630	- .660	.340	- .790	2.010
T_3	1	.13	- .105	- .085	.068	.067	.087
Error	6	2.12	2.900	- 1.030	.535	- 1.320	5.025
Treatment + Error	8	3.26	9.500	- 1.60	.820	- 4.62	55.400
Treatment	2	1.14	6.600	- .570	.285	- 3.30	50.375

and error will be used to check values obtained from the individual treatment groups. The sums of squares for X in treatment group one is $(.5)^2 + (-.3)^2 + (.4)^2 + (.8)^2 - (1.4)^2 (\frac{1}{4}) = .65$, and the cross product for XY in treatment group one is $(3.3)(.5) + (1.5)(-.3) + (2.9)(.4) + (3.8)(.8) - (\frac{1}{4})(11.5)(1.4) = 1.375$. The remainder of the sums of squares and cross products for the individual treatment groups are obtained analogously. The error sums of squares and cross products can be obtained by the application of equations (2.10). The sums of the sums of squares and cross products for the individual treatment groups agree with the values for the error line computed by the usual method.

The regression coefficients, sums of squares of Y regressed on X, sums of squares of Y regressed on Z and the adjusted sums of squares are given in Table 5. Equations (2.8) were used to compute the regression coefficients for the treatment groups. These equations are also applicable in computing the other regression coefficients if the proper values from Table 4 are used. The last equation of equations (2.9) was used to compute the next three columns of Table 5. That is, the value 4.649 is simply (3.381)(1.375), and -1.672 is (2.887)(-.597) and $.002664 = 2.928 - 4.649 - (-1.672)$. The sums of squares for S_1 is the sum of the adjusted sums of squares for the individual treatment groups. The sum of squares for S_2 is the adjusted error sum of squares minus the sum of squares for S_1 or it can also be obtained by summing the sums of squares of Y regressed on X and Y regressed on Z for the individual treatment groups and then subtracting from this sum the sums of squares of Y regressed on X and Y regressed on Z in the error line.

The adjusted sum of squares for treatment is obtained by

TABLE 5

REGRESSION COEFFICIENTS, REGRESSION SUMS OF SQUARES
AND ADJUSTED SUMS OF SQUARES

Source	d.f.	$\beta_{y.x}$	$\beta_{y.z}$	Regression of Y on X	Regression of Y on Z	Adjusted Sum of Squares
Total Block						
T_1	1	3.381	2.887	21.649	-1.622	.002664
T_2	1	1.640	.860	2.673	- .679	.016200
T_3	1	- .895	- .133	+ .040	- .009	.001936
S_1	3					.020800
Error	4	2.618	2.573	7.592	-3.396	.829160
S_2	1					.808360
Treatment + Error	6	3.510	1.226	33.345	-5.664	27.719120
Treatment	2					26.88996
S_3	1	0	0	0	0	50.375000

subtracting the error sum of squares from the treatment plus error sum of squares. The sum of squares for S_3 is obtained by using the treatment sums of squares line of Table 4 to compute the regression sums of squares. These are then subtracted from 50.375, the treatment sum of squares for Y. The regression sums of squares were zero for this particular case so no reduction was obtained.

Tables 4 and 5 constitute an analysis of covariance for these data. From equation (2.18 the test of the null hypothesis that the regression coefficients are homogeneous is

$$F = \frac{S_2/\text{df for } S_2}{S_1/\text{df for } S_1} = \frac{.180836/1}{.0208/3} = 116.59.$$

Since the tabular F for 1 and 3 degrees of freedom at the .05 level of significance is 10.13, the null hypothesis can be rejected.

There are two other hypotheses which could be tested by using values from Table 5. The first is a test for differences among adjusted treatment means. The standard test for this is

$$F = \frac{\text{Adjusted treatment SS/df}}{\text{Adjusted error SS/df}} = \frac{26.88996/2}{.82916/4} = 64.86.$$

The second hypothesis which can be tested is whether the regression for means is linear. The test for this is

$$F = \frac{\text{Adjusted SS for } S_3/\text{df}}{\text{Adjusted error SS/df}} = \frac{50.375/1}{.82916/4} = 243.02.$$

However, neither of the latter tests are valid for these data since the hypothesis of homogeneity of regression coefficients was rejected.

Determining Sources of Heterogeneity

The t test developed in the last portion of CHAPTER II will be applied to the regression coefficients derived from the data in Table 2 to determine sources of heterogeneity indicated by the large F value. The regression coefficients are given in Table 5. The null hypothesis for these data is that specified linear combinations of the vectors

$$\begin{pmatrix}\beta_{11}\\ \\ \beta_{12}\end{pmatrix}, \quad \begin{pmatrix}\beta_{21}\\ \\ \beta_{22}\end{pmatrix}, \quad \begin{pmatrix}\beta_{31}\\ \\ \beta_{32}\end{pmatrix}, \text{ are zero.}$$

The only restriction in the formation of the vector of the linear combination of the regression coefficients is that the sum of the coefficients be zero. In the present example the orthogonal contrasts (-1, 0, 1) and (-2, 1, 1) were chosen for illustrative purposes. The investigator may have reason to believe that the response to a given treatment is substantially greater (or less than) other treatments. He would therefore want to test this belief by using a suitable set of coefficients. Although any choice of scalar multipliers such that their sum is zero can be used, sets consisting of small integers greatly simplifies the procedure for computing the Z vector and its variance.

There are three treatments and two covariates in this hypothetical set of data from which the vectors of regression coefficients

$$\begin{pmatrix}3.381\\ 2.887\end{pmatrix}, \quad \begin{pmatrix}1.640\\ .860\end{pmatrix} \text{ and } \begin{pmatrix}-.895\\ -.133\end{pmatrix} \text{ are computed.}$$

Using equation (2.22), the Z vectors, which are specified linear combinations of the regression coefficients, are

$$Z_1 = \begin{pmatrix} (1)\ 3.381 + (0)\ (1.640) + (-1)\ (-.895) \\ (1)\ 2.887 + (0)\ (\ .860) + (-1)\ (-.133) \end{pmatrix}$$

and

$$Z_2 \quad \begin{pmatrix} (-2)(3.381) + (1)(1.640) + (1)(-.895) \\ -2\ (2.887) + (1)(\ .860) + (1)(-.133) \end{pmatrix}$$

from which, $\hat{u}_1 = 7.296$, and $\hat{u}_2 = -11.064$

The next step is to compute the variance of $\hat{u}_1$ and $\hat{u}_2$. The $[X*'\ X]^{-1}$ matrices to be used are

$$[X*'\ X]^{-1}_1 = \begin{bmatrix} .65 & -.285 \\ & .127 \end{bmatrix}^{-1} = \begin{bmatrix} 95.849 & 215.094 \\ & 490.566 \end{bmatrix}$$

$$[X*'\ X]^{-1}_2 = \begin{bmatrix} 1.34 & -.66 \\ & .34 \end{bmatrix}^{-1} = \begin{bmatrix} 17.000 & 33.000 \\ & 67.000 \end{bmatrix}$$

$$[X*'\ X]^{-1}_3 = \begin{bmatrix} .130 & -.085 \\ & .068 \end{bmatrix}^{-1} = \begin{bmatrix} 42.105 & 52.632 \\ & 80.495 \end{bmatrix} .$$

The variance of $\hat{u}$ is given by equation (2.27). Making the proper substitutions into this equation and assuming homogeneity of variance gives

$$\begin{aligned} V\ (\hat{u}) = (.0208/3)\ \{&(1)^2(95.849 + 490.566 - (2)(215.094)) \\ + \quad &(0)^2(17.000 + 67.000 - 2\ (33.000) \\ + \quad &(-1)^2(42.105 + 80.495 - 2\ (52.632))\} \\ = .120336 & \end{aligned}$$

Equation (2.33), which is Student's t, is to be used to test the null hypothesis of homogeneity of regression coefficients for treatment planes one and three. The test is

$$w = \frac{\hat{u}_1}{\sqrt{V(\hat{u}_1)}} \qquad \frac{7.296}{\sqrt{1.20336}} = 6.6 \ .$$

Student's t with 3 degrees of freedom at the .05 significance level is 3.18, therefore reject the hypothesis that the chosen sets of regression coefficients are homogeneous.

Two other comparisons between pairs of treatment planes are possible using the contrasts (-1, 1, 0) and (0, 1, -1). This appears to be analogous to the least significant difference test in the analysis of variance [7, p. 106].

To test the assumption that twice the vector of coefficients for treatment group one equals the sum of the vectors for treatment groups two and three use

$$w = \hat{u}_2 / \sqrt{V(\hat{u}_2)}$$

$$= -11.064 / \sqrt{4.577269} = -5.1 \ .$$

Reject the null hypothesis. This test would seem applicable when the regression coefficients for treatment groups two and three are equal thus allowing the pooling of their effects to test the vector for treatment group one.

CHAPTER IV

SUMMARY

In CHAPTER II two methods for testing the hypothesis of homogeneity of regression coefficients for the analysis of covariance, randomized complete block design were developed. Until now there was no known statistical technique for testing this hypothesis.

The first method developed parallels the standard test for this hypothesis for the analysis of covariance, completely random design.

This method is based upon the assumption that the sum of squares about an overall treatment line is no larger than the sum of the sums of squares about the individual treatment group lines. This is not an unreasonable assumption for if the slopes differ among the different treatment groups, then the overall line will have a larger variance than can be expected by chance alone.

Although the computations involved are not difficult, they do lengthen the analysis, however, this increase is offset to some extent by the fact that it is not necessary to compute the total and block sums of squares. This is due to the fact that the error sums of squares can be obtained easily by summing the sums of squares for the individual treatment groups. It is also possible to test whether the regression for treatment means is linear. This test is not available in the standard analysis of this design.

The second method developed is of value in testing specified linear contrasts of the vectors of regression coefficients for homogeneity. If the chosen treatment groups react at a different rate over a given range of concomitant variables then this test will indicate the nature of this difference. Standard theoretical techniques were used to develop this test. The statistic used for testing the null hypotheses is student's t with (r-1-p)t degrees of freedom, where r is the number of blocks, p is the number of covariates and t is the number of treatment groups.

These two tests fulfill the need which has existed for a method of testing the assumption of homogeneity of regression coefficients in the analysis of covariance, randomized complete block design. In the past an investigator either ignored the possibility of violating this assumption and the associated consequences when testing for differences among adjusted treatment means or used some other analytical method if it was available. It is no longer necessary to sidestep the issue.

The fact that it is now possible to test this assumption makes this a worthwhile tool to an investigator, since it can add to the understanding of the responses elicited by various treatments over a range of concomitant variables.

REFERENCES

1. Cochran, William G., "Analysis of Covariance: Its Nature and Uses", Biometrics, Vol. 13 (1957), pp. 261-281.

2. Graybill, Franklin A., An Introduction to Linear Statistical Models, Vol. 1. New York: McGraw-Hill Book Company, Inc., 1961.

3. Harkins, Ralph D., "A Method for Testing Homogeneity of Regression Coefficients in the Analysis of Covariance: Two or More Covariates", Unpublished Thesis for the Masters Degree, The University of Oklahoma, (1964).

4. Mood, Alexander M. and Graybill, Franklin A., Introduction to the Theory of Statistics. New York: McGraw-Hill Book Company, Inc., 1963.

5. Scheffe, Henry, The Analysis of Variance. New York: John Wiley and Sons, Inc., 1959.

6. Snedecor, George W. and Cochran, William G., Statistical Methods Applied to Experiments in Agriculture and Biology. 5th Edition. Ames, Iowa: The Iowa State College Press, 1957.

7. Steel, Robert G. D. and Torrie, James H., Principles and Procedures of Statistics with Special Reference to the Biological Sciences. New York: McGraw-Hill Book Company, Inc., 1960.

8. Winer, B. J., Statistical Principles in Experimental Design: New York: McGraw-Hill Book Company, Inc., 1962.