Covariance Analysis for Split Plot and Split Block Designs

WALTER T. FEDERER and MICHAEL P. MEREDITH[1]

BU-1131-M[2] June, 1991

Proper methodology for the analysis of covariance for experiments designed in a split plot or split block design is not found in the statistical literature. Analyses for these designs are often performed incompletely or even incorrectly. This is especially true when popular statistical computer-software packages are used for the analysis of these designs. This paper provides several appropriate models, ANOVA tables, and standard errors for comparisons from experiments arranged in a standard split plot, split-split plot, or split block design where a covariate has been measured on the smallest sized experimental unit.

---

1. INTRODUCTION

The philosophical nature, concepts, population structures, and usage of multiple error terms in the general families of split plot, split-split plot, split block (two-way whole plots), and more complex designs are not well documented in statistical textbook literature. Likewise, information on covariance analyses for these designs is limited. Yates (1937) described one member of the family of split plot designs and this example is the one most frequently presented and discussed in the statistical literature. Some information (see Federer 1955, 1975, 1977) on these topics is available but does not appear to be widely known or used. This failure to use available information carries over into computer-software

---

[1] Walter T. Federer is a Consulting Statistician in the Mathematical Sciences Institute and Professor Emeritus in the Biometrics Unit, Cornell University, Ithaca, NY 14853. Michael P. Meredith is a statistical consultant at the Miami Valley Laboratories, The Procter & Gamble Company, P. O. Box 398707, Cincinnati, OH 45239.

[2]In the Technical Report Series of the Biometrics Unit, Cornell University, Ithaca, NY 14853.

Keywords: Analysis of covariance, Split split plot design, Variances, Repeated measures, Approximate degrees of freedom, Response models, Two-way whole plots.

programs (see Federer *et al.* 1979, 1987a, 1987b, 1987c, Searle *et al.* 1982a, 1982b, 1982c, Miles-McDermott *et al.* 1988, and Meredith *et al.* 1988). In an analysis of covariance (ANCOVA), there will be as many covariate regressions as there are error terms in the analysis of variance (ANOVA). Multiple error terms and multiple error regressions cause difficulty in the statistical analysis of data.

Herein we shall discuss the ANCOVA for three specific designs, i.e.,

(i) the standard split plot design where the whole plot treatments are in a randomized complete block design and split plot treatments are randomized within each whole plot,

(ii) a split-split plot design as in (i) except that the split plot is further split to have whole plot treatments, split plot treatments, and split-split plot treatments, and

(iii) a split block design or two-way whole plot design where each set of treatments are in a randomized complete block design arrangement.

2. SPLIT PLOT EXPERIMENT DESIGNS

The almost universal split plot experiment design discussed in statistics textbooks is that with whole plot treatments in a randomized complete block design and split plots are completely randomized within each whole plot. Denote this as the standard design. Federer (1955, 1976) has pointed out that there is a vast variety of split plot experiment designs used in practice. There are many different experiment designs for whole plot treatments as well as for split plot treatments. Each of these designs may be used to model data arising from repeated measures experiments. The split units may be temporal or spatial.

For brevity, only the standard split plot experiment design will be considered in detail. Many response models may be used for the wide variety of experiments designed as a split plot but we shall confine ourselves to the linear model in Federer (1955). Let the hij*th* observation Y_{hij} with an associated covariate Z_{hij} be represented as follows:

$$Y_{hij} = \mu + \rho_h + \tau_i + \delta_{hi} + \alpha_j + \alpha\tau_{ij} + \beta_1(\bar{Z}_{hi.} - \bar{Z}_{...}) + \beta_2(Z_{hij} - \bar{Z}_{hi.}) + \epsilon_{hij} \, , \quad (1)$$

where μ is an overall mean effect, τ_i is the effect of the i*th* whole plot treatment, α_j is the effect of the j*th* subplot treatment, $\alpha\tau_{ij}$ is the interaction effect for the ij*th* combination of whole plot treatment i and split plot treatment j, ρ_h is a random block effect distributed with mean zero and variance σ^2_ρ, δ_{hi} is a random whole plot error effect normally distributed with mean zero and variance σ^2_δ, and ϵ_{hij} is a random split plot error effect normally distributed and with mean zero and variance σ^2_ϵ. The errors are assumed to be independent. $\bar{Z}_{hi.}$ is the mean of the covariate for the hi*th* whole plot, $\bar{Z}_{...}$ is the overall mean of the covariate (i.e., the usual dot and bar notation), $h = 1,\cdots,r$, $i = 1,\cdots,a$, $j = 1,\cdots,s$, β_1 is a whole plot linear regression coefficient of the Y whole plot residuals on the Z whole plot residuals, and β_2 is a split plot linear regression of the Y split plot residuals on the Z split plot residuals. Estimates of β_1 and β_2, $\hat{\beta}_1$ and $\hat{\beta}_2$, are necessary to correctly adjust means. The purpose of using covariates is to reduce both the variation and bias in observed Y means by measuring an associated covariate. The reduction must then occur in an error or residual line in the ANOVA.

For response model equation (1), the ANCOVA is given in Table 1. The sums of products are computed in the usual manner. For example, $T_{yz} = \sum_h \sum_i \sum_j Y_{hij} Z_{hij}$, $A_{yz} = \sum_h \sum_i \hat{\delta}_{yhi} \hat{\delta}_{zhi}$, where $\hat{\delta}_{yhi}$ is the residual for the variable Y alone and ϵ_{zhig} and $\hat{\delta}_{zhi}$ is the residual for the variable Z alone, and $B_{yz} = \sum_h \sum_i \sum_j \hat{\epsilon}_{yhij} \hat{\epsilon}_{zhij}$ where the $\hat{\epsilon}_{yhij}$ and $\hat{\epsilon}_{zhig}$ are the computed split plot residuals for variables y and z. The above computations still hold for nonorthogonal or unbalanced experiment designs. The mean squares in ANCOVA are obtained by dividing by the appropriate degrees of freedom. If, in addition to an ANCOVA, it is desired to obtain F-statistics, then the ratios W'_{yy}(ar-r-a)/A'_{yy}(a-1), S'_{yy}[a(r-1)(s-1)-1]/B'_{yy}(s-1), and I'_{yy}[a(r-1)(s-1)-1]/B'_{yy}(a-1)(s-1) may be computed (see Table 1). Given that the δ_{hi} and ϵ_{hij} are NID, the probability of obtaining a larger F-statistic may be obtained from prepared tables or computer programs. Even if normality does not hold, the probabilities will be approximately correct for many situations, provided treatments are randomized for each size of experimental unit.

The various Y means adjusted for the covariate Z are:

$$\bar{Y}_{.i.}(\text{adj.}) = \bar{Y}_{.i.} - \hat{\beta}_1(\bar{Z}_{.i.} - \bar{Z}_{...}) = \bar{Y}^*_{.i.}\ ,$$

$$\bar{Y}_{..j}(\text{adj.}) = \bar{Y}_{..j} - \hat{\beta}_2(\bar{Z}_{..j} - \bar{Z}_{...}) = \bar{Y}^*_{..j}\ ,$$

Table 1. ANCOVA for Equation (1) for a Split Plot Experiment Design with Covariate Measured on Each Split Plot Experimental Unit[1]

Source of Variation	Degrees of Freedom (df)	Sums of Products YY	YZ	ZZ	df	Adjusted Sums of Squares
Total	ars	T_{yy}	T_{yz}	T_{zz}		
Correction for Mean	1	M_{yy}	M_{yz}	M_{zz}		
Block	(r–1)	R_{yy}	R_{yz}	R_{zz}		
Whole Plot = W	(a–1)	W_{yy}	W_{yz}	W_{zz}		
Error (a)	(a–1)(r–1)	A_{yy}	A_{yz}	A_{zz}	(ar–a–r)	$A_{yy} - \frac{A_{yz}^2}{A_{zz}} = A'_{yy}$
Split Plot = S	(s–1)	S_{yy}	S_{yz}	S_{zz}		
S × W	(a–1)(s–1)	I_{yy}	I_{yz}	I_{zz}	(as–a–s)	$I_{yy} - \frac{I_{yz}^2}{I_{zz}} = I'_{yy}$
Error (b)	a(r–1)(s–1)	B_{yy}	B_{yz}	B_{zz}	a(r–1)(s–1)–1	$B_{yy} - \frac{B_{yz}^2}{B_{zz}} = B'_{yy}$

Source of Variation	df	Adjusted Sums of Squares
Whole Plot (adj. for $\hat{\beta}_1$)	(a–1)	$W_{yy} - \frac{(W_{yz} + A_{yz})^2}{W_{zz} + A_{zz}} + \frac{A_{yz}^2}{A_{zz}} = W'_{yy}$
Split Plot (adj. for $\hat{\beta}_2$)	(s–1)	$S_{yy} - \frac{(S_{yz} + B_{yz})^2}{S_{zz} + B_{zz}} + \frac{B_{yz}^2}{B_{zz}} = S'_{yy}$
S × W (adj. for $\hat{\beta}_2$)	(a–1)(s–1)	$I_{yy} - \frac{(I_{yz} + B_{yz})^2}{I_{zz} + B_{zz}} + \frac{B_{yz}^2}{B_{zz}} = I'_{yy}$

[1] Mean squares are obtained by dividing sums of squares by the appropriate degrees of freedom.

and

$$\bar{Y}_{.ij}(\text{adj.}) = \bar{Y}_{.ij} - \hat{\beta}_1(\bar{Z}_{.i.} - \bar{Z}_{...}) - \hat{\beta}_2(\bar{Z}_{.ij} - \bar{Z}_{.i.}) = \bar{Y}^*_{.ij},$$

where $\hat{\beta}_1 = A_{yz}/A_{zz}$, $\hat{\beta}_2 = B_{yz}/B_{zz}$, and the usual dot notation is used for the various means.

Estimated variances of a difference between two adjusted means for $i \neq i'$ and $j \neq j'$ are summarized below:

Variance of a difference between two adjusted whole plot treatment means

$$\hat{V}(\bar{Y}^*_{.i.} - \bar{Y}^*_{.i'.}) = E_a\left[\frac{2}{rs} + \frac{(\bar{Z}_{.i.} - \bar{Z}_{.i'.})^2}{A_{zz}}\right].$$

Variance of a difference between two adjusted split plot treatment means

$$\hat{V}(\bar{Y}^*_{..j} - \bar{Y}^*_{..j'}) = E_b\left[\frac{2}{ar} + \frac{(\bar{Z}_{..j} - \bar{Z}_{..j'})^2}{B_{zz}}\right].$$

Variance of a difference between two adjusted split plot treatment means for the same whole plot

$$\hat{V}(\bar{Y}^*_{.ij} - \bar{Y}^*_{.ij'}) = E_b\left[\frac{2}{r} + \frac{(\bar{Z}_{.ij} - \bar{Z}_{.ij'})^2}{B_{zz}}\right].$$

Variance of a difference between two adjusted whole plot treatment means for the same split plot treatment

$$\hat{V}(\bar{Y}^*_{.ij} - \bar{Y}^*_{.i'j}) = \frac{2}{r}(E_b + \hat{\sigma}^2_\delta) + E_a\frac{(\bar{Z}_{.i.} - \bar{Z}_{.i'.})^2}{A_{zz}} + E_b\frac{(\bar{Z}_{.ij} - \bar{Z}_{.i'j} - \bar{Z}_{.i.} + \bar{Z}_{.i'.})^2}{B_{zz}}.$$

The estimated variance of a difference between means for two different split plot treatments from two different whole plot treatments is of the same form as the last variance. In the above formulae,

$$E_a = (\hat{\sigma}^2_\epsilon + s\hat{\sigma}^2_\delta) = \frac{A'_{yy}}{ar-a-r}, \quad E_b = \hat{\sigma}^2_\epsilon = \frac{B'_{yy}}{a(r-1)(s-1)-1},$$

and

$$\hat{\sigma}_\delta^2 = \frac{(\hat{\sigma}_\epsilon^2 + s\hat{\sigma}_\delta^2) - \hat{\sigma}_\epsilon^2}{s} = \frac{E_a - E_b}{s}.$$

$E_b = \hat{\sigma}_\epsilon^2$ is associated with a(r−1)(s−1)−1 degrees of freedom, $E_a = (\hat{\sigma}_\epsilon^2 + s\hat{\sigma}_\delta^2)$ is associated with ar−r−a degrees of freedom, and the degrees of freedom for the estimated variance $\left[(s-1)E_a + E_b\right]/s = E_b + \hat{\sigma}_\delta^2$ above are approximated as the degree of freedom $\hat{f}$ associated with

$$t_\alpha(\hat{f}) = \frac{(s-1)E_a\, t_\alpha\,(ar-r-a) + E_b\, t_\alpha\,[a(r-1)(s-1)-1]}{(s-1)\,(\hat{\sigma}_\epsilon^2 + s\hat{\sigma}_\delta^2) + \hat{\sigma}_\epsilon^2}$$

where $t_\alpha(\hat{f})$ is the tabulated value of the t-statistic at the α percentage level for $\hat{f}$ degrees of freedom. This approximation generally underestimates the degrees of freedom for this variance (see Cochran and Cox, 1957, and Grimes and Federer, 1984).

A method for approximating the degrees of freedom for $\hat{\sigma}_\epsilon^2 + \hat{\sigma}_\delta^2$ has been given by Satterthwaite (1946) and Gaylor and Hopper (1969) as

$$\hat{f} = \frac{\left[\frac{1}{s} E_a + \frac{(s-1)}{s} E_b\right]^2}{\frac{\left(\frac{1}{s} E_a\right)^2}{ar-a-r} + \frac{\left(\frac{s-1}{s} E_b\right)^2}{a(r-1)(s-1)-1}}.$$

Direct extension of the above formulae for combining three or more variances may be made (Grimes and Federer, 1984). Note that in $V(\bar{Y}^*_{\cdot ij} - \bar{Y}^*_{\cdot ij'})$, the three variances, $(E_b + \hat{\sigma}_\delta^2)$, E_a, and E_b are combined. This also could be considered as combining the two variances E_a and E_b with a variance component $\hat{\sigma}_\delta^2$. The degrees of freedom for a variance component may be approximated by the last formula above (Gaylor and Hopper, 1984). Given the above variances, one may now compare individual pairs of means. More general comparisons and their associated variances may also be calculated.

Some authors (e.g., Cochran and Cox, 1957) consider that there is a correlation between the split plot experimental units. Hence, the whole plot expected error mean square would be given as σ^2 and the split plot error would be written as $\sigma^2(1-\rho) = \sigma_\epsilon^2$ where the correlation ρ is equal to $s\sigma_\delta^2/\sigma^2$

under the present formulation. Although this formulation is useful for many situations it is not of universal application; e.g., when measurement error or competition exists between split plot experimental units but not between whole plot experimental units. Population structure and statistical modeling should be carefully considered for any investigation.

For some experiments and for some variables, the formulation of the response model as in (1) is inappropriate. As formulated (1) has two error effects, the δ_{hi} and ϵ_{hij}. However, when the whole plot treatments represent a random sample of treatments from a population, then the τ_i are distributed with mean zero and variance σ^2_τ. An appropriate error term for the fixed split plot treatment effects α_j would be the whole plot by split plot treatment interaction mean square. The $\alpha\tau_{ij}$ would have $E_{i,j}\alpha\tau_{ij} = 0$ and variance $\sigma^2_{\alpha\tau}$. Likewise in an ANCOVA, the appropriate regression for split plot treatment means would be computed from the interaction line rather than the error (b) line (see Table 1). In other situations, the split plot treatments or both split plot and whole plot treatments could be considered as a random sample of treatments and the effects would be random rather than fixed effects. Appropriate modifications are required for both situations.

3. SPLIT-SPLIT PLOT EXPERIMENT DESIGNS

For this class of designs, various experiment designs may be used for whole plot treatments, for split plot treatments, and for split-split plot treatments. We shall confine our remarks to a single member of this class, i.e., the whole plot treatments are arranged in a randomized complete blocks design, split plot treatments randomly allocated to split plot experimental units within each whole plot unit, and split-split plot treatments randomly assigned to the split-split plot experimental units within each split plot experimental unit. There will be r randomizations for the a whole plot treatments, ra randomizations for the s split plot treatments, and ras randomizations for the p split-split plot treatments. The treatment design considered here is a three-factor factorial with asp combinations, but it should be noted that other treatment designs are possible. The factors are assumed to be fixed effects to simplify presentation.

One possible response model for the above experiment and treatment design for a variable Y with a covariate Z is:

$$Y_{hijk} = \mu + \rho_h + \tau_i + \delta_{hi} + \beta_1 (\bar{Z}_{hi..} - \bar{Z}_{....}) + \alpha_j + \alpha\tau_{ij} + \epsilon_{hij}$$
$$+ \beta_2 (\bar{Z}_{hij.} - \bar{Z}_{hi..}) + \gamma_k + \gamma\tau_{ik} + \alpha\gamma_{jk} + \alpha\gamma\tau_{ijk} + \pi_{hijk} \qquad (2)$$
$$+ \beta_3 (Z_{hijk} - \bar{Z}_{hij.}) ,$$

where the first nine effects are as defined for equation (1), γ_k is the effect of the k*th* split-split plot treatment, $\gamma\tau_{ik}$ is a two-factor interaction effect for combination ik, $\alpha\gamma_{jk}$ is a two-factor interaction effect for combination jk, $\alpha\gamma\tau_{ijk}$ is a three-factor interaction effect for combination ijk, π_{hijk} is a random error effect associated with split-split plot experimental unit hijk, normally distributed with mean zero and variance σ^2_π, and β_3 is a linear regression coefficient of the split-split plot Y residuals on the corresponding Z residuals, $h = 1,\cdots,r$, $i = 1,\cdots,a$, $j = 1,\cdots,s$, and $k = 1,\cdots,p$. An ANCOVA for this design and response model is given in Table 2.

The various adjusted means are computed as:

$$\bar{Y}_{.i..} \text{ (adj.)} = \bar{Y}_{.i..} - \hat{\beta}_1(\bar{Z}_{.i..} - \bar{Z}_{....}) = \bar{Y}^*_{.i..} ,$$

$$\bar{Y}_{..j.} \text{ (adj.)} = \bar{Y}_{..j.} - \hat{\beta}_2(\bar{Z}_{..j.} - \bar{Z}_{....}) = \bar{Y}^*_{..j.} ,$$

$$\bar{Y}_{...k} \text{ (adj.)} = \bar{Y}_{...k} - \hat{\beta}_3(\bar{Z}_{...k} - \bar{Z}_{....}) = \bar{Y}^*_{...k} ,$$

$$\bar{Y}_{.ij.} \text{ (adj.)} = \bar{Y}_{.ij.} - \hat{\beta}_1(\bar{Z}_{.i..} - \bar{Z}_{....}) - \hat{\beta}_2(\bar{Z}_{.ij.} - \bar{Z}_{.i..}) = \bar{Y}^*_{.ij.} ,$$

$$\bar{Y}_{.i.k} \text{ (adj.)} = \bar{Y}_{.i.k} - \hat{\beta}_1(\bar{Z}_{.i..} - \bar{Z}_{....}) - \hat{\beta}_3(\bar{Z}_{.i.k} - \bar{Z}_{.i..}) = \bar{Y}^*_{.i.k} ,$$

$$\bar{Y}_{..jk} \text{ (adj.)} = \bar{Y}_{..jk} - \hat{\beta}_2(\bar{Z}_{..j.} - \bar{Z}_{....}) - \hat{\beta}_3(\bar{Z}_{..jk} - \bar{Z}_{..j.}) = \bar{Y}^*_{..jk} ,$$

and

$$\bar{Y}_{.ijk} \text{ (adj.)} = \bar{Y}_{.ijk} - \hat{\beta}_1(\bar{Z}_{.i..} - \bar{Z}_{....}) - \hat{\beta}_2(\bar{Z}_{.ij.} - \bar{Z}_{.i..})$$
$$- \hat{\beta}_3(Z_{.ijk} - \bar{Z}_{.ij.}) = \bar{Y}^*_{.ijk} ,$$

where

$$\hat{\beta}_1 = A_{yz}/A_{zz}, \quad \hat{\beta}_2 = B_{yz}/B_{zz}, \quad \text{and} \quad \hat{\beta}_3 = C_{yz}/C_{zz} .$$

Table 2. ANCOVA for Equation (2) for a Split-Split Plot Experiment Design when the Covariate is Obtained for Each Split-Split Plot Experimental Unit.[1]

Source of Variation	df	Sums of Products yy	yz	zz	df	Adjusted Sums of Squares
Total	rasp	T_{yy}	T_{yz}	T_{zz}		
Correction for Mean	1	M_{yy}	M_{yz}	M_{zz}		
Block	(r–1)	R_{yy}	R_{yz}	R_{zz}		
Whole Plot = W	(a–1)	W_{yy}	W_{yz}	W_{zz}		
Error (a)	(a–1)(r–1)	A_{yy}	A_{yz}	A_{zz}	(ar–r–a)	$A_{yy}-A^2_{yz}/A_{zz} = A'_{yy}$
Split Plot = S	(s–1)	S_{yy}	S_{yz}	S_{zz}		
S × W	(a–1)(s–1)	I_{yy}	I_{yz}	I_{zz}		
Error (b)	a(r–1)(s–1)	B_{yy}	B_{yz}	B_{zz}	a(r–1)(s–1)–1	$B_{yy}-B^2_{yz}/B_{zz} = B'_{yy}$
Split-Split Plot = P	(p–1)	P_{yy}	P_{yz}	P_{zz}		
W × P	(a–1)(p–1)	Q_{yy}	Q_{yz}	Q_{zz}		
S × P	(p–1)(s–1)	U_{yy}	U_{yz}	U_{zz}		
W × S × P	(a–1)(p–1)(s–1)	V_{yy}	V_{yz}	V_{zz}		
Error (c)	as(r–1)(c–1)	C_{yy}	C_{yz}	C_{zz}	as(r–1)(p–1)–1	$C_{yy}-C^2_{yz}/C_{zz} = C'_{yy}$

W(adj. for $\hat{\beta}_1$)	(a–1)	$W_{yy} - (W_{yz} + A_{yz})^2/(W_{zz} + A_{zz}) + A^2_{yz}/A_{zz} = W'_{yy}$
S(adj. for $\hat{\beta}_2$)	(s–1)	$S_{yy} - (S_{yz} + B_{yz})^2/(S_{zz} + B_{zz}) + B^2_{yz}/B_{zz} = S'_{yy}$
S × W(adj. for $\hat{\beta}_2$)	(a–1)(s–1)	$I_{yy} - (I_{yz} + B_{yz})^2/(I_{zz} + B_{zz}) + B^2_{yz}/B_{zz} = I'_{yy}$
P(adj. for $\hat{\beta}_3$)	(p–1)	$P_{yy} - (P_{yz} + C_{yz})^2/(P_{zz} + C_{zz}) + C^2_{yz}/C_{zz} = P'_{yy}$
W × P(adj. for $\hat{\beta}_3$)	(a–1)(p–1)	$Q_{yy} - (Q_{yz} + C_{yz})^2/(Q_{zz} + C_{zz}) + C^2_{yz}/C_{zz} = Q'_{yy}$
S × P(adj. for $\hat{\beta}_3$)	(p–1)(s–1)	$U_{yy} - (U_{yz} + C_{yz})^2/(U_{zz} + C_{zz}) + C^2_{yz}/C_{zz} = U'_{yy}$
W × S × P(adj. for $\hat{\beta}_3$)	(a–1)(p–1)(s–1)	$V_{yy} - (V_{yz} + C_{yz})^2/(V_{zz} + C_{zz}) + C^2_{yz}/C_{zz} = V'_{yy}$

[1] The various mean squares may be obtained by dividing sums of squares by the appropriate degrees of freedom.

Estimated variances of a difference between two means adjusted for a covariate for $i \neq i'$, $j \neq j'$, and $k \neq k'$, $E_a = A'_{yy}/(ar-r-a)$, $E_b = B'_{yy}/[a(r-1)(s-1)-1]$, and $E_c = C'_{yy}/[as(r-1)(p-1)-1]$ are given below:

<u>Variance of a difference between two whole plot treatment adjusted means</u>

$$\hat{V}(\bar{Y}^*_{.i..} - \bar{Y}^*_{.i'..}) = E_a\left[\frac{2}{rsp} + \frac{(\bar{Z}_{.i..} - \bar{Z}_{.i'..})^2}{A_{zz}}\right].$$

<u>Variance of a difference between two split plot treatment adjusted means</u>

$$\hat{V}(\bar{Y}^*_{..j.} - \bar{Y}^*_{..j'.}) = E_b\left[\frac{2}{arp} + \frac{(\bar{Z}_{..j.} - \bar{Z}_{..j'.})^2}{B_{zz}}\right].$$

<u>Variance of a difference between two split-split plot treatment adjusted means</u>

$$\hat{V}(\bar{Y}^*_{...k} - \bar{Y}^*_{...k'}) = E_c\left[\frac{2}{ars} + \frac{(\bar{Z}_{...k} - \bar{Z}_{...k'})^2}{C_{zz}}\right].$$

<u>Variance of a difference between two adjusted means for combinations ij and ij'</u>

$$\hat{V}(\bar{Y}^*_{.ij.} - \bar{Y}^*_{.ij'.}) = E_b\left[\frac{2}{rp} + \frac{(\bar{Z}_{.ij.} - \bar{Z}_{.ij'.})^2}{B_{zz}}\right].$$

<u>Variance of a difference between two adjusted means for combinations ij and i'j</u>

$$\hat{V}(\bar{Y}^*_{.ij.} - \bar{Y}^*_{.i'j.}) = \frac{2}{rp}\,[\hat{\sigma}^2_\delta + \hat{\sigma}^2_\epsilon] + E_a\,\frac{(\bar{Z}_{.i..} - \bar{Z}_{.i'..})^2}{A_{zz}}$$

$$+ E_b\,\frac{(\bar{Z}_{.ij.} - \bar{Z}_{.i..} - \bar{Z}_{.i'j.} + \bar{Z}_{.i'..})^2}{B_{zz}}.$$

<u>Variance of a difference between two adjusted means for combinations ik and ik'</u>

$$\hat{V}(\bar{Y}^*_{.i.k} - \bar{Y}^*_{.i.k'}) = E_c\left[\frac{2}{rs} + \frac{(\bar{Z}_{.i.k} - \bar{Z}_{.i.k'})^2}{C_{zz}}\right].$$

Variance of a difference between two adjusted means for combinations ik and i′k

$$\hat{V}(\bar{Y}^*_{\cdot i \cdot k} - \bar{Y}^*_{\cdot i' \cdot k}) = \frac{2(s\hat{\sigma}^2_\delta + \hat{\sigma}^2_\epsilon + \hat{\sigma}^2_\pi)}{rs} + E_a \frac{(\bar{Z}_{\cdot i \cdot \cdot} - \bar{Z}_{\cdot i' \cdot \cdot})^2}{A_{zz}}$$

$$+ \frac{E_c\,(\bar{Z}_{\cdot i \cdot k} - \bar{Z}_{\cdot i \cdot \cdot} - \bar{Z}_{\cdot i' \cdot k} + \bar{Z}_{\cdot i' \cdot \cdot})^2}{C_{zz}} .$$

Variance of a difference between two adjusted means for combinations ijk and ijk′

$$\hat{V}(\bar{Y}^*_{\cdot ijk} - \bar{Y}^*_{\cdot ijk'}) = E_c \left[\frac{2}{r} + \frac{(\bar{Z}_{\cdot ijk} - \bar{Z}_{\cdot ijk'})^2}{C_{zz}} \right] .$$

Variance of a difference between two adjusted means for combinations ijk and ij′k

$$\hat{V}(\bar{Y}^*_{\cdot ijk} - \bar{Y}^*_{\cdot ij'k}) = \frac{2}{r}(\hat{\sigma}^2_\epsilon + \hat{\sigma}^2_\pi) + E_b \frac{(\bar{Z}_{\cdot ij \cdot} - \bar{Z}_{\cdot ij' \cdot})^2}{B_{zz}}$$

$$+ E_c \frac{(\bar{Z}_{\cdot ijk} - \bar{Z}_{\cdot ij \cdot} - \bar{Z}_{\cdot ij'k} + \bar{Z}_{\cdot ij' \cdot})^2}{C_{zz}} .$$

Variance of a difference between two adjusted means for combination ijk and i′jk

$$\hat{V}(\bar{Y}^*_{\cdot ijk} - \bar{Y}^*_{\cdot i'jk}) = \frac{2}{r}(\hat{\sigma}^2_\delta + \hat{\sigma}^2_\epsilon + \hat{\sigma}^2_\pi) + E_a \frac{(\bar{Z}_{\cdot i \cdot \cdot} - \bar{Z}_{\cdot i' \cdot \cdot})^2}{A_{zz}}$$

$$+ E_b \frac{(\bar{Z}_{\cdot ij \cdot} - \bar{Z}_{\cdot i \cdot \cdot} - \bar{Z}_{\cdot i'j \cdot} + \bar{Z}_{\cdot i' \cdot \cdot})^2}{B_{zz}}$$

$$+ E_c \frac{(\bar{Z}_{\cdot ijk} - \bar{Z}_{\cdot ij \cdot} - \bar{Z}_{\cdot i'jk} + \bar{Z}_{\cdot i'j \cdot})^2}{C_{zz}} .$$

In the above $E_a = \hat{\sigma}^2_\pi + p\hat{\sigma}^2_\epsilon + ps\hat{\sigma}^2_\delta = A'_{yy}/(ar-a-r)$, $E_b = \hat{\sigma}^2_\pi + p\hat{\sigma}^2_\epsilon = B'_{yy}/[a(r-1)(s-1)-1]$, and $E_c = \hat{\sigma}^2_\pi = C'_{yy}/[as(r-1)(p-1)-1]$. Note that

$$\hat{V}(\bar{Y}^*_{\cdot ijk} - \bar{Y}^*_{\cdot i'jk}) = V(\bar{Y}^*_{\cdot ijk} - \bar{Y}^*_{\cdot i'j'k}) = V(\bar{Y}^*_{\cdot ijk} - \bar{Y}^*_{\cdot i'jk'}) = V(\bar{Y}^*_{\cdot ijk} - \bar{Y}^*_{\cdot i'j'k'})$$

and that

$$V(\bar{Y}^*_{\cdot ijk} - \bar{Y}^*_{\cdot ij'k}) = V(\bar{Y}^*_{\cdot ijk} - \bar{Y}^*_{\cdot ij'k'}) .$$

Most variances above, without the covariate, were given by Federer (1955). Estimates of variance components σ_δ^2, σ_ϵ^2, σ_π^2 are needed to compute the fifth, seventh, ninth, and tenth variances above. Note that $ps(\sigma_\pi^2 + \sigma_\epsilon^2 + \sigma_\delta^2) = s(p-1)\sigma_\pi^2 + (s-1)E_b + E_a$ and $p(\hat{\sigma}_\rho^2 + \sigma_\epsilon^2) = (p-1)\sigma_\pi^2 + E_b = (p-1)E_c + E_b$. The degrees of freedom for these variances require approximation as in the previous section.

4. SPLIT BLOCK EXPERIMENT DESIGN

The experiment design considered here is denoted as a split block design. It has also been called a two-way whole plot, criss-cross, and a strip trial design. This design has received scant attention in statistical textbooks with exceptions being Federer (1955) and Cochran and Cox (1957, Section 7.32). It does occur frequently in practice sometimes but is often not analyzed correctly. The member of this class of designs we shall discuss will be for a two-factor factorial treatment design with the levels of one factor being applied perpendicularly across all levels of the second factor within each replicate or complete block. The levels of each factor will have the same design for our example, that is a randomized complete block design. (The levels of one factor could be in a randomized complete block design and the levels of the second factor could be in a latin square, balanced incomplete block, or other experiment design.) Note that there will be r separate randomizations for the levels of *each* of the factors. The number of levels of factors one and two are a and b, respectively, resulting in an a × b factorial treatment design.

A response model equation as given in Federer (1955) for a variable Y and a covariate Z is:

$$Y_{hij} = \mu + \rho_h + \alpha_i + \delta_{hi} + \gamma_j + \pi_{hj} + \alpha\gamma_{ij} + \epsilon_{hij} + \beta_1(\bar{Z}_{hi.} - \bar{Z}_{...}) + \beta_2(\bar{Z}_{h.j} - \bar{Z}_{...}) + \beta_3(Z_{hij} - \bar{Z}_{hi.} - \bar{Z}_{h.j} + \bar{Z}_{...}) , \quad (3)$$

where μ is a general mean effect, ρ_h is the h*th* block effect, which has mean zero and variance σ_ρ^2, α_i is the effect of the i*th* level of factor one, say A, γ_j is the effect of the j*th* level of factor two, say B, δ_{hi} is a random error effect for the hi*th* whole plot for factor A and has mean zero and variance σ_δ^2, π_{hj} is a random error effect for the hj*th* whole plot for factor B and has mean zero and variance σ_π^2, $\alpha\gamma_{ij}$ is the interaction effect for the ij*th* combination of levels of factors A and B, ϵ_{hij} is a random

Table 3. ANCOVA for Equation (3) for a Split Block Experiment Design When the Covariate is Obtained for Each AB Combination Within a Block

Source of Variation	df	Sums of Products	df	Adjusted Sums of Squares
Total	rab	T_{yy} T_{yz} T_{zz}		
Correction for mean	1	M_{yy} M_{yz} M_{zz}		
Replicate = R	(r–1)	R_{yy} R_{yz} R_{zz}		
Whole Plot A	(a–1)	W_{yy} W_{yz} W_{zz}		
Error (a)	(r–1)(a–1)	A_{yy} A_{yz} A_{zz}	(ra–a–r)	$A_{yy} - \frac{A_{yz}^2}{A_{zz}} = A'_{yy}$
Whole A adjusted for $\hat{\beta}_1 = A_{yz}/A_{zz}$			(a–1)	$W_{yy} - \frac{(W_{yz}+A_{yz})^2}{W_{zz}+A_{zz}} + \frac{A_{yz}^2}{A_{zz}} = W'_{yy}$
Whole plot B	(b–1)	S_{yy} S_{yz} S_{zz}		
Error (b)	(b–1)(r–1)	B_{yy} B_{yz} B_{zz}	(rb–b–r)	$B_{yy} - \frac{B_{yz}^2}{B_{zz}} = B'_{yy}$
Whole plot B adjusted for $\hat{\beta}_2 = B_{yz}/B_{zz}$			(b–1)	$S_{yy} - \frac{(S_{yz}+B_{yz})^2}{S_{zz}+B_{zz}} + \frac{B_{yz}^2}{B_{zz}} = S'_{yy}$
A × B	(a–1)(b–1)	I_{yy} I_{yz} I_{zz}		
Error (ab)	(r–1)(a–1)(b–1)	C_{yy} C_{yz} C_{zz}	(r–1)(a–1)(b–1)–1	$C_{yy} - \frac{C_{yz}^2}{B_{zz}} = C'_{yy}$
Interaction adjusted for $\hat{\beta}_3 = C_{yz}/C_{zz}$			(a–1)(b–1)	$I_{yy} - \frac{(I_{yz}+C_{yz})^2}{I_{zz}+C_{zz}} + \frac{C_{yz}^2}{C_{zz}} = I'_{yy}$

(The various mean squares may be obtained by dividing the sums of squares by their respective degrees of freedom.)

error effect associated with the hij*th* subplot for the A × B interaction and has mean zero and variance σ^2_ϵ, β_1 is the linear regression of Y whole plot residuals on the Z whole plot residuals for factor A, β_2 is the linear regression of the Y whole plot residuals on the Z whole plot residuals for factor B, and β_3 is the linear regression of Y subplot residuals on Z subplot residuals.

An ANCOVA for response model (3) is given in Table 3. For this design and for fixed effects for the a × b factorial, there are three error variances and three error regressions. Given error effects that are NID, the usual F statistics may be used. The adjusted means are given by:

$$\bar{Y}_{.i.}(\text{adjusted}) = \bar{Y}_{.i.} - \hat{\beta}_1(\bar{Z}_{.i.} - \bar{Z}_{...}) = \bar{Y}^*_{.i.}\ ,$$

$$\bar{Y}_{..j}(\text{adjusted}) = \bar{Y}_{..j} - \hat{\beta}_2(\bar{Z}_{..j} - \bar{Z}_{...}) = \bar{Y}^*_{..j}\ ,$$

and

$$\bar{Y}_{.ij}(\text{adjusted}) = \bar{Y}_{.ij} - \hat{\beta}_1(\bar{Z}_{.i.} - \bar{Z}_{...}) - \hat{\beta}_2(\bar{Z}_{..j} - \bar{Z}_{...}) - \hat{\beta}_3(\bar{Z}_{.ij} - \bar{Z}_{.i.} - \bar{Z}_{..j} + \bar{Z}_{...}) = \bar{Y}^*_{.ij}\ ,$$

where the $\hat{\beta}$s are defined in Table 3.

Estimated variances of a difference between adjusted means are given below for $i \neq i'$, $j \neq j'$:

$$\hat{V}(\bar{Y}^*_{.i.} - \bar{Y}^*_{.i'.}) = E_a\left[\frac{2}{rb} + \frac{(\bar{Z}_{.i.} - \bar{Z}_{.i'.})^2}{A_{zz}}\right].$$

$$\hat{V}(\bar{Y}^*_{..j} - \bar{Y}^*_{..j'}) = E_b\left[\frac{2}{ra} + \frac{(\bar{Z}_{..j} - \bar{Z}_{..j'})^2}{B_{zz}}\right].$$

$$\hat{V}(\bar{Y}^*_{.ij} - \bar{Y}^*_{.ij'}) = \frac{1}{r}(\hat{\sigma}^2_\pi + \hat{\sigma}^2_\epsilon) + \frac{E_b}{B_{zz}}(\bar{Z}_{..j} - \bar{Z}_{..j'})^2 + \frac{E_c}{C_{zz}}(\bar{Z}_{.ij} - \bar{Z}_{.ij'} - \bar{Z}_{..j} + \bar{Z}_{..j'})\ ,$$

$$\hat{V}(\bar{Y}^*_{.ij} - \bar{Y}^*_{.i'j}) = \frac{1}{r}(\hat{\sigma}^2_\delta + \hat{\sigma}^2_\epsilon) + \frac{E_a}{A_{zz}}(\bar{Z}_{.i.} - \bar{Z}_{.i'.})^2 + \frac{E_c}{C_{zz}}(\bar{Z}_{.ij} - \bar{Z}_{.i'j} - \bar{Z}_{.i.} + \bar{Z}_{.i'.})\ ,$$

and

$$\hat{V}(\bar{Y}^*_{.ij} - \bar{Y}^*_{.i'j'}) = \frac{1}{r}(\hat{\sigma}^2_\delta + \hat{\sigma}^2_\pi + \hat{\sigma}^2_\epsilon) + \frac{E_a}{A_{zz}}(\bar{Z}_{.i.} - \bar{Z}_{.i'.})^2 + \frac{E_b}{B_{zz}}(\bar{Z}_{..j} - \bar{Z}_{..j'})^2\ ,$$

$$+ \frac{E_c}{C_{zz}}(\bar{Z}_{.ij} - \bar{Z}_{.i'j'} - \bar{Z}_{.i.} + \bar{Z}_{.i'.} - \bar{Z}_{..j} + \bar{Z}_{..j'})^2\ ,$$

where

$$E_a = A'_{yy}/(ar-a-r) = \hat{\sigma}^2_\epsilon + b\hat{\sigma}^2_\delta,\ E_b = B'_{yy}/(br-b-r) = \hat{\sigma}^2_\epsilon + a\hat{\sigma}^2_\pi\ ,$$

and

$$E_c = C'_{yy}/[(a-1)(b-1)(r-1)-1] = \hat{\sigma}^2_\epsilon\ .$$

The degrees of freedom for the last three variances need to be approximated by the method previously given or by another appropriate approximation (e.g., Grimes and Federer, 1984).

5. SOME COMMENTS

Since formulae for many of the above adjusted means and variances do not appear to be in statistical literature, it was deemed appropriate to include them here. As can be seen from the analyses for relatively simple designs from each of the three classes, there are a variety of formulas for adjusted means and variances of differences between two adjusted means. The more complex members of each class may have many more error mean squares and a similar number of regression coefficients. Experiments are conducted wherein some of the factors are arranged in split blocks and others in split plot arrangements. Many different designs may be used for the different factors (see, e.g., Federer, 1955, 1975). The complex experiment design described by Federer and Farden (1955) had several split plot and several split block arrangements with a total of 75 error mean squares and 203 lines in the ANOVA!

One method of aiding investigators with ANOVAs and ANCOVAs of complexly designed experiments is to ascertain how much of a statistical analysis can be obtained with computer packages such as SAS, BMDP, GENSTAT, SPSS, and others. Then, the output can be annotated, i.e., an explanation is appended to the computer output describing what has been computed and how to use the results. Annotated computer outputs for two different split plot designs with a covariate have been completed for SAS, BMDP, and GENSTAT (see Federer *et al.*, 1987a, b, c). These reports may be purchased from the Mathematical Sciences Institute, Cornell University. Some of the detail is presented in Miles-McDermott *et al.* (1988) and Meredith *et al.* (1988). Additional annotated computer output ANCOVA reports of Searle *et al.* (1982a, b, c) may be purchased from the Biometrics Unit, Cornell University, Ithaca, NY 14853.

The analyses have been described for a single covariate. Noting that $A_{yy} - A_{yy}^2/A_{zz} = A_{yy}(1-r_{yz}^2) = A'_{yy}$, one may simply use $A_{yy}(1-R^2) = A'_{yy}$ when there are several covariates and where R^2 is the squared multiple correlation coefficient computed on the error line. If the relationship between a covariate Z and Y is curvilinear, it may be possible to use some function of Z, e.g., log Z, $\sqrt{Z}$, 1/Z, which makes the relation linear. If this can be accomplished both computations and interpretations are simplified. Also, it should be clear from the results presented herein how to formulate and analyze a split plot experiment with one covariate measured on whole plots only and another measured on split plots only. The question of homogeneity of slopes for either the whole or split plot factors has not been discussed. However, appropriate model extensions may be readily included along with the appropriate computations.

A simplification of the estimated variances for differences of means has been given by Yates (1934) and Finney (1946). Instead of computing the quantities $(\bar{Z}_{.i.}-\bar{Z}_{.i'.})^2/A_{zz}$ and $(\bar{Z}_{..j}-\bar{Z}_{..j'})^2/B_{zz}$, e.g., for each pair of means, one may compute a single variance by using $W_{zz}/(a-1)A_{zz}$ or $S_{zz}/(s-1)B_{zz}$, respectively. The quantity $W_{zz}/(a-1)$ is an average of all pairs ii′ of $(\bar{Z}_{.i.}-\bar{Z}_{.i'.})^2$. This simplification and approximation considerably reduces the number of computations for large a and/or s. For the quantities $(\bar{Z}_{.ij}-\bar{Z}_{.i'j}-\bar{Z}_{.i.}+\bar{Z}_{.i'.})^2$ and $(\bar{Z}_{.ij}-\bar{Z}_{.ij'}-\bar{Z}_{..j}+\bar{Z}_{..j'})^2$ it is suggested that $I_{zz}/(a-1)(s-1)B_{zz}$ be used if it is desired to compute only a single variance. This procedure may be useful when the individual squared terms are not too disparate.

REFERENCES

Cochran, W. G., and Cox, G. M. (1957), *Experimental Designs*, 2nd ed., New York: John Wiley.

Federer, W. T. (1955), *Experimental Designs – Theory and Application*, New York: Macmillan (republished by Oxford and IBH Publishing Co., New Delhi, India, 1964, 1974).

Federer, W. T. (1975), "The Misunderstood Split Plot," in *Applied Statistics*, ed. R. P. Gupta, Amsterdam: North-Holland, pp. 1-39.

Federer, W. T. (1977), "Sampling, Blocking, and Model Considerations for Split Plot and Split Block Designs." *Biometrical Journal*, 19, 181-200.

Federer, W. T., and Farden, C. A. (1955), "Analysis of Variance Set-Up for Joint Project 69," BU-67-M in the Technical Report Series of the Biometrics Unit, Cornell University.

Federer, W. T., Feng, Z. D., and Miles-McDermott, N. J. (1987a), "Annotated Computer Output for Split Plot Design: GENSTAT," Annotated Computer Output '87-4, Mathematical Sciences Institute, Cornell University, 201 Caldwell Hall, Ithaca, NY 14853.

Federer, W. T., Feng, Z. D., and Miles-McDermott, N. J. (1987b), "Annotated Computer Output for Split Plot Design: BMDP 2V," Annotated Computer Output '87-5, Mathematical Sciences Institute, 201 Caldwell Hall, Ithaca, NY 14853.

Federer, W. T., Feng, Z. D., Meredith, M. P., and Miles-McDermott, N. J. (1987c), "Annotated Computer Output for Split Plot Design: SAS GLM," Annotated Computer Output '87-8, Mathematical Sciences Institute, Cornell University, 201 Caldwell Hall, Ithaca, NY 14853.

Federer, W. T., and Henderson, H. V. (1979), "Covariance Analysis of Designed Experiments × Statistical Packages: An Update," in *Proceedings, Computer Science and Statistics, 12th Annual Symposium on the Interface*, pp. 228-235.

Finney, D. J. (1946), "Standard Errors of Yields Adjusted for Regression on an Independent Measurement," *Biometrics*, 2, 53-55.

Gaylor, D. W., and Hopper, F. N. (1969), "Estimating the Degrees of Freedom for Linear Combinations of Mean Squares by Satterthwaite's Formula," *Technometrics*, 11, 691-706.

Grimes, B. A., and Federer, W. T. (1984), "Comparison of Means from Populations with Unequal Variances," in *W. G. Cochran's Impact on Statistics*, eds. Rao, P.S.R.S., and J. Sedransk, New York: John Wiley, pp. 353-374.

Meredith, M. P., Miles-McDermott, N. J., and Federer, W. T. (1988), "The Analysis of Covariance for Split-Unit and Repeated Measures Experiments Using the GLM Procedure, Part II," *Proceedings, SAS Users Group International 13th Conference*, pp. 1059-1064.

Miles-McDermott, N. J., Federer, W. T., and Meredith, M. P. (1988), "The Analysis of Covariance for Split-Unit and Repeated Measures Experiments, Part I," *Proceedings, SAS Users Group International 13th Conference*, pp. 1053-1058.

Satterthwaite, F. E. (1946), "An Approximate Distribution of Estimates of Variance Components," *Biometrics Bulletin*, 2, 110-114.

Searle, S. R., Hudson, G. F. S., and Federer, W. T. (1982a), "Annotated Computer Output for Covariance (ACO: COV-GENSTAT)," BU-782-M in the Technical Report Series of the Biometrics Unit, Cornell University, Ithaca, NY.

Searle, S. R., Hudson, G. F. S., and Federer, W. T. (1982b), "Annotated Computer Output for Covariance (ACO: COV-SAS)," BU-783-M in the Technical Report Series of the Biometrics Unit, Cornell University, Ithaca, NY.

Searle, S. R., Hudson, G. F. S., and Federer, W. T. (1982c), "Annotated Computer Output for Covariance (ACO: COV-SPSS)," BU-784-M in the Technical Report Series of the Biometrics Unit, Cornell University, Ithaca, NY.

Yates, F. (1934), "A Complex Pig-Feeding Experiment," *Journal of Agricultural Science*, 24, 511-534.

Yates, F. (1937), "The Design and Analysis of Factorial Experiments," *Imperial Bureau of Soil Science Technical Communications*, 35, 1-95.