ESTIMATION OF THE POPULATION COVARIANCE COEFFICIENT FOR SPLIT-PLOT EXPERIMENTS[1]

JOSÉ RUY PORTO DE CARVALHO[2] and ROGER MEAD[3]

ABSTRACT - In this paper the full Maximum Likelihood Estimator is developed for the true covariance coefficient β, to allow covariance adjustments in split-plot experiments when the main and split-plot residual regression coefficients may be assumed to be equal. Intuitively, pooled estimators should produce the most efficient analysis (as compared with the split-plot regression coefficient, which is frequently used to adjust main and split-plot treatment means). The comparison of the MLE against the Cochran and the split-plot estimators has been investigated. The general conclusion is that, from the practical point of view, the full MLE will perform better than the Cochran's and the split-plot estimators. The Likelihood Ratio Test of the hypothesis that the main-plot and split-plot covariance coefficients are equal, together with the relationship between the observed and asymptotic powers is investigated.

Index terms: maximum likelihood estimation, likelihood ratio test, covariance coefficients, split-plot analysis, covariance analysis, bias, mean squared error.

ESTIMAÇÃO DO COEFICIENTE DE COVARIÂNCIA POPULACIONAL PARA EXPERIMENTOS EM PARCELAS-SUBDIVIDIDAS

RESUMO - Neste trabalho, o estimador de Máxima Verossimilhança é desenvolvido para o verdadeiro coeficiente de covariância β, permitindo ajustamentos pela covariável em experimentos de parcelas subdivididas quando os coeficientes de regressão residual para as parcelas principais e subparcelas são consideradas iguais. Intuitivamente, estimadores conjuntos devem produzir uma análise mais eficiente (quando comparada com o coeficiente de regressão das parcelas subdivididas, o qual é freqüentemente usado para ajustar as médias de tratamentos nas parcelas e subparcelas). A comparação do estimador de Máxima Verossimilhança com os estimadores de Cochran e o da subparcela foi investigada. Como conclusão geral do ponto de vista prático, o estimador de Máxima Verossimilhança apresentou melhor desempenho do que os outros dois estimadores. É estudado o Teste da Razão de Probabilidades para hipóteses de que os coeficientes de covariância para parcelas e subparcelas são iguais, bem como a relação entre poderes assintóticos e observados.

Termos para indexação: estimação de máxima-verossimilhança, teste da razão de máxima-verossimilhança, coeficiente de covariância, parcelas-subdivididas, análise de covariância, bias, quadrado médio do erro.

INTRODUCTION

The Analysis of Covariance introduced by Fisher (1946) is an important statistical method, enabling us to reduce the experimental error by eliminating certain environmental effects not controlled by experimental design. It is usually assumed that these environmental effects or concomitant variables are not related to treatments, but whether or not this is true, it is possible to adjust for them by using the Analysis of Covariance, which should be expected to

[1] Accepted for publication on October 29, 1991.
Extracted from a Dissertation submitted by the senior author in partial fulfillment of the requirements for the Ph.D. degree to the Department of Applied Statistics. The University of Reading. Reading, Berkshire, U.K.

[2] Statistician, Ph.D., EMBRAPA/NMA - P.O. Box 491 - 13001 Campinas, SP, Brazil, Fone: (0192) 52.5977.

[3] Applied Statistics Professor, Department of Applied Statistics - The University of Reading, P.O. Box 217 - Reading, RG6-2AN, U.K.

04

result in an increase of accuracy for treatment comparisons.

The adjustment of the dependent variable for the effect of the independent (or concomitant) variable in the Analysis of Covariance can be regarded as a technique which combines the features of Analysis of Variance and Regression.

For split-plot designs, in where there are two residual terms in the Analysis of Variance, covariance coefficients may be determined for both main and split-plots. If it is assumed that the two coefficients are unequal, treatment means are adjusted using the main-plot coefficient to adjust the main-plot treatment means, and using the split-plot coefficient to adjust the split-plot treatment means. These separate adjustments have been suggested by a large number of authors, such as Federer (1955) and Kempthorne (1975). When the covariance coefficients are approximately equal, it has been suggested, for example, by Bartlett (1937), Cochran (1957) and John & Quenouille (1977) that the split-plot covariance coefficient be used for all adjustments. The choice of the split-plot coefficient is governed by the large variances for main-plot comparisons.

If the regression coefficients β_1 and β_2 are assumed to be identical at main-plot and split-plot levels, each part of the analysis provides only part of the information on the population regression coefficient β. Therefore, greater accuracy should be possible from combining both parts in order to obtain a single estimator of the coefficient, instead of using the split-plot regression coefficient throughout the analysis as is commonly recommended. The use of a single pooled estimator of the regression coefficient has been considered by Cochran (1946), Truitt & Smith (1956) and, more recently, by Dear & Mead (1984).

In the literature there are two estimators of for simulating adjustment of main-plot and split-plot comparisons.. These two estimators are compared in this study. In addition, a third estimator, the full Maximum Likelihood Estimator (MLE) of the population regression coefficient is developed and its performance compared against these two estimators. The question of when to use the single estimator is also considered.

Background

The first estimator, the split-plot covariance coefficient, was originally proposed by Bartlett (1937). The residual covariance coefficient of the dependent variable Y on the concomitant variable X, at split-plot level, is defined as:

$$\hat{\beta}_2 = \frac{E_{YX}}{E_{XX}}, \qquad (1)$$

where E_{YX} is the sum of products of the two variables after removing the effects due to main-plots and treatments; E_{XX} is the corresponding sum of squares of the concomitant variable. Bartlett (1937) argued that, although the adjusted error variance for the main-plot analysis will be less than the value obtained by a direct main-plot covariance adjustment, the use of split-plot coefficient can be quite efficient.

Searching for a more efficient method of adjustment for the main-plot comparisons, Cochran (1946) provided the second estimator used in our study. Although Cochran described this estimator as a Maximum-Likelihood estimator of the true covariance coefficient β, it is simply a linear combination of the main-plot and the split-plot sample covariance coefficients. Cochran's estimator β_c is a weighted mean of the two independent sampling coefficients, the weights being the reciprocals of the variances of the respective coefficients. It is defined as follows:

$$\hat{\beta}_c = \frac{\dfrac{b_1 D_{XX}}{\sigma^2_{12}} + \dfrac{b_2 E_{XX}}{\sigma^2_{22}}}{\dfrac{D_{XX}}{\sigma^2_{12}} + \dfrac{E_{XX}}{\sigma^2_{22}}} \qquad (2)$$

where b_1 and b_2 are the main-plot and split-plot sample residual covariance coefficients; D_{XX} is the residual main-plot sum of squares for the covariate; σ^2_{12} and σ^2_{22} are the theoretical error variances for main-plot and split-plot, respectively. Cochran (1946) commented that derivation of the full MLE of the variances σ^2_{12} and σ^2_{22}, which depends on $\hat{\beta}$, was not simple. Truitt & Smith (1956) suggested using the main-plot and split-plot sampling variances of the observations, after eliminating variation due to the covariate, that is

$$s^2_{12} = \frac{D_{YY} - b_1 D_{XY}}{df_1 - 1} \quad \text{and}$$

$$s^2_{22} = \frac{E_{YY} - b_2 E_{XY}}{df_2 - 1} \qquad (3)$$

as the estimators of the theoretical variances. Here, D_{YY} and E_{YY} are the main-plot and split-plot error sum of squares for the dependent variable with df_1 and df_2 degrees of freedom, and D_{XY}, E_{XY} are the respective sum of products. Their suggestion was based on the assumption that almost all information about the true variances is contained in s^2_{12} and s^2_{22}. They again pointed out the difficulty in determining the full MLE of the true variances, since they are functions of the MLE of β.

The next estimator derived in this paper is the MLE when the covariance coefficients are the same, that is when the null hypothesis H_0 : $\beta_1 = \beta_2 = \beta$ holds. If two samples are drawn from Bivariate Normally distributed populations, the sample main-plot and split-plot variance-covariance matrices are independent and each follows a Wishart distribution. The Likelihood Function (LF) for the combined information is the product of the two separate Likelihood Functions based on the two variance-covariance matrices, conditioning on the estimated block and treatment parameters:

$$LF(\sigma_{11}, \sigma_{12}, \sigma_{21}, \sigma_{22}, \beta) =$$

$$\frac{(df_1+1)^{df_1} s_{11}^{df_1} s_{12} (s^2_{12} - b^2_1 s^2_{11})^{(df_1-3)/2}}{\pi \, \Gamma(df_1-1) \sigma_{11}^{df_1} (\sigma^2_{12} - \beta^2 \sigma^2_{11})^{df_1/2}} \times$$

$$\frac{(df_2+1)^{df_2} s_{21}^{df_2} s_{22} (s^2_{22} - b^2_2 s^2_{21})^{(df_2-3)/2}}{\pi \, \Gamma(df_2-1) \sigma_{21}^{df_2} (\sigma^2_{22} - \beta^2 \sigma^2_{21})^{df_2/2}} \times$$

$$\exp\left[- \frac{(df_1+1)}{2(\sigma^2_{12} - \beta^2 \sigma^2_{11})} \left[\frac{\sigma^2_{12} s^2_{11}}{\sigma^2_{11}} - 2\beta b_1 s^2_{11} + s^2_{12} \right] \times \right.$$

$$\left. - \frac{(df_2+1)}{2(\sigma^2_{22} - \beta^2 \sigma^2_{21})} \left[\frac{\sigma^2_{22} s^2_{21}}{\sigma^2_{21}} - 2\beta b_2 s^2_{21} + s^2_{22} \right] \right] \qquad (4)$$

where s^2_{11}, s^2_{12}, and s^2_{21}, s^2_{22}, are the main-plot and split-plot sample residual variances, respectively of the dependent and concomitant variables; σ^2_{11}, σ^2_{12}, and σ^2_{21}, σ^2_{22}, are the corresponding population variances; b_1 and b_2 are the separate main-plot and split-plot estimators of the covariance coefficient. The MLE equations of the five parameters of the above LF are obtained after the derivatives of ln(LF) are equated to zero. The development of these five equations, reduces to the cubic equation in $\hat{\beta}$:

$$\hat{\beta}^3 - \frac{[df_1(b_1+2b_2) + df_2(b_2+2b_1)]}{df_1+df_2} \hat{\beta}^2 +$$

$$\frac{[df_1 s^2_{11}(2s^2_{21} b_1 b_2 + s^2_{22}) + df_2 s^2_{21}(2s^2_{11} b_1 b_2 + s^2_{12})]}{s^2_{11} s^2_{21}(df_1+df_2)} \hat{\beta}$$

$$\frac{[df_1 s^2_{11} s^2_{22} b_1 + df_2 s^2_{12} s^2_{21} b_2]}{s^2_{11} s^2_{21}(df_1+df_2)} = 0 \qquad (5)$$

The reader is referred to Carvalho (1988) for further details about the development of $\hat{\beta}$. The other four ML estimators are obtained by substituting $\hat{\beta}$ back in the normal equations, as follows:

$$\hat{\sigma}^2_{11} = \frac{df_1+1}{df_1} s^2_{11}, \quad \hat{\sigma}^2_{12} = \frac{df_1+1}{df_1} [2\hat{\beta} s^2_{11} (\hat{\beta}-b_1) + s^2_{12}]$$

$$\hat{\sigma}^2_{21} = \frac{df_2+1}{df_2} s^2_{21}, \quad \hat{\sigma}^2_{22} = \frac{df_2+1}{df_2} [2\hat{\beta} s^2_{21} (\hat{\beta}-b_2) + s^2_{22}] \qquad (6)$$

The Likelihood Ratio Test

A test procedure based on the standard Likelihood Ratio statistic in order to test the null hypothesis of equality of main-plot and split-plot covariance coefficients against the general alternative that they are different, may be defined as follows:

$$\lambda = \frac{LF(\hat{\sigma}_{11}, \hat{\sigma}_{12}, \hat{\sigma}_{21}, \hat{\sigma}_{22}, \hat{\beta})}{LF(\hat{\sigma}_{11}, \hat{\sigma}_{12}, \hat{\beta}_1, \hat{\sigma}_{21}, \hat{\sigma}_{22}, \hat{\beta}_2)} \qquad (7)$$

where the numerator is the LF maximized under $H_0 : \beta_1 = \beta_2 = \beta$ and the denominator is the LF maximized under $H_1 : \beta_1 \neq \beta_2$.

Carvalho (1988) considered two different situations as follows:

i) The first situation is when both residual matrices are associated with different df, that is $df_1 \neq df_2$, leading to the following expression for the Likelihood Ratio Test (LRT):

$$\lambda_{21} = \left[\frac{s^2_{12} - b^2_1 s^2_{11}}{s^2_{12} + \hat{\beta} s^2_{11} (\hat{\beta} - 2b_1)} \right]^{\frac{df_1}{2}} \left[\frac{s^2_{22} - b^2_2 s^2_{21}}{s^2_{22} + \hat{\beta} s^2_{21} (\hat{\beta} - 2b_2)} \right]^{\frac{df_2}{2}}, \qquad (8)$$

where $\hat{\beta}$ is defined by evaluating equation (5).

ii) The second situation is when the residual matp2O ave the same df, that is $df_1 = df_2 = df$, resulting in:

$$\lambda_{22} = \left[\left[\frac{s^2_{12} - b^2_1 s^2_{11}}{s^2_{12} + \hat{\beta} s^2_{11} (\hat{\beta} - 2b_1)} \right] \left[\frac{s^2_{22} - b^2_2 s^2_{21}}{s^2_{22} + \hat{\beta} s^2_{21} (\hat{\beta} - 2b_2)} \right] \right]^{\frac{df}{2}} \qquad (9)$$

where $\hat{\beta}$ is obtained from the simplified version of (5) in which $df_1 = df_2$.

Since the distributions of (8) or (9) are far from simple, it would be convenient to use asymptotic approximations. The general result originally established by Wilks (1938), is that $-2\ln(\lambda)$ is asymptotically distributed as χ^2_c, where c is the additional number of parameters requested for the hypothesis H_0. Under the alternative hypothesis, Wald (1943) derived the asymptotic distribution of $-2\ln(\lambda)$ (when the regularity conditions of asymptotic normality and efficiency of the ML estimators are satisfied), as being a non-central Chi-Square with c df and noncentrality parameter δ.

The noncentrality parameter is obtained by evaluating the Fisher Information matrix I^{-1}, giving the following expression for δ.

$$\delta = [(\hat\beta_1-\hat\beta) \quad (\hat\beta_2-\hat\beta)] \begin{bmatrix} \dfrac{df_1\sigma^2_{11}(\sigma^2_{12}+\beta^2_1\sigma^2_{11})}{(\sigma^2_{12}-\beta^2_1\sigma^2_{11})^2} & 0 \\ 0 & \dfrac{df_2\sigma^2_{21}(\sigma^2_{22}+\beta^2_2\sigma^2_{21})}{(\sigma^2_{22}-\beta^2_2\sigma^2_{21})^2} \end{bmatrix} \cdot \begin{bmatrix} \hat\beta_1-\hat\beta \\ \hat\beta_2-\hat\beta \end{bmatrix}$$

$$= \frac{df_1\sigma^2_{11}(\sigma^2_{12}+\beta^2_1\sigma^2_{11})(\beta_1-\beta)^2}{(\sigma^2_{12}-\beta^2_1\sigma^2_{11})^2} + \frac{df_2\sigma^2_{21}(\sigma^2_{22}+\beta^2_2\sigma^2_{21})(\beta_2-\beta)^2}{(\sigma^2_{22}-\beta^2_2\sigma^2_{21})^2} \quad (10)$$

Based on the fact that the ML estimators converge in probability to the true parameter for large sample sizes, the asymptotic power P may be obtained by making use of the Patnaik approximation (Patnaik 1949) for the non-central X^2 distribution in terms of the central X^2 distribution.

The reliability of the asymptotic power was checked against the observed power obtained from a simulation study with a range of combinations of input parameters. Three different population variance and covariance structures (CS), were chosen. For the first CS_1, the ratios between the main-plot and split-plot variances were 3 : 1 for the concomitant variable, and 2 : 1 for the dependent variable. For CS_2, the ratio was 2 : 1 and 3 : 1 and, for CS_3, the ratios were 2 : 1 and 2 : 1, respectively. For each CS, the population covariance coefficient was fixed to vary from -0.5 to 0.0 by 0.1, and the same three sets of df, $df_1 = df_2 = 3$, $df_1 = 3$, $df_2 = 16$ and $df_1 = 12$, $df_2 = 75$ were used.

The first set, $df_1 = 3$ and $df_2 = 3$, corresponds to a small experiment where there are no blocks, three levels of the main-plot factor and two levels of the split-plot factor. The second set, where $df_1 = 3$ and $df_2 = 16$, corresponds to a routine experiment with two blocks, four levels of main-plot factor and five levels of split-plot factor. Finally, the third set, where $df_1 = 12$ and $df_2 = 75$, corresponds to a large experiment with four blocks, five main-plot levels and six split-plot levels.

The asymptotic approximation was wholly satisfactory only in the regions near the null hypothesis, or near the extreme alternative hypothesis. Otherwise, the power derived from the asymptotic theory was invariably found to be an overestimate, even for the large experiment.

The relationship between asymptotic and observed powers was investigated and it was found that after a logarithmic transformation of both powers the relationship was more reasonably fitted by the straight line regression model, giving the fitted regression:

$$\hat{y} = -0.3400 + 0.9128\,x \quad (11)$$

which predicts that the observed log power (y) increases by 0.9128 for the increase in the corresponding of asymptotic log power (x).

Simulation study

If A is any p x p matrix of rank p, and X is a p dimensional vector with a Multivariate Normal distribution X - MN (O, I_p) then, Y = AX has a Multivariate Normal distribution Y - MN(O, Σ), with zero mean and variance-covariance matrix $\Sigma = A A'$; A is lower triangular and it may be obtained from the Cholesky's decomposition of the population matrix Σ. To obtain two independent sample variance-covariance matrices, representing the main-plot and split-plot variation with any specified correlation structures Σ_1 and Σ_2, we generate independent sample of identifically distributed N(0,1) and calculate the corresponding observation matrix.

For each estimator ($\hat{\beta}_1$, $\hat{\beta}_C$ and $\hat{\beta}$) for each of the 54 experimental situations defined in section 3, a simulated analysis was repeated a thousand times (each time with different starting seed) and the observed mean and variance for each estimator were calculated. In addition, the same sampling statistics were observed for the estimators of the theoretical variances, that is $V(\hat{\beta})_2)$, $V(\hat{\beta}_C)$ and $V(\hat{\beta})$.

If the probability of obtaining successfully a LRT value greater than 3.84 (the critical value of the Chi-Square distribution with one degree of freedom at 5%) is fixed to be 50%, we expect from this size of simulation study, that is one thousand observations for each set of input parameters, to have the standard error for the observed power smaller than 2%. From this we may conclude that the size of the simulation study which was chosen was sufficient to consider the sampling coefficients as good estimators of the true parameter.

Comparison of estimators

Two criteria were used to assess the alternative estimators of the common covariance coefficient β. They are the Bias and the Mean-Squared Error (MSE). The principal condition for comparison of the different estimators using the MSE criterion, is that their distributions must have almost the same shape. As Cox & Hinkley (1974) stated, the MSE statistic may not be a valuable criterion for comparison if the distributions are of different shapes, particularly if some estimator has infinite variance.

Before comparing the results for the different estimators we check that the empirical distributions are really comparable. Each distribution is standardized to have zero mean and unit standard deviation. Then, the agreement among the empirical distributions can be tested through a X^2 statistic. For each of the 54 experimental situations, the one thousand standardized observations were grouped into fifteen classes, with equal width intervals, between the maximum and minimum values for the distribution of $\hat{\beta}$. The same classes were used for grouping the one thousand observations for $\hat{\beta}_C$ and $\hat{\beta}_2$ for the correspondent experimental situation.

To check the normality assumption for the regression coefficients a goodness of fit test was used. The X^2 goodness of fit values for the 54 experimental situations are shown in Table 1 and may be compared with the X^2 distribution on 12 df (5% level 21.03).

The results show that for the small experiment (3,3) the three estimators are not normally distributed in any experimental situation, although, few X^2 values for $\hat{\beta}$ or $\hat{\beta}_C$ are significant at 5%. For the other two combinations of degrees of freedom, the distributions of the three estimators may be accepted as Normal. The estimators $\hat{\beta}_2$, $\hat{\beta}_C$ and $\hat{\beta}$ of the true covariance coefficients β, are compared in Table 2. The Bias and the MSE criteria were evaluated for each experimental situation, except that MSE values given for the (3,3) df case.

Table 2 also presents the values of the test statistic

$$Z = \frac{\text{Bias}}{s/(1000)^{\frac{1}{2}}},$$

used to test if the observed Bias can be considered negligible. The Normal distribution used for (3,3) experimental situations should be treated with suspicion because of the non-normality of the sampling distributions. When results for the small combination of df (3,3) are considered, some observations can be made from Table 2.

Although the statistical test on whether Bias is negligible is dubious, all three estimators suggested evidence of Bias since, for all experimental situations, the values obtained for the Bias statistic are small. There are consistent differences of Bias between $\hat{\beta}$ and $\hat{\beta}_C$ but $\hat{\beta}_2$ tends to show large Biasis.

For (3,16) and (12,75) combinations of df, the results in Table 2 show no significant evidence of Bias. For the (3,16) df the MLE $\hat{\beta}$ is preferred on the basis of the MSE criterion in almost all cases. Although the difference

TABLE 1. **X^2 goodness of fit for testing the agreement among the empirical distributions of the rescaled estimators against the standardized Normal distribution.**

df	β	CS = 1 $X^2_{\hat\beta}$	CS = 1 $X^2_{\hat\beta_c}$	CS = 1 $X^2_{\hat\beta_2}$	CS = 2 $X^2_{\hat\beta}$	CS = 2 $X^2_{\hat\beta_c}$	CS = 2 $X^2_{\hat\beta_2}$	CS = 3 $X^2_{\hat\beta}$	CS = 3 $X^2_{\hat\beta_c}$	CS = 3 $X^2_{\hat\beta_2}$
3,3	-0.5	111.73	84.98	64.14	60.66	57.08	58.54	49.56	28.67	60.27
3,3	-0.4	33.75	28.23	46.89	104.55	77.36	76.37	14.26	15.20	38.72
3,3	-0.3	23.67	26.03	48.71	139.61	65.24	65.42	33.98	29.27	64.06
3,3	-0.2	24.52	28.18	111.64	84.09	65.73	42.53	41.29	33.58	89.51
3,3	-0.1	13.44	22.79	54.59	85.96	73.87	79.85	19.74	13.86	74.81
3,3	0.0	31.72	35.71	82.76	76.38	60.78	69.39	31.48	20.12	94.54
3,16	-0.5	31.38	30.30	14.87	19.29	20.16	11.38	22.48	12.56	18.51
3,16	-0.4	8.97	9.51	11.34	11.51	11.66	7.63	7.59	1.71	20.71
3,16	-0.3	15.54	20.31	19.96	18.26	10.68	11.94	8.59	19.58	11.73
3,16	-0.2	9.41	18.91	6.15	18.12	12.48	11.29	15.17	9.22	18.57
3,16	-0.1	16.45	12.12	23.03	10.46	14.14	13.38	8.13	8.53	7.22
3,16	0.0	18.59	18.36	21.60	15.21	20.03	18.62	20.00	16.79	15.99
12,75	-0.5	13.01	12.45	9.06	19.01	18.57	10.54	6.49	6.68	12.02
12,75	-0.4	13.49	18.62	4.66	7.21	7.18	11.85	16.29	22.98	12.41
12,75	-0.3	9.38	10.56	13.17	7.21	4.52	11.67	22.26	18.45	14.78
12,75	-0.2	4.86	4.17	6.34	17.70	22.47	17.14	3.05	4.41	4.58
12,75	-0.1	7.99	10.19	8.05	11.40	9.78	8.49	10.60	8.61	19.53
12,75	0.0	9.91	10.02	14.91	13.79	11.71	6.15	22.59	12.88	17.37

TABLE 2. **Bias and MSE as criteria for comparing the three estimators of β.**

df	β	CS	Bias (x1000) $\hat\beta_2$	Bias (x1000) $\hat\beta_c$	Bias (x1000) $\hat\beta$	Z Test $\hat\beta_2$	Z Test $\hat\beta_c$	Z Test $\hat\beta$	MSE (x 1000) $\hat\beta_2$	MSE (x 1000) $\hat\beta_c$	MSE (x 1000) $\hat\beta$
3,3	-0.5	1	0.73	0.13	0.50	2.56	3.42	4.55			
3,3	-0.5	2	0.39	0.33	0.22	3.23	3.34	3.24			
3,3	-0.5	3	0.94	0.88	1.53	1.99	6.43	4.34			
3,3	-0.4	1	1.44	0.78	0.30	6.55	7.36	2.31			
3,3	-0.4	2	0.21	0.74	0.73	7.34	8.25	2.41			
3,3	-0.4	3	0.24	0.09	0.30	3.42	1.98	2.52			
3,3	-0.3	1	2.16	0.53	1.37	4.73	3.45	7.22			
3,3	-0.3	2	1.61	1.96	0.13	1.97	4.28	7.25			
3,3	-0.3	3	0.57	0.63	0.14	1.98	3.19	6.42			
3,3	-0.2	1	0.34	0.71	0.56	2.13	3.08	5.13			
3,3	-0.2	2	0.52	0.33	1.34	2.24	2.97	3.14			
3,3	-0.2	3	0.49	0.70	0.26	2.35	2.06	2.12			
3,3	-0.1	1	2.48	0.22	0.21	2.88	7.05	8.33			
3,3	-0.1	2	2.80	1.60	4.00	8.99	2.63	9.55			
3,3	-0.1	3	0.78	0.14	0.20	7.30	4.44	3.16			
3,3	0.0	1	1.36	0.15	0.39	6.46	1.96	5.37			
3,3	0.0	2	0.48	0.44	1.09	3.25	7.48	5.19			
3,3	0.0	3	1.40	0.30	0.22	6.14	5.39	3.20			
3,16	-0.5	1	0.01	0.11	0.04	0.34	0.88	0.95	10.43	11.23	9.11
3,16	-0.5	2	0.15	0.05	0.14	0.47	0.14	0.48	9.54	10.32	9.43
3,16	-0.5	3	0.28	0.28	0.34	0.86	0.85	0.89	10.81	11.04	10.09
3,16	-0.4	1	0.38	0.35	0.36	0.98	1.05	1.15	15.25	11.12	11.14
3,16	-0.4	2	0.19	0.33	0.27	0.50	0.82	0.50	14.98	16.27	14.68
3,16	-0.4	3	0.11	0.10	0.18	0.33	0.30	0.36	12.17	11.16	10.23
3,16	-0.3	1	0.18	0.03	0.13	0.39	0.06	0.44	20.73	16.03	16.34
3,16	-0.3	2	0.92	1.08	0.92	1.99	2.25	2.00	21.64	23.08	21.49
3,16	-0.3	3	0.11	0.24	0.19	0.30	0.76	0.33	12.22	10.47	10.41
3,16	-0.2	1	0.20	0.12	0.08	0.41	0.26	0.46	23.01	19.19	18.52
3,16	-0.2	2	0.83	1.19	0.85	1.73	1.28	1.74	22.93	27.31	22.84
3,16	-0.2	3	0.19	0.08	0.09	0.53	0.22	0.56	12.29	11.85	11.14
3,16	-0.1	1	0.30	0.39	0.32	0.59	0.83	0.64	24.62	22.17	21.04
3,16	-0.1	2	0.98	0.44	0.77	1.91	0.79	1.91	26.34	30.55	26.08
3,16	0.0	3	0.71	0.48	0.59	1.93	1.37	1.04	13.85	12.34	12.41
12,75	-0.5	1	0.00	0.05	0.04	0.00	0.52	0.01	1.96	0.83	0.84
12,75	-0.5	2	0.19	0.18	0.13	1.39	1.29	1.39	1.96	1.96	1.95
12,75	-0.5	3	0.17	0.17	0.17	1.13	1.23	1.23	2.28	1.93	1.92
12,75	-0.4	1	0.04	0.06	0.05	0.24	0.35	0.28	3.25	2.46	2.47
12,75	-0.4	2	0.03	0.02	0.02	0.17	0.12	0.18	3.53	3.49	3.48
12,75	-0.4	3	0.16	0.19	0.19	1.11	1.40	1.20	2.21	1.92	1.91
12,75	-0.3	1	0 33	0.30	0.31	1.64	1.72	1.89	4.10	3.07	3.10
12,75	-0.3	2	0.10	0.10	0.10	0.47	0 49	0.47	4.11	4.08	4.07
12,75	-0.3	3	0.07	0.14	0.14	0.47	0.92	0.51	2.52	2.22	2.22
12,75	-0.2	1	0.13	0.09	0.10	0.60	0.44	0.64	4.84	4.25	4.21
12,75	-0.2	2	0.03	0.05	0.04	0.12	0.22	0.13	4.93	4.84	4.83
12,75	-0.2	3	0.03	0.08	0.07	0.16	0.54	0.17	2.58	2.28	2.25
12,75	-0.1	1	0.14	0.16	0.15	0.58	0.73	0.64	5.50	4.52	4.50
12,75	-0.1	2	0.21	0.21	0.22	0.89	0.88	0.90	5.60	5.51	5.50
12,75	-0.1	3	0.11	0.17	0.16	0.73	1.16	0.78	2.36	2.10	2.08
12,75	0.0	1	0.25	0.29	0.26	1.08	1.38	1.19	5.17	4.35	4.31
12,75	0.0	2	0.44	0.48	0.46	1.88	1.07	1.90	5.41	5.31	5.30
12,75	0.0	3	0.00	0.06	0.04	0.01	0.41	0.01	2.56	2.26	2.25

between the MSE values for $\hat\beta$ and $\hat\beta_c$ is rarely large. For (12,75), the performance of $\hat\beta$ is again slightly better than the other two estimators. Cochran's estimator has the smallest MSE in only three experimental situations out of eighteen, but this its advantage over Cochran' approximation is relatively small.

Comparisons between the variances of the estimators

To use the MLE or Cochran's estimator of the covariance coefficient in practice, it is necessary to know the variance of these estimators. We therefore assess the variances of β_c and β.

i) The estimated variance of the Cochran's estimator is

$$V(\hat\beta_c) = \frac{1}{\dfrac{D_{XX}}{\hat\sigma^2_{12}} + \dfrac{E_{XX}}{\hat\sigma^2_{22}}} = \frac{1}{\dfrac{(df_1-1)s^2_{11}}{s^2_{12}-b^2_1 s^2_{11}} + \dfrac{(df_2-1)s^2_{21}}{s^2_{22}-b^2_2 s^2_{21}}} \qquad (12)$$

where $b^2_1 s^2_{11}$ and $b^2_2 s^2_{21}$ are the reductions in main and split-plot error sums of squares, due to fitting covariance coefficients.

ii) The asymptotic variance of the full MLE is obtained by evaluating the 5 x 5 Fisher Information Matrix (Kendall & Stuart 1967) at the MLE values, giving the following equation:

$$V(\hat{\beta}) = \frac{1}{\dfrac{df_1\, s_{11}^2}{\hat{\beta} s_{11}^2 (\hat{\beta} - 2b_1) + s_{12}^2} + \dfrac{df_2\, s_{21}^2}{\hat{\beta} s_{21}^2 (\hat{\beta} - 2b_2) + s_{22}^2}} \quad (13)$$

Both variances are compared using the same criteria defined in the last section. The split-plot coefficient was neglected because of its poor performance in the previous section.

In the previous simulation study, the mean and variance for each estimator of the theoretical variances under each experimental situation, were computed. The sampling distributions of the estimators are very skewed, though the asymmetry decreases considerably for the third combination of df (12,75). The condition for using the MSE criterion to compare estimated variances through the X^2 statistic was verified, by standardising each sampling distribution and comparing them against the theoretical X^2 distribution with 12 df (since there are two parameters imposed in the stadardisation of the data and there are fifteen pairs of observed frequencies to be compared). The results for the X^2 statistic are presented in Table 3.

TABLE 3. X^2 statistic used to check the shape equivalence of the standardized distributions for estimators of the variance $V(\beta)$.

		CS = 1	CS = 2	CS = 3
df	β	$X^2_{V(\hat\beta)V(\hat\beta_c)}$	$X^2_{V(\hat\beta)V(\hat\beta_c)}$	$X^2_{V(\hat\beta)V(\hat\beta_c)}$
3,3	-0.5	6.66	10.19	6.24
3,3	-0.4	3.60	12.53	16.29
3,3	-0.3	0.67	24.19	5.25
3,3	-0.2	13.48	15.15	5.08
3,3	-0.1	3.76	1.39	4.07
3,3	0.0	9.16	12.68	9.47
3,16	-0.5	6.14	4.01	11.88
3,16	-0.4	13.95	0.87	4.45
3,16	-0.3	12.62	4.51	4.72
3,16	-0.2	10.46	3.45	2.94
3,16	-0.1	9.98	6.24	4.84
3,16	0.0	21.44	3.85	4.89
12,75	-0.5	4.04	0.79	11.89
12,75	-0.4	2.20	0.28	1.53
12,75	-0.3	4.65	0.31	4.62
12,75	-0.2	4.65	0.06	8.95
12,75	-0.1	7.32	0.60	6.54
12,75	0.0	6.23	0.39	3.27

Since in almost all experimental situations the values obtained for the X^2 statistic in Table 3 are smaller than the critical $X^2_{12,5\%} = 21.03$, it is reasonable to believe that the distributions, although not Normal, had the same shape.

The results found for the observed variance, bias and the MSE of $V(\hat\beta)$ and $V(\hat\beta_C)$; are presented in Table 4. Since the true sampling variance of B is unknown, the Bias criterion is defined as follows:

Bias $[V(\hat\beta)]$ = |observed mean of $V(\hat\beta)$ - observed variance of $\hat\beta$|,

where the mean of $V(\hat\beta)$ was found from the sampling distribution of $V(\hat\beta)$, for each experimental situation, and the variance of $\hat\beta$ was found from the sampling distribution of $\hat\beta$. The MSE of $V(\hat\beta_C)$ was assessed as usual.

Table 4 shows that, for (3,3), the Bias for $V(\hat\beta_C)$ is small in all experimental situations. Although it was not possible to construct a Z test, we can see in the same table that the observed variance and MSE statistics, for both estimators, are not far from each other, suggesting a less degree of importance for the Bias criterion in our choice of the best estimator. For this same sample size, the MSE criterion for $V(\hat\beta)$ is minimum in all experimental situations where it could be applied. As Kendall & Stuart (1967) remarked when they discussed the properties of Maximum-Likelihood Estimators, the optimum properties of MLE are asymptotic properties. We have, based on these optimum MLE properties, much inducement to expect better performance for $V(\hat\beta)$ under large combinations of df.

For the other two sets of df, the Bias decreases considerably for both variances. In some situations, $V(\hat\beta)$ is the estimator with smaller Bias than $V(\hat\beta_C)$, mainly when CS = 2. The introduction of Bias in $V(\hat\beta_C)$ is, perhaps, explained by the use of sampling variances in (3) as the best estimators of the true variance. There is a strong indication, from the Bias results shown in Table 4, that the sampling variances do not account for all information needed to be regarded as the best estimators of the theoretical va-

TABLE 4. Comparisons between the variance of the Cochran's estimator and the asymptotic variance of the MLE.

df	β	CS	Obser. Variance $V(\hat{\beta})$	$V(\hat{\beta}_c)$	Bias(x 1000) $V(\hat{\beta})$	$V(\hat{\beta}_c)$	MSE(x1000) $V(\hat{\beta})$	$V(\hat{\beta}_c)$
3,3	-0.5	1	0.048	0.088	0.033	0.012	0.059	0.088
3,3	-0.5	2	8.391	14.478	0.280	0.094	9.158	14.490
3,3	-0.5	3	150.224	258.456	2.400	0.046	207.583	260.168
3,3	-0.4	1	0.654	1.163	0.158	0.027	0.864	1.165
3,3	-0.4	2	12.010	21.979	0.650	0.048	16.268	21.993
3,3	-0.4	3	142.060	269.775	2.433	0.551	200.976	272.841
3,3	-0.3	1	1.641	3.306	0.226	0.023	2.111	3.311
3,3	-0.3	2	19.588	35.855	0.894	0.113	20.456	25.876
3,3	-0.3	3	200.503	380.796	2.736	0.857	275.281	388.092
3,3	-0.2	1	1.745	3.364	0.272	0.105	2.457	3.460
3,3	-0.2	2	27.890	49.976	0.655	0.069	32.177	50.009
3,3	-0.2	3	191.328	377.713	3.301	0.970	300.561	387.113
3,3	-0.1	1	4.095	7.638	0.204	0.010	4.476	7.639
3,3	-0.1	2	41.661	74.264	0.619	0.160	45.435	74.526
3,3	-0.1	3	149.939	281.969	1.746	0.012	180.225	281.969
3,3	0.0	1	2.210	4.195	0.268	0.023	2.871	4.199
3,3	0.0	2	24.169	43.480	0.892	0.194	32.168	43.826
3,3	0.0	3	126.562	237.234	2.418	0.235	184.792	237.768
3,16	-0.5	1	0.000	0.001	0.010	0.089	0.008	0.007
3,16	-0.5	2	0.029	0.033	0.040	0.010	0.029	0.033
3,16	-0.5	3	4.268	4.868	0.233	0.285	4.811	5.678
3,16	-0.4	1	0.030	0.034	0.022	0.062	0.032	0.031
3,16	-0.4	2	0.070	0.079	0.044	0.025	0.070	0.079
3,16	-0.4	3	4.249	4.961	0.188	0.225	4.590	5.469
3,16	-0.3	1	0.053	0.059	0.033	0.015	0.062	0.061
3,16	-0.3	2	0.108	0.123	0.011	0.018	0.110	0.126
3,16	-0.3	3	4.803	5.579	0.190	0.115	5.166	5.711
3,16	-0.2	1	0.067	0.079	0.030	0.023	0.076	0.084
3,16	-0.2	2	0.134	0.149	0.012	0.029	0.134	0.158
3,16	-0.2	3	5.181	6.055	0.276	0.289	5.943	6.891
3,16	-0.1	1	0.080	0.094	0.039	0.036	0.096	0.106
3,16	-0.1	2	0.170	0.189	0.013	0.040	0.171	0.208
3,16	-0.1	3	4.124	4.798	0.387	0.299	5.619	5.695
3,16	0.0	1	0.080	0.094	0.029	0.015	0.067	0.097
3,16	0.0	2	0.171	0.195	0.023	0.039	0.177	0.210
3,16	0.0	3	4.594	5.317	0.429	0.355	6.437	6.578
12,75	-0.5	1	0.000	0.001	0.000	0.000	0.000	0.001
12,75	-0.5	2	0.002	0.003	0.000	0.000	0.002	0.003
12,75	-0.5	3	0.033	0.034	0.002	0.001	0.033	0.034
12,75	-0.4	1	0.002	0.003	0.001	0.001	0.002	0.003
12,75	-0.4	2	0.005	0.006	0.002	0.002	0.005	0.006
12,75	-0.4	3	0.036	0.037	0.009	0.013	0.037	0.039
12,75	-0.3	1	0.004	0.005	0.000	0.000	0.004	0.005
12,75	-0.3	2	0.009	0.010	0.000	0.000	0.009	0.010
12,75	-0.3	3	0.041	0.042	0.002	0.002	0.047	0.046
12,75	-0.2	1	0.007	0.008	0.004	0.004	0.009	0.010
12,75	-0.2	2	0.001	0.002	0.001	0.003	0.001	0.002
12,75	-0.2	1	0.041	0.042	0.002	0.002	0.047	0.048
12,75	-0.1	1	0.008	0.009	0.004	0.004	0.010	0.009
12,75	-0.1	2	0.001	0.002	0.003	0.002	0.001	0.002
12,75	-0.1	3	0.039	0.040	0.002	0.003	0.039	0.041
12,75	0.0	1	0.008	0.009	0.000	0.000	0.009	0.010
12,75	0.0	2	0.001	0.002	0.000	0.000	0.001	0.002
12,75	0.0	3	0.043	0.045	0.015	0.012	0.046	0.047

riances. As a consequence, some Bias are introduced in both Cochran's estimator of the true regression coefficient and in its variance.

For these remaining degrees of freedom, the estimators may be considered as unbiased estimators of the true variance. In both cases, the observed variance is in accordance with the MSE criterion. For the same table, the performance of the asymptotic variance was found, when the MSE criterion was dealt with, better in forty nine out of fifty four experimental situations.

CONCLUSIONS

1. In split-plot designs, two independent residual covariance coefficients can be computed. If the regression coefficients are considered homogenous, the split-plot coefficient is normally used to adjust main and split-plot treatment means. However, the accuracy of the adjustments should be greater if an estimator of β is obtained by combining both sampling regression coefficients. This suggestion was affirmatively verified from the results of our simulation study.

2. Since the estimators may be suggested as unbiased estimators of the true β (with exception when $df_1 = df_2 = 3$), only the MSE criterion may be taken into account in our conclusions. The split-plot coefficient resulted in smaller Bias values only in two experimental situations, within (3,3) sample sizes. Cochran's and the full MLE are obviously better estimators of β. $\hat{\beta}_c$ is better than $\hat{\beta}$ in nine out of eighteen experimental situations when these small and equal sample sizes were considered. As veriefied by Carvalho (1988) through a second simulation study, $\hat{\beta}$ shows better behaviour when the sample sizes, although the same, are increased.

3. For different sample sizes, $\hat{\beta}$ is undoubtedly better estimator than $\hat{\beta}_c$. This statement fails in six out of thirty six experimental situations, being three for either (3,16) and (12,75) sets of df. In all of these failure cases, the advantage observed for $\hat{\beta}_c$ or $\hat{\beta}^2$ is irrelevant, besides the fact that, in all above six cases, the ratio between the main and split-plot variances are larger for the concomitant variable, situation which we do not expect to occur in practice.

4. Since one of the purposes of the analysis of covariance is to estimate and test differences among adjusted treatment means, it is important to recognize the factors that affect the adjustment. The relative size of the main and split-plot samples, the pooled regression coefficient and the mean difference of the covariate, all play an important part in the adjustment process.

Specifically, the pooled regression coefficient should be the one with the smallest variance. As a consequence, the standard error (which is different for every pair of treatments which are being compared) of the difference between two treatment means is smaller, leading to higher values for the statistic criterion used to compare the adjusted means. The results for the Bias and MSE criteria, in Table 4, make it easy to recommend the asymptotic variance of the MLE as the best estimator of the true variance of β. Although the Bias for $V(\hat{\beta})_C$) are smallest for the first combination of df, we may decide to have $V(\hat{\beta})$ as the best estimator of $V(\beta)$. That is due to the fact that in, all experimental situations, $V(\hat{\beta})$ is the estimator with minimum mean-squared error. For the two other combinations, since the Bias may be neglected, the asymptotic variance is the one with predominantly smallest MSE.

REFERENCES

BARTLETT, M.S. Some examples of statistical methods of research in agriculture and applied biology. **Journal of the Royal Statistical Society,** London, v.4, p.137-169, 1937. Suppl.

CARVALHO, J.R.P. **Bivariate analysis in intercropping with two levels of error variation.** Reading: Reading University, 1988. 245p. Ph.D. Thesis.

COCHRAN, W.G. **Analysis of covariance.** Raleigh: [s.n.], 1946. 11p. (Institute of Statistics, mimeographed series 6, University of North Carolina).

COCHRAN, W.G. Analysis of covariance: its nature and uses. **Biometrics**, Washington, v.13, n.3, p.261-281, 1957.

COX, D.R.; HINKLEY, D.V. **Theoretical statistics.** London: Chapman and Hall, 1974. 511p.

DEAR, K.D.G.; MEAD, R. **Testing assumptions and other topics in bivariate analysis.** Reading: [s.n.], 1984. 25p. (Statistics in Intercropping. Technical Report, n.2. Department of Applied Statistics).

FEDERER, W.J. **Experimental design.** New York: MacMillan, 1955. 455p.

FISHER, R.A. **Statistical methods for research workers.** Edinburgh: Oliver and Boyd, 1946. 278p.

JOHN, J.A.; QUENOUILLE, M.H. **Experiments: design and analysis.** London: Charles Griffin & Company Ltda., 1977. 343p.

KEMPTHORNE, O. **The design and analysis of experiments.** New York: Robert E. Krieger Publishing Company, 1975. 631p.

KENDALL, M.G.; STUART, A. **The advanced theory of statistics. Vol. 2.** Inference and Relationships. London: Charles Griffin & Company Ltd, 1967. 690p.

PATNAIK, P.B. The non-central X^2 and F distributions and their applications. **Biometrika**, London, v.36, p.202-232, 1949.

TRUITT, J.T.; SMITH, H.F. Adjustment by covariiance and consequent tests of significance in split-plot experiments. **Biometrics**, Washington, v.12, n.1, p.23-39, 1956.

WALD, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. **Transactions American Mathematical Society**, Washington, v.54, p.426-482, 1943.

WILKS, S.S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. **Annals of Mathematical Statistics**, Washington, v.9, p.60-62, 1938.