PERFORMANCE OF THE TRADITIONAL F TESTS
IN SPLIT-PLOT DESIGNS UNDER COVARIANCE HETEROGENEITY

Huynh Huynh, University of South Carolina

and

Leonard S. Feldt, University of Iowa

1. INTRODUCTION

In split-plot analyses of variance the traditional F tests for the treatment and interaction effects demand that the population covariance matrices exhibit a specific structure (Huynh and Feldt, 1970). When this requirement is not fulfilled, some distortion in the level of significance may be expected for these tests. Greenhouse and Geisser (1958), extending a result by Box (1954b), concluded that where the covariance matrices for the plots are equal, the traditional treatment and interaction mean square ratios (MSR) are approximately distributed as central F variates with reduced degrees of freedom. A correction factor, $\varepsilon \leq 1$, evaluated from the common population covariance matrix, may be used to ascertain the degree to which this matrix conforms to the required structure ($\varepsilon = 1$ for strict conformity). This observation, coupled with the simulation results of Collier et al. (1967), indicates that the traditional F tests in split-plot designs with identical covariance matrix will err on the liberal side, e.g., show a size that is larger than the nominal alpha.

In the present paper a theoretical solution is obtained for the problem of determining Type I error probabilities for the tests of the split-plot design. The problem is solved in its general form. That is, the sampling distributions of the mean square ratios for main effects and interaction are derived under any arbitrary set of covariance matrices for the main plots. This solution, coupled with formulas derived by Imhof (1962), makes it possible to determine the exact size of the traditional tests.

2. DISTRIBUTIONS OF THE RATIOS MSR_A AND MSR_{AB} IN THE SPLIT-PLOT DESIGN

Consider g independent k-component normal variates

$$X_j = (x_{1j}, x_{2j}, \cdots, x_{ij} \cdots, x_{kj})',$$

$j = 1, \cdots g$

with mean vectors

$$\mu_j = (\mu_{1j}, \mu_{2j} \cdots, \mu_{ij}, \cdots, \mu_{kj})'$$

and non-singular covariance matrices Σ_j which need not be equal. Each of the k-components (first subscript) corresponds to a level of treatment category A; each of the g populations (second subscript) corresponds to a level of treatment category B (which is also referred to as "group" or "plot"). Thus, the measures under the levels of A are related; the measures under the levels of B are independent. For each population j a random sample of size n_j is drawn, whose members are denoted by

$$X_{js} = (x_{1js}, x_{2js}, \cdots, x_{ijs}, \cdots, x_{kjs})',$$

$s = 1, 2, \cdots, n_j$. The vector X_{js} can be conceptualized as the score vector of the sth member drawn at random from the jth plot.

Let $n = \sum_j n_j$ be the total number of cases (or observation vectors). The effect of the ith treatment of the A category and the interaction of the ith treatment with the jth plot are respectively $\mu_{i.} - \mu_{..}$ and $\mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..}$. The dot (.) notation refers to weighted means. The null hypotheses of interest are

$$H_A: \mu_{i.} - \mu_{..} = 0 \text{ for all } i$$

$$H_{AB}: \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..} = 0 \text{ for all } i, j.$$

The sums of squares associated with the treatments, the interaction and the residual or within error are defined as

$$SS_A = \sum_{i=1}^{k} n(x_{i..} - x_{...})^2$$

$$SS_{AB} = \sum_{i=1}^{k} \sum_{j=1}^{g} n_j(x_{ij.} - x_{i..} - x_{.j.} + x_{...})^2$$

$$SS_{error(w)} = \sum_{j=1}^{g} [\sum_{i=1}^{k} \sum_{s=1}^{n_j} (x_{ijs} - x_{.js} - x_{ij.} + x_{.j.})^2].$$

The mean square ratios normally used to test H_A and H_{AB} are

$$MSR_A = MS_A/MS_{error(w)} = (n - g)SS_A/SS_{error(w)}$$

$$MSR_{AB} = MS_{AB}/MS_{error(w)}$$

$$= (n - g)SS_{AB}/(g - 1)SS_{error(w)} .$$

To obtain the distribution for MSR_A and MSR_{AB}, let $D = I-\underline{1}\underline{1}'/k$ where I denotes an appropriate identity matrix and $\underline{1}$ is the vector having k components all equal to 1. It may be verified that

(1) $$MSR_A = (n-g) \sum_{i=1}^{k-1} \nu_i \chi_i^2(1;\ \delta_i^2) / \sum_{j=1}^{g} \sum_{i=1}^{k-1} \lambda_{ji} \chi_{ji}^2(n_j - 1)$$

where the ν_i's are the eigenvalues of $D \sum_{j=1}^{g} n_j \Sigma_j / n$, the λ_{ji}'s are those of matrices $D\Sigma_j$, and all of the chi-squares are independent. Moreover, the chi-squares in the numerator are central if and only if the hypothesis H_A is true.

A particular case of interest is represented by the situation in which all the covariance matrices Σ_j are equal to Σ. Then $\sum_{j=1}^{g} n_j \Sigma_j / n = \Sigma$ and $\lambda_{ji} = \nu_i$ for all j. Hence, for this case

(2) $$MSR_A = (n-g) \sum_{i=1}^{k-1} \nu_i \chi_i^2(1;\ \delta_i^2) / \sum_{i=1}^{k-1} \nu_i \chi_i^2(n-g)$$

Consider now two matrices. The first matrix, Σ^*, may be formed by g^2 submatrices. Those on the main "diagonal" are $\Sigma_1/n_1, \ldots, \Sigma_g/n_g$ and the others are all zero. The second matrix, G, is also formed by g^2 submatrices. Those on the "diagonal" are equal to $n_j(1-n_j/n)D$, $1 \leq j \leq g$. The submatrix on the ith "row" of the jth "column" is equal to $-n_j n_j\ D/n (1 \leq i \neq j \leq g)$. It may then be verified that

(3) $$MSR_{AB} = \frac{n-g}{g-1} \sum_{i=1}^{(k-1)(g-1)} \gamma_i \chi_i^2\ (1;\ \delta_i^2) \sum_{j=1}^{g} \sum_{i=1}^{k-1} \lambda_{ji} \chi_{ji}^2(n_j-1)$$

where the γ_i's are the positive eigenvalues of the matrix $\Sigma^* G$ and the chi-squares are independent. As before, the non-centrality parameters δ_i^2 are zero if and only if the hypothesis H_{AB} is true.

For the particular case in which all the covariance matrices Σ_j are equal to Σ, the positive eigenvalues of $\Sigma^* G$ are the eigenvalues (ν_i) of $D\Sigma$, each with order of multiplicity $(g - 1)$. Hence,

(4) $$MSR_{AB} = (n-g) \sum_{i=1}^{k-1} \nu_i \chi_i^2(g-1;\ \delta_i^2)/(g-1) \sum_{i=1}^{k-1} \nu_i \chi_i^2(n-g).$$

Formulas (1), (2), (3), and (4), coupled with computational techniques outlined in the next sections, make it possible to compute the probability that a mean square ratio exceeds the critical values of the traditional tests.

3. COMPUTING THE EIGENVALUES

The matrices whose eigenvalues govern the distribution of the mean square ratio are always of the form $(I - \underline{1}\underline{1}'/k)\Sigma$. Let $B = \Sigma^{-1}$ and $D = I - \underline{1}\underline{1}'/k$. Then $(I - \underline{1}\underline{1}'/k)\Sigma = DB^{-1}$. Here D and B are symmetric and B is positive definite. In the present study computation of the eigenvalues was performed via the IBM-supplied subroutine NROOT (1971). The obtained values are accurate up to probably the fifth decimal. This degree of accuracy is sufficient for the purposes under consideration here.

4. COMPUTING THE EXACT PROBABILITIES

The probabilities of the Type I error associated with the traditional tests of the split-plot design can always be written in the form $Pr(Q \geq 0)$ where $Q = \sum_{i=1}^{m} \alpha_i \chi_i^2(h_i)$, all the chi-squares being mutually independent. Imhof (1962) showed that

(5) $$Pr(Q > 0) = 1/2 + \pi^{-1} \int_0^\infty \frac{\sin \theta(u)}{u\rho(u)}\, du$$

where $$\theta(u) = \sum_{i=1}^{m} [h_i \tan^{-1}(\alpha_i u)]/2$$

$$\rho(u) = \prod_{i=1}^{m} (1 + \alpha_i^2 u^2)^{h_i/4} .$$

He also showed that

$$\lim_{u \to 0} \sin\theta(u)/u\rho(u) = (\sum_{i=1}^{m} \alpha_i h_i)/2$$

and that $u\rho(u)$ increases monotonically toward infinity. This allows the numerical integration in (5) to be carried out only over the finite range $0 \leq u \leq U$. The upper limit U was set large enough so that the error due to the truncated interval of integration was sufficiently small. All the integrations were performed via the Gaussian quadrature with 32 points. In this scheme the integrating function was replaced by an appropriate polynomial of degree 63, and the integration was performed as if the function were the polynomial. This method of integration was carried out with the IBM subroutine DQG32 (1971). It was set in such a way

Table 1

Some Population Covariance Matrices Used in the Study (k = 5)

Description	Elements*				
A, ε = .388	1.00				
Source: computer-simulated	.86	1.00			
	.96	.86	1.00		
	.64	.88	.66	1.00	
	.44	.77	.60	.91	1.00
B, ε = .420	1.00				
Source: computed from data in	.85	1.00			
Lindquist (1962, page 167)	.48	.32	1.00		
	.34	.47	.83	1.00	
	.83	.71	.88	.76	1.00
C, ε = .522	1.00				
Source: fictitious	.80	1.00			
	.60	.80	1.00		
	.40	.60	.80	1.00	
	.30	.40	.60	.80	1.00
D, ε = .752	1.00				
Source: Wechsler (1958, page	.81	1.00			
100, Table 20, Variables:	.74	.70	1.00		
Voc., Inf., Sim., BD, OA)	.53	.58	.52	1.00	
	.43	.45	.39	.61	1.00
E, ε =.831	1.00				
Source: Thurstone and	.62	1.00			
Thurstone (1938)	.62	.67	1.00		
	.54	.53	.62	1.00	
	.29	.38	.48	.62	1.00

*All correlations are rounded to the second decimal.

that all the reported probabilities were accurate up to the last tabulated decimal.

5. SITUATIONS CONSIDERED IN THE STUDY OF TYPE I ERROR

In the present study the number of treatments (A) was set at k = 5, and the number of main plots (B) at g = 3. The total number of sampling units was set at n = 18 and 33. It may be recalled that when the covariance matrices are equal, the distributions of the mean square ratios do not depend on the plot (group) sizes n_j per se, but only on their sum, n. It is interesting to note that when n increases indefinitely, each mean square ratio tends stochastically to a linear combination of chi-squares. Therefore, it should be expected that various probabilities associated with large values of n would not vary markedly.

To simplify the study, only covariance matrices with equal variances (1.0 in every case) were used in the study. Under this condition, the traditional tests are valid only when the covariances or correlations are equal. Five matrices with heterogeneous correlations were considered. These matrices had correction factors ε = .388, .420, .522, .752, and .831, respectively. They are displayed in Table 1. In other phases of the study symmetric matrices were employed. These matrices, designated S_ρ, had homogeneous variances of 1.0 and homogeneous correlations indicated by the subscript value. Thus, $S_{.30}$ represents a matrix with variances of 1.0 and all correlations equal to .30.

6. RESULTS FOR THE CASE OF EQUAL COVARIANCE MATRICES

The true probabilities of Type I error, computed as described earlier, are presented in Table 2. They suggest the following trends:

(a) The traditional tests always err on the liberal side, especially when ε and n are small, and α = 2.5 or 1 per cent. Increasing the sample size leads, in most cases, to a slight reduction in the actual probability of Type I error.

(b) Failure of the common covariance matrix to exhibit the required structure has less effect

Table 2

Exact Per Cents of Type I Error Associated with the Traditional Tests in the Split-Plot Design with Equal Σ

Matrix	ε	n	α(%) for Test of Treatment Effect				α(%) for Test of Interaction Effect			
			10	5	2.5	1	10	5	2.5	1
A	.388	18	14.80	10.46	7.54	5.00	16.97	12.08	8.85	5.80
		33	14.50	10.22	7.38	4.90	16.56	11.77	8.53	5.67
		∞	14.13	9.97	7.20	4.78	16.04	11.42	8.28	5.52
B	.420	18	14.79	10.19	7.17	4.60	16.65	11.60	8.21	5.29
		33	14.50	9.95	7.00	4.50	16.27	11.29	8.00	5.16
		∞	14.36	10.12	7.29	4.44	15.76	10.95	7.76	5.02
C	.552	18	13.02	8.40	5.55	3.29	14.40	9.35	6.18	3.64
		33	12.84	8.29	5.50	3.28	14.19	9.22	6.13	3.65
		∞	12.58	8.15	5.43	3.26	13.85	9.05	6.05	3.63
D	.752	18	11.40	6.60	3.91	2.01	12.10	7.04	4.16	2.11
		33	11.33	6.56	3.91	2.03	12.03	7.00	4.17	2.15
		∞	11.18	6.50	3.90	2.04	11.84	6.95	4.17	2.18
E	.831	18	10.86	6.02	3.40	1.64	11.33	6.30	3.56	1.70
		33	10.82	5.99	3.41	1.66	11.30	6.29	3.57	1.73
		∞	10.70	5.96	3.41	1.68	11.17	6.26	3.59	1.76

on the size of the test of H_A than on the size of the test of H_{AB}.

7. RESULTS FOR THE CASE OF UNEQUAL COVARIANCE MATRICES

Preliminary computation indicated that when equality of the covariance matrices does not hold, variation in the plot sizes and the range of the correlations play a major role. Therefore, this part of the study was subdivided into three phases. First, in order to assess the effect of unequal plot sizes, the covariance matrices were restricted to type S_ρ (for which $\varepsilon = 1$). Extreme cases were included to dramatize this effect. Next were considered matrices with wide ranges for the correlations. Finally, matrices with moderate ranges of correlations and different correction factors were considered.

(a) Effect of Unequal Plot Sizes. The data reported in Table 3 confirm the salutary effect of equal plot sizes for the test of interaction. Inequality of plot sizes has little effect on the test of treatment effects. However, variation in plot sizes may seriously invalidate the test of no interaction. The results for the test of H_A are consistent with those of the Box studies (1954a). Box found that, in the case of the completely randomized design, inequality of variance has little effect on the F test so long as the sample sizes are kept equal.

In view of these results, subsequent investigation was made only for the case of equal plot size.

(b) Effect of High Correlations. Exact probabilities of Type I error were also computed for experiments with matrices involving very high correlations. Matrices D, E, and $S_{.99}$ for the three plots was one such configuration. In these situations the probability of Type I error rose markedly above the nominal level, particularly for the test of interaction.

The advantage of the split-plot design over the factorial design depends on the size of the correlation between measures within plots. The higher the correlation, the

Table 3
Exact Per Cents of Type I Error Associated with the Traditional Tests in the Split-Plot Design with Unequal Σ_i: Effect of Unequal Plot Sizes

Matrix for Plot 1	2	3	Size for Plot 1	2	3	α(%) for Test of Treatment Effect 10	5	2.5	1	α(%) for Test of Interaction Effect 10	5	2.5	1
$S_{.10}$	$S_{.90}$	$S_{.90}$	11	11	11	10.78	5.61	2.95	1.28	13.91	8.60	5.41	2.99
			8	11	14	11.78	6.31	3.42	1.53	30.56	22.28	16.35	10.91
			11	14	8	10.78	5.61	2.95	1.28	13.91	8.60	5.42	2.98
			3	15	15	15.15	8.63	4.95	2.39	72.35	64.76	57.82	49.56
			29	2	2	8.56	4.10	1.98	0.75	0.00	0.00	0.00	0.00
$S_{.30}$	$S_{.50}$	$S_{.70}$	11	11	11	10.11	5.06	2.55	1.03	10.50	5.43	2.84	1.21
			8	11	14	10.36	5.22	2.64	1.07	15.87	9.00	5.12	2.43
			8	14	11	10.22	5.13	2.58	1.04	13.04	7.10	3.90	1.77
			3	15	15	10.62	5.38	2.74	1.12	22.94	24.27	8.86	4.70
			29	2	2	9.33	4.56	2.24	0.88	0.85	0.26	0.09	0.02
$S_{.40}$	$S_{.50}$	$S_{.60}$	11	11	11	10.05	5.02	2.51	1.01	10.16	5.11	2.58	1.05
			8	11	14	10.17	5.09	2.55	1.03	12.66	6.70	3.55	1.54
			14	11	8	9.95	4.95	2.47	0.99	8.05	3.84	1.85	0.71
			3	15	15	10.29	5.16	2.60	1.05	15.59	8.68	4.84	2.22
			29	2	2	9.62	4.74	2.35	0.93	3.18	1.24	0.49	0.15

smaller the residual error variance, and the greater is the power of the test. However, when the assumption about covariance matrices is not fulfilled (or only approximately so, as in the case of the matrices E with $\varepsilon = .831$ and D with $\varepsilon = .752$), high correlations may result in a much greater chance of Type I error than would be anticipated.

(c) Effect of Heterogeneity of the Correction Factors. The data reported in Table 4 reveal that departures from the nominal values of α become more serious as the correction factors ε decrease. The effect cannot be ignored when $\varepsilon < .75$. Deviations at $\alpha = 10$ or 5 per cent are not intolerably large when all of the $\varepsilon > .75$. Extremely heterogeneous covariance matrices (with ε in the neighborhood of .5 or .4) almost completely invalidate the traditional tests.

8. CONCLUDING REMARKS

Data are presented in this study describing the performance of the traditional F tests for the split-plot design when nonstandard conditions hold for the covariance matrices. In all situations under investigation, the test for interaction proved to be more vulnerable than the one for treatment effects, especially when the plot sizes are not equal. When heterogeneity of covariance matrices is suspected, or homogeneity appears to hold but $\varepsilon < .8$ for each matrix, multivariate procedures or approximate F tests should be considered. These give better control of Type I error (Arnold, 1973; Huynh and Feldt, 1976; Buynh, in press).

References

Arnold, S. F. (1973) Application of the theory of products of problems to certain patterned covariance matrices. Annals of Statistics, 1, 682-699.

Box, G. E. P. (1954a) Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effects of inequality of variance in the one-way classification. Annals of Mathematical Statistics. 25, 290-302.

Box, G. E. P. (1954b) Some theorems on quadratic forms applied in the study of analysis of variance problems. II. Effects of inequality of variance and of correlation between errors in the two-way classification. Annals of Mathematical Statistics. 25, 484-498.

Collier, R. O., F. B. Mandeville, G. K. & Hayes, T. F. (1967) Estimates of test sizes for several test procedures based on conventional variance ratios in the repeated measures design. Psychometrika, 32, 339-353.

Geisser, S. & Greenhouse, S. (1958) An extension of Box's results on the use of the F distribution in multivariate analysis. Annals of Mathematicsl Statistics, 29, 885-891.

Table 4

Exact Per Cents of Type I Error Associated with the Traditional Tests in the Split-Plot Design with Unequal Σ_i: Effect of Heterogeneity of the Correction Factors ε

Matrix and ε for Plot*			Size for Plot			α(%) for Test of Treatment Effect				α(%) for Test of Interaction Effect			
1	2	3	1	2	3	10	5	2.5	1	10	5	2.5	1
D	E	$S_{.50}$	6	6	6	10.52	5.58	3.00	1.34	10.97	5.92	3.23	1.47
.752	.831	1	11	11	11	10.47	5.53	2.98	1.34	10.93	5.88	3.22	1.48
D	E	E	6	6	6	10.93	6.10	3.48	1.70	11.52	6.48	3.70	1.80
.752	.831	.831	11	11	11	10.84	6.06	3.48	1.72	11.43	8.46	3.71	1.83
D	D	E	6	6	6	11.10	6.29	3.65	1.82	11.78	6.73	3.90	1.93
.752	.752	.831	11	11	11	11.00	6.25	3.64	1.84	11.67	6.69	3.91	1.97
B	C	C	6	6	6	12.09	7.23	4.43	2.38	13.74	8.62	5.52	3.12
.420	.522	.752	11	11	11	11.80	7.00	4.28	2.29	13.43	8.40	5.39	3.06
A	A	E	6	6	6	12.36	7.86	5.14	3.00	14.47	9.36	6.16	3.61
3.88	.388	.831	11	11	11	12.07	7.69	5.03	2.96	14.14	9.15	6.04	3.56
A	B	C	6	6	6	12.72	7.94	5.08	2.89	15.13	10.06	6.81	4.15
.388	.420	.522	11	11	11	12.26	7.57	4.81	2.72	14.65	9.68	6.55	4.00

*The average correlations are .76 for A, .69 for B, .61 for C, .58 for D and .54 for E.

Huynh, H. & Feldt, L. S. (1970) Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. Journal of American Statistical Association, 65. 1582-1589.

Huynh, H. & Feldt, L. S. (1976) Estimation of the Box correction for degrees of freedom from sample data in the randomized block and split-plot designs. Journal of Educational Statistics, 1, 69-82.

Huynh, H. (1978) Some approximate tests for repeated measurement designs. Psychometrika, in press.

Imhof, J. P. (1961) Computing the distribution of quadratic forms in normal variables. Biometrika, 48, 419-426.

IBM Application Program, System/360. (1971) Scientific subroutines package (360-CM-03X) Version III, Programmer's manual. White Plains, New York: IBM Corporation Technical Publication Department.

Lindquist, E. F. (1962) Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin Co.

Thurstone, L. L. & Thurstone, T. F. (1938) Primary mental abilities, Psychometric Monograph, No. 1, University of Chicago Press.

Wechsler, D. (1958) The measurement and appraisal of adult intelligence. Baltimore: Williams and Wilkins.