ADDITIONAL ANALYSES FOR TWO-WAY LAY-OUT*

BU-157-M Walter T. Federer March, 1963

In a previous paper (see Federer [1962], BU-146-M) a number of analyses of variance for eliminating heterogeneity in two directions were described for an experiment laid out in eight columns and seven rows such that the eight columns were the eight blocks of a randomized complete block design with seven treatments. Thus, treatments and blocks (columns) are orthogonal but rows and treatments are not. The analyses described involved fitting linear, quadratic, cubic, quartic, quintic, and sextic regressions within each block and computing the corresponding sums of squares from successively fitting higher degree polynomials. In the other analyses described the successive fitting linear, quadratic, ..., sextic (= total row effects) degree polynomials to row effects was discussed (see Outhwaite and Rutherford, Biometrics 11:431, 1955). The former set of analyses is very costly with degrees of freedom while the latter procedure of removing some or all of the row componants is suitable only when there are negligible interactions between row and column components. This, by the way, is true for any row-column or latin square type lay-out; the appropriateness of standard textbook analyses for latin square type designs depends upon the uniformity and the orientation of row and column gradients with respect to the experimental lay-out.

The purpose of the present paper is to describe alternative analyses involving row and column components and their interactions for five different yield characters recorded for the experiment. Analyses of variance are given for all five characters removing block effects only and removing both block and row effects. The residual mean squares in the latter analyses of variance are used as a base for comparison with the residual mean squares in analyses of variance involving row and column components (e.g., linear, quadratic, etc.) and interactions of the components.

The yield data for the five characters in the experimental arrangement are presented in Table 1, along with row, column, and grand totals. The Z_1, Z_2, Z_3, and Z_4 covariates represent the orthogonal polynomial coefficients for n=8 (e.g., see Fisher, R. A., and Yates, F., Statistical tables for biological, agricultural

*Biometrics Unit, Department of Plant Breeding, Cornell University.

Table 1. Five measurements* on tobacco plants for 7 treatments (A,B,C,D,E,F,G) in eight replicates of a randomized complete block design (H.H.Smith and C. S. Schlottfeldt, 1951)

	Block number and block gradients (Z)								
	1	2	3	4	5	6	7	8	
	Z_1 -7	-5	-3	-1	1	3	5	7	
	Z_2 7	1	-3	-5	-5	-3	1	7	
	Z_3 -7	5	7	3	-3	-7	-5	7	
X_1 X_2 X_3 X_4	Z_4 7	-13	-3	9	9	-3	-13	7	Totals
	F	B	A	F	F	G	B	D	
-3 5 -1 3	1299.2	1369.2	1169.5	1219.1	1120.0	1031.5	1076.4	1099.6	9384.5
	1439.0	1445.0	1413.5	1411.5	1426.0	1386.5	1451.5	1347.0	11320.0
	688.2	717.9	687.6	686.9	669.9	674.3	665.8	690.6	5481.2
	822.0	844.1	833.9	823.5	812.0	823.9	853.2	837.4	6650.0
	408.0	433.6	405.0	414.0	393.5	396.9	405.1	394.2	3250.3
	G	E	E	A	G	B	A	E	
-2 0 1 -7	875.9	844.2	975.8	971.7	827.0	846.5	917.9	947.4	7206.4
	1264.0	1278.5	1300.5	1245.0	1192.0	1206.0	1277.5	1248.0	10011.5
	602.8	588.8	617.7	619.5	572.4	572.9	605.2	635.1	4814.4
	751.7	747.7	766.9	753.7	716.7	727.6	762.8	762.8	5989.9
	353.5	340.5	358.2	356.1	321.5	323.0	343.7	367.6	2764.1
	D	F	C	G	D	D	E	B	
-1 -3 1 1	960.7	968.7	873.4	607.6	671.9	667.8	627.6	787.1	6164.8
	1272.6	1305.3	1191.0	1099.0	1132.0	1163.0	1084.5	1143.0	9390.4
	605.5	612.4	581.8	542.3	530.6	541.0	522.3	569.7	4505.6
	736.1	744.6	719.0	673.2	668.9	684.3	663.0	685.9	5575.0
	350.0	349.5	334.6	302.6	296.2	298.9	290.2	307.7	2529.7
	C	G	G	D	C	C	F	A	
0 -4 0 6	1004.0	975.5	797.8	1000.0	972.2	853.6	776.4	898.3	7277.8
	1343.0	1344.5	1307.0	1335.0	1290.5	1324.5	1244.0	1236.0	10424.5
	618.9	625.9	593.8	604.2	597.1	578.9	570.1	590.8	4779.7
	742.2	769.1	741.7	724.8	724.7	731.1	721.4	699.9	5854.9
	359.8	372.2	347.1	343.8	339.2	330.5	323.7	325.7	2742.0
	A	C	B	B	A	A	C	C	
1 -3 -1 1	1173.2	1322.4	1069.7	1343.3	1083.7	1087.1	960.4	1174.9	9214.7
	1444.5	1458.4	1472.5	1456.3	1383.0	1314.5	1307.5	1365.5	11202.2
	674.6	691.3	670.4	671.8	633.4	630.8	626.6	664.0	5262.9
	806.5	802.5	809.5	798.3	785.2	750.2	760.1	788.2	6300.5
	398.7	400.1	405.4	396.3	371.7	361.5	361.8	375.3	3070.8
	B	A	F	E	B	E	G	F	
2 0 -1 -7	1031.9	1172.6	1093.3	999.4	1146.9	990.2	852.4	1003.3	8290.0
	1452.0	1449.5	1419.0	1339.5	1327.0	1362.0	1300.0	1346.0	10995.0
	654.3	677.6	645.8	619.8	623.8	609.8	605.2	638.5	5074.8
	802.8	801.4	780.5	756.2	753.7	753.6	761.5	750.1	6159.8
	386.7	403.4	380.5	355.4	360.4	354.2	344.0	365.8	2950.4
	E	D	D	C	E	F	D	G	
3 5 1 3	1421.1	1418.9	1169.6	1181.3	993.8	1021.9	1006.2	947.6	9160.4
	1482.0	1477.9	1375.3	1326.5	1275.0	1296.5	1320.0	1334.5	10887.7
	707.8	701.3	664.3	660.3	627.2	653.2	640.2	655.4	5309.7
	811.9	803.1	783.7	765.9	780.7	795.4	793.6	790.9	6325.2
	418.8	414.8	384.9	381.0	358.1	369.1	363.2	381.2	3071.1
0 0 0 0	7766.0	8071.5	7149.1	7322.4	6815.5	6498.6	6217.3	6858.2	56698.6
	9697.1	9759.1	9478.8	9212.8	9025.5	9053.0	8985.0	9020.0	74231.3
	4552.1	4615.2	4461.4	4404.8	4254.4	4260.9	4235.4	4444.1	35228.3
	5473.2	5512.5	5435.2	5295.6	5241.9	5266.1	5315.6	5315.2	42855.3
	2675.5	2714.1	2615.7	2549.2	2440.6	2434.1	2431.7	2517.5	20378.4

*The treatment yields for the five characters are given in the following order: plant height on 7/13/51 in cm.; plant height on 8/14/51 in in.; length of longest leaf on 7/13/51 in cm.; length of longest leaf on 8/14/51 in in.; and width of widest leaf on 7/13/51 in cm. The x and Z values are orthogonal polynomial values for linear through quadratic coefficients obtained from Fisher and Yates [1938].

and medical research workers, Oliver and Boyd, London [1938]) and the covariates X_1, X_2, X_3, and X_4 represent the orthogonal polynomial coefficients for n=7. The interaction covariates are obtained as products of the x_i with the Z_j.

The seven yield equations for the seven treatment effects involving the first character listed in Table 1 are given on page 4 of BU-146-M. Since 8 $Q_{\cdot i \cdot}$ are the values required to obtain solutions for the $\hat{\tau}_i$ these values along with the $Y_{\cdot i \cdot}$ = treatment totals are given in Table 2. With the inverted matrix on page 4 of BU-146-M and the 8 $Q_{\cdot i \cdot}$ values in Table 2, the $\hat{\tau}_i$ are computed for each of the 5 characters. Then, $\sum_{i=1}^{7} \hat{\tau}_i Q_{\cdot i \cdot}$ = the sum of squares for treatment effects eliminating all row effects. The row sums of squares ignoring treatment effects are computed as $\sum_{h=1}^{7} \frac{Y^2_{\cdot \cdot h}}{8} - \frac{Y^2_{\cdot \cdot \cdot}}{56}$; the residual sum of squares may then be computed by subtraction.

The mean squares in the analyses of variance when block effects only are removed are presented in the top part of Table 3. Here we note that the coefficients of variation are relatively large for all characters indicating extraneous variation in the residual mean squares.

In the lower part of Table 3 analyses of variance are presented for the case where both row and column (block) effects are removed. Here we note the drastic reduction in both the treatment and residual mean squares and the relatively large mean square attributable to rows ignoring treatment effects. The row eliminating treatment effects may be obtained by subtraction of appropriate mean squares. (This is because the degrees of freedom are equal; otherwise, sums of squares should be used.) For example, the rows eliminating treatment mean square for 7/31/51 plant height is 193,179-(45646-19954)=167,487 which is a relatively large mean square. The removal of row effects considerably reduced the residual mean squares and, consequently, the coefficients of variation.

Table 2. Treatment totals ($Y_{\cdot i \cdot}$) and $Q_{\cdot i \cdot}$ values and $\hat{\tau}_i$ values removing row effects for the data of Table 1

Treatment	Plant height 7/13/51 $Y_{\cdot i \cdot}$	$8Q_{\cdot i \cdot}$	$\hat{\tau}_i$	Plant height 8/14/51 $Y_{\cdot i \cdot}$	$8Q_{\cdot i \cdot}$	$\hat{\tau}_i$	Length of leaf 7/13/51 $Y_{\cdot i \cdot}$	$8Q_{\cdot i \cdot}$	$\hat{\tau}_i$
A	8474.0	782.8	17.135	10763.5	-261.1	-5.564	5119.5	202.8	4.032
B	8671.0	2218.4	40.205	10953.3	1190.1	20.473	5146.6	215.0	4.117
C	8342.2	1934.9	38.968	10606.9	-303.0	-6.438	5018.9	208.1	4.748
D	7994.7	1319.7	19.802	10422.8	803.6	14.536	4977.7	114.8	1.364
E	7799.5	-288.8	-14.989	10370.0	-230.3	-5.113	4928.5	-289.8	-6.379
F	8501.9	678.7	10.555	10887.3	445.8	9.719	5165.0	131.8	2.159
G	6915.3	-6645.7	-111.675	10227.5	-1645.1	-27.611	4872.1	-582.7	-10.040

Treatment	Leaf length 8/14/51 $Y_{\cdot i \cdot}$	$8Q_{\cdot i \cdot}$	$\hat{\tau}_i$	Leaf width 7/13/51 $Y_{\cdot i \cdot}$	$8Q_{\cdot i \cdot}$	$\hat{\tau}_i$	$\Sigma\hat{\tau}_i Q_{\cdot i \cdot}$ for the five characters in order listed
A	6193.6	2.8	0.416	2965.8	43.1	1.293	119,723.4
B	6275.1	415.3	7.252	3018.2	308.8	5.582	11,297.8
C	6033.7	-96.8	-1.652	2882.3	19.2	0.541	1,353.9
D	6031.9	49.7	0.335	2846.0	-26.7	-1.080	533.9
E	6042.8	-172.3	-3.203	2843.0	-121.0	-2.581	420.4
F	6249.5	-28.7	-0.037	3004.1	38.3	0.833	
G	6028.7	-170.0	-3.109	2819.0	-261.7	-4.588	

Table 3. Analyses of variance for data of Tables 1 and 2

Source of variation	df	Mean squares: Plant height on 7/13/51	Plant height on 8/14/51	Leaf length 7/13/51	Leaf length 8/14/51	Leaf width
Blocks	7	55474	14662	2913	1480	1783
Treatments	6	45646	9522	1622	1577	860
Blocks x treatments	42	30228	9127	2143	2246	1088
Coefficient of variation (%)		17	7	7	6	9
Source	**df**					
Blocks	7	55474	14662	2913	1480	1783
Rows (ignoring treatments)	6	193179	61666	15045	15445	7570
Treatments (eliminating rows)	6	19954	1883	226	89	70
Residual	36	7352	1644	225	294	139
Coefficient of variation (%)		8	3	2	2	3

Since the coefficients of variation are relatively small for four of the five characters it would appear that alternative analyses have little to offer in further reducing the residual mean square; however, for the first character listed it would appear that some alternative analysis of variance would be useful in reducing the residual mean square to a size consistent with the remaining four characters. There appear to be no immediate biological considerations leading to greater variability for the 7/13/51 plant height.

Some criteria need to be set up when heterogeneity in two-way lay-outs is to be eliminated with row and column components and their interactions. Obviously if all row, column, and interaction components are removed there is no variation left, not even for treatments. One criterion would be to eliminate as high a degree polynomial for both rows and columns and interactions as would be required to utilize the same number of degrees as associated with rows and columns. In our case there are 6+7=13 degrees of freedom associated with row and column mean squares. Now, two cases utilizing 15 and 14 degrees of freedom come to mind. These would involve using (i) linear, quadratic, and cubic row and column components and all interactions, and (ii) linear, quadratic, cubic, and quartic row and column components and only those interaction components for which the sum of the polynomial coefficients i+j in the $X_i Z_j$ is less than five. Also, other schemes could be devised.

Another criterion would be to use a form of either (i) or (ii) above such that there are approximately 20 degrees of freedom remaining in the residual square. One such procedure that comes to mind is to fit linear, quadratic, cubic, and quartic row and column components and interactions. This scheme would leave 25 degrees of freedom associated with the residual mean square. Many other schemes could be devised to meet this criterion.

Still another criterion could be to successively fit row linear, column linear, and row linear x column linear; then fit row quadratic, column quadratic, row linear x column quadratic, row quadratic x column linear, and row quadratic x column quadratic; etc. After this, or some other combination, a number of rules could be devised. For example, each degree of freedom could be partitioned as above and then a pooling procedure on each degree of freedom could be devised (e.g., Bozivich, Helen, et al., Annals Math. Stat. 27:1017-1043,[1956]) with only relatively large contributions being omitted from the residual mean square.

Obviously considerable work is still required in the determination of objective and appropriate criteria for determining which degrees of freedom should be used to remove extraneous variation in an experiment. When such criteria and allied procedures become available, then it will be possible to use appropriate analyses for randomized experiments no matter how they are designed. If the first two criteria listed above are not objectionable, and we see no reason why they should be, we may proceed with an analysis for the data of Table 1.

For the analyses of variance in Table 4, the row, column, and row-column interaction effects were removed as covariates. The 15 covariates in Table 4 are: X_1, X_2, X_3, Z_1, Z_2, Z_3, X_1Z_1, X_1Z_2, X_1Z_3, X_2Z_1, X_2Z_2, X_2Z_3, X_3Z_1, X_3Z_2, and X_3Z_3. The first six covariate values for any particular yield are obtained directly from Table 1; the remaining nine values are obtained by multiplication of the six values given in Table 1. The covariance analyses in Table 4 are simply among and within covariance analyses with multiple correlation coefficients being computed on the total and on the within lines in the analysis of covariance.

Table 4. Analyses of variance using row and column linear, quadratic, and cubic gradients and their interactions to remove heterogeneity for the data in Table 1

Source of variation	df	Sums of squares				
		Plant height on		Leaf length on		Leaf width on
		7/13/51	8/14/51	7/13/51	8/14/51	7/13/51
Total	55	1,931,777	543,099	120,120	114,162	63,339
Treatments	6	273,876	57,132	9,734	9,464	5,162
Within treat.	49	1,657,901	485,967	110,386	104,698	58,177
$(1-R^2 \text{ total})\Sigma y^2$	40	554,141	90,280	20,850	19,080	9.902
$(1-R^2 \text{ within})\Sigma y^2$	34	349,858	69,410	14,157	15,276	7,447
Treats. adj.	6	204,283	20,870	6,693	3,804	2,455
		Adjusted mean squares				
Within adj.	34	10,290	2041	416	449	219
Treat. adj.	6	34,047	3478	1116	634	409

A similar set of covariance analyses with the following 14 covariates is presented in Table 5: X_1, X_2, X_3, X_4, Z_1, Z_2, Z_3, Z_4, X_1Z_1, X_1Z_2, X_1Z_3, X_2Z_1, X_2Z_2, and X_3Z_1. The computation procedure is identical to that used in Table 4. Likewise, Table 6 is computed using the 14 covariates for Table 5 plus the following ten covariates: X_1Z_4, X_2Z_3, X_2Z_4, X_3Z_2, X_3Z_3, X_3Z_4, X_4Z_1, X_4Z_2, X_4Z_3, and X_4Z_4.

These computations were performed on a CDC 1604 high speed computer and hence the inversion of the 15x15, 14x14, and 24x24 matrices presented no computational difficulties.

Table 5. Analyses of variance using row and column linear, quadratic, cubic and quartic gradients and lower degree interactions to remove heterogeneity for the data of Table 1

Source of variation	df	Sums of squares				
		Plant height on		Leaf length on		Leaf width on
		7/13/51	8/14/51	7/13/51	8/14/51	7/13/51
$(1-R^2$ total$)\Sigma y^2$	41	540,601	88,459	19,403	17,647	9,573
$(1-R^2$ within$)\Sigma y^2$	35	392,450	72,070	15,224	14,233	7,665
Treat. adj.	6	148,151	16,389	4,179	3,414	1,908
		Adjusted mean squares				
Within adj.	35	11,213	2,059	435	407	219
Treat. adj.	6	24,692	2,731	696	569	318

Table 6. Analyses of variance using row and column linear, quadratic, cubic, and quartic gradients and all interactions to remove heterogeneity for the data of Table 1

Source of variation	df	Sums of squares				
		Plant height on		Leaf length on		Leaf width on
		7/13/51	8/14/51	7/13/51	8/14/51	7/13/51
$(1-R^2$ total$)\Sigma y^2$	31	444,213	71,230	17,408	16,319	8,016
$(1-R^2$ within$)\Sigma y^2$	25	269,245	55,865	11,123	11,830	5,177
Treat. adj.	6	174,968	15,365	6,285	4,489	2,839
		Adjusted mean squares				
Within adj.	25	10,770	2,235	445	473	207
Treat. adj.	6	29,161	2,561	1048	748	473

The fact that the within mean squares adjusted for the various covariates in Tables 4, 5, and 6 are all larger than the corresponding ones in Table 3 is rather surprising. It was expected that the residual mean square for 7/13/51 plant height would be reduced and that the others would be approximately equal to those in Table 3. However, upon further scrutiny of Table 3 we note that the treatment mean squares for 8/14/51 leaf length and for leaf width are considerably smaller than the residual mean squares. This fact is disturbing and apparently reflects the effect of confounding the large row effects with treatment effects and perhaps of treatment x environment interactions. Thus, from some standpoints the analyses in Tables 4 to 6 appear more realistic than the ones in Table 3. Also, the consistency of the error mean squares in Tables 4 to 6 should be noted. The F ratios in Tables 4 and 6 are quite consistent; the F ratios in Table 5 are consistently smaller than those in Tables 4 and 6.

A study of the low yields for 7/13/51 plant height revealed that eight "unusually" low yields fell within rows 3 and 4 and appeared together. Hence, it would appear that removal of row and column effects as in Table 3 would suffice. The five "unusually" highest yields appeared in blocks 1, 2 and 4 with three being in block 2; they were scattered rather than being grouped indicating some previous plot treatments were still effective.

Thus, we conclude that the piece of land used was not really suitable for a comparative experiment, but that despite this analyses were obtained which yield consistent and unambiguous results for all five characters, i.e., the analyses in Tables 4, 5 and 6. The treatment by environment (row) interaction may have been occasioned by the past history of the plots.