AN EXPERIMENTAL DESIGN USED TO ESTIMATE THE OPTIMUM PLANTING DATE FOR COTTON

R. L. ANDERSON

Institute of Statistics, Raleigh, N. C.

H. L. MANNING

Empire Cotton Growing Corporation and Agricultural Department
Uganda, British East Africa

INTRODUCTION

TWO EXPERIMENTS were conducted in 1946–1947 by the Agricultural Research Station of the Empire Cotton Growing Corporation in the Uganda Protectorate of British East Africa to determine the optimum planting date for cotton. One experiment was laid out at the Kawanda Station, where the main rains occur during the first half of the year; the other experiment was laid out at Kabula, where the main rains occur during the second half of the year. Since it was suspected that interference from the insect, *Lygus simonyi*, would be a potent factor, it was not considered advisable to plant a long series of sowing dates in one locality. Lygus is a small capsid bug, which breeds upon grain crops planted with the first rains in a given locality; when these early crops are harvested, the bug migrates on to newly planted cotton, if such is available. There has also been some scanty evidence that where young cotton is planted in contiguous plots, the Lygus bugs would tend to migrate to the younger cotton.

In these experiments, it was decided to plant at three successive dates, separated by intervals of two weeks (one fortnight), at each of eight localities, with a two-thirds overlap between successive localities. That is, dates 1, 2 and 3 were to be used at the first locality; dates 2, 3 and 4 at the second locality; dates 3, 4 and 5 at the third locality; etc. This experimental design allowed us to use ten planting dates, spread over an 18 week period. Four replications were used at each locality, with the three dates at each locality being used in each replicate. Hence the yields at each locality could be analyzed on the basis of a 4×3 randomized blocks design. The design and actual plot yields (kilograms of seed cotton per plot) of the experiment at the Kawanda Station are presented in Table 1. The same design was used at the Kabula Station, except that the planting dates ranged from September 13 through January 17.

TABLE 1

DESIGN AND ACTUAL YIELDS (KGMS/PLOT) OF SEED COTTON FOR A PLANTING DATE TRIAL AT THE KAWANDA STATION*

	Dates										
Locality	1 May 1	2 May 15	3 May 29	4 June 12	5 June 26	6 July 10	7 July 24	8 Aug. 7	9 Aug. 21	10 Sept. 4	Total
1	3.35	3.86	1.99								
	1.49	2.71	2.89								
	2.44	2.18	1.68								
	2.44	1.95	2.13								
	9.72	10.70	8.69								29.11
2		8.84	7.02	7.02							
		6.23	7.93	7.69							
		7.05	6.60	6.61							
		6.55	6.12	5.19							
		28.67	27.67	26.51							82.85
3			6.12	6.34	3.83						
			5.67	5.67	4.30						
			5.22	5.66	5.20						
			6.10	5.17	4.33						
			23.11	22.84	17.66						63.61
4				2.45	3.17	1.94					
				2.80	1.50	1.39					
				2.47	1.95	1.54					
				2.71	1.96	3.19					
				10.43	8.58	8.06					27.07
5					3.85	4.94	3.85				
					4.57	3.53	3.21				
					4.56	4.53	3.82				
					4.70	5.27	3.72				
					17.68	18.27	14.60				50.55
6						5.21	3.75	3.25			
						4.31	4.62	3.00			
						3.81	3.40	3.38			
						4.14	4.17	2.96			
						17.47	15.94	12.59			46.00
7							5.66	5.17	4.94		
							6.06	4.17	3.58		
							5.66	5.94	4.26		
							7.01	5.58	4.62		
							24.39	20.86	17.40		62.65
8								5.11	3.84	2.76	
								4.51	3.55	3.05	
								4.75	3.93	3.39	
								4.18	3.70	2.99	
								18.55	15.02	12.19	45.76
Total	9.72	39.37	59.47	59.78	43.92	43.80	54.93	52.00	32.42	12.19	407.60

*The plot size was 12′ x 42′, approximately 1/86 acre.

The problem is to determine the best planting date, and to set up some kind of confidence limits on this date. One also might want to consider the relative efficiency of this design relative to other designs, assuming that the insect interference were actually negligible. For example, what could one gain if he planted at four successive dates at each locality, giving a three-fourth overlap? Also what is the loss in efficiency as compared to a completely balanced design, whereby all dates are used at each locality, assuming no insect interference?

ANALYSIS OF THE KAWANDA DATA

(a) **Analysis of Variance.** In order to evaluate the true differences between the yields at successive planting dates, it is necessary to adjust the average yields for the differences among the localities. This necessitates a rather complicated least squares solution of the date and locality effects, which we could avoid if there was good evidence of the non-existence of real locality differences. As a preliminary step, we shall consider the analysis of variance for the Kawanda data. It will be granted that we have not demonstrated the non-existence of locality effects even though the analysis of variance shows no significant differences; however, the experimentalist probably would be willing to neglect small locality effects in order to forego the necessity of carrying out the least squares solution.

From Table 1 we note that there are 23 degrees of freedom for the locality, date and residual (locality-date interaction) constants. Since there are 7 degrees of freedom for localities and 9 for dates, there must be 7 remaining for the locality-date interaction. The sum of squares for these 7 degrees of freedom can be determined from the sum of 7 independent squares, each representing one degree of freedom. These 7 independent squares are formed by squaring various linear combinations of the 24 locality-date totals. Let y_{tj} represent the total yield of the 4 plots planted at the t-th date and the j-th locality, where $t = 1, 2, \cdots, 10$ and $j = 1, 2, \cdots, 8$. A linear combination C can be represented as

$$C = \sum a_{tj} y_{tj} \,. \tag{1}$$

Then $C^2/4 \sum a_{tj}^2$ is a quantity which can be used to represent the sum of squares with one degree of freedom. In forming the 7 independent linear combinations, 7 different sets of a_{tj} are needed. In order that any two forms C and C' be independent,

$$\sum a_{tj} a'_{tj} = 0.$$

TABLE 2

THE COEFFICIENTS TO FORM A SET OF 7 INDEPENDENT LINEAR COMBINATIONS FOR THE LOCALITY-DATE INTERACTIONS

Linear Combinations*

Locality	Date	Order of Planting	Yield	1	2	3	4	5	6	7
1	2	2	10.70	+1				+1		+1
	3	3	8.69	−1				−1		−1
2	2	1	28.67	−1				−1		−1
	3	2	27.67	+1				−3		−3
	4	3	26.51					4		4
3	3	1	23.11					4		4
	4	2	22.84		+1			−3		11
	5	3	17.66		−1			−1		−15
4	4	1	10.43		−1			−1		−15
	5	2	8.58		+1			1		−41
	6	3	8.06							56
5	5	1	17.68							56
	6	2	18.27			+1			1	−41
	7	3	14.60			−1			−1	−15
6	6	1	17.47			−1			−1	−15
	7	2	15.94			+1			−3	11
	8	3	12.59						4	4
7	7	1	24.39						4	4
	8	2	20.86				+1		−3	−3
	9	3	17.40				−1		−1	−1
8	8	1	18.55				−1		−1	−1
	9	2	15.02				+1		+1	+1
Total (C)				1.01	3.33	2.14	−0.07	0.78	2.79	17.99
Divisor (D)				16	16	16	16	224	224	43,456
C^2/D				1.0433				0.0375		0.0074

*Blank spaces represent 0 coefficients. Each yield is the total yield of 4 plots at the Kawanda Station.

In addition, these combinations must be independent of date and locality effects. In order to fulfill this condition, the a_{ij} of a given C for any date or locality must also sum to 0.

One set of 7 independent linear combinations for the residual or interaction sum of squares is given in Table 2. The first four combinations represent first-order interaction effects. For example consider

$$C_1 = (y_{21} - y_{31}) - (y_{22} - y_{32}) = y_{21} - y_{31} - y_{22} + y_{32} \, . \tag{2}$$

Each part of the right-hand side measures the difference between the yields at dates 2 and 3 but at two different locations, 1 and 2. Hence the difference between these two parts measures the change in the date effect from one location to another, which we designate as the date-location interaction. We note that C_1 is independent of date and locality effects, since the sum of the coefficients for each date and locality is zero. Also the first four combinations are obviously independent, since they have no plots in common.

TABLE 3
ANALYSIS OF VARIANCE FOR THE KAWANDA DATA

Source	Degrees of Freedom	Sum of Squares	Mean Square
Interaction	7	1.0882	.1555
Localities	7	201.1345	
Dates (adj.)	9	21.7354	2.4150**
Dates	9	45.2186	
Localities (adj.)	7	177.6513	25.3787**
Replications	24	11.6219	0.4842
Error	48	17.3125	0.3606

**Significant at the 1% probability level.

In this experiment these first four combinations also represent some effect of the order of planting on yield (and consequently of Lygus migration from the early to the late plantings). The effect of the Lygus migration from the early to the late plantings is a quadratic effect, since the sums (C) are sums of single yields at the first and third plantings subtracted from two yields at the second planting. The sum of the first four combinations could be used to measure this quadratic effect of the order of planting with one degree of freedom, and the remaining three degrees of freedom would then represent the interaction effect, adjusted for order of planting. The sum of squares for the order of planting would be $(6.41)^2/4(16) = .6420$, leaving 0.4013 for the three interaction degrees of freedom. None of these effects is significantly different from zero, using the error variance given in Table 3. Since there is obviously no date-location interaction, we can use each of the first four C's as a measure of the effect of order of planting. The standard error of each of these C's is 2.40, indicating that none of them is significantly different from zero. It is possible to compute a separate error term for each C with 6 degrees of freedom, but this does not alter any of the above conclusions.

Combinations 5 and 6 are each based on a comparison of 4 dates and 4 localities. For example

$$C_5 = y_{21} - y_{22} - y_{31} - 3y_{32} + 4y_{33} \quad (3)$$
$$+ 4y_{42} - 3y_{43} - y_{44} - y_{53} + y_{54} .$$

Again the sum of the coefficients for any date or locality is 0. In addition the sum of the products of these coefficients with those for any of the first 4 combinations is 0. Combination 7 is an over-all comparison, which was constructed so as to be independent of all the other combinations as well as dates and localities.

The divisors are the denominators, $4 \sum a_{ij}^2$, mentioned above. For the first 4 combinations the divisors are $4(1 + 1 + 1 + 1) = 16$. For the next two, the divisors are $2[4(1 + 1 + 1 + 9 + 16)] = 224$. And finally for the last comparison, the divisor is $4(10864) = 43456$. The combination totals and contributions to the residual sum of squares are given at the bottom of Table 2, and $SSR = 1.0882$.

Next we compute the total sum of squares for the 23 degrees of freedom (SST) and the sum of squares for the unadjusted date effects (SSD), and the sum of squares for the unadjusted locality effects (SSL). These are computed as follows:

$$\text{Correction for mean} = C = (407.60)^2/96 = 1730.6017$$

$$SST = [(9.72)^2 + (10.70)^2 + \cdots + (12.19)^2]/4 - C$$
$$= 223.9581$$

$$SSL = [(29.11)^2 + (82.85)^2 + \cdots + (45.76)^2]/12 - C \quad (4)$$
$$= 201.1345$$

$$SSD = \frac{(9.72)^2}{4} + \frac{(39.33)^2}{8} + \frac{(59.47)^2}{12} + \cdots + \frac{(12.19)^2}{4}$$
$$- C = 45.2186$$

In order to make exact tests of significance, we require the sum of squares for localities adjusted for dates [SSL(adj.)] and the sum of squares for dates adjusted for localities [SSD(adj.)]. A simple identity can be used to evaluate these adjusted sums of squares:

$$(5)\quad \begin{aligned} SSL + SSD(\text{adj.}) &= SSD + SSL(\text{adj.}) \\ &= SST - SSR = 222.8699 \end{aligned}$$

Hence

$$(6)\quad \begin{aligned} SSD(\text{adj.}) &= 222.8699 - 201.1345 = 21.7354 \\ SSL(\text{adj.}) &= 222.8699 - 45.2186 = 177.6513. \end{aligned}$$

Finally the error sum of squares (SSE) is given by the replication by date sum of squares summed over the 8 locations, with $3 \times 2 \times 8 = 48$ degrees of freedom. Because of the large locality differences, it might be suspected that the error variance would not be the same for all locations. Bartlett's χ^2 test for the homogeneity of variance was applied, giving $\chi^2 = 7.364$ with 7 degrees of freedom. The probability of obtaining this or a larger value of χ^2, assuming equal error variances at all localities, is about .40, indicating that it is quite reasonable to assume equal error variances at all localities and to pool the individual estimates as mentioned above. It might be argued that we should also pool the location $\times$ date interaction with the error variance, giving a total of 55 error degrees of freedom. We have not done this in the analysis which follows, because the additional degrees of freedom were not thought necessary in this case.

The complete analysis of variance is given in Table 3.

This analysis indicates that there are highly significant differences among both the dates and localities. Since there are real differences between the mean yields for different localities, the mean yields at each planting date should be adjusted for locality effects. Two types of analysis are suggested at this point. Either we determine the adjusted average yield for each planting date, or we fit a regression curve of some kind to the adjusted average yields.

(b) Average Yields for Each Planting Date, Adjusted for Locality Effects.

We can represent the total yield y_{tj} of the 4 plots at a given date, t, and locality, j, as follows:

$$(7)\qquad y_{tj} = 4(m + d_t + l_j) + r_{tj} ,$$

where m is the general mean, d_t is the effect of the t-th planting date ($t = 1, 2, \cdots, 10$) and l_j is the effect of the j-th locality ($j = 1, 2, \cdots, 8$), r_{tj} is the residual after accounting for the effects of the constants—m, d_t, and l_j. It should be noted that we have multiplied the right-hand side by 4 in order to put the results on a per-plot basis.

The constants will be estimated by the method of least squares. First we set up the error equation

$$\text{(8)} \qquad SSR = \sum [y_{ti} - 4(m + d_t + l_i)]^2/4$$

where SSR indicates the interaction or residual sum of squares, the sum of the 24 squared residuals, $r_{ti}^2/4$. The constants are estimated by minimizing SSR with respect to each of the constants. For d_1, we set

$$\text{(9)} \qquad \frac{\partial SSR}{\partial d_1} = 0.$$

This gives

$$\text{(10)} \qquad \sum_i [y_{1i} - 4(m + d_1 + l_i)] = 0,$$

where the summation is made over only the one locality (4 plots) having the first planting date. Equation (10) simplifies to

$$\text{(11)} \qquad \sum_i y_{1i} = D_1 = 4d_1 + 4l_1 + 4m$$

Similar equations can be obtained for each d_t.

Similarly if we minimize SSR with respect to l_1, we have

$$\text{(12)} \qquad \sum_i y_{t1} = L_1 = 4(d_1 + d_2 + d_3) + 12l_1 + 12m.$$

The equation for m is

$$\text{(13)} \qquad \sum_{ti} y_{ti} = G = 4d_1 + 8d_2 + 12(d_3 + \cdots + d_8) + 8d_9 + 4d_{10} + 12(l_1 + \cdots + l_8) + 96m$$

The coefficients of all 19 least squares equations are presented in Table 4. The reduced equations, after eliminating the locality constants, are given at the bottom of Table 4. In this table the yield totals, adjusted for localities, are indicated by A_t. The G equation indicates that if m is to be the general mean, we should assume that $\sum_i l_i = 0$ and that

$$\text{(14)} \qquad [d_1 + d_{10} + 2(d_2 + d_9) + 3(d_3 + \cdots + d_8)] = 0.$$

In order to obtain standard errors to test for the difference between the adjusted average yields at any two dates, it is necessary to invert the coefficient matrix for the adjusted date effects. The inverted matrix is presented in the upper part of Table 5. The elements of this matrix (which are indicated by the symbol m_{ti}, where t stands for the row and

TABLE 4

COEFFICIENTS OF LEAST SQUARES EQUATIONS FOR ESTIMATION OF PLANTING DATE MEANS

Independent Variables

Plot Totals*	Dates (d_t) 1	2	3	4	5	6	7	8	9	10	Localities (l_j) 1	2	3	4	5	6	7	8	m
D_1 : 9.72	4										4								4
D_2 : 39.37		8									4	4							8
D_3 : 59.47			12								4	4	4						12
D_4 : 59.78				12								4	4	4					12
D_5 : 43.92					12								4	4	4				12
D_6 : 43.80						12								4	4	4			12
D_7 : 54.93							12								4	4	4		12
D_8 : 52.00								12								4	4	4	12
D_9 : 32.42									8								4	4	8
D_{10} : 12.19										4								4	4
L_1 : 29.11	4	4	4								12								12
L_2 : 82.85		4	4	4								12							12
L_3 : 63.61			4	4	4								12						12
L_4 : 27.07				4	4	4								12					12
L_5 : 50.55					4	4	4								12				12
L_6 : 46.00						4	4	4								12			12
L_7 : 62.65							4	4	4								12		12
L_8 : 45.76								4	4	4								12	12
G : 407.60	4	8	12	12	12	12	12	12	8	4	12	12	12	12	12	12	12	12	96

EQUATIONS FOR DATE EFFECTS ADJUSTED FOR LOCALITIES

Adjusted Totals*		Coefficients of Adjusted Date Effects (d_t) 1	2	3	4	5	6	7	8	9	10
$A_1 = 3D_1 - L_1$	= 0.05	8	−4	−4							
$A_2 = 3D_2 - L_1 - L_2$	= 6.15	−4	16	−8	−4						
$A_3 = 3D_3 - L_1 - L_2 - L_3$	= 2.84	−4	−8	24	−8	−4					
$A_4 = 3D_4 - L_2 - L_3 - L_4$	= 5.81		−4	−8	24	−8	−4				
$A_5 = 3D_5 - L_3 - L_4 - L_5$	= −9.47			−4	−8	24	−8	−4			
$A_6 = 3D_6 - L_4 - L_5 - L_6$	= 7.78				−4	−8	24	−8	−4		
$A_7 = 3D_7 - L_5 - L_6 - L_7$	= 5.59					−4	−8	24	−8	−4	
$A_8 = 3D_8 - L_6 - L_7 - L_8$	= 1.59						−4	−8	24	−8	−4
$A_9 = 3D_9 - L_7 - L_8$	= −11.15							−4	−8	16	−4
$A_{10} = 3D_{10} - L_8$	= −9.19								−4	−4	8

*Plot Totals at Kawanda.

j the column number) can be used to determine both the values of the adjusted date effects and their variances. Before inverting the 10-row coefficient matrix at the bottom of Table 4, it is noted that the coefficients always sum to 0 for a given row or column. Such a matrix is

TABLE 5

INVERSE (M_{tj}) MATRIX FOR ADJUSTED DATE EFFECTS*

j

	1	2	3	4	5	6	7	8	9	10
1	.48884									
2	.37500	.40550								
3	.35268	.34450	.36086							
4	.30580	.30799	.30362	.31892						
t 5	.26562	.26504	.26622	.26208	.27744					
6	.22322	.22337	.22306	.22414	.22015	.23503				
7	.18304	.18299	.18308	.18276	.18396	.17949	.19615			
8	.13616	.13617	.13615	.13621	.13601	.13675	.13398	.14434		
9	.11384	.11383	.11385	.11379	.11399	.11325	.11602	.10566	.14434	
10	0	0	0	0	0	0	0	0	0	
A_j**	0.05	6.15	2.84	5.81	−9.47	7.78	5.59	1.59	−11.15	−9.19
d'_t	4.300	4.478	4.110	3.856	2.981	3.022	2.387	1.598	0.699	0
$(d_t + m)$	5.691	5.869	5.501	5.247	4.372	4.413	3.778	2.989	2.090	1.391
$s^2(d'_t)$	0.5288	0.4387	0.3904	0.3450	0.3001	0.2543	0.2122	0.1561	0.1740	0
$s(d'_t)$	0.7272	0.6623	0.6248	0.5874	0.5478	0.5043	0.4607	0.3951	0.4171	0

*Since the matrix is symmetrical, only the lower left section is reproduced here.
**Data from the Kawanda experiment.

called a *singular matrix* and cannot be inverted. This difficulty could have been taken care of by dropping the last equation and eliminating d_{10} by use of equation (14). However, we decided on a simpler procedure, namely to assume $d_{10} = 0$, and hence drop the 10-th row and column from the matrix. Then we merely inserted a column and a row of zeros in the inverted matrix. We have indicated the date effects under this assumption as d'_t .

For those not accustomed to the matrix notation, we might add that a matrix is simply a two-way array of figures such as the coefficients in Table 4. We are here dealing with symmetrical matrices. An inverted symmetrical matrix consists of another two-way array which has the following property: If you multiply the elements of any row (or column) of the original symmetrical matrix by the corresponding elements of the same row (or column) of the inverted matrix, the sum of the products is unity; if you multiply the corresponding elements of different rows or columns, the sum of the products is zero.

The values of the adjusted date effects are then computed from the M and A values as follows:

$$d'_t = \sum_{j=1}^{10} M_{tj} A_j \tag{15}$$

For example, $d_1' = [(0.48884)(0.05) + \cdots + (0.11384)(-11.15)] = 4.3005$. Note that all $M_{t,10}$ values are 0. The complete set of d_t' values is presented in Table 5.

In order to obtain the original d_t values, we subtract a constant k from each d_t' so that the new values $d_t' - k = d_t$ fulfill equation (14). We see that

$$k = [d_1' + 2d_2' + 3(d_3' + \cdots + d_8') + 2d_9' + d_{10}']/24 = 2.855. \tag{16}$$

In most cases, it is desired to replace the date effects by adjusted yields at each date by adding m ($= 4.246$) to each of the d_t . Hence we can combine both steps by adding

$$m - k = 4.246 - 2.855 = 1.391 \tag{17}$$

to each of the d_t' . These adjusted yields for each date ($d_t + m$) are also given in Table 5.

The variance of any adjusted date effect (d_t') is simply $3M_{tt}\sigma^2$, where σ^2 is estimated by the error variance, 0.3606. The factor, 3, is required because the D_t were multiplied by 3 in determining the A_t '. The variance of d_1' , for example, is

$$3(0.48884)(0.3606) = 0.5288$$

The variances and standard errors of d_t' are also given in Table 5.

However, we are generally more interested in the variances of the differences between the adjusted mean yields for successive dates. For example, the difference between the average yields for the first two dates is 0.18. The variance of the difference between the adjusted mean yields at two dates, i and k is given by the formula

$$3(M_{ii} - 2M_{ik} + M_{kk}]\sigma^2 \tag{18}$$

where σ^2 is the error variance, estimated to be 0.3606. Hence the variance of the difference between the average yields for the first two dates is

$$3(0.48884 - 0.75000 + 0.40550)(0.3606) = (0.43301)(0.3606) = 0.1561.$$

Hence we conclude that there was no significant difference between the two adjusted mean yields. The coefficients of the error variance (for example, 0.43301) are given in Table 6.

Some of the differences between adjusted mean yields for successive dates and their standard errors are given in the second half of Table 6 on the next page.

TABLE 6

COEFFICIENTS OF VARIANCES OF DIFFERENCES BETWEEN ADJUSTED DATE MEANS

Date

	2	3	4	5	6	7	8	9	10
1	.43301	.43301	.58846	.70508	.83232	.95676	1.08256	1.21649	1.46652
2		.23205	.32532	.45860	.58137	.70701	.83250	.96651	1.21649
3			.21762	.31758	.44928	.57253	.69866	.83250	1.08256
Date 4				.21658	.31703	.44866	.57253	.70701	.95676
5					.21652	.31703	.44928	.58137	.83232
6						.21658	.31758	.45860	.70508
7							.21762	.32532	.58846
8								.23205	.43301
9									.43301

(19)

Dates	Mean Difference	Standard Error
1–2	−0.18	0.40
2–3	0.37	0.29
3–4	0.25	0.28
4–5	0.88	0.27
5–6	−0.04	0.27
6–7	0.63	0.27
7–8	0.79	0.28
8–9	0.90	0.29
9–10	0.70	0.40
4–6	0.84	0.34
5–7	0.59	0.34
2–4	0.62	0.34
2–5	1.50	0.41
2–6	1.46	0.46

From these averages, we conclude that: (i) The optimum planting date is somewhere near the first planting date (May 1). This optimum is poorly determined, as shown by the fact that there is no significant decrease in the adjusted yield until the fifth planting date (June 26). (ii) There is an unaccountable plateau at the fifth and sixth planting dates. Except for this, after the fourth date there appears to be a general significant decrease in yield between successive planting dates until we reach the last date (September 4). There is a tendency to flatten out at this last date, as expected, because the yields cannot fall below zero.

(c) Determination of the regression of cotton yields (adjusted for locality differences) on planting date. An examination of the adjusted mean yields ($d_t + m$) in Table 5 suggests that a regression equation of the following type should be used:

$$20) \qquad y_{ti} = 4[a + b(t - \bar{t}) + c(t - \bar{t})^2 + l_i] + p_{ti} \; ,$$

where $\bar{t}$ is the mean time period (= 5.5) and p_{ij} is the residual after accounting for the parabolic trend and the effect of the j-th locality, a, b and c are the coefficients of the regression curve, to be estimated from the data. As before, we shall assume that $\sum_j l_j = 0$.

The constants are estimated by least-squares. For example, the least-squares equation for b is

$$(21) \qquad \sum [y_{ij}t' - 4(at' + bt'^2 + ct'^3 + l_j t'_j)] = 0;$$

where the summation extends over all plots, $t' = t - \bar{t}$, and t'_j implies the values of t' used at the j-th locality. Since $\sum t' = \sum t'^3 = 0$ over all plots, equation (21) simplifies as follows:

$$(22) \qquad \sum yt' = 568b - (42l_1 + 30l_2 + 18l_3 + 6l_4 - 6l_5 - 18l_6 - 30l_7 - 42l_8).$$

TABLE 7

COEFFICIENTS OF LEAST SQUARES EQUATIONS FOR ESTIMATION OF OPTIMUM PLANTING DATE OF COTTON

Independent Variables

Plot Totals*	l_1	l_2	l_3	l_4	l_5	l_6	l_7	l_8	a	b	c
L_1 : 29.11	12	0	0	0	0	0	0	0	12	−42	155
L_2 : 82 85	0	12	0	0	0	0	0	0	12	−30	83
L_3 : 63.61	0	0	12	0	0	0	0	0	12	−18	35
L_4 : 27.07	0	0	0	12	0	0	0	0	12	− 6	11
L_5 : 50.55	0	0	0	0	12	0	0	0	12	6	11
L_6 : 46.00	0	0	0	0	0	12	0	0	12	18	35
L_7 : 62.65	0	0	0	0	0	0	12	0	12	30	83
L_8 : 45.76	0	0	0	0	0	0	0	12	12	42	155
G : 407.60*	12	12	12	12	12	12	12	12	96	0	568
B : −39.22*	−42	−30	−18	− 6	6	18	30	42	0	568	0
C : 2299.62*	155	83	35	11	11	35	83	155	568	0	6742

Solution for b and c

(1) $B + (7L_1 + 5L_2 + 3L_3 + L_4 - L_5 - 3L_6 - 5L_7 - 7L_8)/2 = -32.32 = 64b$

(2) $C - (155L_1 + 83L_2 + 35L_3 + 11L_4 + 11L_5 + 35L_6 + 83L_7 + 155L_8)/12 = -64.68 = 4096c/3$

(3) $G = 96a + 568c$

*G = Grand Total; $B = \Sigma y(t - \bar{t})$; $C = \Sigma y(t - \bar{t})^2$; Plot totals at Kawanda.

The coefficients for all of the equations and the requisite data from the Kawanda experiment are given in Table 7. If we multiply the L equations by the appropriate constants, as shown at the bottom of Table 7, and add to the b and c equations, we have at once the equations to estimate the values of b and c adjusted for locality effects,

$$(23) \qquad b = -0.5050, \qquad c = -0.04737.$$

And then $a = \bar{y} - 568c/96 = 4.5261$.

From these final equations, we also see that b and c are independent and that the variances of b and c are

$$(24) \qquad \sigma^2(b) = \sigma^2/64; \qquad \sigma^2(c) = 3\sigma^2/4096,$$

where σ^2 is estimated from the error variance, in this case 0.3606. Hence the estimated variances and standard errors of b and c are

$$(25) \qquad \begin{aligned} s^2(b) &= 0.005634; & s(b) &= 0.07506 \\ s^2(c) &= 0.0002641; & s(c) &= 0.01625 \end{aligned}$$

These results indicate that both b and c are significantly different from 0.

A comparison of the adjusted mean yields based on the quadratic regression equation (20) and on the original equation (7) (Table 5) is given below.

ADJUSTED MEAN YIELD

	Date	(7)	(20)	Deviation
	1	5.691	5.839	−.148
	2	5.869	5.713	.156
	3	5.501	5.492	.009
	4	5.247	5.177	.070
(26)	5	4.372	4.767	−.395
	6	4.413	4.262	.151
	7	3.778	3.662	.116
	8	2.989	2.968	.021
	9	2.090	2.178	−.088
	10	1.391	1.294	.097
	Average*	4.247	4.247	.000

*The average is computed as $[d_1 + d_{10} + 2(d_2 + d_9) + 3(d_3 + \cdots + d_8)]/24$.

From (26), we see that there is no pronounced trend in the deviations, such as consistent positive or negative deviations at the early and late dates or in the middle. The only large deviation is at the fifth date, for which we previously noted the unexplained sharp drop. The chief difference between the two series of adjusted yields is that the maximum point is at the first date using the quadratic regression as compared to the second date for the original adjusted yields.

It might be advisable to check the adequacy of our quadratic prediction equation in estimating the adjusted yields. We find that the reduction in total sum of squares due to the quadratic regression is given by

$$(27) \qquad (-32.32)(-0.5050) + (-64.6758)(-0.04737) = 19.3853.$$

Hence the remaining sum of squares for the other 7 degrees of freedom for dates is

$$(28) \qquad 21.7354 - 19.3853 = 2.3501$$

This mean square is $2.3501/7 = 0.3357$, which is even less than the error mean square.

The planting date to give the maximum yield can be estimated by differentiating the estimating quadratic equation (20) with respect to t and equating the result to 0. This gives as a result the maximum planting date[1]

$$(29) \qquad t_{max} = \bar{t} - b/2c = 0.17$$

The variance of this estimate can only be approximated. If only the first order terms of the Taylor expansion of the differential of $b/2c$ is used, we find that

$$(30) \qquad \begin{aligned} \sigma^2(b/2c) &= \frac{\sigma^2(b)}{4c^2} + \frac{b^2\sigma^2(c)}{4c^4} = \left(\frac{b}{2c}\right)^2\left[\frac{\sigma^2(b)}{b^2} + \frac{\sigma^2(c)}{c^2}\right] \\ &= \left(\frac{b}{2c}\right)^2\left[\frac{1}{64b^2} + \frac{3}{4096c^2}\right]\sigma^2, \end{aligned}$$

where σ^2 is estimated by s^2. Hence

$$(31) \qquad s^2(b/2c) = 3.994; \qquad s(b/2c) = 2.00.$$

The standard error of t_{max} is the same as $s(b/2c)$, because t is fixed. Hence the maximum planting date $\pm$ two standard errors is given by 0.17 ± 4.00. As expected, the confidence interval is very large, indicating that the difference between the two maximum points in (26) is unimportant. The variance of the estimate of the optimum planting is large for two reasons: (i) There is a serious loss of information in making the adjustments for locality differences. For example, the coefficient of b is reduced from 568 to 64 in making the adjustments (Table 7). (ii) The optimum point comes before or near the first planting date used in the experiment. If the optimum planting date were near $\bar{t}$, so that b would be small, $s^2(b/2c)$ would be materially reduced.

(d) Effects of Lygus Infestation. Unfortunately this design does not furnish any means of determining if there is any linear decrease in the

[1]This is a maximum only if c is negative.

yield for the last planting at a given locality because of a Lygus migration from the earlier planted plots. If we include a constant in equation (20) to represent the difference between the yields of the first and third plantings at all localities, this constant proves to be the same as the linear regression coefficient, b, of yield on planting date, adjusted for localities. As indicated before, the quadratic effect of the Lygus migration can be measured by taking the differences between the yield for the second planting and the average of the yields for the first and third plantings. This quadratic effect was shown to be non-significant.

TABLE 8

INDEX OF LYGUS DAMAGE PER COTTON PLANT FOR THE KAWANDA EXPERIMENT

Locality	1			2			3			4		
Date	1	2	3	2	3	4	3	4	5	4	5	6
	2.16	1.16	2.89	4.02	2.00	1.54	3.66	2.17	1.46	1.78	2.13	2.47
	2.49	1.69	2.22	3.20	3.35	1.65	3.64	1.91	1.56	3.59	3.04	2.27
	1.71	1.60	2.89	3.69	4.11	1.09	3.84	1.52	1.56	2.17	2.87	2.76
	1.62	1.33	1.67	3.00	4.55	2.50	3.73	2.11	1.53	2.26	2.31	3.20
Total	7.98	5.78	9.67	13.91	14.01	6.78	14.87	7.71	6.11	9.80	10.35	10.70
	23.43			34.70			28.69			30.85		

Locality	5			6			7			8		
Date	5	6	7	6	7	8	7	8	9	8	9	10
	3.47	4.24	2.34	3.16	3.25	2.04	2.62	2.37	0.78	1.71	0.89	1.16
	3.58	3.89	2.31	3.85	2.05	1.67	1.51	2.39	1.11	2.46	1.49	1.51
	3.24	4.26	2.00	3.78	4.05	1.73	2.20	2.04	0.82	1.89	2.40	1.16
	4.96	3.69	3.07	4.00	2.16	2.02	1.62	2.04	1.20	2.60	1.13	0.71
Total	15.25	16.08	9.72	14.79	11.51	7.46	7.95	8.84	3.91	8.66	5.91	4.54
	41.05			33.76			20.70			19.11		

In order to check if there was an adverse effect on the cotton yield by Lygus migration, an index of the Lygus damage per plant was determined for each plot of the Kawanda experiment. The Lygus damage data are given in Table 8. Using these data, the cotton yields (Y) were adjusted for the Lygus damage (X) in the following analysis of covariance, where $x = X - \bar{X}$ and $y = Y - \bar{Y}$.

TABLE 9

ANALYSIS OF COVARIANCE FOR THE KAWANDA EXPERIMENT

Source	Degrees of Freedom	Sy^2	Sxy	Sx^2
Interaction	7	1.0882	0.4357	3.0523
Dates (adj.)	9	21.7354	19.8782	36.7967
Localities (adj.)	7	177.6513	29.7493	27.0614
Replications	24	11.6219	−2.5228	5.4713
Error	48	17.3125	−2.3957	14.0422
Error + Dates (adj.)	57	39.0479	17.4825	50.8389
Error + Loc. (adj.)	55	194.9638	27.3536	41.1036

Error	47	16.9038	.3597
Error + Dates (adj.)	56	33.0360	
Dates (adj.)	9	16.1322	1.7925**
Error + Loc. (adj.)	54	176.7605	
Localities (adj.)	7	159.8567	22.8367**

**Significant at the 1% probability level.

From this analysis, we conclude that there was no significant over-all regression of yield on Lygus damage, as shown by the non-significant reduction in the error sum of squares by the use of X-variate (Lygus damage). Hence we conclude that there was no serious insect migration to the last-planted cotton at a given locality, when *only three planting dates* were used at each locality.

The regression coefficient, $b_{Y \cdot X}$, is found from the error row in the above table:

$$(32) \qquad b_{Y \cdot X} = \frac{-2.3957}{14.0422} = -0.1706.$$

Each yield figure could be adjusted for the Lygus damage by use of the equation

$$(33) \qquad Y(\text{adjusted}) = Y + 0.1706(X - \overline{X}).$$

Then the optimum planting date could be determined from these adjusted yields by use of the methods given above.

It would not be useful in general to bother with such a small adjustment, but we will illustrate the method which would be used. In Table 7, the locality yield totals (adjusted for Lygus damage) are

(34) 28.15, 83.82, 63.55, 27.38, 52.60, 46.81, 61.23, 44.07.

Similarly the adjusted values of B and C are -51.46 and 2261.44. Using these adjusted values, we obtain

$$b = -0.5965; \qquad c = -0.05113. \tag{35}$$

The variances of these estimates will be the same as before except that σ^2 is estimated as the average variance of $Y - b_{Y \cdot X}(X - \bar{X})$.[2] This average variance is $0.3597(1 + 1/96) = 0.3634$. There is no significant change in the value of either b or c.

ANALYSIS OF THE KABULA DATA

As stated previously, the same experimental design was used at the Kabula Experiment Station, except that the planting dates ranged from September 13 through January 17. The average yield for this experiment was 4.463 kgms./plot. No data were gathered on the Lygus damage. The analysis of variance for the Kabula data is given in Table 10.

TABLE 10

ANALYSIS OF VARIANCE FOR THE KABULA COTTON YIELD DATA

Source	Degrees of Freedom	Mean Square
Interaction	7	0.7999
Dates (adj.)	9	3.6573**
Localities (adj.)	7	26.5159**
Replications	24	1.6730
Error	48	1.1448

**Significant at the 1% significance level.

The unadjusted mean yields per plot for each planting date at Kabula, the mean yields per plot adjusted for localities by equation (7) and the estimated adjusted mean yields per plot based on the quadratic regression equation (20) are presented in Table 11.

[2]This is only an approximation to the best estimate of σ^2. The correct estimate would involve weighting each observation according to its use in the estimation of b and c.

TABLE 11
MEAN COTTON YIELDS FOR EACH PLANTING DATE AT KABULA

Date	t'	Unadjusted Mean Yield	Mean Yields Adjusted for Localities Equation (7)	Equation (20)	Equation (38)
1 (9-13)	−4.5	6.498	6.808	7.279	6.665
2 (9-27)	−3.5	6.412	6.997	6.479	7.151
3 (10-11)	−2.5	6.018	6.631	5.751	6.525
4 (10-25)	−1.5	4.816	5.661	5.096	5.428
5 (11-8)	−0.5	2.544	3.827	4.514	4.325
6 (11-22)	0.5	2.362	3.774	4.003	3.503
7 (12-6)	1.5	3.466	3.065	3.566	3.077
8 (12-20)	2.5	4.611	2.939	3.200	2.983
9 (1-3)	3.5	6.033	2.996	2.907	2.985
10 (1-17)	4.5	4.286	2.644	2.686	2.666
Average		4.463	4.463	4.463	4.463

Again we note that there is no significant decrease in the adjusted mean yield until we reach the fourth or fifth planting date. Some of the differences between the adjusted mean yields, using equation (7), and their standard errors are given below.

	Date	Mean Difference	Standard Error
	2–4	1.336	.610
(36)	4–5	1.834	.498
	5–10	1.183	.976

It would appear that the optimum planting date is somewhere between September 13 and October 11.

When the quadratic trend was fitted to the mean yields at the successive dates by equation (20), we obtained the following estimates of b and c $\pm$ their standard errors:

$$(37) \qquad \begin{aligned} b &= -0.5103 \pm 0.1338 \\ c &= 0.0362 \pm 0.0290 \end{aligned}$$

We note that only the linear coefficient, b, is significantly different from 0. Also since b and c are of opposite signs, this regression curve will be concave upwards, indicating a minimum rather than a maximum at the point: $t = \bar{t} - b/2c$. Hence no optimum planting date can be estimated by use of the quadratic equation. Extrapolation would be impossible becausé the adjusted means estimated by the regression will increase on both tails when c is positive. The minimum point is not reached until $t = 12.55$; hence, Table 11 does not indicate the increasing yields at the later dates. However, we do see that the regression estimate is beginning to diverge quite noticeably at the first date.

It might be useful to investigate the possibility that equation (20) should be changed to include third and fourth degree terms, as follows:

$$y_{ti} = 4[a + bt' + ct'^2 + dt'^3 + et'^4] + p_{ti} \;. \tag{38}$$

The degree must be even if we are to have the downward trend at both ends. The estimates of b and d (adjusted for localities) are independent of c and e. However b and d are correlated with each other and so are c and e.

The estimating equations for b and d are:

$$\begin{aligned} 64b + 1072d &= -32.658 \\ 1072b + 31060d &= -299.868 \end{aligned} \tag{39}$$

The inverse of the coefficient matrix is

$$\begin{bmatrix} .03703545 & -.00127824 \\ -.00127824 & .0000763126 \end{bmatrix} \tag{40}$$

The estimated values of b and d $\pm$ their standard errors are:

$$\begin{aligned} b &= -0.82622 \pm 0.2059 \\ d &= 0.018862 \pm 0.00935. \end{aligned} \tag{41}$$

Hence we conclude that both b and d are significantly different from 0.

Similarly for c and e, the estimating equations are:

$$\begin{aligned} (4{,}096c + 84{,}736e)/3 &= 49.441 \\ (84{,}736c + 2{,}038{,}528e)/3 &= 326.086. \end{aligned} \tag{42}$$

The inverse of the (c, e) coefficient matrix is:

$$\begin{bmatrix} .00522869 & -.000217342 \\ -.000217342 & .0000105060 \end{bmatrix} \tag{43}$$

Hence the estimated values of c and e with their standard errors are:

$$\begin{aligned} c &= .18764 \pm .07739 \\ e &= -.0073197 \pm .003467 \end{aligned} \tag{44}$$

The adjusted mean yields estimated by the quartic regression curve (38) are also given in Table 11. The agreement between these estimated mean yields and those estimated from equation (7) is quite good. The estimated mean yields (by quartic regression) are beginning to drop off at both ends, and there are no pronounced series of plus or minus deviations as with the estimated mean yields found by using only the quadratic regression. The remaining date sum of squares after fitting equation (38) is 4.6986 with 5 degrees of freedom, giving a mean square of 0.9397, which is smaller than the error mean square.

The problem of estimating the optimum planting date is complicated by the fact that there are two maximum points and one minimum point. These points are solutions of the equation

$$b + 2ct' + 3dt'^2 + 4et'^3 = 0. \tag{45}$$

The maximum point with the largest average adjusted yield is at $t' = -3.665$, which is slightly before the second date (September 27). No attempt has been made to estimate the standard error of this estimated optimum planting date.

RELATIVE EFFICIENCY OF THIS EXPERIMENTAL DESIGN

If there were no adverse effects from the Lygus migration to the newly planted cotton, it would be advisable to use as many planting dates as possible at each locality. The analysis of the Kawanda data seems to indicate that more than 3 successive planting dates might have been used at a given locality without incurring any serious insect interference. We shall consider the relative efficiency of the following field designs:

(i) The present "staggered" 3-date plan.

(ii) A staggered 4-date plan, with only 6 locations as follows:

Location	Dates	Location	Dates
1	1,2,3,4	4	5,6,7,8
2	2,3,4,5	5	6,7,8,9
3	3,4,5,6	6	7,8,9,10

(Notice that the sequence (4, 5, 6, 7) is omitted in order to have a design which would have the same number of replications per date as for the 3-date plan.)

(iii) A balanced incomplete blocks design with 3 planting dates per block, which requires 30 blocks with each planting date being replicated 9 times. (The same number of plots could not be planted at each location if this design were used.)

(iv) A complete blocks design with all 10 planting dates at each locality, the locality being a complete block.

Under designs (i) and (ii), 96 plots would be used, but only 90 plots would be used for (iii) and only 80 plots for (iv). Hence the expected error variances will have to be adjusted to an equal number of plots. The most efficient of these designs would be the complete blocks design (iv), because no adjustments for localities would be required in determining the average yield for a given planting date. The balanced incomplete design (iii) has been shown to be 74% as efficient as the complete blocks design (iv); that is, 4 replications of each planting date using (iii) would be required to give the same accuracy as 3 replications using (iv).

The relative efficiency of design (i) will also be assessed with respect to (iv) by comparing the average variances of the difference between the mean yields at successive planting dates. If the error variance per plot is designated as σ^2, then the variance of the difference between *any* two date means for the complete blocks design would be $2\sigma^2/8 = .25\ \sigma^2$. This variance can be compared with any of those in Table 6 by adjusting for the different number of total plots used in the two designs. Since 96 plots were used in the "staggered" 3-date plan (i), the variances in Table 6 should be multiplied by $96/80 = 1.2$ to put them on the same basis as that of the complete blocks design.

The lowest variance in Table 6 is that for the difference in yield between dates 5 and 6, $0.21652\ \sigma^2$. For comparison purposes, we multiply this variance by 1.2, giving $.26\ \sigma^2$. Hence this comparison is 96% efficient as compared with the same comparison if design (iv) were used. If we omit the comparisons for the first two and the last two planting dates, the average of the efficiencies for planting dates separated by one time interval (a fortnight) would be 94%. The comparisons for the first two and the last two dates are quite inefficient, because these planting dates are not used at many locations under design (i); the relative efficiency is only 48%. The average efficiency for all nine of the one fortnight comparisons is 84%. The average efficiencies for all the comparisons of average yields given in Table 6 are presented in Table 12 below. It should be noted that these efficiencies are based on the premise that σ^2 is the same for all designs. If the increase of block size to 10 plots per block in design (iv) results in an increase in σ^2, the relative efficiencies would be higher than those given in Table 12. The incomplete blocks design (iii) would have the same block size; hence,

σ^2 for this design would be expected to be the same as for the 3-date plan (i). The relative efficiency of (i) as compared to (iii) would be $1/.74 = 1.35$ times as great as the efficiencies given in Table 12.

As stated above, we also consider the efficiency of the 4-date plan (ii), assuming that σ^2 would be the same for this design as for (i), even though only 6 instead of 8 localities are considered. The least square equation for (ii) was set up and a variance table such as Table 6 was then constructed. The relative efficiencies of this design (ii) compared with the complete blocks design (iv) are also presented in Table 12. It might be mentioned that a balanced incomplete blocks design with 4 planting dates per block can be constructed, using 15 blocks. This balanced incomplete blocks design is 83% as efficient as the complete blocks design, assuming σ^2 does not increase with the increase of block size for the complete blocks design.

TABLE 12

AVERAGE PERCENTAGE EFFICIENCIES OF THE 3 AND 4-DATE PLANS (i and ii) COMPARED WITH A COMPLETE BLOCKS DESIGN (iv)

Number of Fortnights Separating Planting Dates

Design	1	2	3	4	5	6	7	8	9	Avg.
i*	94	65	46	36	30	25	22			57
ii†	106	89	68	52	44	41	34			74
i†	84	61	43	34	28	23	20	17	14	47
ii†	93	80	63	48	41	35	31	26	21	60

*Omitting the first and last planting dates.
†Using all planting dates.

From Table 12, we see that the 3-date plan (design i) is only about 3/4 as efficient as the 4-date plan (design ii). That is we could have obtained about the same variances of the adjusted mean differences with 4 replications at each of the 6 locations for (ii) as with the 4 replications at each of the 8 locations for (i).

It might be noted that in this particular experiment, the most important comparison for the determination of the range of the optimum planting dates was that between dates 1 and 5. The efficiency of the estimate of the difference in the mean yield between these two dates is only 30% for the 3-date plan (i). Actually since the optimum planting date could have come before the first planting date used in the experiments, this range should have been about twice as large; the efficiency of this estimate would be even lower than 30%. Hence we can conclude that the 3-date design is a very inefficient design for determining the optimum planting date.

The efficiency of the determination of the optimum planting date can also be estimated from the variances of b and c in the quadratic regression equation (20). The variances of b and c for the 3-date plan (using 96 plots) are given in equation (24).

The variances of b and c for the four designs (i, ii, iii and iv), based on 80 plots, are the following coefficients of σ^2:

		(*i*)	(*ii*)	(*iii*)	(*iv*)
(46)	$\sigma^2(b)$	.01875	.01000	.00214	.00152
	$\sigma^2(c)$	.000882	.000514	.000320	.000237

These results show that the 3-date plan (i) is only about 8% efficient in the estimation of b and 27% efficient in the estimation of c as compared to the complete blocks design, if we can assume that σ^2 will not increase with the latter. When compared with the incomplete blocks design, the efficiency is about 1.5 times as great. Comparing the 3 and 4-date plans, we note that the efficiency is almost doubled for the 4-date plan.

From equation (30), we see that the variance of the optimum planting date (if the regression is quadratic) is

$$\sigma^2(t_{max}) = \frac{1}{4c^2}\left[\sigma^2(b) + \frac{b^2}{c^2}\,\sigma^2(c)\right], \tag{47}$$

where $\sigma^2(b)$ and $\sigma^2(c)$ are found by multiplying the constants given in (46) by σ^2. In order to increase the efficiency of the experiment, we must either reduce σ^2, redesign the experiment so as to obtain lower coefficients for $\sigma^2(b)$ and $\sigma^2(c)$, or plan the experiment so as to have the optimum planting date fall near the middle of the dates used in the experiment (make b small).

If we do set up the experiment so that b is expected to be small, then the efficiency of the experiment largely depends on $\sigma^2(b)$. From (46), we see that $\sigma^2(b)$ is very large for the "staggered" designs as compared to the randomized blocks designs (iii and iv). It has been suggested that we might increase the efficiency of the "staggered" 3-plan design by redesigning the experiment. For example we might use only the following sequences of dates—1, 2, 3; 2, 3, 4; 7, 8, 9; 8, 9, 10—planting each sequence at two locations. Although $\sigma^2(c)$ is reduced from $3\sigma^2/4096$ to $3\sigma^2/7168$, $\sigma^2(b)$ remains at $\sigma^2/64$ (for 96 total plots). This design is not recommended for two reasons: (i) It does not improve the efficiency of the estimation of b, which is the chief contribution to the inefficiency of the "staggered" designs, (ii) There should be some estimate of the yield for all planting dates, especially since the hypothesis of a quadratic regression of yield on planting date may be false. If some other regression equation is used, the reduction in the variance of c may be offset by an increase in the variance of some other estimate.

If we were to use the sequences—1, 2, 3; 3, 4, 5; 5, 6, 7; 7, 8, 9—all planting dates, except the tenth, would be represented. However $\sigma^2(b)$ would still be $\sigma^2/64$ and $\sigma^2(c)$ would be increased to $3\sigma^2/3904$.

SUMMARY AND CONCLUSION

This paper presents a new type of experimental design, the "staggered" design, for use with experimental material which can have but few consecutive plantings at a given locality. Two experiments involving the determination of the optimum planting date of cotton have been conducted in British East Africa. The "staggered" design was used here because of a fear that an insect, *Lygus simonyi*, would tend to transfer from earlier planted plots to newly planted ones, hence distorting a proper assessment of the relationship between cotton yield and planting date. At one of the experiments, an index of the Lygus damage was determined for each plot. If this index truly reflected the Lygus damage, it appeared that there was no important migration for planting dates separated by four weeks or less.

We have shown that the 3-date "staggered" design is decidedly inefficient in estimating the optimum planting date. The adjustments for locality effects are so great that there tends to be a very long range of indeterminacy of the optimum planting date. The following suggestions are offered to improve the efficiency of the estimation of the optimum planting date under conditions of uncertainty with respect to insect migration:

(i) Test the adequacy of the Lygus damage index as a true indicator of the infestation by the Lygus bug. If this index is reliable, or can be made reliable, then it appears that more planting dates should be used at each locality and that adjustments should be made for the Lygus damage by covariance techniques rather than by use of the "staggered" design.

(ii) Even if the index is shown to be inadequate, especially when more than 3 planting dates are used at each locality, we would advise at least trying 4 successive planting dates at each locality.

(iii) A great improvement would result if the localities did not differ so widely in their fertility. This "staggered" design would be much more efficient if the locality differences were not so pronounced as in these African experiments.

(iv) It is advisable to plan the experiment so that the optimum planting date is near the middle date used in the experiment.

(v) Ordinarily the experimenter hopes to secure some information on the correct allocation of experimental material as to changes in the number of locations and replications within locations. In this case the small date-location interaction as compared to the date-replication interaction leads us to infer that one could not lose any information by using fewer localities and more replications at each locality. Since this result was obtained in both experiments, we feel that the use of more planting dates at a given locality with fewer localities being used probably would not materially alter the error variance.