A SHORTER METHOD FOR EVALUATING THE ABILITY OF SELECTIONS TO YIELD CONSISTENTLY OVER LOCATIONS[1]

R. L. PLAISTED[2]

Once the potato breeder has obtained his true seed, he must satisfy two requirements for success. First, he must maintain disease free planting stock and second he must carry out his selection in an efficient and effective manner. In the more difficult final stages of selection the breeder tests his selections in regional yield trials. These not only give information about yield in specific locations and seasons but can also be used as a measure of the consistency of their cropping ability. A method was presented in 1959 (2) for making numerical estimates of this consistency of cropping ability using the regional yield trial data already being assembled. An alternate method is proposed which produces the same results with less computational effort.

METHOD

For illustrative purposes let us assume that the regional yield trial consists of 2 locations each with the same 4 varieties in 2 replications planted in a randomized complete block design. Most breeders substitute new entries into their regional trials each year. The analysis of variance model for this type of trial is

$$Y_{ijkl} = U + (VY + V)_i + L_j + (VLY + VL)_{ij} + R_{jk} + E_{ijkl}$$

where Y_{ijkl} is the yield on the $_{ijkl}$th plot, U is the mean effect common to all observations, $(VY + V)_i$ is the confounded effect of the $_i$th variety and its interaction with the particular year in which the trial is conducted, L_j is the effect of the $_j$th location, $(VLY + VL)_{ij}$ is the confounded effect of the interaction of the $_i$th variety and the $_j$th location in the single year the trial is conducted, R_{jk} is the effect of the $_k$th replicate at the $_j$th location, and E_{ijkl} is experimental error variation associated with the $_{ijkl}$th plot.

In practice, the source $(VY + V)_i$ probably should be considered fixed in that the population of varieties of interest are the ones contained in the trial. The other sources would probably be considered random since they represent a sample of a larger range that is of interest, thus they would be random, independent variates distributed around zero. To facilitate understanding the method of analysis we shall impose the restriction on all sources that the sum of the effects in a given source equal zero. This is the type of restriction given to fixed effects.

[1]Accepted for publication October 26, 1959.
[2]Plant Breeding Department, Cornell University, Ithaca, N. Y., Paper No. 382.

Let us assign the following effects to the model:

Mean effect	Location effects	Variety × Location interaction effects	
$U = 8$	$L_1 = +1$	$(VLY + VL)_{A1} = +1$	$(VLY + VL)_{A2} = -1$
	$L_2 = -1$	$(VLY + VL)_{B1} = +2$	$(VLY + VL)_{B2} = -2$
		$(VLY + VL)_{C1} = 0$	$(VLY + VL)_{C2} = 0$
		$(VLY + VL)_{D1} = -3$	$(VLY + VL)_{D2} = +3$

Replicate effects	Variety effects
$R_{11} = +1$	$(VY + V)_A = -1$
$R_{12} = -1$	$(VY + V)_B = -2$
$R_{21} = +1$	$(VY + V)_C = 0$
$R_{22} = -1$	$(VY + V)_D = +3$

Experimental error effects

$e_{A11} = -3$	$e_{A12} = +3$	$e_{A21} = +2$	$e_{A22} = -2$
$e_{B11} = +2$	$e_{B12} = -2$	$e_{B21} = +1$	$e_{B22} = -1$
$e_{C11} = +1$	$e_{C12} = -1$	$e_{C21} = -3$	$e_{C22} = +3$
$e_{D11} = 0$	$e_{D12} = 0$	$e_{D21} = 0$	$e_{D22} = 0$

When these values are substituted into the model, the results given in table 1 are obtained.

TABLE 1.—*Data.*

	Location 1			Location 2			
Var.	Rep. 1	Rep.2	Sum.	Rep. 1	Rep. 2	Sum.	Var. Total
A	7	11	18	8	2	10	28
B	12	6	18	5	1	6	24
C	11	7	18	5	9	14	32
D	10	8	18	14	12	26	44
Totals	40	32	72	32	24	56	128

The first step in the analysis is to compute a separate analysis of variance (AOV) for each location. In this case the results are given in table 2.

The second step is to combine these separate AOV over all locations. Table 3 shows how this may be accomplished. As a precaution, it should be noted that the interpretation of the results of the combined analysis is more meaningful if the error variances of the separate AOV are random variables of the same population.

To show that the sum of squares obtained in table 3 are clearly a function of the effects from which the data were derived let us derive the variety × location sum of squares. If all other effects are ignored, the variety × location effects were assigned as follows:

TABLE 2.—*Separate AOV for each location.*

	Location 1		Location 2	
	df	SS	df	SS
Total	7	36	7	148
Reps	1	8	1	8
Varieties	3	0	3	112
Error	3	28	3	28
Computational data				
Correction factor		648		392
Uncorrected total SS		684		540

TABLE 3.—*Separate AOV combined over locations.*

	df	SS	
Total	15	200	684 + 540 — 1024
Location	1	16	648 + 392 — 1024
Reps in location	2	16	8 + 8
Var. in location	6	112	0 + 112
Var	3	56	$(28^2 + 24^2 + 32^2 + 44^2)/_4 - 1024$
Var × loc	3	56	112 — 56
Error	6	56	28 + 28 = 200 — 16 — 16 — 112

Correction factor = $128^2/_{16} = 1024$

d.f.:

Total	= total number of plots — 1
Location	= number of locations — 1
Reps in loc	= sum of d.f. for reps in all the separate AOV
Var in loc	= sum of d.f. for varieties in all the separate AOV
Var	= number of varieties — 1
Var × loc	= d.f. for var × d.f. for location
Error	= sum of d.f. for error in all the separate AOV.

SS:

Total SS	= sum of all uncorrected total SS from individual analyses — overall correction factor.
Location	= sum of all correction factors from separate AOV — new overall correction factor
Reps in loc	= sum of all SS for reps in separate AOV
Var in loc	= sum of all SS for var in separate AOV
Var	= sum of squares of variety totals divided by total number of observations in each total — correction factor
Var × loc	= variation in location SS — var SS
Error	= by subtraction or by adding error SS from individual AOV.

If the F for the variety × location Mean Square divided by the error M.S. is significant (in this example, it is not), then proceed with the next step.

Construct a table similar to table 4. This table simplifies subsequent computations.

	Location 1		Location 2	
Var	*Rep 1*	*Rep 2*	*Rep 1*	*Rep 2*
A	+1	+1	—1	—1
B	+2	+2	—2	—2
C	0	0	0	0
D	—3	—3	+3	+3
	0	0	0	0

Note that $4\,[(\pm1)^2 + (\pm2)^2 + (\pm0)^2 + (\pm3)^2] = 56$

TABLE 4.—*Sums of squares accumulated by variety.*

Var	Contribution to Uncorrected Total SS	Contribution to Uncorrected Var × Loc SS
A	$7^2 + 11^2 + 8^2 + 2^2 =$ 238	$(18^2 + 10^2)/_2 = 212$
B	$12^2 + 6^2 + 5^2 + 1^2 =$ 206	$(18^2 + 16^2)/_2 = 180$
C	$11^2 + 7^2 + 5^2 + 9^2 =$ 276	$(18^2 + 14^2)/_2 = 260$
D	$10^2 + 8^2 + 14^2 + 12^2 =$ 504	$(18^2 + 26^2)/_2 = 500$
	1224	1152

Next compute a combined AOV, each time omitting a different variety. In our example, let us omit variety A. A new set of totals are needed. They are as follows:

	Location 1			Location 2			Grand
	Rep 1	Rep 2	Total	Rep 1	Rep 2	Total	Total
Varieties (B + C + D)	33	21	54	24	22	46	100

Each of these combined AOV omitting in turn a successive variety may be computed as outlined for the combination of all varieties. However, the method outlined in table 5 is somewhat shorter.

TABLE 5.—*Combined AOV omitting variety A.*

	df	SS	Mean Square	Expected Mean Square
Total	11	152.7		
Location	1	5.4		
Reps in loc	2	24.6		
Var in loc	4			
Var	2	50.7		
Var × loc	2	50.6	25.3	$\sigma^2 + k(\sigma^2_{VLY} + \sigma^2_{VL})$
Error	4	21.4	5.35	σ^2

k = number of reps per loc.

c.f. $= 100^2/12 = 833.3$

Total SS = total uncorrected SS from combined AOV of all varieties — contribution of the omitted var. — this new c.f.
$= 1224 - 238 - 833.3 = 152.7$

Loc SS = sum of each location total squared divided by number of observations in each total — c.f.
$54^2 + 46^2/6 - 833.3 = 5.4$

Reps in loc SS = sum of each rep total squared divided by the number of varieties — correction factor — Loc SS

$= (33^2 + 21^2 + 24^2 + 22^2)/3 - 833.3 - 5.4 = 24.6$

Var SS = sum of each variety total squared divided by number of observations in each total — c.f.

$= (24^2 + 32^2 + 44^2)/4 - 833.3 = 50.7$

$(V \times L)$SS = (total uncorrected $(V \times L)$SS — contribution of the omitted variety) — Var SS — Loc SS — c.f.

Error SS = by subtraction

152.7 — 5.4 — 24.6 — 50.7 — 50.6

Again let us use the variety $\times$ location sum of squares to illustrate how the assigned effects have produced the SS given in the analysis. If all other effects are ignored and variety A is omitted, the following variety $\times$ location effects were assigned to the data analysed in table 5:

	Location 1		Location 2	
Var	*Rep 1*	*Rep 2*	*Rep 1*	*Rep 2*
B	+2	+2	—2	—2
C	0	0	0	0
D	—3	—3	+3	+3
	—1	—1	+1	+1

Note that the sum of the effects within replications within locations is not zero. Therefore the effect of locations in this analysis has been changed by the average of this amount, $\pm 1/3$. Let us remove this effect from the variety $\times$ location effects:

	Location 1		Location 2	
Var	*Rep 1*	*Rep 2*	*Rep 1*	*Rep 2*
B	$+2 + 1/3 = 2\,1/3$	$2\,1/3$	$-2 - 1/3 = -2\,1/3$	$-2\,1/3$
C	$0 + 1/3 = 1/3$	$1/3$	$0 - 1/3 = -1/3$	$-1/3$
D	$-3 + 1/3 = -2\,2/3$	$-2\,2/3$	$3 - 1/3 = 2\,2/3$	$2\,2/3$
	0	0	0	0

Obtain the sum of squares of these effects.

$$4[(\pm 2\,1/3)^2 + (\pm 1/3)^2 + (\pm 2\,2/3)^2] = 50.7$$

The difference between this and the 50.6 of Table 5 is due to rounding.

Table 5 also gives the expected mean squares for the variety $\times$ location and error sources of variation. These are derived from the model and will be characteristic for analyses of regional yield trial data repeated over locations, but not seasons. If the number of replications at each location is not the same the divisors in computing the sums of squares will be effected and the coefficient k of the interaction component will be an average value which can be calculated according to the formula:

$$\frac{1}{v-1}\left(\Sigma r_i - \frac{\Sigma r_i^2}{\Sigma r_i}\right)$$

The use of this formula is given by Federer (1) on page 105.

The next step is to compute the estimate of the variety $\times$ location component of variance ($\sigma^2_{VLY} + \sigma_{VL}^2$). This is done by subtracting the error mean square from the variety $\times$ location mean square and dividing the coefficient k. For example:

$$\widehat{(\sigma^2_{VLY} + \sigma_{VL}^2)}_A = (25.3 - 5.35)/_2 = 10.0$$

Similarly the estimates when the other varieties are omitted can be obtained. These are given in table 6.

TABLE 6.—*Estimates of the remainder* ($\sigma^2_{VLY} + \sigma_{VL}^2$) *when successive varieties are omitted from the analysis.*

Variety Omitted	Contribution to the V $\times$ L interaction	$\widehat{(\sigma^2_{VLY} + \sigma^2_{VL})}$
A	$\pm$ 1	10.0
B	$\pm$ 2	3.3
C	0	10.3
D	$\pm$ 3	$-$ 5.0

DISCUSSION

The analysis of this small example shows how the final results recover the information on interaction put into the model. The larger the contribution of a variety to the variety $\times$ location interaction, the smaller will be the estimate of the remainder interaction component of variance. Therefore in the analysis of real data with estimates of the remainder interaction component of variance ranked from high to low, the dependability of the variety will also be ranked from high to low.

This method gives results identical with those obtained with the method outlined in the first paper. Whereas the latter method of approach is not as logically straightforward, it involves less computational time and it is possible to apply a test of significance to the estimates obtained.

When this method was applied to the 4 years' data described in 1959, the results were identical with those described for those analyses. Of the three varieties repeated in all the years, Green Mountain was the most variable of all the varieties except in one year, Cobbler was very low in its variability except in one year, and Katahdin was also low in its variability. These results are in agreement with past experience in this State with these varieties. They also point out the need for estimates based on more than one year's data due to the magnitude of the interaction of varieties $\times$ locations $\times$ years. It is not uncommon that this second order interaction is larger than either of the first order interactions, varieties × locations and varieties × years. However, this precaution imposes no real burden on the potato breeder since he would certainly include a selection in more than one or two years' regional yield trials before considering it for release.

Literature Cited

1. Federer, W. T. 1955. Experimental design. The Macmillan Co., New York.
2. Plaisted, R. L. and L. C. Peterson. 1959. A technique for evaluating the ability of selections to yield consistently in different locations and seasons. Am. Potato J. 36: 381-385.