

L.). Arboretum Kórnickie 23: 161—169 (1978). — JACKSON, D. I. and SWEET, G. B.: Flower initiation in temperate woody plants. Hort. Abstr. 42: 9—22 (1972). — LONGMAN, K. A. and WAREING, P. F.: Effect of gravity on flowering and shoot growth in Japanese larch (*Larix leptolepis* MURRAY). Nature 182: 379—381 (1958). — LONGMAN, K. A., NASR, T. A. A. and WAREING, P. F.: Gravimorphism in Trees. 4. The Effect of Gravity on Flowering. Annals of Botany, N.S. 29: 459—473 (1965). — ROMBERGER, J. A. and GREGORY, R. A.: Analytical mor-

phogenesis and the physiology of flowering in trees. Proceedings of the Third American Forest Biology Workshop. 9—42 September 1974, Colorado State University: 132—147 (1974). — SMITH, W. H. and STANLEY, R. G.: Cone distribution in crown of Slash pine (*Pinus elliottii* ENGELM. in relation to stem, crown and wood increment. Silvae Genetica 18: 86—91 (1969). — WAREING, P. F.: Reproductive development in *Pinus sylvestris*. In: The Physiology of Forest Trees, Ed. K. V. TAIMANN. Ronald Press, N.Y.: 643—654 (1958).

## Improved Estimators for Provenance Breeding Values

By F. H. KUNG<sup>1)</sup>

(Received September 1978 / March 1979)

### Summary

The provenance mean of a trait is a sample estimate for the breeding value of that provenance. However, if there are more than three provenances under test, a shrinking factor should be applied to each sample mean to obtain the breeding value. The best shrinking factor to be used with the prediction depends on the F-statistics and the number of provenances in the plantation. Advantages of such adjustment by a shrinking factor are illustrated by a computer simulated example.

Key words: Provenance test, selection differential, genetic gain, STEIN'S estimator.

### Zusammenfassung

Der Mittelwert eines gegebenen Pflanzenmerkmals für ein bestimmtes Ursprungsgebiet ist ein Proben-Schätzwert für den Zuchtwert dieses Gebietes. Werden mehr als drei Gebiete getestet, so sollte ein Schrumpffaktor auf jeden Proben-Mittelwert angewandt werden, um den Zuchtwert zu erhalten. Der beste Schrumpffaktor für eine Voraussage hängt von der F-Statistik und der Anzahl der Ursprungsgebiete ab, welche bei der Anpflanzung in einer Plantage herangezogen werden. Die Vorteile einer derartigen Anpassung durch einen Schrumpffaktor werden anhand eines computer-simulierten Beispiels erläutert.

### Introduction

Provenance testing is useful in screening the naturally available genetic variation and in providing the best available types for reforestation or future breeding work. It is particularly necessary to do provenance testing prior to more intensive breeding work when dealing with an exotic (WRIGHT, 1975). Besides providing basic genetic information, selection of superior geographical sources can be done and the test plantation may be converted to a seedling seed orchard (NANSON, 1972).

When selection is made based on provenance test data, should we use the provenance mean as the breeding value for that provenance? In comparing provenance performances, should we consider the superiority as genetic gain or merely as selection differential? For example, if provenance A averaged 50 m and the plantation mean was 40 m, can we expect that the seed collected from provenance A will be 50 m, or 10 m better than the population mean?

<sup>1)</sup> Associate Professor, Dept. of Forestry, Southern Illinois University, Carbondale, Illinois, U.S.A. 62901.

Some people think so. The argument is based on the fact that the genetic difference, not the environmental difference, among provenances, can be transmitted from the native range to the test site. Thus, the provenance mean is a sample estimate of the breeding value of that provenance. The superiority is the genetic gain.

Statistically speaking, the sample mean is the best predictor for a population mean under random sampling. It has minimum variance among all unbiased estimators and likewise among all translation invariant estimators. Moreover, it is minimax (minimizes the maximum expected squared error), admissible (no competing estimation rule has smaller squared error for all values of the true mean), and it is the maximum likelihood estimator (choosing the population mean equal to sample mean maximizes, among all values of true mean, the probability of obtaining the value of sample mean actually observed). However, statisticians lately have begun to realize that biased estimation rules have definite advantages over the usual unbiased estimators (EFRON, 1975). If there are more than three population means to be predicted, sample means are not the best predictors (STEIN, 1955). The best one involved a shrinking factor for each mean (JAMES and STEIN, 1961).

Provenance testing is not without experimental error. The experimental error pushes the sample mean above or below its true breeding value. If we have many provenances being tested, there are more provenances near the plantation average than near the extreme. In other words, the distribution of provenance breeding value will most likely follow a normal distribution rather than a uniform distribution. So, there are more central genotypes being pushed outward from the center. The outcome is a flatter distribution of phenotypic means. In order to reconstruct the breeding value from phenotypic provenance means, the provenance means need to be pushed (shrunk) back toward the center to resemble the original distribution of provenance breeding values.

### The Shrinking Factor

The breeding value of a provenance Z can be expressed as:

$$Z = \bar{Y} + C(Y - \bar{Y}) \quad (1)$$

where Y = Plantation mean, or the average of all provenance means.

Y = Provenance mean observed in the plantation.

C = Shrinking factor.

The above in the same as:

$$Z = Y - (1 - C)(Y - \bar{Y}) \quad (2)$$

As shown in (2) each provenance mean  $Y$  is shrunk by the amount  $(1 - C)(Y - \bar{Y})$ . The larger the  $Y$  is, the greater is the reduction.

The shrinking factor given by JAMES and STEIN (1961) is:

$$C = 1 - (k - 3) D/S \quad (3)$$

where  $k$  is the number of provenances.

$D$  is estimated variance of the theoretical normal distribution of provenance breeding values. It is calculated from error variance divided by number of observations within provenance.

$S$  is the sum of (Provenance mean — Plantation mean)<sup>2</sup>.

The shrinking factor given by BOX and TIAO (1973) is:

$$C = 1 - (1/F) \quad (4)$$

where  $F$  is the observed variance ratio, MSQ among provenances divided by MSQ within provenance in an analysis of variance table.

The shrinkage factor given by VINOD (1976) is:

$$C = 1 - (1 - R^2) / \{(n - p - 1) R^2\} \quad (5)$$

where  $R^2$  is the degree of determination of provenance means or the regression SSQ due to means divided by total SSQ,  $n$  is total number of observations and  $p$  is the number of provenances.

The comparison of the above three shrinking factors is facilitated if we assume equal variance within provenance and rewrite them as follows:

JAMES and STEIN  $C = 1 - \{(k - 3) / (k - 1)\} (1/F)$

BOX and TIAO  $C = 1 - (1/F)$

VINOD  $C = 1 - (1/k) (1/F)$

Thus, for a given set of provenance test data, the BOX and TIAO formula gives the smallest breeding value and VINOD gives the largest. The JAMES and STEIN estimator is close to that of BOX and TIAO, especially when  $k$  is large.

### Which Shrinking Factor is the Best

No uniformly best estimator has yet been found (EFRON and MORRIS, 1977). Therefore, the best shrinking factor depends on the noise to signal ratio (error variance component/provenance variance component) and the experimental design (no. of provenances and no. of observations) (VINOD, 1976).

When the  $F$  ratio is less than 1, we accept the hypothesis of no provenance effect. This is equivalent to using a shrinking factor of 0.0 to set all provenance breeding values back to the plantation mean. In that case, no selection should be made. When there is a significant  $F$  ratio, then we have to look at the number of provenances involved before we estimate provenance breeding values. If there are only two or three provenances, we should use the mean directly as the breeding value for that provenance. This is equivalent to using a shrinking factor of 1.0. When there were three or four provenances, the shrinking factor by VINOD was superior to others (VINOD, 1976). Beyond that, the JAMES — STEIN estimator should be used.

The JAMES — STEIN estimator is especially suitable for plantations with heteroscedasticity (unequal variance within provenance). Each provenance may have a different shrinking factor. A provenance with larger within variance will shrink more than a provenance with smaller within variance (EFRON and MORRIS, 1977). Therefore, unlike other shrinking factors, the ranking of provenance may change.

With equal variance and large data base, the BOX — TIAO shrinking factor is suitable. It also gives the same value as provenance heritability (KUNG and BEY, 1977). Therefore, the average breeding value of a group of provenances is the genetic gain for that group.

To summarize, the optimal shrinking factor may be set as follows:

Provenance Test Result		Shrinking Factor to be Chosen
$F \leq 1.0$		$C = 0.0$
$1.0 < F,$	$k < 3$	$C = 1.0$
$1.0 < F,$	$k = 3, 4$	VINOD $C$ in (5)
$1.0 < F,$	$4 < k$	Heteroscedasticity, JAMES — STEIN $C$ in (3)
		Homoscedasticity, BOX — TIAO $C$ in (4)

### Computer Simulation

Because the true breeding value for each provenance is unknown, it would be impossible to illustrate the advantages of using a shrinking factor. Therefore, computer simulation is used here to substitute the real plantation data for demonstration.

The model used in assembling data points is as follows:

$$X_{ijkl} = U + P_i + G_j + B_k + E_l$$

where  $U$  is the population mean, for simplicity it is set to zero.

$P_i$  is the breeding value for  $i$ th provenance,  $P_i$  was drawn from a normal distribution  $N(0,5)$ .

$G_j$  is the genetic deviation from provenance breeding value, and was drawn from normal distribution  $N(0,15)$ .

$B_k$  is the  $k$ th block effect and was drawn from the normal distribution  $N(0,5)$ .

$E_l$  is the environmental deviation from the block and was drawn from the normal distribution  $N(0,75)$ .

$X_{ijkl}$  is the  $j$ th seed from the  $i$  provenance planted in the  $l$ th location of the  $k$ th block.

Table 1. — Simulated seedling heights for a randomized complete block provenance test plantation.

Prov. no.	Block no.	Prov. effect	Gene dev.	Block effect	Environ. dev.	Simulated height
		(1)	(2)	(3)	(4)	(1+2+3+4)
1	1	4.19	-7.73	1.48	-16.08	-18.14
1	1	4.19	-2.14	1.48	- 8.26	- 4.73
1	1	4.19	-0.50	1.48	9.26	14.43
1	1	4.19	-1.61	1.48	-10.28	- 6.22
1	2	4.19	-0.79	0.25	- 2.54	1.11
1	2	4.19	2.82	0.25	0.72	7.98
1	2	4.19	0.74	0.25	- 2.45	2.73
1	2	4.19	-1.05	0.25	11.11	14.50
1	3	4.19	1.35	-2.82	6.49	9.21
1	3	4.19	-0.17	-2.82	- 0.15	1.05
1	3	4.19	3.53	-2.82	10.27	15.17
1	3	4.19	-2.56	-2.82	5.67	4.48
1	4	4.19	2.42	-1.81	5.20	10.00
1	4	4.19	6.70	-1.81	4.58	13.66
1	4	4.19	-3.97	-1.81	5.72	4.13
1	4	4.19	-4.39	-1.81	- 8.94	-10.95
1	5	4.19	1.77	-3.12	7.97	10.81
1	5	4.19	-4.72	-3.12	5.60	1.95
1	5	4.19	-0.47	-3.12	-12.55	-11.95
1	5	4.19	-2.31	-3.12	6.52	3.28
1	6	4.19	-5.66	5.79	5.35	9.67
1	6	4.19	5.41	5.79	8.31	23.70
1	6	4.19	-3.08	5.79	-11.26	- 4.36
1	6	4.19	-0.46	5.79	3.04	12.56
2	1	-0.97	1.19	1.48	- 1.86	- 0.16
:	1	-0.97	-1.19	1.48	9.27	8.59
:	1	-0.97	3.29	1.48	7.60	11.40
:	1	-0.97	0.97	1.48	4.07	5.55
:	2	-0.97	1.36	0.25	- 4.71	- 4.07
:	:	:	:	:	:	:
:	:	:	:	:	:	:
25	6	2.81	-1.56	5.79	- 9.82	- 2.78

Table 2. — Comparison between two estimators for the true provenance mean. The symbol \* in the table indicates that the sample (test) mean is closer to the true mean than the Box & Tiao estimator.

Provenance	True Mean	Test Mean	Box & Tiao estimator	Deviation between	
				Test - True	Box - True
1	4.19	4.42	2.84	0.23*	-1.35
2	-0.97	-0.56	-0.64	0.41	0.33
3	-0.59	-0.97	-0.93	-0.38	-0.34
4	-1.58	2.86	1.75	4.44	3.33
5	-1.18	0.34	-0.01	1.52	1.17
6	1.15	1.18	0.57	0.03*	-0.58
7	-0.61	-2.21	-1.80	-1.60	-1.19
8	1.56	-1.21	-1.10	-2.77	-2.66
9	-4.63	-5.83	-4.33	-1.20	0.30
10	-1.09	-0.62	-0.69	0.47	0.40
11	0.66	2.23	1.31	1.57	0.65
12	-2.08	-1.39	-1.23	0.69*	0.85
13	-0.14	-1.81	-1.52	-1.67	-1.38
14	-1.03	-0.62	-0.69	0.41	0.34
15	-0.34	1.95	1.11	2.29	1.45
16	5.78	5.12	3.33	-0.66*	-2.45
17	-1.60	-4.87	-3.66	-3.27	-2.06
18	-1.29	0.37	0.01	1.66	1.30
19	-0.66	-2.29	-1.86	-1.63	-1.20
20	-2.65	-7.17	-5.27	-4.52	-2.62
21	2.54	3.33	2.08	0.79	-0.46
22	-3.73	-3.89	-2.98	-0.16*	0.75
23	-3.76	-5.86	-4.36	-2.10	-0.60
24	-4.64	-7.12	-5.24	-2.48	-0.60
25	2.81	3.59	2.26	0.78	-0.55
Average	-0.55	-0.84	-0.84	-0.29	-0.29
Average absolute deviation				1.51	1.16
Sum of square deviation				94.22	50.87

The simulated plantation involved 25 provenances, 6 blocks, 4 trees per plot and a total of 600 trees. A portion of the plantation data are illustrated in Table 1.

The model used in the analysis of variance now changes to:

$$X_{ijk} = U + P_i + B_k + E_{ijk}$$

The change of model is necessary because the genetic deviation from provenance breeding value an environmental deviation from block effect cannot be separated by analysis of variance. Results are as follows:

Source	df	SSQ	MSQ	F
Block	5	4661.19		
Provenance	24	7170.31	298.76	3.34
Error	570	51027.13	89.52	
Total	599	62858.63		

Because the F ratio is significantly larger than 1.0, and the within variances are equal among provenances, Box —

Tiao shrinking factor  $1 - (1/F)$  is used. Given the plantation means as  $-.84$  (Table 2), and the shrinking factor  $1 - (1/F)$  as  $1 - (1/3.34) = .70$ , breeding value, Z, according to equation (1) can be expressed as:

$$Z = -.84 + .70(Y + .84) \\ = .70Y - .25$$

It can be seen in Table 2 that the estimated breeding values by Box and Tiao are closer to the true provenance effects 20 times out of 25 comparisons. Only in provenance 1, 6, 12, 16 and 22 are the provenance test means better predictors. By using the Box — Tiao shrinking factor, the average absolute deviation from the true effect was reduced from 1.51 to 1.16, and the sum of square of deviation was reduced from 94.22 to 50.87, a great improvement in accuracy.

### Conclusion

Since most provenance testing involves many seedlots, the phenotypic seedlot mean should be adjusted by a shrinking factor to represent its breeding value. The breeding value of a provenance Z can be expressed as:

$$Z = Y - (1/F)(Y - \bar{Y})$$

where Y = Phenotypic mean observed in the provenance test.

F = F value for the provenance test.

$\bar{Y}$  = Plantation mean.

### Literature Cited

- Box, G. E. P. and G. C. Tiao: Bayesian inference in statistical analysis. Addison-Wesley Publishing Co. Reading, Mass. 588 pp. (1973). — EFRON, B.: Biased versus unbiased estimation. *Advances in Mathematics* 16: 259—277 (1975). — EFRON, B. and C. MORRIS: STEIN'S paradox in statistics. *Scientific American* 236 (5): 119—127 (1977). — JAMES, W. and C. STEIN: Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium*. University of California Press 1: 361—379 (1961). — KUNG, F. H. and C. F. BEY: Heritability construction for provenance and family selection. *Proceedings of the 13th Lake State Forest Tree Improvement Conference*. pp 136—146 (1977). — NANSON, A.: The provenance seedling seed orchard. *Silvae Genet.* 21: 243—249 (1972). — STEIN, C.: Inadmissability of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. pp. 197—206. University of California Press (1955). — VINOD, H. D.: Simulation and extension of a minimum, mean squared error estimator in comparison with STEIN'S. *Technometrics* 18 (4): 491—496 (1976). — WRIGHT, J. W.: Introduction to forest genetics. Academic Press, N.Y., N.Y. 463 pp. (1975).