

A METHOD OF EVALUATING SEEDING UNIFORMITY

H. Pasternak, U. M. Peiper, and J. Putter

Israel Agricultural Research Organization, Bet-Dagan, Israel

Received 10 July 1985, accepted 14 April 1986

Pasternak, H., U. M. Peiper, and J. Putter. 1987. A method of evaluating seeding uniformity. Can. Agric. Eng. 29: 35-37.

The coefficient of variation (CV) of the distances (x) between successive seeds is an accepted criterion of seeding uniformity. The standard method of seeder evaluation estimates CV by the corresponding coefficient of variation in the sample run. In this paper, we propose to estimate CV using the regression of the seed location on its serial number in the sample run. It is shown that this method gives more accurate estimates of the mean and variance of x , and therefore also of CV and hence may reduce the costs of measuring seeding uniformity. A more appropriate approach to the problem of detecting missing and multiple seedings is also discussed. This approach is based on taking into consideration the distance pattern of all the sample run in the process of elimination suspected seedings rather than the suspected distances.

INTRODUCTION

The inaccuracy of a seeder, i.e., the lack of uniformity of the distances between successive seeds, is known to have an adverse effect on the yield. The parameter commonly used to express this inaccuracy is the coefficient of variation (CV = standard deviation/mean) of the distances between successive seeds. In practice, this CV cannot be exactly determined, and has to be estimated from sample data. The estimation problem is complicated by the fact that in some "seeding actions" of the seeder either several seeds or no seeds may be sown. Thus, successive seeds do not always correspond to successive seeding actions. This also contributes to the inaccuracy of the seeder, but it represents another aspect of the problem, and may require modifications other than those needed to increase the uniformity of the distances between seeds sown in successive seeding actions. Therefore, the location of multiple or missing seeds should, as far as possible, be identified and corrected, before the estimator of the CV parameter is calculated.

The standard procedure of the International Organization for Standardization (1982) (I.S.O.) is approximately as follows. First, all the distances between successive seeds (in the sample run) are recorded and compared with the claimed seeding distance R (according to the manufacturer specifications). The distances that are either smaller than $R/2$ or greater than $3R/2$ are ignored. The sample mean (\bar{x}) and the sample variance (v_x) of the remaining distances are then calculated, and the CV parameter is estimated by the ratio

$$c = \frac{\sqrt{v_x}}{\bar{x}} \quad (1)$$

The purpose of this paper is to propose a different CV estimator, based on the

regression of the seeding location on the serial number of the seeding action. For the ideal case of no multiple or missing seeds, we show that the proposed estimator is more accurate than the I.S.O. estimator. We also suggest a different approach to the problem of detecting and correcting for the missing and multiple seedings.

THE REGRESSION METHOD

The Model

Consider the case in which no multiple or missing seeding actions occur. Denote by d_i the location (measured from an arbitrary origin) of the seed sown in the i th seeding action. If the seeding were completely uniform, we would have

$$d_i = a + i \times \mu$$

where a is an arbitrary intercept and μ is the (constant) distance between successive seeds. Actually, due to mechanical and other disturbances, the seed locations deviate from these "ideal" locations by unknown (positive or negative) amounts y_i , i.e.

$$d_i = a + i\mu + y_i \quad (i = 1, 2, \dots, n) \quad (2)$$

where n is the number of seeding actions in the sample run.

We assume that each y_i is a random variable whose distribution does not depend on i . We can assume that the expected value of y_i is

$$E(y_i) = 0$$

We also assume that the deviations y_i are independent, i.e., y_i is not affected by the preceding deviation y_{i-1} or by any other y_j ($j \neq i$). It follows from Eq. (2) that the distances between successive seeds are

$$x_i = d_{i+1} - d_i = \mu + y_{i+1} - y_i \quad (i = 1, 2, \dots, n-1) \quad (3)$$

Thus, the expected value of x_i is

$$E(x_i) = \mu$$

and the variance of x_i is

$$\text{var}(x_i) = 2\sigma_y^2$$

where σ_y^2 is the variance of y_i . The CV parameter, whose value we want to estimate, is therefore

$$\text{CV} = \frac{\sqrt{2\sigma_y}}{\mu} \quad (4)$$

The Proposed Estimator

We propose to estimate Eq. (4) by

$$c1 = \frac{\sqrt{2s_y}}{b} \quad (5)$$

where b and s_y^2 are the usual regression estimators of, respectively, the slope μ and the residual variance σ_y^2 pertaining to the regression Eq. (2) (see page 36). Thus,

$$b = \frac{\sum_{i=1}^n i(d_i - \bar{d})}{\sum_{i=1}^n (i - \bar{i})^2} \quad (6)$$

and

$$s_y^2 = \frac{s_{d \cdot i}^2}{n-2} \quad (7)$$

where

$$s_{d \cdot i}^2 = (1 - r_{d \cdot i}^2) \sum_{i=1}^n (d_i - \bar{d})^2$$

$r_{d \cdot i}$ denoting the sample correlation between d and i . To compare the estimators b and $2s_y^2$ with the corresponding I.S.O. estimators (\bar{x} and v_x), we derived their variances under the assumption that the distribution of the deviations y_i is approximately normal. The derivations are described in Appendix A. It is shown that:

$$\text{var}(x)/\text{var}(b) \cong n/6 \quad (8)$$

and

$$\text{var}(v_x)/\text{var}(2s_y^2) \cong 3/2. \quad (9)$$

Thus, unless the sample run is very short, b estimates μ with a considerably higher precision than x , and $2s_y^2$ estimates $\text{var}(x_i) = 2\sigma_y^2$ with somewhat higher precision than v_x . It may therefore be expected that the proposed estimator (5) will estimate the CV parameter (4) with an appreciably higher precision than the I.S.O. estimator (1).

MISSING AND MULTIPLE SEEDINGS

The problem of eliminating suspected cases of missing and multiple seedings has been discussed in the literature, but full and satisfactory solutions have been found only for simple special cases. The I.S.O. method (described in the introduction) is simple and easy to apply, but is open to criticism, because, in deciding to eliminate a suspected distance (between two successive seeds), it does not take into account the pattern of the neighboring distances. For example, when a small distance is followed by a large one, both discrepancies may be due to a negative deviation of the middle seed rather than to a double seed followed by a missing one. The same objection applies to the method of Burema et al. (1980), which is based on the distribution of the distances in the sample run.

We propose a different approach to the problem, which eliminates suspected seeding locations rather than distances. The approach is based on trying to fit the sequence of seeds to a sequence of seeding actions. The fitting procedure proceeds by trial and error, as follows.

First, it is assumed that the "target location" for the i th seed is

$$(i - 1/2)R$$

where R is the projected between-seeds distance. Thus, the initial target location for the first seed is taken as $R/2$, corresponding to the assumption that the sampling run starts, and the seed location are measured, from a randomly chosen origin.

Second, each actual seed is assigned to the initial target location which is either closest to it or neighboring (immediately preceding or following) the closest one. In choosing the sequence of assigned locations we try to achieve, as far as possible, the following objectives: at least one seed should be assigned to each target (first priority); the number of cases in which several seeds are assigned to one target should be minimal (second priority); and the number of seeds assigned to targets

TABLE I. ELIMINATION OF SUSPECTED MISSING AND MULTIPLE SEEDING, BY THE I.S.O. METHOD AND BY THE SUGGESTED METHOD IN A SAMPLE RUN OF 20 SEEDS

Initial serial number of seeding action	Measured distance from origin (cm)	Distance from preceding seed (cm)	Initial target location (cm)	Assigned distance from origin (cm)	Modified serial number of seeding action
1	25		12	25	1
2	80	55	37		
3	92	12	62	80	3
4	117	25	87	92	4
5	146	29	112	117	5
6	165	19	137	146	6
7	186	21	162	165	7
8	209	23	187	186	8
9	236	27	212	209	9
10	260	24	237	236	10
11	291	31	262	260	11
12	304	13	287	291	12
13	330	26	312	304	13
14	360	30	337	330	14
15	376	16	362	360	15
16	399	23	387	376	16
17	433	34	412	399	17
18	439	6	437		
19	453	14	462	453	19
20	475	22	487	475	20

other than the ones closest to them should be minimal (third priority).

Third, the seeding locations to which either no seed or several seeds have been fitted are eliminated from the regression calculations. However, the serial numbers of the remaining seeding locations are not changed, i.e., the sequence of serial numbers of seeding actions is a sequence of consecutive numbers with some numbers missing. If the number of remaining locations is m , then the estimator of σ_y^2 is $s_{d \cdot i}^2 / (m - 2)$ (see example).

AN EXAMPLE

Table I shows a sample run of 20 seeds, on which the I.S.O. procedure and the suggested procedure were carried out.

The projected between-seeds distances was $R = 25$ cm. The calculations by the I.S.O. method give

$$x = 23.5 \text{ cm}, v_x = 34 \text{ cm}^2, c = 24.9\%$$

The calculation by the suggested method gives

$$b = 23.6 \text{ cm}, r_{d \cdot i} = 0.9996, s_y = 4.0 \text{ cm}, c_1 = 23.8\%$$

CONCLUSIONS

The coefficient of variation of the distances between successive seeding locations is the accepted criterion to measure seeder accuracy. A new method to estimate this criterion based on the regression of the seeding location on the serial number of the seeding action has been developed.

It is shown that this method has an appreciably higher precision than the existing one and hence may reduce the costs of seeder evaluation. Its relative efficiency is increased when analyzing results of a small sample run.

A new approach, which eliminates suspected seeding location rather than distances, is proposed. This approach is based on fitting seed locations to seeding actions and seems to have advantages over the existing methods. When this approach is applied the relative efficiency of the proposed estimation method is increased.

SYMBOLS

- a = intercept.
- b = estimator of μ the target distance between successive seeds.
- c, c_1 = estimator of coefficient of variation.
- d_1 = distance between seed location and the beginning of the interval.
- $r_{d \cdot i}$ = sample correlation between d and i .
- s = sample standard deviation.
- s^2 = sample variance.
- $s_{d \cdot i}^2$ = variance of the deviation of d for the corresponding value of the regression line.
- v = sample variance.
- x = distances between successive seeds.
- y_i = deviation of the seed location from the ideal one.

μ = target constant distance between successive seeds.
 σ^2 = population variance.

Subscripts

cov = covariance.
 CV = coefficient of variation.
 E = expected value.
 I.S.O. = International Organisation for Standardization.
 var = variance.

REFERENCES

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. 1982. Draft standard methods of test for sowing equipment Part 1. Single seed drafts. (ISO/Dis 7256/1). Geneva, Switz.
 BUREMA, H. S., E. N. C. MEISER, and M. G. TELLE 1980. Development and study of a self-recording sticky belt. Research Report 80-3. Institute of Agricultural Engineering, Mansholtlaan 10-12. Wageningen, The Netherlands.

APPENDIX A. DERIVATIONS

Derivation of Eq. (8)

Equation (3) implies that

$$\bar{x} = \mu + (y_n - y_1)/(n - 1). \quad (10)$$

Therefore,

$$\text{var}(\bar{x}) = \frac{2\sigma_y^2}{(n - 1)^2} \quad (11)$$

The variance of the regression estimator (6) is known to be

$$\text{var}(b) = \frac{\sigma_y^2}{\sum_{i=1}^n (i - \bar{i})^2}$$

It follows from the known formulae

$$\sum_{i=1}^n i = \frac{n(n + 1)}{2}$$

and

$$\sum_{i=1}^n (i)^2 = \frac{n(n + 1)(2n + 1)}{6}$$

that

$$\sum_{i=1}^n (i - \bar{i})^2 = n(n^2 - 1)/12$$

Therefore,

$$\text{var}(b) = \frac{12\sigma_y^2}{n(n - 1)(n + 1)} \quad (12)$$

From Eqs. (11) and (12) we obtain

$$\begin{aligned} \text{var}(\bar{x})/\text{var}(b) &= n(n + 1)/(6(n - 1)) \\ &\cong n/6. \end{aligned}$$

Derivation of Eq. (9)

The I.S.O. estimator of

$$\text{var}(x) = 2\sigma_y^2$$

is

$$v_x = \frac{\sum_{i=1}^{n-1} (x_i - \bar{x})^2}{(n - 1)}$$

In terms of y_i , we have

$$x_i - \bar{x} = y_{i+1} - y_i - (y_n - y_1)/(n - 1)$$

and

$$\sum_{i=1}^{n-1} (x_i - \bar{x})^2 = \sum_{i=1}^{n-1} (y_{i+1} - y_i)^2 - (y_n - y_1)^2/(n - 1) \quad (13)$$

The term

$$(y_n - y_1)^2/(n - 1)$$

is negligible relatively to the other terms on the right-hand side of (13). Therefore

$$\sum_{i=1}^{n-1} (x_i - \bar{x})^2 = \sum_{i=1}^{n-1} z_i^2$$

where

$$z_i = y_{i+1} - y_i$$

and

$$\text{var}\left(\sum_{i=1}^{n-1} (x_i - \bar{x})^2\right) \cong S_1 + S_2, \quad (14)$$

where

$$S_1 = \sum_{i=1}^{n-1} \text{var}(z_i^2)$$

and

$$S_2 = \sum_{i=1}^{n-1} \sum_{j \neq i} \text{cov}(z_i^2, z_j^2)$$

To evaluate S_1 , we note that

$$\text{var}(z_i^2) = E(z_i^4) - (E(z_i^2))^2$$

and that

$$E(z_i^2) = \text{var}(z_i) = 2\sigma_y^2$$

If the distribution of y_i (and therefore also of z_i) is approximately normal, then

$$E(z_i^4) \cong 3(\text{var}(z_i))^2 = 12\sigma_y^4$$

and therefore

$$\text{var}(z_i^2) = 8\sigma_y^4$$

Hence,

$$S_1 \cong 8(n - 1)\sigma_y^4 \quad (15)$$

To evaluate S_2 , we note that z_i and z_j are independent, and therefore $\text{cov}(z_i^2, z_j^2) = 0$, unless $|i - j| = 1$. Therefore

$$S_2 = 2 \sum_{i=1}^{n-1} \text{cov}(z_i^2, z_{i+1}^2)$$

Expressing z_i in terms of y_i , we obtain by simple algebra

$$E(z_i^2 z_{i+1}^2) = 6\sigma_y^4$$

and therefore

$$\begin{aligned} \text{cov}(z_i^2, z_{i+1}^2) &= E(z_i^2 z_{i+1}^2) - E(z_i^2)E(z_{i+1}^2) \\ &\cong 2\sigma_y^4 \end{aligned}$$

if the deviations y_i are approximately normal. Hence,

$$S_2 \cong 4(n - 2)\sigma_y^4 \quad (16)$$

Combining Eqs. (14), (15) and (16), we have

$$\text{var}(v_x) \cong \frac{4(3n - 4)\sigma_y^4}{(n - 1)^2} \quad (17)$$

The regression estimator of

$$2\sigma_y^2 \text{ is } 2s_{d,i}^2/(n - 2)$$

If y_i are normal, then

$$s_{d,i}^2/\sigma_y^2$$

known to be distributed as a χ^2 variable with $n - 2$ degrees of freedom, whose variance is $2(n - 2)$. Therefore, the variance of the proposed estimator of $2\sigma_y^2$ is

$$\text{var}(2S_y^2) \cong \frac{8\sigma_y^4}{(n - 2)}. \quad (18)$$

Combining Eqs. (17) and (18), we obtain (9).