Listy Biometryczne - Biometrical Letters
Vol. 38(2001), No. 1, 41-47

Optimal number of replicates and measurements in a single DUST trial

Wiesław Pilarczyk, Bogna Zawieja

Department of Mathematical and Statistical Methods, Agricultural University,
Wojska Polskiego 28, 60-637 Poznań

SUMMARY

The problem of optimal relationship between the number of replicates and measurements in a single DUST (distinctness, uniformity and stability) trial is discussed. The solution is searched in which the discriminating power of characteristics is taken into account. All the problems discussed are illustrated with the use of the results of the DUST trial on winter rye conducted at Słupia Wielka experimental station in which eight characteristics were observed.

KEY WORDS: distinctness, uniformity, stability, optimality, discriminating power.

1. Introduction

DUST trials aim at checking distinctness, uniformity and stability of new varieties of plants before registration. They are conducted according to some fixed principles. Most often, these trials last three years for each variety. Single trials are carried out in randomised complete block designs with two or three replicates and twenty measurements per plot. Observations are done on single plants, their parts or whole plots. In one trial there can be from a few to a few dozens of measured characteristics. There can also be a few dozens to a few hundreds of compared varieties as usually in these trials apart from new varieties all previously registered varieties are sown.

Hence it is evident that costs of DUST trials are high. Therefore, it is important to plan trials in such a way that the funds assigned are used in optimum manner without losing precision of comparisons between varieties. This paper reports results obtained by pointing out winter rye characteristics with the highest discriminating power as well as results of searching of optimal relations between the number of replicates and the number of measurements for particular characteristics, to define an optimal trial plan.

All the calculations presented in the paper were conducted for the results of the trial on winter rye carried out in the Experimental Station for Cultivar Testing at Słupia Wielka in 1999. Results for measurable characteristics were taken into account.

2. Method

The mathematical model of the observation, in the case of the analysis of single DUST trial data (one year and one place data), is as follows:

$$\begin{aligned} y_{ijm} &= \mu + \rho_i + v_j + \eta_{ij} + e_{ijm}, \\ i &= 1,2,...,r; j = 1,2,...,v; m = 1,2,...,k, \end{aligned} \tag{1}$$

where y_{ijm} denotes observed value of characteristic for the m-th measurement of the j-th variety, in the i-th replication; μ denotes the general mean, ρ_i – random effect of the i-th replication, v_j – fixed effect of the j-th variety, η_{ij} – random effect of experimental (plot) error and e_{ijm} – random measurement error. Random components of the model are assumed to have independent normal distributions with expectation equal to zero and variances σ_ρ^2, σ_η^2 and σ_e^2, respectively.

To analyse the results of single trials the analysis of variance is applied. It leads to a conclusion concerning significance of differences between varieties. Successively, all differences among varieties are checked. Usually, the Fisher test for a single comparison or Tukey multiple comparisons test are used. Taking into consideration the result of one of these tests, the characteristic with the highest discriminating power can be found as well as the surplus characteristics (Zawieja, Pilarczyk, 2000). The characteristic with the highest discriminating power is the one that distinguishes the highest number of pairs of varieties. The surplus characteristic is the one that points no additional pairs of varieties (all varieties distinguished by this characteristic have already been declared distinct by other characteristics).

The characteristic having the highest discriminating power will be treated as the basis for calculating such parameters as the number of replications and the number of observations. When calculating these parameters, both precision of comparisons between varieties and costs of conducting experiments have been taken into account.

It is known that the precision of comparisons between varieties can be increased by decreasing variances of these comparisons. For the assumed mathematical model (1) the variance of the comparison between two variety means is (Oktaba 1986)

$$Var(\bar{y}_{.j.} - \bar{y}_{.j'.}) = 2\left(\frac{\sigma_\eta^2}{r} + \frac{\sigma_e^2}{rk}\right). \tag{2}$$

This formula clearly shows that a reduction of variance can be obtained if the number of replications r and the number of observations k is increased. However, increased values of these parameters involve increasing trial costs. Therefore, minimisation of variance ought to be confined by available funds (Pilarczyk 1998). In the case of a single trial the function that should be minimised is

$$F(r,k) = \frac{\sigma_\eta^2}{r} + \frac{\sigma_e^2}{rk} \tag{3}$$

with the confining condition concerning costs of elements of the trial

$$W = w_1 + rw_2 + rkcw_3, \tag{4}$$

where W denotes the amount of money at disposal, w_1 – organising costs connected with the establishment of a new trial, w_2 – organising costs of one replication, w_3 – cost of one measurement of one characteristic on a plot. The meaning of symbols r and k is the same as in formula (2), whereas c is the total number of characteristics. In order to simplify the calculations, it has been assumed that costs of taking measurements are the same for all characteristics, which is not always true.

The Lagrange multipliers method is used to find the minimum of formula (3) (Fichtenholz 1985). Thus, the minimum of the following formula is searched:

$$\min_{r,k,\lambda} F1(r,k,\lambda) = \min_{r,k,\lambda} [F(r,k) + \lambda(W - w_1 - rw_2 - rkcw_3)]. \tag{5}$$

The solution obtained is of the form

$$k^* = \sqrt{\frac{w_2\sigma_e^2}{cw_3\sigma_\eta^2}}, \qquad r^* = \frac{W - w_1}{w_2 + k^*cw_3}. \tag{6}$$

Under the assumption that costs W for a three-year series of trials are triplicate costs W for a single trial and the value w_3 is c times less, the formulas for k^* and r^* are identical to formulas obtained in the paper by Pilarczyk (1998).

Assuming that the number of years of research l is fixed and that funds W' for l years are $W' = lW$, that is, the costs W are the same for each year, the values r^* and k^* calculated for one year can be generalised for l years, because

$$k^{*\prime} = k^*, \qquad r^{*\prime} = \frac{W' - lw_1}{lw_2 + lkcw_3} = r^*. \tag{7}$$

If amount W is fixed irrespective of the number of years of research, then for calculation of k^* and r^* for a fixed l one can use the following formulas:

$$k^* = \sqrt{\frac{w_2\sigma_e^2}{cw_3\sigma_\eta^2}}, \qquad r^* = \frac{W - lw_1}{lw_2 + lkcw_3} \tag{8}$$

These formulas show that with the increase of the number of years the number of replications decreases, but the number of measurements on plot does not change as it is independent of the number of years of the research.

3. Analysis of results

In order to find the optimal relationship between the number of replications and the number of single measurements made at every plot, the analysis of variance is performed as the first step. Next, the estimates of variance components are calculated. This is illustrated in Table 1, in which the analysis of variance for characteristic 32 (length between upper node and ear) made for a single trial on rye performed in 1999 at experimental station in Słupia Wielka is presented.

Table 2 presents the estimates of variance components obtained with the use of restricted maximum likelihood method (REML) of Genstat 5 package for eight characteristics involved.

Table 1. Analysis of variance for characteristic 32

Source of variation	Degrees of freedom	Sums of squares	Mean squares	F-value
Total	3473	161829.97		
Replicates	2	90.19	45.094	0.25
Varieties	72	121707.11	1690.377	9.36**
Plot error	144	26003.25	180.578	
Measurement error	3260	75453.68	23.145	

** denotes significance at $\alpha = 0.01$

Table 2. Estimates of variance components for plot (σ_η^2) and measurement error (σ_e^2) obtained in single trial on winter rye

No	Characteristic	σ_η^2	σ_e^2
31	Plant height	92.486	140.869
32	Length between upper node and ear	3.303	23.145
33	Length of ear	0.187	1.854
41	Length of blade of leaf next to flag leaf	0.896	9.883
42	Width of blade of leaf next to flag leaf	0.004	0.049
51	Number of spikelets	1.627	7.046
52	Length of rachis	14.670	54.378
53	Ear density	0.707	5.093

Using variance components from Table 2 and taking into account the total number of characteristics ($c = 8$) the values of k^* and r^* were calculated according to formulas (6) assuming that $W = 20000; w_1 = 10000; w_2 = 1500, 2000, 2500, 3000, 3500$ and $w_3 = 1$, respectively. Rounded results are given in Table 3.

Even superficial inspection of the results allows to notice that different values r^* and k^* have been obtained for different characteristics. Thus, the problem arises which value to use when planning a new experiment. A possible solution would be to use average values of k^* and r^* for all analysed characteristics. But a better approach seems to be the use of the results for the characteristic with the highest discriminating power among all characteristics involved.

Table 3. Optimum number of replicates and measurements

Characteristic	$W = 20000$									
	$w_2 = 1500$		$w_2 = 2000$		$w_2 = 2500$		$w_2 = 3000$		$w_2 = 3500$	
	r^*	k^*	r^*	k^*	r^*	k^*	r^*	k^*	r^*	k^*
31	7	17	5	20	4	22	4	24	3	26
32	**6**	**22**	**5**	**25**	**4**	**28**	**4**	**30**	**3**	**33**
33	6	25	5	29	4	33	4	36	3	39
41	6	46	5	53	4	59	4	65	3	70
42	6	48	5	55	4	62	4	68	3	73
51	6	29	5	33	4	37	4	41	3	44
52	6	27	5	31	4	35	4	38	3	41
53	6	37	5	43	4	48	4	52	3	57

In this paper the latter approach has been applied. Namely, after using Tukey's method of multiple comparisons at 5% level for all characteristics in turn, the characteristic possessing the highest discriminating power has been chosen (Zawieja, Pilarczyk 2000). The results are summed up in Table 4. As characteristic 32 distinguishes the largest number of variety pairs, establishing the "optimum" number of replicates and measurements for that characteristic is the most meaningful. In all tables which follow the results for this characteristic are given in bold.

Table 4. Number of pairs of varieties declared as different for all characteristics

Characteristic	31	32	33	41	42	51	52	53
Number of pairs	218	**1162**	27	22	100	8	13	59

4. Conclusions

The results obtained for a single experiment within a year can easily be generalised for many years but under a rather unrealistic assumption that the costs of experiments are the same for all years. The results obtained depend on the relations among costs of W, w_1 and w_2. Those are rarely known. Only the values of W are relatively easy to assess. Therefore, it seems to be reasonable to compare ratios of k^*/r^* obtained in the paper for different combinations of costs with the ratios k/r used in practice. The values of k^*/r^* obtained for various costs w_2 are given in Table 5.

Table 5. Ratio of the number of measurements to the number of replicates

Characteristic	k^*/r^*				
	$w_2 = 1500$	$w_2 = 2000$	$w_2 = 2500$	$w_2 = 3000$	$w_2 = 3500$
31	2	4	6	6	9
32	**4**	**5**	**7**	**8**	**11**
33	4	6	8	9	13
41	8	11	15	16	23
42	8	11	16	17	24
51	5	7	9	10	15
52	5	6	9	10	14
53	6	9	12	13	19

It is easy to notice that this ratio depends heavily on values of w_2 (the cost of a single replicate). For characteristic 32, for $w_2 = 2000$, that ratio equals 5. It means that the optimal ratio k^*/r^* is 5 for assumed relations between the costs. In the experiment on winter rye, which results were used for estimating the variance components, there were 3 replicates and 20 measurements, so the ratio k/r was about 7. As the relations between different costs can be different in different experimental stations, a detailed solution must be searched for every particular situation.

References

Fichtenholz G.M. (1985). *Rachunek różniczkowy i całkowy.* Tom I, PWN, Warszawa.

Oktaba W. (1986). *Metody statystyki matematycznej w doświadczalnictwie.* PWN, Warszawa.

Pilarczyk W. (1998). Optymalne planowanie dowiadczeń OWT. Zagadnienie liczby powtórzeń i liczby pomiarów osobniczych. *XXVIII Colloquium Biometryczne*, 15-26.

Zawieja B., Pilarczyk W. (2000). Próba oceny mocy dyskryminacyjnej cech mierzalnych żyta ozimego przy badaniu odrębności odmian. *XXX Colloquium Biometryczne*, 207-217.

Received 12 January 2001; revised 23 May 2001

Optymalna liczba replikacji oraz liczby pomiarów osobniczych w pojedynczym doświadczeniu OWT

STRESZCZENIE

W pracy rozważane jest zagadnienie optymalnej liczby pomiarów osobniczych i liczby replikacji w doświadczeniu pojedynczym dotyczącym badania odrębności, wyrównania i trwałości odmian. W proponowanych rozwiązaniach uwzględnia się też problem mocy dyskryminacyjnej uwzględnionych w analizie cech. Prowadzone rozważania ilustrowane są wynikami doświadczenia z żytem ozimym, przeprowadzonego w Stacji Doświadczalnej Oceny Odmian w Słupi Wielkiej, w którym badano osiem cech mierzalnych.

SŁOWA KLUCZOWE: odrębność, wyrównanie, trwałość, optymalność, moc dyskryminacyjna.