E

TWC/19/2
ORIGINAL: English
DATE: April 10, 2001

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

# TECHNICAL WORKING PARTY ON AUTOMATION AND COMPUTER PROGRAMS

## Nineteenth Session
## Prague, June 4 to 7, 2001

THREE-STAGE TESTING UNIFORMITY OF VARIETIES OF SELF-FERTILIZED PLANTS IN DUS TRIALS

*Document prepared by experts from Poland*

# THREE-STAGE TESTING UNIFORMITY OF VARIETIES OF SELF-FERTILIZED PLANTS IN DUS TRIALS

by *BOGNA ZAWIEJA*

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

Summary

In DUST (distinctness, uniformity, stability) testing uniformity of new candidate varieties is examined, among other things. For this purpose, special DUST field trials are conducted. In every DUST trial, both candidate varieties and reference (old) varieties are compared. In testing uniformity of self-fertilized varieties, the off-type plants are counted and compared with the maximum number of off-types allowed. There is always a risk of two wrong decisions: (1) rejecting uniform variety; (2) declarations that the variety is uniform when in fact it is not uniform.

In this paper, in order to find a better size of trial, a three-stage test is described. The errors of both kinds were calculated and compared. In this calculation, the parameters, established in the UPOV Test Guidelines, were taken into account. The method is applicable for self-fertilized and vegetatively propagated species only.

## Introduction

In the DUST trials, conducted in the Experimental Stations for Cultivar Testing, Poland, the degree of uniformity of new varieties is tested. The criteria of uniformity (homogeneity) are established on the basis of UPOV Test Guidelines. This organization coordinates international cooperation for working out research methods applied before registration and protection of new varieties of plant. The decision concerning uniformity of a variety is based on a comparison of variability between plants of a new variety with variability of registered varieties. Uniformity is checked for all characteristics for which distinctness is tested. A new variety is required to have uniformity of a degree not smaller than the other varieties used for comparison (Pilarczyk 1993).

As the costs of DUST trials are high, it is important to have the sample size as small as possible. It is required, however, that the new variety is uniform in degree not smaller than other well-known varieties of a given species. In this paper we tried to establish the optimal sample size depending on the fixed border of probability of the error of 1st kind while controlling the probability of the error of 2nd kind.

A similar consideration was conducted previously (Zawieja 2000) for the case of research uniformity of self-fertilized varieties of plants where the single test and the two-stage test were compared. Since DUST trials in Poland are usually conducted in three-year series, a three-stage test of research of uniformity of varieties will be considered. Afterwards the optimal sample size resulting from the conducted calculations will be pointed out.

## Method

During the assessment of varietal uniformity there is a risk of making two wrong decisions. The first one is to declare the rejection of a variety, when in fact it is uniform – this is the risk of making the error of 1st kind. The second one is to declare that the variety is uniform when in fact it is not uniform. This decision is connected with the risk of making the error of 2nd kind. In this paper, a three-stage method of checking uniformity of rye (inbred lines) varieties was applied in order to find a sample size guaranteeing the smallest possible probability of the error of 2nd kind, simultaneously controlling level of probability of the error of 1st kind. It means that the difference between the probability of both kinds of errors should be as small as possible and the probability of the error of 1st kind cannot exceed allowed values (Kristensen 1998). These considerations are conducted in order to reduce the costs of trials through the reduction of a sample size.

According to UPOV Test Guidelines, uniformity of new self-fertilized and vegetatively propagated species of plants consists of counting off-types in a sample of n size. The result is compared with Tables (Kristensen 1998). In these Tables, maximum numbers of off-types allowed for different sample sizes, different population standards and acceptance probability are given. The construction of these Tables is based on the fact that appearance of off-types is subject to binomial distribution of probability.

The population standard, marked as p, is understood as the maximum fraction of off-types in varieties, which are recognized as uniform. Very often the population standard is expressed in per cent values. The acceptance probability is the probability of accepting a variety as uniform with the population standard p (p% off-types). These parameters are established for each species of plant separately, on the basis of observation of known varieties of given species. They are given in guidelines of DUST trials. In this paper, the parameters concerning inbred lines of winter rye are applied.

For varieties of self-fertilized plants the population standard, established by the registering organization, is: p=0,02. The acceptance probability should not be smaller than 95%. For the established values of these parameters, different sample sizes were appointed and from appropriate Tables (Kristensen 1998) the maximum accepted numbers of off-type plants were taken. Having established these parameters the search for the optimal calculation could begin.

In the case of a three-stage test, the decision scheme given in Table 1 was applied (Kordoński 1967, Oderfeld 1954). Every year the sample of n elements has been taken and the number of off-types has been counted. In Table 1, the observed number of off-types in i-th year of trial is designated by $k_i$ for i=1,2,3. The maximum acceptable number of off-types for sample size i*n, found in the Table, is designated by $r_i$ ,and $a_i$ is fixed (by researcher), number of off-types for which the variety is accepted for given (i-th) stage (year) of trials.

Table 1

Decision scheme for three-stage test

| stage | sample size | acceptance decision | rejection decision | passing on the next stage |
|---|---|---|---|---|
| 1 | n | $k_1 < a_1$ | $k_1 > r_1$ | $a_1 \le k_1 \le r_1$ |
| 2 | 2n | $k_1+k_2 < a_2$ | $k_1+k_2 > r_2$ | $a_2 \le k_1+k_2 \le r_2$ |
| 3 | 3n | $k_1+k_2+k_3 \le r_3$ | $k_1+k_2+k_3 > r_3$ | ____________ |

The probability of making errors of both kinds and the expected sample size were calculated using Bernoulli's distribution and conditional probability. The formulae for calculating the probability of the error of 1st kind are:

$$\alpha = P(k_1 > r_1) + P(k_1 + k_2 > r_2 \mid k_1) + P(k_1 + k_2 + k_3 > r_3 \mid k_1 + k_2); \qquad (1)$$

$$\alpha = P(k_1 > r_1) + P(k_2 > r_2 - k_1 \mid k_1) + P(k_3 > r_3 - k_1 - k_2 \mid k_1 + k_2); \qquad (2)$$

then:

$$\alpha = \sum_{i=r_1+1}^{n} P_i + \sum_{i=a_1}^{r_1} P_i \sum_{j=r_2-i+1}^{n} P_j + \sum_{i=a_1}^{r_1} P_i \sum_{\substack{j=a_2-i \\ j\ge 0}}^{r_2-i} P_j \sum_{\substack{l=r_3+1-i-j \\ l\ge 0}}^{n} P_l ; \qquad (3)$$

where n is the sample size, and:

$$P_s = \binom{n}{s} p^s (1-p)^{n-s}. \qquad (4)$$

The probability of the error of 2nd kind is calculated for three different degrees of heterogeneity of the variety, that is assuming that the number of off-types, in a given variety, is greater than $r_i$ for various extent. The calculations were done for a fraction of off-types $p_q$ equal to 2, 5, 10 times the standard population (which was assumed to be lower than or equal to 0,1 that is $p \le 0{,}1$; $p_q \le 1$). Then:

$$\beta = P(k_1 < a_1) + P(k_1 + k_2 < a_2 \mid k_1) + P(k_1 + k_2 + k_3 < r_3 \mid k_1 + k_2); \qquad (5)$$

$$\beta = P(k_1 < a_1) + P(k_2 < a_2 - k_1 \mid k_1) + P(k_3 < r_3 - k_1 - k_2 \mid k_1 + k_2); \qquad (6)$$

and:

$$\beta = \sum_{i=0}^{a_1-1} P_{iq} + \sum_{i=a_1}^{r_1} P_{iq} \sum_{j=0}^{a_2-i-1} P_{jq} + \sum_{i=a_1}^{r_1} P_{iq} \sum_{\substack{j=a_2-i \\ j\ge 0}}^{r_2-i} P_{jq} \sum_{l=0}^{r_3-i-j} P_{lq} ; \qquad (7)$$

where:

$$P_{sq} = \binom{n}{s} p_q^s (1-p_q)^{n-s}; \qquad (8)$$

and: $p_q = qp$; where: $q = 2,5,10$. (9)

So the expected sample size is:

$n_e = n(1-P(a_1 \le k_1 \le r_1)) + 2n(P(k_1+k_2<a_2 \mid k_1)+P(k_1+k_2>r_2 \mid k_1)) + 3n(P(a_2 \le k_1+k_2 \le r_2 \mid k_1))$ (10)

that is:

$$n_e = n\left(1 + \sum_{i=a_1}^{r_1} P_i + \sum_{i=a_1}^{r_1} P_i \sum_{\substack{j=a_2-i \\ j\ge 0}}^{r_2-i} P_j\right); \qquad (11)$$

where: $P_s$ - is given in formula (4).

Example calculations

In the DUST trials, as mentioned before, there exist some established principles concerning the number of replications and number of measurements per plot, thereby affecting the number of researched plants of a given variety in one year. For trials conducted within the UPOV member States, the number of replications is often three, but in Poland, apart from three, there are also applied two replications. Therefore for calculation purposes, the sample sizes which were multiplications of these numbers were selected. Results are given in Table 2. In this Table, for each sample size, the values $r_1$, $r_2$ and $r_3$ (as in the Tables) are given and three different levels of $a_1$ and $a_2$ of a fraction of the off-types chosen by a researcher are given. Expected sample sizes, given in Table 2, are rounded to integer number according to ordinary principles.

Table 2

The probabilities of 1st and 2nd kind of error for three-stage test

| lp. | n | $a_1$ | $a_2$ | $r_1$ | $r_2$ | $r_3$ | $1^{st}$ error | $2^{nd}$ error | | | $n_e$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | 4% | 10% | 20% | |
| 1 | 34 | 0 | 2 | 2 | 3 | 5 | 6.44% | 64.22% | 3.89% | 0.00% | 78 |
| 2 | | 1 | 2 | | | | 614% | 66.05% | 5.30% | 0.05% | 59 |
| 3 | | 2 | 2 | | | | 4.91% | 73.38% | 14.24% | 0.48% | 41 |
| 4 | 38 | 0 | 2 | 2 | 4 | 5 | 6.06% | 62.34% | 2.01% | 0.00% | 90 |
| 5 | | 1 | 2 | | | | 5.87% | 64.04% | 3.39% | 0.02% | 69 |
| 6 | | 2 | 4 | | | | 4.89% | 72.60% | 11.28% | 0.22% | 44 |
| 7 | 40 | 0 | 2 | 2 | 4 | 5 | 7.07% | 58.15% | 1.35% | 0.00% | 95 |
| 8 | | 1 | 2 | | | | 6.83% | 60.04% | 2.53% | 0.01% | 74 |
| 9 | | 2 | 3 | | | | 5.87% | 67.68% | 8.63% | 0.15% | 49 |
| 10 | 42 | 0 | 3 | 3 | 4 | 5 | 5,00% | 60.08% | 1.40% | 0.00% | 92 |
| 11 | | 1 | 2 | | | | 4.97% | 60.45% | 2.02% | 0.01% | 82 |
| 12 | | 2 | 3 | | | | 3.81% | 68.82% | 7.30% | 0.10% | 55 |
| 13 | 48 | 0 | 2 | 3 | 4 | 6 | 5.78% | 56.04% | 0.62% | 0.00% | 121 |
| 14 | | 1 | 3 | | | | 5.54% | 57.96% | 1.26% | 0.00% | 88 |
| 15 | | 2 | 3 | | | | 4.72% | 63.95% | 4.36% | 0.03% | 66 |
| 16 | 54 | 0 | 3 | 3 | 5 | 6 | 6.05% | 50.53% | 0.28% | 0.00% | 125 |
| 17 | | 1 | 4 | | | | 5.48% | 54.54% | 0.80% | 0.00% | 96 |
| 18 | | 2 | 4 | | | | 4.87% | 59.69% | 2.63% | 0.01% | 74 |
| 19 | 57 | 0 | 3 | 3 | 5 | 7 | 5,24% | 54.57% | 0.30% | 0.00% | 133 |
| 20 | | 1 | 3 | | | | 5,16% | 55.40% | 0.51% | 0.00% | 113 |
| 21 | | 2 | 3 | | | | 4,87% | 59.69% | 1.98% | 0.00% | 85 |
| 22 | 60 | 0 | 3 | 3 | 5 | 7 | 6.36% | 49.33% | 0.16% | 0.00% | 141 |
| 23 | | 2 | 4 | | | | 5.66% | 55.64% | 1.52% | 0.00% | 85 |
| 24 | | 3 | 4 | | | | 4.57% | 65,43% | 5.35% | 0.02% | 68 |
| 25 | 69 | 2 | 4 | 3 | 6 | 8 | 6.55% | 50.60% | 0.66% | 0.00% | 105 |
| 26 | | 3 | 5 | | | | 5.85% | 59.10% | 2.66% | 0.00% | 79 |

The probabilities of the error of 1st and 2nd kind depend on established fractions of off-types $a_1$ and $a_2$. Namely, along with increasing $a_1$ and $a_2$ values, the probability of the error of 1st kind decreases but the probability of the error of 2nd kind increases. In order to illustrate this dependency better, in Table 3, for sample size 42, the values of probability of the error for all possible values of $a_1$ and $a_2$ are given.

Table 3

The probabilities of 1st and 2nd kind of error for n=42

| $a_1$ | $a_2$ | 1st error | 2nd error | | | $n_e$ | $a_1$ | $a_2$ | 1st error | 2nd error | | | $n_e$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 4% | 10% | 20% | | | | | 4% | 10% | 20% | |
| 0 | 0 | 5.31% | 58.08 % | 0.96 % | 0.00 % | 124 | 1 | 3 | 4.77 % | 61.72 % | 2.30% | 0.01% | 73 |
| 0 | 1 | 5.30% | 58.10 % | 0.96 % | 0.00 % | 117 | 1 | 4 | 4.10 % | 66.36 % | 3.63% | 0.01% | 68 |
| 0 | 2 | 5.26% | 58.39 % | 1.02 % | 0.00 % | 103 | 2 | 2 | 3.87 % | 68.40 % | 7.21% | 0.10% | 58 |
| 0 | 3 | 5.00% | 60.08 % | 1.40 % | 0.00 % | 92 | 2 | 3 | 3.81 % | 68.82 % | 7.30% | 0.10% | 55 |
| 0 | 4 | 4.23% | 65.35 % | 2.91 % | 0.00 % | 86 | 2 | 4 | 3.42 % | 71.45 % | 8.05% | 0.10% | 52 |
| 1 | 1 | 4.99% | 60.31 % | 1.99 % | 0.01 % | 88 | 3 | 3 | 2.24 % | 80.82 % | 19.63 % | 0.56% | 45 |
| 1 | 2 | 4.97% | 60.45 % | 2.02 % | 0.01 % | 82 | 3 | 4 | 2.15 % | 81.45 % | 19.81 % | 0.56% | 44 |

In Tables 2 and 3 expected sample sizes are also given, which inform about the sample size when the trial concerns many varieties simultaneously. It means that by multiplying $n_e$ by the number of tested varieties, the estimated number of individual measurements will be found, which is necessary within a three year series of trials.

## Conclusion

After inspection of counted probabilities of the error of 1st kind (Table 2), it is obvious that not all presented arrangements of parameters fulfill the requirements indicated under the previous headings. In some schemes the probability of the error of 1st kind is greater than 5%, thus the acceptance probability is lower than required. Among them there are also sample sizes 60 and 40, which are in accordance with the plans of trials established for DUST trials. The sample sizes 34 and 38 fulfill the above given criteria when $a_1$=$r_1$; it means that only in the case when the number of off-types is equal to $r_1$ can we pass to the second stage. Such a situation is not very interesting because, although the accepted sample size is small, the probability of the error of 2nd kind is then relatively large. In other words, the decision on uniformity would be often taken after one year of trialing but also the relatively heterogeneous varieties would often be accepted as uniform. Apart from this, the researcher has no freedom in choosing the value of $a_1$.

While the sample size 42 (which is in accordance with examining 14 plants from every plot in a trial with 3 replications or 21 plants in one with 2 replications respectively) and 57, which only slightly exceeds 5%, fulfill all required criteria, these sample sizes lead to decreasing the costs of conducting the research because the sample size was decreased (when compared with 60).

Observing expected sample size (Table 2) it can be stated that the real number of measured plants in the case of a three-stage test will be lower than in the case of traditional annual tests of uniformity of varieties. It can also be noticed that expected sample size depends on fractions of off-types $a_1$ and $a_2$ established by a researcher. It means that, in trying to establish these parameters, one should take into consideration the probability of errors of both kinds and the expected sample size, thus trying to find a balance between the probability of errors of two kinds and the cost of trialing.

References

1. Kristensen K., 1996, Testing of uniformity of self-fertilized and vegetatively propagated species using off-types, UPOV, TC\14\3, Geneva.
2. Pilarczyk W., 1993, Badanie odrębności, wyrównania i trwałości odmian. Zakres stosowania metod statystycznych w badaniach OWT. Dwudzieste Trzecie Colloquium Metodologiczne z Agrobiometrii, PAN, str. 74-85.
3. Kordoński Ch. B., 1967, Zastosowanie rachunku prawdopodobieństwa w technice, Wydawnictwo Naukowo Techniczne, Warszawa.
4. Fisz M., 1954, Rachunek prawdopodobieństwa i statystyka matematyczna. PWN, Warszawa.
5. Oderfeld J., 1954, Zarys statystycznej kontroli jakości. PWN, Warszawa.
6. Zawieja B., 2000, Porównanie dwóch sposobów testowania wyrównania odmian roślin samopylnych. Bibliotheca Fragmenta Agronomika, t.7, str. 193-202.

[End of document]