E

TGP/8.6 Draft 2
ORIGINAL: English
DATE: May 6, 2004

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

Associated Document
to the
General Introduction to the Examination
of Distinctness, Uniformity and Stability and the
Development of Harmonized Descriptions of New Varieties of Plants (document TG/1/3)

DOCUMENT TGP/8

"USE OF STATISTICAL PROCEDURES IN

DISTINCTNESS, UNIFORMITY AND STABILITY TESTING"

Section TGP/8.6: Examining DUS in Bulk Samples

Document prepared by experts from Denmark

to be considered by the

Technical Working Party on Automation and Computer Programs (TWC),
at its twenty-second session, to be held in Tsukuba, Japan, from June 14 to 17, 2004

SECTION 8.6

EXAMINING DUS IN BULK SAMPLES

8.6.1 Introduction and abstract

1. In some crops, samples are bulked before certain characteristics are examined. The term "bulk sampling" is used here for the process of merging some or all individual plants before recording a characteristic. There are different degrees of bulking ranging from: 1) merging pairs of plants, 2) merging 3 or 4 up to all plants within a plot up to 3) merging all plants within a variety. The degree of bulking may play an important role in the efficiency of the tests. Bulking is usually only applied where the measurement of the characteristic is very expensive or very difficult to obtain from individual plants. Some examples are seed weight in field peas and field beans, and erucic acid content in rapeseed. This document describes some of the consequences of bulk sampling. It is shown that the test of distinctness (using COYD) may be expected to be relatively insensitive to the degree of bulking, but that the efficiency of the tests for uniformity (using COYU) must be expected to decrease when the data are bulked. The COYU test for uniformity cannot be carried out if all plants within a plot are bulked.

8.6.2 Distinctness

2. In the COYD method for examining distinctness, the basic values to be used in the analyses are the annual variety means. As bulk sampling also gives at least one value for each variety per year, it will usually still be possible to use the COYD method for distinctness purposes for any degree of bulking, as long as at least one value is recorded for each variety in each year and that the bulk samples are representative for the variety. However, some problems may be foreseen: the assumption of data being normal distributed may be better fulfilled when the mean of many individual measurements are analyzed instead of the mean of fewer measurements or, in the extreme, just a single measurement.

3. The efficiency of the test of distinctness may be expected to be lower when based on bulked samples than when it is based on the mean of all individual plants in a year. The loss will be from almost zero upwards, depending on the importance of the different sources of variations. The variation which is relevant for the efficiency of variety comparisons is formulated in the following model.

$$\sigma_{total}^2 = \sigma_{vy}^2 + \sigma_p^2 + \sigma_i^2 + \sigma_m^2$$

where

σ_{total}^2 is the total variance of a characteristic used for comparing varieties

The total variance is regarded as being composed of four sources of variations:

1: σ_{vy}^2 the year in which the variety is measured

2: σ_p^2 the plot in which the measurement was taken

3: σ_i^2 the plant on which the measurement was taken

4: σ_m^2 the inaccuracy in the measurement process

4. In cases where the data is not bulked, the variance of the difference between two variety means, σ^2_{diff}, becomes:

$$\sigma^2_{dif} = 2\left\{\frac{\sigma^2_{vy}}{a} + \frac{\sigma^2_p}{ab} + \frac{\sigma^2_i}{abc} + \frac{\sigma^2_m}{abc}\right\}$$

where

a is the number of years used in the COYD method

b is the number of replicates in each trial

c is the number of plants in each plot

5. Assuming that each bulk sample has been composed in such a way that it represents an equal amount of material from all the individual plants which have been bulked into that sample, the variance between two varieties based on k bulked samples (each of l plants) becomes:

$$\sigma^2_{dif} = 2\left\{\frac{\sigma^2_{vy}}{a} + \frac{\sigma^2_p}{ab} + \frac{\sigma^2_i}{abkl} + \frac{\sigma^2_m}{abk}\right\}$$

where

k is the number of bulk samples

l is the number of plants in each bulk sample

6. Thus if all plants in each plot are divided in k groups of l plants each and an average measurement is taken for each of the k groups, then only the last term in the expression for σ^2_{diff} has increased (as kl is equal to c). For many characteristics it is found that the variance caused by the measurements process is small and hence the bulking of samples will only have a minor effect on the conclusions reached by the COYD method. Only if the variance caused by the measurement process is relatively large can bulking have a substantial effect on the distinctness tests using COYD.

7. To illustrate the effect, the variances for comparing varieties were estimated (by the use of estimated variance components) for different degrees of bulking. The calculations were based on the weight of 100 seeds of 145 pea varieties grown in Denmark during 1999 and 2000. In this example, the contribution to the variance caused by the measurement process was relatively very small, which means that bulking will have a low influence on the test for distinctness. In a 3 year test with 30 plants in each of 2 blocks, the variance on a difference between two varieties was estimated to be 2.133 and 2.135, for no bulking and a single bulk sample per plot, respectively. It should be noted that tests for uniformity are impossible if only 1 bulk per plot is used (see section 6.3).

8. For other variables the variance component due to the measurement process may be relatively more important. However, it is likely that in most practical cases this variance component will be relatively small.

9. In some cases, each bulk sample is not drawn from a specific set of plants (say, plant 1 to 5 in bulk sample 1, plant 6 to 10 in bulk sample 2 etc.), but bulk samples are formed from mixed samples of all plants in a plot. This means that different bulk samples may contain material from the same plants. It must be expected that similar results apply here, although,

in this situation, the effect of bulking may have an increased effect because there is no guarantee that all plants will be equally represented in the bulk samples.

8.6.3 Uniformity

Bulking within plot

10. In COYU the test is based on the standard deviation between individual plants (within plots) as a measurement of uniformity. The log of the standard deviations plus one are analyzed in an over-years analysis; i.e. the values $Z_{vy} = \log(s_{vy} + 1)$ are used in the analyses. The variance on these Z_{vy} values can be regarded as arising from two sources, a component that depends on the variety-by-year interaction and a component that depends on the number of degree of freedom used for estimating the standard deviation, s_{vy} (the fewer degrees of freedom the more variable the standard deviation will be). This can be written (note that the same symbols as used in the distinctness section will be used here with different meaning):

$$Var(Z_{vy}) = \sigma_{vy}^2 + \sigma_f^2$$

where this varaince can is regarded as being composed of two sources of variations:

1: σ_{vy}^2 the year in which the variety is measured

2: σ_f^2 the number of degrees of freedom using in estimating s_{vy}

σ_f^2 is approximately $\frac{1}{2\nu}\left(\frac{\sigma}{\sigma+1}\right)^2$ when then recorded variable is normally distributed and the standard deviations do not vary too much. This last expression reduces to $0.5/\nu$ when $\sigma >> 1$. Here σ is the mean value of the s_{vy} values and ν is the number of degrees of freedom used in the estimation of s_{vy}.

11. The variance caused by the year in which the variety is measured may be assumed to be independent of whether the samples are bulked or not, whereas the variance caused by the number of degrees of freedom will be increase when bulked samples are used because a lower number of degrees of freedom is available.

12. The variance of a difference between a Z_{vy} for a candidate variety and the mean of the reference varieties' Z_{vy} values may be written:

$$\sigma_{dif}^2 = \left(\sigma_{vy}^2 + \sigma_f^2\right)\left(\frac{1}{a} + \frac{1}{ar}\right)$$

where

a is the number of year used in the test

r is the number of refference varieties

13. To illustrate the effect of bulking in the test for uniformity, an estimate was made using the same data as for the illustration in Section 6.2, paragraph 7. For a test using 50 reference varieties in 3 years with 30 plants per variety in each of 2 plots per trial the variance for comparing the Z_{vy} value for a candidate variety and the mean of the reference varieties' Z_{vy} will be 0.0004 if no bulking is done. This can be compared to 0.0041, 0.0016 and 0.0007

when 2, 4 and 10 bulk samples per plot were used. Thus, in this example, the effect of bulking has a great influence on the test for uniformity. The variance increased, approximately by a factor of 10 when changing from individual plant records to just 2 bulk samples per plot. This means that the assessed degree of non-uniformity must be much higher for it to be detected when 2 bulk samples are used instead of individual plant records.

Bulking across plots

14. Bulking across plots means that part of the between-plot (and block) variation will be included in the estimated standard deviation between bulked samples. If this variation is relatively large it will tend to mask any differences in uniformity between varieties. In addition some noise may also be added because the ratio of material from the different plots may vary from bulk to bulk. Finally the assumptions for the present recommended method, COYU, may not be fulfilled in such cases. Therefore it is recommended to bulk only within plots.

Taking just one bulk sample per plot

15. In general, if all plants in a plot are bulked in such a way that only a single sample is available for each plot, it becomes impossible to calculate the within-plot variability, in which cases no tests for uniformity can be performed. In rare cases, where non-uniformity may be judged from values that can only be found in mixtures, non-uniformity may be detected even where a single bulk sample for each plot is used. For example, in the characteristic "erucic acid" in oil seed rape, values between 2% and 45% can only arise because of a lack of uniformity. However this only applies in certain special cases and even here the non-uniformity may only be assessed under certain circumstances.

[End of document]