TWC/VIII/9 Rev.
Original: English
Date: May 22, 1990

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

TECHNICAL WORKING PARTY ON AUTOMATION AND COMPUTER PROGRAMS

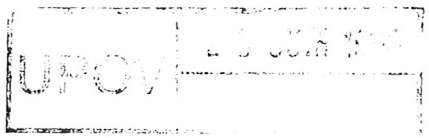
Eighth Session
Belfast, United Kingdom, June 6 to 8, 1990

STATISTICAL ASPECTS OF
MINIMUM DISTANCE BETWEEN VARIETIES

DOCUMENT PREPARED BY EXPERTS FROM THE UNITED KINGDOM

(ORIGINAL)

STATISTICAL ASPECTS OF MINIMUM DISTANCE BETWEEN VARIETIES



INTRODUCTION

At the sixth session of the TWC in 1988 the question of minimum distance between varieties was discussed. It was agreed to seek the views of experts from other groups on what they saw as the problems of establishing minimum distance in the separate species. The response to this request raised a number of specific questions some of which were addressed at the seventh session of the TWC in 1989. At that session it was agreed that the question of minimum distance should be studied further. This paper aims to provide a basis for such a study. The paper summarises the difficulties experienced by experts and it suggests some statistical principles which may be relevant.

BACKGROUND

The main legal basis for minimum distance is set out in Article 6(1)(a) of the UPOV Convention which states:

> "Whatever may be the origin, artificial or natural, of the initial variation from which it has resulted, the variety must be clearly distinguishable by one or more important characteristics from any other variety whose existence is a matter of common knowledge at the time when protection is applied for. Common knowledge may be established by reference to various factors such as: cultivation or marketing already in progress, entry in an official register of varieties already made or in the course of being made, inclusion in a reference collection, or precise description in publication. The characteristics which permit a variety to be defined and distinguished must be capable of precise recognition and description."

Two key notions contained in this statement are"clearly distinguishable" and "important characteristics". It has been decided that the term "important" in this context means important for the purpose of establishing distinctness. and does not refer to economic or practical value of the character.

With regard to the term, "clearly distinguishable" the Revised General Introduction to the Guidelines for the conduct of Tests for Distinctness, Homogeneity and Stability of New Varieties of Plants, states:

"20. Two varieties have to be considered distinct if the difference:

- has been determined at least in one testing place,
- is clear, and
- is consistent"

"21. In the case of true qualitative characteristics the difference between two varieties has to be considered clear if the respective characteristics show expressions which fall into two different states. In the case of other qualitatively handled characteristics an eventual fluctuation has to be taken into account in establishing distinctness."

"22. When distinctness depends on measured characteristics, the difference has to be considered clear if it occurs with one per cent probability of an error, for example, on the basis of the method of the Least Significant Difference. The differences are consistent, if they occur with the same sign in two consecutive, or in two out of three, growing seasons."

2 X 1% AND COY

The 2 x 1% rule has been used for sometime as a measure of minimum distance. However, the 2 x 1% rule has been criticized because a difference that was nearly significant each year contributes no more to the assessment of distinctness than, for example, a zero difference in each year. In 1984 the TWC recommended the adoption of the combined-over-years (COY) criterion which measures variety differences against variation in differences over years. COY has several desirable characteristics:

- it provides a check on the reproducibility of variety differences over years;
- it uses all the available information;
- the risks of making incorrect decisions is constant for each character.

COY also has the advantage of providing a measure of minimum distance in the same units as are used for measurement of the plants.

Since COY was introduced the TWC has been working on the details of its implementation including the most appropriate probability levels to ensure a smooth transition from the 2 x 1% rule.

PROBLEMS WITH MINIMUM DISTANCE

The problems experienced by experts in implementing the "clearly distinguishable" or minimum distance principle are set out below together with tentative solutions, where possible.

a) LSD < MINIMUM DISTANCE

Several experts expressed their concern at minimum distances less than the observational unit, eg. less than one day for ear emergence if observations are recorded in units of days.

Where observations are averaged over several plants, plots and trials, then it is possible to obtain, and valid to use, LSDs which are less than the observational unit. In such cases it must be for the expert to decide the minimum unit to be used for distinctness testing. As long as the minimum distance is greater than the LSD then there should be no difficulty.

b) LSD > MINIMUM DISTANCE

A number of examples have been reported where LSDs are greater than the largest minimum distance.

Some characteristics are by their nature less consistent than others, eg. percentages and counts. This lack of discriminating power might be compensated for by increasing replication or by improving the experimental technique.

In some circumstances the LSD may be greater than the target minimum distance because the variances used in calculating the LSD are based on data which contain unusual or aberrant values. It should be possible to minimise the effects of such occurrences by the use of, what are known as, robust estimation techniques.

c) ESTIMATING MINIMUM DISTANCE FROM SMALL DATA SETS

When there are only a few varieties in trials, then there are few degrees of freedom for estimating variances and the Student's t multiplier used in calculating the LSD can be large.

An advantage of using the COY method is that data from years prior to those in which we are immediately interested can be used to provide an estimate of LSD to apply to current data. Such LSDs should be more stable and more representative. Details of this method have been supplied in TWC/VII/6.

d) MAINTENANCES OF THE SAME VARIETY

An example has been supplied of a case where differences between two stocks of the same variety exceeded the LSD.

There may be several reasons for such an occurrence: it may represent a genuine difference and the stocks may have drifted apart in the multiplication process; alternatively, one must expect occasionally to obtain a significant difference when there is no real difference.

The example supplied by the expert was results for one trial only. Before one could judge which of the possible occurrences is the more likely, it would be necessary to obtain data from at least one further trial.

e) SHAPE CHARACTERS

Examples have been presented of shape characters for which minimum distances need to be established, eg. strawberry cores.

Several approaches are possible. One approach, already in use, is to express shape in terms of a single shape coefficient, eg. the ratio of length to breadth, and to treat the coefficient as a univariate measure. A second possibility is to derive a measure of distance between varieties based on a multivariate analysis of size measurements. Another method is to overlay the images of samples from two varieties and to compute a goodness of match coefficient.

With the introduction of vision analysis techniques for distinctness testing this subject would seem to warrant more detailed study by the TWC.

f) BIOCHEMICAL TECHNIQUES

Electrophoretic techniques pose special problems. Difference between varieties on the basis of presence or absence of defined bands are clear-cut. However, for some species such differences are not sufficient for separating established varieties. In these cases one must rely on comparing the intensity of individual bands. This can be done on a subjective basis but more objective procedures which take account of variation in the electrophoretic process would seem to be required. Some work on this aspect is currently being undertaken in the U.K.

g) MULTIVARIATE MINIMUM DISTANCE

Current UPOV guidelines on minimum distance require that characters be treated individually. However concern has been expressed that such an approach is weighting the scales in favour of potential plagiarists, particularly for species where there are many positively correlated characters.

One way of tackling this problem is to derive a multivariate measure of distance between all varieties. A measure which combines information from all characters into a single measure is the Mahalonobis D^2 statistic. This statistic allows both for the effects of different scales of measurement for the characters and also for the correlations between characters.

To derive a minimum distance for D^2 in a manner that is analogous to the LSD for a univariate measure, it would be possible to use the information that D^2 follows the F-distribution and to derive confidence limits on this basis.

There are several difficulties to be overcome with the application of the D^2 statistic to establishing protection zones around varieties:

there may already exist established variety pairs which lie within the confidence region;

a species may include umbrella varieties and two levels of protection may be needed - within and between varieties;

the scale on which the D^2 statistic is expressed is not easy to interpret as it is dependent on the number of characters and the number of replicates;

how does one introduce into the statistic an additional character suggested by the breeder as essential for distinguishing two otherwise similar varieties.

However none of these are insuperable difficulties and could be overcome if it was thought appropriate to use a measure which combines information from all measured characters.

CONCLUSION

A number of issues concerning statistical aspects of minimum distance have been touched on in this paper. For some problems solutions are to hand; for some the TWC is working towards a solution; others remain to be addressed. Of the latter the most important may be the application of computational methods in the following areas:

i) shape characters;
ii) electrophoretic data;
iii) multivariate distance measures.

M Talbot
Scottish Agricultural
Statistics Service
Edinburgh
May 1990

[End of document]