E

TGP/8/1 Draft 4
ORIGINAL: English only
DATE: June 8, 2006

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

DRAFT

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]/[TRIAL DESIGN AND TECHNIQUES USED IN THE EXAMINATION OF DISTINCTNESS, UNIFORMITY AND STABILITY]

Document prepared by the Office of the Union

to be considered by the

Technical Working Party for Vegetables (TWV), at its fortieth session, to be held in Guanajuato, Guanajuato State, Mexico, June 12 to 16, 2006

Technical Working Party on Automation and Computer Programs (TWC), at its twenty-fourth session, to be held in Nairobi, Kenya, from June 19 to 22, 2006

Technical Working Party for Agricultural Crops (TWA), at its thirty-fifth session, to be held in Beijing, China, from July 3 to 7, 2006

Technical Working Party for Fruit Crops (TWF), at its thirty-seventh session, to be held in Salvador, Bahia State, Brazil, from August 21 to 25, 2006

Technical Working Party for Ornamental Plants and Forest Trees (TWO), at its thirty-ninth session, to be held in Fortaleza, Ceará State, Brazil, from August 28 to September 1, 2006

TABLE OF CONTENTS PAGE

INTRODUCTION

The purpose of this document is to provide guidance on trial design and data analysis, and to provide information on certain techniques used for the examination of DUS. This document is structured as follows:

PART I: DUS TRIAL DESIGN AND DATA ANALYSIS: this part of the document provides guidance on trial design, data validation, and assumptions to be fulfilled for statistical analysis.

PART II: TECHNIQUES USED IN DUS EXAMINATION: this part of the document provides details on certain techniques referred to in TGP/9 “Examining Distinctness”, and TGP/10 Examining Uniformity.

An overview of the parts of the process of examining distinctness in which trial design and techniques covered in this document are relevant is provided in [the schematic overview of the process of examining distinctness provided in document TGP/9 “Examining Distinctness”, Section 1*[cross ref.]*].

PART I: DUS TRIAL DESIGN AND DATA ANALYSIS

1. REASON FOR USING STATISTICS

~~To be developed by the TWC~~

2. TRIAL DESIGN

2.1. Introduction

2.1.1 The UPOV Convention requires that a variety be examined for compliance with the distinctness, uniformity and stability criteria. The 1991 Act of the UPOV Convention clarifies that, "In the course of the examination, the authority may grow the variety or carry out other necessary tests, cause the growing of the variety or the carrying out of other necessary tests, or take into account the results of growing tests or other trials which have already been carried out."

2.1.2 Guidance for conducting the examination is provided in the Test Guidelines. In that respect the General Introduction states:

> "2.2.1 Where UPOV has established specific Test Guidelines for a particular species, or other group(s) of varieties, these represent an agreed and harmonized approach for the examination of new varieties and, in conjunction with the basic principles contained in the General Introduction, should form the basis of the DUS test.
>
> 2.2.2 Where UPOV has not established individual Test Guidelines relevant to the variety to be examined, the examination should be carried out in accordance with the principles in this document and, in particular, the recommendations contained in Chapter 9, "Conduct of DUS Testing in the Absence of Test Guidelines." In particular, the recommendations in Chapter 9 are based on the approach whereby, in the absence of Test Guidelines, the DUS examiner proceeds in the same general way as if developing new Test Guidelines."
>
> [………]
>
> "The design of the growing trial or other tests, with regard to aspects such as the number of growing cycles, layout of the trial, number of plants to be examined and method of observation, is largely determined by the nature of the variety to be examined. Guidance on design is a key function of the Test Guidelines……"

2.1.3 In addition it is expected that the examiner conducting the tests should understand the objective of the DUS test and have good knowledge of the growing conditions for the species and the factors that can affect the expressions of the characteristics of the variety.

2.1.4 The purpose of Part I "DUS Trial Design and Data Analysis" is to provide guidance relative to DUS trials and data analysis, including guidance in the development and implementation of Test Guidelines.

2.2 Number of growing cycles

2.2.1 Introduction

2.2.1.1 A key consideration with regard to growing trials is to determine the appropriate number of growing cycles. In that respect, document TGP/7, Annex I: TG Template, Section 4.1.2, states:

> "4.1.2 Consistent Differences
>
> The differences observed between varieties may be so clear that more than one growing cycle is not necessary. In addition, in some circumstances, the influence of the environment is not such that more than a single growing cycle is required to provide assurance that the differences observed between varieties are sufficiently consistent. One means of ensuring that a difference in a characteristic, observed in a growing trial, is sufficiently consistent is to examine the characteristic in at least two independent growing cycles."

2.2.1.2 The UPOV Test Guidelines, where available, specify the recommended number of growing cycles. When making the recommendation, the experts drafting the UPOV Test Guidelines take into account factors such as the number of varieties to be compared in the growing trial, the influence of the environment on the expression of the characteristics, and the degree of variation within varieties taking into account the features of propagation of the variety e.g. whether it is a vegetatively propagated, self-pollinated, cross-pollinated or a hybrid variety.

2.2.2 The notion of independent growing cycles

2.2.2.1 As indicated in Section 3.1 *[cross ref.]*, one means of ensuring that a difference in a characteristic, observed in a growing trial, is sufficiently consistent is to examine the characteristic in at least two independent growing cycles. The notion of independence is of particular relevance for the use of statistical procedures. In general, the assessment of independence is based on the experience of experts.

2.2.2.2 When varieties are grown in successive years and the layout of the plants in the trial is randomized (at least partly), the independence of the growing cycles is usually considered to be satisfied.

2.2.2.3 For some perennial crops, for example in perennial ryegrass, the age of the plants may significantly influence the expression of characteristics of varieties in subsequent years. In such cases, it is appropriate to observe two independent growing cycles in the form of two separate plantings. However, in some other perennial crops, for example fruit trees, the two independent growing cycles can be achieved by examining the same plants over two successive years.

2.2.2.4 In the case of plants grown in greenhouses, provided the time between two sowings is not "too short" and the layout of the plants in the trial is randomized (at least partly), two growing cycles can overlap and still be considered as independent.

2.2.2.5 Where two growing cycles are conducted in the same year and at the same time, a suitable distance or a suitable difference in growing conditions between two locations may satisfy the requirement for independence.

2.2.2.6 Where the two growing cycles are in the same location and the same year, a suitable time period between plantings may satisfy the requirement for independence.

2.2.3 Use of multiple locations in the examination of distinctness

Document TGP/7, "Development of Test Guidelines", (see Annex I, TG Template, Section 3.2) clarifies that "Tests are normally conducted at one place". In cases where more than one place is used, the factors below should be taken into account:

2.2.3.1 Purpose

It may be considered appropriate to conduct tests at more than one place for the following purposes:

2.2.3.1.1 Minimizing the overall testing period

More than one location may be used on a routine basis, for example, as a means of achieving more than one independent growing cycle in the same year, as set out in Section 3.2.5 *[cross ref.]*. This could reduce the overall length of the testing period and facilitate a quicker decision.

2.2.3.1.2 Reserve trial

Authorities may designate a primary location, but organize an additional reserve trial in a separate location. In general, only the data from the primary location would be used, but in cases where that location failed, the reserve trial would be available to prevent the loss of one year's results, provided there was no significant variety-by-location interaction.

2.2.3.1.3 Different agro-climatic conditions

Different types of varieties may require different agro-climatic growing conditions. In such cases, the breeder would be required to specify the candidate variety type, to allow the variety to be distributed to the appropriate testing location. Section 3.4 "Additional Tests" *[cross ref.]* addresses the situation where a variety needs to be grown in a particular environment for certain characteristics to be examined, e.g. winter hardiness. However, in such cases each variety will be tested in one location. The possibilities include:

2.2.3.2 Use of information from multiple locations[1]

Where more than one location is used, it is important to establish decision rules with regard to the use of data from the different locations for the assessment of distinctness and for the establishment of variety descriptions. The possibilities include:

(a) distinctness established independently at all growing trial locations

In general, a requirement for distinctness to be established at all growing trial locations would not be appropriate for the purposes set out in paragraph 2.2.3.1

[1] The TC agreed that consideration should be given to moving Section 3.3.2 to Section 5, "Assessing Distinctness Based on the Growing Trial".

(b) distinctness established using characteristics examined at different locations

For example, additional tests (see Section 2.2.3.1.4) [cross ref.] may be carried out to examine particular characteristics e.g. greenhouse tests for disease resistance, laboratory tests for chemical constituents etc.. In such cases, the data for particular characteristics can be obtained at a different location to the main growing trial. In addition, reserve trial data may be available for some or all characteristics which could not be observed in the growing trial at the primary location. In cases where the data for the characteristic(s) is obtained exclusively from the reserve trial, the situation is similar to that for an additional test, although it would be important to record that the variety description for the characteristics concerned was not based on the normal (primary) location. The situation where data from different locations (i.e. the primary location and reserve location) for the same characteristic are combined is covered in paragraph (d).

(c) distinctness established on the basis of data for the same characteristics examined at different locations

In order to minimize the overall testing period where two growing cycles are recommended (see Section 3.3.1.1), a second location might be used to check the consistency of a difference observed in the first location (see Section 3.2.5). Such cases would normally apply where the assessment of distinctness is based on Notes (see Sections 5.2.1.1(b) and 5.2.3) and the assessment of distinctness and the variety description could be considered to be based on the first location. In general, because of the influence of the environment on variety descriptions, it is advisable to produce variety descriptions based on a single location for each characteristic and not to calculate an average across locations.

In cases where the assessment of distinctness is based on statistical analysis of growing trial data obtained in two or more independent growing cycles (see Sections 5.2.1.1(c) and 5.2.4) it might be considered desirable to combine data from different locations, instead of different years, in order to minimize the overall testing period or to be able to use data from a reserve trial. The suitability of such an approach would depend on the features of the crop concerned. In particular, careful consideration would need to be given to check if the necessary assumptions would be satisfied. For example, the COYD criterion was developed for combining data over different years and not for combining data from different locations. In such cases, a decision would also need to be made on whether to develop a variety description based on a single location or all locations.

2.2.3.1.4 Additional Tests

Document TGP/7, "Development of Test Guidelines", explains that, in addition to the main growing trial, additional tests may be established for the examination of relevant characteristics.

2.3 Type of plot for observation

The UPOV Test Guidelines may specify the type/s of plot for the growing trial (e.g. spaced plants, row plot, drilled plot, etc.) in order to examine distinctness as well as uniformity and stability.

2.4 Organizing the growing trial layout

2.4.1 Type of trial layout

The organization of the trial layout is, in the first instance, determined by whether the trial will have replicated plots and whether it will be randomized, or whether it will be organized such that similar varieties are kept together in order to facilitate side-by-side visual comparison in the growing trial. ~~The following sections focus on the situation where the growing trial is to be organized to facilitate side-by-side visual comparison. Information concerning replicated and randomized trial designs is provided in document TGP/8 *[cross ref.]*~~ The type of trial layout will be determined by the approach for the assessment of distinctness.

2.4.2 Approaches for assessment of distinctness

Document TGP/9 "Examining Distinctness, Section 5.2.1 further explains that:

"5.2.1 Introduction

"5.2.1.1 Approaches for assessment of distinctness based on the growing trial can be summarized as follows:

(a) Side-by-side visual comparison in the growing trial;

(b) Assessment by Notes / single variety records ("Notes"): the assessment of distinctness is based on the recorded state of expression of the variety for a characteristic;

(c) Statistical analysis of growing trial data: the assessment of distinctness is based on a statistical analysis of the data obtained from the growing trial. This approach requires that, for a characteristic, there are a sufficient number of records for a variety.

"5.2.1.2 The choice of approach for the assessment of distinctness will depend on the method of observation and type of record (VG, MG, VS or MS), which is influenced by the features of propagation of the variety and the type of expression of the characteristic. The common situations are summarized by the table in Section 4.5 *[cross ref.]*. The purpose of the following sections is to consider how the assessment of distinctness is conducted for those different situations."

2.4.2.1 Side by side (visual) comparison

TGP/9 explains the following:

"5.2.2.1 Side-by-side visual comparison means that the assessment of distinctness is based on a direct visual comparison of varieties, side-by-side in the growing trial. This approach requires that the characteristics can be observed visually and indicates that the expression of the characteristic for a variety can be represented by a single record. It also requires that all similar varieties can be the subject of a direct side-by-side comparison in the growing trial. Such a requirement can be difficult to meet if the growing trial contains a large number of varieties and there are limited possibilities for ensuring that all similar varieties are grouped together in the growing trial.

[………..]

"5.2.2.3 In the case of vegetatively propagated and self-pollinated varieties, there is relatively little variation within varieties and visual assessment of distinctness is particularly suitable. However, where the range of variation within a variety is larger, because of the features of its propagation, and in particular for cross-pollinated and some types of hybrid varieties, determining distinctness on the basis of side-by-side visual comparison would require particular care."

2.4.2.2 Assessment of distinctness by notes/single variety records.

TGP/9 Section 5.2.3 explains the following:

"5.2.3.1 Assessment by Notes / single variety records means that, for a particular characteristic, the assessment of distinctness is based on the recorded state of expression of a variety, obtained from the growing trial. The record may, for example, be in the form of: a Note corresponding to a state of expression in the UPOV Test Guidelines (e.g. 1, 2, 3 etc.); a value (e.g. RHS Colour Chart reference number); a measurement (e.g. length (cm), weight (g), date (18-12-2005), count (3) etc.); an image etc.. The Notes / single variety records approach can be used for characteristics which are visually observed or measured, but requires that the expression of the characteristic for a variety can be represented by a single record for the purpose of the assessment of distinctness (VG, MG, mean of MS, mean of VS).

"5.2.3.2 Where the requirements for distinctness assessment by Notes / single variety records are met it would usually also be possible to make a side-by-side visual comparison. However, in the case of assessment by Notes / single variety records, such proximity is not required, which is a particular advantage where the growing trial contains a large number of varieties and where there are limited possibilities for ensuring that all similar varieties are grouped together in the growing trial. On the other hand, because the varieties are not the subject of a side-by-side visual comparison, the difference required between varieties as a basis for distinctness is, with the exception of qualitative characteristics (see below), somewhat greater."

2.4.2.3 Assessment by statistical analysis of growing trial data

TGP/9, Section explains:

"5.2.4.1 Where appropriate, the assessment of distinctness can be based on a statistical analysis of the data obtained from the growing trial. This approach requires that there is a sufficient number of records for a variety, e.g. records for a number of single, individual plants or parts of plants, whether obtained by measurement (MS) or by visual observation (VS). In most cases, when a single record is obtained by visual observation or measurement of a group of plants (VG / MG), this results in a single record per variety, in which case it is not possible or necessary to apply statistical methods for the assessment of distinctness. However, in some cases, e.g. where there are several repetitions or plots, or more than one growing trial, more than one record per variety may be obtained, in which case statistical methods can be applied, although it is particularly relevant to check if the data obtained meets the assumptions required for a statistical procedure to be applied.

"5.2.4.2 The assessment of distinctness by Notes / single variety records or side-by-side visual comparison is generally quicker and cheaper than the use of statistical analysis. However, as explained above, those approaches require that the expression of the characteristic for a variety can be represented by a single record. That requirement implies that there should be very little variation within varieties, which is usually met for all characteristics of vegetatively propagated varieties and self-pollinated varieties and for qualitative and pseudo-qualitative characteristics for cross-pollinated and hybrid varieties, except in cases of segregating characteristics. Thus, the most common use of statistical analysis of growing trial data is for quantitative characteristics of cross-pollinated and some hybrid varieties. ”

2.5 Trial Elements

The two most important aspects of deciding on a trial layout are the random allocation of varieties to plots, and the control of local variation in conditions, which might otherwise influence the results of the trials. These issues are discussed in the following sections.

2.5.1 Plots and the allocation of varieties to plots

2.5.1.1 A plot is the experimental unit to which the varieties are allocated. A plot contains plants from the same variety. Depending on the type of growing trial, a plot may be an area of land, or a group of plots.

2.5.1.2 In cases other than for some side-by-side comparisons, varieties must be allocated to plots at random. It ensures that there is no subjectivity in the allocation of varieties to plots.

2.5.1.3 There are further advantages of randomization if there are replications of plots or more than one growing trial, and if variety means are to be calculated, such as when distinctness is assessed by statistical analysis of growing trial data. Random allocation ensures that on average the effects of other factors influencing the plants’ characteristics, such as soil conditions, are expected to cancel out when the variety means are compared.

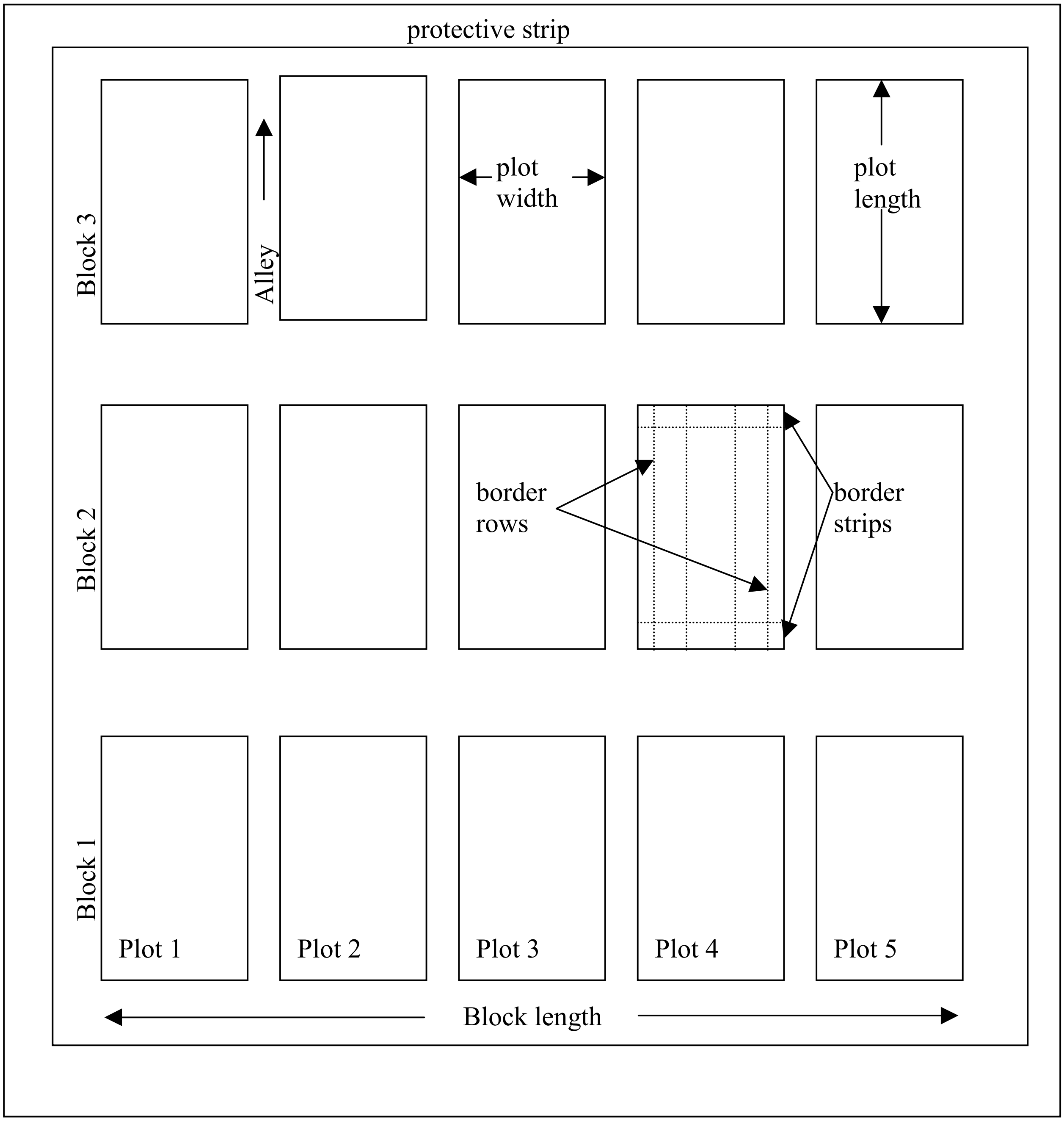
2.5.1.4 A block is a group of plots within which the varieties are either allocated at random, or, in the case of some side-by side comparisons (see section 2.6.5)*[cross ref.]*, pair-wise randomly allocated. A growing trial may contain just one block or it may contain more than one block.

2.5.2 Plot size, shape and configuration

2.5.2.1 In deciding on trial layout, it important that local variation in conditions are controlled. For this decisions on the following are needed.

- plot size
- shape of the plots
- alignment of the plots
- barrier rows and border strips and
- protective strips

2.5.2.2 The following figure may be helpful to give some explanations of the particular trial elements.



2.5.2.3 For the assessment of distinctness unbiased observation of characteristics are necessary. In some cases it is necessary to have border rows and strips to minimize bias caused by inter-plot interference, i.e interference between plants on different plots, and other special border effects, such as shading and soil moisture. Also, protective strips on the border of the trial are often used to reduce the chance of external influences biasing one plot in favour of another. When observing characteristics on the plants on a plot it is usual to exclude the plot's border rows and border strips.

2.5.2.4 The Test Guidelines indicate the type of record required for the assessment of distinctness and uniformity (single record for a group of plants or parts of plants (G), or records for a number of single individual plants or parts of plants (S)). These will determine the sample size, i.e. the number of plants which must be observed, and hence determine the minimum effective size of the plot. To decide on the actual plot size, allowance must be made for any necessary border rows and strips.

2.5.2.5 The plot size and the plot shape also depend on the soil and other conditions, irrigation equipment, or on the sowing and harvesting machinery. The shape of the plot can be defined as the ratio of plot length divided by plot width. This ratio can be important to mitigate variation in conditions within the block (e.g. caused by soil variation).

2.5.2.6 Square plots have the smallest total length of the borders (circumference). From the theoretical point of view the square shape is optimal to minimize the interference of different phenotypes. Grouping the varieties can also help minimize this interference.

2.5.2.7 Narrow and long plots are preferred from the technological point of view. The best length to width ratio lies between 5:1 and 15:1 and depends on the plot size and the number of varieties. The larger the number of varieties in a block the narrower the plots - but not so narrow that the inter-plot competition becomes a problem.

2.5.3 Independence of plots

2.5.3.1 One of the most important requirements of experimental units is independence. This is particularly important when distinctness and uniformity are to be assessed by statistical analysis of the growing trial data.

2.5.3.2 Independence of plots means that observations made on a plot are not influenced by the circumstances in other plots. For example, if tall varieties are planted next to short ones there could be a negative influence of the tall ones on the short ones and a positive influence in the other direction. In such a case, in order to avoid this dependency an additional row of plants can be planted on both sides of the plot, i.e. border rows and strips. Another possibility to minimize this influence is to group similar varieties together. Where grouping is used it should be done according to the principles set out in document TGP/9 Section 3*[cross ref.*]

2.5.4 The arrangement of the plants within the plot

The Test Guidelines indicate the arrangement of the plants within the plot. This may be:

- Rows of plants: This type of arrangement is used for many self-pollinated species, such as cereals. Most characteristics are assessed in an overall observation – usually using the notes stated in the Test Guidelines. In some cases it may be necessary to remove some plants from the plot in order to record some characteristics; and in that case the size of the plot should allow the removal of plants without prejudicing the observations which must be made up to the end of the growing cycle including the assessment of uniformity (see document TGP/7, ASW 6 *[cross ref.]*).
- Ear rows: This type of arrangement is frequently used for the assessment of uniformity in self-pollinated varieties.
- Spaced plants: This type of arrangement is used in many cross-pollinated and vegetatively propagated varieties.

2.6 Aspects of trial design relevant to when statistical analysis will be used

This section describes a number of concepts that are relevant when designing growing trials for which distinctness and uniformity are to be assessed by statistical analysis of the growing trial data. Firstly, if there are to be replicate plots of each variety in the growing trial, decisions must be made as to whether the replicate plots should be grouped into blocks and how the plots should be aligned within a block, i.e. the Experimental Design. This determines how local, unwanted or nuisance variation is controlled and hence how precisely distinctness and uniformity can be assessed. Then there is the notion that variation arises from different sources, and how this can affect the choice of sample sizes, which again impacts on precision. Precision is important because it in turn impacts on the decision making. If data are relatively imprecise and decisions are based on this data, there is an appreciable chance that inappropriate or wrong decisions will get made. This is discussed below in terms of the hypotheses being tested, and chosen between, when decisions are made.

2.6.1 The hypotheses under test

2.6.1.1 When statistical analysis of growing trial data is to be used to assess distinctness and uniformity, the purpose of the growing trial is to get precise and unbiased averages of characteristics for each variety and also to judge the within-variety variability by calculating the standard deviation. Decisions about the distinctness of varieties are made based on the characteristic averages. Decisions about the uniformity of a variety are based on the standard deviations in the case of quantitative characteristics, and in the case of qualitative characteristics on the number of off-types present in a sample.

2.6.1.2 In making each of these decisions we test a Null Hypothesis and either accept or reject it. If we reject it, we accept an Alternative Hypothesis. The Null and Alternative Hypotheses for the Distinctness and Uniformity decisions are given in the following table:

	Null hypothesis (H0)	Alternative Hypothesis (H1)
Distinctness	two varieties are not distinct for the characteristic	two varieties are distinct
Uniformity	a variety is uniform for the characteristic	a variety is not uniform

2.6.1.3 We make each decision by computing a test statistic from the observations using a formula. If the test statistic is greater than its chosen critical value, the null hypothesis H0 is rejected, the alternative hypothesis H1 is accepted, and the test is called significant. If the test statistic is not greater than its chosen critical value, the null hypothesis H0 is accepted. The choice of the critical value that the test statistic is compared with is explained below.

2.6.1.4 In making a decision based on a test statistic, because it is a test statistic based on a sample and hence subject to variability, there is a chance that the wrong conclusion is arrived at. Such “statistical errors” can occur in two ways, let us first consider distinctness decisions :-

- The decision based on the test statistic, i.e. from the DUS trial, is that two varieties are distinct, when in reality, i.e. if all plants of the two varieties were examined, they are not distinct. This is known as a Type I error and its risk is denoted by α.

- The decision based on the test statistic, i.e. from the DUS trial, is that two varieties are not distinct, when in reality, i.e. if all plants of the two varieties were examined, they are distinct. This is known as a Type II error and its risk is denoted by β.

2.6.1.5 The two types of statistical error that can be made when testing for distinctness are shown in the following table:

	Decision based on test statistic	
Situation in Reality	Varieties are not distinct (H0 true)	Varieties are distinct (H1 true)
Varieties are not distinct (H0 true)	Correct decision	Incorrect decision, Type I error, made with probability α
Varieties are distinct (H1 true)	Incorrect decision, Type II error, made with probability β	Correct decision

2.6.1.6 Likewise, it is possible when deciding on uniformity based on a test statistic, i.e. from the DUS trial, to decide that a variety is not uniform, when in reality if all plants of the variety were examined, it is uniform, i.e. a Type I error (α). Alternatively, a Type II error (β) is the decision that a variety is uniform, when in reality it is not uniform. The following table shows the two types of statistical error that can be made when testing for uniformity:

	Decision based on test statistic	
Situation in Reality	Varieties are uniform (H0 true)	Varieties are not uniform (H1 true)
Varieties are uniform (H0 true)	Correct decision	Incorrect decision, Type I error, made with probability α
Varieties are not uniform (H1 true)	Incorrect decision, Type II error, made with probability β	Correct decision

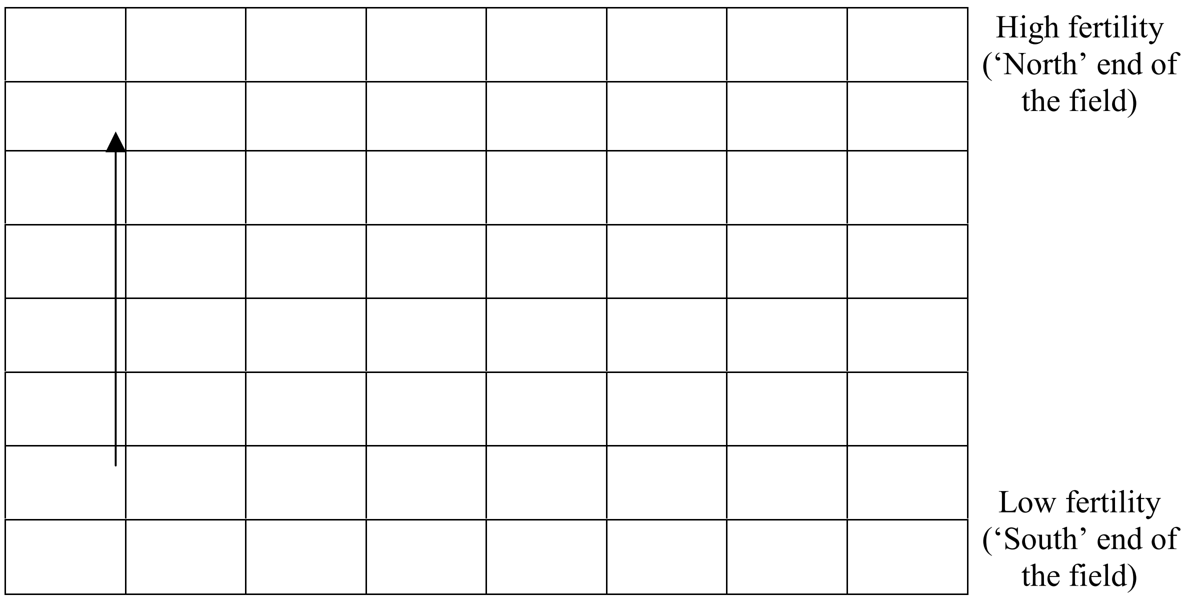
2.6.1.7 The risk of making a type I error can be controlled easily by choice of α, which determines the critical value that the test statistic is compared against. α is also known as the size of the test. The risk of making a type II error is more difficult to control as it depends, for example in the case of distinctness, on the size of the real difference between the varieties, the chosen α, and the precision of the test in terms of the number of replicates and the random variability. The Crop Expert can reduce the risk of making a type II error by increasing the precision by increasing the number of replicates and reducing the random variability by choice of number of plants per plot (or sample size), and by controlling local, unwanted or nuisance variation through careful choice of experimental design.

2.6.2 Sources of variation

When the same variety is assigned to a number of different plots, the observations on the different plots may vary. The variation between these observations is called the 'between-plot variability'. This variability is a mixture of different sources of variation: different plots, different plants, different times of observation, different errors of measurement and so on. It is not possible to distinguish between these sources of variation. When there are observations of more than one, say n, plants per plot it is possible to compute two variance components: the “within-plot” or “plant” component and the “plot” component.

2.6.3 Completely randomized design and randomized complete block design

2.6.3.1 In designing an experiment it is important to choose an area of land that is as homogeneous as possible in order to minimize the variation between plots of the same variety, i.e. the random variation. Assume that we have a field where it is known that the largest variability is in the ‘north-south’ direction, e.g. as in the following figure:



2.6.3.2 Let’s take an example where four varieties are to be compared with each other in an experiment within this field where each of the varieties is assigned to 4 different plots. It is important to randomize the varieties over the plots. If varieties are arranged systematically, not all varieties would necessarily be under the same conditions (see following figure).

Variety A	Variety A	Variety A	Variety A	Variety B	Variety B	Variety B	Variety B	Higher fertility row
Variety C	Variety C	Variety C	Variety C	Variety D	Variety D	Variety D	Variety D	Lower fertility row

If the fertility of the soil decreases from the north to the south of the field, the plants of variety A and B have grown on more fertile plots than the other varieties. The comparison of the varieties is influenced by a difference in fertility of the plots. Differences between varieties are said to be confounded with differences in fertility.

2.6.3.2 To avoid systematic errors it is advisable to randomize varieties across the site. A complete randomization of the four varieties over the sixteen plots could have resulted in the following layout:

Variety C	Variety A	Variety A	Variety B	Variety C	Variety D	Variety B	Variety C	Higher fertility row
Variety C	Variety A	Variety D	Variety A	Variety D	Variety B	Variety D	Variety B	Lower fertility row

2.6.3.3 However, looking at the design we find that variety C occurs three times in the top row (with high fertility) and only once in the second row (with lower fertility). For variety D we have the opposite situation. Because we know that there is a fertility gradient, this is still not a good design, but it is better than the first systematic design.

2.6.3.3 When we know that there are certain systematic sources of variation like the fertility gradient in the paragraphs before, we may take that information into account by making so-called blocks. The blocks should be formed so that the plots within each block are as homogeneous as possible. With the assumed gradients we may choose either two blocks each consisting of one row or we may choose four blocks – two blocks in each row with four plots each. In larger trials (more plots) the latter will most often be the best, as there will also be some variation within rows even though the largest gradient is between rows.

Block I				Block II				
Variety A	Variety C	Variety D	Variety B	Variety A	Variety C	Variety D	Variety B	Higher fertility row
Variety B	Variety C	Variety A	Variety D	Variety C	Variety A	Variety D	Variety B	Lower fertility row
Block III				Block IV				

An alternative way of reducing the effect of any gradient between the columns is to use plots that are half the width, but which extend over two rows, i.e. by using long and narrow plots:

Block I				Block II				Block III				Block IV			
Var A	Var C	Var D	Var B	Var A	Var C	Var D	Var B	Var B	Var C	Var A	Var D	Var C	Var A	Var D	Var B

In both designs above the 'north-south' variability will not affect the comparisons between varieties.

2.6.3.4 In a randomized complete block design the number of plots per block equals the number of varieties. All varieties are present once in each block and the order of the varieties within each block is randomized. The advantage of a randomized complete block design is that the standard deviation between plots (varieties), a measure of the random variation , does not contain variation due to differences between blocks. The main reason for the random allocation is that it ensures that the results are unbiased and so represent the varieties being compared. In other words, the variety means will, on-average, reflect the true variety effects, and will not be inflated or deflated by having been allocated to inherently better or worse plots. An interesting feature of the randomization is that it makes the observations from individual plots 'behave' as independent observations (even though they may not be so). There is usually no extra cost associated with blocking, so it is recommended to arrange the plots in blocks.

2.6.3.5 Blocking is introduced here on the basis of differences in fertility. Several other systematic sources of variation could have been used as the basis for blocking. Although it is not always clear how heterogeneous the field is, and therefore it is unknown how to arrange the blocks, it is usually a good idea to create blocks for other reasons. When there are different sowing machines, different observers, different observation days, such effects are included in the residual standard deviation if they are randomly assigned to the plots. However, these effects can be eliminated from the residual standard deviation if all the plots within each block have the same sowing machine, the same observer, the same observation day, and so on.

2.6.3.6 Management may influence the choice of the form of the plots. In some crops it may be easier to handle long and narrow plots than square plots. Long narrow plots are usually considered to be more influenced by varieties in adjacent plots than square plots. The size of the plots should be chosen in such a way that the necessary number of plants for sampling is available. For some crops it may be necessary also to have guard plants (areas) in order to avoid large competition effects. However, overly large plots require more land and will often increase the random variability between plots. Grouping of the varieties according to e.g. height may also reduce the competition between adjacent plots. If nothing is known about the fertility of the area, then layouts with compact blocks (i.e. almost square blocks) will often be most appropriate because the larger the distance between two plots the more different they will usually be. In both designs above, the blocks can be placed as shown or they could be placed 'on top of each other' (see following figure). This will usually not change the variability between plots considerably – unless one of the layouts, forces the crop expert to use more heterogeneous soil.

Variety A	Variety C	Variety D	Variety B	Block I	Higher fertility row
Variety A	Variety C	Variety D	Variety B	Block II	
Variety B	Variety C	Variety A	Variety D	Block III	
Variety C	Variety A	Variety D	Variety B	Block IV	Lower fertility row

2.6.4 Randomized incomplete block designs

2.6.4.1 If the number of varieties becomes very large (>20-40), it may be impossible to construct complete blocks that would be sufficiently homogeneous. In that case it might be advantageous to form smaller blocks, each one containing only a fraction of the total number of varieties. Such designs are called incomplete block designs. Several types of incomplete block designs can be found in the literature for example, balanced incomplete block designs and unbalanced incomplete block designs such as Lattice designs and Row and column designs. One of the most familiar types for variety trials is a lattice design. The generalized lattice designs (also called α-designs) are very flexible and can be constructed for any number of varieties and for a large range of block sizes and number of replicates. One of the features of generalized lattice designs is that some of the incomplete blocks can be (and usually are) collected to form a whole replicate. This means that such designs will be at least as good as randomized complete block designs, since the analysis can be performed using either a lattice model or a randomized complete block model. The lattice model should be preferred if conditions are fulfilled.

2.6.4.2 Incomplete blocks need to be constructed in such a way that it is possible to compare all varieties in an efficient way. An example of an α-design is shown in the following figure:

Block	Sub-block	Variety			
3	5	6	5	15	19
	4	13	8	10	20
	3	2	3	4	7
	2	12	1	18	14
	1	17	11	16	9
Block	Sub-block	Variety			
2	5	4	16	6	1
	4	18	5	10	2
	3	14	7	17	8
	2	11	19	13	3
	1	15	9	20	12
Block	Sub-block	Variety			
1	5	4	20	5	17
	4	2	13	1	9
	3	3	6	12	8
	2	18	7	11	15
	1	16	10	14	19

In the example above, 20 varieties are to be grown in a trial with three replicates. In the design the 5 sub-blocks of each block form a complete replicate. Thus each replicate contains all varieties whereas any pair of varieties occurs either once or not at all in the same subblock.

2.6.4.3 The incomplete block design is most suitable for trials where grouping characteristics are not available. If grouping characteristics are available then some modification may be advantageous for trials with many varieties, such as using grouping characteristics to form separate trials rather than a single trial, see document TGP/9 Section 3.6.2.1 Grouping characteristics.

2.6.5 Design for pair-wise comparisons between particular varieties

2.6.5.1 When a close comparison is needed between a pair of varieties by means of statistical analysis, it may be good to grow them in neighbouring plots. A similar theory to that used in split-plot designs may be used for setting up a design where the comparisons between certain pairs of varieties are to be optimized. When setting up the design, the pairs of varieties are treated as the whole plot factor and the comparison between varieties within each pair is the sub-plot factor. As each whole plot consists of only two sub-plots, the comparisons within pairs will be (much) more precise than if a randomized block design was used.

2.6.5.2 If, for example, four pairs of varieties (A-B, C-D, E-F and G-H) have to be compared very precisely, then this can be done using the following design of 12 whole plots each having 2 subplots:

Pair 1 variety A	Pair 3 variety E	Pair 4 variety H
Pair 1 variety B	Pair 3 variety F	Pair 4 variety G
Pair 3 variety F	Pair 2 variety D	Pair 1 variety A
Pair 3 variety E	Pair 2 variety C	Pair 1 variety B
Pair 4 variety G	Pair 1 variety B	Pair 2 variety C
Pair 4 variety H	Pair 1 variety A	Pair 2 variety D
Pair 2 variety D	Pair 4 variety H	Pair 3 variety E
Pair 2 variety C	Pair 4 variety G	Pair 3 variety F

In this design each column represents a replicate. Each of these is then divided into four incomplete blocks (whole plots) each consisting of two (sub)plots. The four pairs of varieties are randomized to the incomplete blocks within each replicate and the order of varieties are randomized within each incomplete block. The comparison between varieties of the same pair is made more precise at the cost of the precision of the comparison between varieties of a different pair.

2.6.6 The effect of sample size on precision and decision making

2.6.6.1 The Test Guidelines will usually define the sample size of one experiment. However, the precision of a test does not depend on sample size alone. The final precision of a test based on the observations of one experiment depends for quantitative characteristics on at least three sources of variation:

- the variation between individual plants within a plot
- the variation between the plots within a block
- the variation caused by the environment, i.e. the variation in the expression of characteristics from year to year (or from location to location)

2.6.6.2 To estimate the optimal sample size when developing new Test Guidelines it is necessary to know the standard deviations of the above sources of variation, expected differences between the varieties which should be significant, the number of varieties and the number of blocks in the trial. Additionally, the crop expert has to determine the type I (α) and type II error (β). In cooperation with a statistician the crop expert can compute the optimal sample size for some characteristics and then he can determine the optimal sample size for this trial for all characteristics. Especially for the assessment of uniformity, the type II error is sometimes more important than the type I error. In some cases the type II error could be greater than 50 % and becomes unacceptable.

2.6.7 The impact of precision on analyses over years or cycles

2.6.7.1 The comparison between varieties may be based on observations from two to three years or cycles. Therefore, the number of replicates and the number of plants per plot in a single trial have some effect on the variability which is used to test distinctness and uniformity in the over-year or over-cycle statistical analyses (see Part II: Sections 2.1 and 2.2 [cross ref.]). Before performing these analyses the means of the variety means and (log) standard deviations per year or cycle are calculated and then the analysis is performed on these means in the two-way variety-by-year or variety-by-cycle layout. The residual variation in these analyses is the variety-by-year or variety-by-cycle interaction.

2.6.7.2 The precision of the variety means in one year's or one cycle's experiment depend on the number of replicates, the number of plants per plot, and the Experimental design. When these means are used in the over-year or over-cycle analysis for COY-D for example, their precision is only of benefit indirectly, because the standard deviation in that analysis is based on the interaction between the varieties and the years or cycles. Further, if the differences between the varieties over the years or cycles are very large, the precision of the means per experiment are relatively unimportant.

3. VALIDATION OF DATA AND ASSUMPTIONS

[TWC Chairperson: to be reviewed to clarify that aspects of this section may be relevant when statistics are not used]

3.1 Introduction

3.1.1 When data are observed on plots in the growing trial it is important that the data are correct. This is the case whether the data are notes, single variety records, or data for the assessment of distinctness and uniformity by statistical analysis. The first of the following sections describes how the data can be validated or checked. These preliminary checks can be done on all data, whether or not they are subsequently analysed by statistical methods.

3.1.2 If the data are to be statistical analysed, the assumptions behind the theory on which the statistical methods are based must be met - at least approximately. The second of the following sections describes the assumptions behind the most common statistical analysis methods used in DUS testing. The third of the following sections is on the validation of assumptions, and describes how these assumptions may be evaluated. Because mistakes in the data effectively negate the assumptions behind the statistical analysis, the methods used to validate the assumptions can often also serve to identify mistakes in the data that were not identified in the initial validation of the data.

3.1.3 The assumptions and methods of validation described here are for the analyses of single experiments (randomized blocks). However, the principles are the same when analyzing data from several experiments over years. Instead of plot means, the analyses are then carried out on variety means per year (and blocks then become equivalent to years). Thus the methods of validation can be used with the COYD and COYU analyses for quantitative characteristics, which are over-year analyses based on variety means per year for COYD, and variety means of the (logarithm of the) between-plants standard deviation per year for COYU.

Throughout this section data of 'Leaf: Length' (in mm) are used from an experiment laid out in 3 blocks of 26 plots with 20 plants per plot. Within each block, 26 different oilseed rape varieties were randomly assigned to each plot.

3.2 Validation of data

3.2.1 In order to avoid mistakes in the interpretation of the results the data should always be inspected so that the data are logically consistent and not in conflict with prior information about the ranges likely to arise for the various characteristics. This inspection can be done manually (usually visually) or automatically.

3.2.2 Table 1 shows an extract of some recordings for 10 plants from a plot of field peas. For 'Seed: shape' the notes are visually scored on a scale with values 1, 2, 3, 4, 5 or 6. For 'Stem: length' the measurements are in cm and from past experience it is known that the length in most cases will be between 40 and 80 cm. The 'Stipule: length' is measured in mm and will in most cases be between 50 and 90 mm. The table shows 3 types of mistakes which occasionally occur when making manual recordings: for plant 4, 'Seed: shape' the recorded value, 7, is not among the allowed notes and must, therefore, be due to a mistake. It might be caused by a misreading a hand-written "1". The 'Stem: length' of plant 6 is outside the expected range and could be caused by changing the order of the figures, so 96 has been keyed instead of 69. The 'Stipule: length' of 668 mm is clearly wrong. It might be caused by accidentally repeating the figure 6 twice. In all cases a careful examination needs to be carried out in order to find out what the correct values should be.

Table 1 Extract of recording sheet for field peas

Plant no	Seed: shape (UPOV 1)	Stem: length (UPOV 12)	Stipule: length (UPOV 31)
1	1	43	80
2	2	53	79
3	1	50	72
4	7	43	668
5	2	69	72
6	1	96	72
7	1	51	70
8	2	64	63
9	1	44	62
10	2	49	62

3.2.3 Examination of frequency distributions of the characteristics to look for small groups of discrepant observations.

3.2.4 Examination of scatter plots of pairs of characteristics likely to be highly related. This may often detect discrepant observations very efficiently.

3.2.5 Other types of [plot] may also be used to validate the quality of the data. A so-called Box-plot is an efficient way to get an overview of the data. In a Box-plot a box is drawn for each group (plot or variety). In Figure 1, all 60 Leaf Lengths of each of the 26 varieties are taken together. (If there are large block differences a better Box-plot can be produced by taking the differences with respect to the plot mean). The box shows the range for the largest part of the individual observations (usually 75%). A horizontal line through the box and a symbol indicates the median and mean, respectively. At each end of the box, vertical lines are drawn to indicate the range of possible observations outside the box, but within a reasonable distance (usually 1.5 times the height of the box). Finally, more extreme observations are shown individually. In Figure 1, it is seen that one observation of variety 13 is clearly much larger than the remaining observations of that variety. Also it is seen that variety 16 has large leaf lengths and that about 4 observations are relatively far from the mean. Among other things that can be seen from the figure are the variability and the symmetry of the distribution. So it can be seen that the variability of variety 15 is relatively large and that the distribution is slightly skewed for this variety (as the mean and median are relatively far apart).

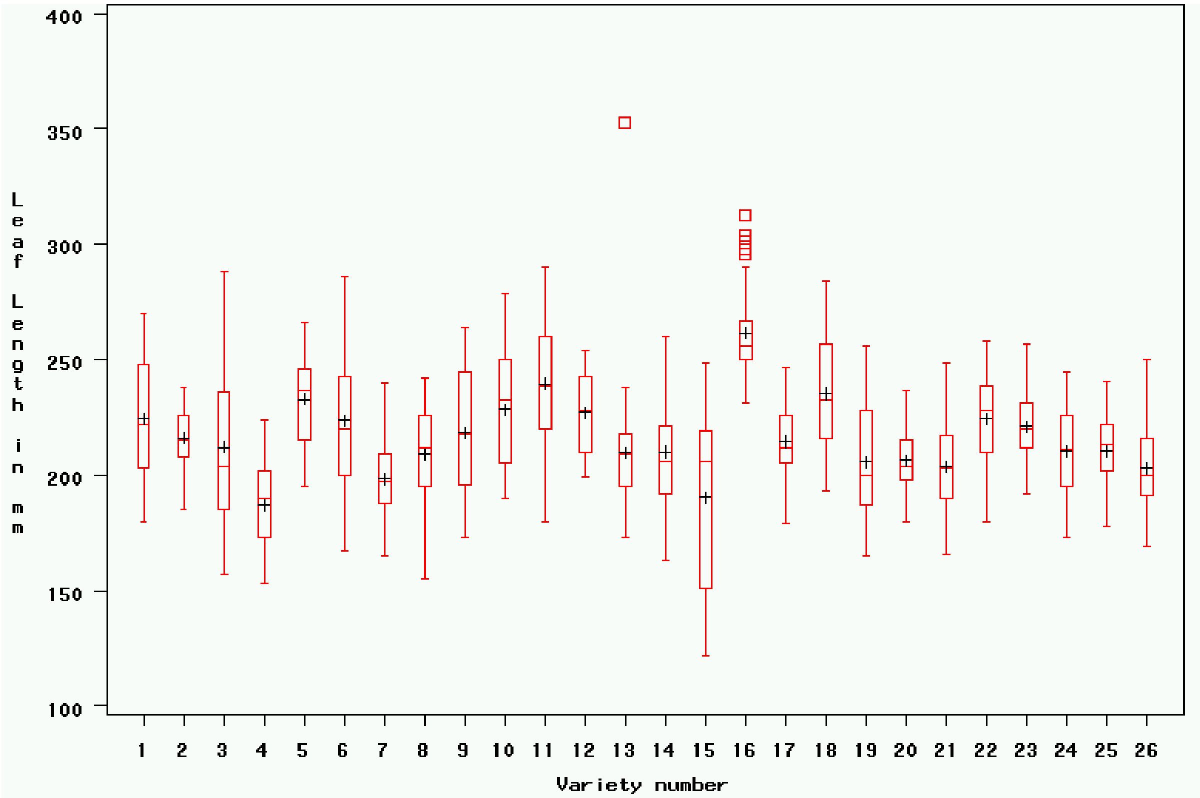


Figure 1. Box-plot for Leaf Length of 26 varieties of oil seed rape

3.2.6 When discrepant observations are found, it is important to try to find out why the observations are deviating. In some cases it may be possible to go back to the field and to check if the plant or plot is damaged by external factors (e.g. rabbits) or a measurement mistake has occurred. In the latter case a correction is possible. In other cases, it may be necessary to look in previous notes (or on other measurements from the same plant/plot) in order to find the reason for the discrepant observation. Generally observations should only be removed when there are good reasons.

3.3 Assumptions

3.3.1 Introduction

3.3.1.1 Firstly, it is essential that the growing trial/experiment is designed properly and involve randomisation. The most important assumptions of analysis of variance methods are:

independent observations

variance homogeneity
additivity of block and variety effects for a randomized block design and additivity of year and variety effects for COYD
normally distributed observations (residuals)

3.3.1.2 In addition, one could state that there should be no mistakes in the data. However, most mistakes (at least the largest) will usually also mean that the above assumptions are not met, as the observations are not normally distributed and they have different variances (non-homogeneity of variances).

3.3.1.3 The assumptions mentioned here are most important when the statistical methods are used to test hypotheses. When statistical methods are used only to estimate effects (means), the assumptions are less important and the assumption of normal distributed observations is not necessary.

3.3.2 Independent observations

This is a very important assumption. It means that no records may depend on other records in the same analysis (dependence between observations may be built into the model, but has not been built into COYD and COYU or the other methods included in TGP/8). Dependency may be caused by e.g. competition between neighboring plots, lack of randomisation or improper randomisation. More details on ensuring independence of observations may be found in Part I: Section 2.1 *[cross ref.]* "Experimental Design Practices."

3.3.3 Variance homogeneity

Variance homogeneity means that the variance of all observations should be identical apart from random variation. Typical deviations from the assumption of variance homogeneity fall most often into one of the following two groups:

The variance depends on the mean, e.g. the larger the mean value the larger the standard deviation is. In this case the data may often be transformed such that the variances on the transformed scale may be approximately homogeneous. Some typical transformations of characteristics are: the logarithmic transformation (where the standard deviation is approximately proportional to the mean), the square-root transformation (where the variance is approximately proportional to the mean, e.g. counts), and the angular transformation (where the variance is low at both ends of the scale and higher in between, typical for percentages).

The variance depends on for example, variety, year or block. If the variances depend on such variables in a way that is not connected to the mean value, it is not possible to obtain variance homogeneity by transformation. In such cases it might be necessary either to use more sophisticated statistical methods that can take unequal variances into account or to exclude the group of observations with deviant variances (if only a few observations have deviant variances). To illustrate the seriousness of variance heterogeneity: imagine a trial with 10 varieties where varieties A, B, C, D, E, F, G and H each have a variance of 5, whereas varieties I and J each have a variance of 10. The real probability of detecting differences between these varieties when, in fact, they have the same mean is shown in Table 2. In Table 2, the variety comparisons are based on the pooled variance as is normal in traditional ANOVA. If they are compared using the 1% level of significance, the probability that the two varieties with a variance of 10 become significantly different from each other is almost 5 times larger (4.6%) than it should be. On the other hand, the probability of significant differences between two varieties with a variance of 5 decreases to 0.5%, when it should be 1%. This means that it becomes too difficult to detect differences between two varieties with small variances and too easy to detect differences between varieties with large variances.

Table 2. Real probability of significant difference between two identical varieties in the case where variance homogeneity is assumed but not fulfilled (varieties A to H have a variance of 5 and varieties I and J have a variance of 10.)

Comparisons, variety names	Formal test of significance level	
	1%	5%
A and B	0.5%	3.2%
A and I	2.1%	8.0%
I and J	4.6%	12.9%

3.3.4 Normal distributed observations

The residuals should be approximately normally distributed. The ideal normal distribution means that the distribution of the data is symmetric around the mean value and with the characteristic bell-shaped form (see Figure 2). If the residuals are not approximately normally distributed, the actual level of significance may deviate from the nominal level. The deviation may be in both directions depending on the way the actual distribution of the residuals deviates from the normal distribution. However, deviation from normality is usually not as serious as deviations from the previous two assumptions.

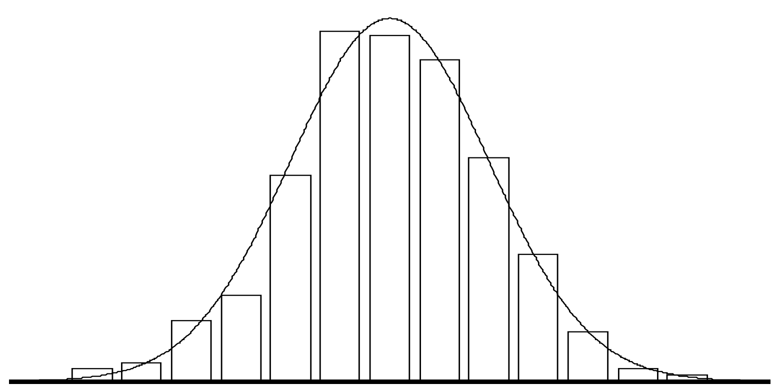


Figure 2. Histogram for normal distributed data with the ideal normal distribution shown as a curve

3.3.5 Additivity of block and variety effects

3.3.5.1 The effects of blocks and varieties are assumed to be additive because the error term is the sum of random variation and the interaction between block and variety. This means that the effect of a given variety is the same in all blocks. This is demonstrated in Table 3 where plot means of artificial data (of Leaf Length in mm) are given for two small experiments with three blocks and four varieties. In experiment I, the effects of blocks and varieties are additive because the differences between any two varieties are the same in all blocks, e.g. the differences between variety A and B are 4 mm in all three blocks. In experiment II, the effects are not additive, e.g. the differences between variety A and B are 2, 2 and 8 mm in the three blocks.

Table 3. Artificial plot means of Leaf Length in mm from two experiments showing additive block and variety effects (left) and non-additive block and variety effects (right)

Experiment I			
Variety	Block		
	1	2	3
A	240	242	239
B	244	246	243
C	245	247	244
D	241	243	240

Experiment II			
Variety	Block		
	1	2	3
A	240	242	239
B	242	244	247
C	246	244	243
D	241	242	241

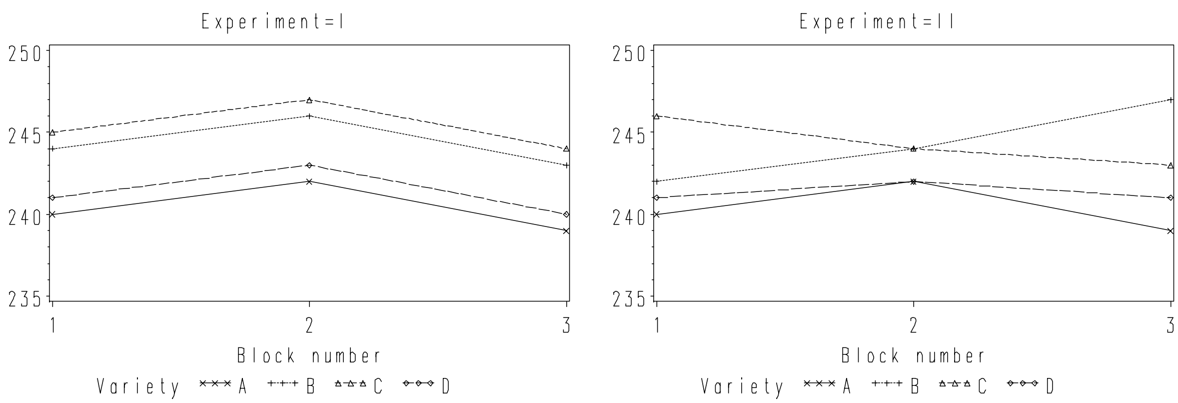


Figure 3. Artificial plot means from two experiments showing additive block and variety effects (left) and non-additive block and variety effects (right) using same data as in table 2

3.3.5.2 In Figure 3 the same data are presented graphically. Plotting the means versus block numbers and joining the observations from the same varieties by straight lines produces the graphs. Plotting the means versus variety names and joining the observations from the same blocks could also have been used (and may be preferred especially if many varieties are to be shown in the same figure). The assumption on additivity is fulfilled if the lines for the varieties are parallel (apart from random variation). As there is just a single data value for each variety in each block, it is not possible to separate interaction effects and random variation. So in practice the situation is not as nice and clear as here because the effects may be masked by random variation.

3.4 Validation of assumptions

[TWC Chairperson: To review the content of this section for overlap with section 2.5.2]

3.4.1 Introduction

3.4.1.1 The purpose of validation is partly to check that the data are without mistakes and that the assumptions underlying the statistical analyses are fulfilled. The main purpose of validation is to check that the assumptions underlying the statistical analyses are fulfilled. However, it also serves as a secondary check that the data are without mistakes.

3.4.1.2 There are different methods to use when validating the assumptions. Some of these are:

- look through the data to verify the assumptions
- produce plots or figures to verify the assumptions
- make formal statistical tests for the different types of assumptions. In the literature several methods to test for outliers, variance homogeneity, additivity and normality may be found. Such methods will not be mentioned here partly because many of these depend on assumptions that do not affect the validity of COYD and COYU seriously and partly because the power of such methods depends heavily on the sample size (this means that serious lack of assumptions may remain undetected in small datasets, whereas small and unimportant deviations may become statistically significant in large datasets)

3.4.2 Looking through the data

In practice this method is only applicable when a few observations have to be checked. For large datasets this method takes too much time, is tedious and the risk of overlooking suspicious data increases as one goes through the data. In addition, it is very difficult to judge the distribution of the data and to judge the degree of variance homogeneity when using this method.

3.4.3 Using Figures

2.5.4.3.1 Different kinds of figures can be prepared which are useful for the different aspects to be validated. Many of these consist of plotting the residuals in different ways. (The residuals are the differences between the observed values and the values predicted by the statistical model).

3.4.3.2 The plot of the residuals versus the predicted values may be used to judge the dependence of the variance on the mean. If there is no dependence, then the observations should fall approximately (without systematic deviation) in a horizontal band symmetric around zero (Figure 4). In cases where the variance increases with the mean, the observations will fall approximately in a funnel with the narrow end pointing to the left. Outlying observations, which may be mistakes, will be shown in such a figure as observations that clearly have escaped from the horizontal band formed by most other observations. In the example used in figure 4, no observations seem to be outliers (the value at the one bottom left corner where the residual is about -40 mm may at first glance look so, but several observations have positive values of the same numerical size). Here it is important to note that an outlier is not necessarily a mistake and also that a mistake will not necessarily show up as an outlier.

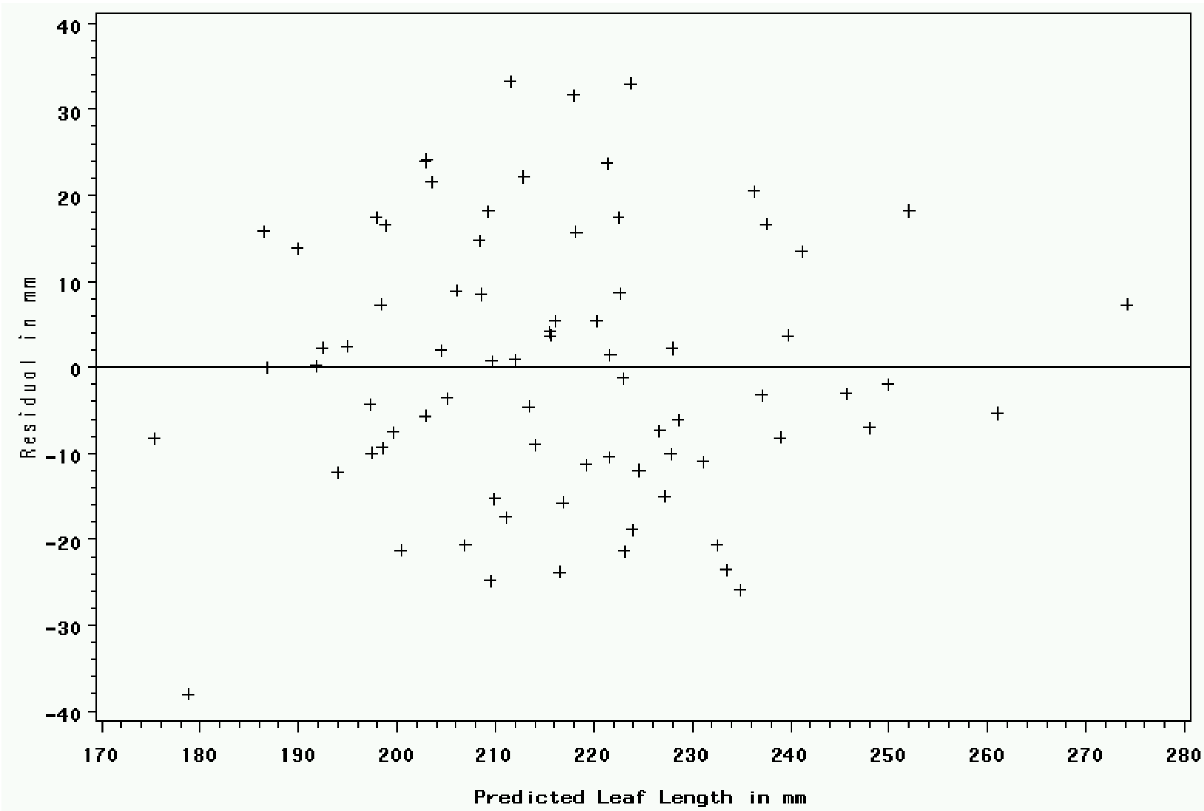


Figure 4. Plot of residuals versus plot predicted values for Leaf Length in 26 oil seed rape varieties in 3 blocks

3.4.3.3 The residuals can also be used to form a histogram, like Figure 2, from which the assumption about the distribution can be judged.

3.4.3.4 The range (maximum value minus minimum value) or standard deviation for each plot may be plotted versus some other variables such as the plot means, variety number or plot number. Such figures (Figure 5) may be useful to find varieties with an extremely large variation (all plots of the variety with a large value) or plots where the variation is extremely large (maybe caused by a single plant). It is clearly seen that the range for one of variety 13's plots is much higher than in the other two plots. Also the range in one of variety 3's plots seems to be relatively large.

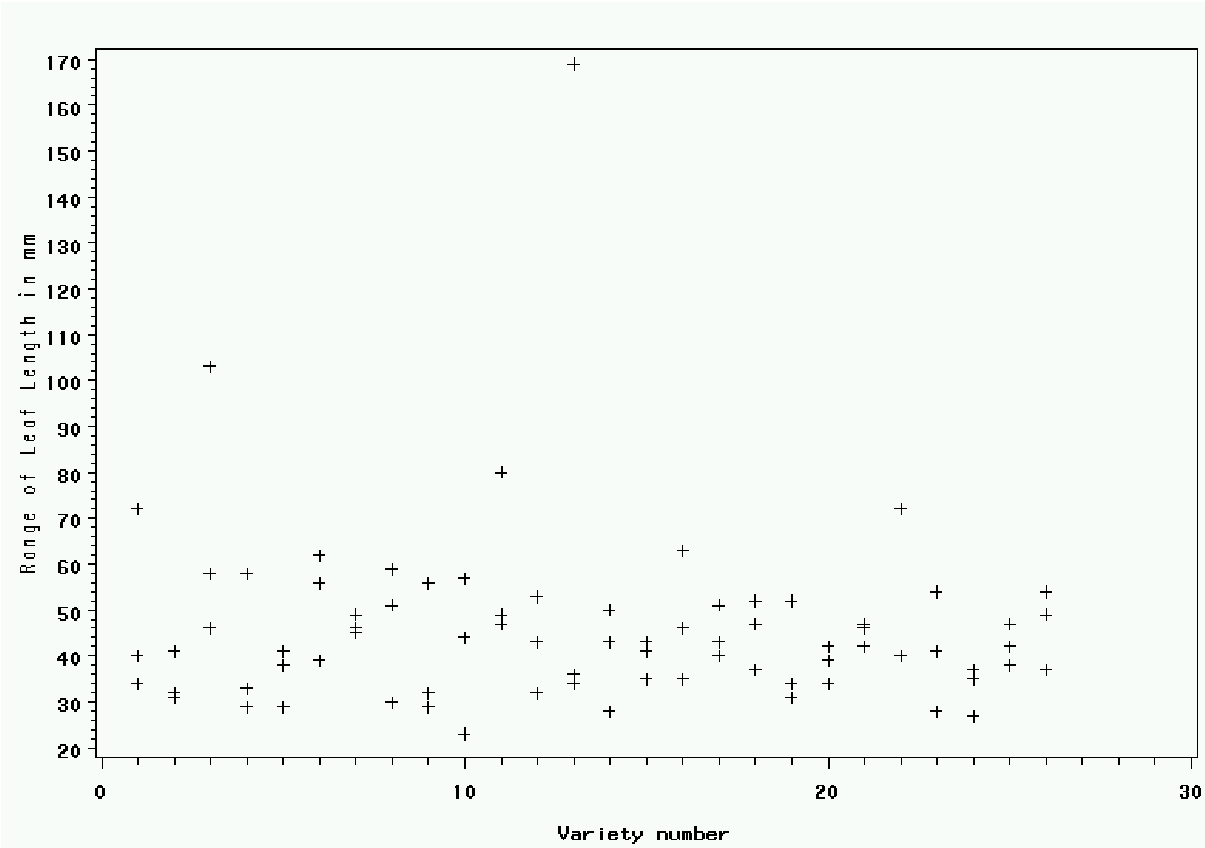


Figure 5. Differences between minimum and maximum of 20 leaf lengths for 3 plots versus oil seed rape variety number

3.4.3.5 A figure with the plot means (or variety adjusted means) versus the plot number can be used to find out whether the characteristic depends on the location in the field (Figure 6). This, of course, requires that the plots are numbered such that the numbers indicate the relative location. In the example shown in Figure 6, there is a clear trend showing that the leaf length decreases slightly with plot number. However most of the trend over the area used for the trial will - in this case - be explained by differences between blocks (plot 1-26 is block 1, plot 27-52 is block 2 and plot 53-78 is block 3).

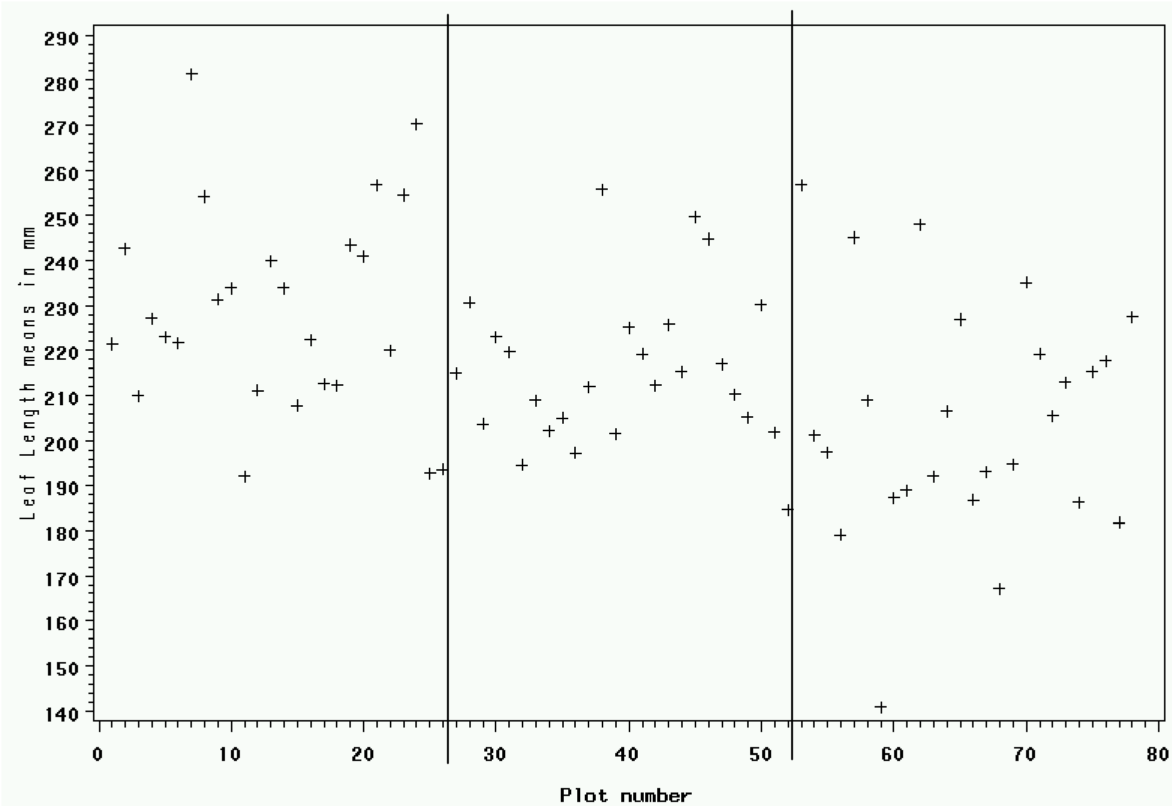


Figure 6. Plot means of 20 Leaf Lengths versus plot numbers

3.4.3.6 The plot means can also be used to form a figure where the additivity of block and variety effects can be visually checked at (see Figure 3).

3.4.3.7 Normal Probability Plots (Figure 7). This type of graph is used to evaluate to what extent the distribution of the variable follows the normal distribution. The selected variable will be plotted in a scatter plot against the values "expected from the normal distribution." The standard normal probability plot is constructed as follows. First, the residuals (deviations from the predictions) are rank ordered. From these ranks the program computes the expected values from the normal distribution, hereafter called z-values. These z-values are plotted on the X-axis in the plot. If the observed residuals (plotted on the Y-axis) are normally distributed, then all values should fall onto a straight line. If the residuals are not normally distributed, then they will deviate from the line. Outliers may also become evident in this plot. If there is a general lack of fit, and the data seem to form a clear pattern (e.g. an S shape) around the line, then the variable may have to be transformed in some way.

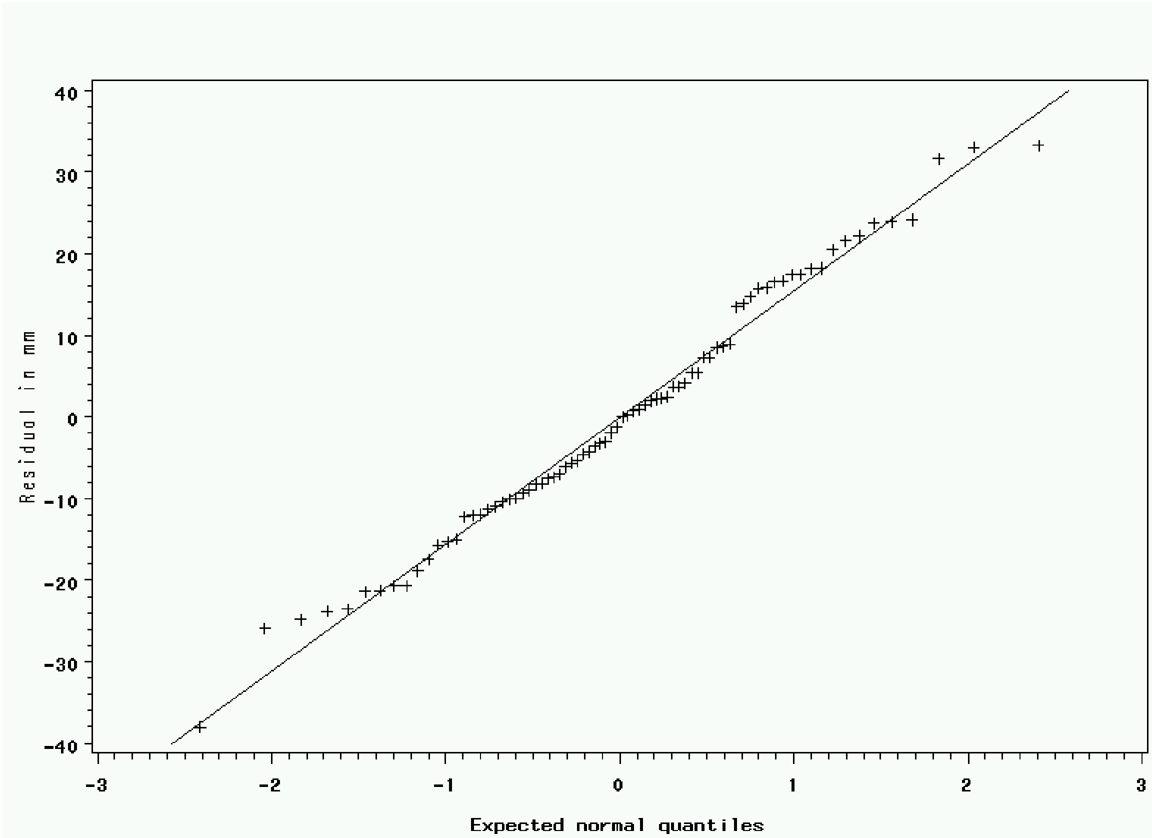


Figure 7. Normal probability plot for the residuals of Leaf Length in 26 oil seed rape varieties in 3 blocks

4. TYPES OF CHARACTERISTICS AND THEIR SCALE LEVELS

[TWC Chairperson: TWC to consider whether this section helps crop experts to better understand the statistical basis for the examination of characteristics]

4.1 Introduction

4.1.1 The General Introduction makes the following recommendations with regard to the use of statistical methods in the assessment of distinctness:

"5.5 Interpretation of Observations for the Assessment of Distinctness with the Application of Statistical Methods

"5.5.1 General

"5.5.1.1 For measured characteristics as well as for visually assessed[*] characteristics statistical methods can be applied. Appropriate methods have to be chosen for the interpretation of observations. The data structure and the type of scale from a statistical point of view (nominal, ordinal, interval or ratio) is decisive for the choice of appropriate methods. The data structure depends on the method of assessment[*] (visual assessment[*] or measurements, observation of plots or single plants) which is influenced by the type of characteristic, the features of propagation of the variety, the experimental design and other factors. DUS examiners should be aware of certain basic rules of statistics and especially the fact that their use is linked to mathematical assumptions and the use of experimental design practices, such as randomization. Therefore, those assumptions should be verified before applying statistical methods. Some statistical methods are quite robust, however, and can be used, with some caution, even if some assumptions are not fully met.

"5.5.1.2 Document TGP/8, "Use of Statistical Procedures in DUS Testing," provides guidance on some appropriate statistical procedures for DUS assessment and includes keys for the choice of methods in relation to the data structure.

[…]

"5.5.2 Visually Assessed[*] Characteristics

"Non-parametric statistics may be used when visually assessed[*] characteristics have been recorded on a scale that does not fulfill the assumptions of the usual parametric statistics. The calculation of the mean value, for example, is only permitted if the Notes are taken on a graded scale which shows equal intervals throughout the scale. In the case of non-parametric procedures, the use of a scale that has been established on the basis of example varieties representative of the different states of the characteristics is recommended. The same variety should then always receive about the same Note and thereby facilitate the interpretation of data. More details on the handling of visually assessed[*] characteristics are given in document TGP/9, "Examining Distinctness"."

([*] the term "observed" would be more consistent with the use of the terms "observed" and "assessed" in TGP/9)

4.1.2 For the revision of UPOV Test Guidelines or for establishing new ones, and in order to understand the relations between the different steps of work of the crop experts during the DUS test, it is necessary to have an answer to the following questions:

1. What is a characteristic?
2. What is a process level?
3. What is a scale level of a characteristic?
4. What is the influence of the scale level on the :
   - planning of a trial,
   - recording of data,
   - determination of distinctness and uniformity and
   - description of varieties.

4.2 Different levels to look at a characteristic

Characteristics can be considered in different levels of process (Table 1). The characteristics as expressed in the trial (type of expression) are considered as process level 1. The data taken from the trial for the assessment of distinctness, uniformity and stability are defined as process level 2. These data are transformed into states of expression for the purpose of variety description. The variety description is process level 3.

Table 1: Definition of different process levels to consider characteristics

Process level	Description of the process level
1	characteristics as expressed in trial
2	data for evaluation of characteristics
3	variety description

From the statistical point of view the information level decreases from process level 1 to 3. Statistical analysis is only applied in level 2.

Sometimes for crop experts it seems that there is no need to distinguish between different process levels. The process level 1, 2 and 3 could be identical. However, in general, this is not the case.

4.2.1 Understanding the need for process levels

4.2.1.1 The crop expert may know from UPOV Test Guidelines or his own experience that, for example, 'Length of plant' is a good characteristic for the examination of DUS. There are varieties which have longer plants than other varieties. Another characteristic could be 'Variegation of leaf blade'. For some varieties, variegation is present and for others not. The crop expert has now two characteristics and he knows that 'Plant length' is a quantitative characteristic and 'Variegation of leaf blade' is a qualitative characteristic (definitions: see Part I: Section 4.3.2 *[cross ref.]* below). This stage of work can be described as **process level 1.**

4.2.1.2 The crop expert then has to plan the trial and to decide on the type of observation for the characteristics. For characteristic 'Variegation of leaf blade', the decision is clear. There are two possible expressions: 'present' or 'absent'. The decision for characteristic 'Plant length' is not specific and depends on expected differences between the varieties and on the variation within the varieties. In many cases, the crop expert will decide to measure a number of plants (in cm) and to use special statistical procedures to examine distinctness and uniformity. But it could also be possible to assess the characteristic 'Plant length' visually by using expressions like 'short', 'medium' and 'long', if differences between varieties are large enough (for distinctness) and the variation within varieties is very small or absent in this characteristic. The continuous variation of a characteristic is assigned to appropriate states of expression which are recorded by notes (see TGP/9, Section 4*)[cross. ref]*. The crucial element in this stage of work is the recording of data for further evaluations. It is described as **process level 2**.

4.2.1.3 At the end of the DUS test, the crop expert has to establish a description of the varieties using notes from 1 to 9 or parts of them. This phase can be described as **process level 3**. For 'Variegation of leaf blade' the crop expert can take the same states of expression (notes) he recorded in process level 2 and the three process levels appear to be the same. In cases where the crop expert decided to assess 'Plant length' visually, he can take the same states of expression (notes) he recorded in process level 2 and there is no obvious difference between process level 2 and 3. If the characteristic 'Plant length' is measured in cm, it is necessary to assign intervals of measurements to states of expressions like 'short', 'medium' and 'long' to establish a variety description. In this case, for statistical procedures, it is important to be clearly aware of the relevant level and to understand the differences between characteristics as expressed in the trial, data for evaluation of characteristics and the variety description. This is absolutely necessary for choosing the most appropriate statistical procedures in cooperation with statisticians or by the crop expert.

4.3 Types of expression of characteristics

4.3.1 Characteristics can be classified according to their types of expression. The consideration of the type of expression of characteristics corresponds to process level 1. The following types of expression of characteristics are defined in 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, the "General Introduction", Chapter 4.4):

4.3.2 Qualitative characteristics" are those that are expressed in discontinuous states (e.g. sex of plant: dioecious female (1), dioecious male (2), monoecious unisexual (3), monoecious hermaphrodite (4)). These states are self-explanatory and independently meaningful. All states are necessary to describe the full range of the characteristic, and every form of expression can be described by a single state. The order of states is not important. As a rule, the characteristics are not influenced by environment.

4.3.3 "Quantitative characteristics" are those where the expression covers the full range of variation from one extreme to the other. The expression can be recorded on a one-dimensional, continuous or discrete, linear scale. The range of expressions is divided into a number of states for the purpose of description (e.g. length of stem: very short (1), short (3), medium (5), long (7), very long (9)). The division seeks to provide, as far as practical, an even distribution across the scale. The Test Guidelines do not specify the difference needed for distinctness. The states of expression should, however, be meaningful for DUS assessment.

4.3.4 In the case of "pseudo-qualitative characteristics" the range of expression is at least partly continuous, but varies in more than one dimension (e.g. shape: ovate (1), elliptic (2), circular (3), obovate (4)) and cannot be adequately described by just defining two ends of a linear range. In a similar way to qualitative (discontinuous) characteristics – hence the term "pseudo-qualitative" – each individual state of expression needs to be identified to adequately describe the range of the characteristic.

4.4 Types of scales of data

The possibility to use specific procedures for the assessment of distinctness, uniformity and stability depends on the scale level of the data which are recorded for a characteristic. The scale level of data depends on the type of expression of the characteristic and on the way of recording this expression. The type of scale may be quantitative or qualitative.

4.4.1 Quantitatively scaled data (metric or ordinal scaled data)

Quantitatively scaled data are all data which are recorded by measuring or counting. Weighing is a special form of measuring. Quantitatively scaled data can have a continuous or a discrete distribution. Continuous data result from measurements. They can take every value out of the defined range. Discrete quantitative data result from counting.

Examples

Quantitatively scaled data	Example	Example number
- continuous	Plant length in cm.	1
- discrete	Number of stamens	2

For description of the states of expression, see Table 6.

The continuous quantitatively scaled data for the characteristic "Plant length" are measured on a continuous scale with defined units of assessment. A change of unit of measurement e.g. from cm into mm is only a question of precision and not a change of type of scale.

The discrete quantitatively scaled data of the characteristic "Number of stamens " are assessed by counting (1, 2, 3, 4, and so on). The distances between the neighboring units of assessment are constant and for this example equal to 1. There are no real values between two neighboring units but it is possible to compute an average which falls between those units.

In biometrical terminology, quantitative scales are referred to as metric scales or cardinal scales. Quantitative scales can be subdivided into ratio scales and interval scales.

4.4.1.1 Ratio scale

[TWC Chairperson: To review if this paragraph is relevant for DUS testing]

A ratio scale is a quantitative scale with a defined absolute zero point. There is always a constant non-zero distance between two adjacent expressions. Ratio scaled data may be continuous or discrete.

The absolute zero point:

The definition of an absolute zero point makes it possible to define meaningful ratios. This is a requirement for the construction of index numbers (e.g. the ratio of length to width). An index is the combination of at least two characteristics. In the General Introduction, this is referred to as a combined characteristic (see document TG/1/3, Section 4.6.3).

It is also possible to calculate ratios between the expression of different varieties. For example, in the characteristic 'Plant length' assessed in cm, there is a lower limit for the expression which is '0 cm' (zero). It is possible to calculate the ratio of length of plant of variety 'A' to length of plant of variety 'B' by division:

[TWC Chairperson: To review if this paragraph is relevant for DUS testing]

Length of plant of variety 'A' = 80 cm
Length of plant of variety 'B' = 40 cm

Ratio = Length of plant of variety 'A' / Length of plant of variety 'B'
= 80 cm / 40 cm
= 2.

So it is possible in this example to state that plant 'A' is double the length of plant 'B'. The existence of an absolute zero point ensures an unambiguous ratio.

The ratio scale is the highest classification of the scales (Table 2). That means that ratio scaled data include the highest information about the characteristic and it is possible to use many statistical procedures (Chapter 7 *[cross ref.]*).

The examples 1 and 2 (Table 6) are examples for characteristics with ratio scaled data.

4.4.1.2 Interval scale

An Interval scale is a quantitative scale without a defined absolute zero point. There is always a constant non-zero distance between two adjacent expressions. Interval scaled data may be distributed continuously or discretely.

An example for a discrete interval scaled characteristic is 'Time of beginning of flowering' measured as date which is given as example 6 in Table 6. This characteristic is defined as the number of days from April 1. The definition is useful but arbitrary and April 1 is not a natural limit. It would also be possible to define the characteristic as the number of days from January 1.

It is not possible to calculate a meaningful ratio between two varieties which should be illustrated with the following example:

Variety 'A' begins to flower on May 30 and variety 'B' on April 30

Case I) Number of days from April 1 of variety 'A' = 60
Number of days from April 1 of variety 'B' = 30

$$\text{Ratio}_\text{I} = \frac{\text{Number of days from April 1 of variety 'A'}}{\text{Number of days from April 1 of variety 'B}} = \frac{\text{60 days}}{\text{30 days}} = 2$$

Case II) Number of days from January 1 of variety 'A' = 150
Number of days from January 1 of variety 'B' = 120

$$\text{Ratio}_\text{II} = \frac{\text{Number of days from January 1 of variety 'A'}}{\text{Number of days from January 1 of variety 'B}} = \frac{\text{150 days}}{\text{120 days}} = 1.25$$

$\text{Ratio}_\text{I} = 2 > 1.25 = \text{Ratio}_\text{II}$

It is impossible to state that the time of flowering of variety 'A' is twice that of variety 'B'. The ratio depends on the choice of the zero point of the scale. This kind of scale is defined as an "Interval scale": a quantitative scale without a defined absolute zero point.

The interval scale is lower classified than the ratio scale (Table 2). Fewer statistical procedures can be used with interval scaled data than with ratio scaled data (see Part I: Section 4.7 *[cross ref.]*). The interval scale is theoretically the minimum scale level to calculate arithmetic mean values.

4.4.2 Qualitatively scaled data

Qualitatively scaled data are data which can be arranged in different discrete qualitative categories. Usually they result from visual assessment. Subgroups of qualitative scales are ordinal and nominal scales.

4.4.2.1 Ordinal scale

[TWC Chairperson: example for a non-quantitative characteristic to be provided]

Ordinally scaled data are qualitative data of which discrete categories can be arranged in an ascending or descending order. They result from visually assessed quantitative characteristics.

Example:

Qualitative data	Example	Example number
- ordinal	Intensity of anthocyanin	3

For description of the states of expressions, see Table 6.

An ordinal scale consists of numbers which correspond to the states of expression of the characteristic (notes). The expressions vary from one extreme to the other and thus they have a clear logical order. It is not possible to change this order, but it is not important which numbers are used to denote the categories. In some cases ordinal data may reach the level of discrete interval scaled data or of discrete ratio scaled data (Chapter 6 *[cross ref.]*).

The distances between the discrete categories of an ordinal scale are not exactly known and not necessarily equal. Therefore, an ordinal scale does not fulfil the condition to calculate arithmetic mean values, which is the equality of intervals throughout the scale.

The ordinal scale is lower classified than the interval scale (Table 2). Less statistical procedures can be used for ordinal scale than for each of the higher classified scale data (see Part I: Section 3.7 *[cross ref.]*).

4.4.2.2 Nominal scale

Nominal scaled qualitative data are qualitative data without any logical order of the discrete categories.

Examples:

Qualitative data	Example	Example number
- nominal	Sex of plant	4
- nominal with two states	Leaf blade: variegation	5

For description of the states of expressions, see Table 6.

A nominal scale consists of numbers which correspond to the states of expression of the characteristic, which are referred to in the Test Guidelines as notes. Although numbers are used for designation there is no inevitable order for the expressions and so it is possible to arrange them in any order.

Characteristics with only two categories (dichotomous characteristic) are a special form of nominal scales.

The nominal scale is the lowest classification of the scales (Table 2). Few statistical procedures are applicable for evaluations (Chapter 7 *[cross ref.]*).

The different types of scales are summarized in the following table.

Table 2: Types of scales and scale levels

[TWC Chairperson: To modify the table for consistency with the subsequent paragraphs]

Type of scale		Description	Distribution	Data recording	Scale Level
quantitative (metric)	ratio	constant distances with absolute zero point	Continuous	Absolute Measurements	High
			Discrete	Counting	
	interval	constant distances without absolute zero point	Continuous	Relative measurements	↑
			Discrete	Date	
qualitative with underlying quantitative variable	ordinal	Ordered expressions with varying distances	Discrete	Visually assessed notes	↑
qualitative	nominal	No order, no distances	Discrete	Visually assessed notes	Low

From the statistical point of view a characteristic is only considered at the level of data which has been recorded, whether for analysis or for describing the expression of the characteristic. Therefore, characteristics with quantitative data are denoted as quantitative characteristics and characteristics with ordinal and nominal scaled data as qualitative characteristics.

4.5 Scale levels for variety description

The description of varieties is based on the states of expression (notes) which are given in the Test Guidelines for the specific crop. In the case of visual assessment, the notes from the Test Guidelines are usually used for recording the characteristic as well as for the assessment of DUS. The notes are distributed on a nominal or ordinal scale (see Part I: Section 4.4.2 *[cross ref.]*). For measured or counted characteristics, DUS assessment is based on the recorded values and the recorded values are transformed into states of expression only for the purpose of variety description.

4.6 Relation between types of expression of characteristics and scale levels of data

4.6.1 Records taken for the assessment of qualitative characteristics are distributed on a nominal scale, for example “Sex of plant”, “Leaf blade: variegation” (Table 6, examples 4 and 5).

4.6.2 For quantitative characteristics the scale level of data depends on the method of assessment. They can be recorded on a quantitative or ordinal scale. For example, “Length of plant” can be recorded by measurements resulting in ratio scaled continuous quantitative data. However, visual assessment on a 1 to 9 scale may also be appropriate. In this case, the recorded data are qualitatively scaled (ordinal scale) because the size of intervals between the midpoints of categories is not exactly the same.

Remark: In some cases visually assessed data on quantitative characteristics may be handled as measurements. The possibility to apply statistical methods for quantitative data depends on the precision of the assessment and the robustness of the statistical procedures. In the case of very precise visually assessed quantitative characteristics the usually ordinal data may reach the level of discrete interval scaled data or of discrete ratio scaled data.

4.6.3 A pseudo-qualitative type of characteristic is one in which the expression varies in more than one dimension. The different dimensions are combined in one scale. At least one dimension is quantitatively expressed. The other dimensions may be qualitatively expressed or quantitatively expressed. The scale as a whole has to be considered as a nominal scale (e.g. “Shape”, “Flower color”; Table 6, examples 7 and 8).

4.6.4 In the case of using the off-type procedure for the assessment of uniformity the recorded data are nominally scaled. The records fall into two qualitative classes: plants belonging to the variety (true-types) and plants not belonging to the variety (off-types). The type of scale is the same for qualitative, quantitative and pseudo-qualitative characteristics.

4.6.5 The relation between the type of characteristics (process level 1) and the type of scale of data recorded for the assessment of distinctness and uniformity is described in Table 3. A qualitative characteristic is recorded on a nominal scale for distinctness (state of expression) and for uniformity (true-types vs. off-types). Pseudo-qualitative characteristics are recorded on a combined scale for distinctness (state of expression) and on a nominal scale for uniformity (true-types vs. off-types). Quantitative characteristics are recorded on an ordinal, interval or ratio scale for the assessment of distinctness depending on the characteristic and the method of assessment. If the records are taken from single plants the same data may be used for the assessment of distinctness and uniformity. If distinctness is assessed on the basis of a single record of a group of plants, uniformity has to be judged with the off-type procedure (nominal scale).

Table 3: Relation between type of characteristic and type of scale of assessed data

Procedure	Type of scale (level 2)	Distribution	Type of characteristic (level 1)		
			Quantitative	Pseudo-qualitative	Qualitative
Distinctness	ratio	Continuous	▪		
		Discrete	▪		
	interval	Continuous	▪		
		Discrete	▪		
	ordinal	Discrete	▪		
	combined	Discrete		▪	
	nominal	Discrete			▪
Uniformity	ratio	Continuous	▪		
		Discrete	▪		
	interval	Continuous	▪		
		Discrete	▪		
	ordinal	Discrete	▪		
	combined	Discrete	▪		
	nominal	Discrete	▪	▪	▪

4.7 Relation between method of observation of characteristics, scale levels of data and recommended statistical procedures

4.7.1 TGP/9, Section 4.4 provides the following in respect of the method of observation:

4.4 Recommendations in the UPOV Test Guidelines

The indications used in UPOV Test Guidelines for the method of observation and the type of record are as follows:

Method of observation

M: to be measured (an objective observation against a calibrated, linear scale e.g. using a ruler, weighing scales, colorimeter, dates, counts, etc.);

V: to be observed visually (includes observations where the expert uses reference points (e.g. diagrams, example varieties, side-by-side comparison) or non-linear charts (e.g. color charts)

Type of record(s)

G: single record for a variety, or a group of plants or parts of plants;

S: records for a number of single, individual plants or parts of plants

4.5 Summary

The following table summarizes the common method of observation and type of record for the assessment of distinctness, although there may be exceptions:

	Type of expression of characteristic		
Method of propagation of the variety	QL	PQ	QN
Vegetatively propagated	VG	VG	VG/MG/MS
Self-pollinated	VG	VG	VG/MG/MS
Cross-pollinated	VG/(VS*)	VG/(VS*)	VS/VG/MS/MG
Hybrids	VG/(VS*)	VG/(VS*)	**

* records of individual plants only necessary if segregation is to be recorded
** to be considered according to the type of hybrid

[TWC Chairperson: To update these paragraphs in accordance with any changes to TGP/7 and TGP/9]

4.7.3 Established statistical procedures can be used for the assessment of distinctness and uniformity considering the scale level and some further conditions such as the degree of freedom or unimodality (Tables 4 and 5).

4.7.4 The relation between the expression of characteristics and the scale levels of data for the assessment of distinctness and uniformity is summarized in Table 6.

Table 4: Statistical procedures for the assessment of distinctness

Type of scale	Distribu-tion	Observa-tion method	Procedure[1)] and further Conditions	Reference document
ratio	continuous	MS MG (VS) [1)]	COY-D Normal distribution, df >=20 long term LSD Normal distribution, df<20 2 out of 3 method (LSD 1%) Normal distribution, df>=20	TGP/9
	discrete			
interval	continuous			
	discrete			
ordinal	discrete	VG VS	See explanation for QN characteristics in TGP/9 Sections 5.2.2 and 5.2.3, See explanation for QN characteristics in TGP/9 Section 5.2.4	TGP/9 TWC/ 14/12
Combina-tion of ordinal or ordinal and nominal scales	discrete	VG (VS) [32]	See explanation for PQ characteristics in TGP/9 Sections 5.2.2 and 5.2.3	TGP/9
nominal	discrete	VG (VS) [2)]	See explanation for QL characteristics in TGP/9 Sections 5.2.2 and 5.2.3	TGP/9

1) see remark in Chapter 6 *[cross ref.]*
2) normally VG but VS would be possible

Table 5: Statistical procedures for the assessment of uniformity

Type of scale	Distribu-tion	observa-tion method	Procedure[1)] and Further Conditions	Refe-rence document
ratio	continuous	MS	COY-U Normal distribution 2 out of 3 method (s^2_c<=1.6s^2_s)) Normal distribution LSD for untransformed percentage of off-types	TGP/10
	discrete	MS		
interval	continuous	VS		
	discrete			
ordinal	discrete	VS	threshold model	TWC/ 14/12
Combina-tion of ordinal or ordinal and nominal scales	discrete		There is no case where uniformity is assessed on combined scaled data	
nominal	discrete	VS	off-type procedure for dichotomous (binary) data	TGP/10

Table 6: Relation between expression of characteristics and scale levels of data for the assessment of distinctness and uniformity

Example	Name of characteristic	Distinctness			Uniformity		
		Unit of assess-ment	Description (states of expression)	Type of scale	Unit of assess-ment	Description (states of expression)	Type of scale
1	Length of plant	cm	assessment in cm without digits after decimal point	ratio scaled continuous quantitative data	cm	assessment in cm without digits after decimal point	ratio scaled continuous quantitative data
					True-type Off-type	Number of plants belonging to the variety Number of off-types	nominally scaled qualitative data
2	Number of stamens	counts	1, 2, 3, ... , 40,41, ...	ratio scaled discrete quantitative data	counts	1, 2, 3, ... , 40,41, ...	ratio scaled discrete quantitative data
3	Intensity of anthocyanin	1 2 3 4 5 6 7 8 9	very low very low to low low low to medium medium medium to high high high to very high very high	ordinally scaled qualitative data (with an underlying quantitative variable)	True-type Off-type	Number of plants belonging to the variety Number of off-types	nominally scaled qualitative data
4	Sex of plant	1 2 3 4	dioecious female dioecious male monoecious unisexual monoecious hermaphrodite	nominally scaled qualitative data	True-type Off-type	Number of plants belonging to the variety Number of off-types	nominally scaled qualitative data

Example	Name of characteristic	Distinctness			Uniformity		
		Unit of assessment	Description (states of expression)	Type of scale	Unit of assessment	Description (states of expression)	Type of scale
5	Leaf blade: variegation	1 9	absent present	nominally scaled qualitative data	True-type Off-type	Number of plants belonging to the variety Number of off-types	nominally scaled qualitative data
6	Time of beginning of flowering	date	e.g. May 21, 51st day from April 1	interval scaled discrete quantitative data	date	e.g. May 21, 51st day from April 1	interval scaled discrete quantitative data
					True-type Off-type	Number of plants belonging to the variety Number of off-types	nominally scaled qualitative data
7	Shape	1 2 3 4 5 6 7	deltate ovate elliptic obovate obdeltate circular oblate	combination of ordinal and nominal scaled discrete qualitative data	True-type Off-type	Number of plants belonging to the variety Number of off-types	nominally scaled qualitative data
8	Flower color	1 2 3 4 5 6 7 8 9 10	dark red medium red light red white light blue medium blue dark blue red violet violet blue violet	combination of ordinal and nominal scaled discrete qualitative data	True-type Off-type	Number of plants belonging to the variety Number of off-types	nominally scaled qualitative data

PART II: TECHNIQUES USED IN DUS EXAMINATION

1. METHODS FOR ASSESSING UNIFORMITY ON THE BASIS OF OFF-TYPES

1.1 Fixed Population Standard

1.1.1 Introduction

1.1.1.1 TGP/10 Section 4 provides guidance on the uniformity assessment on the basis of off-types. This section describes the method of assessing uniformity by comparing the number of off-types observed to a fixed population standard. In particular, the tables included in Section 3.1.12 provide the recommendation on the acceptable number of off-types in different sample sizes for specific population standards and acceptance probabilities within acceptable levels of errors.

> 1.1.1.2 This document also outlines procedures for when more than a single test (more than one year for instance) is used and explains the possibility of using sequential tests to minimize testing effort.

1.1.1.3 When testing for uniformity on the basis of a sample, there will always be some risk of making a wrong decision. The risks can be reduced by increasing the sample size but at a greater cost. The aim of the statistical procedure described here is to achieve an acceptable balance between risks.

1.1.1.4 The procedures described below require the user to define an acceptance standard (called the *population standard*) for the crop in question. The methods described then show how to determine the sample size and the maximum number of off-types allowed for various levels of risks.

1.1.1.5 The population standard is the maximum percentage of off-types that would be accepted if all individuals of the variety could be examined.

1.1.2 Recommendations on the fixed population standard method of assessing uniformity by the number of off-types

1.1.2.1 This method is recommended for use in assessing the uniformity by number of off-types with a fixed population standard.

1.1.2.2 The sample size and acceptable number of off-types employed depend on the crop.

1.1.3 Errors in testing for off-types

[TWF: to review whether such an explanation is appropriate in the light of sample sizes used in the TG. The document should reflect the positive experience of UPOV in the existing sample sizes]

[TWC Chairperson: to review the terms used and, in particular, terms such as “wrong decisions”]

1.1.3.1 As mentioned, there will be some risk of making wrong decisions. Two types of error exist:

(a) Declaring that the variety lacks uniformity when it in fact meets the standard for the crop. This is known as "type I error."

(b) Declaring that the variety is uniform when it in fact does not meet the standard for the crop. This is known as "type II error."

1.1.3.2 The types of error can be summarized in the following table:

True state of the variety	Decision made on variety	
	Acceptance as uniform	Rejection as non-uniform
uniform	correctly accepted	type I error
non-uniform	type II error	correctly rejected

1.1.3.3 The probability of correctly accepting a uniform variety is called the acceptance probability and is linked to the probability of type I error by the relation:

"Acceptance probability" + "probability of type I error" = 100%

1.1.3.4 The probability of type II error depends on "how non-uniform" the candidate variety is. If it is much more non-uniform than the population standard then the probability of type II error will be small and there will be a small probability of accepting such a variety. If, on the other hand, the candidate variety is only slightly more non-uniform than the standard, there is a large probability of type II error. The probability of acceptance will approach the acceptance probability for a variety with a level of uniformity near to the population standard.

1.1.3.5 Because the probability of type II error is not fixed but depends on "how non-uniform" the candidate variety is, this probability can be calculated for different degrees of non-uniformity. This document gives probabilities of type II error for three degrees of non-uniformity: 2, 5 and 10 times the population standard.

1.1.3.6 In general, the probability of making errors will be decreased by increasing the sample size and increased by decreasing the sample size.

1.1.3.7 For a given sample size, the balance between the probabilities of making type I and type II errors may be altered by changing the number of off-types allowed.

1.1.3.8 If the number of off-types allowed is increased, the probability of type I error is decreased but the probability of type II error is increased. On the other hand, if the number of off-types allowed is decreased, the probability of type I errors is increased while the probability of type II errors is decreased.

1.1.3.9 By allowing a very high number of off-types it will be possible to make the probability of type I errors very low (or almost zero). However, the probability of making type II errors will now become (unacceptably) high. If only a very small number of off-types is allowed, the result will be a small probability of type II errors and an (unacceptably) high probability of type I errors. This will be illustrated by examples.

1.1.4 Examples

Example 1

1.1.4.1 From experience, a reasonable standard for the crop in question is found to be 1%. So the population standard is 1%. Assume that a single test with a maximum of 60 plants is used. From tables 4, 10 and 16 (chosen to give a range of target acceptance probabilities), the following schemes are found:

Scheme	Sample size	Target acceptance probability*	Maximum number of off-types
a	60	90%	2
b	53	90%	1
c	60	95%	2
d	60	99%	3

1.1.4.2 From the figures 4, 10 and 16, the following probabilities are obtained for the type I error and type II error for different percentages of off-types (denoted by P_2, P_5 and P_{10} for 2, 5 and 10 times the population standard).

Scheme	Sample size	Maximum number of off-types	Probabilities of error (%)			
			Type I	Type II		
				$P_2 = 2\%$	$P_5 = 5\%$	$P_{10} = 10\%$
a	60	2	2	88	42	5
b	53	1	10	71	25	3
c	60	2	2	88	42	5
d	60	3	0.3	97	65	14

1.1.4.3 The table lists four different schemes and they should be examined to see if one of them is appropriate to use. (Schemes a and c are identical since there is no scheme for a sample size of 60 with a probability of type I error between 5 and 10%). If it is decided to ensure that the probability of a type I error should be very small (scheme d) then the probability of the type II error becomes very large (97, 65 and 14%) for a variety with 2.5 and 10% of off-types, respectively. The best balance between the probabilities of making the two types of error seems to be obtained by allowing one off-type in a sample of 53 plants (scheme b).

---

* See paragraph 54

Example 2

1.1.4.4 In this example, a crop is considered where the population standard is set to 2% and the number of plants available for examination is only 6.

1.1.4.5 Using the tables and the figures 3, 9 and 15, the following schemes a-d are found:

Scheme	Sample size	Acceptance probability	Maximum number of off-types	Probability of error (%)			
				Type I	Type II		
					$P_2 = 4\%$	$P_5 = 10\%$	$P_{10} = 20\%$
a	6	90	1	0.6	98	89	66
b	5	90	0	10	82	59	33
c	6	95	1	0.6	98	89	66
d	6	99	1	0.6	98	89	66
e	6		0	11	78	53	26

1.1.4.6 Scheme e of the table is found by applying the formulas (1) and (2) shown later in this document.

1.1.4.7 This example illustrates the difficulties encountered when the sample size is very low. The probability of erroneously accepting a non-uniform variety (a type II error) is large for all the possible situations. Even when all five plants must be uniform for a variety to be accepted (scheme b), the probability of accepting a variety with 20% of off-types is still 33%.

1.1.4.8 It should be noted that a scheme where all six plants must be uniform (scheme e) gives slightly smaller probabilities of type II errors, but now the probability of the type I error has increased to 11%.

1.1.4.9 However, scheme e may be considered the best option when only six plants are available in a single test for a crop where the population standard has been set to 2%.

Example 3

1.1.4.10 In this example we reconsider the situation in example 1 but assume that data are available for two years. So the population standard is 1% and the sample size is 120 plants (60 plants in each of two years).

1.1.4.11 The following schemes and probabilities are obtained from the tables and figures 4, 10 and 16:

Scheme	Sample size	Acceptance probability	Maximum number of off-types	Probability of error (%)			
				Type I	Type II		
					$P_2 = 2\%$	$P_5 = 5\%$	$P_{10} = 10\%$
a	120	90	3	3	78	15	<0.1
b	110	90	2	10	62	8	<0.1
c	120	95	3	3	78	15	<0.1
d	120	99	4	0.7	91	28	1

1.1.4.12 Here the best balance between the probabilities of making the two types of error is obtained by scheme c, i.e. to accept after two years a total of three off-types among the 120 plants examined.

1.1.4.13 Alternatively a two-stage testing procedure may be set up. Such a procedure can be found for this case by using formulae (3) and (4) later in this document.

1.1.4.14 The following schemes can be obtained:

Scheme	Sample size	Acceptance probability	Largest number for acceptance after year 1	Largest number before reject in year 1	Largest number to accept after 2 years
e	60	90	can never accept	2	3
f	60	95	can never accept	2	3
g	60	99	can never accept	3	4
h	58	90	1	2	2

1.1.4.15 Using the formulas (3), (4) and (5) the following probabilities of errors are obtained:

Scheme	Probability of error (%)				Probability of testing in a second year
	Type I	Type II			
		$P_2 = 2\%$	$P_5 = 5\%$	$P_{10} = 10\%$	
e	4	75	13	0.1	100
f	4	75	13	0.1	100
g	1	90	27	0.5	100
h	10	62	9	0.3	36

1.1.4.16 Schemes e and f both result in a probability of 4% for rejecting a uniform variety (type I error) and a probability of 13% for accepting a variety with 5% off-types (type II error). The decision is:

– Never accept the variety after 1 year
– More than 2 off-types in year 1: reject the variety and stop testing
– Between and including 0 and 2 off types in year 1: do a second year test
– At most 3 off-types after 2 years: accept the variety
– More than 3 off-types after 2 years: reject the variety

1.1.4.17 Alternatively, scheme h may be chosen but scheme g seems to have a too large probability of type II errors compared with the probability of type I error.

1.1.4.18 Scheme h has the advantage of often allowing a final decision to be taken after the first test (year) but, as a consequence, there is a higher probability of a type I error.

[TWC: paragraphs 1.1.5.17 and 1.1.5.18 to be elaborated further]

Example 4

1.1.4.19 In this example, we assume that the population standard is 3% and that we have 8 plants available in each of two years.

1.1.4.20 From the tables and figures 2, 8 and 14, we have:

Scheme	Sample size	Acceptance probability	Maximum number of off-types	Probability of error (%)			
				Type I	Type II		
					$P_2 = 6\%$	$P_5 = 15\%$	$P_{10} = 30\%$
a	16	90	1	8	78	28	3
b	16	95	2	1	93	56	10
c	16	99	3	0.1	99	79	25

1.1.4.21 Here the best balance between the probabilities of making the two types of error is obtained by scheme a.

1.1.5 Introduction to the tables and figures

1.1.5.1 In the TABLES AND FIGURES section (Part II: Section 3.1.12 *[cross ref.]*), there are 21 table and figure pairs corresponding to different combinations of population standard and acceptance probability. These are design to be applied to a single off-type test. An overview of the tables and the figures are given in table A.

1.1.5.2 Each table shows the maximum numbers of off-types (k) with the corresponding ranges in sample sizes (n) for the given population standard and acceptance probability. For example, in table 1 (population standard 5%, acceptance probability ≥ 90%), for a maximum set at 2 off-types, the corresponding sample size (n) is in the range from 11 to 22. Likewise, if the maximum number of off-types (k) is 10, the corresponding sample size (n) to be used should be in the range 126 to 141.

1.1.5.3 For small sample sizes, the same information is shown graphically in the corresponding figures (figures (1 to 21). These show the actual risk of rejecting a uniform variety and the probability of accepting a variety with a true proportion of off-types 2 times (2P), 5 times (5P) and 10 times (10P) greater than the population standard. (To ease the reading of the figure, lines connect the risks for the individual sample sizes, although the probability can only be calculated for each individual sample size).

Table A. Overview of table and figure 1 to 18.

Population standard %	Acceptance probability %	See table and figure no.
10	>90	19
10	>95	20
10	>99	21
5	>90	1
5	>95	7
5	>99	13
3	>90	2
3	>95	8
3	>99	14
2	>90	3
2	>95	9
2	>99	15
1	>90	4
1	>95	10
1	>99	16
0.5	>90	5
0.5	>95	11
0.5	>99	17
0.1	>90	6
0.1	>95	12
0.1	>99	18

1.1.5.4 When using the tables the following procedure is suggested:

[TWC Chairperson: to be revised in accordance with the use of the tables set out in TGP/10 and with established practice]

(a) Choose the relevant population standard.

(b) Write down the different relevant decision schemes (combinations of sample size and maximum number of off-types), with the probabilities of type I and type II errors read from the figures.

(c) Choose the decision scheme with the best balance between the probabilities of errors.

1.1.5.5 The use of the tables and figures is illustrated in the example section.

1.1.6 Detailed description of the method for one single test

The mathematical calculations are based on the binomial distribution and it is common to use the following terms:

(a) The percentage of off-types to be accepted in a particular case is called the "population standard" and symbolized by the letter P.

(b) The "acceptance probability" is the probability of accepting a variety with P% of off-types. However, because the number of off-types is discrete, the actual probability of accepting a uniform variety varies with sample size but will always be greater than or equal to the "acceptance probability." The acceptance probability is usually denoted by 100 - α, where α is the percent probability of rejecting a variety with P% of off-types (i.e. type I error probability). In practice, many varieties will have less than P% off-types and hence the type I error will in fact be less than α for such varieties.

(c) The number of plants examined in a random sample is called the sample size and

P and P_q are expressed here as proportions, i.e. percents divided by 100.

denoted by n.

(d) The maximum number of off-types tolerated in a random sample of size n is denoted by k.

(e) The probability of accepting a variety with more than P% off-types, say P_q% of off-types, is denoted by the letter β or by β_q.

(f) The mathematical formulae for calculating the probabilities are:

$$\alpha = 100 - 100\sum_{i=0}^{k}\binom{n}{i}P^{i}(1-P)^{n-i} \qquad (1)$$

$$\beta_q = 100\sum_{i=0}^{k}\binom{n}{i}P_q^{i}(1-P_q)^{n-i} \qquad (2)$$

P and P_q are expressed here as proportions, i.e. percents divided by 100.

1.1.7 More than one single test (year)

1.1.7.1 Often a candidate variety is grown in two (or three years). The question then arises of how to combine the uniformity information from the individual years. Two methods will be described:

(a) Make the decision after two (or three) years based on the total number of plants examined and the total number of off-types recorded. (A combined test).

(b) Use the result of the first year to see if the data suggests a clear decision (reject or accept). If the decision is not clear then proceed with the second year and decide after the second year. (A two-stage test).

1.1.7.2 However, there are some alternatives (e.g. a decision may be made in each year and a final decision may be reached by rejecting the candidate variety if it shows too many off-types in both (or two out of three years)). Also there are complications when more than one single year test is done. It is therefore suggested that a statistician should be consulted when two (or more) year tests have to be used.

1.1.8 Detailed description of the methods for more than one single test

1.1.8.1 Combined Test

The sample size in test i is n_i. So after the last test we have the total sample size $n = \Sigma n_i$. A decision scheme is set in exactly the same way as if this total sample size had been obtained in a single test. Thus, the total number of off-types recorded through the tests is compared with the maximum number of off-types allowed by the chosen decision scheme.

1.1.8.2 Two-stage Test

1.1.8.2.1 The method for a two-year test may be described as follows: In the first year take a sample of size n. Reject the candidate variety if more than r_1 off-types are recorded and accept the candidate variety if less than a_1 off-types are recorded. Otherwise, proceed to the second year and take a sample of size n (as in the first year) and reject the candidate variety if the total number of off-types recorded in the two years' test is greater than r. Otherwise, accept the candidate variety. The final risks and the expected sample size in such a procedure may be calculated as follows:

$$\alpha = P(K_1 > r_1) + P(K_1 + K_2 > r \mid K_1)$$
$$= P(K_1 > r_1) + P(K_2 > r\text{-}K_1 \mid K_1)$$

$$= \sum_{i=r_1+1}^{n} \binom{n}{i} P^i (1-P)^{n-i} + \sum_{i=\alpha_1}^{r_1} \binom{n}{i} P^i (1-P)^{n-i} \sum_{j=r-i+1}^{n} \binom{n}{i} P^j (1-P)^{n-j} \qquad (3)$$

$$\beta_q = P(K_1 < \alpha_1) + P(K_1 + K_2 \le r \mid K_1)$$
$$= P(K_1 < \alpha_1) + P(K_2 \le r\text{-}K_1 \mid K_1)$$

$$= \sum_{i=0}^{\alpha_1-1} \binom{n}{i} P_q^i (1-P_q)^{n-i} + \sum_{i=\alpha_1}^{r_1} \binom{n}{i} P_q^i (1-P_q)^{n-i} \sum_{j=0}^{r-i} \binom{n}{i} P_q^j (1-P_q)^{n-j} \qquad (4)$$

$$n_e = n\left(1 + \sum_{i=\alpha_1}^{r_1} \binom{n}{i} P^i (1-P)^{n-i}\right) \qquad (5)$$

where

P = population standard
α = probability of actual type I error for P
β_q = probability of actual type II error for q P
n_e = expected sample size
r_1, a_1 and r are decision-parameters
P_q = q times population standard = q P
K_1 and K_2 are the numbers of off-types found in years 1 and 2 respectively.

The decision parameters, a_1, r_1 and r, may be chosen according to the following criteria:

(a) α must be less than α_0, where α_0 is the maximum type I error, i.e. α_0 is 100 minus the required acceptance probability
(b) β_q (for q=5) should be as small as possible but not smaller than α_0
(c) if β_q (for q=5) < α_0 n_e should be as small as possible.

1.1.8.2.2 However, other strategies are available. No tables/figures are produced here as there may be several different decision schemes that satisfy a certain set of risks. It is suggested that a statistician should be consulted if a 2-stage test (or any other sequential tests) is required.

1.1.8.3 Sequential tests

The two-stage test mentioned above is a type of sequential test where the result of the first stage determines whether the test needs to be continued for a second stage. Other types of sequential tests may also be applicable. It may be relevant to consider such tests when the practical work allows analyses of off-types to be carried out at certain stages of the examination. The decision schemes for such methods can be set up in many different ways and it is suggested that a statistician should be consulted when sequential methods are to be used.

1.1.9 Note on type I and type II errors

1.1.9.1 We cannot in general obtain type I-errors that are nice pre-selected values because the number of off-types is discrete. The scheme a of example 2 with 6 plants above showed that we could not obtain an α of 10% - our actual α became 0.6%. Changing the sample size will result in varying α and β values. Figure 3 - as an example - shows that α gets closer to its nominal values at certain sample sizes and that this is also the sample size where β is relatively small.

1.1.9.2 Larger sample sizes are generally beneficial. With same acceptance probability, a larger sample will tend to have proportionally less probability of type II errors. Small sample sizes result in high probabilities of accepting non-uniform varieties. The sample size should therefore be chosen to give an acceptably low level of type II errors. However small increases in the sample size may not always be advantageous. For instance, a sample size of five gives $\alpha = 10\%$ and $\beta_2 = 82\%$ whereas a sample size of six gives $\alpha = 0.6\%$ and $\beta_2 = 98\%$. It appears that the sample sizes, which give α-values in close agreement with the acceptance probability are the largest in the range of sample sizes with a specified maximum number of off-types. Thus, the largest sample sizes in the range of sample sizes with a given maximum number of off-types should be used.

1.1.10 Definition of statistical terms and symbols

The statistical terms and symbols used have the following definitions:

Population standard. The percentage of off-types to be accepted if all the individuals of a variety could be examined. The population standard is fixed for the crop in question and is based on experience.

Acceptance probability. The probability of accepting a uniform variety with P% of off-types. Here P is population standard. However, note that the actual probability of accepting a uniform variety will always be greater than or equal to the acceptance probability in the heading of the table and figures. The probability of accepting a uniform variety and the probability of a type I error sum to 100%. For example, if the type I error probability is 4%, then the probability of accepting a uniform variety is 100 – 4 = 96%, see e.g. figure 1 for n=50). The type I error is indicated on the graph in the figures by the sawtooth peaks between 0 and the upper limit of type I error (for instance 10 on figure 1). The decision schemes are defined so that the actual probability of accepting a uniform variety is always greater than or equal to the acceptance probability in the heading of the table.

Type I error: The error of rejecting a uniform variety.

Type II error: The error of accepting a variety that is too non-uniform.

P Population standard

P_q The assumed true percentage of off-types in a non-uniform variety. $P_q = q\,P$.

In the present document q is equal to 2, 5 or 10. These are only 3 examples to help the visualization of type II errors. The actual percentage of off-types in a variety may take any value. For instance we may examine different varieties which in fact may have respectively 1.6%, 3.8%, 0.2%, … of off-types.

n Sample size
k Maximum number of off-types allowed
α Probability of type I error
β Probability of type II error

1.1.3 Tables and figures

Table and figure 1: **Population Standard = 5%**
Acceptance Probability ≥90%
n=sample size, k=maximum number of off-types

n			k
1	to	2	0
3	to	10	1
11	to	22	2
23	to	35	3
36	to	49	4
50	to	63	5
64	to	78	6
79	to	94	7
95	to	109	8
110	to	125	9
126	to	141	10
142	to	158	11
159	to	174	12
175	to	191	13
192	to	207	14
208	to	224	15
225	to	241	16
242	to	258	17
259	to	275	18
276	to	292	19
293	to	310	20
311	to	327	21
328	to	344	22
345	to	362	23
363	to	379	24
380	to	397	25
398	to	414	26
415	to	432	27
433	to	449	28
450	to	467	29
468	to	485	30
486	to	503	31
504	to	520	32
521	to	538	33
539	to	556	34
557	to	574	35
575	to	592	36
593	to	610	37
611	to	628	38
629	to	646	39
647	to	664	40
665	to	682	41
683	to	700	42
701	to	718	43
719	to	736	44
737	to	754	45
755	to	772	46
773	to	791	47
792	to	809	48
810	to	827	49
828	to	845	50
846	to	864	51
865	to	882	52
883	to	900	53
901	to	918	54
919	to	937	55
938	to	955	56
956	to	973	57
974	to	992	58
993	to	1010	59

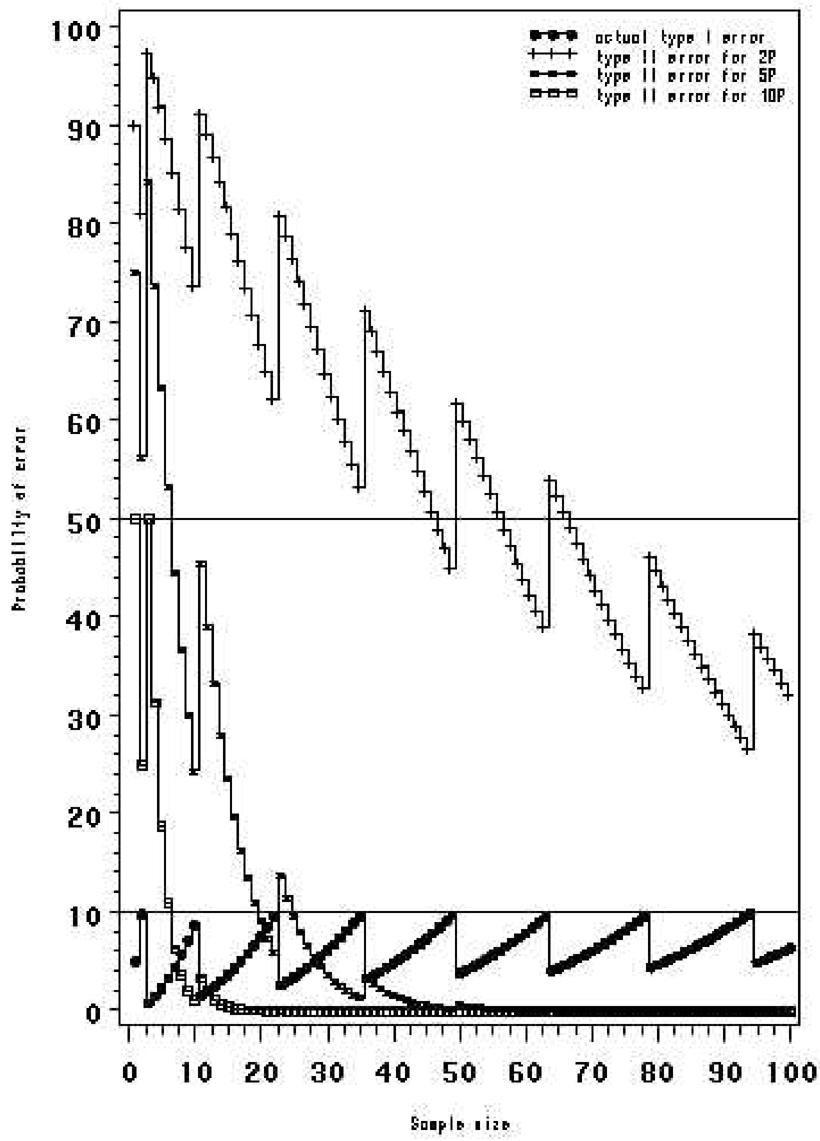


Table and figure 2: **Population Standard = 3%**
Acceptance Probability ≥90%
n=sample size, k=maximum number of off-types

n			k
1	to	3	0
4	to	17	1
18	to	37	2
38	to	58	3
59	to	81	4
82	to	105	5
106	to	130	6
131	to	156	7
157	to	182	8
183	to	208	9
209	to	235	10
236	to	262	11
263	to	289	12
290	to	317	13
318	to	345	14
346	to	373	15
374	to	401	16
402	to	429	17
430	to	457	18
458	to	486	19
487	to	515	20
516	to	543	21
544	to	572	22
573	to	601	23
602	to	630	24
631	to	659	25
660	to	689	26
690	to	718	27
719	to	747	28
748	to	777	29
778	to	806	30
807	to	836	31
837	to	865	32
866	to	895	33
896	to	925	34
926	to	955	35
956	to	984	36
985	to	1014	37
1015	to	1044	38
1045	to	1074	39
1075	to	1104	40
1105	to	1134	41
1135	to	1164	42
1165	to	1195	43
1196	to	1225	44
1226	to	1255	45
1256	to	1285	46
1286	to	1315	47
1316	to	1346	48
1347	to	1376	49
1377	to	1406	50
1407	to	1437	51
1438	to	1467	52
1468	to	1498	53
1499	to	1528	54

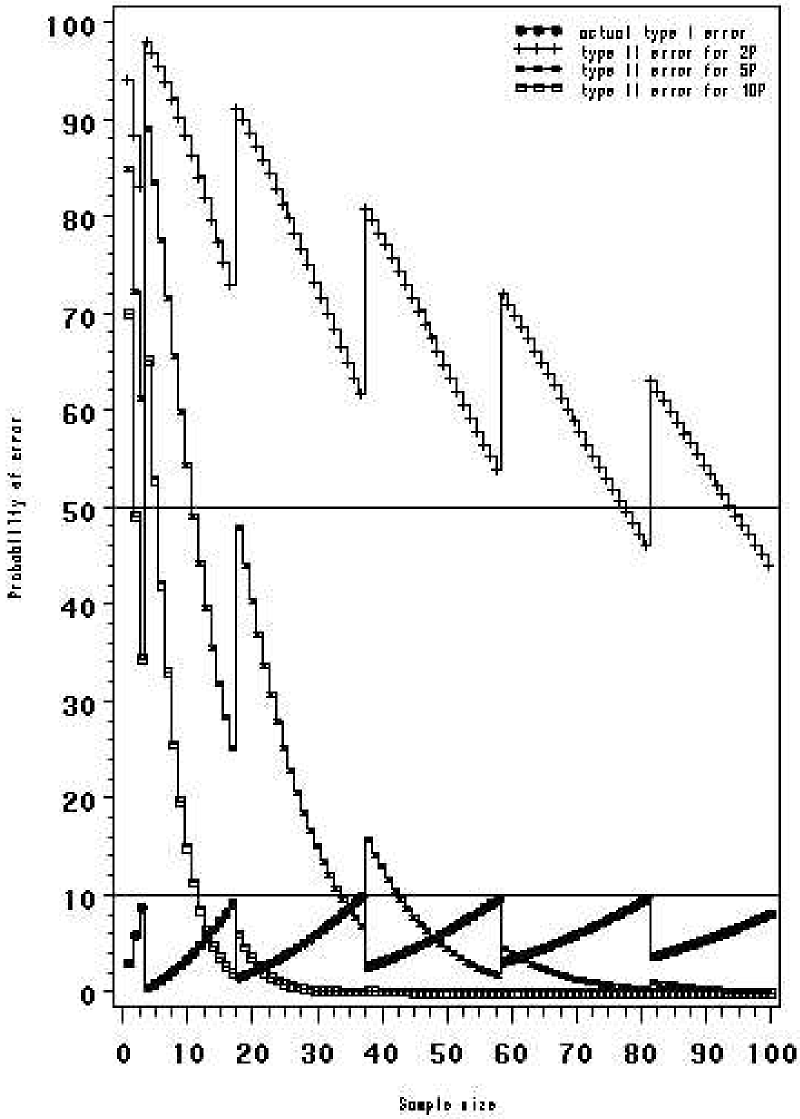


Table and figure 3: **Population Standard = 2%**
Acceptance Probability ≥90%
n=sample size, k=maximum number of off-types

n			k
1	to	5	0
6	to	26	1
27	to	55	2
56	to	87	3
88	to	122	4
123	to	158	5
159	to	195	6
196	to	233	7
234	to	272	8
273	to	312	9
313	to	352	10
353	to	393	11
394	to	433	12
434	to	475	13
476	to	516	14
517	to	558	15
559	to	600	16
601	to	643	17
644	to	685	18
686	to	728	19
729	to	771	20
772	to	814	21
815	to	857	22
858	to	901	23
902	to	944	24
945	to	988	25
989	to	1032	26
1033	to	1076	27
1077	to	1120	28
1121	to	1164	29
1165	to	1208	30
1209	to	1252	31
1253	to	1297	32
1298	to	1341	33
1342	to	1386	34
1387	to	1431	35
1432	to	1475	36
1476	to	1520	37
1521	to	1565	38
1566	to	1610	39
1611	to	1655	40
1656	to	1700	41
1701	to	1745	42
1746	to	1790	43
1791	to	1835	44
1836	to	1881	45
1882	to	1926	46
1927	to	1971	47
1972	to	2000	48

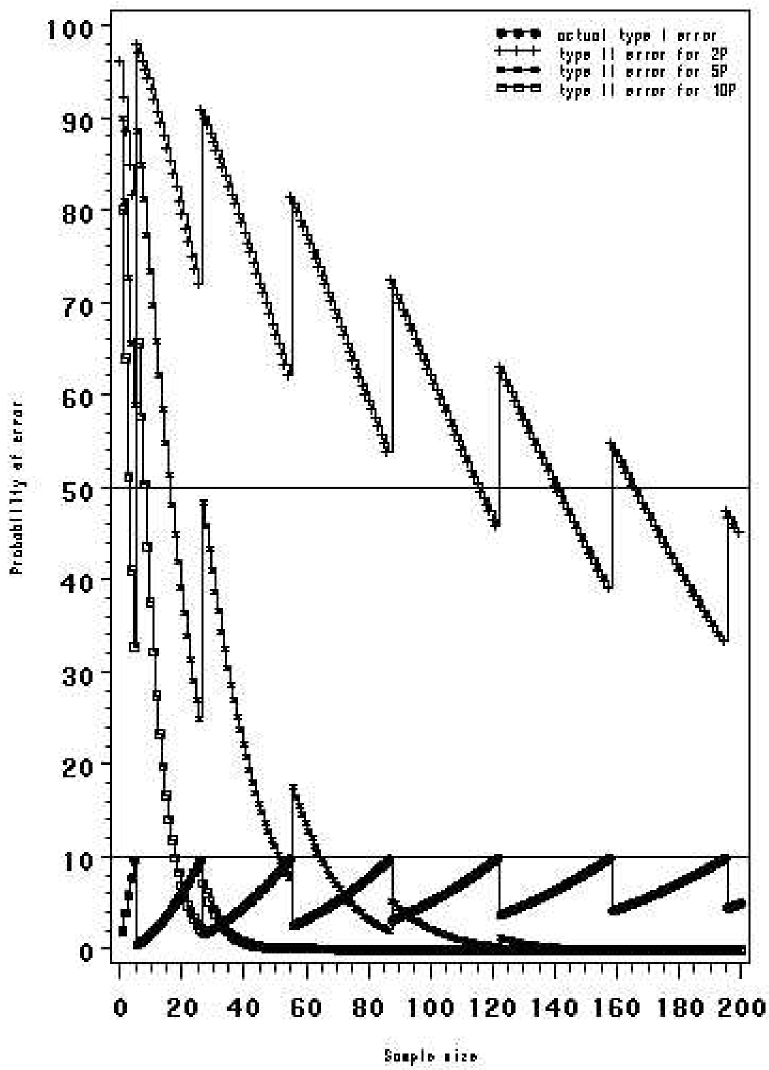


Table and figure 4: **Population Standard = 1%**
Acceptance Probability ≥90%
n=sample size, k=maximum number of off-types

	n		k
1	to	10	0
11	to	53	1
54	to	110	2
111	to	175	3
176	to	244	4
245	to	316	5
317	to	390	6
391	to	466	7
467	to	544	8
545	to	623	9
624	to	703	10
704	to	784	11
785	to	866	12
867	to	948	13
949	to	1031	14
1032	to	1115	15
1116	to	1199	16
1200	to	1284	17
1285	to	1369	18
1370	to	1454	19
1455	to	1540	20
1541	to	1626	21
1627	to	1713	22
1714	to	1799	23
1800	to	1887	24
1888	to	1974	25
1975	to	2061	26
2062	to	2149	27
2150	to	2237	28
2238	to	2325	29
2326	to	2414	30
2415	to	2502	31
2503	to	2591	32
2592	to	2680	33
2681	to	2769	34
2770	to	2858	35
2859	to	2948	36
2949	to	3000	37

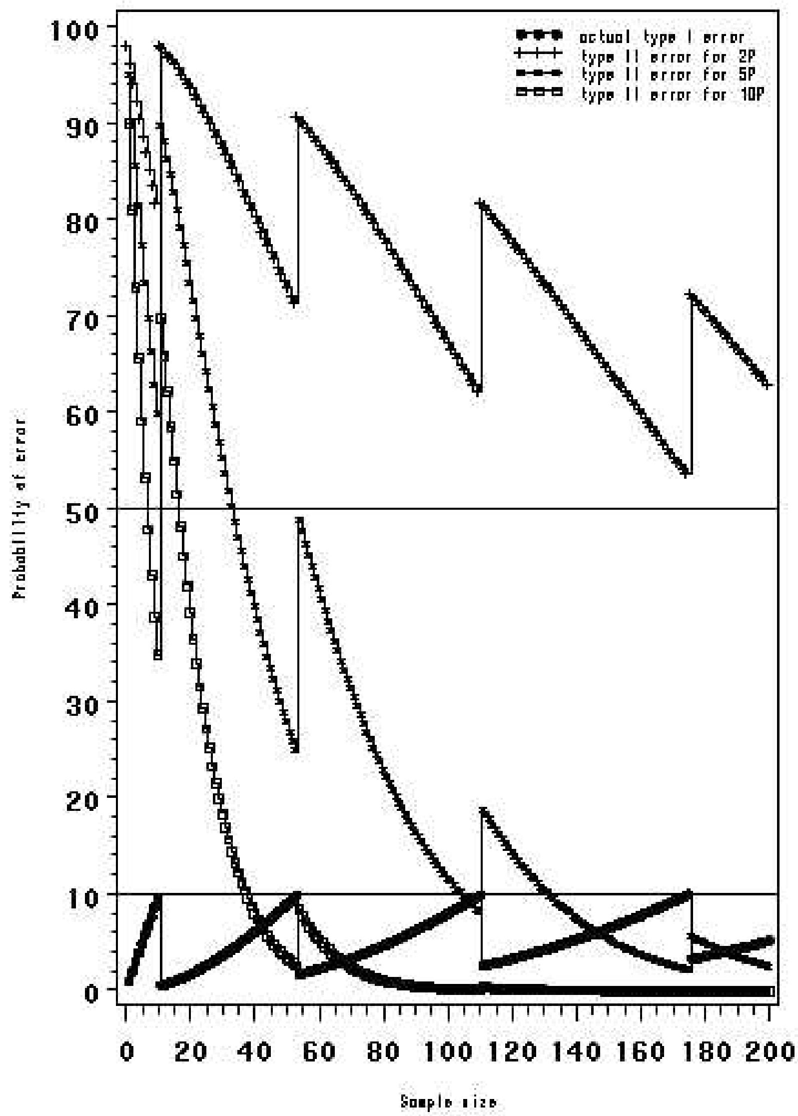


Table and figure 5: **Population Standard = .5%**
Acceptance Probability ≥90%
n=sample size, k=maximum number of off-types

n			k
1	to	21	0
22	to	106	1
107	to	220	2
221	to	349	3
350	to	487	4
488	to	631	5
632	to	780	6
781	to	932	7
933	to	1087	8
1088	to	1245	9
1246	to	1405	10
1406	to	1567	11
1568	to	1730	12
1731	to	1895	13
1896	to	2061	14
2062	to	2228	15
2229	to	2397	16
2398	to	2566	17
2567	to	2736	18
2737	to	2907	19
2908	to	3000	20

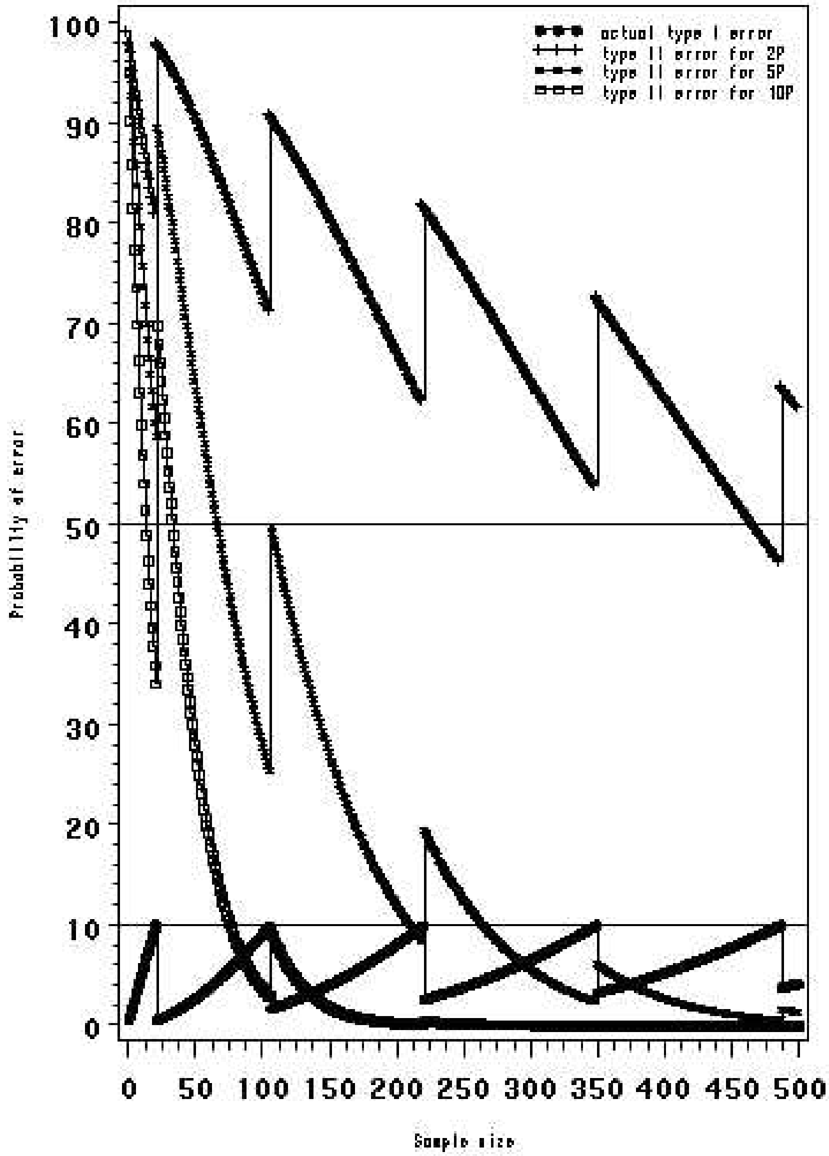


Table and figure 6: **Population Standard = .1%**
Acceptance Probability ≥90%
n=sample size, k=maximum number of off-types

	n		k
1	to	105	0
106	to	532	1
533	to	1102	2
1103	to	1745	3
1746	to	2433	4
2434	to	3000	5

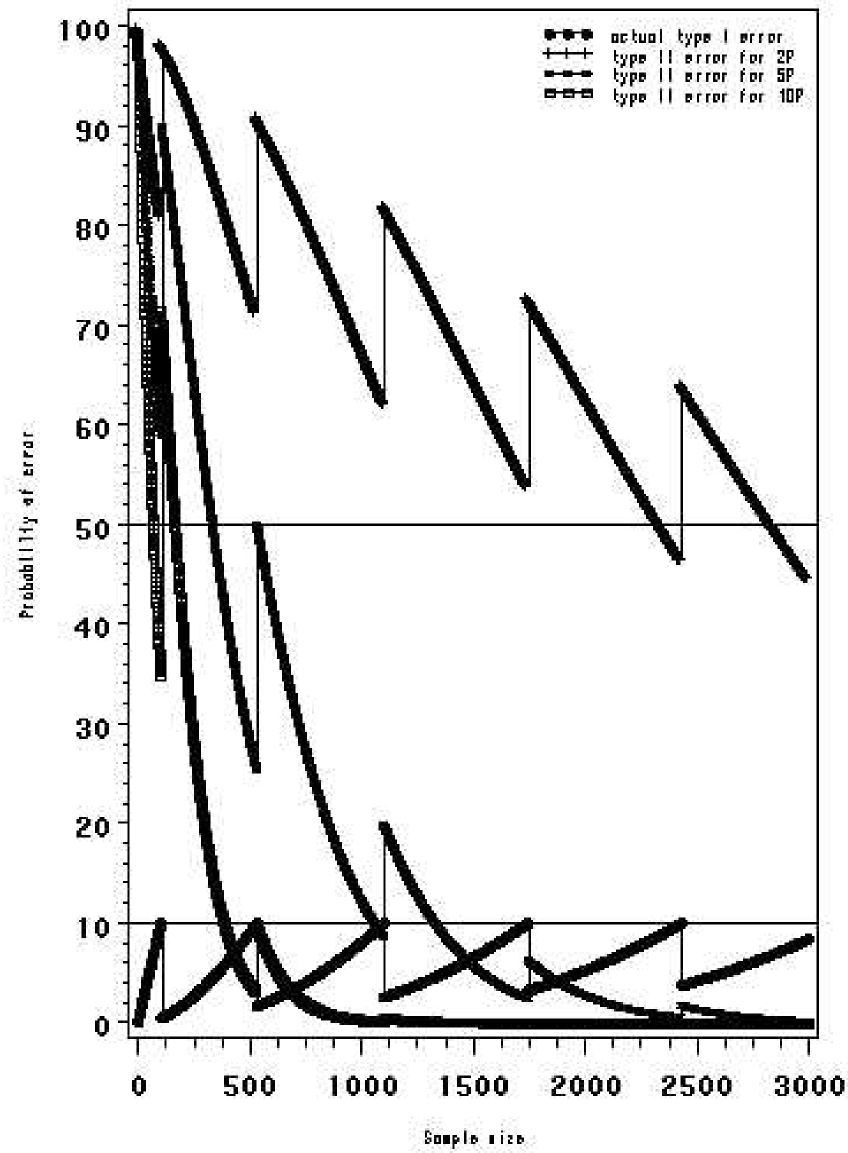


Table and figure 7: **Population Standard = 5%**
Acceptance Probability ≥95%
n=sample size, k=maximum number of off-types

n			k
1	to	1	0
2	to	7	1
8	to	16	2
17	to	28	3
29	to	40	4
41	to	53	5
54	to	67	6
68	to	81	7
82	to	95	8
96	to	110	9
111	to	125	10
126	to	140	11
141	to	155	12
156	to	171	13
172	to	187	14
188	to	203	15
204	to	219	16
220	to	235	17
236	to	251	18
252	to	268	19
269	to	284	20
285	to	300	21
301	to	317	22
318	to	334	23
335	to	351	24
352	to	367	25
368	to	384	26
385	to	401	27
402	to	418	28
419	to	435	29
436	to	452	30
453	to	469	31
470	to	487	32
488	to	504	33
505	to	521	34
522	to	538	35
539	to	556	36
557	to	573	37
574	to	590	38
591	to	608	39
609	to	625	40
626	to	643	41
644	to	660	42
661	to	678	43
679	to	696	44
697	to	713	45
714	to	731	46
732	to	748	47
749	to	766	48
767	to	784	49
785	to	802	50
803	to	819	51
820	to	837	52
838	to	855	53
856	to	873	54
874	to	891	55
892	to	909	56
910	to	926	57
927	to	944	58
945	to	962	59
963	to	980	60
981	to	998	61

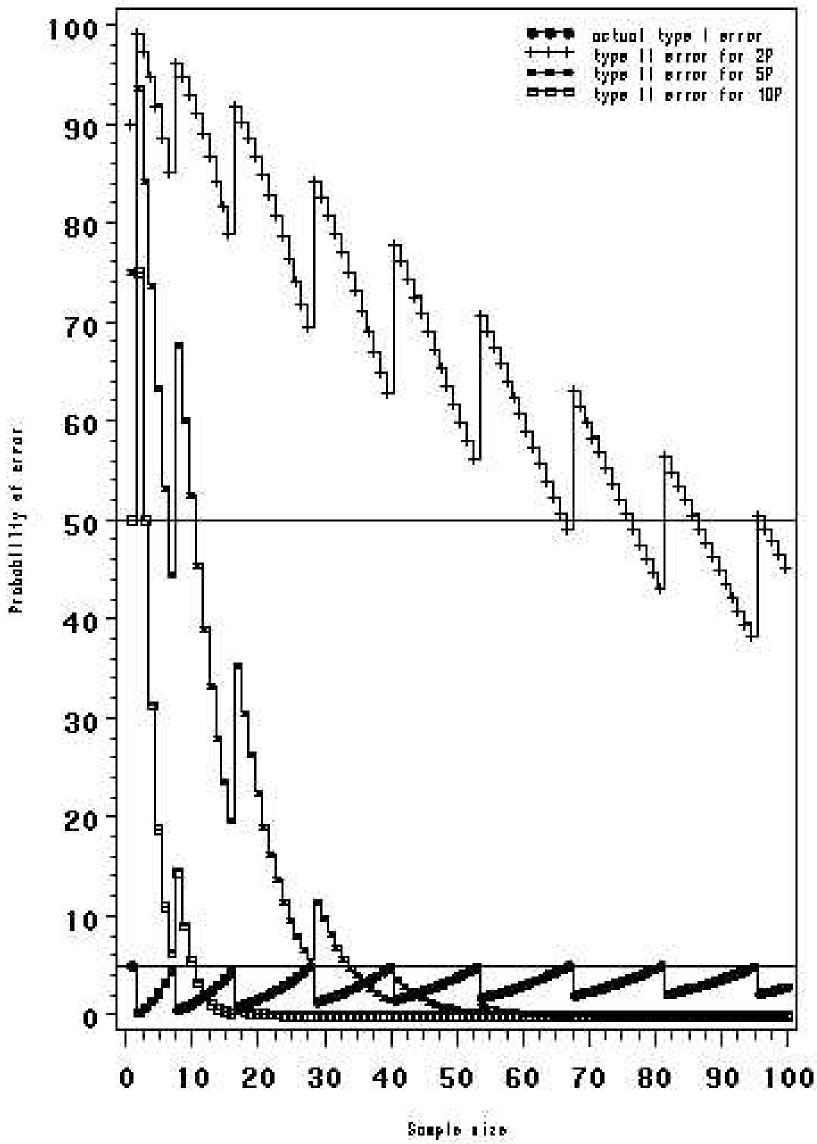


Table and figure 8: **Population Standard = 3%**
Acceptance Probability ≥95%
n=sample size, k=maximum number of off-types

n			k
1	to	1	0
2	to	12	1
13	to	27	2
28	to	46	3
47	to	66	4
67	to	88	5
89	to	110	6
111	to	134	7
135	to	158	8
159	to	182	9
183	to	207	10
208	to	232	11
233	to	258	12
259	to	284	13
285	to	310	14
311	to	337	15
338	to	363	16
364	to	390	17
391	to	417	18
418	to	444	19
445	to	472	20
473	to	499	21
500	to	527	22
528	to	554	23
555	to	582	24
583	to	610	25
611	to	638	26
639	to	666	27
667	to	695	28
696	to	723	29
724	to	751	30
752	to	780	31
781	to	809	32
810	to	837	33
838	to	866	34
867	to	895	35
896	to	924	36
925	to	952	37
953	to	981	38
982	to	1010	39
1011	to	1040	40
1041	to	1069	41
1070	to	1098	42
1099	to	1127	43
1128	to	1156	44
1157	to	1186	45
1187	to	1215	46
1216	to	1244	47
1245	to	1274	48
1275	to	1303	49
1304	to	1333	50
1334	to	1362	51
1363	to	1392	52
1393	to	1422	53
1423	to	1451	54
1452	to	1481	55
1482	to	1511	56
1512	to	1541	57
1542	to	1570	58
1571	to	1600	59
1601	to	1630	60
1631	to	1660	61
1661	to	1690	62
1691	to	1720	63
1721	to	1750	64
1751	to	1780	65
1781	to	1810	66
1811	to	1840	67
1841	to	1870	68
1871	to	1900	69

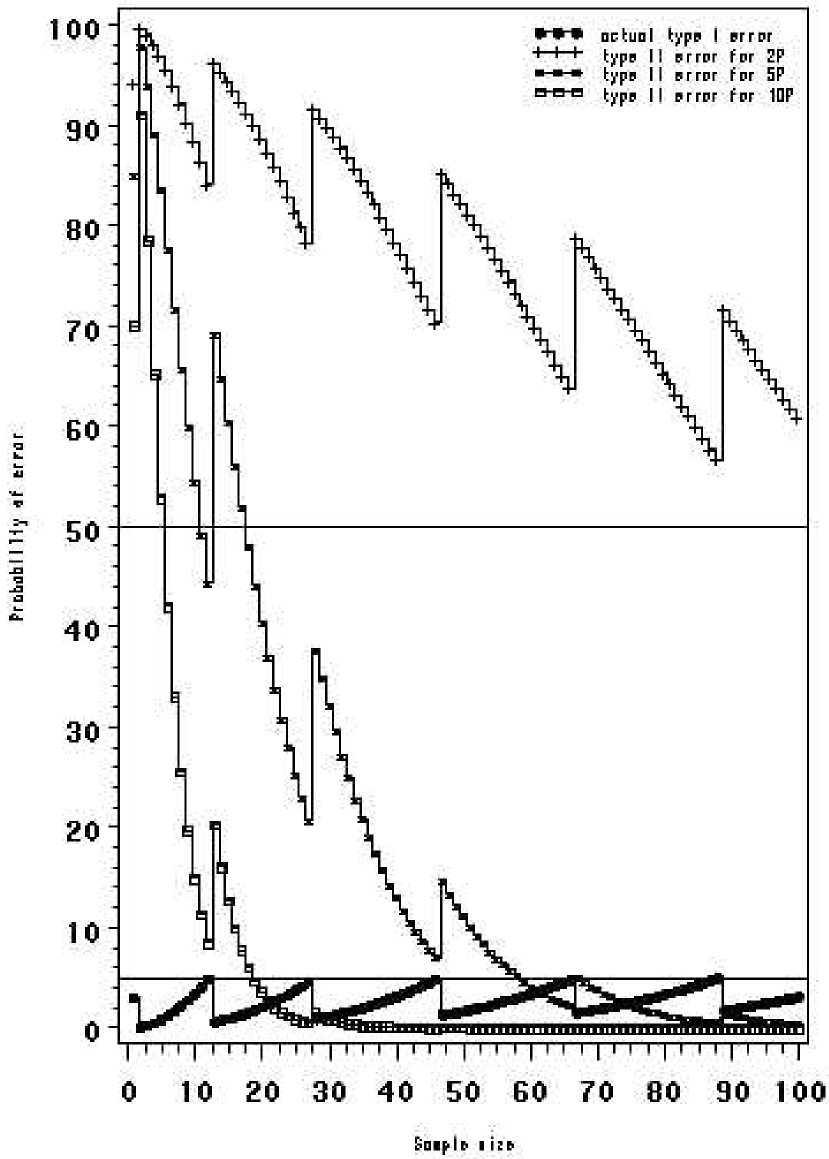


Table and figure 9: **Population Standard = 2%**
Acceptance Probability ≥95%
n=sample size, k=maximum number of off-types

	n		k
1	to	2	0
3	to	18	1
19	to	41	2
42	to	69	3
70	to	99	4
100	to	131	5
132	to	165	6
166	to	200	7
201	to	236	8
237	to	273	9
274	to	310	10
311	to	348	11
349	to	386	12
387	to	425	13
426	to	464	14
465	to	504	15
505	to	544	16
545	to	584	17
585	to	624	18
625	to	665	19
666	to	706	20
707	to	747	21
748	to	789	22
790	to	830	23
831	to	872	24
873	to	914	25
915	to	956	26
957	to	998	27
999	to	1040	28
1041	to	1083	29
1084	to	1126	30
1127	to	1168	31
1169	to	1211	32
1212	to	1254	33
1255	to	1297	34
1298	to	1340	35
1341	to	1383	36
1384	to	1427	37
1428	to	1470	38
1471	to	1514	39
1515	to	1557	40
1558	to	1601	41
1602	to	1645	42
1646	to	1689	43
1690	to	1732	44
1733	to	1776	45
1777	to	1820	46
1821	to	1864	47
1865	to	1909	48
1910	to	1953	49
1954	to	1997	50
1998	to	2000	51

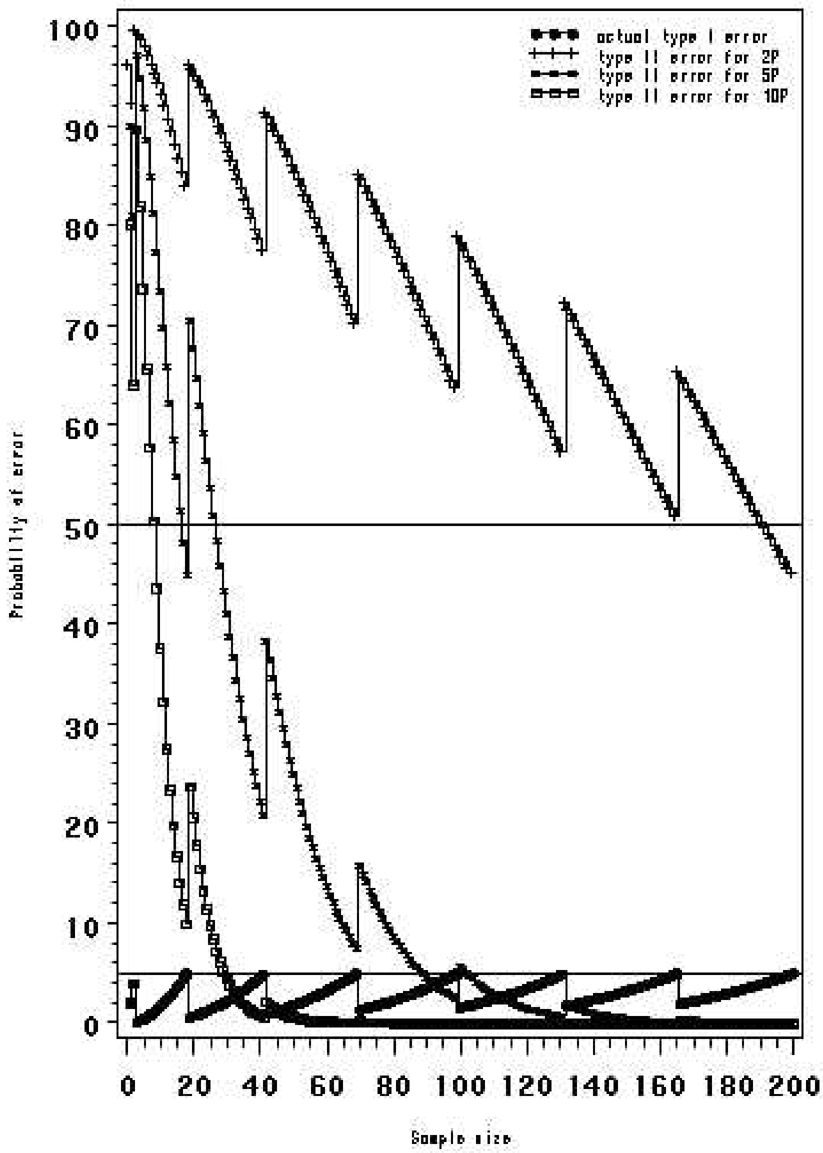


Table and figure 10: **Population Standard = 1%**
Acceptance Probability ≥95%
n=sample size, k=maximum number of off-types

	n		k
1	to	5	0
6	to	35	1
36	to	82	2
83	to	137	3
138	to	198	4
199	to	262	5
263	to	329	6
330	to	399	7
400	to	471	8
472	to	544	9
545	to	618	10
619	to	694	11
695	to	771	12
772	to	848	13
849	to	927	14
928	to	1006	15
1007	to	1085	16
1086	to	1166	17
1167	to	1246	18
1247	to	1328	19
1329	to	1410	20
1411	to	1492	21
1493	to	1575	22
1576	to	1658	23
1659	to	1741	24
1742	to	1825	25
1826	to	1909	26
1910	to	1993	27
1994	to	2078	28
2079	to	2163	29
2164	to	2248	30
2249	to	2333	31
2334	to	2419	32
2420	to	2505	33
2506	to	2591	34
2592	to	2677	35
2678	to	2763	36
2764	to	2850	37
2851	to	2937	38
2938	to	3000	39

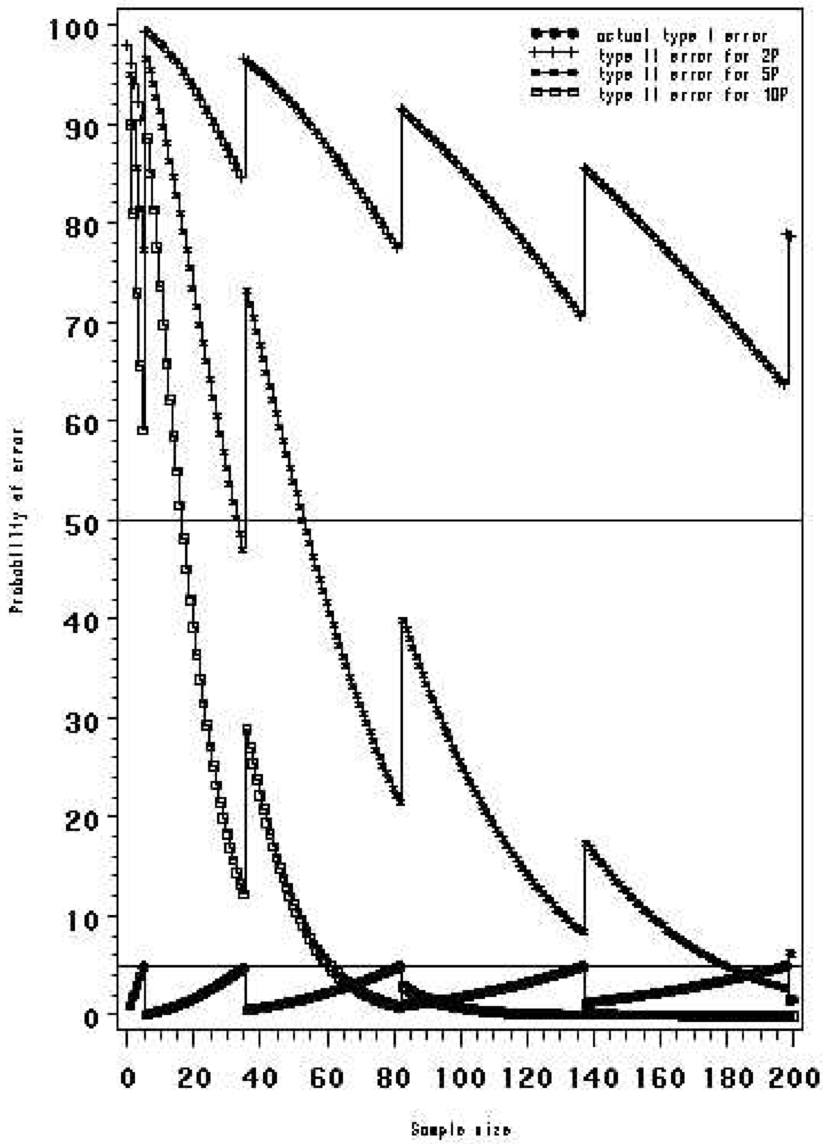


Table and figure 11: **Population Standard = .5%**
Acceptance Probability ≥95%
n=sample size, k=maximum number of off-types

	n		k
1	to	10	0
11	to	71	1
72	to	164	2
165	to	274	3
275	to	395	4
396	to	523	5
524	to	658	6
659	to	797	7
798	to	940	8
941	to	1086	9
1087	to	1235	10
1236	to	1386	11
1387	to	1540	12
1541	to	1695	13
1696	to	1851	14
1852	to	2009	15
2010	to	2169	16
2170	to	2329	17
2330	to	2491	18
2492	to	2653	19
2654	to	2817	20
2818	to	2981	21
2982	to	3000	22

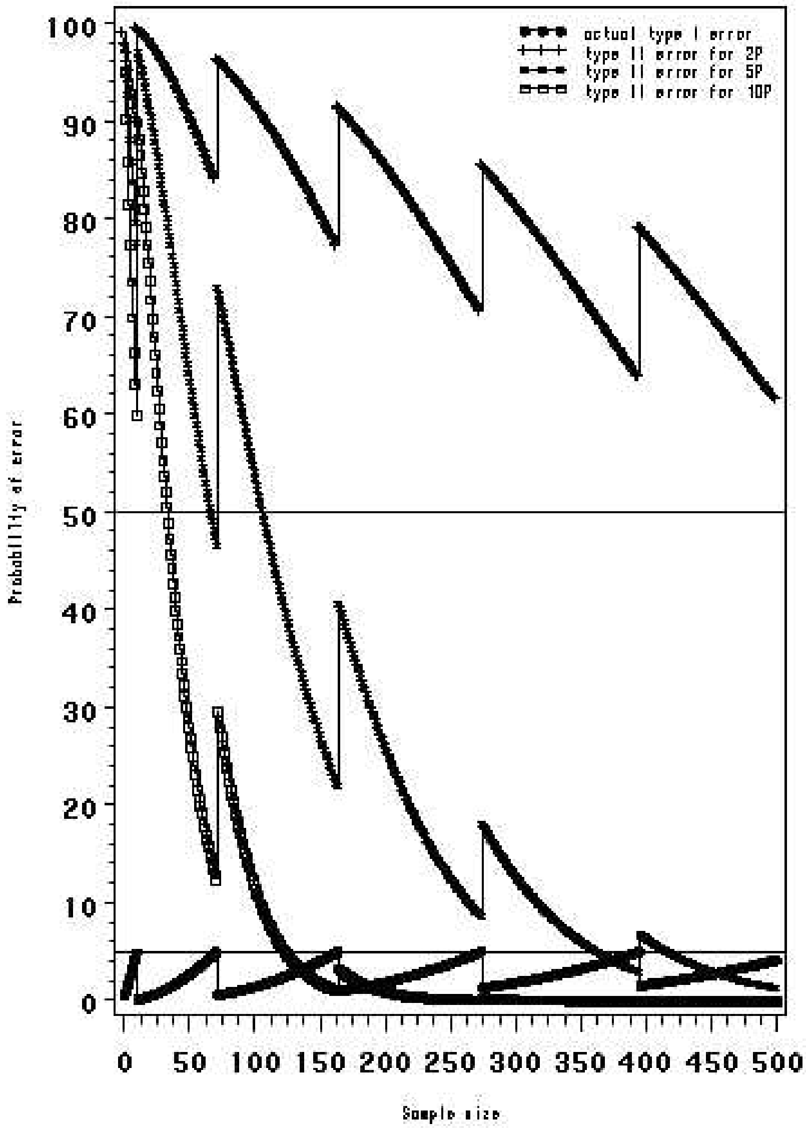


Table and figure 12: **Population Standard = .1%**
Acceptance Probability ≥95%
n=sample size, k=maximum number off-types

n			k
1	to	51	0
52	to	355	1
356	to	818	2
819	to	1367	3
1368	to	1971	4
1972	to	2614	5
2615	to	3000	6

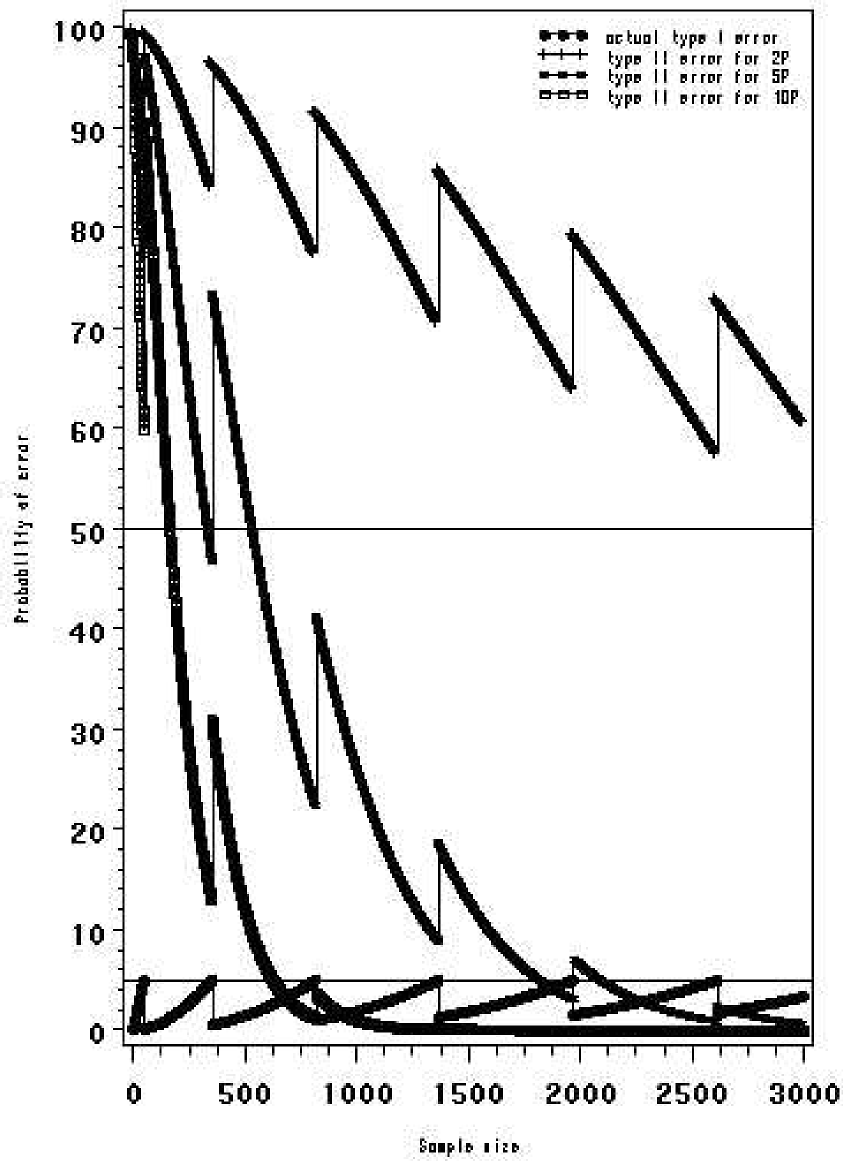


Table and figure 13: **Population Standard = 5%**
Acceptance Probability ≥99%
n=sample size, k=maximum number of off-types

n			k
1	to	3	1
4	to	9	2
10	to	17	3
18	to	26	4
27	to	37	5
38	to	48	6
49	to	60	7
61	to	72	8
73	to	85	9
86	to	98	10
99	to	111	11
112	to	124	12
125	to	138	13
139	to	152	14
153	to	167	15
168	to	181	16
182	to	196	17
197	to	210	18
211	to	225	19
226	to	240	20
241	to	255	21
256	to	270	22
271	to	286	23
287	to	301	24
302	to	317	25
318	to	332	26
333	to	348	27
349	to	364	28
365	to	380	29
381	to	395	30
396	to	411	31
412	to	427	32
428	to	444	33
445	to	460	34
461	to	476	35
477	to	492	36
493	to	508	37
509	to	525	38
526	to	541	39
542	to	558	40
559	to	574	41
575	to	591	42
592	to	607	43
608	to	624	44
625	to	640	45
641	to	657	46
658	to	674	47
675	to	690	48
691	to	707	49
708	to	724	50
725	to	741	51
742	to	758	52
759	to	775	53
776	to	792	54
793	to	809	55
810	to	826	56
827	to	843	57
844	to	860	58
861	to	877	59
878	to	894	60
895	to	911	61
912	to	928	62
929	to	945	63
946	to	962	64
963	to	979	65
980	to	997	66
998	to	1014	67

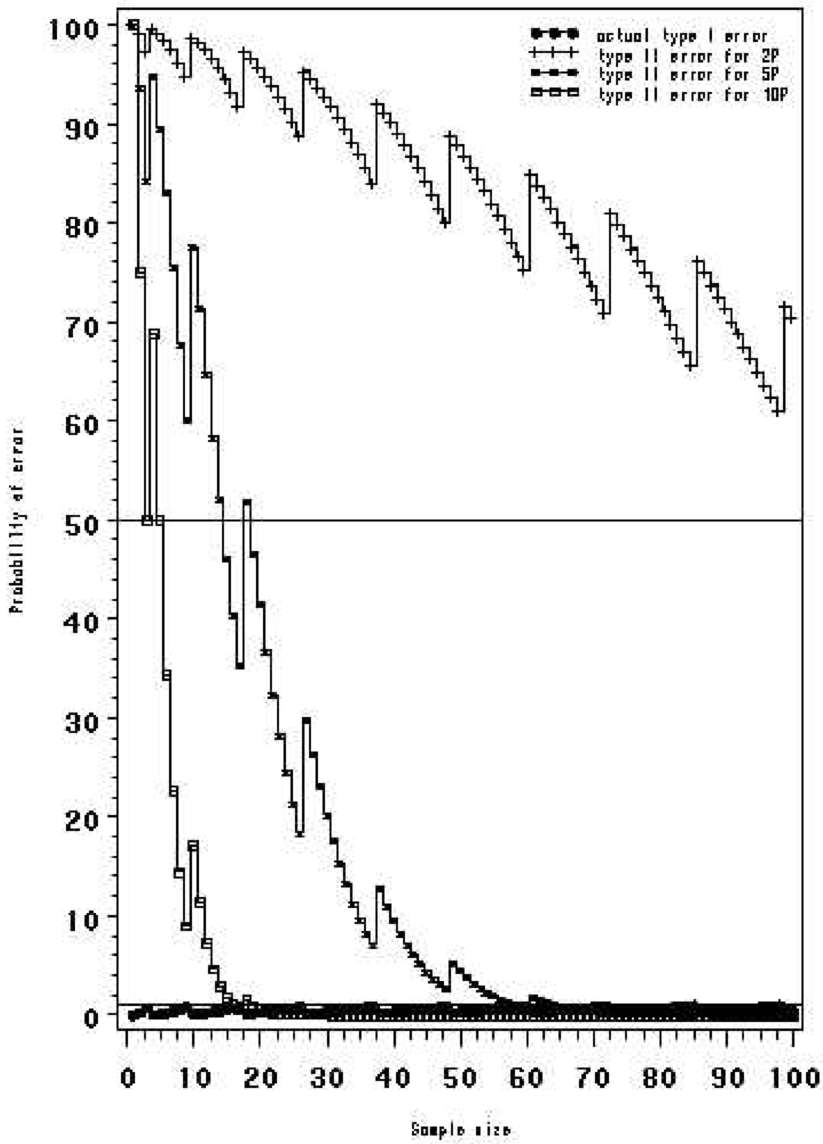


Table and figure 14: **Population Standard = 3%**
Acceptance Probability ≥99%
n=sample size, k=maximum number of off-types

	n		k
1	to	5	1
6	to	15	2
16	to	28	3
29	to	44	4
45	to	61	5
62	to	79	6
80	to	98	7
99	to	119	8
120	to	140	9
141	to	161	10
162	to	183	11
184	to	206	12
207	to	229	13
230	to	252	14
253	to	276	15
277	to	300	16
301	to	324	17
325	to	348	18
349	to	373	19
374	to	398	20
399	to	423	21
424	to	448	22
449	to	474	23
475	to	499	24
500	to	525	25
526	to	551	26
552	to	577	27
578	to	603	28
604	to	629	29
630	to	656	30
657	to	682	31
683	to	709	32
710	to	736	33
737	to	763	34
764	to	789	35
790	to	816	36
817	to	844	37
845	to	871	38
872	to	898	39
899	to	925	40
926	to	953	41
954	to	980	42
981	to	1008	43
1009	to	1035	44
1036	to	1063	45
1064	to	1091	46
1092	to	1119	47
1120	to	1146	48
1147	to	1174	49
1175	to	1202	50
1203	to	1230	51
1231	to	1258	52
1259	to	1286	53
1287	to	1315	54
1316	to	1343	55
1344	to	1371	56
1372	to	1399	57
1400	to	1428	58
1429	to	1456	59
1457	to	1484	60
1485	to	1513	61

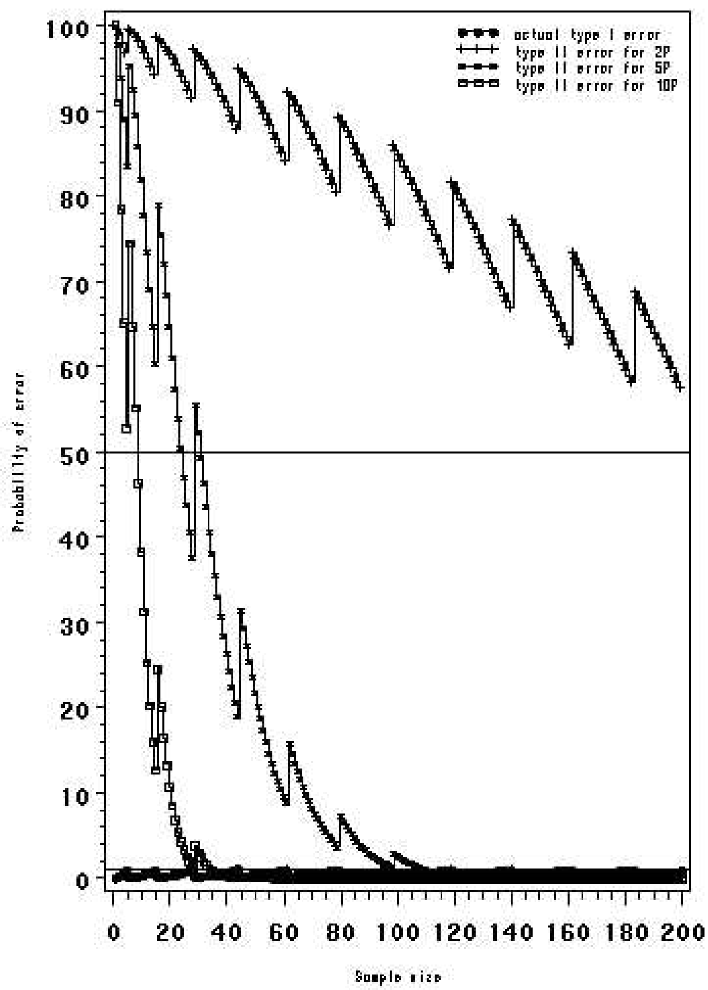


Table and figure 15: **Population Standard = 2%**
Acceptance Probability ≥99%
n=sample size, k=maximum number of off-types

	n		k
1	to	7	1
8	to	22	2
23	to	42	3
43	to	65	4
66	to	90	5
91	to	118	6
119	to	147	7
148	to	177	8
178	to	208	9
209	to	241	10
242	to	274	11
275	to	307	12
308	to	342	13
343	to	377	14
378	to	412	15
413	to	448	16
449	to	484	17
485	to	521	18
522	to	558	19
559	to	595	20
596	to	632	21
633	to	670	22
671	to	708	23
709	to	747	24
748	to	785	25
786	to	824	26
825	to	863	27
864	to	902	28
903	to	942	29
943	to	981	30
982	to	1021	31
1022	to	1061	32
1062	to	1101	33
1102	to	1141	34
1142	to	1182	35
1183	to	1222	36
1223	to	1263	37
1264	to	1303	38
1304	to	1344	39
1345	to	1385	40
1386	to	1426	41
1427	to	1467	42
1468	to	1509	43
1510	to	1550	44
1551	to	1591	45
1592	to	1633	46
1634	to	1675	47
1676	to	1716	48
1717	to	1758	49
1759	to	1800	50
1801	to	1842	51
1843	to	1884	52
1885	to	1926	53
1927	to	1968	54
1969	to	2000	55

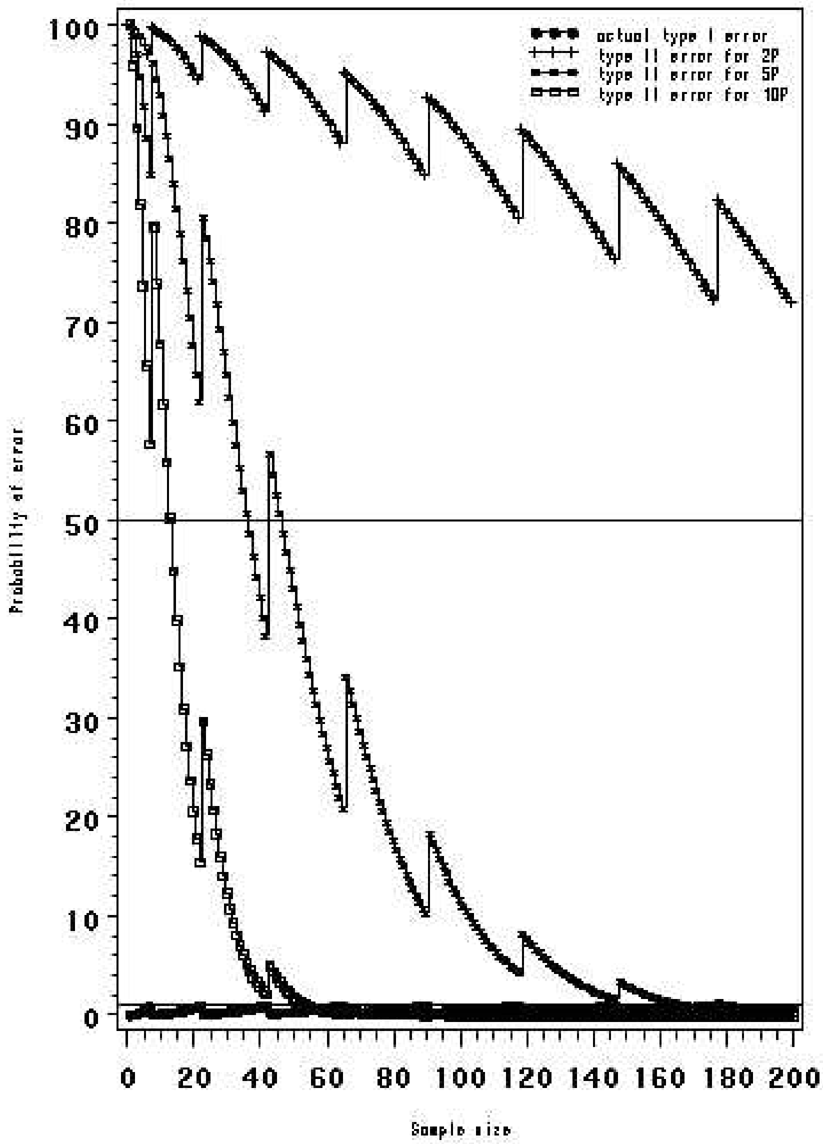


Table and figure 16: **Population Standard = 1%**
Acceptance Probability ≥99%
n=sample size, k=maximum number of off-types

	n		k
1	to	1	0
2	to	15	1
16	to	44	2
45	to	83	3
84	to	129	4
130	to	180	5
181	to	234	6
235	to	292	7
293	to	353	8
354	to	415	9
416	to	479	10
480	to	545	11
546	to	612	12
613	to	681	13
682	to	750	14
751	to	821	15
822	to	893	16
894	to	965	17
966	to	1038	18
1039	to	1112	19
1113	to	1186	20
1187	to	1261	21
1262	to	1337	22
1338	to	1413	23
1414	to	1489	24
1490	to	1566	25
1567	to	1644	26
1645	to	1722	27
1723	to	1800	28
1801	to	1879	29
1880	to	1958	30
1959	to	2037	31
2038	to	2117	32
2118	to	2197	33
2198	to	2277	34
2278	to	2358	35
2359	to	2439	36
2440	to	2520	37
2521	to	2601	38
2602	to	2683	39
2684	to	2764	40
2765	to	2846	41
2847	to	2929	42
2930	to	3000	43

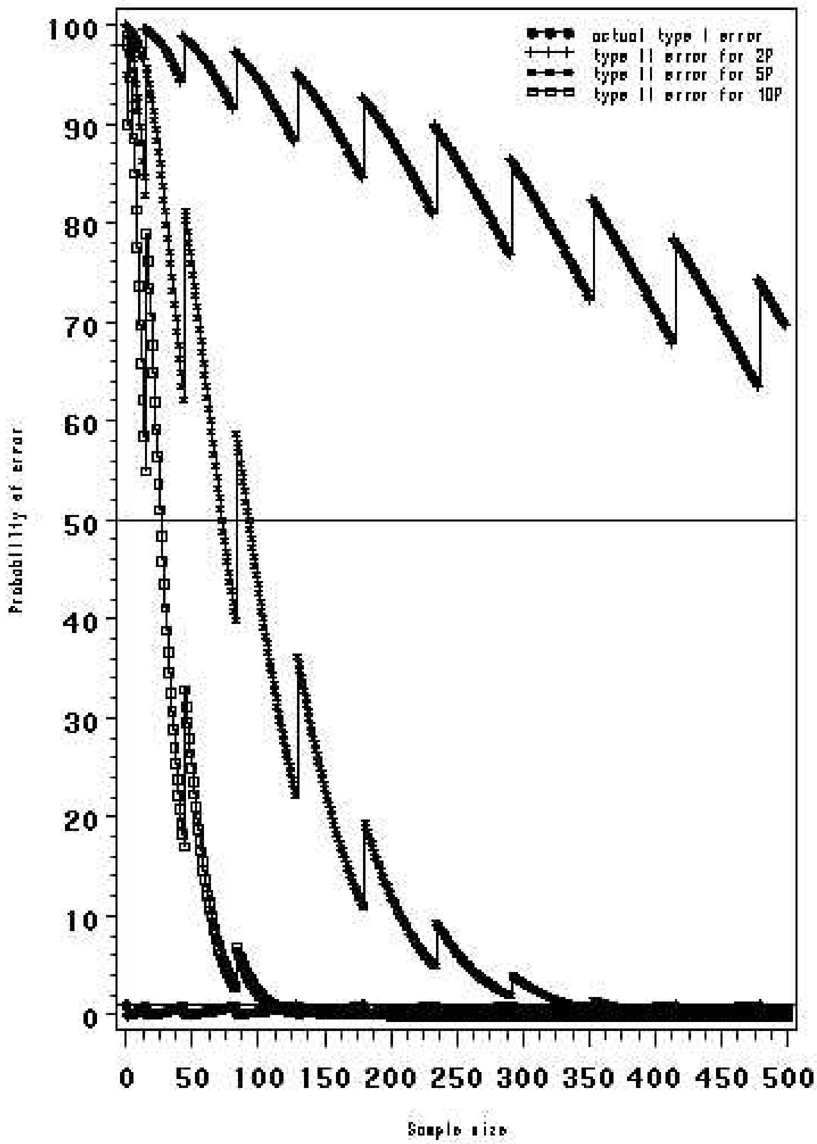


Table and figure 17: **Population Standard = .5%**
Acceptance Probability ≥99%
n=sample size, k=maximum number of off-types

n			k
1	to	2	0
3	to	30	1
31	to	87	2
88	to	165	3
166	to	257	4
258	to	358	5
359	to	467	6
468	to	583	7
584	to	703	8
704	to	828	9
829	to	956	10
957	to	1088	11
1089	to	1222	12
1223	to	1359	13
1360	to	1498	14
1499	to	1639	15
1640	to	1782	16
1783	to	1926	17
1927	to	2072	18
2073	to	2220	19
2221	to	2369	20
2370	to	2519	21
2520	to	2670	22
2671	to	2822	23
2823	to	2975	24
2976	to	3000	25

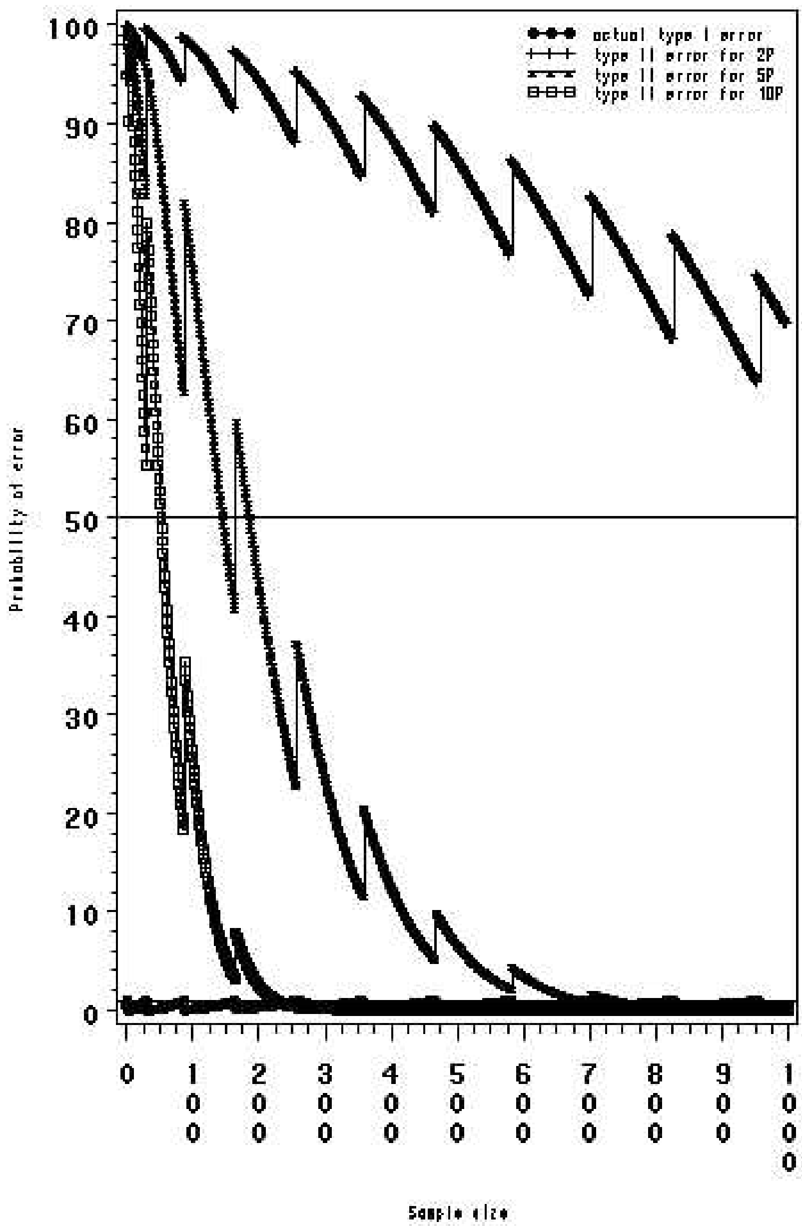


Table and figure 18: **Population Standard = .1%**
Acceptance Probability ≥99%
n=sample size, k=maximum number of off-types

n			k
1	to	10	0
11	to	148	1
149	to	436	2
437	to	824	3
825	to	1280	4
1281	to	1786	5
1787	to	2332	6
2333	to	2908	7
2909	to	3000	8

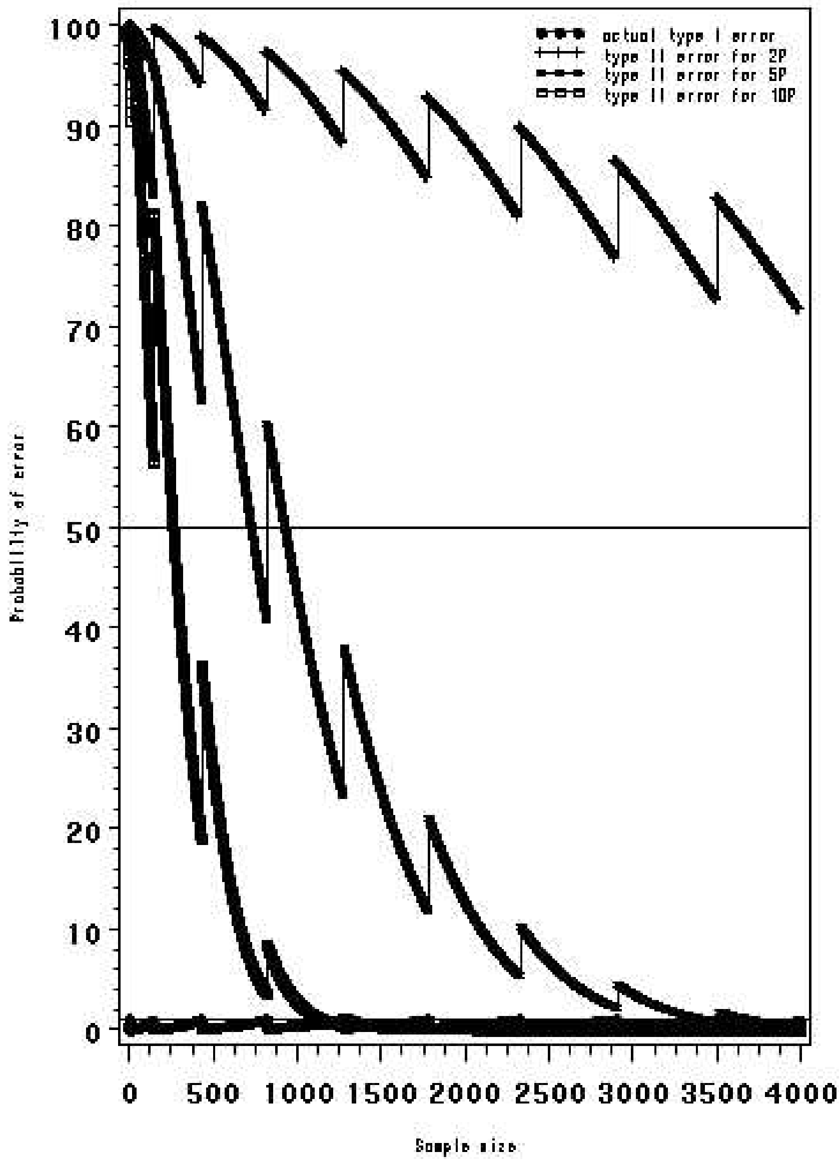


Table and figure 19: **Population Standard = 10%**
Acceptance Probability ≥ 90%
n=sample size, k=maximum number of off-types

n			k
1	to	1	0
2	to	5	1
6	to	11	2
12	to	18	3
19	to	25	4
26	to	32	5
33	to	40	6
41	to	47	7
48	to	55	8
56	to	63	9
64	to	71	10
72	to	79	11
80	to	88	12
89	to	96	13
97	to	104	14
105	to	113	15
114	to	121	16
122	to	130	17
131	to	138	18
139	to	147	19
148	to	156	20
157	to	164	21
165	to	173	22
174	to	182	23
183	to	191	24
192	to	199	25
200	to	200	26

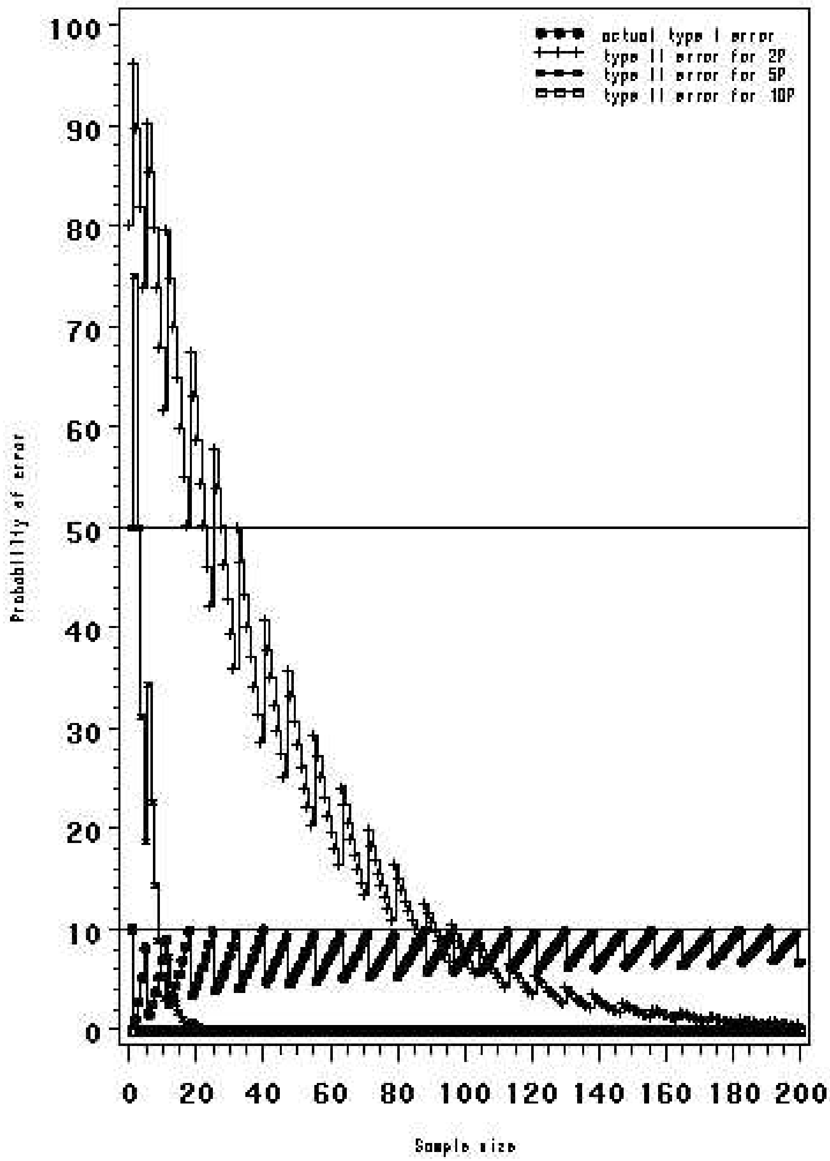


Table and figure 20: **Population Standard = 10%**
Acceptance Probability ≥ 95%
n=sample size, k=maximum number of off-types

n			k
1	to	3	1
4	to	8	2
9	to	14	3
15	to	20	4
21	to	27	5
28	to	34	6
35	to	41	7
42	to	48	8
49	to	56	9
57	to	63	10
64	to	71	11
72	to	79	12
80	to	86	13
87	to	94	14
95	to	102	15
103	to	110	16
111	to	119	17
120	to	127	18
128	to	135	19
136	to	143	20
144	to	152	21
153	to	160	22
161	to	168	23
169	to	177	24
178	to	185	25
186	to	194	26
195	to	200	27

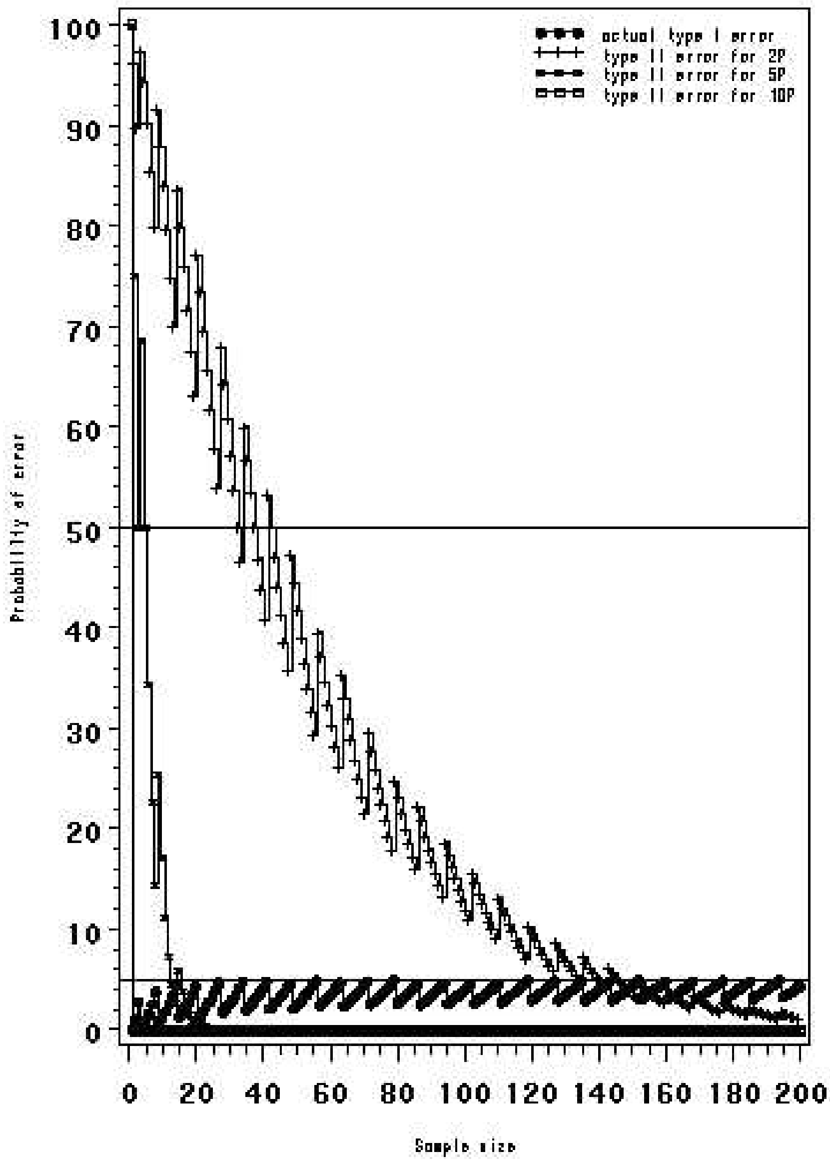
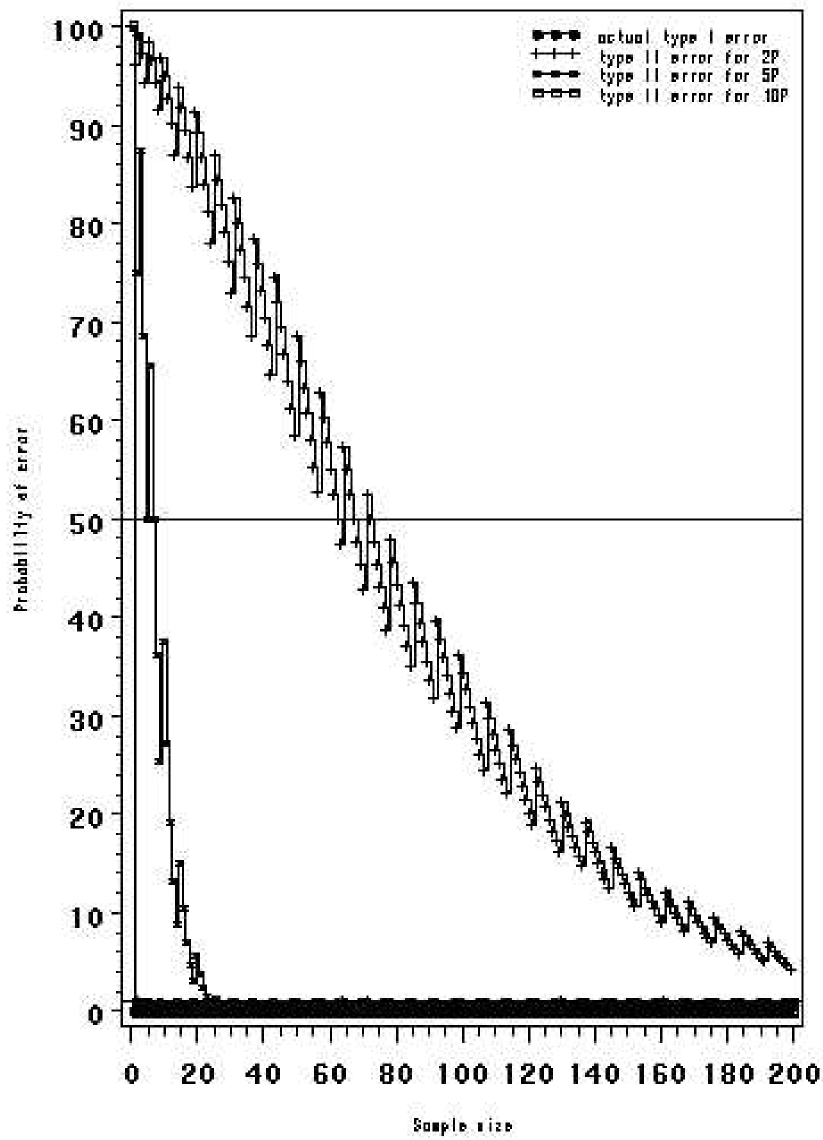


Table and figure 21 : **Population Standard = 10%**
Acceptance Probability ≥ 99%
n=sample size, k=maximum number of off-types

n			k
1	to	2	1
3	to	5	2
6	to	9	3
10	to	14	4
15	to	19	5
20	to	25	6
26	to	31	7
32	to	37	8
38	to	43	9
44	to	50	10
51	to	57	11
58	to	64	12
65	to	71	13
72	to	78	14
79	to	85	15
86	to	92	16
93	to	99	17
100	to	107	18
108	to	114	19
115	to	122	20
123	to	130	21
131	to	137	22
138	to	145	23
146	to	153	24
154	to	161	25
162	to	168	26
169	to	176	27
177	to	184	28
185	to	192	29
193	to	200	30



2. LSD

to be provided by the TWC (see extract below from TGP/9/1 Draft 5)

[5.2.4.10 The General Introduction clarifies the situation with regard to measured, quantitative characteristics for vegetatively propagated and self-pollinated varieties as follows:

> "5.5.3 Measured Characteristics
>
> The following paragraphs provide guidance on the typical methods for examining distinctness according to the particular features of propagation of the variety:
>
> […]
>
> "5.5.3.1 Self-Pollinated and Vegetatively Propagated Varieties
>
> "UPOV has endorsed several statistical methods for the handling of measured quantitative characteristics. One method established for self-pollinated and vegetatively propagated varieties is that varieties can be considered clearly distinguishable if the difference between two varieties equals or exceeds the Least Significant Difference (LSD) at a specified probability level with the same sign over an appropriate period, even if they are described by the same state of expression. This is a relatively simple method but is considered appropriate for self-pollinated and vegetatively propagated varieties because the level of variation within such varieties is relatively low. Further details are provided in document TGP/9, "Examining Distinctness"."

5.2.4.11 Information on the Least Significant Difference (LSD) method is provided in TGP/8 ***[cross ref.]***.]

[TWC Chairperson: to find users and crops in which this approach is used]

3. THE COMBINED OVER-YEARS CRITERION FOR DISTINCTNESS AND UNIFORMITY

3.1 The Combined Over-Years Distinctness Criterion (COYD)

3.1.1 Summary

3.1.1.1 Document TGP/9, Section 5.2.4.14 explains that "To assess distinctness for varieties on the basis of a quantitative characteristic it is possible to calculate a minimum distance between varieties such that, when the distance calculated between a pair of varieties is greater than this minimum distance, they may be considered as "distinct" in respect of that characteristic. Amongst the possible ways of establishing minimum distances is the method known as the Combined-Over-Years Distinctness (COYD). The COYD analysis takes into account variations between years. Its main use is for cross-pollinated, including synthetic, varieties but, if desired, it can also be used for self-pollinated and vegetatively propagated varieties in certain circumstances. This method requires the size of the differences to be sufficiently consistent over the years and takes into account the variation between years." ***[cross ref.]***.

3.1.1.2 The COYD method involves:

– for each characteristic, taking the variety means from the two or three years of trials for candidates and established varieties and producing over-year means for the varieties;

– calculate a least significant difference (LSD), based on variety-by-years variation, for comparing variety means.

– if the over-years mean difference between two varieties is greater than or equal to the LSD then the varieties are said to be distinct in respect of that characteristic.

3.1.1.3 The main advantages of the COYD method are:

– it combines information from several seasons into a single criterion (the "COYD criterion") in a simple and straightforward way;

– it ensures that judgements about distinctness will be reproducible in other seasons; in other words, the same genetic material should give similar results, within reasonable limits, from season-to-season;

– the risks of making a wrong judgement about distinctness are constant for all characteristics.

3.1.2 Introduction

3.1.2.1 The following sections describe:

– the principles underlying the COYD method;

– UPOV recommendations on the application of COYD to individual species;

– details of ways in which the procedure can be adapted to deal with special circumstances. This includes when there are small numbers of varieties in trial;

– the computer software which is available to apply the procedure.

3.1.3 The COYD Method

3.1.3.1 The COYD method aims to establish for each characteristic a minimum difference, or distance, which, if achieved by two varieties in trials over a period of two or three years, would indicate that those varieties are distinct with a specified degree of confidence.

3.1.3.2 The method uses variation in variety expression of a characteristic from year-to-year to establish the minimum distance. Thus, characteristics which show consistency in variety ranking between years will have smaller minimum distances than those with marked changes in ranking.

3.1.3.3 Calculation of the COYD criterion involves analysing the variety-by-year table of means for each characteristic to get an estimate of the varieties-by-years variation, which is used in the next step: to calculate an LSD. Usually data for all candidate and established varieties which appeared in trials over the two or three test years are included in the table, the analysis is by analysis of variance, the varieties-by-years mean square is used as the estimate of the varieties-by-years variation, and the resulting LSD is known as the COYD LSD. However, where there are small numbers of varieties in trial, the approach is different.

3.1.3.4 Where there are small numbers of varieties in trial, the table used to calculate of the COYD criterion is expanded with means from other varieties and earlier years, a different method of analysis is used to get a varieties-by-years mean square to estimate the varieties-by-years variation, and the resulting LSD is known as the Long-Term LSD. This is discussed later.

3.1.3.5 Equation [1]

$$LSD_p = t_p \times \sqrt{2} \times SE(\bar{x})$$

where $SE(\bar{x})$ is the standard error of a variety's over-year mean calculated as:

$$SE(\bar{x}) = \sqrt{\frac{\text{varieties - by - years mean square}}{\text{number of test years}}}$$

and t_p is the value in Student's t table appropriate for a two-tailed test with probability p and with degrees of freedom associated with the variety-by-years mean square. The probability level p that is appropriate for individual species is discussed under UPOV RECOMMENDATIONS ON COYD below.

3.1.3.6 An example of the application of COYD to a small data set is given in Figure 1. Statistical details of the method are in Part I: Section 2.1.8 *[cross ref.]* . Further information about the COYD criterion can be found in Patterson and Weatherup (1984).

3.1.4 Use of COYD

3.1.4.1 COYD is an appropriate method for assessing the distinctness of varieties where:

– the characteristic is quantitative;

– there are some differences between plants (or plots) of a variety.

– observations are made on a plant (or plot) basis over two or more years;

3.1.4.2 A pair of varieties is considered to be distinct if their over-years means differ by at least the COYD LSD in one or more characteristics.

3.1.4.3 The UPOV recommended probability level p for the t_p value used to calculate the COYD LSD differs depending on the crop and for some crops depends on whether the test is over two or three years. The testing schemes that usually arise in distinctness testing are described in [……] *[cross ref.]* .

3.1.5 Adapting COYD to special circumstances

3.1.5.1 Differences between years in the range of expression of a characteristic. Occasionally, marked differences between years in the range of expression of a characteristic can occur. For example, in a late spring, the heading dates of grass varieties can converge. To take account of this effect it is possible to fit extra terms, one for each year, in the analysis of variance. Each term represents the linear regression of the observations for the year against the variety means over all years. The method is known as modified joint regression analysis (MJRA) and is recommended in situations where there is a statistically significant ($p \leq 1\%$) contribution from the regression terms in the analysis of variance. Statistical details, and a computer program to implement the procedure, are described in Part II Sections 2.1.8 and 2.1.9 *[cross ref.]*.

3.1.5.2 Small numbers of varieties in trials: Long-Term COYD

3.1.5.2.1 It is recommended that there should be at least 20 degrees of freedom for the varieties-by-years mean square in the COYD analysis of variance. This is in order to ensure that the varieties-by-years mean square is based on sufficient data to be a reliable estimate of the varieties-by-years variation for the LSD. Twenty degrees of freedom corresponds to 11 varieties common in three years of trials, or 21 varieties common in two years. Trials with fewer varieties in common over years are considered to have small numbers of varieties in trial.

3.1.5.2.2 In such trials the variety-by-year tables of means can be expanded to include means for earlier years, and if necessary, other established varieties. As not all varieties are present in all years, the resulting tables of variety-by-year means are not balanced. Consequently, each table is analysed by the least squares method of fitted constants (FITCON) or by REML, which produces an alternative varieties-by-years mean square as a long-term estimate of variety-by-years variation. This estimate has more degrees of freedom as it is based on more years and varieties.

$$\text{degrees of freedom} = \begin{pmatrix} \text{No. values in expanded} \\ \text{variety - by - year table} \end{pmatrix} - (\text{No. varieties}) - (\text{No. years}) + 1$$

3.1.5.2.3 The alternative varieties-by-years mean square is used in equation [1] above to calculate an LSD. This LSD is known as a “Long-Term LSD” to distinguish it from COYD LSD based on just the test years and varieties. The Long-Term LSD is used in the same way as the COYD LSD is used to assess the distinctness of varieties by comparing their over-year (the test years) means. The act of comparing the means of varieties using a “Long-Term LSD” is known as “Long-Term COYD”.

3.1.5.2.4 Long-Term COYD should only be applied to those characteristics lacking the recommended minimum degrees of freedom. However, when there is evidence that a characteristic’s LSD fluctuates markedly across years, it may be necessary to base the LSD for that characteristic on the current two or three-years of data, even though it has few degrees of freedom.

3.1.5.2.5 Figure 2 gives an example of the application of Long-Term COYD to the Italian ryegrass characteristic “Growth habit in spring”. A flow diagram of the stages and DUST modules used to produce Long-Term LSD’s and perform Long-Term COYD is given in Figure B2 in Part II: Section 2.1.9. *[cross ref.]*

3.1.5.3 Marked year-to-year changes in an individual variety's characteristic

Occasionally, a pair of varieties may be declared distinct on the basis of a t-test which is significant solely due to a very large difference between the varieties in a single year. To monitor such situations a check statistic is calculated, called F_3, which is the variety-by-years mean square for the particular variety pair expressed as a ratio of the overall variety-by-years mean square. This statistic should be compared with F-distribution tables with 1 and g, or 2 and g, degrees of freedom, for tests with two or three years of data respectively where g is the degrees of freedom for the variety-by-years mean square. If the calculated F_3 value exceeds the tabulated F value at the 1% level then an explanation for the unusual result should be sought before making a decision on distinctness.

3.1.6 Implementing COYD

The COYD method can be applied using the DUST package for the statistical analysis of DUS data, which is available from Dr. Sally Watson, Biometrics Division, Department of Agriculture for Northern Ireland (DANI), Newforge Lane, Belfast BT9 5PX, United Kingdom. Sample outputs are given in Part II: Section 2.1.9. *[cross ref.]*

3.1.7 References

DIGBY, P.G.N. (1979). Modified joint regression analysis for incomplete variety x environment data. J. Agric. Sci. Camb. 93, 81-86.

PATTERSON, H.D. & WEATHERUP, S.T.C. (1984). Statistical criteria for distinctness between varieties of herbage crops. J. Agric. Sci. Camb. 102, 59-68.

TALBOT, M. (1990). Statistical aspects of minimum distances between varieties. UPOV TWC Paper TWC/VIII/9, UPOV, Geneva.

Figure 1: Illustrating the calculation of the COYD criterion

Characteristic: Days to ear emergence in perennial ryegrass varieties

Varieties	Years 1	2	3	Over Year Means	*Difference (Varieties compared to C2)*	
Reference		Means				
R1	38	41	35	38	*35*	*D*
R2	63	68	61	64	*9*	*D*
R3	69	71	64	68	*5*	*D*
R4	71	75	67	71	*2*	
R5	69	78	69	72	*1*	
R6	74	77	71	74	*-1*	
R7	76	79	70	75	*-2*	
R8	75	80	73	76	*-3*	
R9	78	81	75	78	*-5*	*D*
R10	79	80	75	78	*-5*	*D*
R11	76	85	79	80	*-7*	*D*
Candidate						
C1	52	56	48	52	*21*	*D*
C2	72	79	68	73	*0*	-
C3	85	88	85	86	*-13*	*D*

ANALYSIS OF VARIANCE

Source	df	Mean square
Years	2	174.93
Variety	13	452.59
Variety-by-years	26	2.54

$$LSD_p = t_p * \sqrt{2} * SE(\overline{X})$$

$$LSD_{0.01} = 2.779 * 1.414 * \sqrt{(2.54/3)} = 3.6$$

Where t_p is taken from Student's t table with $p = 0.01$ (two-tailed) and 26 degrees of freedom.

To assess the distinctness of a candidate, the difference in the means between the candidate and all other varieties is computed. In practice a column of differences is calculated for each candidate. In this case, varieties with mean differences greater than, or equal to, 3.6 are regarded as distinct (marked *D* above).

Figure 2: Illustrating the application of Long-Term COYD

Characteristic: Growth habit in spring in italian ryegrass varieties

Varieties	Years 1	2	3*	4*	5*	Mean over test years	*Difference (Varieties compared to C2)*	
Reference			Means					
R1	43	42	41	44				
R2		39	45					
R3	43	38	41	45	40	42	*6*	*D*
R4	44	40	42	48	44	44.7	*3.3*	*D*
R5	46	43	48	49	45	47.3	*0.7*	
R6	51	48	52	53	51	52	*-4*	*D*
Candidate								
C1			43	45	44	44	*4*	*D*
C2			49	50	45	48	*0*	
C3			48	53	47	49.3	*-1.3*	

* indicates a test year

The aim is to assess the distinctness of the candidate varieties C1, C2 & C3 grown in the test years 3, 4 & 5.

The trial has a small number of varieties in trial because there are just seven varieties in common over the test years 3, 4 & 5 (data marked by a black border).

FITCON analysis of the variety-by-years table of means expanded to nine varieties in five years gives: varieties-by-years mean square = 1.924, on 22 degrees of freedom

$$\text{Long-term LSD}_p = t_p * \sqrt{2} * SE(\overline{X})$$

$$\text{Long-term LSD}_{0.01} = 2.819 * 1.414 * \sqrt{(1.924/3)} = 3.19$$

Where t_p is taken from Student's t table with $p = 0.01$ (two-tailed) and 22 degrees of freedom

To assess the distinctness of a candidate, the difference in the means between the candidate and all other varieties is computed. In practice a column of differences is calculated for each candidate. In the case of variety C2, varieties with mean differences greater than, or equal to 3.19 are regarded as distinct (marked *D* above).

3.1.8 COYD statistical methods

3.1.8.1 Analysis of variance

The standard errors used in the COYD criterion are based on an analysis of variance of the variety-by-years table of a characteristic's means. For m years and n varieties this analysis of variance breaks down the available degrees of freedom as follows:

Source	Df
Years	m-1
Varieties	n-1
Varieties-by-years	(m-1)(n-1)

3.1.8.2 Modified joint regression analysis (MJRA)

3.1.8.2.1 As noted above, the COYD criterion bases the SE of a variety mean on the varieties-by-years variation as estimated by the varieties-by-years mean square. Systematic variation can sometimes be identified as well as non-systematic variation. This systematic effect causes the occurrence of different slopes of the regression lines relating variety means in individual years to the average variety means over all years. Such an effect can be noted for the heading date characteristic in a year with a late spring: the range of heading dates can be compressed compared with the normal. This leads to a reduction in the slope of the regression line for variety means in that year relative to average variety means. Non-systematic variation is represented by the variation about these regression lines. Where only non-systematic varieties-by-years variation occurs, the slope of the regression lines have the constant value 1.0 in all years. However, when systematic variation is present, slopes differing from 1.0 occur but with an average of 1.0. When MJRA is used, the SE of a variety mean is based on the non-systematic part of the varieties-by-year variation.

3.1.8.2.2 The difference between the total varieties-by-years variation and the varieties-by-years variation adjusted by MJRA is illustrated in Figure B1, where variety means in each of three years are plotted against average variety means over all years. The variation about three parallel lines fitted to the data, one for each year, provides the total varieties-by-years variation as used in the COYD criterion described above. These regression lines have the common slope 1.0. This variation may be reduced by fitting separate regression lines to the data, one for each year. The resultant residual variation about the individual regression lines provides the MJRA-adjusted varieties-by-years mean square, on which the SE for a variety mean may be based. It can be seen that the MJRA adjustment is only effective where the slopes of the variety regression lines differ between years, such as can occur in heading dates.

3.1.8.2.3 The use of this technique in assessing distinctness has been included as an option in the computer program which applies the COYD criterion in the DUST package. It is recommended that it is only applied where the slopes of the variety regression lines are significantly different between years at the 1% significance level. This level can be specified in the computer program.

3.1.8.2.4 To calculate the adjusted variety means and regression line slopes the following model is assumed.

$$y_{ij} = u_j + b_j\, v_i + e_{ij}$$

where y_{ij} is the value for the i^{th} variety in the j^{th} year.

u_j is the mean of year j (j = 1, ..., m)

b_j is the regression slope for year j

v_i is the effect of variety i (i = 1, ..., n)

e_{ij} is an error term.

3.1.8.2.5 From equations (6) and (7) of Digby (1979), with the meaning of years and varieties reversed, the following equations relating these terms are derived for the situation where data are complete:

$$\sum_{i=j}^{n} v_i y_{ij} = b_j \sum_{i=1}^{n} v_i^2$$

$$\sum_{j=1}^{m} b_j y_{ij} = v_i \sum_{j=1}^{m} b_j^2$$

3.1.8.2.6 These equations are solved iteratively. All b_j values are taken to be 1.0 as a starting point in order to provide values for the v_i's. The MJRA residual sum of squares is then calculated as:

$$\sum_{j=1}^{m} \sum_{i-1}^{n} \left(y_{ij} - u_j - b_j v_i\right)^2$$

3.1.8.2.7 This sum of squares is used to calculate the MJRA-adjusted varieties-by-years mean square on $(m-1)(n-1) - m + 1$ degrees of freedom.

3.1.8.3 Comparison of COYD with other criteria

3.1.8.3.1 It can be shown that, for a three-year test, the COYD criterion applied at the 1% probability level is of approximately the same stringency as the 2x1% criterion for a characteristic where the square root of the ratio of the variety-by-years mean square to the variety-by-replicates-within-trials mean square (λ) has a value of 1.7. The COYD criterion applied at the 1% level is less stringent than the 2x1% criterion if $\lambda < 1.7$, and more stringent if $\lambda > 1.7$.

3.1.9 COYD software

3.1.9.1 An example of the output from the computer program in the DUST package which applies the COYD criterion is given in Tables B 1 to 3. It is taken from a perennial ryegrass (diploid) trial involving 40 reference varieties (R1 to R40) and 9 candidate varieties (C1 to C9) in 6 replicates on which 8 characteristics were measured over the years 1988, 1989 and 1990.

3.1.9.2 Each of the 8 characteristics is analysed by analysis of variance. As this analysis is of the variety-by-year-by-replicate data, the mean squares are 6 (= number of replicates) times the size of the mean squares of the analysis of variance of the variety-by-year data referred to in the main body of this paper. The results are given in Table B 1. Apart from the over-year variety means there are also presented:

YEAR MS:	the mean square term for years
VARIETY MS:	the mean square term for varieties
VAR.YEAR MS:	the mean square for varieties-by-years interaction
F1 RATIO:	ratio of VARIETY MS to VAR.YEAR MS (a measure of the discriminating power of the characteristic - large values indicate high discriminating power)
VAR.REP MS:	average of the variety-by-replicate mean squares from each year
LAMBDA VALUE (λ):	square root of the ratio of VAR.YEAR MS to VAR.REP MS
BETWEEN SE:	standard error of variety means over trials on a plot basis i.e. the square root of the VAR.YEAR MS divided by 18 (3 years x 6 replicates)
WITHIN SE:	the standard error of variety means within a trial on a plot basis i.e. the square root of the VAR.REP MS divided by 18
DF:	the degrees of freedom for varieties-by-years
MJRA SLOPE:	the slope of the regression of a single year's variety means on the means over the three years
REGR F VALUE:	the mean square due to MJRA regression as a ratio of the mean square about regression
REGR PROB:	the statistical significance of the REGR F VALUE
TEST:	indicates whether MJRA adjustment was applied (REG) or not (COY).

3.1.9.3 Each candidate variety is compared with every other candidate and reference variety. The mean differences between pairs of varieties are compared with the LSD for the characteristic. The results for the variety pair R1 and C1 are given in Table B 2. The individual within year t-values are listed to provide information on the separate years. Varieties R1 and C1 are considered distinct since, for at least one characteristic, a mean difference is COYD significant at the 1% level. If the F_3 ratio for characteristic 8 had been significant at the 1% level rather than the 5% level, the data for characteristic 8 would have been investigated, and because the differences in the three years are not all in the same direction, the COYD significance for characteristic 8 would not have counted towards distinctness.

3.1.9.4 The outcome in terms of the tests for distinctness of each candidate variety from all other varieties is given in Table B 3, where D indicates “distinct” and ND denotes “not distinct.”

Table B 1: An example of the output from the COYD program showing variety means and analysis of variance of characteristics

PRG (DIPLOID) EARLY N.I. UPOV 1988-90

VARIETY MEANS OVER YEARS

	5	60	8	10	11	14	15	24
	SP.HT	NSPHT	DEEE	H.EE	WEE	LFL	WFL	LEAR
1 R	45.27	34.60	67.87	45.20	70.05	20.39	6.85	24.54
2 R2	42.63	31.84	73.85	41.96	74.98	19.68	6.67	24.44
3 R3	41.57	27.40	38.47	27.14	57.60	17.12	6.85	22.57
4 R4	33.35	21.80	77.78	30.77	78.04	18.25	6.40	21.09
5 R5	37.81	25.86	50.14	27.24	62.64	16.41	6.41	16.97
6 R6	33.90	21.07	78.73	32.84	79.15	19.44	6.46	21.79
7 R7	41.30	31.37	73.19	41.35	71.87	20.98	6.92	24.31
8 R8	24.48	19.94	74.83	32.10	62.38	15.22	6.36	19.46
9 R9	46.68	36.69	63.99	44.84	68.62	18.11	7.02	22.58
10 R10	25.60	20.96	75.64	32.31	57.20	14.68	5.51	20.13
11 R11	41.70	30.31	74.60	40.17	76.15	19.45	6.79	22.72
12 R12	28.95	21.56	66.12	27.96	59.56	14.83	5.53	20.55
13 R13	40.67	29.47	70.63	36.81	74.12	19.97	7.04	24.05
14 R14	26.68	20.53	75.84	34.14	63.29	15.21	6.37	20.37
15 R15	26.78	20.18	75.54	30.39	66.41	16.34	6.01	20.94
16 R16	42.44	27.01	59.03	30.39	72.71	17.29	6.47	22.48
17 R17	27.94	21.58	76.13	32.53	68.37	16.72	6.11	22.03
18 R18	41.34	30.85	69.80	37.28	69.52	20.68	7.09	25.40
19 R19	33.54	23.43	73.65	30.35	75.54	18.97	6.37	22.43
20 R20	44.14	34.48	68.74	42.60	64.17	18.63	6.56	22.02
21 R21	27.77	21.53	80.52	31.59	69.41	16.81	5.81	22.35
22 R22	38.90	27.83	75.68	43.25	75.08	19.63	7.46	23.99
23 R23	42.43	31.80	72.40	42.07	74.77	20.99	6.78	23.57
24 R24	38.50	27.73	73.19	37.12	75.76	19.28	6.91	22.77
25 R25	43.84	29.60	68.82	39.79	74.83	20.63	7.08	22.65
26 R26	49.48	36.53	63.45	42.01	70.46	22.14	7.84	25.91
27 R27	25.61	19.25	78.78	29.81	56.81	15.81	5.07	18.94
28 R28	26.70	20.31	79.41	32.75	66.54	16.92	6.00	21.91
29 R29	27.90	20.94	72.66	29.85	67.14	16.85	6.28	21.79
30 R30	43.07	30.34	70.53	40.51	73.23	19.49	7.28	23.70
31 R31	38.18	25.47	74.23	36.88	80.23	20.40	7.09	25.21
32 R32	35.15	27.56	71.49	37.26	63.10	18.18	6.80	23.13
33 R33	42.71	31.09	67.58	39.14	70.36	19.85	7.12	23.35
34 R34	23.14	18.05	72.09	24.29	59.37	13.98	5.63	18.91
35 R35	32.75	25.41	77.22	38.90	67.07	17.16	6.42	21.49
36 R36	41.71	31.94	77.98	44.33	73.00	19.72	7.09	23.45
37 R37	44.06	32.99	74.38	45.77	71.59	20.88	7.40	24.06
38 R38	42.65	32.97	74.76	44.42	74.13	20.29	7.38	24.32
39 R39	28.79	22.41	76.83	35.91	64.52	16.85	6.34	22.24
40 R40	44.31	31.38	72.24	43.83	74.73	21.53	7.60	25.46
41 C1	42.42	31.68	64.03	40.22	67.02	20.73	6.90	26.16
42 C2	41.77	32.35	86.11	46.03	75.35	20.40	6.96	22.99
43 C3	41.94	31.09	82.04	43.17	74.04	19.06	6.26	23.44
44 C4	39.03	28.71	78.63	45.97	70.49	21.27	6.67	23.37
45 C5	43.97	30.95	72.99	39.14	77.89	19.88	6.68	25.44
46 C6	37.56	27.14	83.29	39.16	81.18	19.47	6.97	25.25
47 C7	38.41	28.58	83.90	42.53	76.44	19.28	6.00	23.47
48 C8	40.08	27.25	83.50	43.33	80.16	22.77	7.92	26.81
49 C9	46.77	34.87	51.89	37.68	61.16	19.25	6.92	24.82
YEAR MS	1279.09	3398.82	3026.80	2278.15	8449.20	672.15	3.36	51.32
VARIETY MS	909.21	476.72	1376.10	635.27	762.41	80.21	6.44	74.17
VAR.YEAR MS	23.16	18.86	14.12	23.16	46.58	4.76	0.28	2.73
F1 RATIO	39.26	25.27	97.43	27.43	16.37	16.84	22.83	27.16
VAR.REP MS	8.83	8.19	4.59	11.95	23.23	1.52	0.15	1.70
LAMBDA VALUE	1.62	1.52	1.75	1.39	1.42	1.77	1.37	1.27
BETWEEN SE	1.13	1.02	0.89	1.13	1.61	0.51	0.13	0.39
WITHIN SE	0.70	0.67	0.50	0.81	1.14	0.29	0.09	0.31
DF	96	94	96	96	96	96	96	96
MJRA SLOPE 88	0.90	0.86	0.99	0.91	0.99	1.09	0.97	0.95
MJRA SLOPE 89	1.05	1.08	1.01	0.99	1.06	0.97	1.02	0.98
MJRA SLOPE 90	1.05	1.06	1.00	1.10	0.95	0.94	1.01	1.07
REGR F VAL	4.66	6.17	0.06	4.48	0.76	1.62	0.29	1.91
REGR PROB	1.17	0.30	93.82	1.39	47.08	20.27	74.68	15.38
TEST	COY	REG	COY	COY	COY	COY	COY	COY

Table B 2: An example of the output from the COYD program showing a comparison of varieties R1 and C1

PRG (DIPLOID) EARLY N.I. UPOV 1988-90

41 C1 VERSUS 1 R1 *** USING REGR
WHERE SIG ***

(T VALUES + VE IF 41 C1 > 1 R1)

		SIG LEVELS YEARS 88	89	90		COYD T	PROB%	SIG	T VALUES YEARS 88	89	90	TSCORE	F3	
5	SP.HGHT	-	-	-1	ND	-1.78	7.88	NS	-1.05	-1.34	-2.64	-2.64	0.23	NS
60	NATSPHT	-	-1	-	ND	-2.02	4.61	*	-1.58	-2.61	-1.17	-2.61	0.22	NS
8	DATEEE	-1	-1	+	D	-3.06	0.29	**	-4.14	-6.33	0.80	-6.74	3.99	*
10	HGHT.EE	-1	-1	-5	D	-3.11	0.25	**	-2.79	-2.69	-2.06	-7.55	0.06	NS
11	WIDTHEE	-	-	-	ND	-1.33	18.58	NS	-1.47	-1.80	-0.21	0.00	0.32	NS
14	LGTHFL	+	+	-	ND	0.47	63.61	NS	0.17	1.83	-0.67	0.00	0.56	NS
15	WIDTHFL	+	-	+	ND	0.27	78.83	NS	0.31	-0.41	0.67	0.00	0.17	NS
24	EARLGTH	5	1	+	ND	2.93	0.42	**	2.10	3.33	1.01	5.43	0.84	NS

<u>Notes</u>

1. The three "COYD" columns headed, T PROB% and SIG give the COYD t value, its significance probability and significance level. The t value is the test statistic formed by dividing the mean difference between two varieties by the standard error of that difference. The t value can be tested for significance by comparing it with appropriate values from Students t-table. Calculating and testing a t value in this manner is equivalent to deriving an LSD and checking to see if the mean difference between the two varieties is greater than the LSD.

2. The two right-hand "F3" columns give the F_3 variance ratio statistic and its significance level. The F_3 statistic is defined in Part II, Section 2.1.5.2 *[cross ref.]*.

3. The sections in boxes refer to earlier distinctness criteria. The three "T VALUES, YEARS" columns headed 88, 89 and 90 are the individual within year t-test values (the Student's two-tailed t test of the variety means with standard errors estimated using the plot residual mean square), and the three "SIG LEVELS, YEARS" columns headed 88, 89 and 90 give their direction and significance levels. The column containing D and ND gives the distinctness status of the two varieties by the 2 x 1% criterion described in Section 5.2.4.18 of document TGP/8 [cross ref.]. The column headed T SCORE gives the obsolete T Score statistic and should be ignored.

Table B 3: An example of the output from the COYD program showing the distinctness status of the candidate varieties

PRG (DIPLOID) EARLY N.I. UPOV 1988-90

SUMMARY FOR COYD AT 1.0% LEVEL *** USING REGR ADJ WHEN SIG ***

CANDIDATE VARIETIES		C1	C2	C3	C4	C5	C6	C7	C8	C9
1	R1	D	D	D	D	D	D	D	D	D
2	R2	D	D	D	D	ND	D	D	D	D
3	R3	D	D	D	D	D	D	D	D	D
4	R4	D	D	D	D	D	D	D	D	D
5	R5	D	D	D	D	D	D	D	D	D
6	R6	D	D	D	D	D	D	D	D	D
7	R7	D	D	D	D	D	D	D	D	D
8	R8	D	D	D	D	D	D	D	D	D
9	R9	D	D	D	D	D	D	D	D	D
10	R10	D	D	D	D	D	D	D	D	D
11	R11	D	D	D	D	D	D	D	D	D
12	R1	D	D	D	D	D	D	D	D	D
13	R13	D	D	D	D	ND	D	D	D	D
14	R14	D	D	D	D	D	D	D	D	D
15	R15	D	D	D	D	D	D	D	D	D
16	R16	D	D	D	D	D	D	D	D	D
17	R17	D	D	D	D	D	D	D	D	D
18	R18	D	D	D	D	D	D	D	D	D
19	R19	D	D	D	D	D	D	D	D	D
20	R20	D	D	D	D	D	D	D	D	D
21	R21	D	D	D	D	D	D	D	D	D
22	R22	D	D	D	D	D	D	D	D	D
23	R23	D	D	D	D	D	D	D	D	D
24	R24	D	D	D	D	D	D	D	D	D
25	R25	D	D	D	D	D	D	D	D	D
26	R26	D	D	D	D	D	D	D	D	D
27	R27	D	D	D	D	D	D	D	D	D
28	R28	D	D	D	D	D	D	D	D	D
29	R29	D	D	D	D	D	D	D	D	D
30	R30	D	D	D	D	D	D	D	D	D
31	R31	D	D	D	D	D	D	D	D	D
32	R32	D	D	D	D	D	D	D	D	D
33	R33	D	D	D	D	D	D	D	D	D
34	R34	D	D	D	D	D	D	D	D	D
35	R35	D	D	D	D	D	D	D	D	D
36	R36	D	D	D	ND	D	D	D	D	D
37	R37	D	D	D	D	D	D	D	D	D
38	R38	D	D	D	D	D	D	D	D	D
39	R39	D	D	D	D	D	D	D	D	D
40	R40	D	D	D	D	D	D	D	D	D
41	C1	-	D	D	D	D	D	D	D	D
42	C2	D	-	D	D	D	D	D	D	D
43	C3	D	D	-	D	D	D	ND	D	D
44	C4	D	D	D	-	D	D	D	D	D
45	C5	D	D	D	D	-	D	D	D	D
46	C6	D	D	D	D	D	-	D	D	D
47	C7	D	D	ND	D	D	D	-	D	D
48	C8	D	D	D	D	D	D	D	-	D
49	C9	D	D	D	D	D	D	D	D	-
NO OF ND VARS		0	0	1	1	2	0	1	0	0
DISTINCTNESS		D	D	ND	ND	ND	D	ND	D	D
CANDIDATE VAR		C1	C2	C3	C4	C5	C6	C7	C8	C9

Figure B1. Heading date yearly variety means against over-year variety means

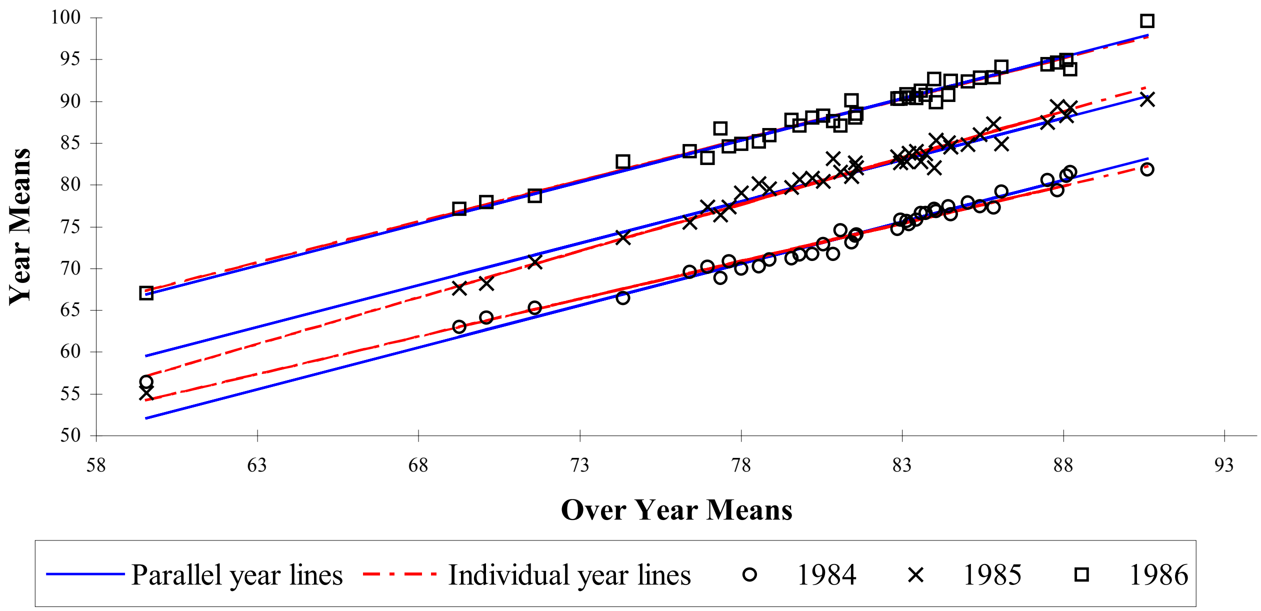
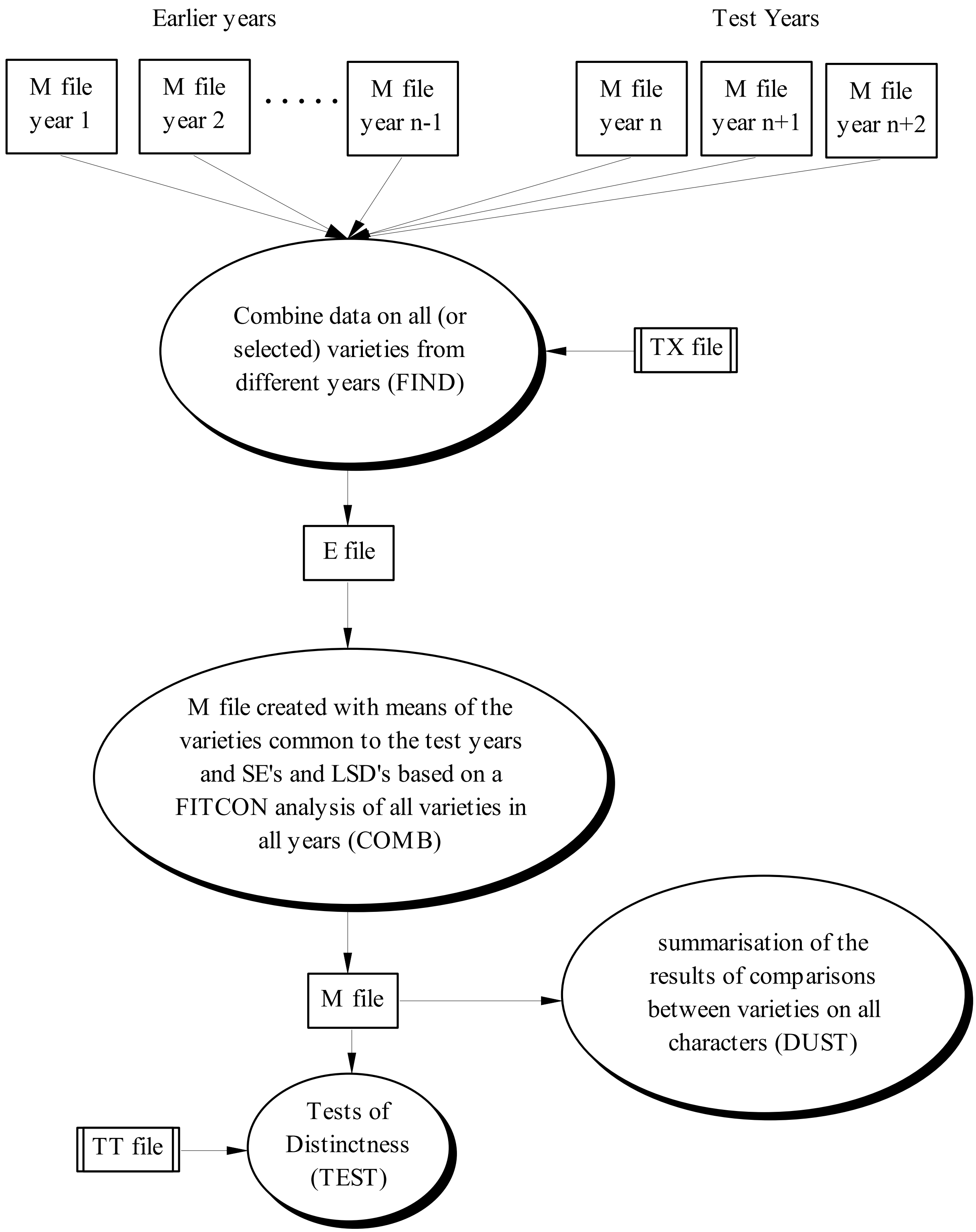


Figure B2. Flow Diagram of the stages and DUST modules used to produce long-term LSD's and perform long-term COYD



3.2 The Combined-Over-Years Uniformity Criterion (COYU)

3.2.1 Summary

3.2.1.1 TGP/10 explains that when the off-type approach for the assessment of uniformity is not appropriate for the assessment of uniformity, the standard deviation approach can be used. It further states the following with respect to determination of the acceptable level of variation.

> 5.2 Determining the acceptable level of variation
>
> 5.2.1.1 The comparison between a candidate variety and comparable varieties is carried out on the basis of standard deviations, calculated from individual plant observations. UPOV has proposed several statistical methods for dealing with uniformity in measured quantitative characteristics. One method, which takes into account variations between years, is the Combined Over Years Uniformity (COYU) method. The comparison between a candidate variety and comparable varieties is carried out on the basis of standard deviations, calculated from individual plant observations. This COYU procedure calculates a tolerance limit on the basis of comparable varieties already known i.e. uniformity is assessed using a relative tolerance limit based on varieties within the same trial with comparable expression of characteristics.

3.2.1.2 Uniformity is often related to the expression of a characteristic. For example, in some species, varieties with larger plants tend to be less uniform in size than those with smaller plants. If the same standard is applied to all varieties then it is possible that some may have to meet very strict criteria while others face standards that are easy to satisfy. COYU addresses this problem by adjusting for any relationship that exists between uniformity, as measured by the plant-to-plant SD, and the expression of the characteristic, as measured by the variety mean, before setting a standard.

[TWC Chairperson: Explanation of “reference” varieties to be developed]

3.2.1.3 The technique involves ranking reference and candidate varieties by the mean value of the characteristic. Each variety’s SD is taken and the mean SD of the most similar varieties is subtracted. This procedure gives, for each variety, a measure of its uniformity expressed relative to that of comparable varieties. The term reference varieties here refers to the non-candidate varieties selected for inclusion in the growing trial.

3.2.1.4 The results for each year are combined in a variety-by-years table of adjusted SDs and analysis of variance is applied. The mean adjusted SD for the candidate is compared with the mean for the reference varieties using a standard t-test.

3.2.1.5 COYU, in effect, compares the uniformity of a candidate with that of the reference varieties most similar in relation to the characteristic being assessed. The main advantages of COYU are that all varieties can be compared on the same basis and that information from several years of testing may be combined into a single criterion.

3.2.2 Introduction

3.2.2.1 Uniformity is sometimes assessed by measuring individual characteristics and calculating the standard deviation (SD) of the measurements on individual plants within a plot. The SDs are averaged over all replicates to provide a single measure of uniformity for each variety in a trial.

3.2.2.2 This section outlines a procedure known as the combined-over-years uniformity (COYU) criterion. COYU assesses the uniformity of a variety relative to reference varieties based on SDs from trials over several years. A feature of the method is that it takes account of possible relationships between the expression of a characteristic and uniformity.

3.2.2.3 This section describes:

- The principles underlying the COYU method.
- UPOV recommendations on the application of COYU to individual species.
- Mathematical details of the method with an example of its application.
- The computer software that is available to apply the procedure.

3.2.3 The COYU Criterion

3.2.3.1 The application of the COYU criterion involves a number of steps as listed below. These are applied to each characteristic in turn. Details are given under Part II: Section 2.2.4 *[cross ref.]* below.

- Calculation of within-plot SDs for each variety in each year.
- Transformation of SDs by adding 1 and converting to natural logarithms.
- Estimation of the relationship between the SD and mean in each year. The method used is based on moving averages of the log SDs of reference varieties ordered by their means.
- Adjustments of log SDs of candidate and reference varieties based on the estimated relationships between SD and mean in each year.
- Averaging of adjusted log SDs over years.
- Calculation of the maximum allowable SD (the uniformity criterion). This uses an estimate of the variability in the uniformity of reference varieties derived from analysis of variance of the variety-by-year table of adjusted log SDs.
- Comparison of the adjusted log SDs of candidate varieties with the maximum allowable SD.

3.2.3.2 The advantages of the COYU criterion are:

- It provides a method for assessing uniformity that is largely independent of the varieties that are under test.
- The method combines information from several trials to form a single criterion for uniformity.

- Decisions based on the method are likely to be stable over time.
- The statistical model on which it is based reflects the main sources of variation that influence uniformity.
- Standards are based on the uniformity of reference varieties.

3.2.4 Recommendations on COYU

3.2.4.1 COYU is recommended for use in assessing the uniformity of varieties

- For quantitative characteristics.
- When observations are made on a plant basis over two or more years.
- When there are some differences between plants of a variety, representing quantitative variation rather than presence of off-types.

3.2.4.2 A variety is considered to be uniform for a characteristic if its mean adjusted log SD does not exceed the uniformity criterion.

3.2.4.3 The probability level "p" used to determine the uniformity criterion depends on the crop. Recommended probability levels are given in [.....] *[cross ref.]*

3.2.4.4 The uniformity test may be made over two or three years. If the test is normally applied over three years, it is possible to choose to make an early acceptance or rejection of a variety using an appropriate selection of probability values.

3.2.4.5 It is recommended that there should be at least 20 degrees of freedom for the estimate of variance for the reference varieties formed in the COYU analysis. This corresponds to 11 reference varieties for a COYU test based on two years of trials and 8 reference varieties for three years. In some situations, there may not be enough reference varieties to give the recommended minimum degrees of freedom. Advice is being developed for such cases.

3.2.5 Mathematical details

Step 1: Derivation of the within-plot standard deviation

3.2.5.1 Within-plot standard deviations for each variety in each year are calculated by averaging the plot between-plant standard deviations, SD_j, over replicates:

$$SD_j = \sqrt{\frac{\sum_{i=1}^{n}\left(y_{ij} - \mathbf{y}_j\right)^2}{(n-1)}}$$

$$SD = \frac{\sum_{j=1}^{r} SD_j}{r}$$

where y_{ij} is the observation on the i^{th} plant in the j^{th} plot, $\mathbf{y}_j$ is the mean of the observations from the j^{th} plot, n is the number of plants measured in each plot and r is the number of replicates.

Step 2: Transformation of the SDs

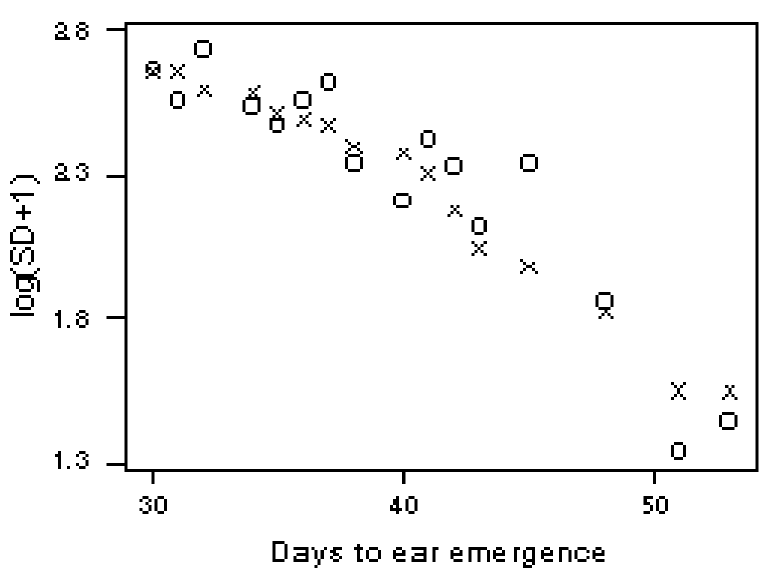
3.2.5.2 Transformation of SDs by adding 1 and converting to natural logarithms. The purpose of this transformation is to make the SDs more amenable to statistical analysis.

Step 3: Estimation of the relationship between the SD and mean in each year

3.2.5.3 For each year separately, the form of the average relationship between SD and characteristic mean is estimated for the reference varieties. The method of estimation is a 9-point moving average. The log SDs (the Y variate) and the means (the X variate) for each variety are first ranked according to the values of the mean. For each point ($X_{i,}$ Y_i) take the trend value T_i to be the mean of the values Y_{i-4}, Y_{i-3}, ... , Y_{i+4} where i represents the rank of the X value and Y_i is the corresponding Y value. For X values ranked 1^{st} and 2^{nd} the trend value is taken to be the mean of the first three values. In the case of the X value ranked 3^{rd} the mean of the first five values are taken and for the X value ranked 4^{th} the mean of the first seven values are used. A similar procedure operates for the four highest-ranked X values.

3.2.5.4 A simple example in Figure 1 illustrates this procedure for 16 varieties. The points marked “0” in Figure 1a represent the log SDs and the corresponding means of 16 varieties. The points marked “X” are the 9-point moving-averages, which are calculated by taking, for each variety, the average of the log SDs of the variety and the four varieties on either side. At the extremities the moving average is based on the mean of 3, 5, or 7 values.

Figure 1: Association between SD and mean – days to ear emergence in cocksfoot varieties (*symbol O is for observed SD, symbol X is for moving average SD*)



Step 4: Adjustment of transformed SD values based on estimated SD-mean relationship

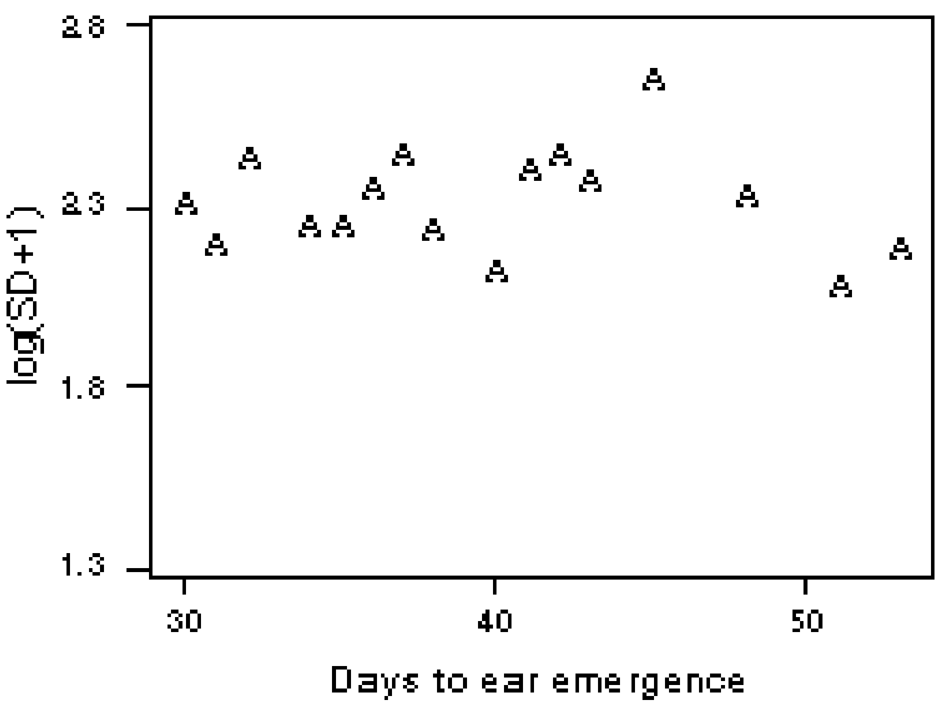
3.2.5.5 Once the trend values for the reference varieties have been determined, the trend values for candidates are estimated using linear interpolation between the trend values of the nearest two reference varieties as defined by their means for the characteristic. Thus if the trend values for the two reference varieties on either side of the candidate are T_i and T_{i+1} and the observed value for the candidate is X_c, where $X_i \lesseqqgtr X_c \lesseqqgtr X_{i+1}$, then the trend value T_c for the candidate is given by

$$T_c = \frac{(X_C - X_i)T_{i+1} + (X_{i+1} - X_C)T_i}{X_{i+1} - X_i}$$

3.2.5.6 To adjust the SDs for their relationship with the characteristic mean the estimated trend values are subtracted from the transformed SDs and the grand mean is added back.

3.2.5.7 The results for the simple example with 16 varieties are illustrated in Figure 2.

Figure 2: Adjusting for association between SD and mean – days to ear emergence in cocksfoot varieties (*symbol A is for adjusted SD*)



Step 5: Calculation of the uniformity criterion

3.2.5.8 An estimate of the variability in the uniformity of the reference varieties is derived by applying a one-way analysis of variance to the adjusted log SDs, i.e. with years as the classifying factor. The variability (V) is estimated from the residual term in this analysis of variance.

3.2.5.9 The maximum allowable standard deviation (the uniformity criterion), based on k years of trials, is

$$UC_p = SD_r + t_p \sqrt{V\left(\frac{1}{k} + \frac{1}{Rk}\right)}$$

where SD_r is the mean of adjusted log SDs for the reference varieties, V is the variance of the adjusted log SDs after removing year effects, t_p is the one-tailed t-value for probability p with degrees of freedom as for V, k is the number of years and R is the number of reference varieties.

3.2.6 Early decisions for a three-year test

3.2.6.1 Decisions on uniformity may be made after two or three years depending on the crop. If COYU is normally applied over three years, it is possible to make an early acceptance or rejection of a candidate variety using an appropriate selection of probability values.

3.2.6.2 The probability level for early rejection of a candidate variety after two years should be the same as that for the full three-year test. For example, if the three-year COYU test is applied using a probability level of 0.2%, a candidate variety can be rejected after two years if its uniformity exceeds the COYU criterion with probability level 0.2%.

3.2.6.3 The probability level for early acceptance of a candidate variety after two years should be larger than that for the full three-year test. As an example, if the three-year COYU test is applied using a probability level of 0.2%, a candidate variety can be accepted after two years if its uniformity does not exceed the COYU criterion with probability level 2%.

3.2.6.4 Some varieties may fail to be rejected or accepted after two years. In the example set out in paragraphs 26 and 27, a variety might have a uniformity that exceeds the COYU criterion with probability level 2% but not the criterion with probability level 0.2%. In this case, such varieties should be re-assessed after three years.

3.2.7 Example of COYU calculations

3.2.7.1 An example of the application of COYU is given here to illustrate the calculations involved. The example consists of days to ear emergence scores for perennial ryegrass over three years for 11 reference varieties (R1 to R11) and one candidate (C1). The data is tabulated in Table 1.

Table 1: Example data-set – days to ear emergence in perennial ryegrass

Variety	Character Means			Within Plot SD			Log (SD+1)		
	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3
R1	38	41	35	8.5	8.8	9.4	2.25	2.28	2.34
R2	63	68	61	8.1	7.6	6.7	2.21	2.15	2.04
R3	69	71	64	9.9	7.6	5.9	2.39	2.15	1.93
R4	71	75	67	10.2	6.6	6.5	2.42	2.03	2.01
R5	69	78	69	11.2	7.5	5.9	2.50	2.14	1.93
R6	74	77	71	9.8	5.4	7.4	2.38	1.86	2.13
R7	76	79	70	10.7	7.6	4.8	2.46	2.15	1.76
R8	75	80	73	10.9	4.1	5.7	2.48	1.63	1.90
R9	78	81	75	11.6	7.4	9.1	2.53	2.13	2.31
R10	79	80	75	9.4	7.6	8.5	2.34	2.15	2.25
R11	76	85	79	9.2	4.8	7.4	2.32	1.76	2.13
C1	52	56	48	8.2	8.4	8.1	2.22	2.24	2.21

3.2.7.2 The calculations for adjusting the SDs in year 1 are given in Table 2. The trend value for candidate C1 is obtained by interpolation between values for varieties R1 and R2, since the characteristic mean for C1 (i.e. 52) lies between the means for R1 and R2 (i.e. 38 and 63). That is

$$T_c = \frac{(X_C - X_i)T_{i+1} + (X_{i+1} - X_C)T_i}{X_{i+1} - X_i} = \frac{(52-38)x2.28 + (63-52)x2.28}{63-38} = 2.28$$

Table 2: Example data-set – calculating adjusted log(SD+1) for year 1

Variety	Ranked mean (X)	Log (SD+1) (Y)	Trend Value T	Adj. Log (SD+1)
R1	38	2.25	(2.25 + 2.21 + 2.39)/3 = 2.28	2.25 - 2.28 + 2.39 = 2.36
R2	63	2.21	(2.25 + 2.21 + 2.39)/3 = 2.28	2.21 - 2.28 + 2.39 = 2.32
R3	69	2.39	(2.25 + . . . + 2.42)/5 = 2.35	2.39 - 2.35 + 2.39 = 2.42
R5	69	2.50	(2.25 + . . . + 2.48)/7 = 2.38	2.50 - 2.38 + 2.39 = 2.52
R4	71	2.42	(2.25 + . . . + 2.32)/9 = 2.38	2.42 - 2.38 + 2.39 = 2.43
R6	74	2.38	(2.21 + . . . + 2.53)/9 = 2.41	2.38 - 2.41 + 2.39 = 2.36
R8	75	2.48	(2.39 + . . . + 2.34)/9 = 2.42	2.48 - 2.42 + 2.39 = 2.44
R7	76	2.46	(2.42 + . . . + 2.34)/7 = 2.42	2.46 - 2.42 + 2.39 = 2.43
R11	76	2.32	(2.48 + . . . + 2.34)/5 = 2.43	2.32 - 2.43 + 2.39 = 2.28
R9	78	2.53	(2.32 + 2.53 + 2.34)/3 = 2.40	2.53 - 2.40 + 2.39 = 2.52
R10	79	2.34	(2.32 + 2.53 + 2.34)/3 = 2.40	2.34 - 2.40 + 2.39 = 2.33
Mean	70	2.39		
C1	52	2.22	2.28	2.22 – 2.28 + 2.39 = 2.32

3.2.7.3 The results of adjusting for all three years are shown in Table 3.

Table 3: Example data-set – adjusted log(SD+1) for all three years with over-year means

Variety	Over-Year Means Char. mean	 Adj. Log (SD+1)	Adj. Log (SD+1) Year 1	 Year 2	 Year 3
R1	38	2.26	2.36	2.13	2.30
R2	64	2.10	2.32	2.00	2.00
R3	68	2.16	2.42	2.10	1.95
R4	71	2.15	2.43	1.96	2.06
R5	72	2.20	2.52	2.14	1.96
R6	74	2.12	2.36	1.84	2.16
R7	75	2.14	2.43	2.19	1.80
R8	76	2.02	2.44	1.70	1.91
R9	78	2.30	2.52	2.16	2.24
R10	78	2.22	2.33	2.23	2.09
R11	80	2.01	2.28	1.78	1.96
Mean	70	2.15	2.40	2.02	2.04
C1	52	2.19	2.32	2.08	2.17

3.2.7.4 The analysis of variance table for the adjusted log SDs is given in Table 4 (based on reference varieties only). The variability in the uniformity of reference varieties is estimated from this (V=0.0202).

Table 4: Example data set – analysis of variance table for adjusted log (SD+1)

Source	Degrees of freedom	Sums of squares	Mean squares
Year	2	1.0196	0.5098
Varieties within years (=residual)	30	0.6060	**0.0202**
Total	32	1.6256	

3.2.7.5 The uniformity criterion for a probability level of 0.2% is calculated thus:

$$UC_p = SD_r + t_p\sqrt{V\left(\frac{1}{k}+\frac{1}{Rk}\right)} = 2.15 + 3.118x\sqrt{0.0202x\left(\frac{1}{3}+\frac{1}{3x11}\right)} = 2.42$$

where t_p is taken from Student's t table with p=0.002 (one-tailed) and 30 degrees of freedom.

3.2.7.6 Varieties with mean adjusted log (SD + 1) less than, or equal to, 2.42 can be regarded as uniform for this characteristic. The candidate variety C1 satisfies this criterion.

3.2.8 Implementing COYU

The COYU criterion can be applied using the DUST software package for the statistical analysis of DUS data. This is available from the Dr. Sally Watson, Biometrics Division, Department of Agriculture for Northern Ireland, Newforge Lane, Belfast BT9 5PX, UK .

3.2.9 COYU Software

3.2.9.1 DUST Computer program

3.2.9.1.1 The main output from the DUST COYU program is illustrated in Table A1. This summarises the results of analyses of within-plot SDs for 49 perennial ryegrass varieties assessed over a three-year period. Supplementary output is given in Table A2 where details of the analysis of a single characteristic, date of ear emergence, are presented. Note that the analysis of variance table given has an additional source of variation; the variance, V, of the adjusted log SDs is calculated by combining the variation for the variety and residual sources.

3.2.9.1.2 In Table A1, the adjusted SD for each variety is expressed as a percent of the mean SD for all reference varieties. A figure of 100 indicates a variety of average uniformity; a variety with a value less than 100 shows good uniformity; a variety with a value much greater than 100 suggests poor uniformity in that characteristic. Lack of uniformity in one characteristic is often supported by evidence of poor uniformity in related characteristics.

3.2.9.1.3 The symbols "*" and "+" to the right of percentages identify varieties whose SDs exceed the COYU criterion after 3 and 2 years respectively. The symbol ":" indicates that after two years uniformity is not yet acceptable and the variety should be considered for testing for a further year. Note that for this example a probability level of 0.2% is used for the three-year test. For early decisions at two years, probability levels of 2% and 0.2% are used to accept and reject varieties respectively. All of the candidates had acceptable uniformity for the 8 characters using the COYU criterion.

3.2.9.1.4 The numbers to the right of percentages refer to the number of years that a within-year uniformity criterion is exceeded. This criterion has now been superseded by COYU.

3.2.9.1.5 The program will operate with a complete set of data or will accept some missing values, e.g. when a variety is not present in a year.

Table A1: Example of summary output from COYU program

**** OVER-YEARS UNIFORMITY ANALYSIS SUMMARY ****

WITHIN-PLOT STANDARD DEVIATIONS AS % MEAN OF REFERENCE VARIETY SDS

CHARACTERISTIC NUMBER

	5	60	8	10	11	
R1	100	100	95 1	100	97	97
R2	105	106	98	99	104	101
R3	97	103	92 1	103	96	98
R4	102	99	118 2	105	101	101
R5	102	99	116 3	95	104	110
R6	103	102	101	99	97	104
R7	100	95	118 2	102 1	98	99
R8	97	98	84	95	97	93
R9	97	105	87	99	101	99
R10	104	100	96	105 1	96	102
R11	99	96	112	99	101	98
R12	100	97	99 1	103	105	106
R13	95	96	101	100	96	101
R14	105	103	90	97	101	97
R15	102	100 1	89	105	105 1	101
R16	99	98	92 1	98	102	98
R17	97	101	98	101	101	95
R18	99	97	96	96	102	99
R19	103	101	105	102	100	98
R20	104	99	93	91	100	102
R21	97	94	103	97	100	102
R22	101	110*1	112	107 1	103 1	101
R23	94	101	107	99	104	97
R24	99	97	95	99	100	103
R25	104 1	103	93 1	99	101	96
R26	98	97	111 2	96	102 1	106
R27	102	99	106 1	99	103	107
R28	101	106	90	95	101	101
R29	101	105	83	102	94	93
R30	99	96	97	99	95	100
R31	99	102	107	107 1	102	99
R32	98	93	111 2	102	98	103
R33	104	102 1	107 1	103	100	97
R34	95	94	82	95	97	96
R35	100	102	95	100	99	94
R36	99	98	111 1	99	100	103
R37	100	107 1	107	101	100	107
R38	95	97	102	107 1	97	101
R39	99	99	90	98	101	100
R40	104	102	112 1	100	101	97
C1	100 1	106	113 2	104 1	106 1	106
C2	103	101	98	97	101	109
C3	97	93	118 2	98	99	109
C4	102	101	106	103	99	101
C5	100	104	99	103	100	107
C6	101	102	103	100	103	107
C7	96	98	106	97	102	103
C8	101	105 1	116 2	103	103	93
C9	99	99	90 2	91	97	98

CHARACTERISTIC

5	SPRING	60	NATURAL SPRIN
8	DATE OF EAR	10	HEIGHT AT EAR
11	WIDTH AT EAR	14	LENGTH OF FLA
15	WIDTH OF FLAG	24	EAR LENGTH

SYMBOLS

* - SD EXCEEDS OVER-YEARS CRITERION AFTEI
\+ - SD EXCEEDS OVER-YEARS CRITERION AFTEI
: - SD NOT YET ACCEPTABLE AFTER 2 YEARS W
1,2,3 - THE NUMBER OF OCCASIONS THE WITHIN-YE

**** UNIFORMITY ANALYSIS OF BETWEEN-PLANT STANDARD DEVIATIONS (SD) ****

VARIETY	OVER-YEARS CHAR. MEAN	OVER-YEARS ADJ. LOG SD	OVER-YEARS UNADJ LOG SD	CHAR. MEAN 88	CHAR. MEAN 89	CHAR. MEAN 90	LOG (SD+1) 88	LOG (SD+1) 89	LOG (SD+1) 90	ADJ LOG(SD+1) 88	ADJ LOG(SD+1) 89	ADJ LOG(SD+1) 90
REFERENCE												
R3	38.47	1.823	2.179	39.07	41.21	35.12	2.02	2.18	2.34X	1.73	1.78	1.96
R5	50.14	2.315	2.671	48.19	53.69	48.54	2.52X	2.74X	2.76X	2.23	2.33	2.39
R16	59.03	1.833	2.179	57.25	63.33	56.50	2.28X	2.24	2.01	1.96	1.73	1.81
R26	63.44	2.206	2.460	61.00	66.53	62.81	2.50X	2.75X	2.13	2.18	2.33	2.11
R9	63.99	1.739	1.994	62.92	68.32	60.72	2.21	2.03	1.74	1.96	1.64	1.62
R12	66.12	1.964	2.086	67.89	65.35	65.12	2.07	2.58X	1.60	1.97	2.14	1.78
R33	67.58	2.124	2.254	66.66	71.54	64.53	2.55X	2.26	1.95	2.32	1.92	2.12
R1	67.87	1.880	1.989	69.07	70.64	63.90	1.60	2.45X	1.93	1.60	2.08	1.96
R20	68.74	1.853	1.893	67.17	74.31	64.74	2.05	1.95	1.68	1.92	1.75	1.89
R25	68.82	1.853	1.905	68.28	72.38	65.81	1.83	2.39X	1.49	1.75	2.09	1.72
R18	69.80	1.899	1.853	68.61	75.22	65.58	1.88	1.84	1.84	1.82	1.80	2.08
R30	70.53	1.919	1.864	70.36	75.08	66.15	2.04	1.84	1.71	2.00	1.78	1.98
R13	70.63	2.005	2.000	70.23	75.00	66.66	1.97	2.03	2.01	1.91	1.86	2.24
R32	71.49	2.197	2.238	70.03	74.98	69.44	2.32X	2.45X	1.94	2.31	2.27	2.01
R34	72.09	1.630	1.545	71.32	77.35	67.59	1.57	1.49	1.58	1.54	1.58	1.78
R40	72.24	2.222	2.178	72.71	75.07	68.95	2.25X	2.26	2.03	2.29	2.16	2.22
R23	72.40	2.122	2.058	69.72	78.39	69.10	2.11	2.14	1.93	2.16	2.14	2.06
R29	72.66	1.657	1.580	73.13	75.80	69.04	1.46	1.63	1.65	1.47	1.69	1.81
R7	73.19	2.341	2.342	72.23	75.80	71.52	2.62X	2.30X	2.10	2.61	2.30	2.11
R24	73.19	1.888	1.796	74.00	76.37	69.20	1.62	1.84	1.93	1.71	1.91	2.04
R19	73.65	2.083	2.049	73.32	76.06	71.57	1.96	2.05	2.14	1.96	2.13	2.16
R2	73.85	1.946	1.897	72.98	78.16	70.42	1.76	1.96	1.97	1.79	2.02	2.03
R31	74.23	2.119	2.012	73.73	78.23	70.71	2.05	1.86	2.13	2.25	1.94	2.17
R37	74.38	2.132	2.020	74.87	76.95	71.32	1.97	2.04	2.04	2.23	2.11	2.06
R11	74.60	2.224	2.150	73.87	78.07	71.87	2.21	2.08	2.16	2.36	2.10	2.21
R38	74.76	2.029	1.916	76.11	78.24	69.93	1.84	2.15	1.75	1.98	2.24	1.87
R8	74.83	1.677	1.593	74.27	78.77	71.45	1.62	1.55	1.61	1.75	1.64	1.64
R15	75.54	1.760	1.682	75.72	78.68	72.22	1.53	1.79	1.73	1.64	1.84	1.80
R10	75.64	1.915	1.847	73.47	79.24	74.23	1.87	1.66	2.00	1.99	1.78	1.98
R22	75.68	2.228	2.133	74.57	79.17	73.32	2.18	2.21	2.01	2.40	2.26	2.03
R14	75.84	1.797	1.688	74.53	79.56	73.43	1.54	1.63	1.90	1.70	1.76	1.93
R17	76.13	1.942	1.832	75.34	79.09	73.96	1.65	2.04	1.81	1.90	2.10	1.83
R39	76.83	1.781	1.676	75.49	80.50	74.50	1.56	1.51	1.96	1.72	1.70	1.92
R35	77.22	1.886	1.773	76.67	80.85	74.15	1.73	1.67	1.92	1.88	1.85	1.93
R4	77.78	2.349	2.268	76.80	81.22	75.33	2.36X	2.13	2.31X	2.52	2.33	2.20
R36	77.98	2.209	2.173	78.97	79.85	75.11	2.13	2.15	2.25X	2.24	2.21	2.18
R6	78.73	2.009	1.935	77.53	82.88	75.78	2.00	1.75	2.06	2.03	2.09	1.91
R27	78.78	2.116	2.098	77.61	80.03	78.69	1.80	2.25	2.24X	1.87	2.39	2.09
R28	79.41	1.785	1.722	78.28	81.99	77.97	1.68	1.43	2.05	1.79	1.67	1.89
R21	80.52	2.045	1.950	77.43	85.02	79.11	1.98	1.75	2.13	2.07	2.09	1.98
CANDIDATE												
C1	64.03	2.252	2.438	63.85	63.33	64.92	2.49X	2.81X	2.02	2.25	2.29	2.21
C2	86.11	1.940	1.837	84.83	88.63	84.85	1.79	1.71	2.01	1.90	2.05	1.87
C3	82.04	2.349	2.248	82.26	87.45	76.40	2.37X	2.03	2.35X	2.48	2.37	2.20
C4	78.63	2.104	2.033	78.01	82.17	75.72	2.05	2.01	2.04	2.15	2.27	1.90
C5	72.99	1.973	1.869	71.98	79.40	67.59	1.95	1.78	1.88	1.93	1.90	2.08
C6	83.29	2.050	1.947	84.10	85.57	80.21	2.05	1.69	2.10	2.16	2.03	1.96
C7	83.90	2.100	1.997	84.12	87.99	79.60	1.93	1.95	2.11	2.04	2.29	1.97
C8	83.50	2.304	2.201	82.43	85.98	82.08	2.27X	2.00	2.34X	2.38	2.33	2.20
C9	51.89	1.788	2.157	52.35	55.77	47.56	1.83	2.34X	2.31X	1.52	1.91	1.93
MEAN OF REFERENCE	71.47	1.988		70.78	74.97	68.65	1.97	2.03	1.96	1.99	1.99	1.99

UNIFORMITY CRITERION

		PROB. LEVEL
3-YEAR REJECTION	2.383	0.002
2-YEAR REJECTION	2.471	0.002
2-YEAR ACCEPTANCE	2.329	0.020

**** ANALYSIS OF VARIANCE OF ADJUSTED LOG(SD+1) *** *

	DF	MS	F RATIO
YEARS	2	0.06239	
VARIETIES	39	0.11440	5.1
RESIDUAL	78	0.02226	
TOTAL	119	0.05313	

SYMBOLS

* - SD EXCEEDS OVER-YEARS UNIFORMITY CRITERION AFTER 3 YEARS.
+ - SD EXCEEDS OVER-YEARS UNIFORMITY CRITERION AFTER 2 YEARS.
: - SD NOT YET ACCEPTABLE ON OVER-YEARS CRITERION AFTER 2 YEARS.
X - SD EXCEEDS 1.265 TIMES MEAN OF REFERENCE VARIETIES

3.3 Standard probability levels used for COYD and COYU

The following four cases are those which, in general, represent the different situations which may arise where COYD and COYU are used in DUS testing:

Scheme A. Test is conducted over 2 independent growing cycles and decisions made after 2 growing cycles (a growing cycle could be a year and is further on denoted by cycle)

Scheme B. Test is conducted over 3 independent growing cycles and decisions made after 3 cycles

Scheme C. Test is conducted over 3 independent growing cycles and decisions made after 3 cycles, but a variety may be accepted after 2 cycles

Scheme D. Test is conducted over 3 independent growing cycles and decisions made after 3 cycles, but a variety may be accepted or rejected after 2 cycles

The stages at which the decisions are made in Cases A to D are illustrated in figures 1 to 4 respectively. These also illustrate the various standard probability levels (p_{d2}, p_{nd2}, p_{d3}, p_{u2}, p_{nu2} and p_{u3}) which are needed to calculate the COYD and COYU criteria depending on the case. These are defined as follows:

Probability Level	**Used to decide whether a variety is :-**
p_{d2}	distinct after 2 cycles
p_{nd2}	non-distinct in a characteristic after 2 cycles
p_{d3}	distinct after 3 cycles
p_{u2}	uniform in a characteristic after 2 cycles
p_{nu2}	non-uniform after 2 cycles
p_{u3}	uniform in a characteristic after 3 cycles

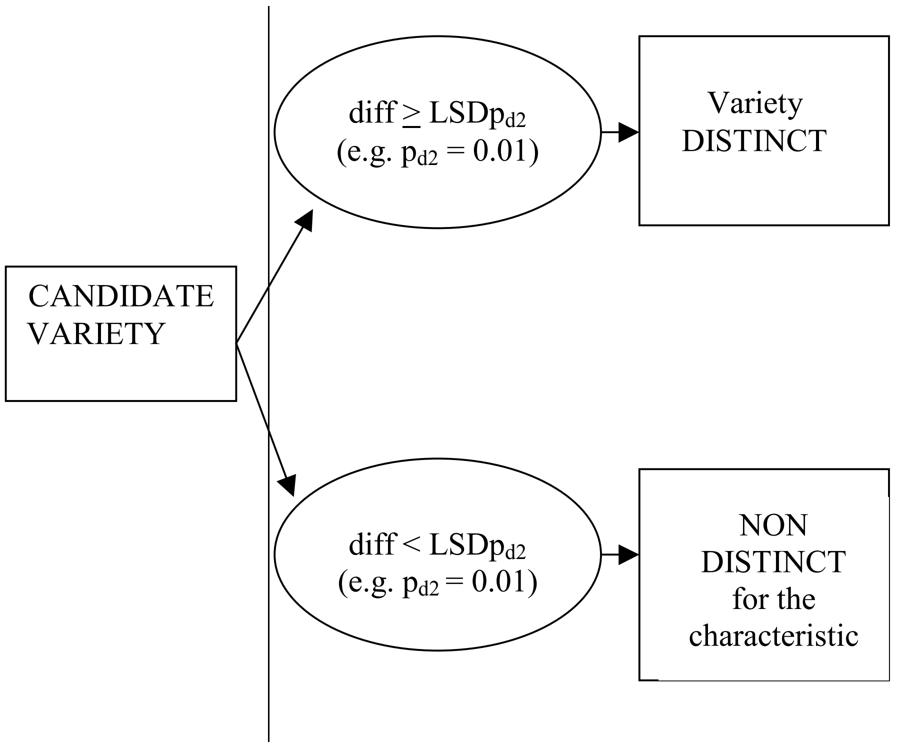
In figures 1 to 4 the COYD criterion calculated using say the probability level p_{d2} is denoted by LSDp_{d2} etc., and the COYU criterion calculated using say the probability level p_{u2} is denoted by UCp_{u2} etc. The term “diff” represents the difference between the means of a candidate variety and another variety for a characteristic, while “U” represents the mean adjusted log(SD+1) of a variety for a characteristic.

Table 1 summarises the various standard probability levels needed to calculate the COYD and COYU criteria in each of Cases A to D. For example, in Case B only two probability levels are needed (p_{d3} and p_{u3}), whereas Case C requires four (p_{d2}, p_{d3}, p_{u2} and p_{u3}).

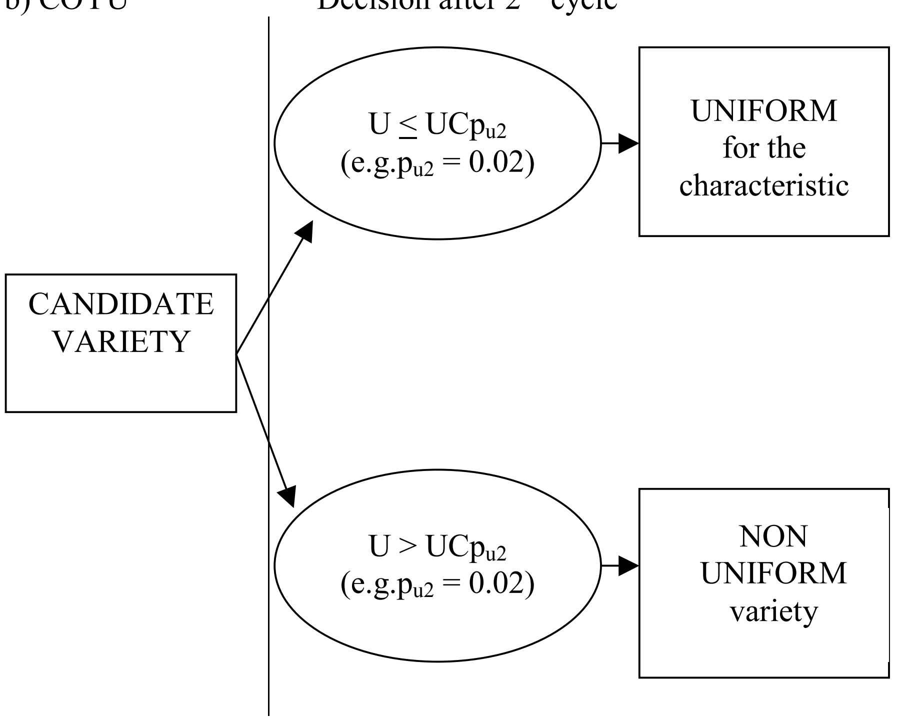
Table 1	COYD			COYU		
CASE	p_{d2}	p_{nd2}	p_{d3}	p_{u2}	p_{nu2}	p_{u3}
A						
B						
C						
D						

Figure 1. COYD and COYU decisions and standard probability levels (p_i) in Case A

a) COYD Decision after 2nd cycle



b) COYU Decision after 2nd cycle



NOTE:-

“diff” is the difference between the means of the candidate variety and another variety for the characteristic.

LSDp is the COYD criterion calculated at probability level p.

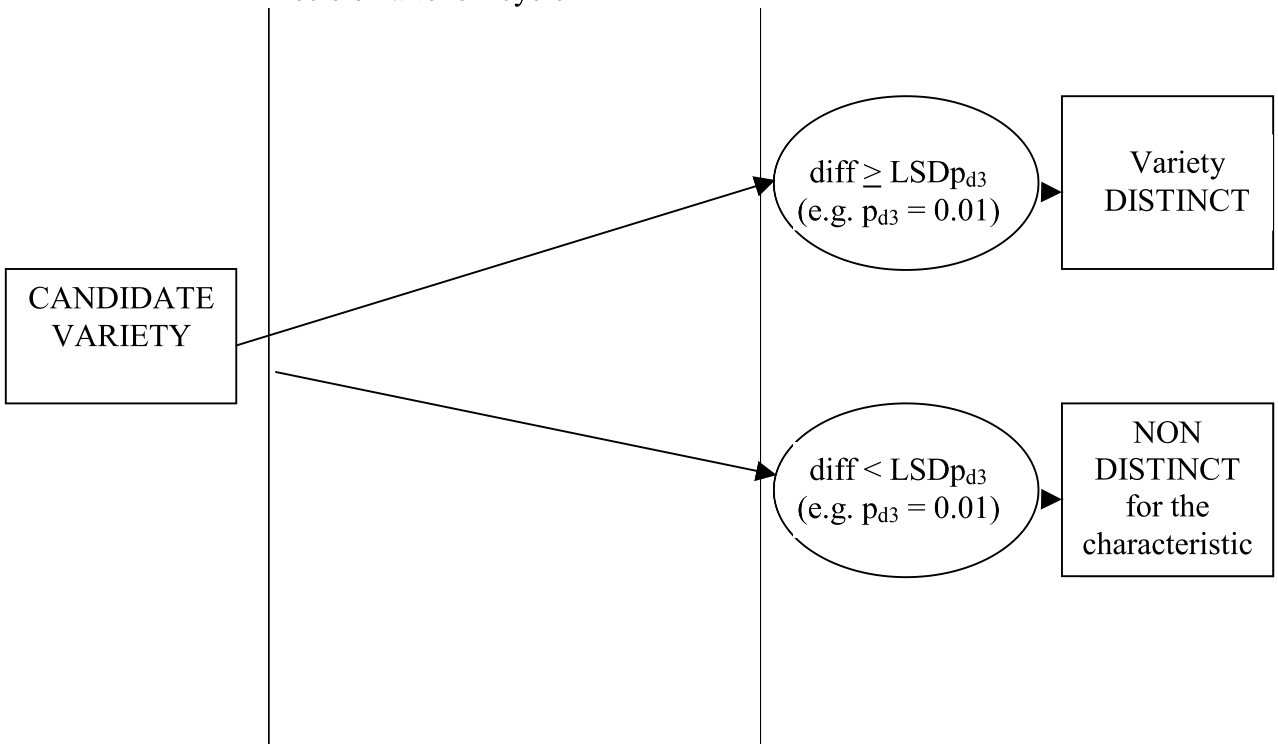
“U” is the mean adjusted log(SD+1) of the candidate variety for the characteristic.

UCp is the COYU criterion calculated at probability level p.

Figure 2. COYD and COYU decisions and standard probability levels (p_i) in Case B

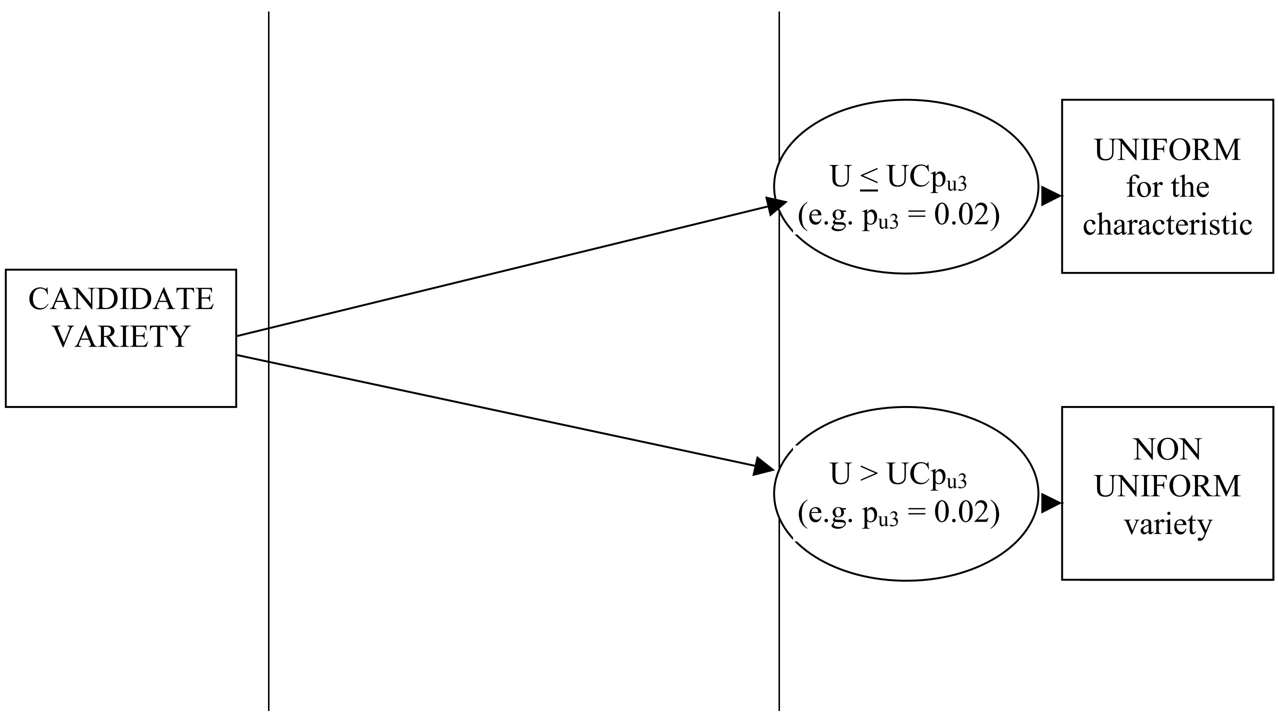
a) COYD

Decision after 3rd cycle



b) COYU

Decision after 3rd cycle

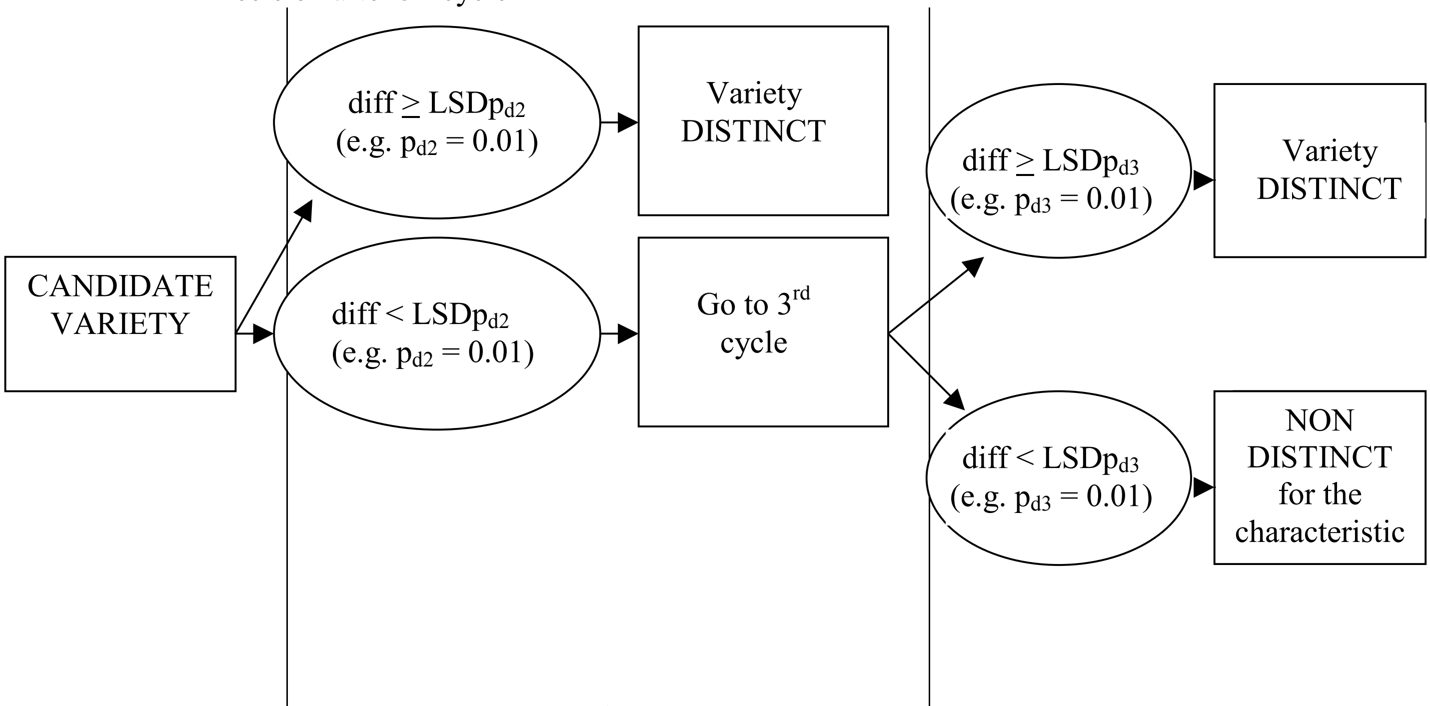


NOTE:-

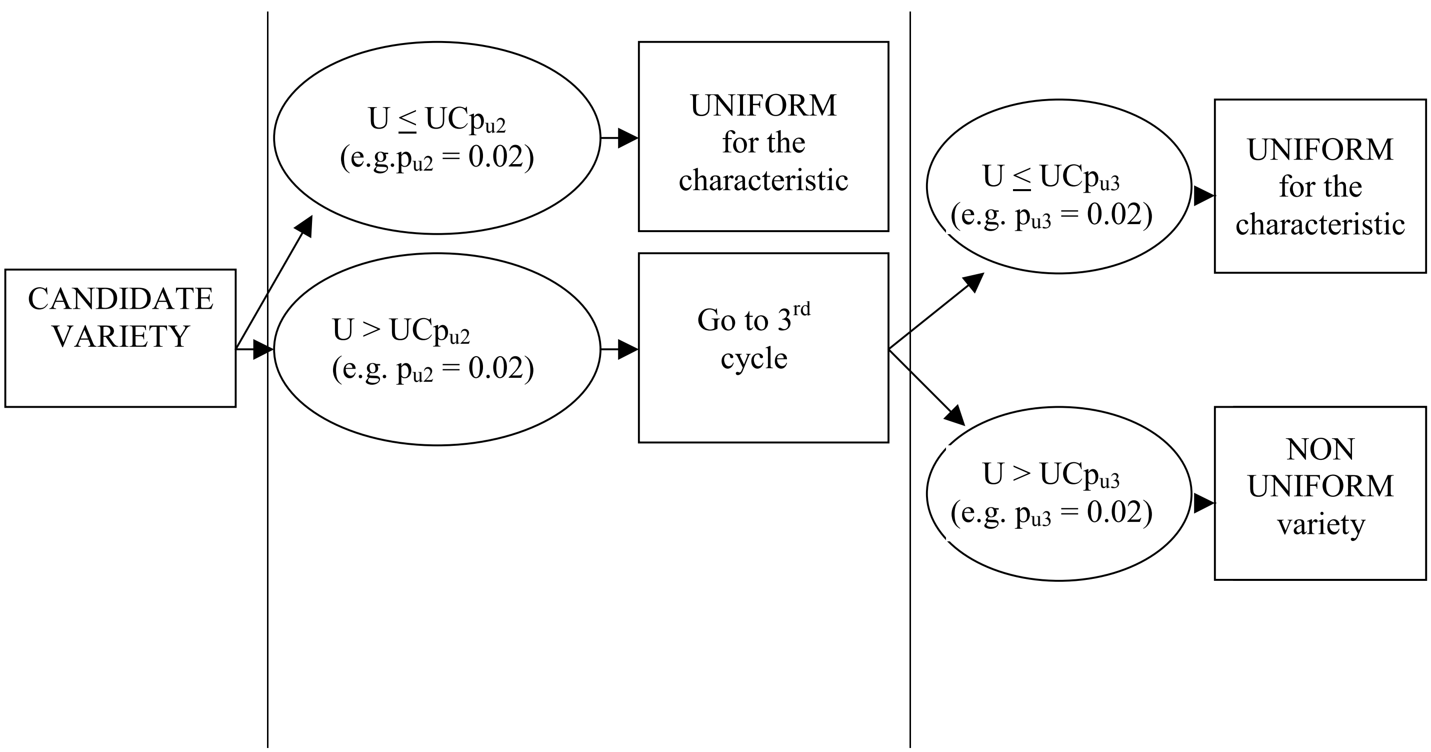
"diff"	is the difference between the means of the candidate variety and another variety for the characteristic.
LSDp	is the COYD criterion calculated at probability level p.
"U"	is the mean adjusted log(SD+1) of the candidate variety for the characteristic.
UCp	is the COYU criterion calculated at probability level p.

Figure 3. COYD and COYU decisions and standard probability levels (p_i) in Case C

a) COYD Decision after 2nd cycle
Decision after 3rd cycle



b) COYU Decision after 2nd cycle
Decision after 3rd cycle

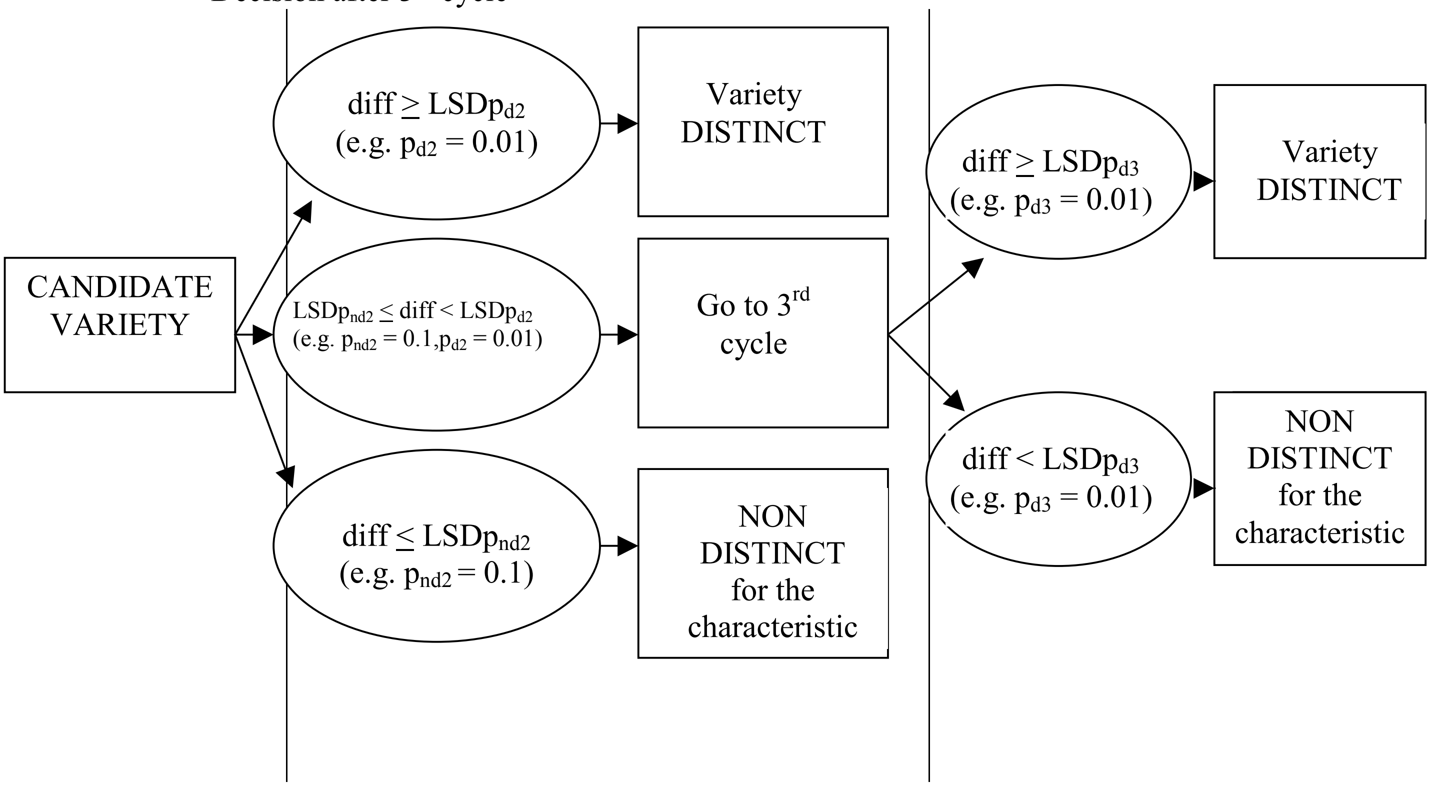


NOTE:-

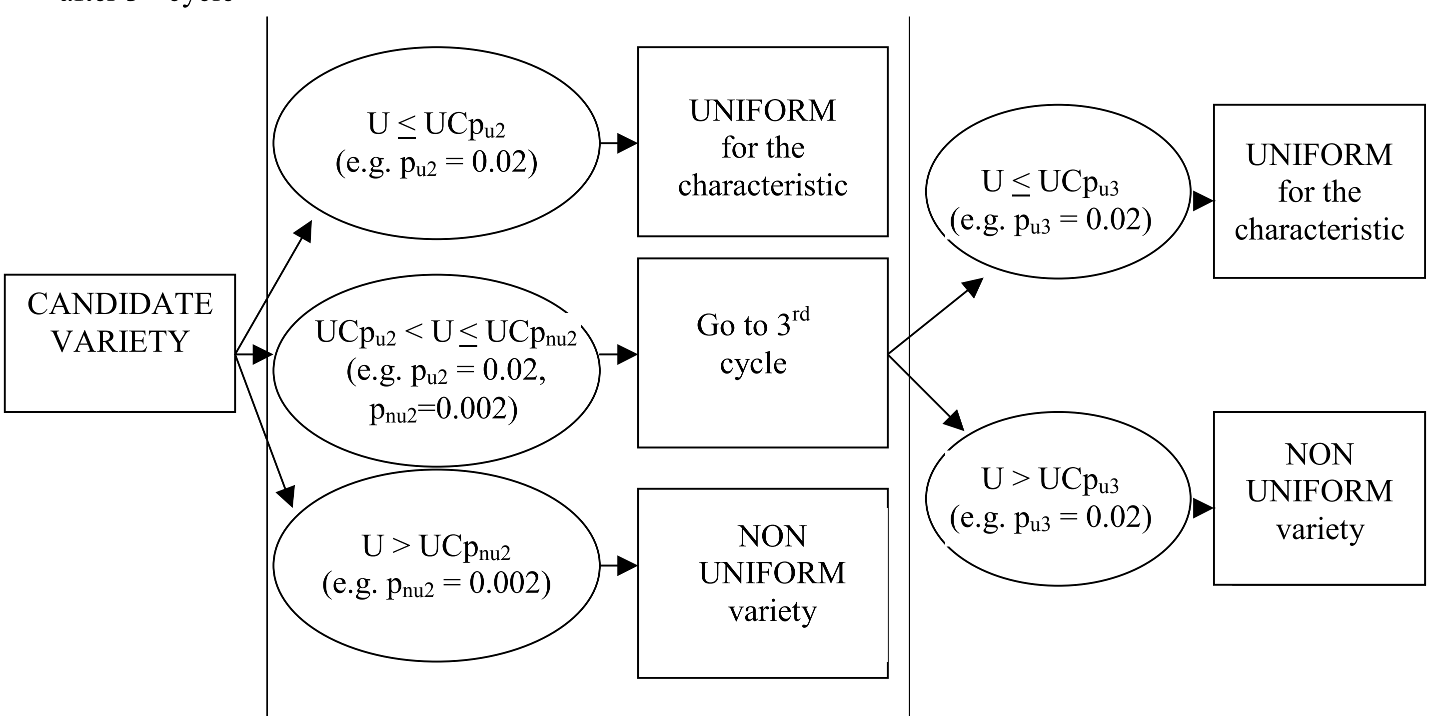
"diff"	is the difference between the means of the candidate variety and another variety for the characteristic.
LSDp	is the COYD criterion calculated at probability level p.
"U"	is the mean adjusted log(SD+1) of the candidate variety for the characteristic.
UCp	is the COYU criterion calculated at probability level p.

Figure 4. COYD and COYU decisions and standard probability levels (p_i) in Case D

a) COYD Decision after 2nd cycle
Decision after 3rd cycle



b) COYU Decision after 2nd cycle Decision
after 3rd cycle



NOTE:-
"diff" is the difference between the means of the candidate variety and another variety for the characteristic.
LSDp is the COYD criterion calculated at probability level p.
"U" is the mean adjusted log(SD+1) of the candidate variety for the characteristic
UCp is the COYU criterion calculted at probability level p.

4. PARENT FORMULA OF HYBRID VARIETIES

4.1 Introduction

The use of the parental formula requires that the difference between parent lines is sufficient to ensure that the hybrid obtained from those parents is distinct. The method is based on the following steps:

(i) description of parent lines according to the Test Guidelines;

(ii) checking the originality of those parent lines in comparison with the variety collection, based on the table of characteristics in the Test Guidelines, in order to identify similar parent lines;

(iii) checking the originality of the hybrid formula in relation to the hybrids in the variety collection, taking into account the most similar parent lines; and

(iv) assessment of distinctness at the hybrid level for varieties with a similar formula.

4.2 Requirements of the method

The application of the method requires:

(i) a declaration of the formula and submission of plant material of the parent lines of hybrid varieties;

(ii) inclusion in the variety collection of the parent lines used as parents in the hybrid varieties of the variety collection (for guidance on the constitution of a variety collection see document TGP/4 section 1) and a list of the formulae of the hybrid varieties;

(iii) application of the method to all varieties in the variety collection. This condition is important to obtain the full benefit; and

(iv) a rigorous approach to assess the originality of any new parent line in order to be confident on the distinctness of the hybrid variety based on that parent line.

4.3 Assessing the originality of a new parent line

4.3.1 The originality of a parental line is assessed using the characteristics included in the relevant Test Guidelines.

4.3.2 The difference between parent lines must be sufficient to be sure that hybrids produced using different parent lines will be distinct. For example:

Characteristic 1: a characteristic having two states of expression (absent/present), which are determined by two alleles of a single gene, with one dominant allele (+) for the expression "present" and one recessive allele (-) for the expression "absent".

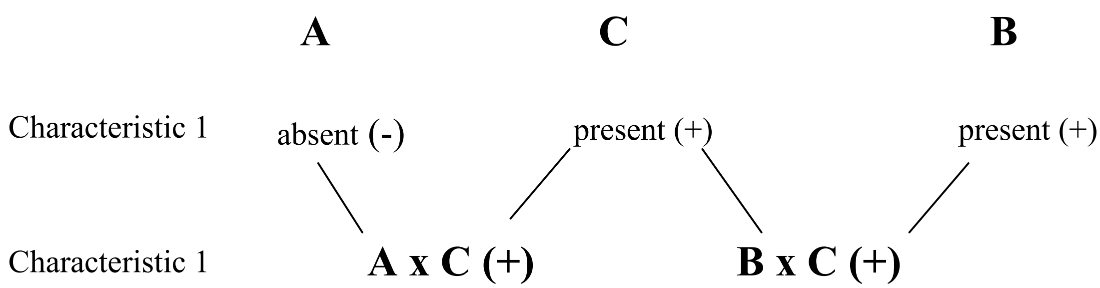
Three parent lines:

A: with the recessive allele (-) with expression "absent"
B: with the dominant allele (+) with expression "present"
C: with the dominant allele (+) with expression "present"

Crossing the above-mentioned parent lines to obtain the following F1 hybrids:

(A x C): having expression "present" for Characteristic 1
(B x C): having expression "present" for Characteristic 1

The following diagram shows the ways the two different crossings result in the same expression of Characteristic 1 (i.e. "present" in both hybrids), although parent line A(-) and parent line B(+) have different expressions.



4.3.3 Although the parent lines A and B are clearly different for characteristic 1, the two hybrid varieties A x C and B x C have the same expression. Thus, a difference between A and B for Characteristic 1 is not sufficient.

4.3.4 With a more complex genetic control involving several genes, not precisely described, the interaction between the different alleles of each gene and between genes might also lead to similar expression at the level of the hybrid varieties. In such cases, a larger difference is appropriate to establish distinctness between two parent lines.

4.3.5 Determining the difference required is mainly based on a good knowledge of the species, of the characteristics and, when available, on their genetic control.

4.4 Verification of the formula

4.4.1 The aim of verifying the formula is to check if the candidate hybrid variety has been produced by crossing the parent lines declared and submitted by the applicant.

4.4.2 Different characteristics can be used to perform this check when the genetic pattern of each parent can be identified in the hybrid. Generally, characteristics based on polymorphism of enzymes or of some storage proteins can be used.

4.4.3 If no suitable characteristics are available, the only possibility is to cross the parent lines using the plant material submitted by the applicant and to compare the hybrid variety seedlots (the sample submitted by the applicant and the sample harvested after the cross).

4.5 Uniformity and stability of parent lines

4.5.1 The uniformity and stability of the parent lines should be assessed according to the appropriate recommendations for the variety concerned. The uniformity and stability of the parent lines are important for the stability of the hybrid. Another requirement for the stability of the hybrid is the use of the same formula for each cycle of the hybrid seed production.

4.5.2 A check of the uniformity on the hybrid should also be done, even if distinctness of the hybrid has been established on the basis of the parent lines.

4.6 Description of the hybrid

4.6.1 A description of the hybrid variety should be established, even where the distinctness of the hybrid has been established on the basis of the parent formula.

5. THE GAIA METHODOLOGY

GAIA method has been developed to optimize trials, by avoiding to unnecessarily grow some reference varieties. The principle is to compute a phenotypic distance between each pair of varieties, this distance being a sum of distances on each individual observed characteristic.
The originality of the method relies on the possibility given to the crop expert to express his confidence on the differences observed, by giving weights to the difference for each observed characteristic.

5.1 Some reasons to sum and weight observed differences

5.1.1 When assessing distinctness, a DUS examiner first observes a variety characteristic-by-characteristic. In the case of similar varieties, the DUS examiner also considers all observed differences as a whole. The GAIA software helps the DUS examiner to assess differences characteristic-by-characteristic and for all characteristics together.

5.1.2 A DUS examiner may see that two varieties are so distinct after the first growing cycle that it is not necessary to repeat the comparison. Those two varieties, which are “ distinct plus” (see Section Section 2.6.2.1.2 *[cross ref.]*), are obviously distinct.

5.1.3 A DUS examiner may have a situation where two varieties receive a different note (e.g. Variety A is Note 3 for a given characteristic and Variety B is Note 4), but the two varieties are considered by the examiner to be similar. The difference could be due to the fact that the varieties were not grown very close each other (i.e. had different environmental conditions), or to variability of the observer when assessing the notes, etc.

5.1.4 Characteristics vary in their susceptibility to environmental conditions and the precision with which they are observed (i.e. visual observation/measurement). For characteristics which are susceptible to environmental conditions and which are not assessed very precisely, the examiner requires a large difference between Variety A and Variety B to be confident that the observed difference indicates distinctness.

5.1.5 For characteristics which are independent of environmental conditions and which are assessed precisely, the examiner can be confident in a smaller difference between Variety A and Variety B.

5.1.6 In the GAIA method, the examiner decides the appropriate weights for the observed differences for each observed characteristic. The software computes the sum of the weightings and indicates to the crop examiner which pairs of varieties are “distinct plus” and which are not. The examiner can then decide which of the varieties of common knowledge can be excluded from the subsequent growing cycle(s), because they are already obviously distinct from all candidate varieties.

5.2 Computing GAIA phenotypic distance

5.2.1 The principle of the GAIA method is to compute a phenotypic distance between two varieties, being the total distance between a pair of varieties resulting from the addition of the weightings of all characteristics. Thus, the GAIA phenotypic distance is:

$$dist(i,j) = \sum_{k=1,nchar} W_k(i,j)$$

where:

$dist(i,j)$ is the computed distance between variety i and variety j.

k is the k^{th} characteristic, from the *nchar* characteristics selected for computation.

$W_k(i,j)$ is the weighting of characteristics k, which is a function of the difference observed between variety *i* and variety *j* for that characteristic k.

$W_k(i,j) = f(|OV_{ki} - OV_{kj}|)$

where OV_{ki} is the observed value on characteristic *k* for variety *i*.

5.2.2 Detailed information on e is provided in section 5.2.

5.2 Detailed informatin on the GAIA methodology

5.2.1. Weighting of characteristics

5.2.1.1 Weighting is defined as the contribution in a given characteristic to the total distance between a pair of varieties. For each species, this system must be calibrated to determine the weight which can be given to each difference and to evaluate the reliability of each characteristic in a given environment and for the genetic variability concerned. For that reason the role of the crop expert is essential.

5.2.1.2 Weighting depends on the size of the difference and on the individual characteristic. The weightings are defined by the crop expert on the basis of its expertise in the crop and on a "try-and-check" (see Diagram 3 at the end of this annex) learning process. The expert can give zero weighting to small differences, thus, even if two varieties have different observed values in many characteristics, the overall distance might be zero. For a given difference, the same weighting is attributed to any pair of varieties for a given characteristic.

5.2.1.3 The weighting should be simple and consistent. For instance the crop expert can base the weights for a characteristic only with integer values, i.e. 0, 1, 2, 3, (or more).

If so,

- a weight of 0 is given to observed differences which for this characteristic are considered by the crop expert as possibly caused by environment effects or lack of precision in measure.
- a weight of 1 is the minimum weight which can contribute as a non zero distance
- a weight of 3 is considered to be about 3 times greater in term of confidence or distance than a weight of 1.

5.2.1.4 The distinctness plus threshold will be defined as a value for which the sum of the differences with a non zero weight is great enough to ensure a reliable obvious distinction.

5.2.1.5 Diagram 3 is a flowchart which describes how an iterative “try and learn” process can be used to obtain step by step a satisfactory set of weights for a given crop.

5.2.1.6 The following simple example on Zea mays shows the computation of the distance between two varieties:

Example: taking the characteristic “Shape of ear”, observed on a 1 to 3 scale, the crop expert has attributed weighting to differences which they consider significant:

Shape of ear:
1 = conical
2 = conico-cylindrical
3 = cylindrical

Comparison between difference in notes and weighting		
	Different in notes	Weighting
conical (1) vs. conical (1)	0	0
conical (1) vs. conico-cylindrical (2)	1	2
conical (1) vs. cylindrical (3)	2	6
conico-cylindrical (2) vs. conico-cylindrical (2)	0	0
conico-cylindrical (2) vs. cylindrical (3)	1	2
cylindrical (3) vs. cylindrical (3)	0	0

When the crop expert compares a variety 'i' with conical ear (note 1) to a variety 'j' with cylindrical ear (note 3), he attributes a weighting of 6 etc. The weightings are summarized in the form of a weighting matrix:

Weighting matrix 'i'		Variety i		
		1	**2**	**3**
Variety 'j'	**1**	0	2	6
	2		0	2
	3			0

When the crop expert compare a variety i with conical ear (note 1) to a variety j with cylindrical ear (note 3), he attributes a weighting of 6.

5.2.2. Examples of use

5.2.2.1 Determining "Distinctness Plus"

5.2.2..1 The threshold for the phenotypic distance used to eliminate varieties from the growing trial is called "Distinctness Plus" and is settled by the crop expert at a level which is higher than the difference needed to establish distinctness. This ensures that all pairs of varieties having a distance equal or greater than the threshold (Distinctness Plus) would be distinct if they were grown in another trial.

5.2.2.2 The Distinctness Plus threshold must be based on experience gained with the varieties of common knowledge and must minimize the risk of excluding in a next growing trial a pair of varieties which should need to be further compared in the field.

5.2.2.2 Other examples of use

Using phenotypic distance in the first growing cycle

5.2.2.2.1 A crop that has a large variety collection and uses only characteristics on a 1 to 9 scale; GAIA methodology allows the selection of varieties to be included in the growing trial. This can be used to plan the first growing cycle trials as well as the subsequent growing cycles.

5.2.2.2.2 In crops with relatively few candidates and a small variety collection, which enables the crop expert to sow all candidates (e.g. an agricultural crop), and the appropriate reference varieties, in two or three successive growing cycles. The same varieties are sown in growing cycles 1, 2 and 3, in a randomized layout. The software will help to identify the pairs with a small distance, to enable the expert to focus his attention on these particular cases when visiting the field.

Using phenotypic distance after the first growing trial

5.2.2.2.3 After one growing cycle (e.g. in the examination of an ornamental crop), the absolute data and distance computations are an objective way to secure the decision of the expert, because the quality of the observation and reliability of differences observed have been taken into account in the weighting system. If more growing cycles are necessary before a decision is taken, the software helps to identify on which cases the expert will need to focus.

5.2.2.2.4 In cases where there are many candidate and reference varieties and there is a wide variability in the species (e.g. a vegetable crop such as Capsicum); on the one hand there are already obvious differences after only one cycle, but on the other hand some varieties are very similar. In order to be more efficient in their checks, the crop expert wishes to grow "similar" varieties close to each other. The raw results and distances will help to select the "similar" varieties and decide on the layout of the trial for the next growing cycle.

5.2.2.2.5 In crops in which there are many similar varieties, for which it is a common practice to make side-by-side comparisons, GAIA can be used to identify the similar varieties after the first cycle, in particular, when the number of varieties in a trial increases, making it less easy to identify all the problem situations. The software can help to "not miss" the less obvious cases.

5.2.2.2.6 In vegetatively propagated ornamental varieties, the examination lasts for one or two growing cycles: after the first growing cycle, some reference varieties in the trial are obviously different from all candidates, and their inclusion in the second growing cycle is not necessary. When the number of varieties is large, the raw data and distance(s) can help the expert to detect reference varieties for which the second growing cycle is unnecessary.

<u>5.2.3. Computing GAIA phenotypic distance</u>

The principle is to compute a phenotypic distance between two varieties, which is the sum of weightings given by the crop expert to the differences he observed.

GAIA phenotypic distance is:

$$dist(i,j) = \sum_{k=1,nchar} W_k(i,j)$$

where:

$dist(i,j)$ is the computed distance between variety i and variety j.

k is the k^{th} characteristic, from the *nchar* characteristics selected for computation.

$W_k(i,j)$ is the weighting of characteristics k, which is a function of the difference observed between variety *i* and variety *j* for that characteristic k.

$$W_k(i,j) = f(|OV_{ki} - OV_{kj}|)$$

where OV_{ki} is the observed value on characteristic *k* for variety *i*.

This phenotypic distance computations allows to:

- compare two varieties,
- compare a given variety to all other varieties,
- compare all candidate varieties to all [candidate + reference] observed varieties
- compare all possible pair combinations.

5.2.4. GAIA software

5.2.4.1 GAIA software allows the computation of the phenotypic distance using UPOV characteristics of the crop guideline, which can be used alone or in combination. The user can decide on the type of data and the way it is used. He can select all the available characteristics, or different subsets of characteristics.

5.2.4.2 The main use of GAIA is to define a “distinct plus” threshold which corresponds to a reliable and obvious distinction.

5.2.4.3 Remember that all differences with a zero weight do not contribute at all to the distance. Two varieties can have different notes in a number of observed characteristics, and end with a zero distance.

5.2.4.4 Non zero weights are summed in the distance. If the distance is smaller than the distinct plus threshold, even if there are a number of clear differences in notes or measures, the varieties will not be suggested as reliably and obviously distinct.
If the distance is greater than the distinct plus threshold set by the crop expert, this shall correspond to a case where a pair comparison in a further growing trail is un-necessary.

5.2.4.5 GAIA enables the crop expert to use the threshold parameter in two other ways for practical means other than distinctness plus:

- a low threshold helps to find the more difficult cases (to identify similar varieties or close varieties) where expert will have to focus its attention in next cycle
- a very big threshold allows to see all available raw data and the weightings for each characteristic on screens and printouts

5.2.4.6 In practice different thresholds can be used according to the different needs, they can easily be selected before to run a comparison. Different comparisons can be computed, stored and recalled from the database with their appropriate threshold, set of characteristics, set of varieties....

5.2.4.7 The software provides a comprehensive report for each pair-wise comparison and a classification of all pair wise comparisons, from the more distinct to the more similar. Software computes an overall distance, but also provides all the individual absolute values and the distance contribution of each characteristic.

5.2.4.8 In order to minimize computation time, as soon as the threshold is achieved for a comparison between two given varieties, the software proceeds to the next pair of varieties. Remaining characteristics and their raw values will not be shown in the summary output, and will not contribute to the distance.

5.2.4.9 Section 5 of this Annex provides a screen copy of a display tree which shows how the expert can navigate and visualise the results of computations.

5.2.4.10 GAIA software has been developed with WINDEV. The general information (species, characteristics, weighting, etc.), the data collected on the varieties and the results of computations are stored in an integrated database. Import and export facilities allow for other information systems to be used in connection with the GAIA software. ODBC allows access to the GAIA database and to other databases simultaneously.

5.2.4.11 1 or 2 notes per variety can be used. 1 note occurs when one cycle is available. Two notes are present for instance when two trials are made in different locations a given year, or if 2 cycles are obtained in the same location.
For electrophoresis data, only one description can be entered per variety.
For measurements at least 2 values (different trials, repeats, etc.) are necessary and the user can select which to use in the computation.

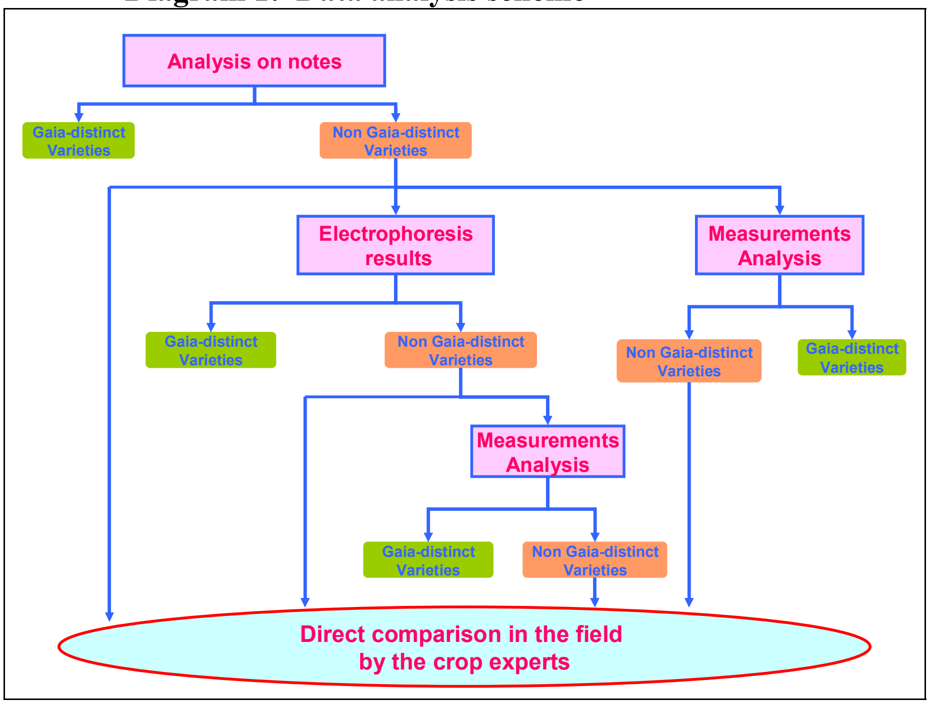
5.2.4.12 GAIA is most suitable for self-pollinated and vegetatively propagated varieties, but can also be used for other types of varieties.

5.2.5 Example with Zea mays data

5.2.5.1 Introduction

The software can use notes, measurements and/or electrophoresis results. These types of data can be used alone or in combination, as shown in Diagram 1.

Diagram 1: Data analysis scheme



In this example, it is assumed that the crop expert has decided to use a Distinctness Plus threshold $\mathbf{S}_{\mathbf{dist}}$ of 10 (see section 2 of this Annex).

5.2.5.2 Analysis of notes

5.2.5.2.1 *In qualitative analysis notes (1 to 9) are used. Notes can come from qualitative, quantitative and pseudo-quantitative characteristics.*

5.2.5.2.2 For each characteristic, weightings according to differences between levels of expression are pre-defined in a matrix of distances.

5.2.5.2.3 "Shape of ear": observed on a 1 to 3 scale, the crop expert has attributed weightings greater than zero to differences which they consider significant:

1 = conical
2 = conico-cylindrical
3 = cylindrical

		Variety 'i'		
Variety 'j'		1	2	3
	1	0	2	6
	2		0	2
	3			0

5.2.5.2.4 When the crop expert compares a variety ‘i’ with conical ear (note 1) to a variety ‘j’ with cylindrical ear (note 3), they attribute a weighting of 6.

5.2.5.2.5 “Length of husks”, observed on a 1 to 9 scale, the crop expert has defined the following weighting matrix:

1 = very short
2 = very short to short
3 = short
4 = short to medium
5 = medium
6 = medium to long
7 = long
8 = long to very long
9 = very long

		Variety ‘i’								
		1	2	3	4	5	6	7	8	9
Variety ‘j’	1	0	0	0	2	2	2	2	2	2
	2		0	0	0	2	2	2	2	2
	3			0	0	0	2	2	2	2
	4				0	0	0	2	2	2
	5					0	0	0	2	2
	6						0	0	0	2
	7							0	0	0
	8								0	0
	9									0

5.2.5.2.6 The weighting between a variety ‘i’ with very short husks (note 1) and a variety ‘j’ with short husks (note 3) is 0. The expert considers a difference of 3 notes is the minimum difference in order to recognise a non-zero distance between two varieties. Even if the difference in notes is greater than 3, the expert keeps the distance weight to 2 while in very reliable characteristics a difference of 1 is given a weight of 6.

5.2.5.2.7 The reason for using a lower weighting for some characteristics compared to others can be that they are less “reliable” or “consistent” (e.g. more subject to the effect of the environment); and/or they are considered to indicate a lower distance between varieties.

5.2.5.2.8 The matrix for a qualitative analysis for 5 characteristics for varieties A and B:

	Ear shape	Husk length	Type of grain	Number of rows of grain	Ear diameter	
Notes for variety A (1 to 9 scale)	1	1	4	6	5	
Notes for variety B (1 to 9 scale)	3	3	4	4	6	
Difference observed	2	2	0	2	1	
Weighting according to the crop expert	*6*	*0*	*0*	*2*	*0*	$D_{qual} = 8$

In this example **D_{qual} = 8 < 10** **(S_{dist}** =10 in this example) varieties A and B are declared “GAIA NON-distinct” on the basis of these 5 characteristics.

5.2.5.3 Electrophoresis analysis

5.2.5.3.1 *In some UPOV Test Guidelines electrophoresis results can be used, as in Zea mays. The software does not allow the use of heterozygous alleles, but only the use of homozygous allele, in conformity with the Guide lines. Results used are 0 (absent) and 1 (present), and the knowledge of chromosome number.*

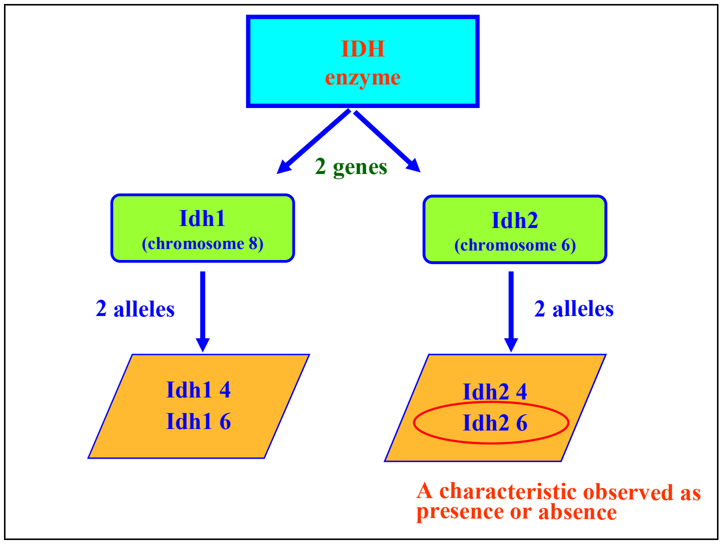


Diagram 2: The Isocitrate Deshydrogenase (IDH) enzyme has two genes (Idh1 and Idh2) located on two different chromosomes. Each of them has two alleles which are observed as 1 (presence) or 0 (absence).

5.2.5.3.2 Electrophoresis results are noted as 0 or 1 (absence or presence). The decision rule, used to give a weighting to two varieties, is the addition of the weighting number of differences observed and the weighting number of chromosomes related to these differences (see example below):

	Chromosome 8		Chromosome 6	
	Idh1 4	Idh1 6	Idh2 4	Idh2 6
Variety A	0	1	1	0
Variety B	0	1	0	1
Difference	0	0	1	1

5.2.5.3.3 In this example, varieties A and B are described for 4 electrophoresis results:

Idh1 4, Idh1 6, Idh2 4 and Idh2 6. The software looks at differences and gives the phenotypic distance using the following computation:

$$\mathbf{D_{elec} = 2 \quad x \quad 0.25 \quad + \quad 1 \quad x \quad 1 \quad = \quad 1.5}$$

2 is the number of differences observed

0.25 is the weighting attributed by experts to the number of differences

1 is the number of chromosome on which differences are observed

1 is the weighting associated by experts to chromosome.

5.2.5.3.4 This formula, which might be difficult to understand, was established by the crop expert in collaboration with biochemical experts. Both the *number of differences* and the *number of chromosomes on which differences are observed* are used. Thus, less importance is attached to differences when these occur on the same chromosome, than when they occur on different chromosomes.

5.2.5.3.5 After qualitative and electrophoretic analysis, the phenotypic distance between varieties A and B is equal to:

$$\mathbf{D = D_{qual} + D_{elec} = 8 + 1.5 = 9.5}$$

5.2.5.3.6 The phenotypic distance is *lower than* $\mathbf{S_{dist}}$ ($\mathbf{S_{dist}}$=10 in this example) *therefore varieties A and B are considered "GAIA NON-distinct".*

5.2.5.3.7 The crop expert can decide he does not want to establish distinctness solely on the basis of electrophoresis analysis. It is necessary to have a minimal phenotypic distance in qualitative analysis in order to take into account the electrophoresis results. This minimal phenotypic distance must also be defined by the crop expert.

5.2.5.4 Analysis of measurements

5.2.5.4.1 *Analysis of measurements computes differences on observed or computed measurements, counts are handled as measurements*

5.2.5.4.2 For each measured characteristic, the comparison of two varieties is made by looking for consistent differences in at least two different experimental units. Experimental units are defined by the user depending on data present in the database. It can, for example, be the data from two geographical locations of the first growing cycle, or 2 or 3 replications from the same trial in the case of a single geographical location, or data from 2 cycles in the same location.

5.2.5.4.3 For a comparison to be made, the two varieties must be present in the same experimental units. The differences observed must be greater than one of the two threshold values (or minimal distances), fixed by the crop expert.

- $\mathbf{D_{min\text{-}inf}}$ is the lower value from which a weighting is attributed,

- $\mathbf{D_{min\text{-}sup}}$ is the higher minimal distance. These values could be chosen arbitrarily or calculated (15% and 20% of the mean for the trial, or LSD at 1% and 5%, etc.)

For each minimal distance a weighting is attributed:

- $\mathbf{D_{min\text{-}inf}}$ a weighting P_{min} is attributed;

- $\mathbf{D_{min\text{-}sup}}$ a weighting P_{max} is attributed;

- the observed difference is lower than $\mathbf{D_{min\text{-}inf}}$ a zero weighting is associated.

5.2.5.4.4 Varieties A and B have been measured for characteristics "Width of blade" and "Length of plant" in two trials.

For each trial, and each characteristic, the crop expert has decided to define (**$D_{min-inf}$**) and **$D_{min-sup}$** by calculating respectively the 15% and 20% of the mean for the trial:

	Width of blade		Length of plant	
	Trial 1	Trial 2	Trial 1	Trial 2
$D_{min-inf}$ = 15% of the trial mean	1.2 cm	1.4 cm	28 cm	24 cm
$D_{min-sup}$ = 20% of the trial mean	1.6 cm	1.9 cm	37 cm	32 cm

For each characteristic, the crop expert has attributed the following weighting:

A weighting **P_{min}** = 3 is attributed when the difference is greater than **$D_{min-inf}$**.

A weighting **P_{max}** = 6 is attributed when the difference is greater than **$D_{min-sup}$**.

	Width of blade		Length of plant		
	Trial 1	Trial 2	Trial 1	Trial 2	
Variety A	9.9 cm	9.8 cm	176 cm	190 cm	
Variety B	9.6 cm	8.7cm	140 cm	152 cm	
Difference	0.3 cm	1.1 cm	36 cm	38 cm	
Weighting according to the crop expert	0	0	3	6	D_{quan} =?

5.2.5.4.5 In this example, for the characteristic "Width of blade", the differences observed are lower than **$D_{min-inf}$**, so no weighting is associated. On the other hand, for the characteristic "Length of plant" one difference is greater than the **$D_{min-inf}$** value and the other is greater than the **$D_{min-sup}$** value. These two differences are attributed different weightings.

5.2.5.4.6 The user must decide which weighting will be used for the analysis:

- the weighting chosen is that attributed to the lowest difference (minimalist option);
- the weighting chosen is that attributed to the highest difference (maximalist option);
- mean option: the weighting chosen is the mean of the others (mean option).

5.2.5.4.7 In this example, the crop expert has decided to choose the lowest of the two weightings, so the phenotypic distance based on measurements is **D_{quan} = 3**.

5.2.5.4.8 In summary, at the end of all analysis, the phenotypic distance between varieties A and B is:

$D = D_{qual} + D_{elec} + D_{quan} = 8 + 1.5 + 3 = 12.5 > S_{dist}$

5.2.5.4.9 The phenotypic distance is greater than the distinction threshold **S_{dist}**, fixed by the crop expert at 10, so varieties A and B are declared "GAIA-distinct".

5.2.5.4.10 In this example, the use of electrophoresis data “confirms” a distance between the two varieties; but on the basis of qualitative and quantitative data alone, the threshold is exceeded (8 + 3 = 11 is greater than 10).

5.2.5.4.11 If the threshold had been set at 6, the difference on the characteristic ear shape would have been sufficient, as variety A is conical and variety B is cylindrical, which is already a clear difference.

1 = conical
2 = conico-cylindrical
3 = cylindrical

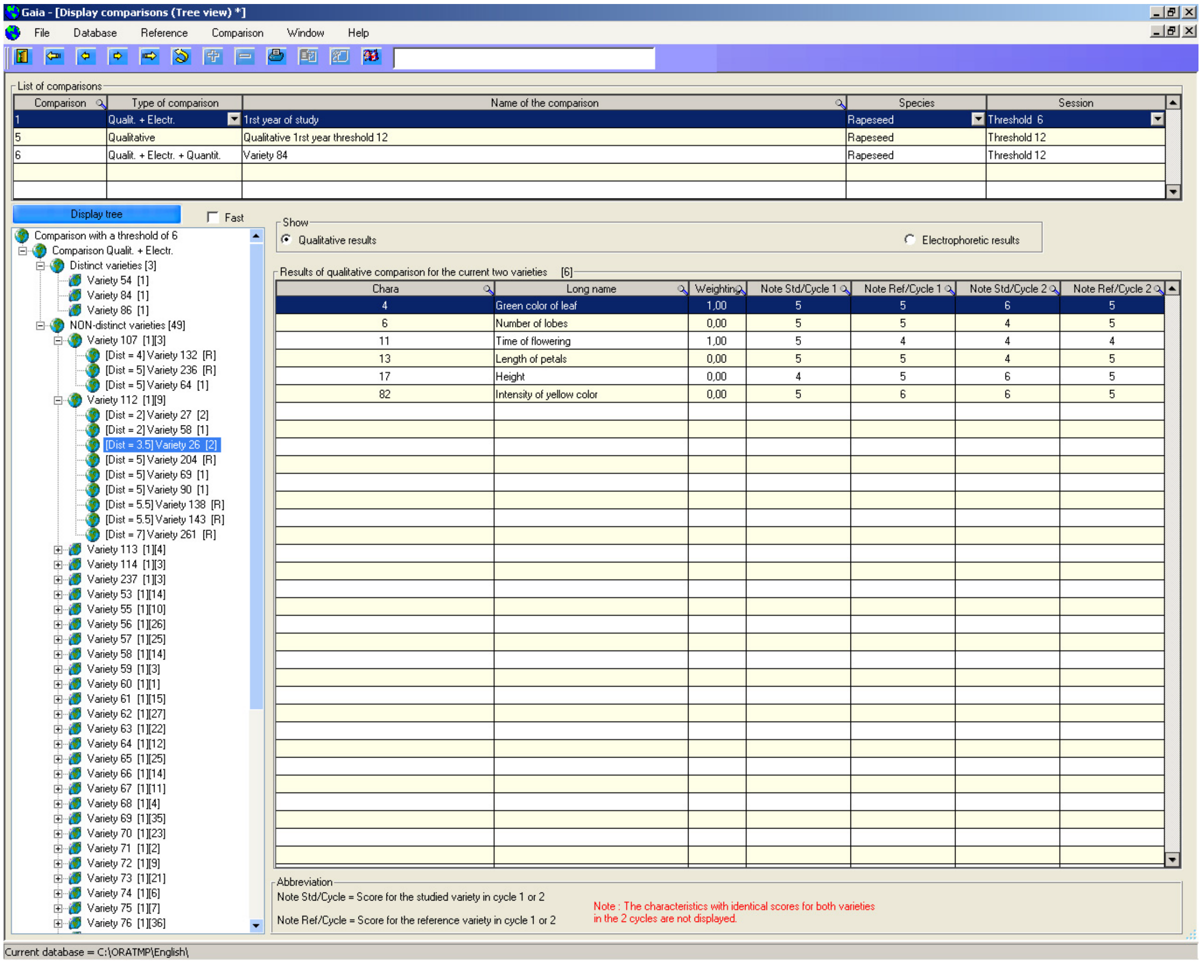
Variety i			
	1	2	3
1	0	2	6
2		0	2
3			0

5.2.5.5 Measurements and 1 to 9 scale on the same characteristic

5.2.5.5.1 For some crops, it is common practice to produce values on a 1 to 9 scale from measurements. Sometimes the transformation process is very simple, sometimes it is complex.

5.2.5.5.2 GAIA can include both as two separate characteristics: the original measurements and the 1 to 9 scale. They are associated in the description of the characteristics. Using the knowledge of this association, when both are present, only one of them is kept, in order to avoid the information being used twice for weighting.

5.2.5 Example of GAIA screen copy



5.2.5.1 The upper part "List of comparisons" shows 3 different computations which have been kept in the database. Comparison 1 is highlighted (selected) and shown on the display tree.

5.2.5.2 The "Display tree" on the left shows results for a [qualitative + electrophoresis at threshold of 6] computation.

5.2.5.3 *Distinct varieties [3]* indicates that 3 varieties were found distinct from all others. There was a total of 52 (49 + 3) varieties in the computation.

5.2.5.4 The display tree is used to navigate through all possible pairs.

5.2.5.5 The user can expand or reduce the branches of the tree according to his needs.

5.2.5.6 *NON-distinct varieties [49]*. Forty-nine varieties were found "not distinct from all others" with a threshold of 6.

5.2.5.7 The first variety, *Variety 107*, has only 3 close varieties, whereas the second, *Variety 112*, has 9 close varieties, the third, *Variety 113*, 4 close varieties, etc.

5.2.5.8 *Variety 112 [1][9]* indicates variety 112 is in the first year of examination *[1]*; and has 9 close varieties according to the threshold of 6 *[9]*.

5.2.5.9 *[dist=3.5]Variety 26 [2]* indicates variety 26 (comparison highlighted=selected) has a GAIA distance of 3.5 from variety 112, which is in second year of examination.

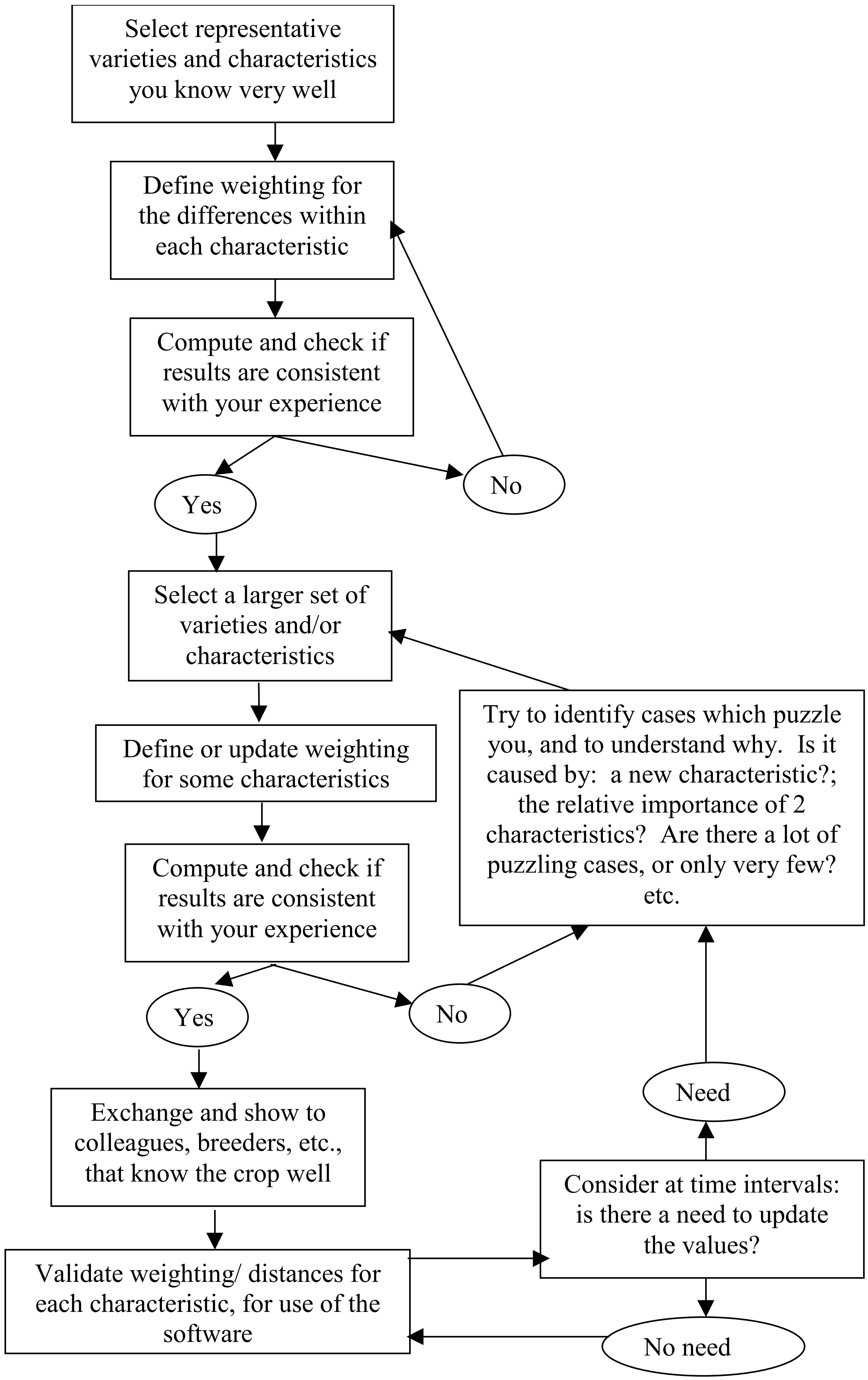
5.2.5.10 On the right of the Display tree, the raw data for *Variety 112* and *Variety 26* are visible for the 6 qualitative characteristics observed on both varieties (two cycles).

5.2.5.11 The third column "weighting" is the weighting according to the pre-defined matrices. The notes for both varieties are displayed for the two available cycles (Std stands for "studied" which are the candidate varieties).

5.2.5.12 As noted in red, if two varieties have the same description on a given characteristic, this characteristic is not displayed.

5.2.5.13 In this screen copy the varieties have been numbered for sake of confidentiality, the crop expert can name the varieties according to their need (lot or application number, name, etc.).

Diagram 3: "Try-and-check" process to define and revise the weightings for a crop



6. EXAMINING DUS IN BULK SAMPLES

6.1 Introduction and abstract

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 for individual plants. Some examples are seed weight in cereals and peas and beans, and erucic acid content in rapeseed. This section describes some of the consequences of bulk sampling. It is shown that the test of distinctness (using COYD, see Part II: Section 2.1 *[cross ref.]*) may be expected to be relatively insensitive to the degree of bulking, but that the efficiency of the tests for uniformity (using COYU, see Part II: Section 2.2 *[cross ref.]*) 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.

6.2 Distinctness

6.2.1 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.

6.2.2 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.

$$\sigma^2_{total} = \sigma^2_{vy} + \sigma^2_p + \sigma^2_i + \sigma^2_m$$

where

σ^2_{total} 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^2_{vy} the year in which the variety is measured

2: σ^2_p the plot in which the measurement was taken

3: σ^2_i the plant on which the measurement was taken

4: σ^2_m the inaccuracy in the measurement process

6.2.3 In cases where the data are 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

6.2.4 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

6.2.5 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.

Example 1

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 of 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.

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.

6.2.6 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.

6.3 Uniformity

6.3.1 Bulking within plot

6.3.1.1 In COYU the test is based on the standard deviation of the individual plant observations (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 variance 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 $\frac{1}{2v}\left(\frac{\sigma}{\sigma+1}\right)^2$ when then recorded variable is normally distributed and the standard deviations do not vary too much. This last expression reduces to $0.5/v$ when $\sigma >> 1$. Here σ is the mean value of the s_{vy} values and v is the number of degrees of freedom used in the estimation of s_{vy}.

6.3.1.2 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.

6.3.1.3 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

Example 2

The effect of bulking in the test for uniformity, an estimate was made using the same data as for Example 1 I Part II, Section 6.2.5 [cross ref.]. 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 degree of non-uniformity must be much higher for it to be detected when 2 bulk samples are used instead of individual plant records.*

6.3.2 Bulking across plots

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.

6.3.3 Taking just one bulk sample per plot

In general, if all plants in a plot are bulked such that only a single sample is available for each plot, it becomes impossible to calculate the within plot variability and in such 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 show up under certain circumstances.

[End of document]