E

TGP/8/1 Draft 6

INTRODUCTION &

PART I: DUS Trial Design and Data Analysis

ORIGINAL: English only

DATE: March 18, 2007

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]

Introduction

Part I: DUS Trial Design and Data Analysis

Document prepared by the Office of the Union

to be considered by the Technical Committee at its forty-third session,
to be held in Geneva, Switzerland, from March 26 to 28, 2007

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

~~*2.5.2.5. Reason for using statisticsi*~~

~~Statistics are generally used when data from the growing trial are subject to variation. Variation tends to obscure differences between varieties, so making comparisons difficult, which in turn could lead to mistaken decisions about a variety's distinctness, uniformity or stability. Statistics permit the Crop Expert to make allowance for the variability in the data and so make decisions about the candidate variety with a certain level of confidence that the decision is the correct one.~~

~~Examples of variation encountered in DUS data include:-~~

- ~~The variation in the observations from plant to plant in a quantitative characteristic for a cross pollinated variety reflects the variety's uniformity in that characteristic. Because of sampling variation this uniformity will vary from plot to plot and from growing trial to growing trial for the one variety. The Crop Expert wishes to use the data on the observations of the characteristic to assess a candidate variety's uniformity and to compare it with an assessment of the uniformity of established varieties, which is also based on growing trial data, and hence is variable.~~

- ~~The variety mean of a quantitative characteristic for a cross pollinated variety will vary from growing trial to growing trial. The Crop Expert wishes to determine whether the difference between the over-year means for two varieties is sufficiently large compared to the year to year variation in the variety means to be able to say that the difference is consistent, and so that the two varieties are distinct in that characteristic.~~

Moved to 2.5.2.7

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."
>
> [………]
>
> **2.3 Design of DUS Test**
>
> "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 2.2.1 *[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. The rationale is that if the observed difference in a characteristic is sufficiently consistent, or unchanging, then that difference should be observed if the varieties were compared again in a similar location. To determine whether this is the case, the variation in variety differences from growing cycle to growing cycle must be considered for the characteristic. Comparison of the observed difference in the characteristic with this variation indicates which of the following is likely to apply

- The observed difference in the characteristic is large compared to the variation in variety differences from growing cycle to growing cycle, i.e. the observed difference is consistent.

or

- The observed difference in the characteristic is not large compared to the variation in variety differences from growing cycle to growing cycle. In which case, the observed difference could have arisen by chance as a result of this growing cycle to growing cycle variation in the growing cycles tested, i.e. the observed difference is not consistent.

2.2.2.2 The variation in variety differences from growing cycle to growing cycle is also known as the variety x growing cycle interaction.

2.2.2.3 The independence of the growing cycles is essential as it allows the variety x growing cycle interactions to be assessed without bias. If the growing cycles are not independent, say for example if the trials were conducted simultaneously in neighboring fields, then the variety x growing cycle interactions are likely to be minimal and the observed differences appear to be more consistent than they actually are.

2.2.2.4 Exactly what constitutes an independent growing cycle depends on what factors interact with the varieties. If the main factors to interact with the varieties are aspects of the environment that change with years, then the independent growing cycles should take place in different years, and two trials within one year would not suffice, as they would not independently sample the yearly effects.

2.2.2.5 In general, the assessment of independence is based on the experience of experts.

2.2.2.6 When a characteristic is observed in a growing trial in two independent growing cycles, it is generally observed on two completely separate trials in the form of two separate plantings or sowings. The exception to this is the case of some perennial crops, such as fruit trees, where the growing cycles take the form of one trial observed in two successive years.

2.2.2.7 When field or greenhouse crop varieties are planted/sown in successive years and the layout of the plants in the trial is randomized (at least partly), the independence of the growing cycles is to be satisfied.

2.2.2.8 In the case of plants grown in greenhouses or other highly controlled environments, 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.

2.2.2.9 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 ~~under certain circumstances~~ satisfy the requirement for independence.

2.2.2.10 Where the two growing cycles are in the same location and the same year, a suitable time period between plantings may ~~under certain circumstances~~ 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:

(a) 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.

(b) 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.

(c) 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 2.3 "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.

2.2.3.2 Use of information from multiple locations

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]/[DUS] 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

(b) distinctness established using characteristics examined at different locations

For example, additional tests (see Section 2.3) *[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) are 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 2.2.3.1(a) *[cross ref.]*), a second location might be used to check the consistency of a difference observed in the first location). Such cases would

---

[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".

normally apply where the assessment of distinctness is based on Notes (see TGP/9 Sections 5.2.1.1(b) and 5.2.3*[cross ref.]*) and the assessment of distinctness and the variety description could be considered as 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 TGP/9 Sections 5.2.1.1(c) and 5.2.4*[cross ref.]*) 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 (see Section 2.2.2.6 *[cross ref.]*). In particular, careful consideration would need to be given to check if the necessary assumptions would be satisfied. Further, the COYD criterion was tested on data over different years and not tested on 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.

2.3 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.

2.4 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.

2.5 Organizing the growing trial layout

2.5.1 Type of trial layout

The type of trial layout will be determined by the approach to be used for the assessment of distinctness. This determines the organization of the trial layout in terms of 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.

2.5.2 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.”

2.5.2.1 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.”

2.5.2.2 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.”

2.5.2.3 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. ”

2.5.2.4 Trial layout for side by side (visual) comparison

Where the assessment of distinctness is based on a direct visual comparison of side-by-side varieties in the growing trial, the trial will comprise of a number of plots. The plots will be grouped into one or more replicates such that each replicate contains at least one plot of each variety. The allocation of the varieties to plots will be decided by the Crop Expert, who will arrange for similar varieties to be grown either side-by-side or in close proximity with other similar varieties. If all similar varieties must be grown side-by-side with all other similar varieties, then some varieties may be present in more than one plot within a replicate.

2.5.2.5 Trial layout for assessment of distinctness by notes/single variety records

Where the assessment of distinctness is based on a single recorded state of expression of a characteristic, i.e. a Note, for a variety obtained from the growing trial, the trial will comprise of a number of plots. The plots will be grouped into one or more replicates such that each replicate contains one plot of each variety. The allocation of the varieties to plots will be decided by the Crop Expert.

2.5.2.6 Trial layout for assessment by statistical analysis of growing trial data

Where the assessment of distinctness is based on statistical analysis of growing trial data, the trial will comprise of a number of plots. The plots will be grouped into one or more replicates such that each replicate contains one plot of each variety. As will be discussed in a later section, the allocation of varieties to plots will involve randomization.

2.5.2.7. Reason for using statistics[ii]

Statistics are generally used when data from the growing trial are subject to variation. Variation tends to obscure differences between varieties, so making comparisons difficult, which in turn could lead to mistaken decisions about a variety's distinctness, uniformity or stability. Statistics permit the Crop Expert to make allowance for the variability in the data and so make decisions about the candidate variety with a certain level of confidence that the decision is the correct one.

Examples of variation encountered in DUS data include:-

- The variation in the observations from plant to plant in a quantitative characteristic for a cross pollinated variety reflects the variety's uniformity in that characteristic. Because of sampling variation this uniformity will vary from plot to plot and from growing trial to growing trial for the one variety. The Crop Expert wishes to use the data on the observations of the characteristic to assess a candidate variety's uniformity and to compare it with an assessment of the uniformity of established varieties, which is also based on growing trial data, and hence is variable.

- The variety mean of a quantitative characteristic for a cross pollinated variety will vary from growing trial to growing trial. The Crop Expert wishes to determine whether the difference between the over-year means for two varieties is sufficiently large compared to the year to year variation in the variety means to be able to say that the difference is consistent, and so that the two varieties are distinct in that characteristic.

2.6 Trial elements

It is important when deciding on trial layout that, where possible, local variation in conditions is controlled by the choice of plot size, shape and configuration. Otherwise this local variation, such as variation in soil conditions, or partial shade from a nearby structure on the trial, might influence the results of the trials. Equally important when distinctness is assessed by statistical analysis of growing trial data are the random allocation of varieties to plots and the independence of the plots. These issues are discussed in the following sections.

2.6.1 Plots and the allocation of varieties to plots

2.6.1.1 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 plant pots

2.6.1.2 Random allocation of varieties to plots is not generally used in side-by-side comparison trials and when arranging varieties into groups in trials where the distinctness is to be assessed by notes or single variety records. However, whenever distinctness is assessed by statistical analysis of growing trial data, depending on the trial design either randomisation or partial randomisation must be used, as it ensures that there is no subjectivity in the allocation.

2.6.1.3 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.

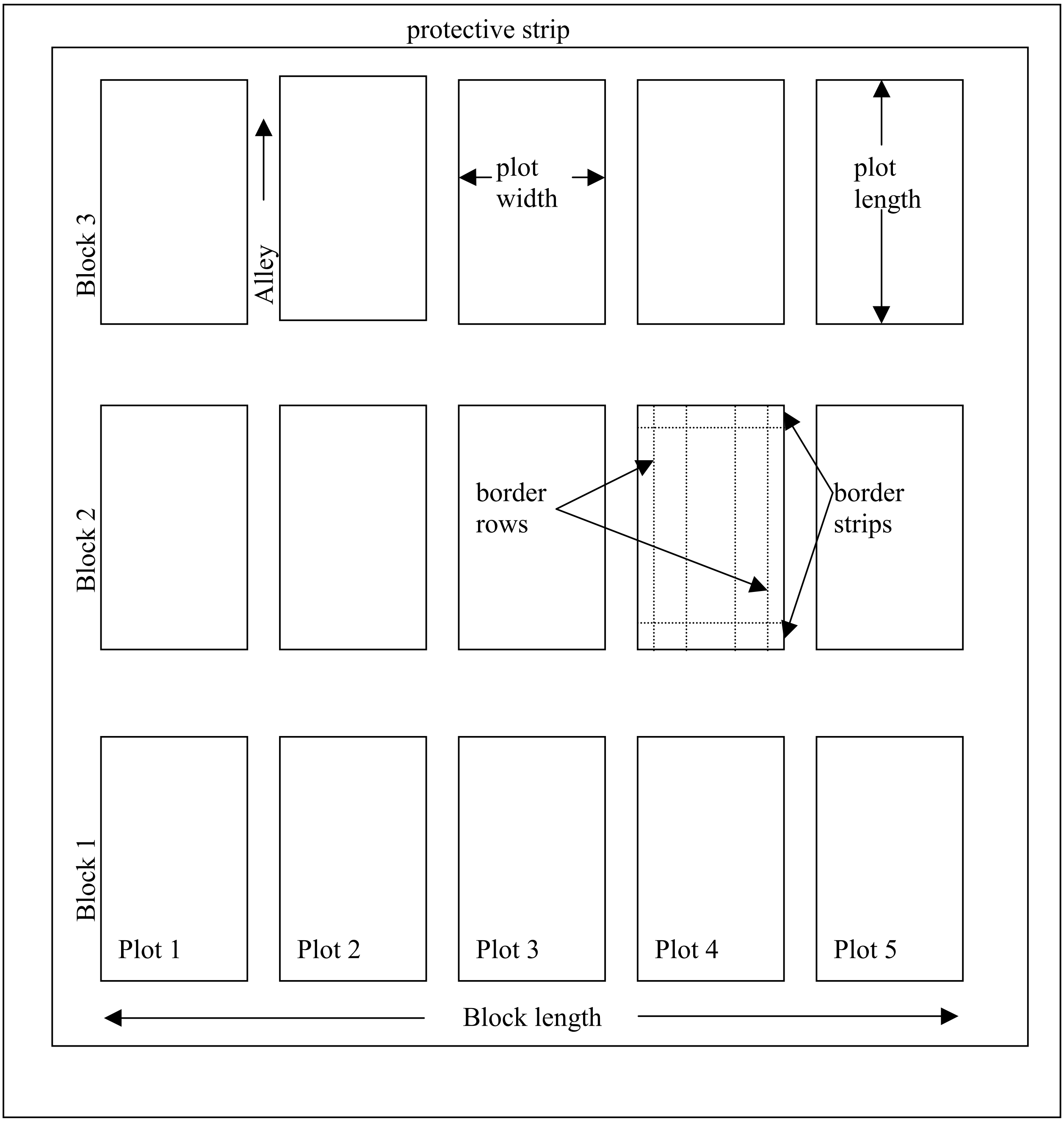
2.6.1.4 A block is a group of plots within which the varieties are allocated at random. A growing trial may contain just one block or it may contain more than one block.

2.6.2 Plot size, shape and configuration

2.6.2.1 In deciding on trial layout, it is important that local variation in conditions are taken into account. 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

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



2.6.2.3 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.

2.6.2.4 The Test Guidelines may 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)). Uniformity, however is assessed on the whole sample under examination by the off-type approach and/or by the standard deviation approach (see TGP/10 Section 3 *[cross ref.]*). 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.

2.6.2.5 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).

2.6.2.6 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.

2.6.2.7 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.

2.6.3 Independence of plots

2.6.3.1 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.

2.6.3.2 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 interfering with 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 grow physically similar varieties together.

2.6.4 The arrangement of the plants within the plot[2]

~~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] To be covered by TGP/14

2.7 Statistical aspects of trial design

2.7.1 Introduction

2.7.1.1 This section describes a number of concepts that are relevant when designing growing trials for which distinctness and/or uniformity are to be assessed by statistical analysis of the growing trial data. TGP/9 section 5 *[cross ref.]* provides guidance where the assessment of distinctness is based on a statistical analysis of the data obtained from the growing trial, and TGP/10 section 2 *[cross ref.]* provides guidance where the assessment of uniformity is on the basis of standard deviations.

2.7.1.2 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.

2.7.2 The hypotheses under test

2.7.2.1 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 some quantitative characteristics, and in the case of qualitative, pseudo-qualitative, and other quantitative characteristics on the number of off-types present in a sample.

2.7.2.2 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

2.7.2.3 We make each decision by computing a test statistic from the observations using a formula. If the absolute value of 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.

2.7.2.4 Note that if the null hypothesis is rejected for distinctness, this leads to the conclusion that the candidate variety is distinct.

2.7.2.5 On the other hand, if the null hypothesis is rejected for uniformity, the candidate variety is considered not uniform.

2.7.2.6 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 a different conclusion is arrived at compared to the conclusion that would be arrived at if all plants of a variety could be examined. 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 if all plants of the two varieties could be examined, they would not be 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 if all plants of the two varieties could be examined, they would be distinct. This is known as a Type II error and its risk is denoted by β.

2.7.2.7 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	
Decision that would be made if all plants of a variety could be examined	Varieties are distinct (H1 true)	Varieties are not distinct (H0 true)
Varieties are distinct (H1 true)	Same decision	Different decision, Type II error, made with probability β
Varieties are not distinct (H0 true)	Different decision, Type I error, made with probability α	Same decision

2.7.2.8 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 if all plants of the variety could be examined, the variety would be considered uniform, i.e. a Type I error (α). Alternatively, a Type II error (β) is the decision based on a test statistic that a variety is uniform when, if all plants of the variety could be examined, the variety would not be considered uniform. The following table shows the two types of statistical error that can be made when testing for uniformity:

	Decision based on test statistic	
Decision that would be made if all plants of a variety could be examined	Variety is uniform (H0 true)	Variety is not uniform (H1 true)
Variety is uniform (H0 true)	Same decision	Different decision, Type I error, made with probability α
Variety is not uniform (H1 true)	Different decision, Type II error, made with probability β	Same decision

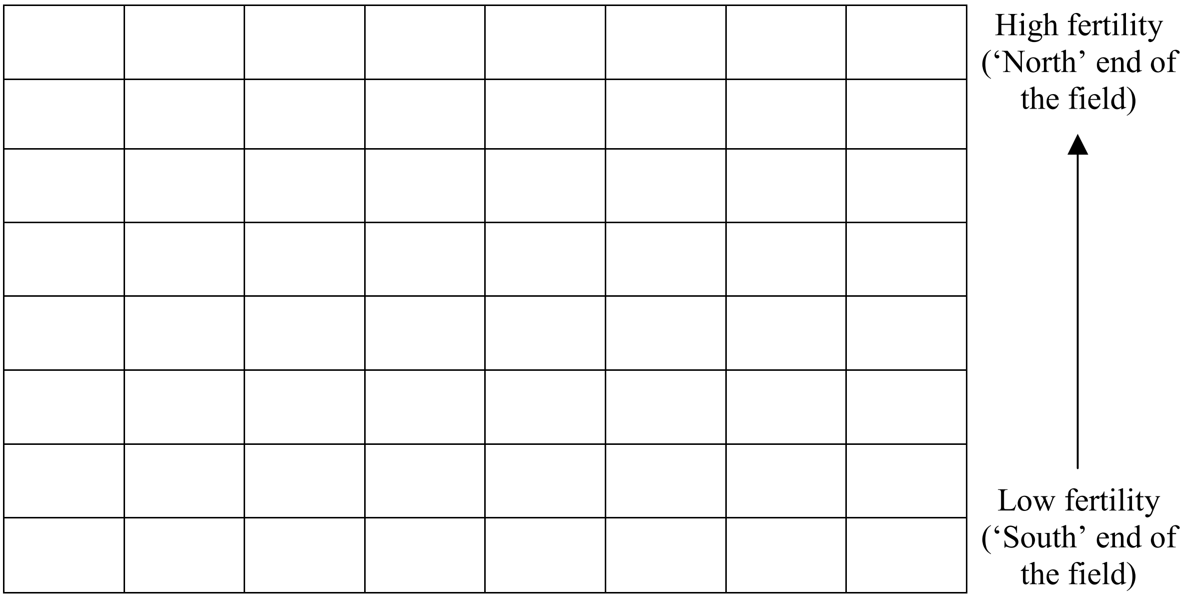
2.7.2.9 The risk of making a type I error can be controlled easily by choice of α, which determines the critical value that the test statistic is compared against. α is also known as the size of the test and the significance level of the test. The risk of making a type II error is more difficult to control 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, e.g. by increasing the number of replicates, by reducing the random variability by choice of number of plants per plot (or sample size), by controlling local, unwanted or nuisance variation through careful choice of experimental design, and by improving the way measurements/observations are made and so reducing the observer error. [Experts are invited to develop last sentence]

2.7.3 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.

2.7.4 Completely randomized design and randomized complete block design

2.7.4.1 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:



2.7.4.2 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.

2.7.4.3 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

2.7.4.4 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.

2.7.4.5 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:

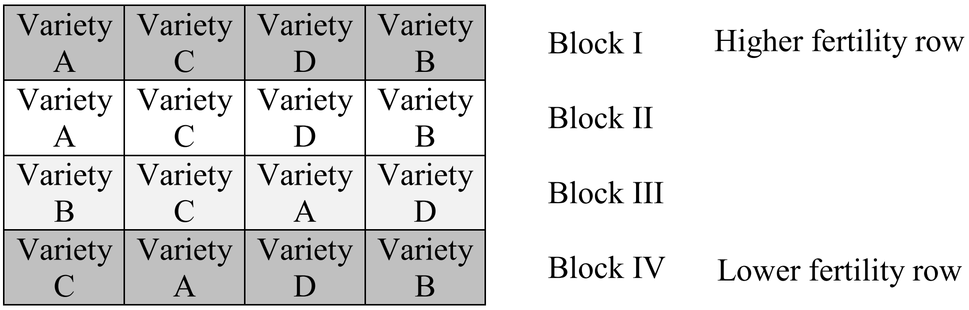
Block I				Block II				Block III				Block IV			
Var A	Var C	Var D	Var B	Var A	Var C	Var D	Var B	Var B	Var C	Var A	Var D	Var C	Var A	Var D	Var B

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

2.7.4.6 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.

2.7.4.7 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.

2.7.4.8 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. Growing physically similar varieties together, e.g. varieties of similar 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

2.7.5 Randomized incomplete block designs

2.7.5.1 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 partially balanced 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.

2.7.5.2 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	F	E	O	S
	4	M	H	J	T
	3	B	C	D	G
	2	L	A	R	N
	1	Q	K	P	I
Block	Sub-block	Variety			
2	5	D	P	F	A
	4	R	E	J	B
	3	N	G	Q	H
	2	K	S	M	C
	1	O	I	T	L
Block	Sub-block	Variety			
1	5	D	T	E	Q
	4	B	M	A	I
	3	C	F	L	H
	2	R	G	K	O
	1	P	J	N	S

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. Note: in the literature, the blocks and sub-blocks are sometimes referred to as super-blocks and blocks.

2.7.5.3 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.

2.7.6 Design for pair-wise comparisons between particular varieties

2.7.6.1 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.

2.7.6.2 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.

2.7.7 The effect of sample size on precision and decision making

2.7.7.1 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, say 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)

2.7.7.2 To estimate the optimal sample size for a quantitative characteristic 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 (β) error probabilities. 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 quantitative 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 % which may be unacceptable.

2.7.7.3 To estimate the optimal sample size for pseudo-qualitative characteristics such as segregating characteristics, the final precision depends on the same sources of variation described above except for that between individual plants within a plot. The process of choosing the optimal sample size is also as described above.

2.7.7.4 There are no recommendations on choosing the optimal sample size for quantitative, qualitative or pseudo-qualitative characteristics in trials where distinctness is assessed by notes or single variety records or by side by side visual comparison. In these cases the optimal sample size is based on the experience of experts[3].[TWP's invited to comment]

2.7.8 The impact of precision on analyses over years or cycles

2.7.8.1 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 3.1 and 3.2 [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.

2.7.8.2 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.

2.8 Aspects of trial design relevant to when statistical analysis are not used

2.8.1 Control of variation due to different observers

[If this section is required, TWPs are invited to contribute guidance on the control of variation due to different observers when statistical analysis is not used to determine distinctness and to consider it in relation to paragraph 2.7.2.9.]

3. VALIDATION OF DATA AND ASSUMPTIONS

3.1 Introduction

3.1.1 When data are observed on plots in the growing trial it is important that the data are representative. 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 analyzed by statistical methods. Thus they are done on data observed on qualitative, pseudo-qualitative and quantitative characteristics, and observed as a single record for a group of plants or parts of plants (G), and records for a number of single individual plants or parts of plants (S).

3.1.2 If the data are to be statistically analyzed (see TGP/9 section 5 *[cross ref.]* for guidance on when the assessment of distinctness is based on a statistical analysis of the data obtained

---

[3] TWC Chairperson: TWPs may contribute to describe the sample size is decide in side-by-side trials

from the growing trial, and TGP/10 section 2 *[cross ref.]* for guidance on when the assessment of uniformity is on the basis of standard deviations), 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.

3.1.4 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 Graphical displays, or plots of the characteristics may help to validate the data. For example, examination of the frequency distributions of the characteristics may identify small groups of discrepant observations. Also, in the case of quantitative characteristics, examination of scatter plots of pairs of characteristics that are likely to be highly related may detect discrepant observations very efficiently.

3.2.4 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 quantitative 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).

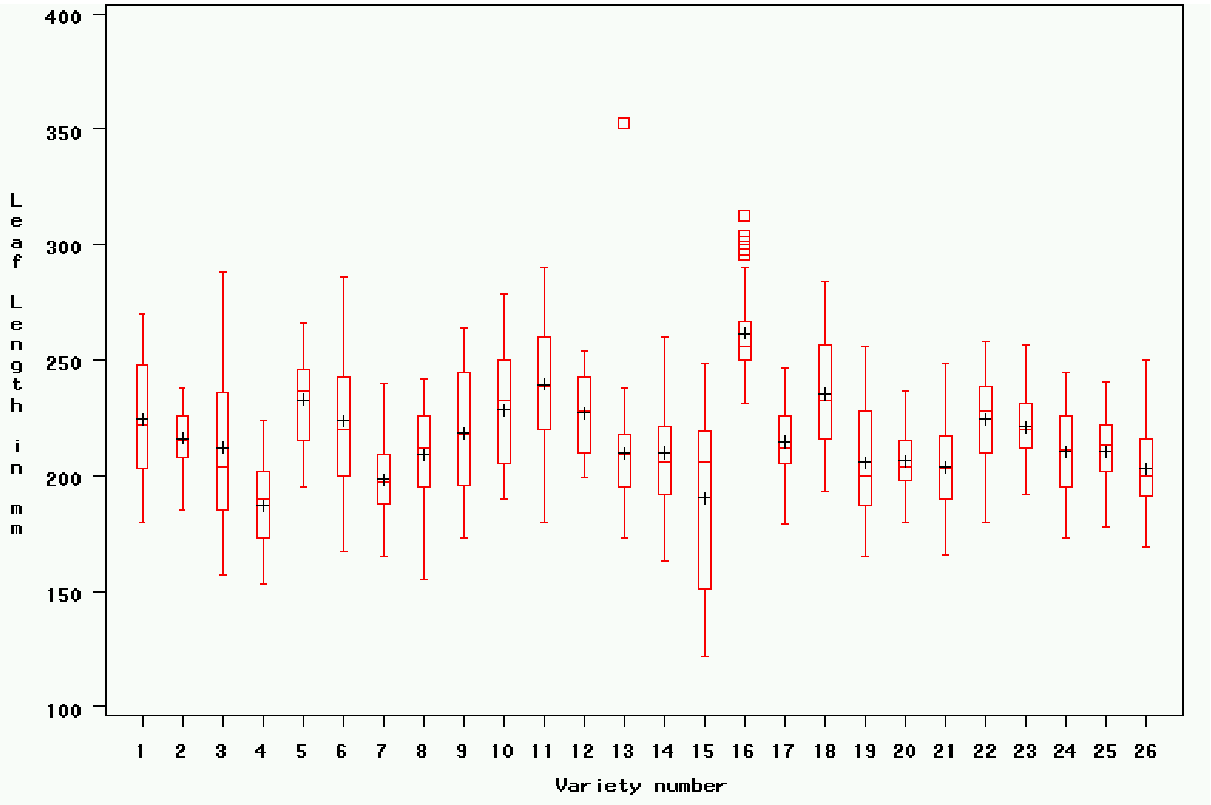


Figure 1. Box-plot for Leaf Length of 26 varieties of oil seed rape

3.2.5 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.

3.2.6 [TWPs are invited to contribute on any other methods of validation used where statistical analysis is not used to determine distinctness.]

3.3 Assumptions necessary if the data are to be statistically analysed

3.3.1 Introduction

3.3.1.1 Firstly, it is essential that the growing trial/experiment is designed properly and involves 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
- 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 based on the Method of Least Squares are used to test hypotheses. When such statistical methods are used only to estimate effects (means), the assumptions are less important and the assumption of normally 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 2.7 *[cross ref.]* “Aspects of trial design relevant to when statistical analysis will be used”

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.

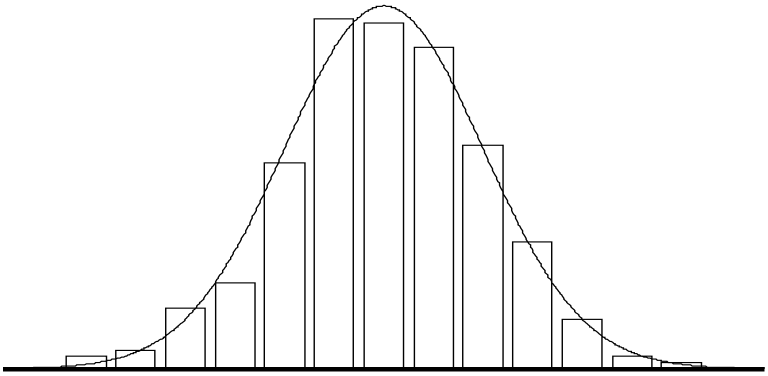


Figure 2. Histogram for normal distributed data with the ideal normal distribution shown as a curve

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

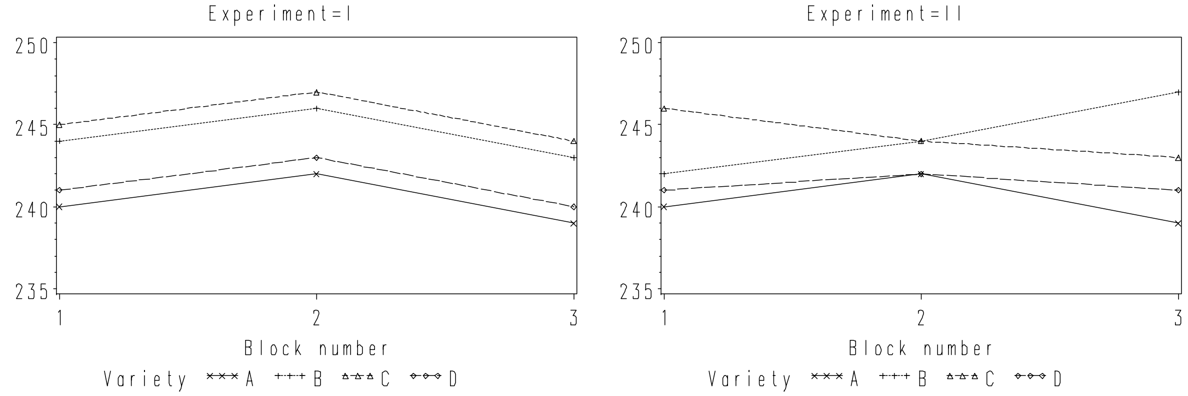


Figure 3. Artificial plot means from two experiments showing additive block and variety effects (left) and non-additive block and variety effects (right) using same data as in table 2

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 necessary if the data are to be statistical analysed

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

3.4.3.1 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.

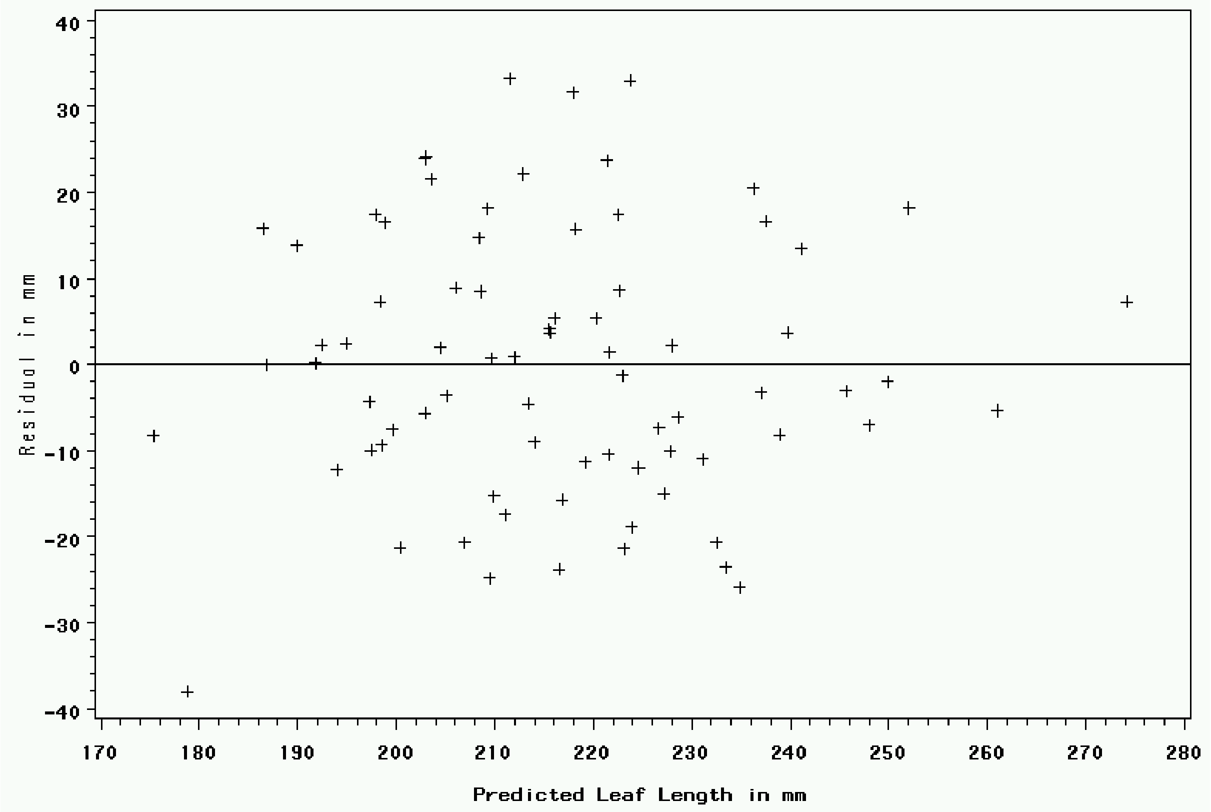


Figure 4. Plot of residuals versus plot predicted values for Leaf Length in 26 oil seed rape varieties in 3 blocks

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.

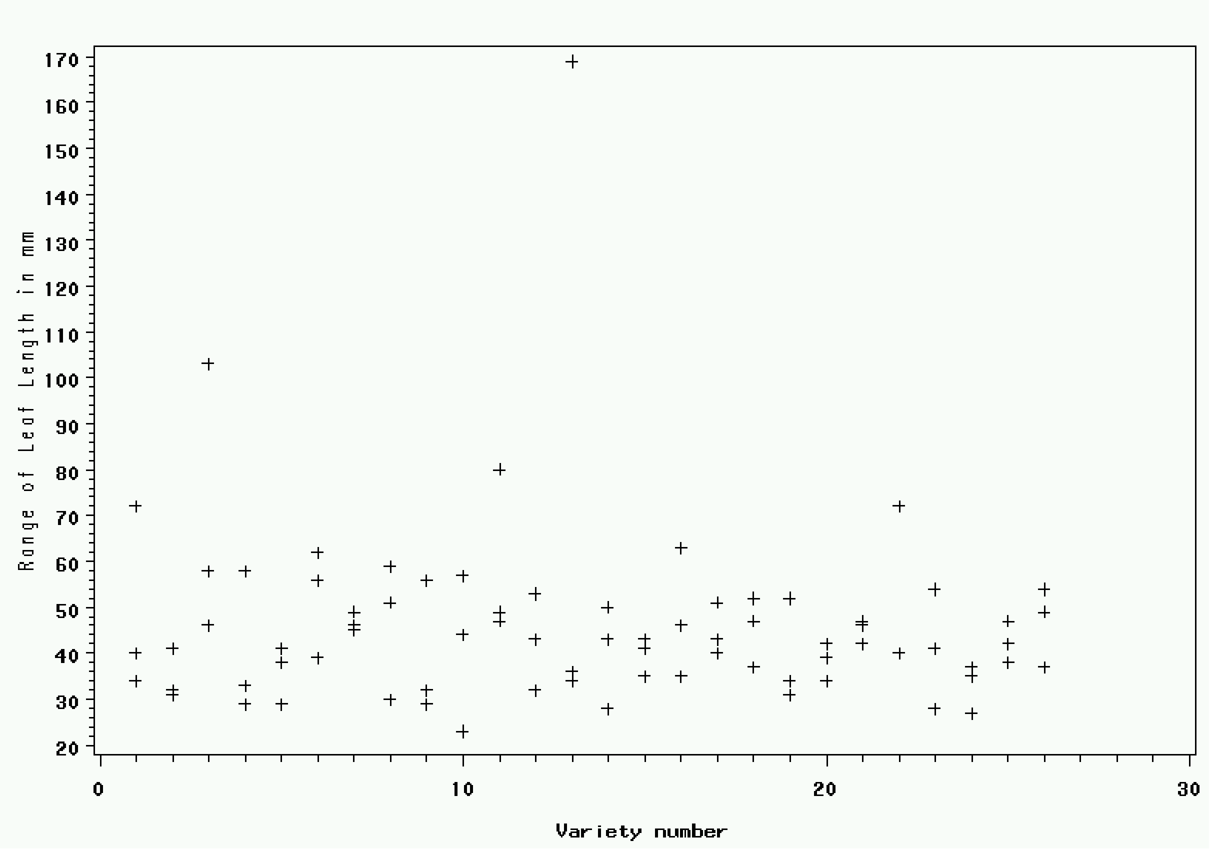


Figure 5. Differences between minimum and maximum of 20 leaf lengths for 3 plots versus oil seed rape variety number

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).

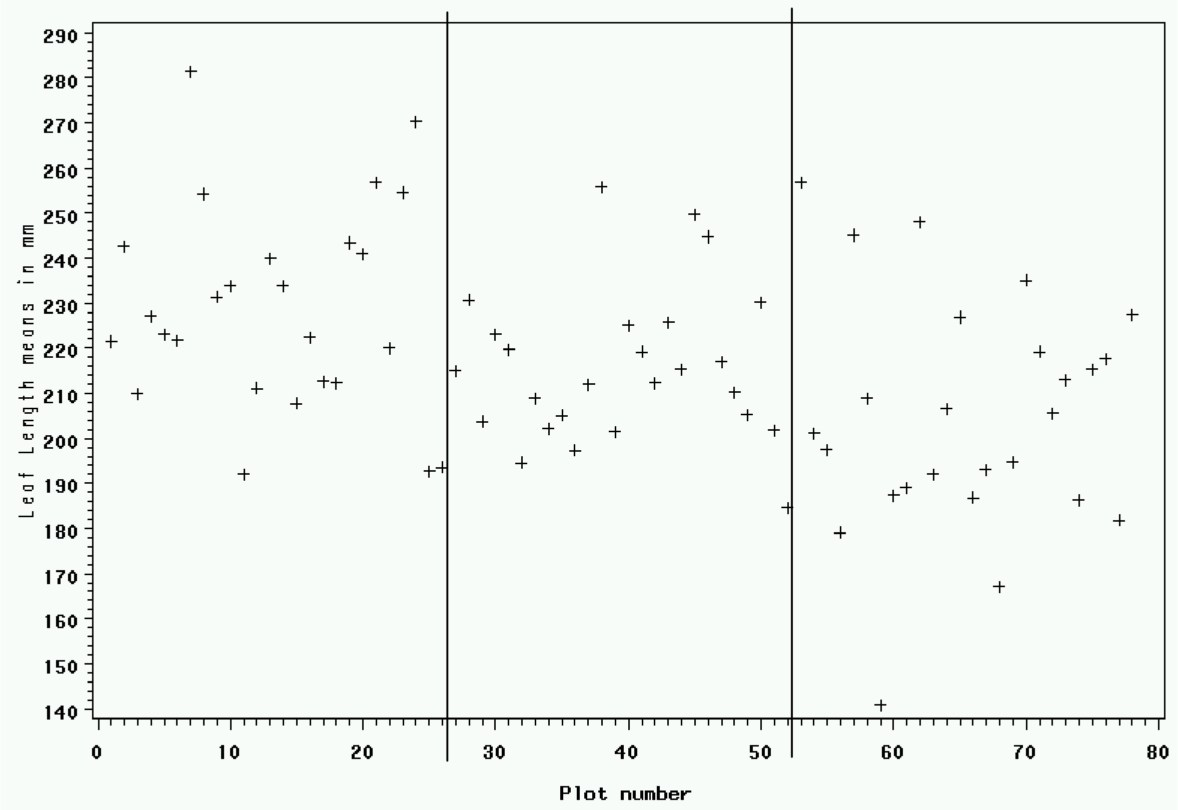


Figure 6. Plot means of 20 Leaf Lengths versus plot numbers

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.

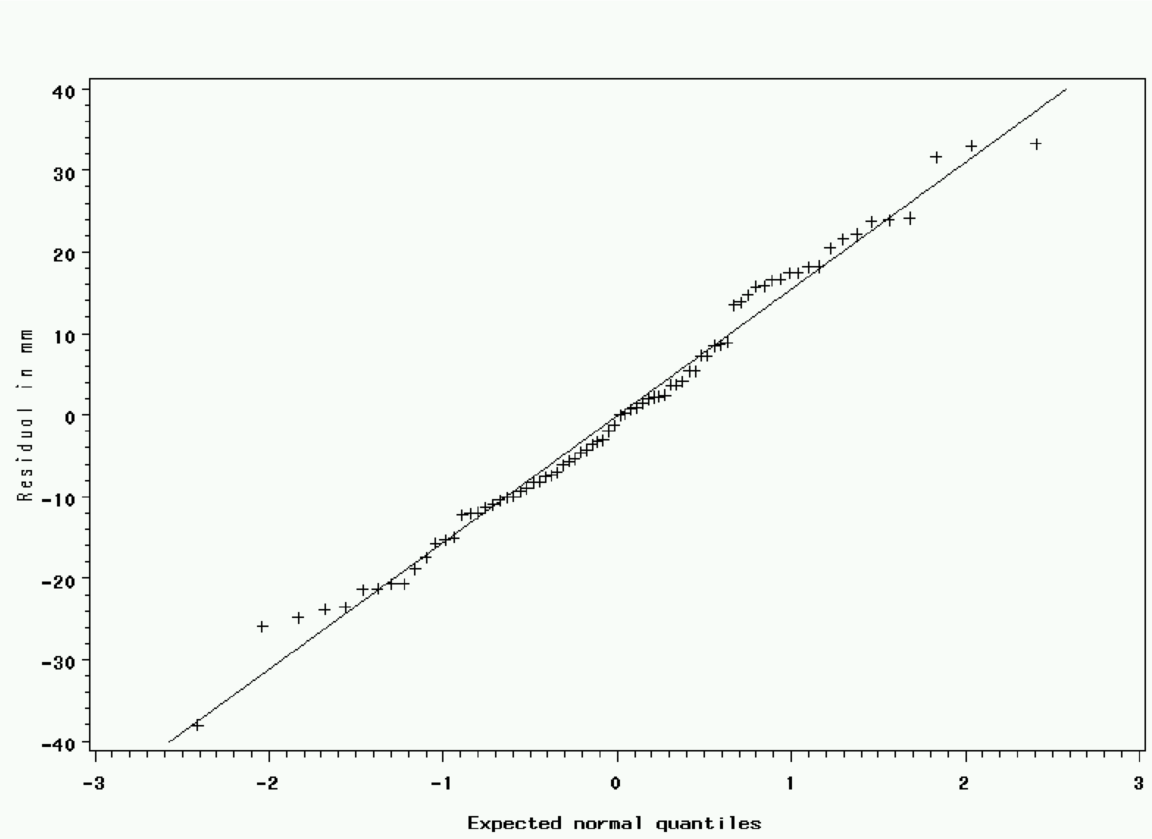


Figure 7. Normal probability plot for the residuals of Leaf Length in 26 oil seed rape varieties in 3 blocks

4. TYPES OF CHARACTERISTICS AND THEIR SCALE LEVELS

[TWC Chairperson: 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

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

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

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

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

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 for the examination of distinctness,**Error! Bookmark not defined.** 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). “Visual” observation refers to the sensory observations of the expert and, therefore, also includes smell, taste and touch.**Error! Bookmark not defined.**

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 …[iii]

For the purposes of distinctness, observations may be recorded as a single record for a group of plants or parts of plants (G), or may be recorded as records for a number of single, individual plants or parts of plants (S). In most cases, “G” provides a single record per variety and it is not possible or necessary to apply statistical methods in a plant-by-plant analysis for the assessment of distinctness.”

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]

4.7.2 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).

4.7.3 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	Distribution	Observation method	Procedure1) 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
Combination 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 Section 4.6 *[cross ref.]*
2) normally VG but VS would be possible

Table 5: Statistical procedures for the assessment of uniformity

Type of scale	Distribution	observation method	Procedure1) and Further Conditions	Reference 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
Combination 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 assess-ment	Description (states of expression)	Type of scale	Unit of assess-ment	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 follows]

---

[i] Former SECTION 1

[ii] Former SECTION 1

[iii] The TWV proposed that the definition of “S” should refer to observation of (at least) the number of single, individual plants or parts of plants recommended in Section 3.5 of the Test Guidelines and to explain that the individual plant data obtained could be used for the assessment of uniformity by statistical analysis.

[End of Part I]