E

TGP/8/1 Draft 9

PART II: **Techniques Used in DUS Examination**

ORIGINAL: English only

DATE: April 2, 2008

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

Part II: Techniques used in DUS Examination

Document prepared by the Office of the Union

to be considered by the

Technical Committee at its forty-fourth session,
to be held in Geneva, from April 7 to 9, 2008

Note for Draft version
Footnotes will be retained in published document
Endnotes are background information to assist in the consideration of this draft and will not appear in the final, published document

TABLE OF CONTENTS PAGE

PART II: TECHNIQUES USED IN DUS EXAMINATION

1. THE METHOD OF UNIFORMITY ASSESSMENT ON THE BASIS OF OFF-TYPES

Note

TWC: The Technical Committee and Technical Working Parties to be invited to consider whether it would be appropriate to make reference to suitable methods (e.g. "seedcalc") for developing tables for combinations of population standards and acceptance probabilities which were not included in section 1.1.11, in order to cover combinations used by UPOV members which did not correspond to a combination in use in UPOV Test Guidelines. Alternatively, to consider whether document TGP/8 should only provide tables for combinations which existed in UPOV Test Guidelines, in which case it was noted that document TGP/8 would need to be revised if a new combination was introduced in UPOV Test Guidelines.

1.1 Fixed Population Standard

[The TWC agreed that the results of the questionnaire in document TWC/25/18 should be reviewed with a view to incorporating guidance in this section of TGP/8]

1.1.1 Introduction

TGP/10 section 4 *[cross ref.]* provides guidance on when it would be appropriate to use the approach of uniformity assessment on the basis of off-types, using a fixed population standard. It also provides guidance on the determination of crop dependent details such as sample size and the acceptable number of off-types. This section describes the off-type approach from the following perspectives:

- Use of the off-type approach to assess uniformity in a crop.
- The issues to be considered when deciding on the crop dependent details for assessing the uniformity of a crop by the method of off-types. These details include the sample size, the acceptable number of off-types, whether to test in more than one year, and whether to use sequential testing.

1.1.2 Using the approach to assess uniformity in a crop

[TWC: to be changed to incorporate the wording from document TGP/10/1 and to use a sample size larger than 20 plants]

1.1.2.1 To use the approach to assess uniformity in a crop, the following crop dependent details are obtained, e.g. from the crop Test Guidelines:

- a sample size, e.g. 20 plants
- a maximum number of off-types to be allowed in the sample, e.g. 1
- a fixed population standard, e.g. 1%
- and an acceptance probability, e.g. at least 95%

1.1.2.2 Next, a sample of the correct size of candidate variety plants is taken and the number of off-types counted. If this number is less than or equal to the maximum allowed, the variety is accepted as uniform, otherwise it is rejected as non-uniform. In making these decisions there are two statistical errors that could be made. The risks of making these errors are controlled by the choice of sample size and the maximum allowed number of off-types.

1.1.2.3 ~~In some circumstances the Crop Expert may decide on a sample size and a maximum number of allowable off-types on the basis of experience with a similar crop, i.e. without explicitly considering the fixed population standard or the acceptance probability. In doing this the Crop Expert should be aware that although he has not explicitly considered them, a population standard and acceptance probability will be implicit in his choice of sample size and maximum number of off-types, as they will have been considered in the choice of sample size and maximum number of off-types for the similar crop.~~[a]

1.1.2.4 The fixed population standard, or "population standard", is the maximum percentage of off-types that would be permitted if all individuals of the variety could be examined. In the example above it is 1%. Varieties with less than the population standard of off-types are uniform, and those with more than the population standard are non-uniform. However, not all individuals of the variety can be examined, and a sample must be examined instead.

1.1.2.5 Consider a variety which, if all individuals of the variety were examined, would have no more than the population standard of off-types. In taking a sample there are two possible outcomes. Either the sample contains no more than the maximum allowed number of off-types, in which case the variety is accepted as uniform, or the sample contains more than the maximum allowed number of off-types and the variety is rejected. In the latter case a statistical error known as a "type I error" would have been made. The probability of accepting this variety and the probability making a type I error are linked as follows:

"probability accept" + "probability make a type I error" = 100%

1.1.2.6 The chances of accepting or rejecting a variety on the basis of a sample depend on the sample size, the maximum allowed number of off-types, and the percentage of off-types that would be found if all individuals of the variety were examined. The sample size and maximum allowed number of off-types are chosen so as to satisfy the "acceptance probability", which is the minimum probability of accepting a variety with the population standard of off-types. Thus for the example above, the sample size and maximum number of off-types have been chosen to give ~~an~~[b] at a least 95% chance of accepting a variety which, if all individuals of the variety were examined, would have 1% off-types.

1.1.2.7 To verify the sample size and maximum number of off-types in the example above, the reader should refer to table A, which lists table 10 and figure 10 as relevant for a population standard of 1% and an acceptance probability of ≥95%. Turning to table 10, the reader will see that a sample size of 20 (between 6 and 35) and a maximum number of off-types of 1 will give an acceptance probability of >95% for a population standard of 1%. Figure 10 gives more detail: the lowest of the four traces gives the probability of a type I error for the different sample sizes and maximum numbers of off-types listed in Table 10. Thus for a population standard of 1%, a sample size of 20, and allowing 0 or 1 off-types, the probability of a type I error is 2%, so the probability of accepting on the basis of such a sample a variety with the population standard, i.e. 1%, of off-types is 100% - 2% = 98%, which is greater than the "acceptance probability" (95%) as required.

1.1.2.8 It can be seen from figure 10 that as the sample size increases, the probability of a type I error increases and the probability of accepting a variety with the population standard, i.e. 1%, of off-types decreases, until this probability becomes too low to satisfy the "acceptance probability", and it becomes necessary to increase the maximum number of off-types in accordance with table 10.

1.1.2.9 Just as a variety with the population standard or fewer off-types can be either accepted or rejected (type I error) on the basis of a sample, so can a variety with more than the population standard of off-types be either accepted or rejected. To accept on the basis of a sample a variety with more than the population standard of off-types is known as a "type II error". The probability of a type II error depends on how non-uniform the variety is. The three upper traces in figure 10 give the probabilities of type II errors for three degrees of non-uniformity for the different sample sizes and maximum numbers of off-types listed in table 10. The three degrees of non-uniformity are 2, 5 and 10 times the population standard. They are represented by the top, middle and bottom of the three upper traces respectively. Thus for a sample size of 20, and allowing 0 or 1 off-types, the probability of accepting a variety with 2% off-types is 94%, that of accepting a variety with 5% off-types is 74%, and that of accepting a variety with 10% off-types is 39%. In general:

- The greater the non-uniformity, the smaller the probability of a type II error.
- For a given maximum number of off-types, as the sample size increases the probability of a type II error decreases.
- The probability of a type II error increases as the maximum number of off-types increases.

1.1.3 Issues to be considered when deciding on the use of the method

1.1.3.1 In the preceding section it has been seen that the probability of accepting a variety with the population standard or fewer off-types, or rejecting it (type I error), and the probability of accepting a variety with more than the population standard of off-types (type II error) or rejecting it all depend on the choice of sample size and maximum allowed number of off-types. The remainder of this chapter is a discussion of how these choices can be used to balance the risks of type I and type II errors. This will be illustrated through a series of examples. The discussion is extended to include the situation where the test is carried out over more than one year, including the possibility of using sequential testing to minimise sampling effort. The reader is provided with tables and figures from which to obtain the type I and type II error probabilities for different combinations of population standard and acceptance probability. The reader is also given details of how to calculate the probabilities directly, both for single year tests and for two or more year tests, including two-stage testing.

1.1.3.2 The two types of error described above can be summarised in the following table:

	Decision based on number of off-types in a sample	
Decision that would be made if all plants of a variety could be examined	Variety is accepted as uniform	Variety is rejected as non-uniform
Variety is Uniform	Same decision	Different decision, type I error
Variety is not uniform	Different decision, type II error	Same decision

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

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

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

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

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

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

1.1.4 Examples

Example 1

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

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

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

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

---

[*] See paragraph 54

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

Example 2

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

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

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

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

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

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

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

Example 3

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

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

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

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

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

1.1.4.14 The following schemes can be obtained:

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

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

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

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

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

1.1.4.17 Alternatively, one of schemes a and h may be chosen. However, scheme g seems to have a too large probability of type II errors compared with the probability of type I error. For example, there is a 1% probability of rejecting a uniform variety (type I error) and a 27% probability of accepting a variety with 5% off-types (type II error).

1.1.4.18 Scheme h has the advantage of often allowing a final decision to be taken after the first test (year) but, as a consequence, there is a higher probability of a type I error. In this case, there is a 10% probability of rejecting a uniform variety (type I error) and a 9% probability of accepting a variety with 5% off-types (type II error).

Example 4

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

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

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

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

1.1.5 Introduction to the tables and figures

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

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

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

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

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

1.1.5.4 When using the tables the following procedure is suggested:

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

(a) Choose the relevant population standard.

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

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

1.1.6 Detailed description of the method for one single test

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

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

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

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

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

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

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

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

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

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

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

1.1.7 More than one single test (year)

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

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

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

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

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

1.1.8.1 Combined test

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

1.1.8.2 Two-stage test

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

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

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

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

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

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

where

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

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

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

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

1.1.8.3 Sequential tests

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

1.1.9 Note on balancing the type I and type II errors

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

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

1.1.10 Definition of statistical terms and symbols

The statistical terms and symbols used have the following definitions:

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

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

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

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

P Population standard

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

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

1.1.11 Tables and figures

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

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

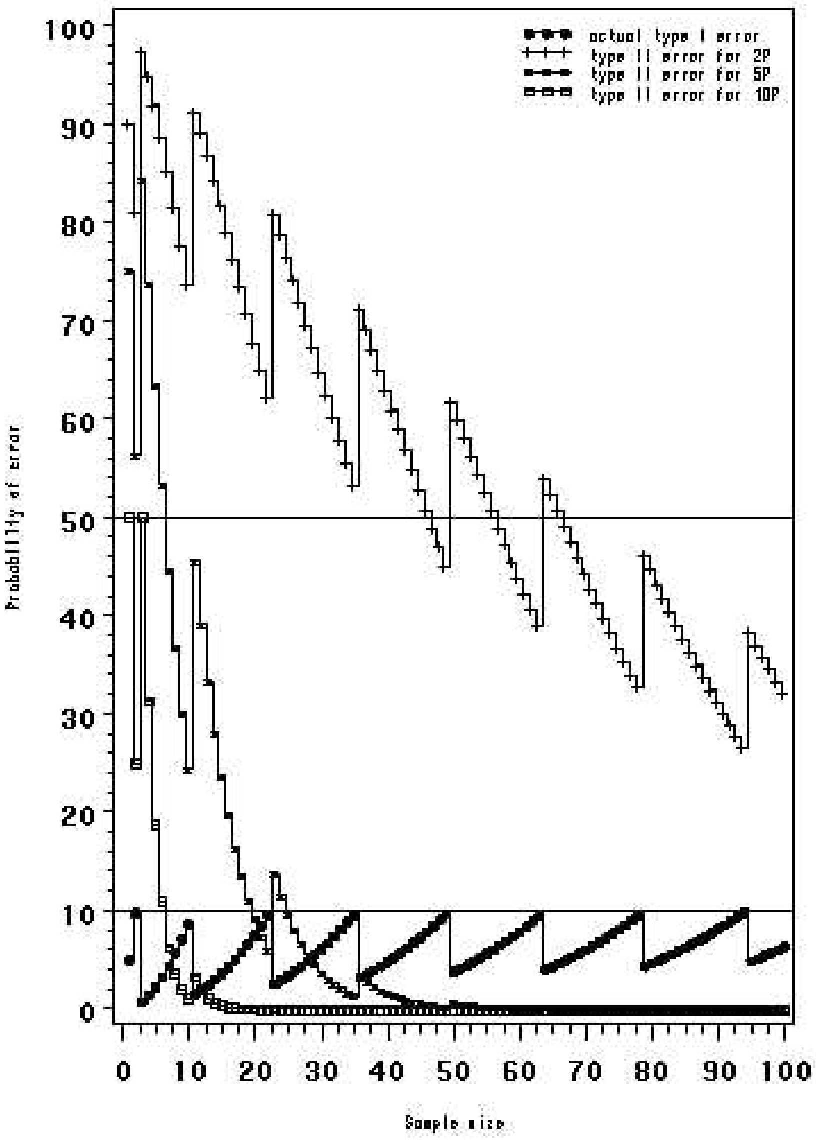


Table and figure 2:

Population Standard = 3%
Acceptance Probability ≥90%
n=sample size, k=maximum number of off-types

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

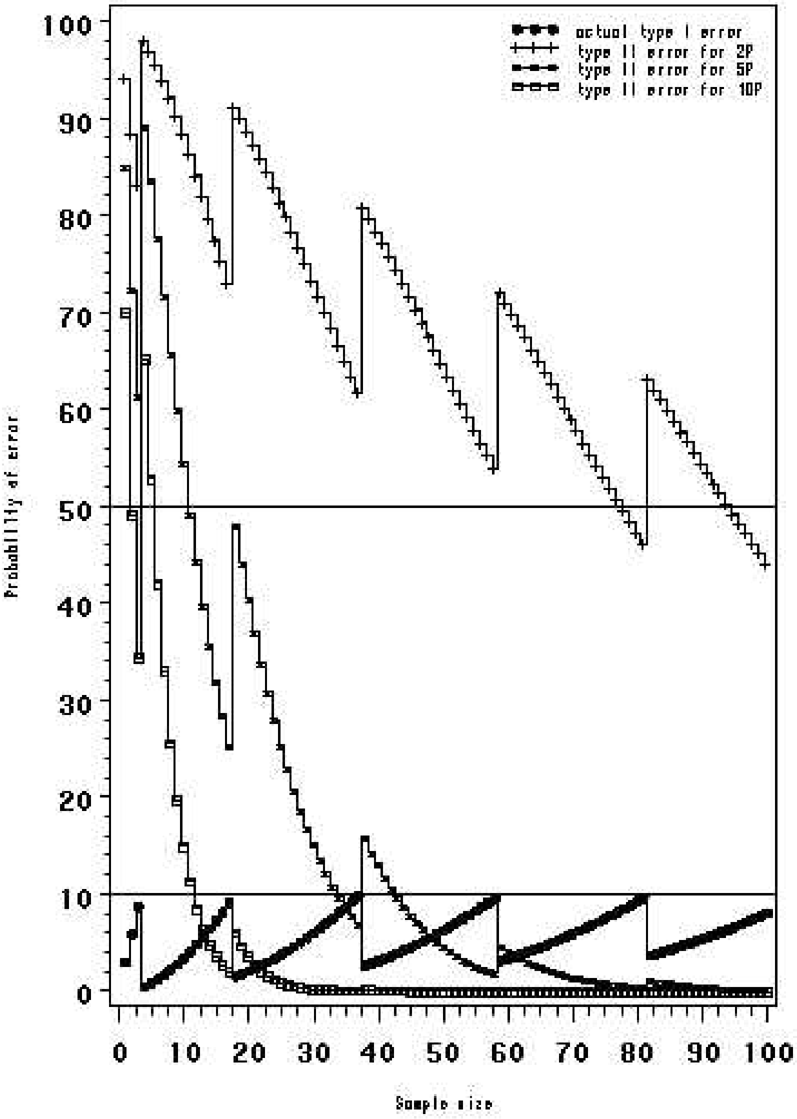


Table and figure 3: **Population Standard = 2%**
Acceptance Probability ≥90%
n=sample size, k=maximum number of off-types

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

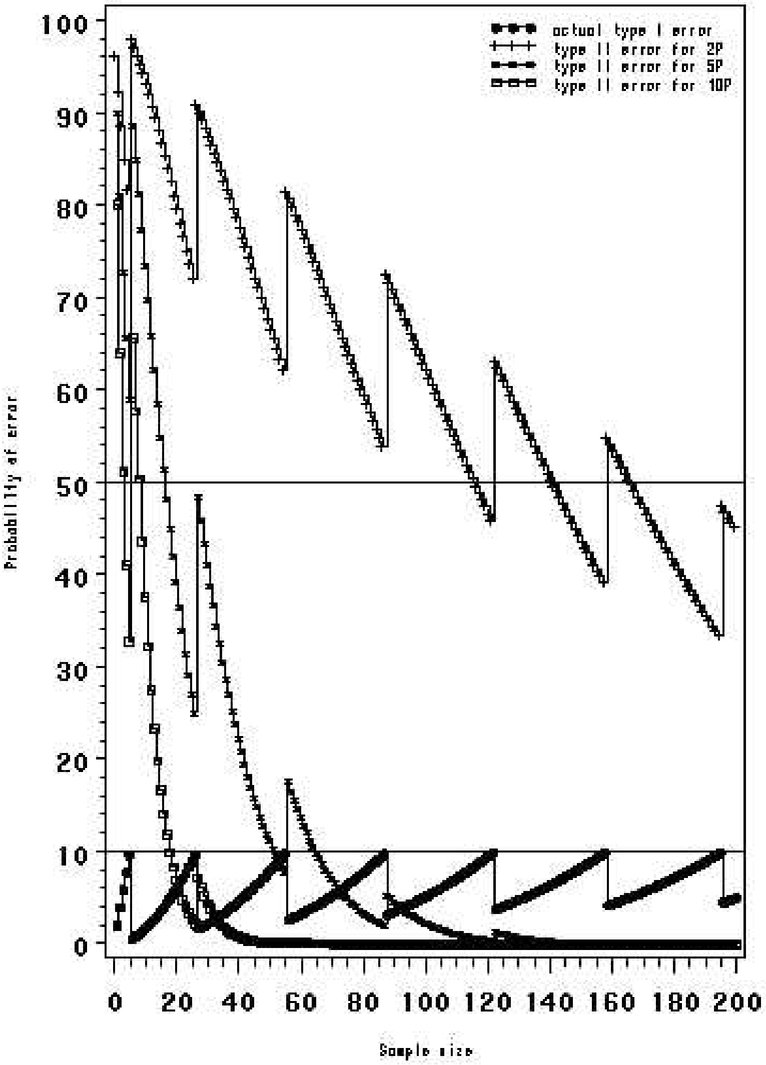


Table and figure 4: **Population Standard = 1%**
Acceptance Probability ≥90%
n=sample size, k=maximum number of off-types

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

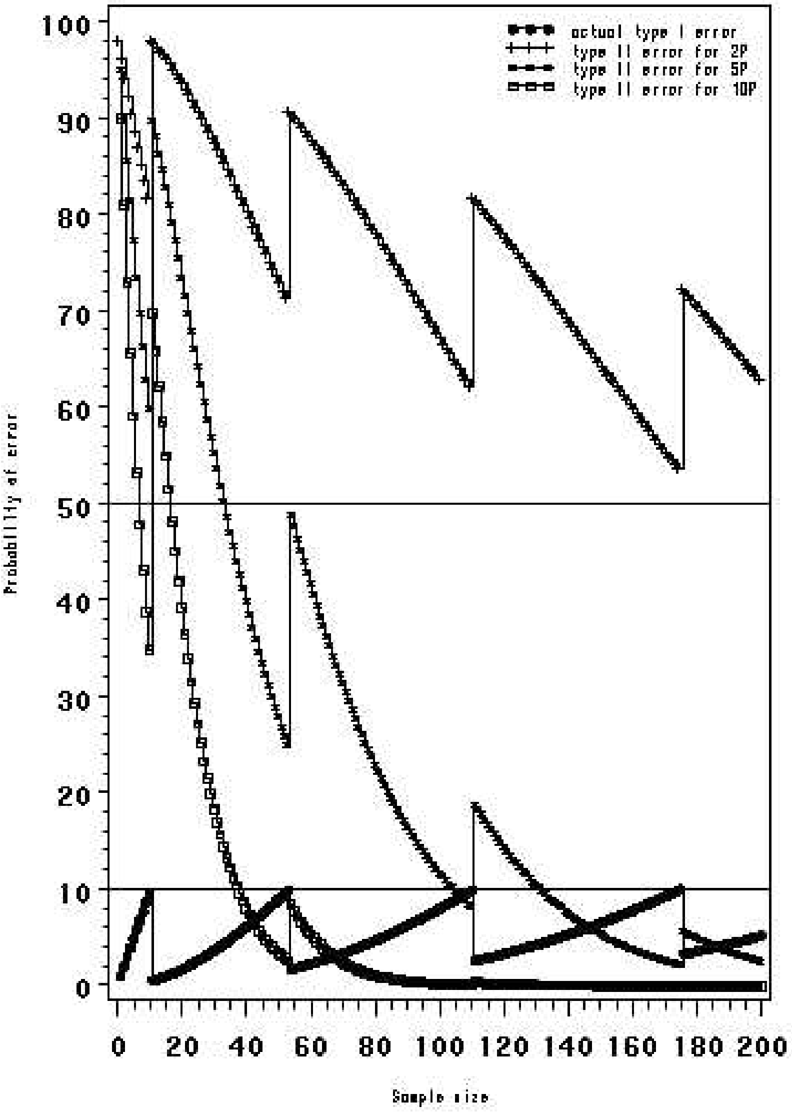


Table and figure 5: **Population Standard = .5%**
Acceptance Probability ≥90%
n=sample size, k=maximum number of off-types

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

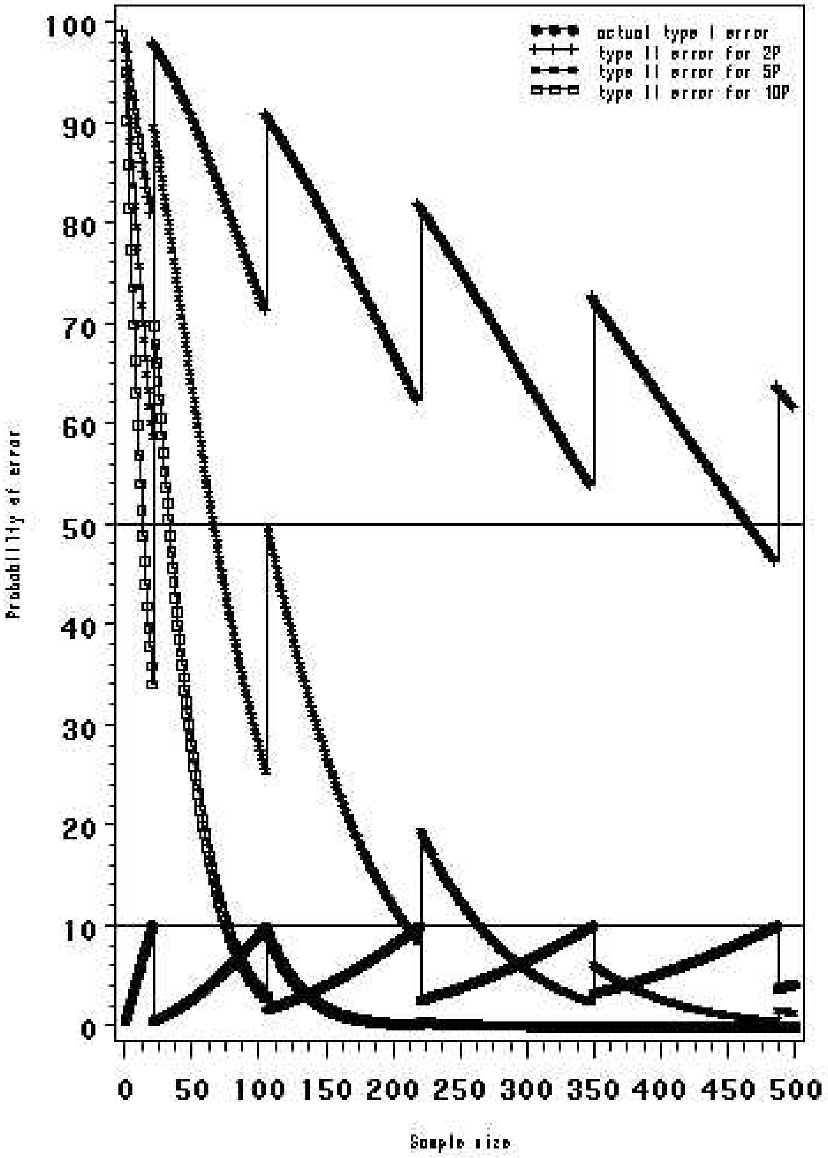


Table and figure 6: **Population Standard = .1%**
Acceptance Probability ≥90%
n=sample size, k=maximum number of off-types

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

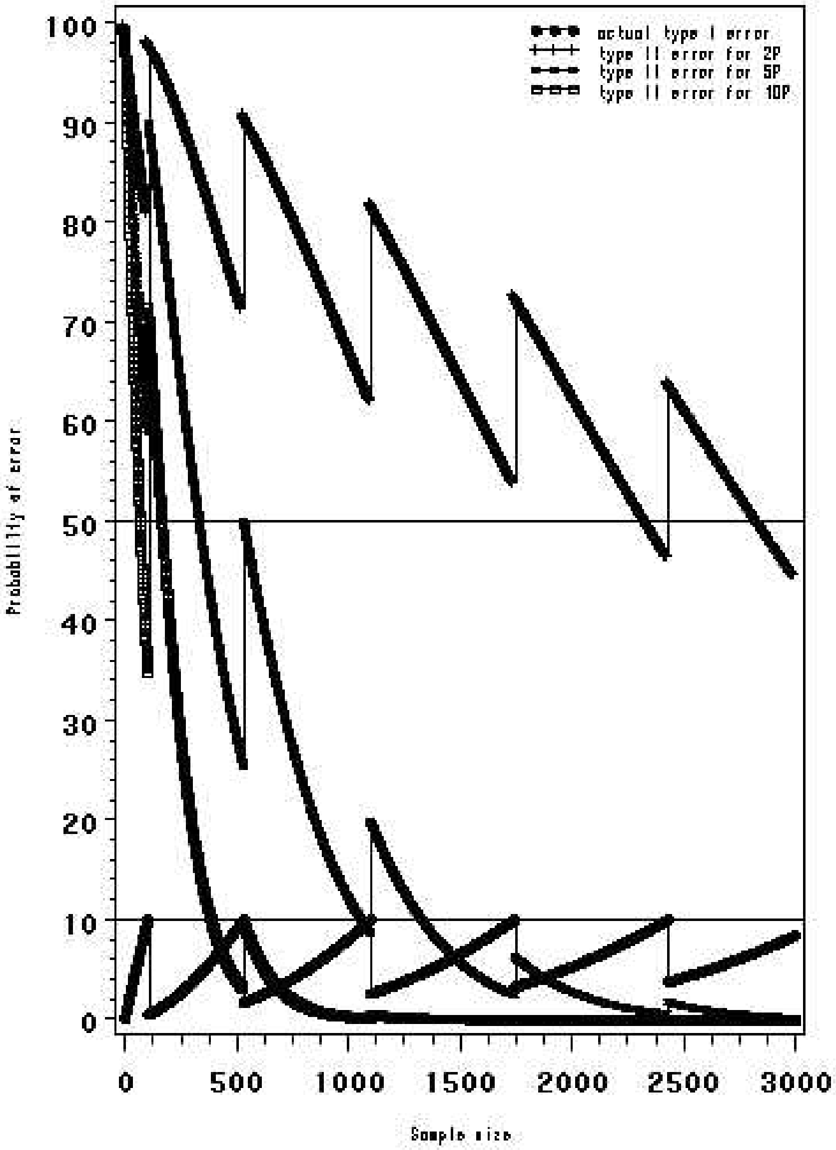


Table and figure 7: **Population Standard = 5%**
Acceptance Probability ≥95%
n=sample size, k=maximum number of off-types

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

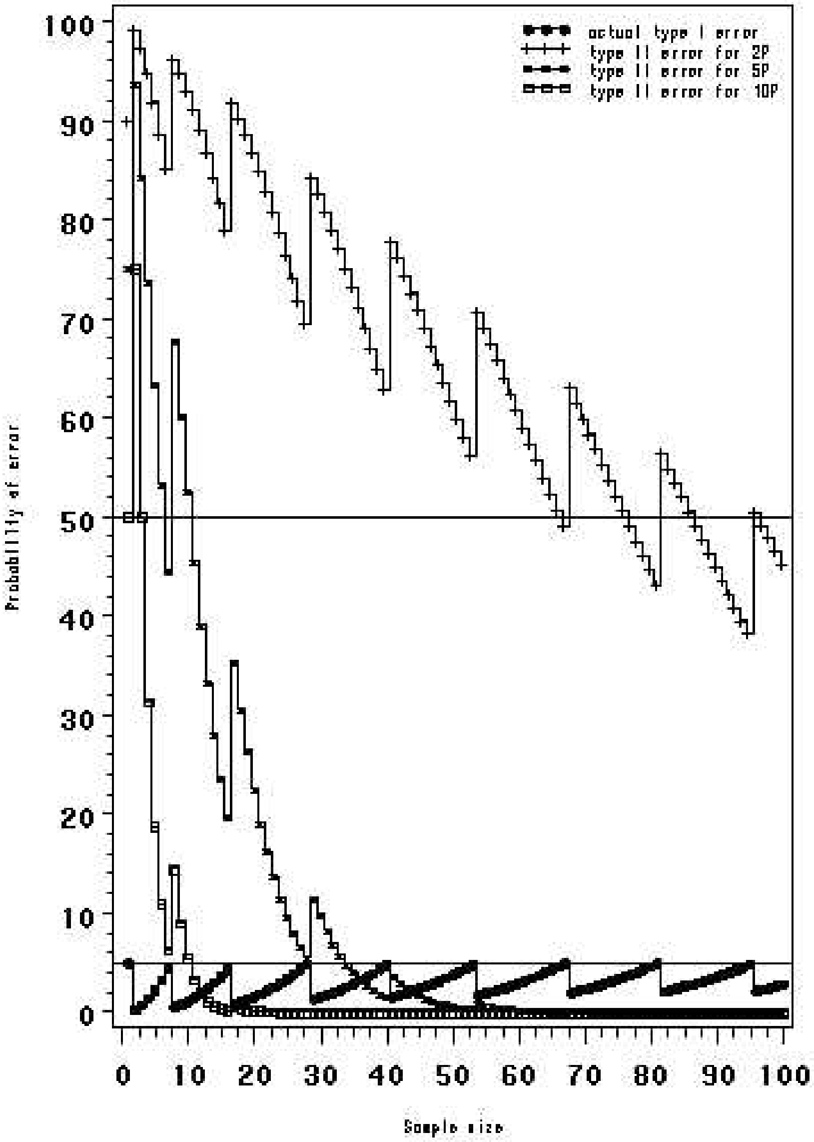


Table and figure 8: **Population Standard = 3%**
Acceptance Probability ≥95%
n=sample size, k=maximum number of off-types

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

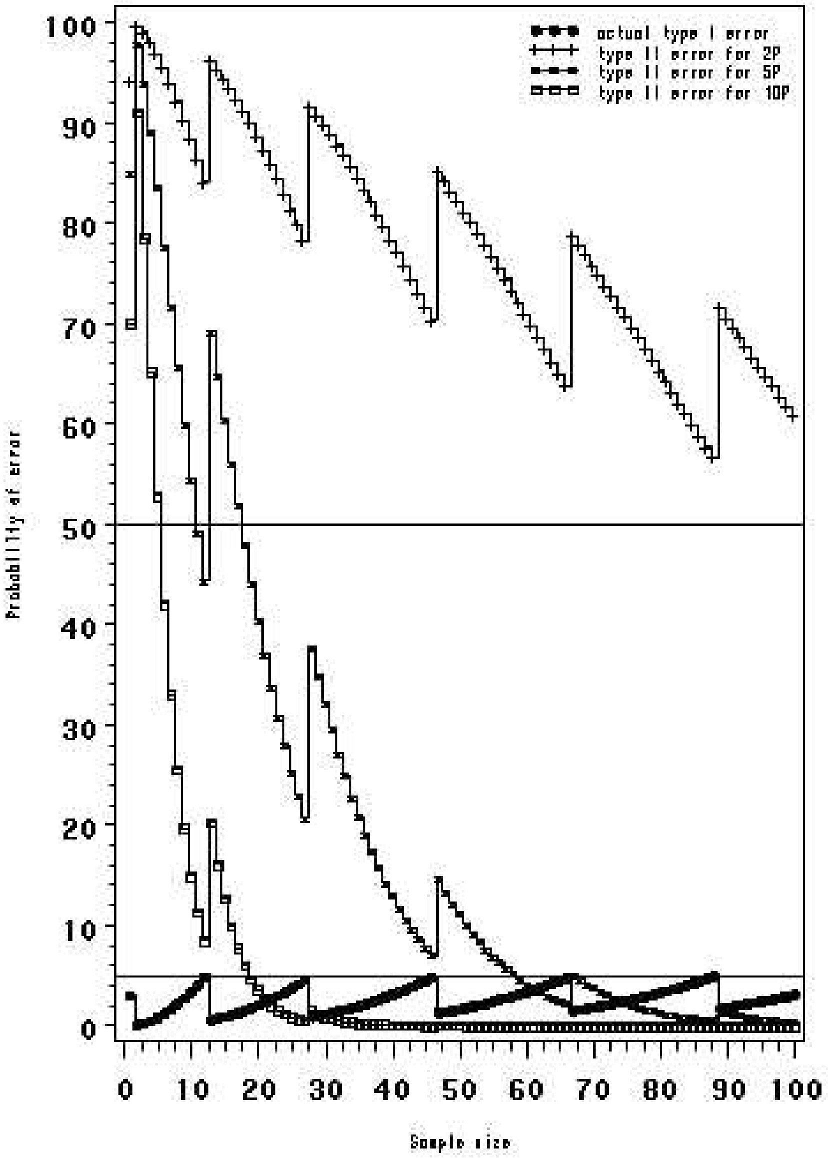


Table and figure 9: **Population Standard = 2%**
Acceptance Probability ≥95%
n=sample size, k=maximum number of off-types

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

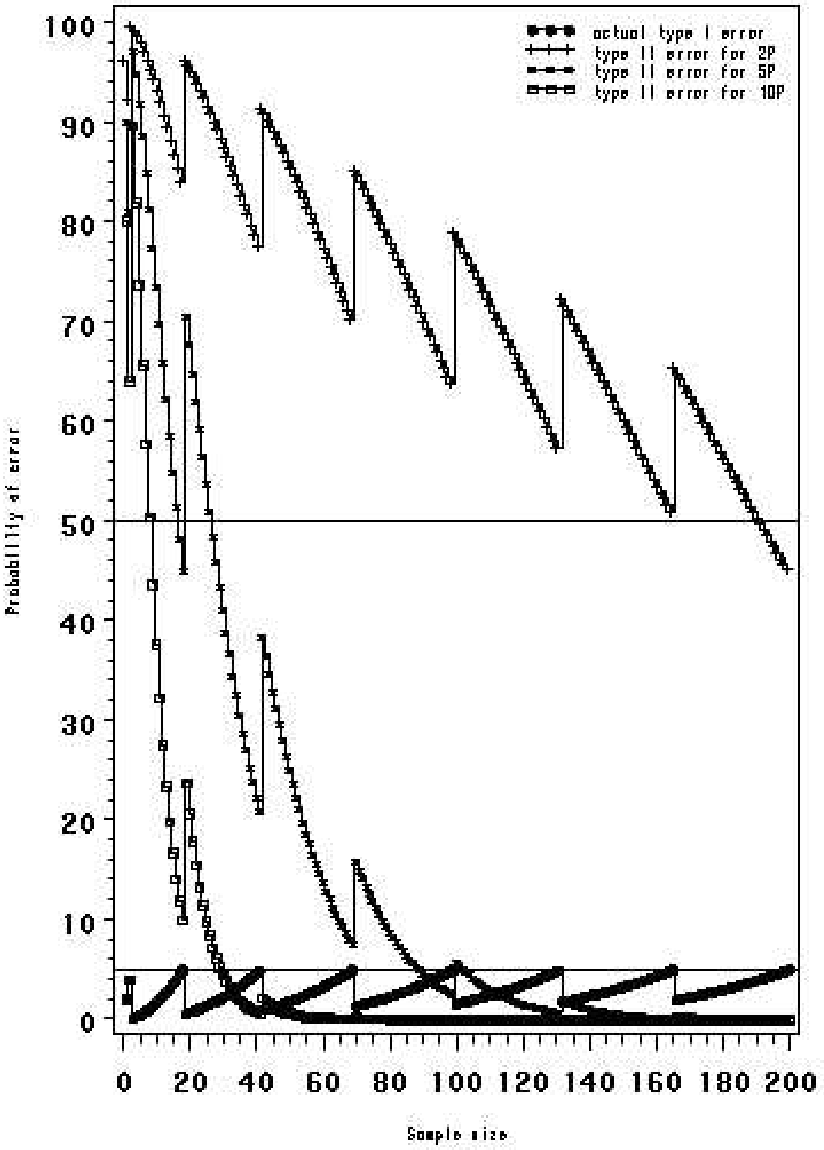


Table and figure 10: **Population Standard = 1%**
Acceptance Probability ≥95%
n=sample size, k=maximum number of off-types

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

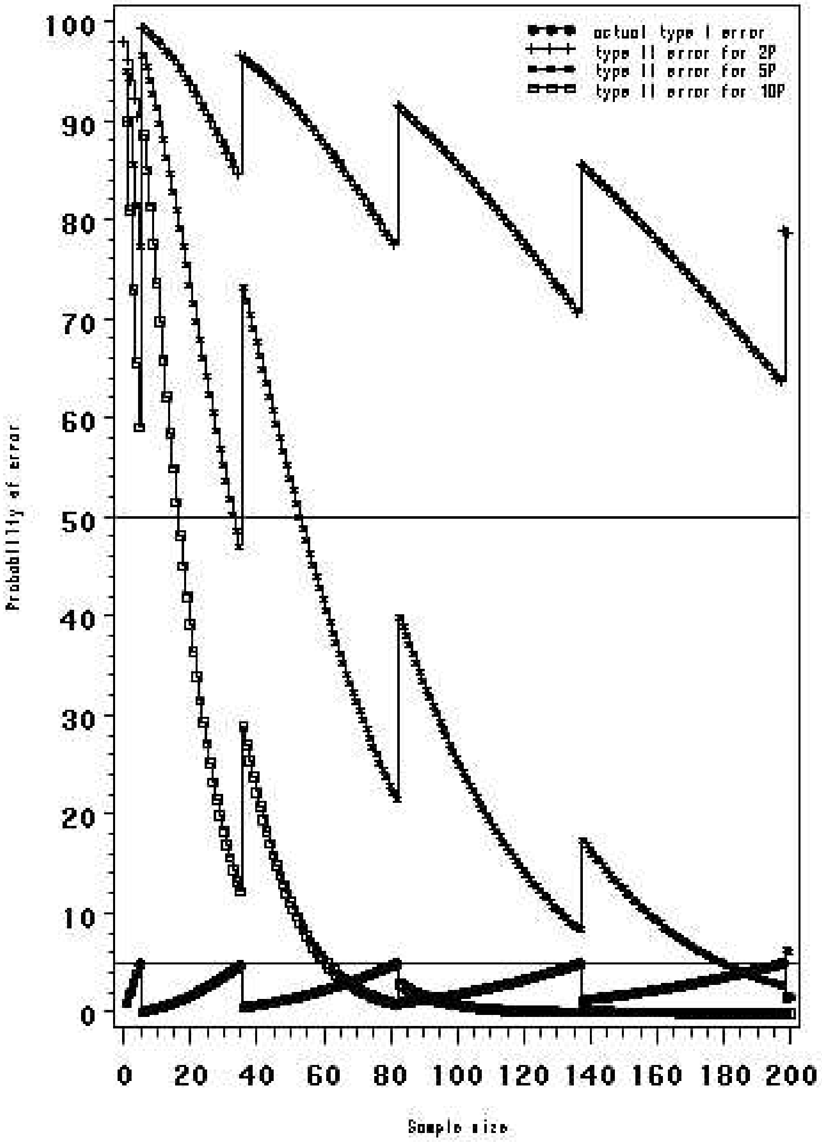


Table and figure 11: **Population Standard = .5%**
Acceptance Probability ≥95%
n=sample size, k=maximum number of off-types

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

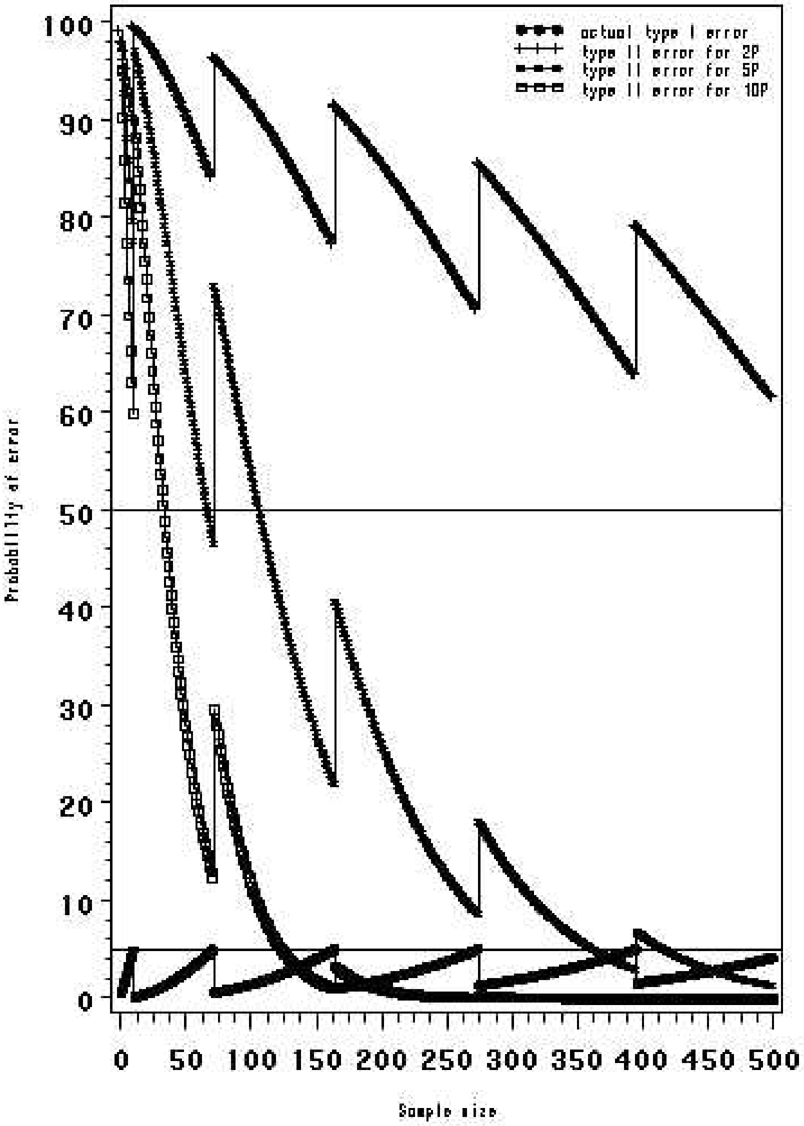


Table and figure 12: **Population Standard = .1%**
Acceptance Probability ≥95%
n=sample size, k=maximum number off-types

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

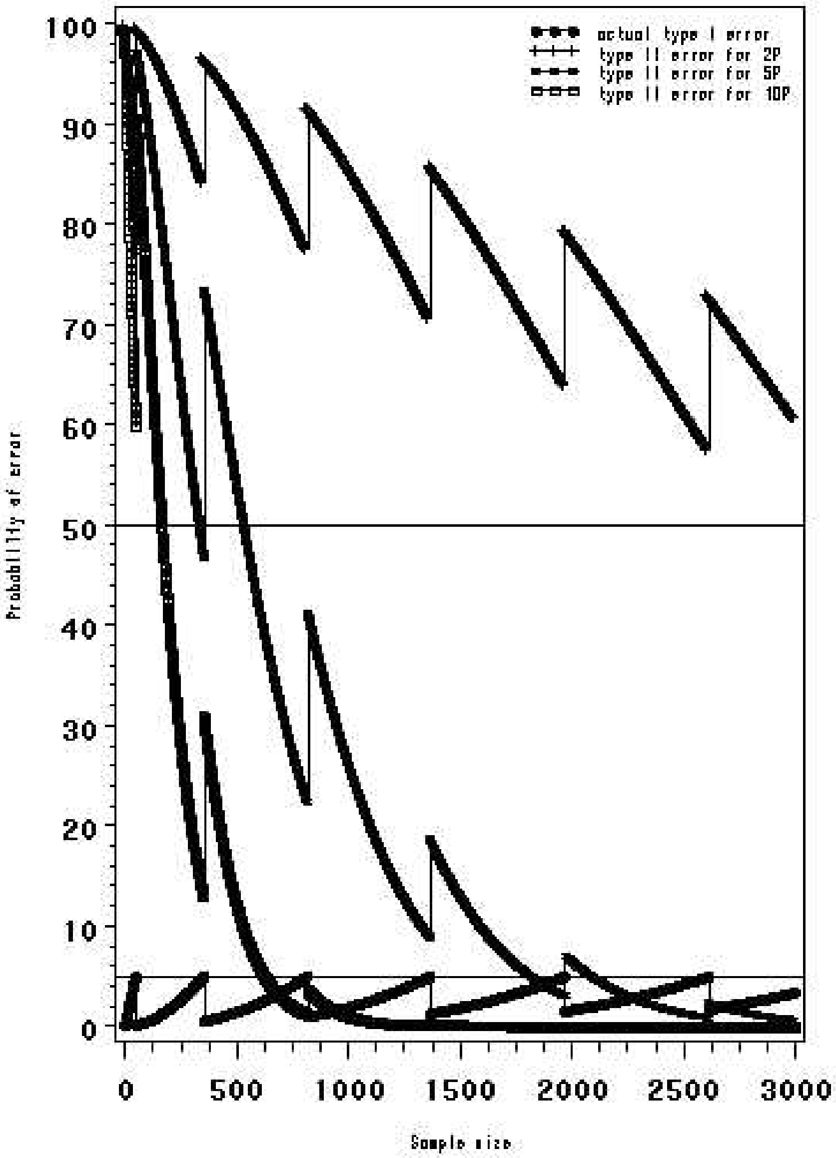


Table and figure 13: **Population Standard = 5%**
Acceptance Probability ≥99%
n=sample size, k=maximum number of off-types

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

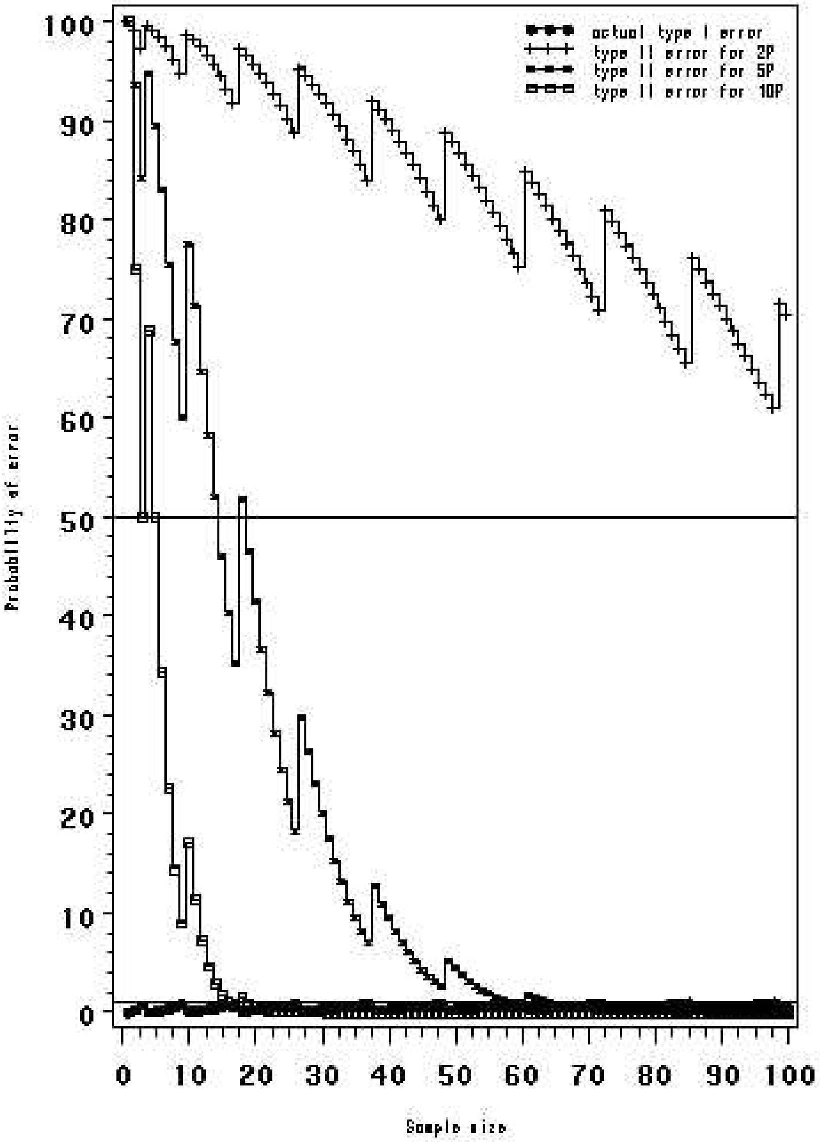


Table and figure 14: **Population Standard = 3%**
Acceptance Probability ≥99%
n=sample size, k=maximum number of off-types

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

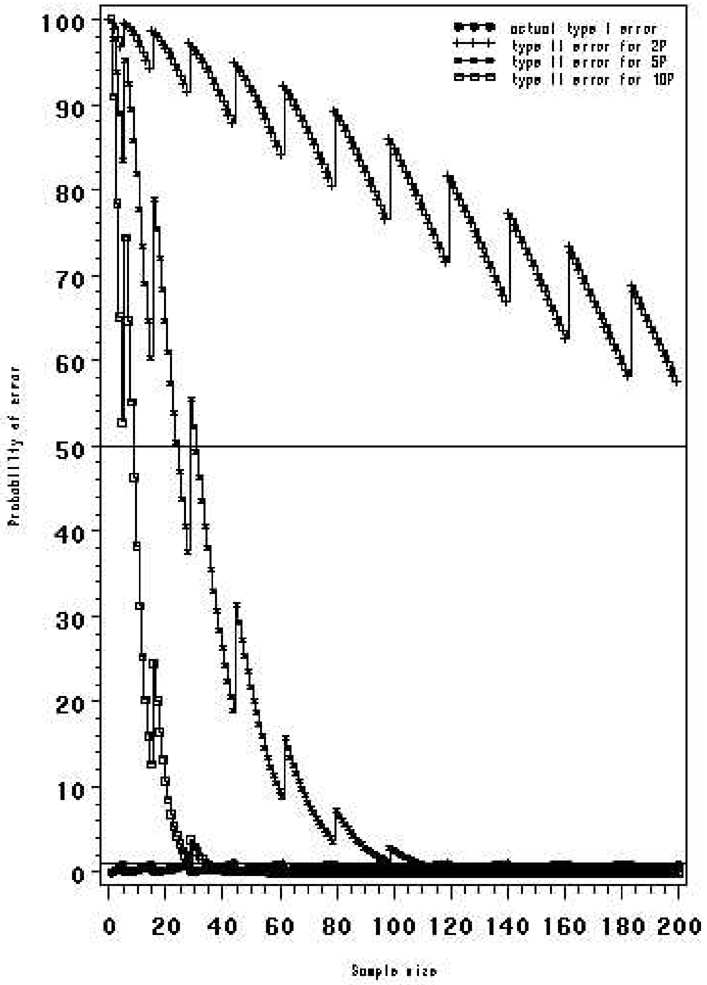


Table and figure 15: **Population Standard = 2%**
Acceptance Probability ≥99%
n=sample size, k=maximum number of off-types

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

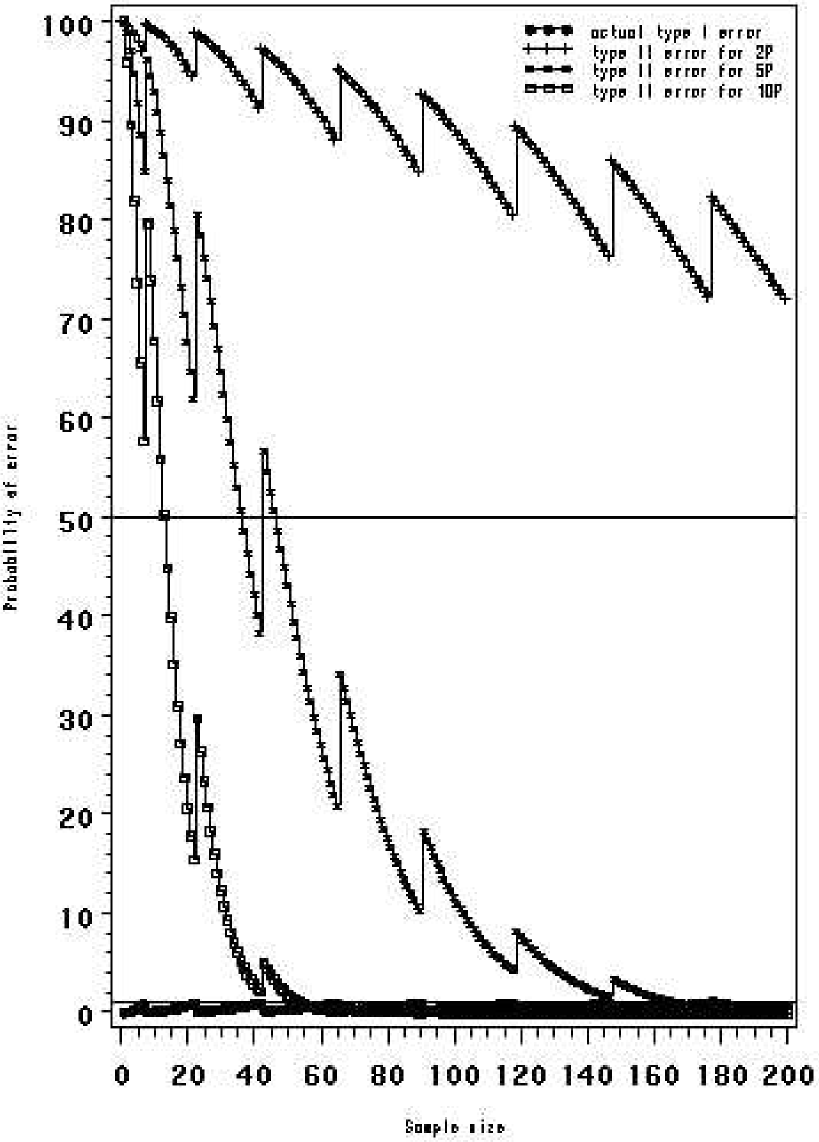


Table and figure 16: **Population Standard = 1%**
Acceptance Probability ≥99%
n=sample size, k=maximum number of off-types

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

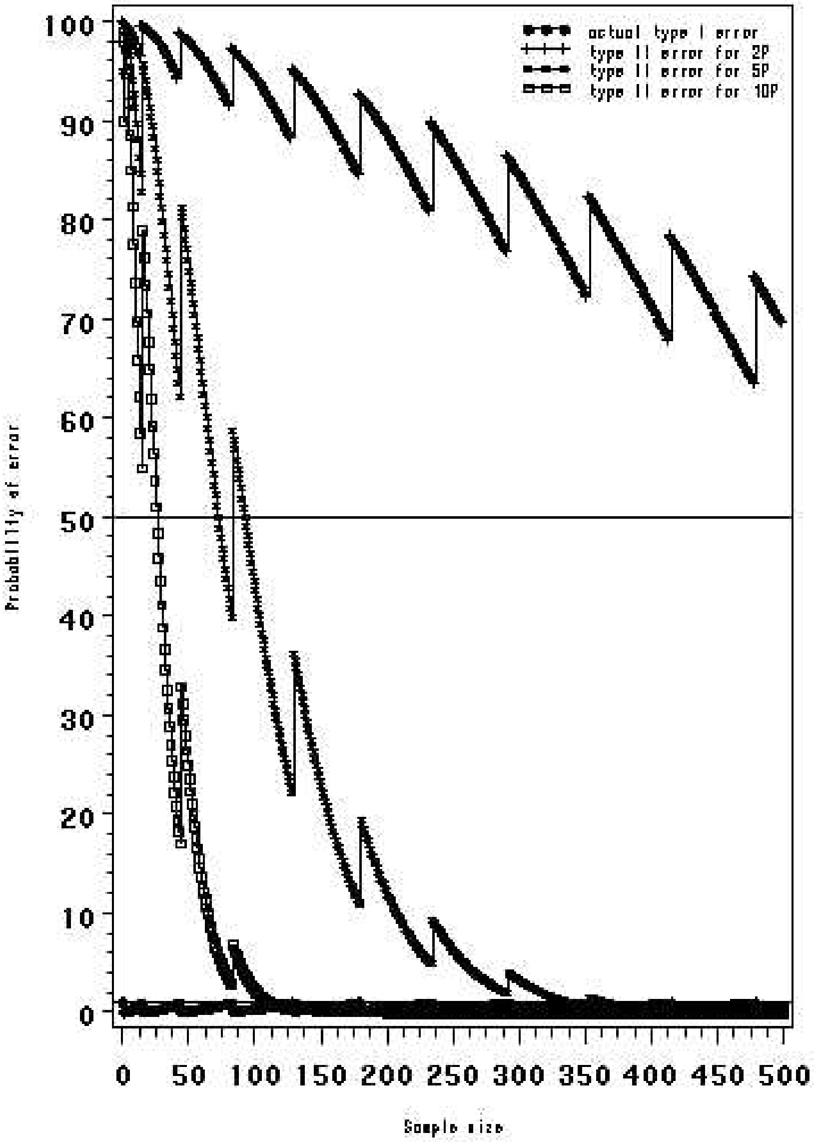


Table and figure 17: **Population Standard = .5%**
Acceptance Probability ≥99%
n=sample size, k=maximum number of off-types

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

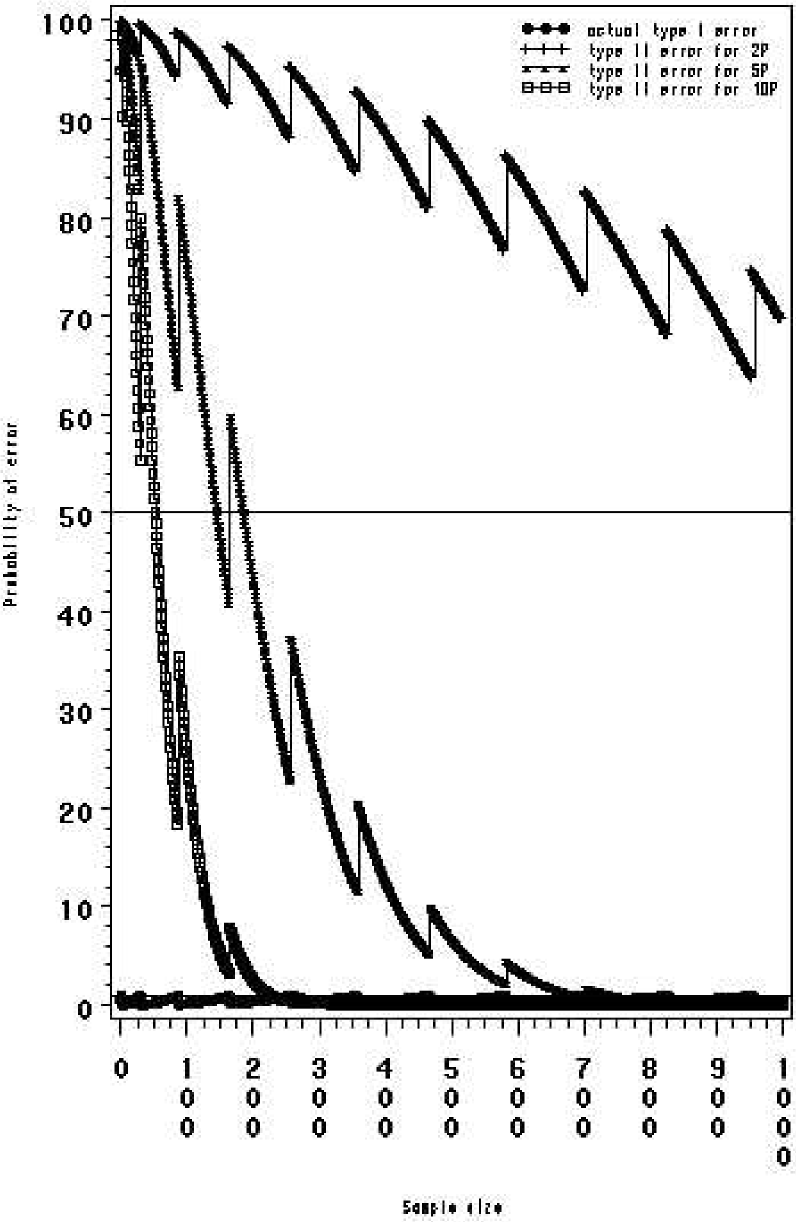


Table and figure 18: **Population Standard = .1%**
Acceptance Probability ≥99%
n=sample size, k=maximum number of off-types

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

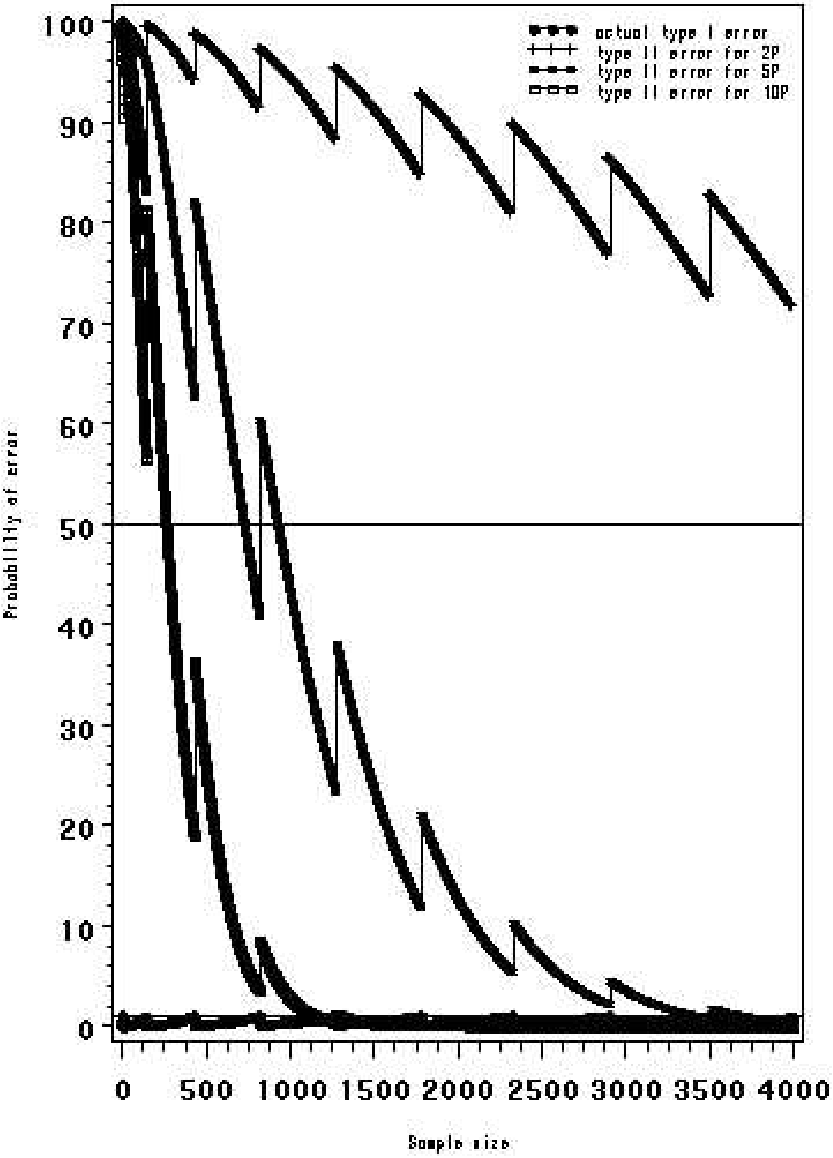


Table and figure 19: **Population Standard = 10%**
Acceptance Probability ≥ 90%
n=sample size, k=maximum number of off-types

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

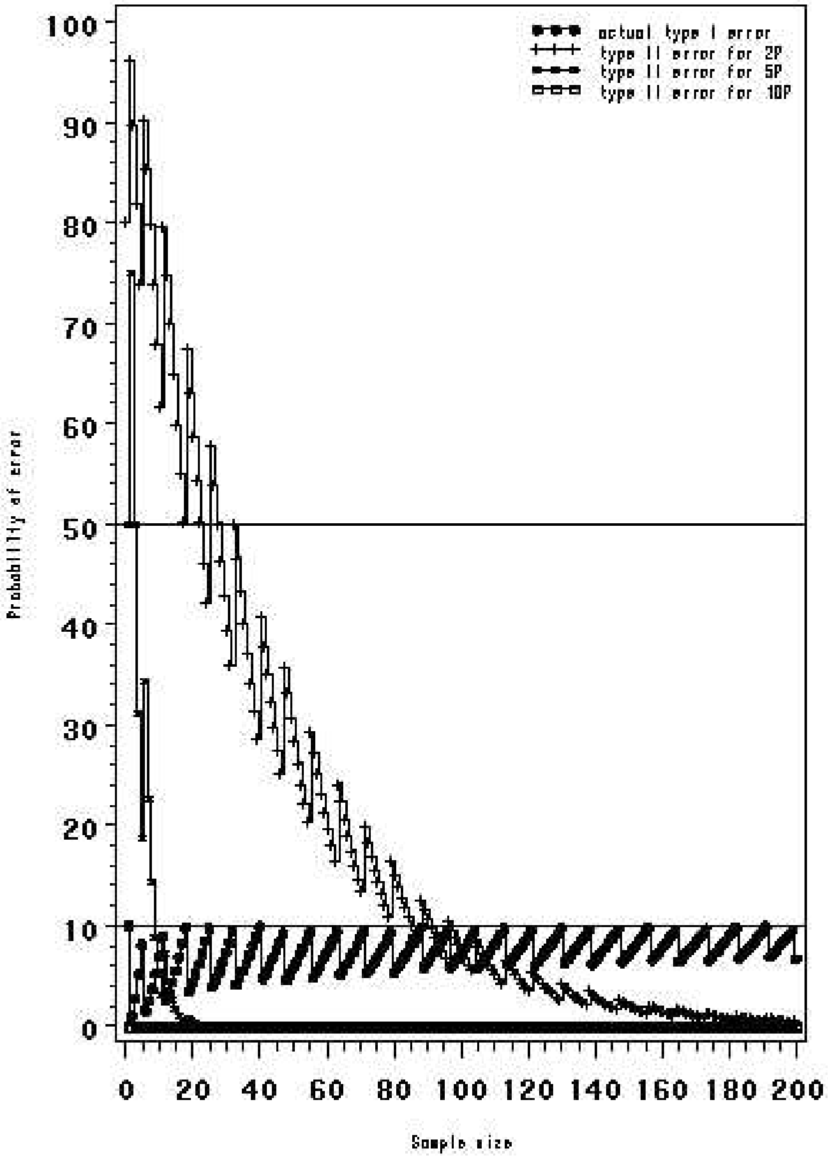


Table and figure 20: **Population Standard = 10%**
Acceptance Probability ≥ 95%
n=sample size, k=maximum number of off-types

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

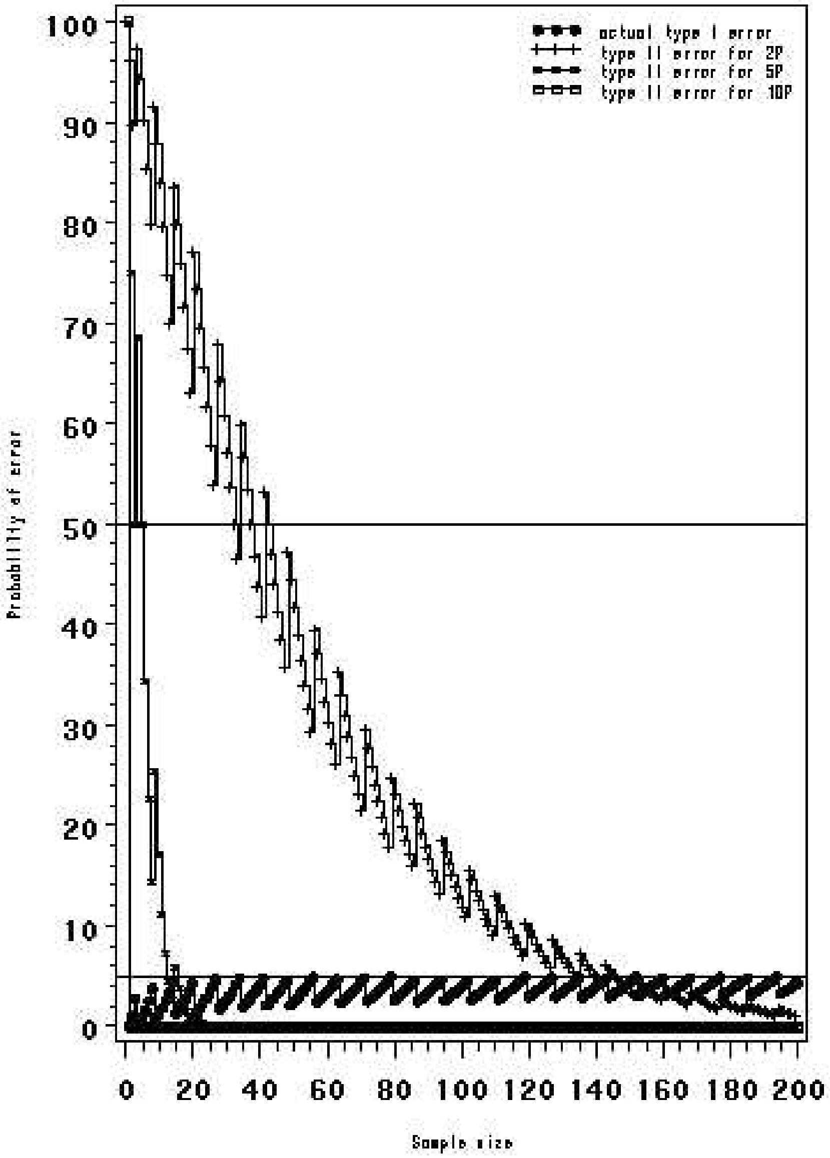
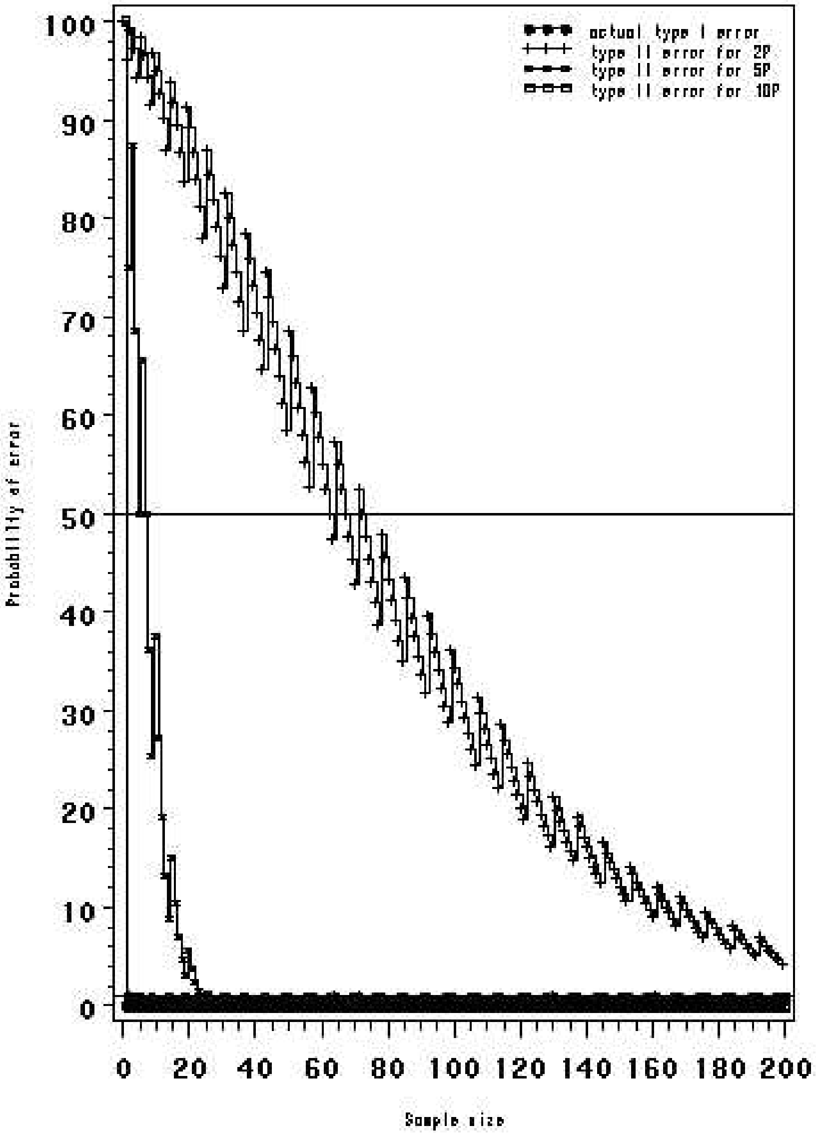


Table and figure 21 : **Population Standard = 10%**
Acceptance Probability ≥ 99%
n=sample size, k=maximum number of off-types

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



2. THE CHOICE OF STATISTICAL METHODS FOR THE ASSESSMENT OF DISTINCTNESS

2.1 Introduction

2.1.1 This section addresses some general considerations when choosing suitable statistical methods for the assessment of distinctness. It contains a discussion of factors influencing the choice of method, and, as the statistical test used by each method is an essential part of that method, it includes a brief discussion of statistical tests, factors influencing their selection and some comments on their usefulness in particular situations.

2.1.2 Statistical methods are most commonly used for the assessment of distinctness of measured quantitative characteristics for cross-pollinated varieties when the data from the growing trial for a variety are subject to variation. Because of this variation, distinctness criteria based on statistical methods are needed in order to separate genuine varietal differences from chance variation and so make decisions about whether the candidate variety is distinct with a certain level of confidence that the decision is the correct one.

2.1.3 The variation may occur for example from plant to plant, from plot to plot and from year to year. Whether a single growing cycle or more than a single growing cycle is needed to provide assurance that the differences observed between varieties are sufficiently consistent will depend on the levels or amounts of variation from these different sources that are observed in a species. Section 1.2 of PART I. provides information on growing cycles.

2.2 Statistical methods for use with two or more independent growing cycles

2.2.1 Introduction

2.2.1.1 A number of different statistical methods have been developed to assess distinctness when there are at least two independent growing cycles. The choice of which method to use depends partly on the species and partly on whether the trial and data requirements for the different statistical methods are met. Where those requirements are not met, such as where only one, or very few, known varieties exist for a taxon, and so a large trial is not possible, then other suitable approaches might be used.

2.2.1.2 The principles common to suitable statistical methods used to assess distinctness when there are at least two independent growing cycles include:

- statistical tests of the differences between variety means are used to determine whether the differences between varieties in the expression of their characteristics are significant.
- a requirement for the differences to be consistent across the different growing cycles. This requirement may be part of the statistical test as in the COYD method, or not part of the statistical test as in the 2x1% and Match methods.

For the sake of brevity in the following the term ‘year’ is used, though for these purposes it is interchangeable with the term ‘independent growing cycle’.

2.2.1.3 Examples of suitable statistical methods include:-

(a) The COYD and long-term COYD methods to assess distinctness, which have been developed by UPOV to analyse data from two or more years of growing trials where there are either at least a certain minimum number of varieties in trial or data from sufficient trials in earlier years. Whether differences are sufficiently consistent is assessed using a statistical test based on a two-tailed LSD to assess whether differences in over-year variety means are significant. Details of the COYD and long-term COYD methods and the requirements for their use are given in TGP/8 Part II section 3.

(b) The 2x1% method to assess distinctness, which has also been developed by UPOV to analyse data from two or more years of growing trials. Unlike the COYD methods, this method has no particular trial size requirements. Differences are assessed in each year using a statistical test based on a two-tailed LSD to compare the within-year variety means. Whether differences are sufficiently consistent is determined by the requirement that two varieties are significantly different in the same direction at the 1% level in both years, or, where trials are conducted in three years, in at least two out of three years. Details of the 2x1% method and how it compares with the COYD method are given in TGP/8 Part II section 4.

(c) The Match method to assess distinctness was developed by Australia for use where the trials are conducted by the breeder in the first year and examined by the testing authority in the second year (see TGP/6 section 2/1). They typically involve relatively small scale trials. The number of candidate and reference varieties in the trial is limited to the most similar varieties of common knowledge by, inter alia, using grouping characteristics from the relevant UPOV Test Guidelines. Whether differences are sufficiently consistent is assessed using a statistical test to gauge whether the within-year variety mean differences in the second year are significant and agree with the “direction of the differences” declared by the breeders in the first year. Thus the statistical test may be based on a one-tailed LSD, if there is one candidate, or on a Multiple Range Test, if there is more than one candidate included in the growing trial. Although these tests are most useful in trials of cross pollinated varieties, they can be similarly applied to trials of self-pollinated and vegetatively propagated varieties provided the relevant criteria are met. An example of the Match method is given in TGP/8 Part II section 2.2 [example to be taken from TWC/25/9 rev & TWC/25/11 on LSD & MRT’s: example may need to be expanded to include the breeder side of the test].

The above methods use different statistical tests to assess whether differences between variety means are significant. The choice of the statistical test that is used has implications for the risks to the breeder and the tester of making statistical errors and is discussed below.

2.2.1.4 The relative discriminating power of two statistical methods used to assess distinctness may be compared by applying them to the same data sets for a number of tests. This may be done retrospectively. It also allows the significance levels of the statistical tests to be adjusted to give as near equivalence as is possible in terms of the resulting decisions. For example, this would be done when it is necessary to change the statistical method used to assess distinctness.

2.2.1.5 The COYD and 2x1% statistical methods have been compared using this approach. The Match method has not yet been compared with other methods. The main advantages and disadvantages of the above statistical methods are summarised in the following table. For greater detail see the section relevant to the statistical method in TGP/8 Part II.

Advantages	**Disadvantages**
COYD method	
• combines information over years in a simple and efficient way • decisions based on it are likely to be reproducible in other years • the risks of wrongly declaring distinctness are constant for all characteristics	• there is a requirement on the number of varieties in trial
Long-term COYD method	
(same as for COYD method) • there is no requirement on the number of varieties in trial	• there is a requirement on data from trials in earlier years
2x1% method	
• there is no requirement on the number of varieties in trial	• does not take account of consistency in the size of differences from year to year • inefficient use of information as decision is accumulated over years
Match method	
• can be used where the trials are conducted by both breeder and testing authority • used with small trials	• does not take account of consistency in the size of differences from year to year

2.3 Statistical methods for use with one growing cycle

2.3.1 As indicated in the General Introduction, when the level of variation within varieties is relatively low the assessment of distinctness may be based on a single growing cycle. In this case the assessment of distinctness of measured quantitative characteristics may use a statistical method based on a statistical test, such as for example an LSD in the case of self-pollinated and vegetatively propagated varieties.

2.3.2 The principle of the statistical method used to assess distinctness when there is a single growing cycle is that the statistical test of variety means is used to determine whether the differences between varieties in the expression of their characteristics are significant in that single growing cycle.

2.3.3 Example of suitable statistical method:-

Single growing cycle method.
[TWP's are invited to provide information on this method.]
In the absence of information on this method it is thought likely that in using this method differences are assessed using a statistical test based on a two-tailed LSD to compare the variety means. Details of the Single Growing Cycle method, when provided, will be found in TGP/8 Part II section 5.

2.4 Statistical tests central to the statistical methods

A number of different statistical methods have been developed to assess distinctness. These methods use statistical tests to assess whether differences between variety means are significant. The choice of the statistical test that is used by a statistical method has implications in terms of risks or chances of making statistical errors. This section describes two statistical tests that are commonly used. These are the Least Significant Difference and the Multiple Range Test.

2.4.1 The Least Significant Difference (LSD)

2.4.1.1 The Least Significant Difference (LSD) is a statistic used to compare variety means from analysis of variance (ANOVA) of a characteristic and to make decisions about whether the varieties are significantly different from each other in that characteristic. In other words it represents the minimum difference between two variety means that the crop expert may declare to be different at a given significance level. The LSD is calculated using an estimate of random variation from the ANOVA.

2.4.1.2 It would be inconsistent with the rest of this document to describe the LSD in detail as descriptions can be found in many statistical text books. However, enough detail will be given to place it in context with the following sections on Multiple Range Tests and their comparison with LSD's

2.4.1.3 The LSD is chosen to give a particular size or significance level of test (α%) when comparing two means using a single characteristic, e.g. 5% or 1%. It means that if an LSD is used to make an a priori comparison, i.e. without knowledge of the data, then there is an α% chance of making a Type I error, i.e. declaring the means of two varieties to be significantly different when, if all plants of the two varieties could be examined, the means would not be different.

2.4.1.4 Although the LSD controls the comparison-wise Type I error chance, it does not control the experiment-wise Type I error chance, ie the chance that in all the comparisons made, the means of at least one pair of varieties are significantly different when, if all plants of the varieties could be examined, the means would not be different. The more comparisons that an LSD is used to make, the greater the experiment-wise Type I error chance. For example, if a 5% LSD is used to compare 14 independent pairs of means, then there is a 51% chance ($=100\% \times (1-(1-0.05)^{14})$) of declaring at least one of the pairs of variety means to be significantly different when, if all plants of the varieties could be examined, the means would not be different.

2.4.1.5 The standard formula for an α% LSD to compare two means made up of n1 and n2 observations respectively is:-

$$LSD\,\alpha\% = t_{(\alpha\%,rdf)} \times \sqrt{s^2 \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Where s^2 is an estimate of random variation taken from the ANOVA, rdf is the degrees of freedom of s^2, and $t_{(\alpha\%,\ rdf)}$ is either the two-sided or the one-sided α% critical value of the Student's t-statistic on rdf degrees of freedom, depending on whether the test is two-tailed or one-tailed.

2.4.1.6 It is important to note that in using an LSD to test the differences between variety means, the crop expert is assessing whether the difference in the variety means is larger than the difference that might reasonably have arisen due to chance or random variation affecting the observations making up the variety means when there was no difference between the varieties. Thus, the source of variation used to estimate random variation (s^2) in the LSD is very important in terms of the conclusions or inferences that can be drawn about the consistency of any differences between varieties declared to be significantly different.

2.4.2 The Multiple Range Test (MRT)

Note
The TWC agreed that in order to have a well-developed draft of the section on Multiple Range Test of TGP/8 for consideration by the TWC at its twenty-sixth session, Mrs. Sally Watson (United Kingdom), in conjunction with the Office and other experts, would develop a first draft text for comment by interested experts before the next draft of document TGP/8 was finalized for consideration by the Technical Working Parties in 2008. Mr. Nik Hulse (Australia), Mr. Uwe Meyer (Germany), Mr. Wieslaw Pilarczyk (Poland), Mr. Adrian Roberts (United Kingdom) and Mr. Gerie van der Heijden (Netherlands) agreed to comment on the first draft text as interested experts.

2.4.2.1 A Multiple Range Test (MRT), also known as a multiple comparison test, is similar to an LSD in that it is:

a statistic used to compare variety means from analysis of variance (ANOVA) of a characteristic and decide about the significance of variety differences
calculated using an estimate of random variation from the ANOVA

2.4.2.2 An MRT differs from an LSD in that it is chosen to give a particular size (α%) of test over all the comparisons for which it is intended for a characteristic. In other words, it controls to an extent the experiment-wise Type I error chance. It does this by reducing the comparison-wise Type I error chance and, as a result, the critical value of an α% MRT is larger than that of an α% LSD. This means that the MRT is usually more conservative than the LSD in that it is less likely to declare as significantly different two variety means where, if all plants of the two varieties could be examined, the means would not be different. On the other hand, the MRT is less powerful than the LSD as its Type II error chance is larger. In other words, there is a smaller chance with the MRT than with the LSD of declaring as significantly different two variety means where, if all plants of the two varieties could be examined, the means would be different.

2.4.2.3 There are a number of different MRT's. The choice of which to use depends partly on the comparisons to be made: for example, if one particular variety mean is to be compared with all others, or if all variety means are to be compared with all others. Descriptions of MRT's can be found in many statistical text books.

2.4.2.4 As with the LSD, the source of variation used to estimate random variation (s^2) in the MRT is very important in terms of the conclusions or inferences that can be drawn about the consistency of any differences between varieties declared to be significantly different.

2.4.3 Comparison of the use of the LSD and the MRT in distinctness testing

2.4.3.1 Comparison of the use of the LSD and the MRT in distinctness testing hinges on a risk that is different to both the experiment-wise and the comparison-wise Type I error chances. It is a risk of particular interest to testers and is called here the 'test-wise Type I error chance'. It is the chance of one or more candidates being significantly different from all other varieties in at least one characteristic when, if all plants of the varieties could be examined, the means would not be different. In other words, it is the chance of one or more candidates being wrongly declared as distinct when they are not distinct.

2.4.3.2 The test-wise Type I error chance increases with the number of candidates and with the number of characteristics used in the comparisons. It decreases with the number of reference varieties and proportionally with the significance level used when comparing varieties on a characteristic by characteristic basis, ie the comparison-wise Type I error chance. It is generally a very small chance, except however when there are few reference varieties, and in particular when there are many characteristics being used for the comparisons. Consequently, when trials are small, ie few reference varieties, and particularly when many characteristics are being used for variety comparisons, it is advantageous to use an MRT in place of an LSD, as the MRT serves to reduce the comparison-wise Type I error chance and hence reduce the test-wise Type I error chance. This effectively protects the tester's interests, as it reduces the chance of incorrectly declaring varieties distinct when they are not. In doing so it marginally penalises the breeders, as the lower power of the MRT makes it harder to detect differences when they do exist. An alternative in these circumstances would be to use an LSD with a smaller significance level. This would also reduce the comparison-wise Type I error chance and hence reduce the test-wise Type I error chance.

2.4.3.3 With larger trials the test-wise Type I error chance is very small, and so the advantage of the MRT over the LSD in controlling it does not exist and the LSD should be used in preference to the MRT as being the more powerful test.

2.4.3.4 Depending on which MRT is used, the minimum difference between two variety means represented by a MRT depends either on the total number of varieties in the trial or on the relative position rank-wise of other varieties with respect to the pair being compared. In either case, the acceptance of a candidate would be affected by the other candidates included in the trial, which may not be considered a fair system for testing. However, as the degree to which the acceptance of a candidate is affected is proportionate to the size of the trial, providing the MRT is used with small trials, its impact on testing is likely to be minor. Nonetheless, if the small trial has relatively many candidate varieties, it may be necessary to take steps in order not to penalise one breeder because another breeder has entered many candidates. In this regard, grouping of varieties as described in TGP/9 may be useful.

2.4.3.5 When using an MRT for distinctness testing not all the comparisons on which the MRT is based are necessarily made. Hence the MRT critical value is larger and the comparison-wise Type I error chance is smaller than are needed to achieve the intended experiment-wise Type I error chance. However, this is not a disadvantage when the MRT is used in small trials, as it still serves to reduce the test-wise Type I error chance.

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

3.1 The Combined Over-Years Distinctness Criterion (COYD)

3.1.1 Summary

3.1.1.1 Document TGP/9, section 5.2.4.5.1.1 *[cross ref.]* explains that "To assess distinctness for varieties on the basis of a quantitative characteristic it is possible to calculate a minimum distance between varieties such that, when the distance calculated between a pair of varieties is greater than this minimum distance, they may be considered as "distinct" in respect of that characteristic. Amongst the possible ways of establishing minimum distances is the method known as the Combined-Over-Years Distinctness (COYD)." The COYD analysis takes into account variation between years. Its main use is for cross-pollinated, including synthetic, varieties but, if desired, it can also be used for self-pollinated and vegetatively propagated varieties in certain circumstances. This method requires the size of the differences to be sufficiently consistent over the years and takes into account the variation between years.

3.1.1.2 The COYD method involves:

- for each characteristic, taking the variety means from the two or three years of trials for candidates and established varieties and producing over-year means for the varieties;
- calculate a least significant difference (LSD), based on variety-by-years variation, for comparing variety means.
- if the over-years mean difference between two varieties is greater than or equal to the LSD then the varieties are said to be distinct in respect of that characteristic.

3.1.1.3 The main advantages of the COYD method are:

- it combines information from several seasons into a single criterion (the "COYD criterion") in a simple and straightforward way;
- it ensures that judgements about distinctness will be reproducible in other seasons; in other words, the same genetic material should give similar results, within reasonable limits, from season-to-season;
- the risks of making a wrong judgement about distinctness are constant for all characteristics.

3.1.2 Introduction

3.1.2.1 The following sections describe:

- the principles underlying the COYD method;

- UPOV recommendations on the application of COYD to individual species;
- details of ways in which the procedure can be adapted to deal with special circumstances. This includes when there are small numbers of varieties in trial;
- the computer software which is available to apply the procedure.

3.1.3 The COYD method

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

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

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

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

3.1.3.5 Equation [1]

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

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

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

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

3.1.3.6 An example of the application of COYD to a small data set is given in Figure 1. Statistical details of the method are in Part II: section 3.1.8 *[cross ref.]*. Further information about the COYD criterion can be found in Patterson and Weatherup (1984).

3.1.4 Use of COYD

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

– the characteristic is quantitative;

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

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

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

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

3.1.5 Adapting COYD to special circumstances

3.1.5.1 Differences between years in the range of expression of a characteristic.

Occasionally, marked differences between years in the range of expression of a characteristic can occur. For example, in a late spring, the heading dates of grass varieties can converge. To take account of this effect it is possible to fit extra terms, one for each year, in the analysis of variance. Each term represents the linear regression of the observations for the year against the variety means over all years. The method is known as modified joint regression analysis (MJRA) and is recommended in situations where there is a statistically significant ($p \leq 1\%$) contribution from the regression terms in the analysis of variance. Statistical details, and a computer program to implement the procedure, are described in Part II sections 3.1.8 and 3.1.9 *[cross ref.]*.

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

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

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

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

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

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

3.1.5.2.5 Figure 2 gives an example of the application of Long-Term COYD to the Italian ryegrass characteristic "Growth habit in spring". A flow diagram of the stages and DUST modules used to produce Long-Term LSD's and perform Long-Term COYD is given in Figure B2 in Part II: section 3.1.9 *[cross ref.]*

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

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

3.1.6 Implementing COYD

Note:
TWC: it was noted that the DUST package contained more statistical methods than just COY and it was agreed that the text should be amended to clarify that aspect and to indicate which part of the DUST package was relevant for COY.

The COYD method can be applied using the DUST package for the statistical analysis of DUS data, which is available from Dr. Sally Watson, [Biometrics Branch, Agri-Food & Biosciences Institute, 18a, Newforge Lane, Belfast BT9 5PX, / to provide a web link] United Kingdom. Sample outputs are given in Part II: section 3.1.9. *[cross ref.]*

3.1.7 References

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

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

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

Figure 1: Illustrating the calculation of the COYD criterion

Characteristic: Days to ear emergence in perennial ryegrass varieties

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

ANALYSIS OF VARIANCE

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

$$\mathrm{LSD}_p = t_p * \sqrt{2} * \mathrm{SE}(\overline{X})$$

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

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

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

Figure 2: Illustrating the application of Long-Term COYD

Characteristic: **Growth habit in spring in Italian ryegrass varieties**

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

* indicates a test year

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

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

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

Long-term $LSD_p = t_p * \sqrt{2} * SE(\bar{X})$

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

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

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

3.1.8 COYD statistical methods

3.1.8.1 Analysis of variance

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

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

3.1.8.2 Modified joint regression analysis (MJRA)

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

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

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

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

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

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

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

b_j is the regression slope for year j

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

e_{ij} is an error term.

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

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

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

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

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

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

3.1.8.3 Comparison of COYD with other criteria

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

3.1.9 COYD software

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

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

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

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

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

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

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

                    VARIETY MEANS OVER YEARS
                       5        60        8         10        11        14        15        24
                    SP.HT     NSPHT     DEEE       H.EE     WEE       LFL       WFL       LEAR
 1 R1               45.27     34.60     67.87     45.20     70.05     20.39      6.85     24.54
 2 R2               42.63     31.84     73.85     41.96     74.98     19.68      6.67     24.44
 3 R3               41.57     27.40     38.47     27.14     57.60     17.12      6.85     22.57
 4 R4               33.35     21.80     77.78     30.77     78.04     18.25      6.40     21.09
 5 R5               37.81     25.86     50.14     27.24     62.64     16.41      6.41     16.97
 6 R6               33.90     21.07     78.73     32.84     79.15     19.44      6.46     21.79
 7 R7               41.30     31.37     73.19     41.35     71.87     20.98      6.92     24.31
 8 R8               24.48     19.94     74.83     32.10     62.38     15.22      6.36     19.46
 9 R9               46.68     36.69     63.99     44.84     68.62     18.11      7.02     22.58
10 R10              25.60     20.96     75.64     32.31     57.20     14.68      5.51     20.13
11 R11              41.70     30.31     74.60     40.17     76.15     19.45      6.79     22.72
12 R12              28.95     21.56     66.12     27.96     59.56     14.83      5.53     20.55
13 R13              40.67     29.47     70.63     36.81     74.12     19.97      7.04     24.05
14 R14              26.68     20.53     75.84     34.14     63.29     15.21      6.37     20.37
15 R15              26.78     20.18     75.54     30.39     66.41     16.34      6.01     20.94
16 R16              42.44     27.01     59.03     30.39     72.71     17.29      6.47     22.48
17 R17              27.94     21.58     76.13     32.53     68.37     16.72      6.11     22.03
18 R18              41.34     30.85     69.80     37.28     69.52     20.68      7.09     25.40
19 R19              33.54     23.43     73.65     30.35     75.54     18.97      6.37     22.43
20 R20              44.14     34.48     68.74     42.60     64.17     18.63      6.56     22.02
21 R21              27.77     21.53     80.52     31.59     69.41     16.81      5.81     22.35
22 R22              38.90     27.83     75.68     43.25     75.08     19.63      7.46     23.99
23 R23              42.43     31.80     72.40     42.07     74.77     20.99      6.78     23.57
24 R24              38.50     27.73     73.19     37.12     75.76     19.28      6.91     22.77
25 R25              43.84     29.60     68.82     39.79     74.83     20.63      7.08     22.65
26 R26              49.48     36.53     63.45     42.01     70.46     22.14      7.84     25.91
27 R27              25.61     19.25     78.78     29.81     56.81     15.81      5.07     18.94
28 R28              26.70     20.31     79.41     32.75     66.54     16.92      6.00     21.91
29 R29              27.90     20.94     72.66     29.85     67.14     16.85      6.28     21.79
30 R30              43.07     30.34     70.53     40.51     73.23     19.49      7.28     23.70
31 R31              38.18     25.47     74.23     36.88     80.23     20.40      7.09     25.21
32 R32              35.15     27.56     71.49     37.26     63.10     18.18      6.80     23.13
33 R33              42.71     31.09     67.58     39.14     70.36     19.85      7.12     23.35
34 R34              23.14     18.05     72.09     24.29     59.37     13.98      5.63     18.91
35 R35              32.75     25.41     77.22     38.90     67.07     17.16      6.42     21.49
36 R36              41.71     31.94     77.98     44.33     73.00     19.72      7.09     23.45
37 R37              44.06     32.99     74.38     45.77     71.59     20.88      7.40     24.06
38 R38              42.65     32.97     74.76     44.42     74.13     20.29      7.38     24.32
39 R39              28.79     22.41     76.83     35.91     64.52     16.85      6.34     22.24
40 R40              44.31     31.38     72.24     43.83     74.73     21.53      7.60     25.46
41 C1               42.42     31.68     64.03     40.22     67.02     20.73      6.90     26.16
42 C2               41.77     32.35     86.11     46.03     75.35     20.40      6.96     22.99
43 C3               41.94     31.09     82.04     43.17     74.04     19.06      6.26     23.44
44 C4               39.03     28.71     78.63     45.97     70.49     21.27      6.67     23.37
45 C5               43.97     30.95     72.99     39.14     77.89     19.88      6.68     25.44
46 C6               37.56     27.14     83.29     39.16     81.18     19.47      6.97     25.25
47 C7               38.41     28.58     83.90     42.53     76.44     19.28      6.00     23.47
48 C8               40.08     27.25     83.50     43.33     80.16     22.77      7.92     26.81
49 C9               46.77     34.87     51.89     37.68     61.16     19.25      6.92     24.82

YEAR MS           1279.09   3398.82   3026.80   2278.15   8449.20    672.15      3.36     51.32
VARIETY MS         909.21    476.72   1376.10    635.27    762.41     80.21      6.44     74.17
VAR.YEAR MS         23.16     18.86     14.12     23.16     46.58      4.76      0.28      2.73
F1 RATIO            39.26     25.27     97.43     27.43     16.37     16.84     22.83     27.16
VAR.REP MS           8.83      8.19      4.59     11.95     23.23      1.52      0.15      1.70
LAMBDA VALUE         1.62      1.52      1.75      1.39      1.42      1.77      1.37      1.27
BETWEEN SE           1.13      1.02      0.89      1.13      1.61      0.51      0.13      0.39
WITHIN SE            0.70      0.67      0.50      0.81      1.14      0.29      0.09      0.31
DF                 96        94        96        96        96        96        96        96
MJRA SLOPE 88        0.90      0.86      0.99      0.91      0.99      1.09      0.97      0.95
MJRA SLOPE 89        1.05      1.08      1.01      0.99      1.06      0.97      1.02      0.98
MJRA SLOPE 90        1.05      1.06      1.00      1.10      0.95      0.94      1.01      1.07
REGR F VAL           4.66      6.17      0.06      4.48      0.76      1.62      0.29      1.91
REGR PROB            1.17      0.30     93.82      1.39     47.08     20.27     74.68     15.38
TEST               COY       REG       COY       COY       COY       COY       COY       COY
```

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

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

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

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

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

Notes

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

2. The two right-hand "F3" columns give the F_3 variance ratio statistic and its significance level. The F_3 statistic is defined in Part II, section 3.1.5.2 *[cross ref.]*.

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

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

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

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

CANDIDATE VARIETIES        C1     C2     C3     C4     C5     C6     C7     C8     C9
1              R1          D      D      D      D      D      D      D      D      D
2              R2          D      D      D      D      ND     D      D      D      D
3              R3          D      D      D      D      D      D      D      D      D
4              R4          D      D      D      D      D      D      D      D      D
5              R5          D      D      D      D      D      D      D      D      D
6              R6          D      D      D      D      D      D      D      D      D
7              R7          D      D      D      D      D      D      D      D      D
8              R8          D      D      D      D      D      D      D      D      D
9              R9          D      D      D      D      D      D      D      D      D
10             R10         D      D      D      D      D      D      D      D      D
11             R11         D      D      D      D      D      D      D      D      D
12             R1          D      D      D      D      D      D      D      D      D
13             R13         D      D      D      D      ND     D      D      D      D
14             R14         D      D      D      D      D      D      D      D      D
15             R15         D      D      D      D      D      D      D      D      D
16             R16         D      D      D      D      D      D      D      D      D
17             R17         D      D      D      D      D      D      D      D      D
18             R18         D      D      D      D      D      D      D      D      D
19             R19         D      D      D      D      D      D      D      D      D
20             R20         D      D      D      D      D      D      D      D      D
21             R21         D      D      D      D      D      D      D      D      D
22             R22         D      D      D      D      D      D      D      D      D
23             R23         D      D      D      D      D      D      D      D      D
24             R24         D      D      D      D      D      D      D      D      D
25             R25         D      D      D      D      D      D      D      D      D
26             R26         D      D      D      D      D      D      D      D      D
27             R27         D      D      D      D      D      D      D      D      D
28             R28         D      D      D      D      D      D      D      D      D
29             R29         D      D      D      D      D      D      D      D      D
30             R30         D      D      D      D      D      D      D      D      D
31             R31         D      D      D      D      D      D      D      D      D
32             R32         D      D      D      D      D      D      D      D      D
33             R33         D      D      D      D      D      D      D      D      D
34             R34         D      D      D      D      D      D      D      D      D
35             R35         D      D      D      D      D      D      D      D      D
36             R36         D      D      D      ND     D      D      D      D      D
37             R37         D      D      D      D      D      D      D      D      D
38             R38         D      D      D      D      D      D      D      D      D
39             R39         D      D      D      D      D      D      D      D      D
40             R40         D      D      D      D      D      D      D      D      D

41             C1          -      D      D      D      D      D      D      D      D
42             C2          D      -      D      D      D      D      D      D      D
43             C3          D      D      -      D      D      D      ND     D      D
44             C4          D      D      D      -      D      D      D      D      D
45             C5          D      D      D      D      -      D      D      D      D
46             C6          D      D      D      D      D      -      D      D      D
47             C7          D      D      ND     D      D      D      -      D      D
48             C8          D      D      D      D      D      D      D      -      D
49             C9          D      D      D      D      D      D      D      D      -

NO OF ND VARS              0      0      1      1      2      0      1      0      0
DISTINCTNESS               D      D      ND     ND     ND     D      ND     D      D
CANDIDATE VAR              C1     C2     C3     C4     C5     C6     C7     C8     C9
```

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

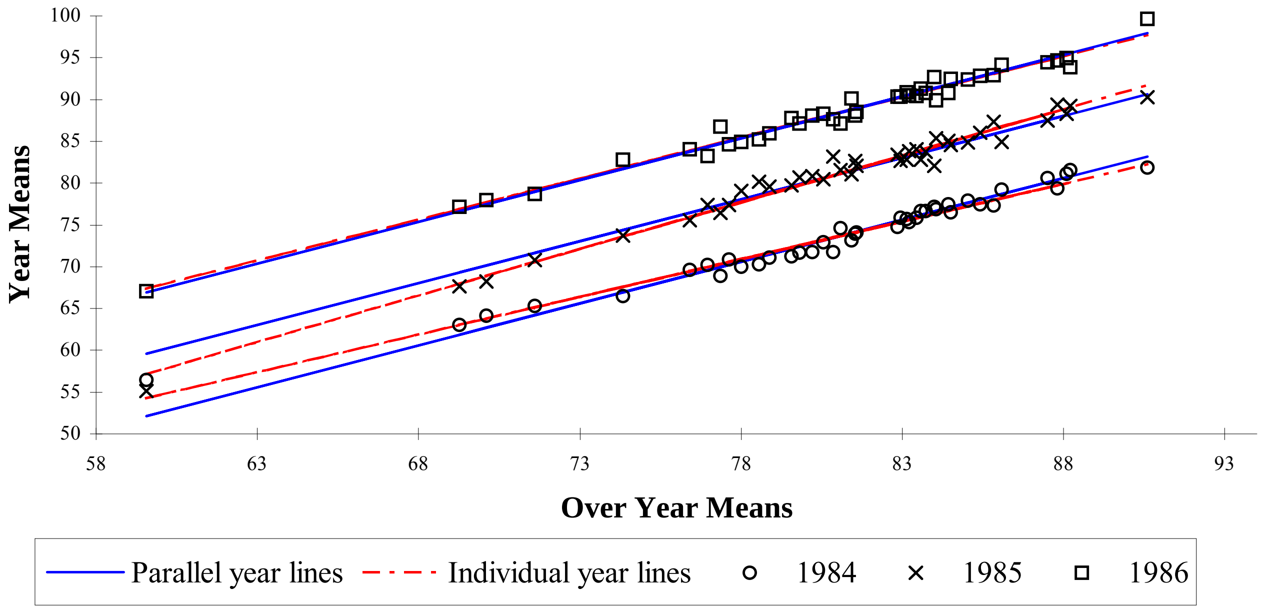
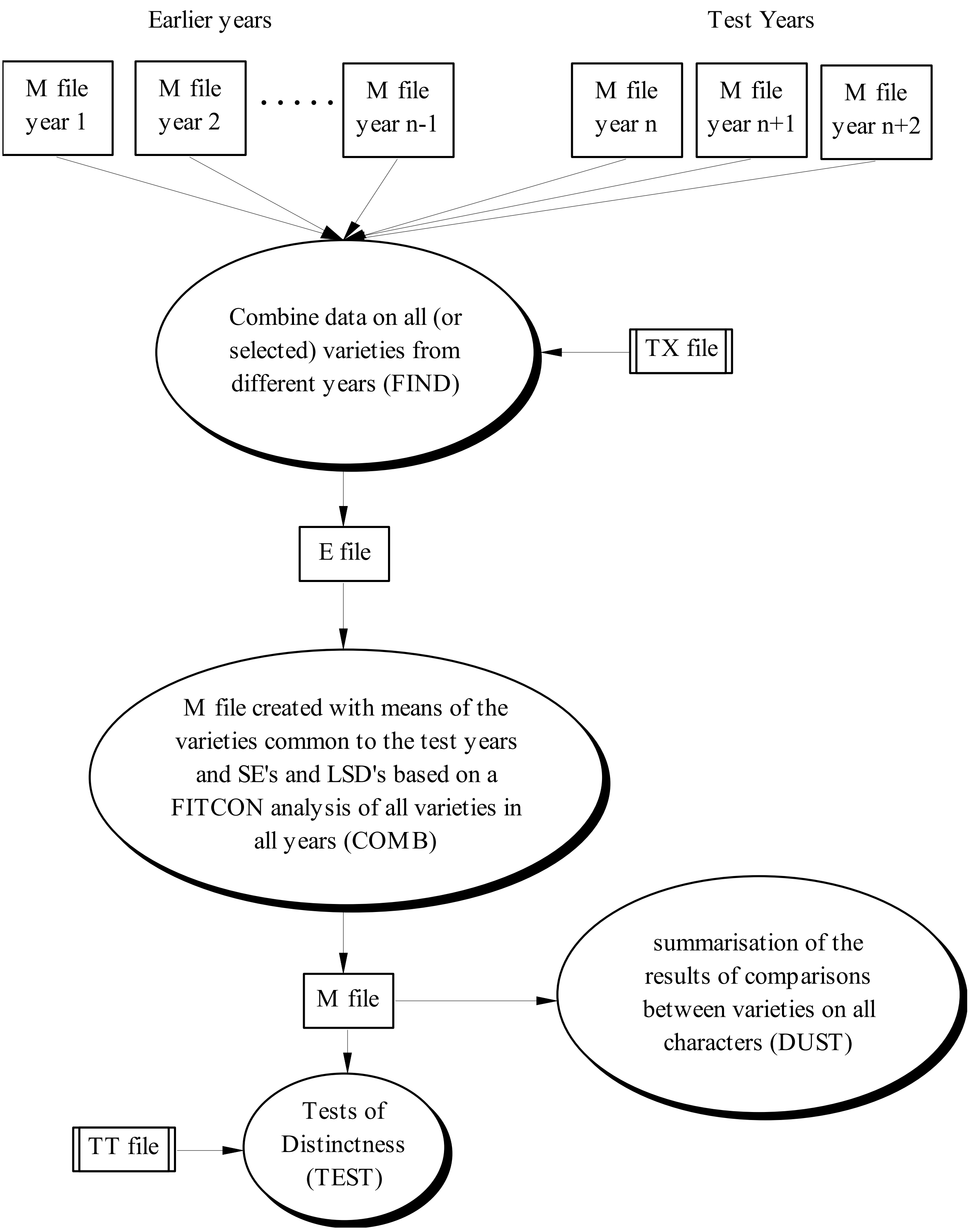


Figure B2. Flow Diagram of the stages and DUST modules used to produce long-term LSD's and perform long-term COYD



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

3.2.1 Summary

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

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

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

3.2.1.3 The technique involves ranking reference and candidate varieties by the mean value of the characteristic. Each variety’s SD is taken and the mean SD of the most similar varieties is subtracted. This procedure gives, for each variety, a measure of its uniformity expressed relative to that of similar varieties. The term reference varieties here refers to established varieties which have been included in the growing trial and which have comparable expression of the characteristics under investigation

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

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

3.2.2 Introduction

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

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

3.2.2.3 This section describes:

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

3.2.3 The COYU Criterion

3.2.3.1 The application of the COYU criterion involves a number of steps as listed below. These are applied to each characteristic in turn. Details are given under Part II: section 3.2.5 *[cross ref.]* below.

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

3.2.3.2 The advantages of the COYU criterion are:

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

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

3.2.4 Recommendations on COYU

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

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

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

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

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

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

3.2.5 Mathematical details

Step 1: Derivation of the within-plot standard deviation

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

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

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

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

Step 2: Transformation of the SDs

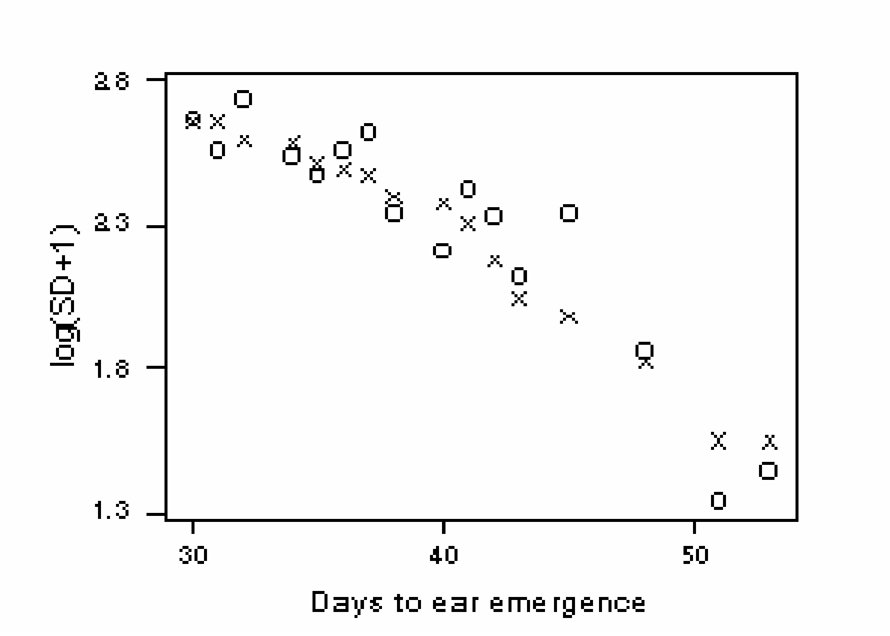
3.2.5.2 Transformation of SDs by adding 1 and converting to natural logarithms. The purpose of this transformation is to make the SDs more amenable to statistical analysis.

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

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

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

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



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

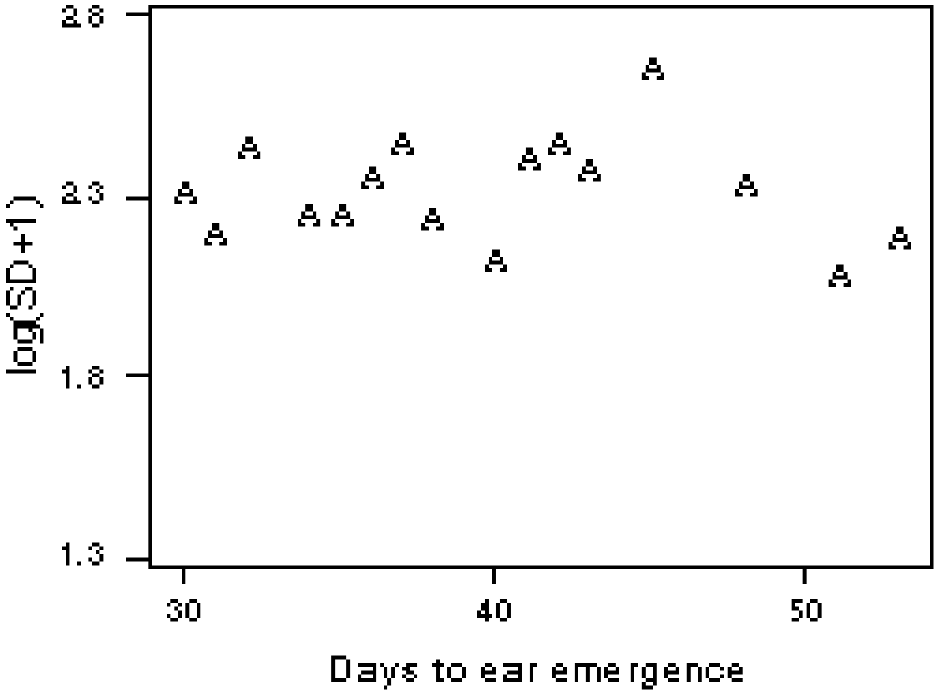
3.2.5.5 Once the trend values for the reference varieties have been determined, the trend values for candidates are estimated using linear interpolation between the trend values of the nearest two reference varieties as defined by their means for the characteristic. Thus if the trend values for the two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

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

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

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

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



Step 5: Calculation of the uniformity criterion

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

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

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

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

3.2.6 Early decisions for a three-year test

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

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

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

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

3.2.7 Example of COYU calculations

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\mathrm{UC_p} = \mathrm{SD_r} + \mathrm{t_p}\sqrt{\mathrm{V}\left(\frac{1}{\mathrm{k}}+\frac{1}{\mathrm{Rk}}\right)} = 2.15 + 3.118\mathrm{x}\sqrt{0.0202\mathrm{x}\left(\frac{1}{3}+\frac{1}{3\mathrm{x}11}\right)} = 2.42$$

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

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

3.2.8 Implementing COYU

The COYU criterion can be applied using the DUST software package for the statistical analysis of DUS data. This is available from the Dr. Sally Watson, Biometrics & Information Systems, Agri-Food & Biosciences Institute, Newforge Lane, Belfast BT9 5PX, UK .

3.2.9 COYU software

3.2.9.1 DUST computer program

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

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

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

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

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

Table A1: Example of summary output from COYU program

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

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

           CHARACTERISTIC NUMBER
                5    60      8     10     11

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

CHARACTERISTIC

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

SYMBOLS

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

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

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

CANDIDATE
C1          64.03  2.252   2.438    63.85  63.33  64.92   2.49X  2.81X  2.02   2.25   2.29   2.21
C2          86.11  1.940   1.837    84.83  88.63  84.85   1.79   1.71   2.01   1.90   2.05   1.87
C3          82.04  2.349   2.248    82.26  87.45  76.40   2.37X  2.03   2.35X  2.48   2.37   2.20
C4          78.63  2.104   2.033    78.01  82.17  75.72   2.05   2.01   2.04   2.15   2.27   1.90
C5          72.99  1.973   1.869    71.98  79.40  67.59   1.95   1.78   1.88   1.93   1.90   2.08
C6          83.29  2.050   1.947    84.10  85.57  80.21   2.05   1.69   2.10   2.16   2.03   1.96
C7          83.90  2.100   1.997    84.12  87.99  79.60   1.93   1.95   2.11   2.04   2.29   1.97
C8          83.50  2.304   2.201    82.43  85.98  82.08   2.27X  2.00   2.34X  2.38   2.33   2.20
C9          51.89  1.788   2.157    52.35  55.77  47.56   1.83   2.34X  2.31X  1.52   1.91   1.93

MEAN OF
REFERENCE   71.47  1.988            70.78  74.97  68.65   1.97   2.03   1.96   1.99   1.99   1.99

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

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

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

 TOTAL      119    0.05313


    SYMBOLS

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

3.3 Schemes used for the application of COYD and COYU

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

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

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

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

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

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

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

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

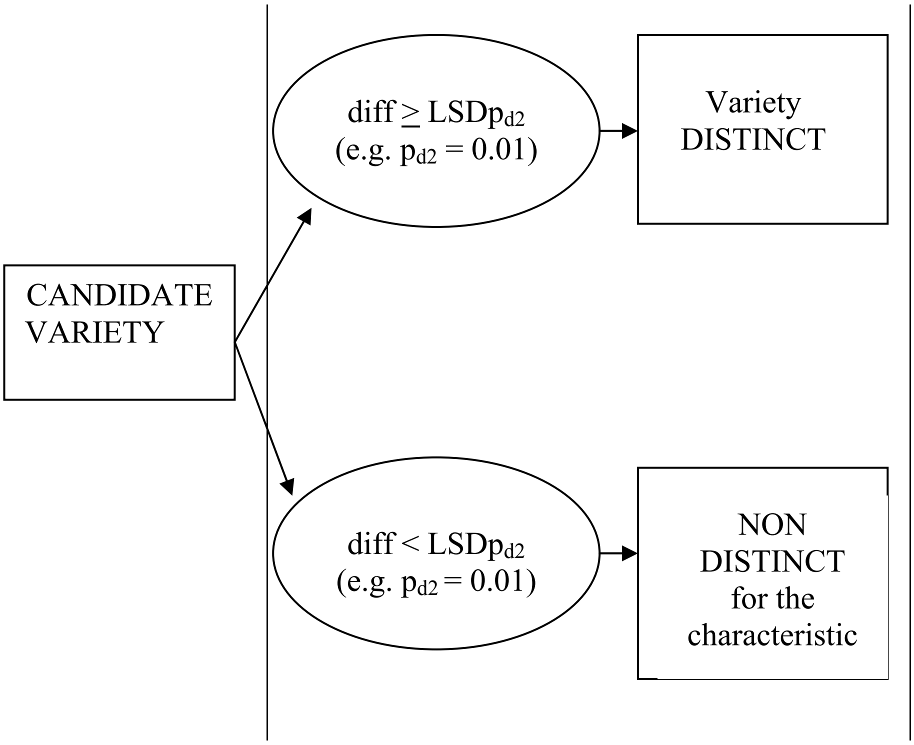
Table 1 summarises the various standard probability levels needed to calculate the COYD and COYU criteria in each of Cases A to D. For example, in Case B only two probability levels are needed (p_{d3} and p_{u3}), whereas Case C requires four (p_{d2}, p_{d3}, p_{u2} and p_{u3}).

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

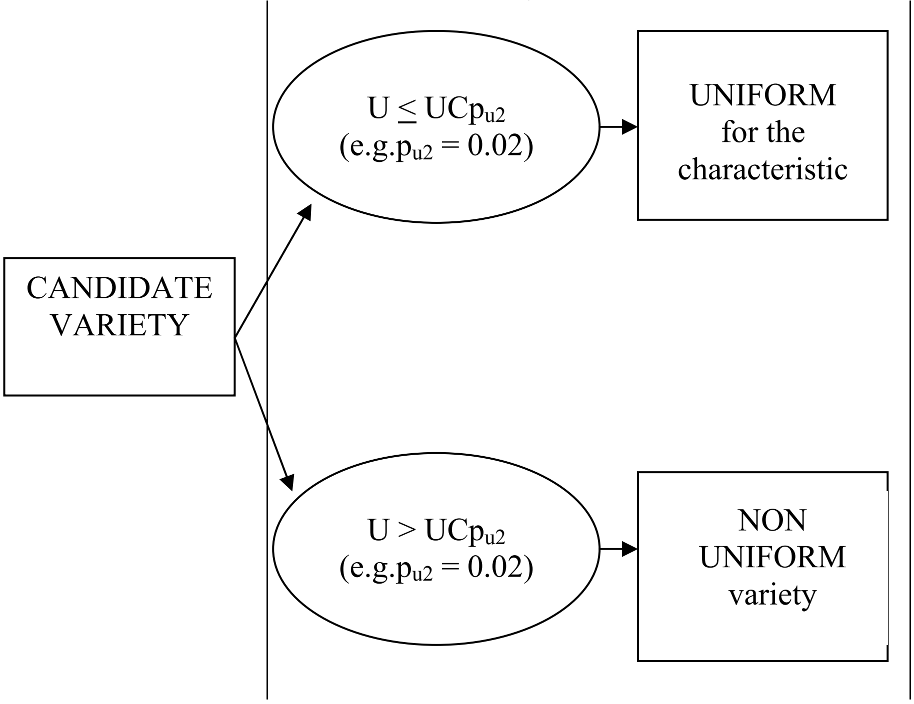
The actual standard probability levels used for the application of COYD and COYU with different crops by various UPOV member states have been ascertained by questionnaire. See document TWC/23/10 (or a more recent version) *[cross ref.]*.

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

a) COYD Decision after 2nd cycle



b) COYU Decision after 2nd cycle

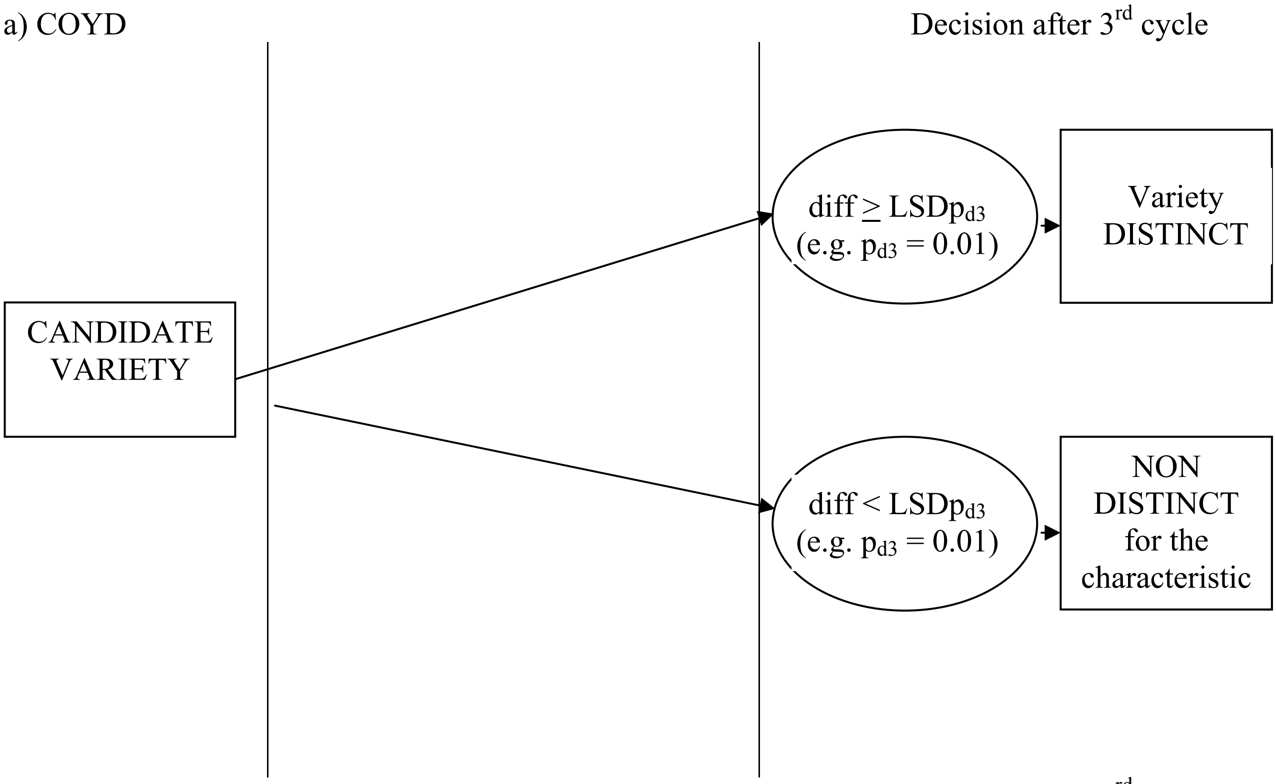


NOTE:-

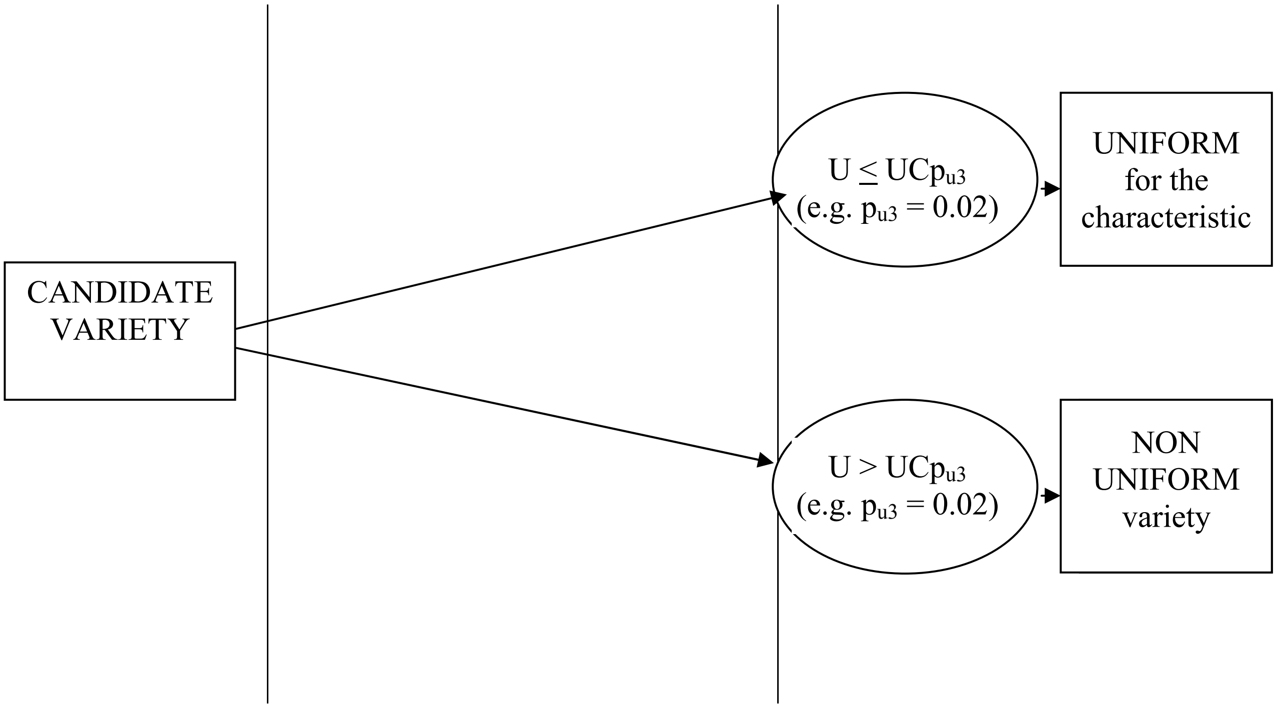
“diff”	is the difference between the means of the candidate variety and another variety for the characteristic.
LSDp	is the COYD criterion calculated at probability level p.
“U”	is the mean adjusted log(SD+1) of the candidate variety for the characteristic.
UCp	is the COYU criterion calculated at probability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

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

a) COYD



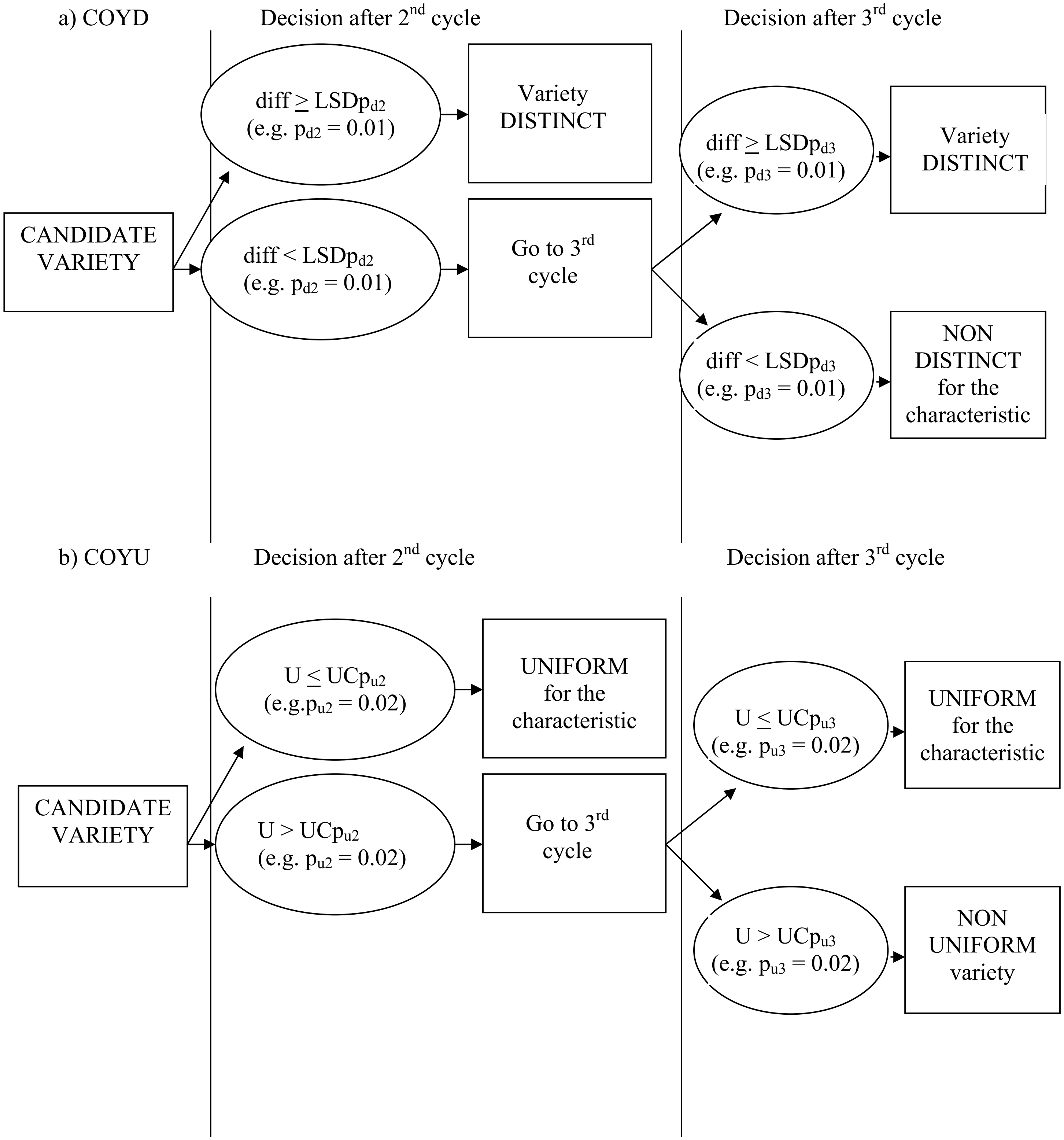
b) COYU



NOTE:-

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

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

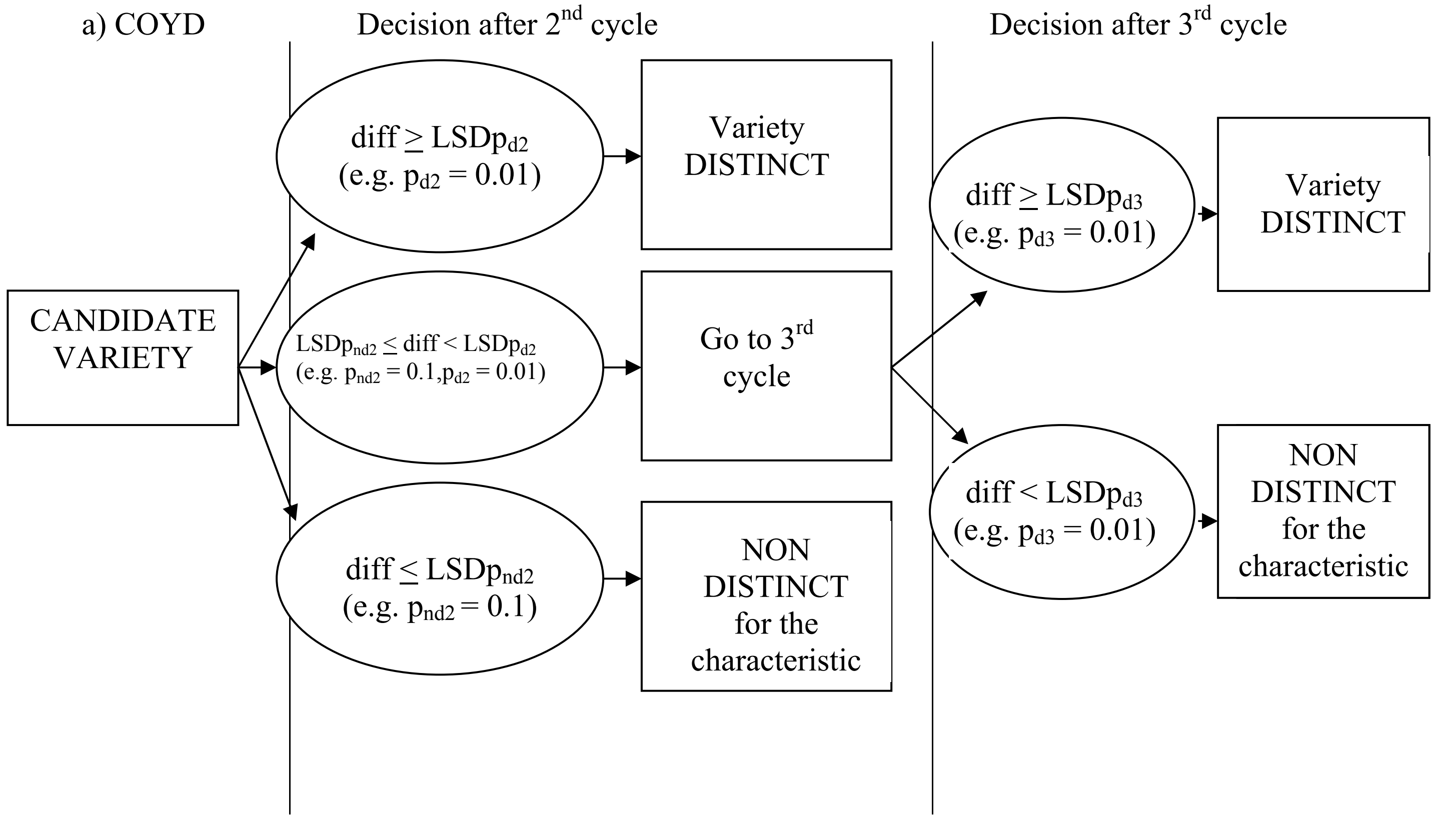


NOTE:-

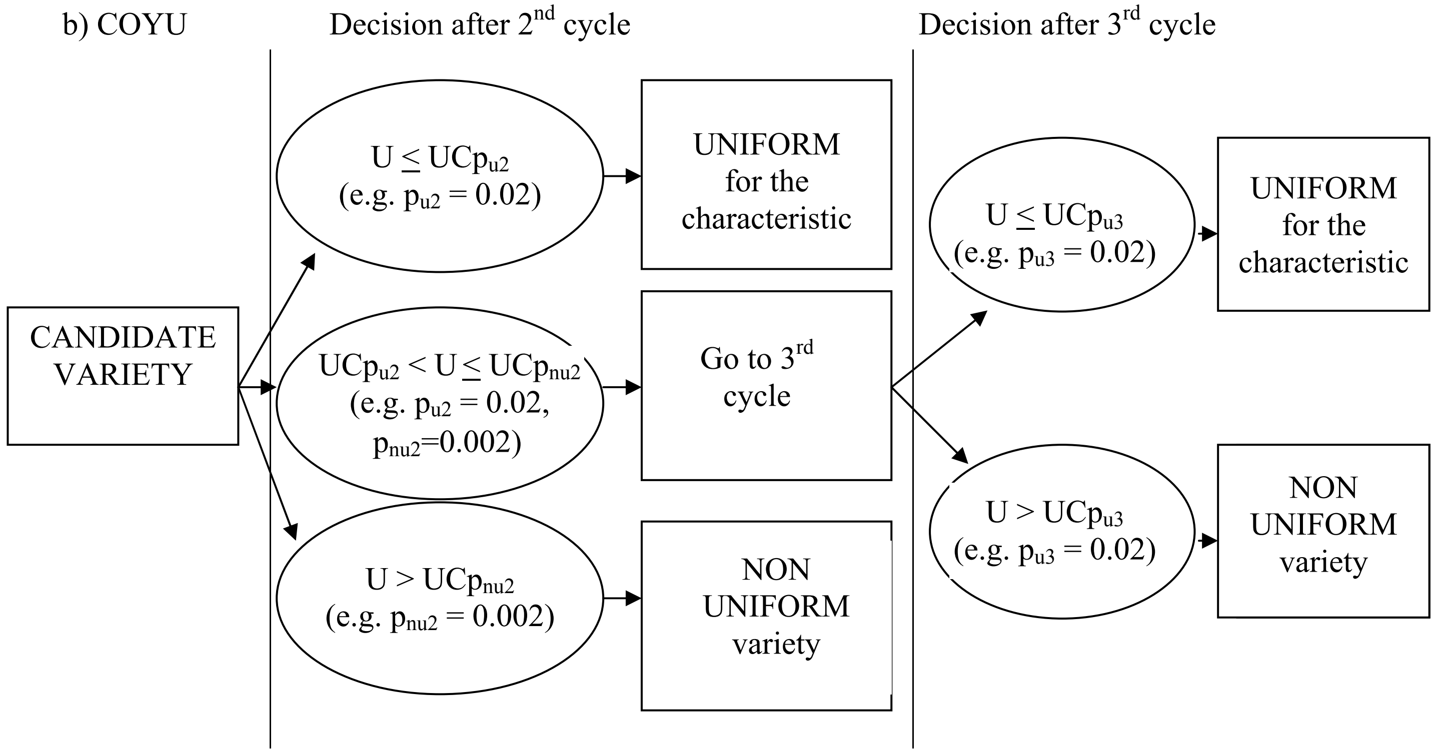
“diff”	is the difference between the means of the candidate variety and another variety for the characteristic.
LSDp	is the COYD criterion calculated at probability level p.
“U”	is the mean adjusted log(SD+1) of the candidate variety for the characteristic.
UCp	is the COYU criterion calculated at probability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

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

a) COYD



b) COYU



NOTE:-

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

4 SECTION ON 2X1% METHOD

[The following material was originally in TGP/9 but has been removed]

4.1 2x1% Criterion (Method)

4.1.1 For two varieties to be distinct using the 2x1% criterion, the varieties need to be significantly different in the same direction at the 1% level in at least two out of three years in one or more measured characteristics. The tests in each year are based on Student's two-tailed t-test of the differences between variety means with standard errors estimated using the plot residual mean square from the analysis of the variety x replicate plot means.

4.1.2 With respect to the 2x1% criterion, compared to COYD, it is important to note that:

- Information is lost because the criterion is based on the accumulated decisions arising from the results of t-tests made in each of the test years. Thus, a difference which is not quite significant at the 1% level contributes no more to the separation of a variety pair than a zero difference or a difference in the opposite direction. For example, three differences in the same direction, one of which is significant at the 1% level and the others at the 5% level would not be regarded as distinct.
- Some characteristics are more consistent over years than others in their expression of differences between varieties. However, beyond requiring differences to be in the same direction in order to count towards distinctness, the 2x1% criterion takes no account of consistency in the size of the differences from year to year. The result is that the risks of wrongly declaring distinctness (declaring distinctness when, if all plants of the varieties could be examined, they would not be distinct) are greater in characteristics that are inconsistent over years than in consistent characteristics.

5 SECTION ON SINGLE GROWING CYCLE METHOD

5.1 Single Growing Cycle Method

[TWP's are invited to provide information on this method.]

5.1.1 In the absence of information on this method it is thought likely that the following applies:-

- For two varieties to be distinct using the Single Growing Cycle method, the varieties need to be significantly different at a given significance level in one or more measured characteristics. Differences can be assessed using a statistical test based on a two-tailed LSD to compare the variety means with standard errors calculated using the plot residual mean square from the analysis of the variety x replicate plot means as the estimate of random variation (s^2).

- The source of variation used to estimate random variation (s^2) in the LSD determines what can be inferred from using the LSD. The Single Growing Cycle method estimates random variation (s^2) in the LSD using the plot residual mean square, which represents the plot-to-plot variation within a variety (allowing for any block effects if blocks are present). As a result, using this LSD the crop expert can conclude that varieties with significantly different means are different relative to the plot-to-plot variation within a variety.

NEW SECTION: CHI-SQUARE TEST

Note

Chi-Square Test: the TWC experts did not have experience of the use of chi-square test for segregating characteristics, but invited experts to consult their colleagues to see if it would be possible to provide an explanation

6. PARENT FORMULA OF HYBRID VARIETIES

6.1 Introduction

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

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

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

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

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

6.2 Requirements of the method

The application of the method requires:

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

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

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

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

6.3 Assessing the originality of a new parent line

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

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

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

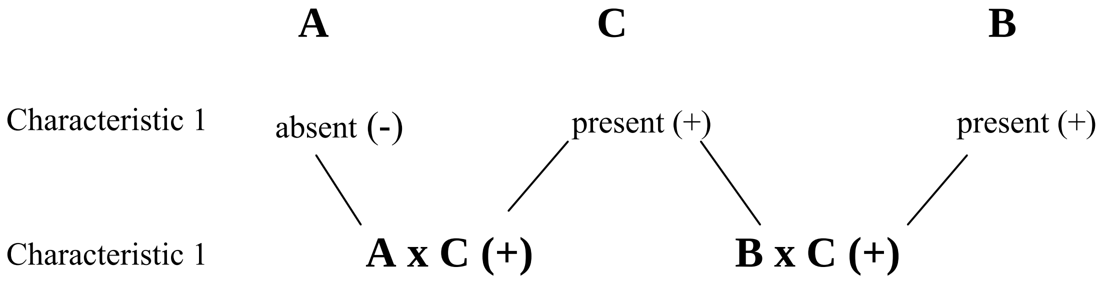
Three parent lines:

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

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

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

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



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

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

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

6.4 Verification of the formula

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

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

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

6.5 Uniformity and stability of parent lines

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

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

6.6 Description of the hybrid

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

7 THE GAIA METHODOLOGY

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

The GAIA methodology is mainly used after a first growing cycle to identify those varieties of common knowledge which can be excluded from the subsequent growing cycle(s) because they are “Distinct Plus” (see section 7.3.2 *[cross ref.]*) from all the candidate varieties. GAIA can also identify similar varieties, on which the DUS examiner will need to focus attention in the subsequent growing cycle(s).

7.1 Some reasons to sum and weight observed differences

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

7.1.2 A DUS examiner may see that two varieties are so distinct after the first growing cycle that it is not necessary to repeat the comparison. Those two varieties, which are “distinct plus” (see section 7.3.2.1 *[cross ref.]*), are obviously distinct.

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

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

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

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

7.2 Computing GAIA phenotypic distance

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

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

where:

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

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

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

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

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

7.2.2 Detailed information on e is provided in section 7.3

7.3 Detailed information on the GAIA methodology

7.3.1 Weighting of characteristics

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

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

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

If so,

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

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

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

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

Example: taking the characteristic "Weighting matrix shape of ear", observed on a 1 to 3 scale, the crop expert has attributed weighting to differences which they consider significant:

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

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

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

Weighting matrix ‘i’				
	Variety i			
		1	**2**	**3**
Variety ‘j’	**1**	0	2	6
	2		0	2
	3			0

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

7.3.2 Examples of use

7.3.2.1 Determining “Distinctness Plus”

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

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

7.3.2.2 Other examples of use

Using phenotypic distance in the first growing cycle

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

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

Using phenotypic distance after the first growing trial

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

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

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

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

7.3.3 Computing GAIA phenotypic distance

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

GAIA phenotypic distance is:

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

where:

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

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

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

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

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

This phenotypic distance computations allows to:

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

7.3.4 GAIA software

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

7.3.4.2 The main use of GAIA is to define a "distinct plus" threshold which corresponds to a reliable and obvious distinction.

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

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

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

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

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

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

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

7.3.4.9 Section 7.3.5 provides a screen copy of a display tree which shows how the expert can navigate and visualise the results of computations.

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

7.3.4.11 1 or 2 notes per variety can be used. 1 note occurs when one cycle is available. Two notes are present for instance when two trials are made in different locations a given year, or if 2 cycles are obtained in the same location.
For electrophoresis data, only one description can be entered per variety.
For measurements at least 2 values (different trials, repeats, etc.) are necessary and the user can select which to use in the computation.

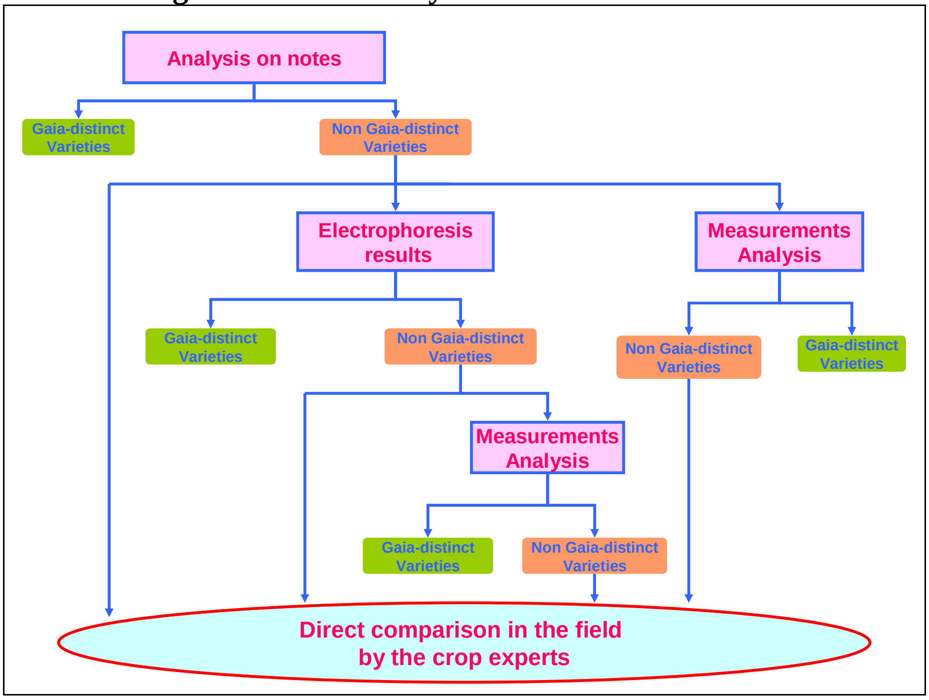
7.3.4.12 GAIA is most suitable for self-pollinated and vegetatively propagated varieties, but can also be used for other types of varieties.

7.3.5 Example with *Zea mays* data

7.3.5.1 Introduction

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

Diagram 1: Data analysis scheme



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

7.3.5.2 Analysis of notes

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

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

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

1 = conical
2 = conico-cylindrical
3 = cylindrical

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

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

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

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

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

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

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

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

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

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

7.3.5.3 Electrophoresis analysis

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

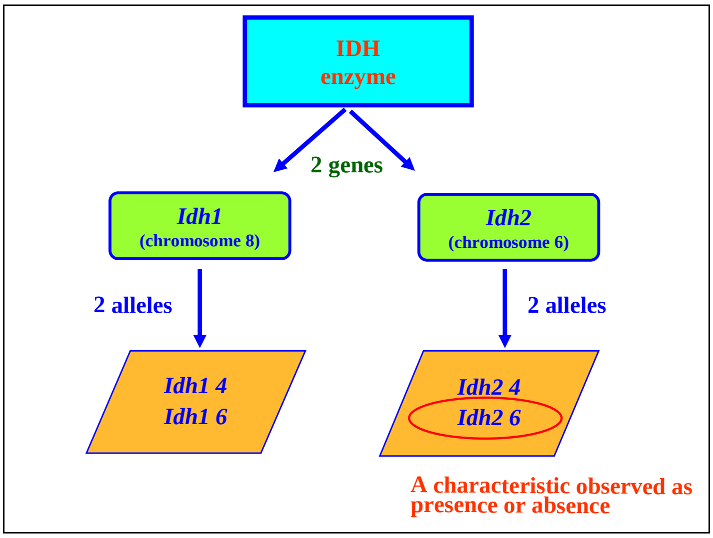


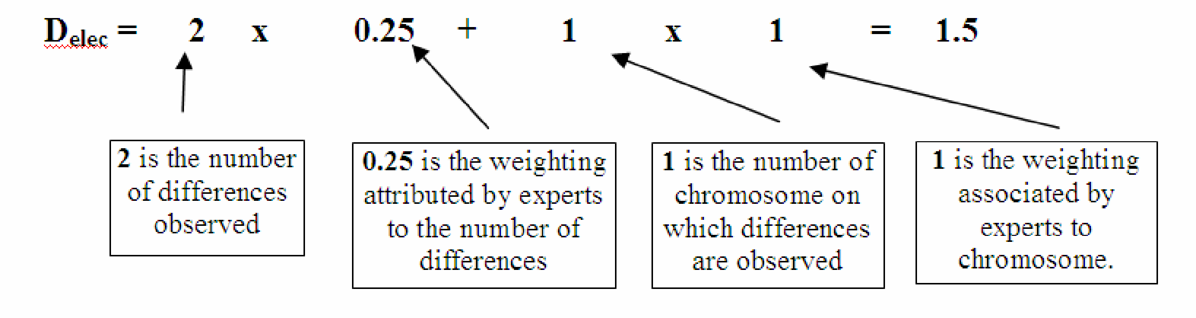
Diagram 2: The Isocitrate Dehydrogenase (IDH) enzyme has two genes (*Idh1* and *Idh2*) located on two different chromosomes. Each of them has two alleles which are observed as 1 (presence) or 0 (absence).

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

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

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

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



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

2 is the number of differences observed

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

1 is the number of chromosome on which differences are observed

1 is the weighting associated by experts to chromosome.

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

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

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

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

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

7.3.5.4 Analysis of measurements

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

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

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

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

For each minimal distance a weighting is attributed:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

7.3.5.4.9 The phenotypic distance is greater than the distinction threshold $\mathbf{S_{dist}}$, fixed by the crop expert at 10, so varieties A and B are declared “GAIA-distinct”.

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

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

1 = conical
2 = conico-cylindrical
3 = cylindrical

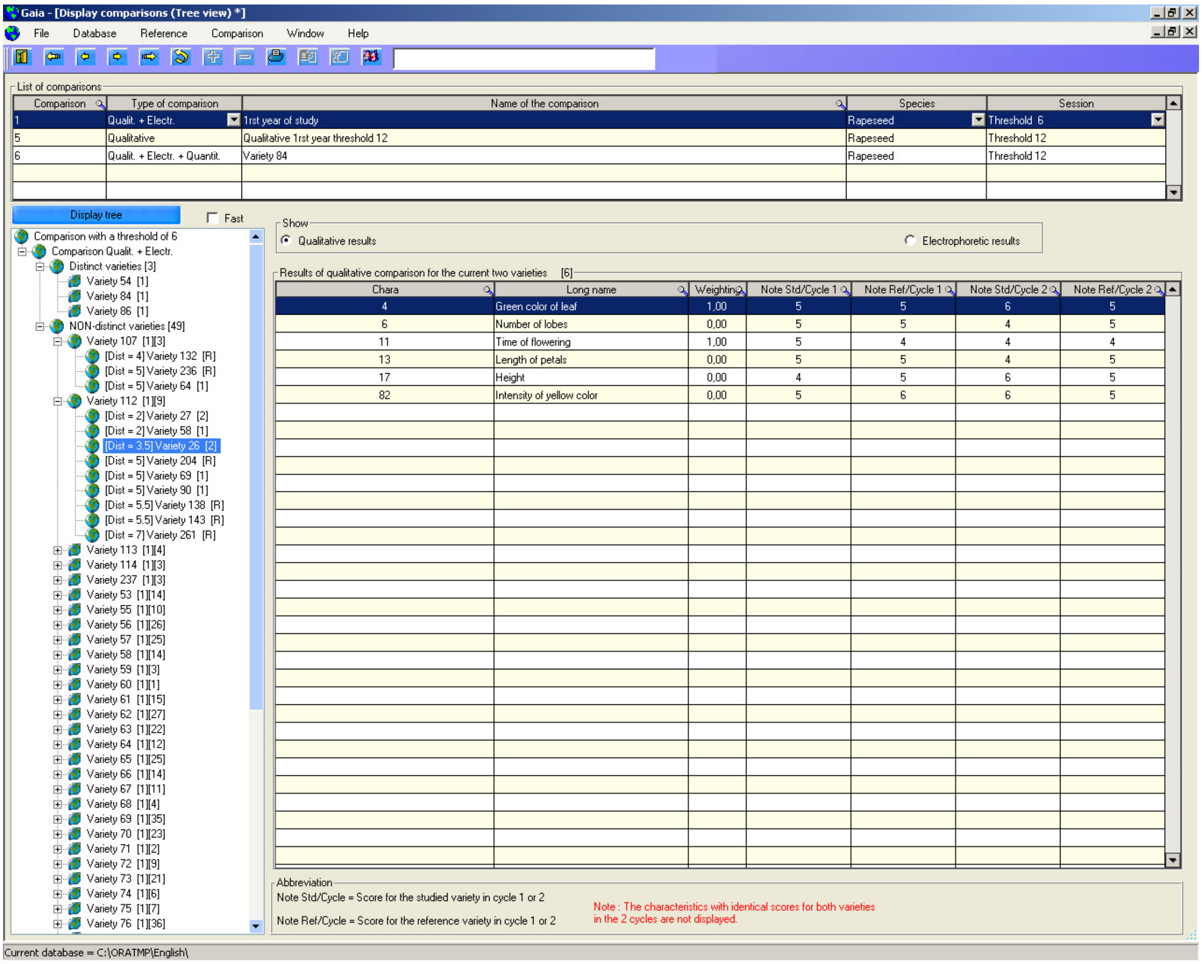
Variety i			
	1	2	3
1	0	2	6
2		0	2
3			0

7.3.5.5 *Measurements and 1 to 9 scale on the same characteristic*

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

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

7.3.6 Example of GAIA screen copy



List of comparisons

Comparison	Type of comparison	Name of the comparison	Species	Session
1	Qualit. + Electr.	1rst year of study	Rapeseed	Threshold 6
5	Qualitative	Qualitative 1rst year threshold 12	Rapeseed	Threshold 12
6	Qualit. + Electr. + Quantit.	Variety 84	Rapeseed	Threshold 12

Comparison with a threshold of 6
- Comparison Qualit. + Electr.
  - Distinct varieties [3]
    - Variety 54 [1]
    - Variety 84 [1]
    - Variety 86 [1]
  - NON-distinct varieties [49]
    - Variety 107 [1][3]
      - [Dist = 4] Variety 132 [R]
      - [Dist = 5] Variety 236 [R]
      - [Dist = 5] Variety 64 [1]
    - Variety 112 [1][9]
      - [Dist = 2] Variety 27 [2]
      - [Dist = 2] Variety 58 [1]
      - [Dist = 3.5] Variety 26 [2]
      - [Dist = 5] Variety 204 [R]
      - [Dist = 5] Variety 69 [1]
      - [Dist = 5] Variety 90 [1]
      - [Dist = 5.5] Variety 138 [R]
      - [Dist = 5.5] Variety 143 [R]
      - [Dist = 7] Variety 261 [R]
    - Variety 113 [1][4]
    - Variety 114 [1][3]
    - Variety 237 [1][3]
    - Variety 53 [1][14]
    - Variety 55 [1][10]
    - Variety 56 [1][26]
    - Variety 57 [1][25]
    - Variety 58 [1][14]
    - Variety 59 [1][3]
    - Variety 60 [1][1]
    - Variety 61 [1][15]
    - Variety 62 [1][27]
    - Variety 63 [1][22]
    - Variety 64 [1][12]
    - Variety 65 [1][25]
    - Variety 66 [1][14]
    - Variety 67 [1][11]
    - Variety 68 [1][4]
    - Variety 69 [1][35]
    - Variety 70 [1][23]
    - Variety 71 [1][2]
    - Variety 72 [1][9]
    - Variety 73 [1][21]
    - Variety 74 [1][6]
    - Variety 75 [1][7]
    - Variety 76 [1][36]

Show
● Qualitative results ○ Electrophoretic results

Results of qualitative comparison for the current two varieties [6]

Chara	Long name	Weighting	Note Std/Cycle 1	Note Ref/Cycle 1	Note Std/Cycle 2	Note Ref/Cycle 2
4	Green color of leaf	1,00	5	5	6	5
6	Number of lobes	0,00	5	5	4	5
11	Time of flowering	1,00	5	4	4	4
13	Length of petals	0,00	5	5	4	5
17	Height	0,00	4	5	6	5
82	Intensity of yellow color	0,00	5	6	6	5

Abbreviation
Note Std/Cycle = Score for the studied variety in cycle 1 or 2
Note Ref/Cycle = Score for the reference variety in cycle 1 or 2

Note : The characteristics with identical scores for both varieties in the 2 cycles are not displayed.

Current database = C:\ORATMP\English\

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

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

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

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

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

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

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

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

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

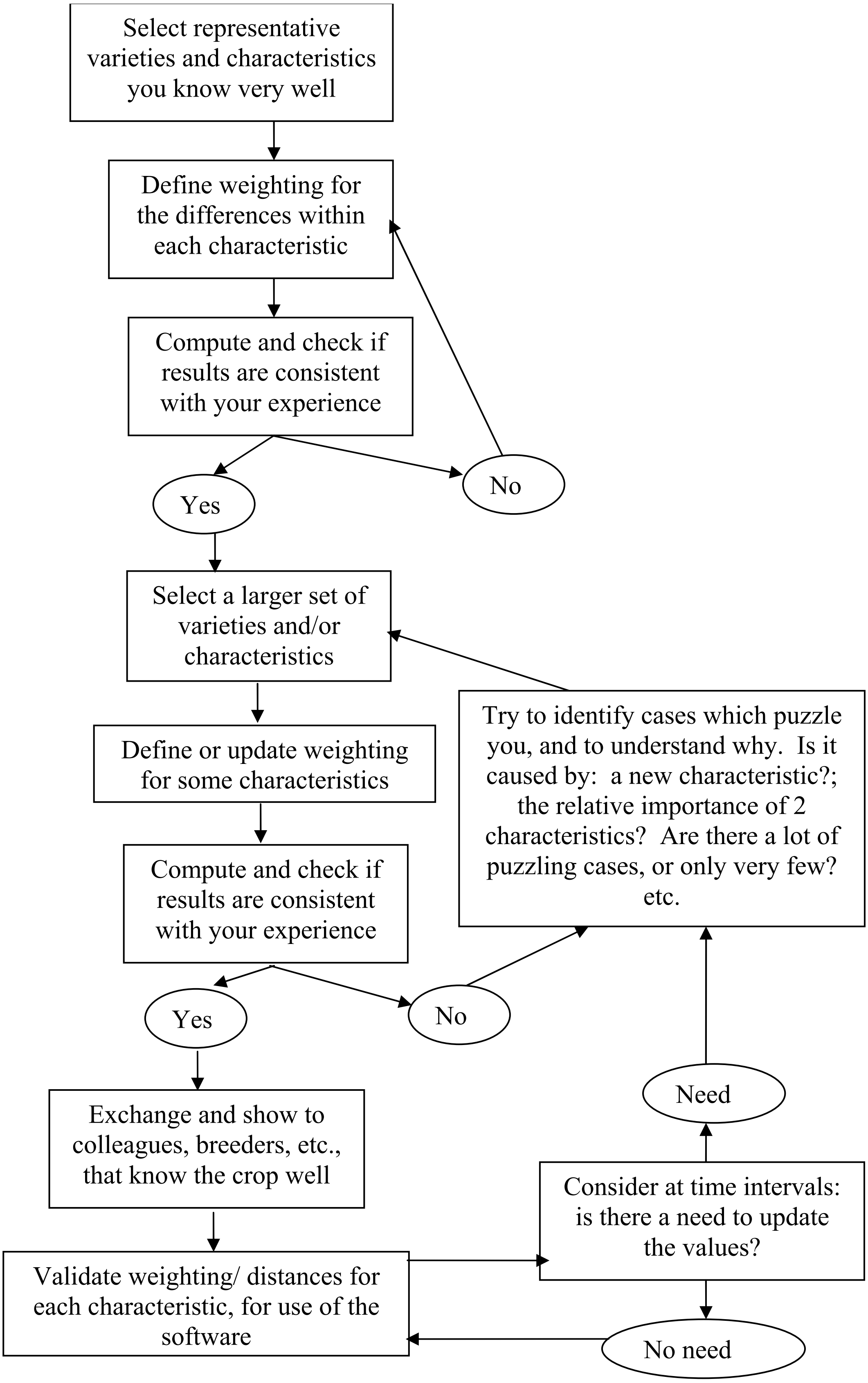
7.3.6.10 On the right of the Display tree, the raw data for *Variety 112* and *Variety 26* are visible for the 6 qualitative characteristics observed on both varieties (two cycles).

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

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

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

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



8. EXAMINING DUS IN BULK SAMPLES

8.1 Introduction and abstract

In some crops samples are bulked before certain characteristics are examined. The term “bulk sampling” is used here for the process of merging some or all individual plants before recording a characteristic. There are different degrees of bulking ranging from: 1) merging pairs of plants, 2) merging 3 or 4 up to all plants within a plot up to 3) merging all plants within a variety. The degree of bulking may play an important role in the efficiency of the tests. Bulking is usually only applied where the measurement of the characteristic is very expensive or very difficult to obtain for individual plants. Some examples are seed weight in cereals and peas and beans, and erucic acid content in rapeseed. This section describes some of the consequences of bulk sampling. It is shown that the test of distinctness (using COYD, see Part II: section 3.1 *[cross ref.]*) may be expected to be relatively insensitive to the degree of bulking, but that the efficiency of the tests for uniformity (using COYU, see Part II: section 3.2 *[cross ref.]*) must be expected to decrease when the data are bulked. The COYU test for uniformity cannot be carried out if all plants within a plot are bulked.

8.2 Distinctness

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

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

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

where

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

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

1: σ_{vy}^2 the variance component due to the year in which the variety is measured

2: σ_p^2 the variance component due to the plot in which the measurement was taken

3: σ_i^2 the variance component due to the plant on which the measurement was taken

4: σ_m^2 the variance component due to the inaccuracy in the measurement process

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

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

where

a is the number of years used in the COYD method

b is the number of replicates in each trial

c is the number of plants in each plot

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

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

where

k is the number of bulk samples

l is the number of plants in each bulk sample

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

Example 1

Variances for comparing varieties were estimated (by the use of estimated variance components) for different degrees of bulking. The calculations were based on the weight of 100 seeds of 145 pea varieties grown in Denmark during 1999 and 2000. In this example, the contribution to the variance caused by the measurement process was relatively very small, which means that bulking will have a low influence on the test for distinctness. In a 3 year test with 30 plants in each of 2 blocks, the variance of a difference between two varieties was estimated to be 2.133 and 2.135, for no bulking and a single bulk sample per plot, respectively.

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

8.2.6 In some cases each bulk sample is not drawn from a specific set of plants (say, plant 1 to 5 in bulk sample 1, plant 6 to 10 in bulk sample 2 etc.), but bulk samples are formed from mixed samples of all plants in a plot. This means that different bulk samples may contain material from the same plants. It must be expected that similar results apply here, although, in this situation, the effect of bulking may have an increased effect because there is no guarantee that all plants will be equally represented in the bulk samples.

8.3 Uniformity

8.3.1 Bulking within plot

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

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

where this variance can be regarded as being composed of two sources of variation:

1: σ_{vy}^2 the variance component due to the year in which the variety is measured

2: σ_f^2 the variance component due to the number of degrees of freedom used in estimating s_{vy}

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

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

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

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

where

a is the number of years used in the test

r is the number of reference varieties

Example 2

The effect of bulking in the test for uniformity, an estimate was made using the same data as for Example 1 I Part II, section 6.2.5 [cross ref.]. For a test using 50 reference varieties in 3 years with 30 plants per variety in each of 2 plots per trial the variance for comparing the Z_{vy} *value for a candidate variety and the mean of the reference varieties'* Z_{vy} *will be 0.0004 if no bulking is done. This can be compared to 0.0041, 0.0016 and 0.0007 when 2, 4 and 10 bulk samples per plot were used. Thus, in this example, the effect of bulking has a great influence on the test for uniformity. The variance increased, approximately by a factor of 10 when changing from individual plant records to just 2 bulk samples per plot. This means that the degree of non-uniformity must be much higher for it to be detected when 2 bulk samples are used instead of individual plant records.*

8.3.2 Bulking across plots

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

8.3.3 Taking just one bulk sample per plot

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

[End of Part II and of document TGP/8]

---

[a] TWC: to be deleted

[b] TWC proposal