Disclaimer: unless otherwise agreed by the Council of UPOV, only documents that have been adopted by the Council of UPOV and that have not been superseded can represent UPOV policies or guidance.

This document has been scanned from a paper copy and may have some discrepancies from the original document.

———

Avertissement: sauf si le Conseil de l'UPOV en décide autrement, seuls les documents adoptés par le Conseil de l'UPOV n'ayant pas été remplacés peuvent représenter les principes ou les orientations de l'UPOV.

Ce document a été numérisé à partir d'une copie papier et peut contenir des différences avec le document original.

———

Allgemeiner Haftungsausschluß: Sofern nicht anders vom Rat der UPOV vereinbart, geben nur Dokumente, die vom Rat der UPOV angenommen und nicht ersetzt wurden, Grundsätze oder eine Anleitung der UPOV wieder.

Dieses Dokument wurde von einer Papierkopie gescannt und könnte Abweichungen vom Originaldokument aufweisen.

———

Descargo de responsabilidad: salvo que el Consejo de la UPOV decida de otro modo, solo se considerarán documentos de políticas u orientaciones de la UPOV los que hayan sido aprobados por el Consejo de la UPOV y no hayan sido reemplazados.

Este documento ha sido escaneado a partir de una copia en papel y puede que existan divergencias en relación con el documento original.

UPOV

TC/XX/5
ORIGINAL: English
DATE: August 1, 1984

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

TECHNICAL COMMITTEE

Twentieth Session
Geneva, November 6 and 7, 1984

THE CASE FOR ADOPTING THE OVER-YEARS CRITERION FOR ASSESSING DISTINCTNESS

Document prepared at the request of the Technical Working Party on Automation and Computer Programs by its Chairman

Background

At its first meeting, in May 1983, the Technical Working Party on Automation and Computer Programs agreed that 'the combined over-years criterion for assessing distinctness seemed to be the most satisfactory, would lead to better discrimination and would diminish the risk of establishing a difference which did not exist'. The Technical Committee requested a detailed explanation of the combined over-years analysis (Report of meeting November 1983, paragraph 34). Dr. Weatherup's paper (the Annex to this document) describes this analysis and evaluates in detail the use of the over-years criterion in comparison with the present UPOV distinctness criterion and the t-score. The following comments summarise the case for adoption of the over-years criterion. Tables referred to are contained in the Annex to this document (identical to document TWC/II/5).

Present Criterion (2 x 1% rule)

The present UPOV criterion for distinctness requires that the difference between varieties should be statistically significant at the 1% probability level in at least 2 years out of 3 for one or more important characteristics. This is called the 2 x 1% rule and the technical experts who have accepted it recognized the importance of seeking repeatability of significant variety differences over years.

The 2 x 1% rule has been criticized because a difference between varieties which just fails to achieve the 1% significance level contributes no more to the assessment of distinctness than, for instance, a zero difference in one year. If the differences between varieties are in the same direction for each of 3 years with one difference significant at the 1% level and two significant at the 5% level this would not, under the present 2 x 1% rule, be sufficient evidence of distinctness. An intuitive interpretation of three such results suggests that the varieties are distinct and that the information from all three years should be taken into account.

t-score method

The t-score method goes some way to meeting the above criticism by providing a system of calculation which uses results from all three years, but does not allow one very extreme result out of three to out-weigh the evidence from the other two years. Table 3 shows an example of a case where the t-score confirms the intuitive view of the results expressed above.

Risks of wrong decisions

Both the 2 x 1% and t-score systems assess differences between varieties against the plot error variation within trials. This does not take account of variety variation over years. It can be shown that some variety plant characters are much less consistent than others over time. Where decisions on distinctness are based on characters which are consistent from year to year (low values of λ in Table 2), there is relatively low risk that repeated tests in a later year would lead to a different decision. Decisions based on less consistent characters have a greater risk that they could not be reproduced. When either the 2 x 1% rule or the t-score is used the risks involved in making decisions are of varying magnitude depending on the consistency over years of the character on which the distinctness decision is based. This is unsatisfactory for both testing authority and breeder.

Use of the over-years criterion, which measures variety differences against variation over years, produces an equal risk no matter which characters provide the basis for decisions on distinctness. Using the proposed over-years criterion the testing authority would, irrespective of characters used in decisions, accept a 1% risk of wrongly declaring distinct two varieties which are in fact not distinct. Using the 2 x 1% rule this risk could be less than 1% for characters which are consistent over years and as high as 9% if based on the least consistent character (see Tester's risk Table 2).

An F-ratio test statistic is included in the over-years criterion to identify any excessive variation in the differences between a candidate variety and a control compared with a general variety x years variation over the 3 years of test. A significant F-ratio might indicate an unstable candidate variety.

Conclusion

The over-years criterion has the following desirable attributes conducive to more reliable decision taking:

- by definition it provides a check on reproducibility of variety differences over years
- it uses all available information contained in the data from all the years
- the risks of making wrong decisions are constant for all characters.

The Technical Working Party on Automation and Computer Programs at its meeting in La Minière this year agreed in principle that the over-years criterion was the most suitable of the distinctness criteria available. The Technical Working Party recommended that member States evaluate it further during the coming year so that detailed proposals for its implementation can be agreed at its next meeting for submission to the Technical Committee.

[Annex follows]

TC/XX/5

ANNEX

Description and evaluation of the combined over-years criterion for distinctness between varieties of herbage crops

by S T C Weatherup

Biometrics Division, Department of Agriculture, Northern Ireland, UK

Introduction

At the first meeting of the UPOV Technical Working Party on Automation and Computer programs it was agreed "that the combined over-year analysis seemed to be the most satisfactory, would lead to better discrimination and would diminish the risk of establishing a difference which did not exist". In the present note the current and the combined over-years criteria for distinctness are described and compared on actual trial data with a view to adding to the understanding of the combined over-years criterion and to showing how it improves on the present criteria. The methods are compared using data from a series of Perennial Ryegrass (Diploid) trials carried out at Crossnacreevy, UK, during 1979-81. Table 1 presents a list of the characters measured in this trial series. Further information on the details of the combined over-years criterion can be found in Patterson H D and Weatherup S T C (1984).

Test criteria

(a) 2 x 1% method

The present criterion for distinctness of varieties is based on separate t-tests in 3 years between a candidate variety and each other variety. These use the t-values defined by

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{2}\ SE(\bar{x})} \qquad \ldots\ldots (1)$$

where $\bar{x}_1$, $\bar{x}_2$ are the means on a measured character of the two varieties being compared and $SE(\bar{x})$ is the standard error of a variety mean estimated from the plot error mean square. The significance level of a calculated t-value can be

obtained from t-tables using the degrees of freedom of the plot error mean square. For each comparison between a candidate variety and some other variety 3 such tests are available from the 3 years that the candidate variety was grown in the trials. Under this criterion the candidate variety is considered to be distinct from another variety if in at least 2 out of the 3 years the t-value is significant at the 1% level in the same direction.

(b) t-score method

The 2 x 1% method can be criticized on the grounds that a within-year difference which just fails to achieve the 1% significance level contributes no more to the separation of a variety pair than a zero difference or even a non-significant difference of the opposite sign. For example, 3 differences in the same direction one significant at the 1% level and the others at the 5% level would not be regarded as sufficient evidence for distinctness. The t-score method was introduced to overcome this weakness. In this method the t-value calculated according to formula (1) is converted to a t-score using constants k_1 and k_2 where these are the tabulated t-scores at the 5% and 0.1% levels respectively. The conversion from t-values to t-scores is then defined by:

(i) t-score = 0 if $-k_1 < t < k_1$

(ii) t-score = t-value if $k_1 \leq t \leq k_2$ or $-k_2 \leq t \leq -k_1$

(iii) t-score = k_2 if $t > k_2$

(iv) t-score = $-k_2$ if $t < -k_2$

These conversion rules are illustrated in Figure 1. Using t-scores two varieties are distinct if the absolute sum of their t-scores over 3 years exceeds the critical value 5.2, which is equivalent to twice the tabulated t value at the 1% significance level with a large number of degrees of freedom. Like the 2x1% criterion the t-score method requires more than a single large t-value for distinctness but the confirmatory evidence need not be so strong. Thus three 5% results, provided they are in the same direction are sufficient to ensure distinctness.

Where a testing authority has a second trial site an additional condition is imposed which requires that if a t-score in excess of 5.2 has been obtained at one site sufficient corroborative evidence must be provided from the second site to make the total t-score 8.5 over both for distinctness to be achieved.

(c) Combined over years criterion

Both the previous criteria are based on the plot error mean square and hence do not take into account variety variation over years. Since variety measurements are less consistent on some characters than on others the use of these criteria results in acceptance standards varying between characters. Inconsistency is indicated by large values of λ where

$$\lambda^2 = \frac{\text{varieties x years mean square}}{\text{plot error mean square}}$$

and consistency by values of λ near to 1. Values of λ from a specimen set of trial data are given in Table 1 and show a range from 1.21 to 2.41. The over years criterion meets the above criticism by using the variety x years mean square to estimate error instead of the plot error mean square. Two varieties are distinct if the absolute value of

$$\frac{\bar{\bar{x}}_1 - \bar{\bar{x}}_2}{\sqrt{2}\ SE(\bar{\bar{x}})}$$

is larger than the critical 1% point in the distribution of t where $\bar{\bar{x}}_1$, $\bar{\bar{x}}_2$ are the means of the varieties over 3 years and $SE(\bar{\bar{x}})$ is the standard error of a variety mean calculated from the variety x years mean square.

Using this criterion the ratio, F, defined as

$$F = \frac{\text{varieties mean square}}{\text{varieties x years mean square}}$$

provides a measure of the discriminating power of a character. Thus Table 1 shows that characters 11 and 20 are the most discriminating characters and 4 and 19 are the least discriminating.

Comparison of the acceptance probability criteria

A convenient measure of the effectiveness of the 3 methods considered above is the probability of declaring distinct two hypothetical varieties whose true difference is known on some particular character. The probability is called the acceptance probability. The simplest case is where the true difference is assumed to be zero on some character ie. the varieties are identical on this measure. Due to sampling errors the actual differences between the means of these varieties obtained in trials will not be zero but will be distributed about zero with a variance which can be calculated from trial data over 3 years. Using this distribution the acceptance probability of a candidate variety can be calculated. In this case since this is the probability of accepting a non-distinct variety, it is known as the Tester's Risk. Values for these risks for the 3 criteria on the characters measured in the specimen trial are also provided in Table 1. These show that for both the 2x1% and the t-score criteria there are marked differences between the characters in their Tester's Risks depending upon their consistency. Thus for character 5 which exhibited inconsistency over years (ie. large λ) the probability of accepting a variety whose true mean is identical to that of another variety is 9.2% and 14.1% for the 2x1% and t-score criteria respectively. Also for character 11 which was consistent over years (ie small λ) the corresponding acceptance probabilities are only 0.2% and 0.5%. By contrast the acceptance probabilities of the combined over-years criterion remains constant at 1% for all characters when the true difference is zero.

The same approach can be used to determine acceptance probabilities at other assumed true differences to produce a curve for a given value of λ of acceptance probability against true difference. Such a curve is known as the operating characteristic, OC, of the character. Curves are provided in Figure 2 for the three criteria considered above giving acceptance probabilities against standardised difference for λ values of 1, 2 and ∞. The slope of an OC at midsection provides

a measure of the efficiency of the testing scheme; the steeper the slope the more efficient the scheme. Thus it can be seen that the t-score method is more efficient than the 2x1% criterion but is also less stringent having higher acceptance probabilities at a given λ. When λ is increased the 2x1% and t-score OC curves become similar and less steep. Efficiency is then poor and the Tester's Risk is very high. Thus although these methods both require consistency over years the standard set is low and bears no relation to the 1% standard adopted for variation between plots. By contrast the OC of the combined over years criterion remains constant for all values of λ giving the same probability of acceptance of any defined true difference regardless of the consistency of the character.

Homogeneity of varieties x years variance

From the previous paragraphs it is recommended that the combined over years criterion be used in preference to the 2x1% or the t-score methods with one qualification. The varieties x years mean square used in the combined over years criterion is a pooled value calculated from a large number of varietal comparisons and hence may not be appropriate to any particular comparison. Hence to ensure that the specific within pair varieties x years mean square with 2 degrees of freedom is not larger than the pooled varieties x years mean square the ratio, F_2 say, should be calculated and tested for significance. The combined over years criterion should be treated with caution where a significant F_2 ratio occurs.

Examples

In many cases there is agreement between the 2x1% and over-years test results but there are instances in which conclusions differ. Examples from the specimen data are in Tables 2, ..., 5. The following are some comments on these.

(a) Table 2. The t-values for AvB on character 20 are significantly different at the 1% level in only 1 year out of 3 and so the varieties are not judged

distinct on this character using the 2x1% criterion. The value of λ for character 20 is small and hence distinctness is difficult to achieve on the 2x1% criterion. One of the t-values was significant at the 0.1% level; although not achieving the high standard required by the 2x1% criterion the other two provide strong confirmatory evidence. The 1% over-years criterion is easily met and the value of F_2 is small. The t-score criterion is also satisfied. It is concluded that the varieties are distinct with variety A producing taller plants than variety B in the aftermath.

(b) Table 3. Neither the 2x1% criterion nor the t-score criterion provide evidence of distinctness between varieties C and D on character 11. However, the over years criterion indicates distinctness and F_2 is smaller than 1 showing no excessive variety pair x year variation. Only one of the individual t-values attains significance at the 1% level but the other two provide confirmatory evidence. The 2x1% criterion ignores the significant 1981 t-value. Again λ is small and it seems reasonable to conclude that the two varieties are distinct.

(c) Table 4. Differences EvF on character 5 are significant at the 1% level in 2 years and hence the varieties are distinct on the 2x1% and t-score criteria but this conclusion is not supported by the over-year analysis. The λ value for character 5 is large, and in consequence the 2x1% and t-score tests are unusually relaxed.

(d) Table 5. All three criteria lead to the conclusion that variety G is distinct from variety H because its plants have longer stems at 30 days after ear emergence (character 17). Year-to-year consistency suggests, however, that the conclusion may be wrong. Results in 1979 contradict the results of 1980 and 1981 and the F_2 ratio is significant at the 1% level.

Multivariate criterion

None of the previous criteria take into account the substantial correlations that exist between the characters. This deficiency can be remedied by using the Mahalanobis generalised distance, D^2 as a measure of distinctness; this distance is defined for a pair of varieties as $D^2 = d^T W^{-1} d$ where d is the vector of differences between the over-year means of the variety pair for all characters, d^T is its transpose and W is the covariance matrix calculated from the varieties x years mean squares and cross products for all characters. Thus the matrix W is the multivariate analogue of the varieties x years mean square used in the over-years criterion.

The critical value for D^2 is given by

$$\frac{2p(m-1)(n-1)}{m(mn-m-n-p+2)} F$$

where m is the number of years, n is the number of varieties, p is the number of characters and F is the F ratio at the 1% significance level with p and mn - m - n - p + 2 degrees of freedom.

Table 7 provides t-values of a comparison between a pair of varieties, X and Y, on a complete set of characters. It can be seen that the 2x1% criterion was not met on any single character. Although the t-score criterion was achieved on character 4 this was not sufficient for distinctness in this case as the varieties were being compared at 2 sites and no supporting evidence was available at the second site on this character. However the D^2 value was 36.43 which easily met the critical D^2 value of 21.5 for this set of trials and hence on a multivariate basis the varieties could be considered distinct. An examination of the signs of the individual t-values shows consistent trends across many of the characters viz. 4, 5, 14, 17, 20 and 24. The first two results indicate a tendency for variety X to be more upright and taller in the spring than variety Y, while results on characters 14, 17 and 24 point to X having shorter leaves than Y.

Conclusion

It is recommended that the 2x1% and t-score criteria are replaced immediately by the combined over years criterion to provide distinctness decisions which being based on over year variation are more likely to remain true in the long term. Further to distinguish between other variety pairs for which there is no single separating character but which demonstrate several consistent differences on related characters it is recommended that Mahalanobis D^2 distances are also used as a distinctness criterion.

References

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

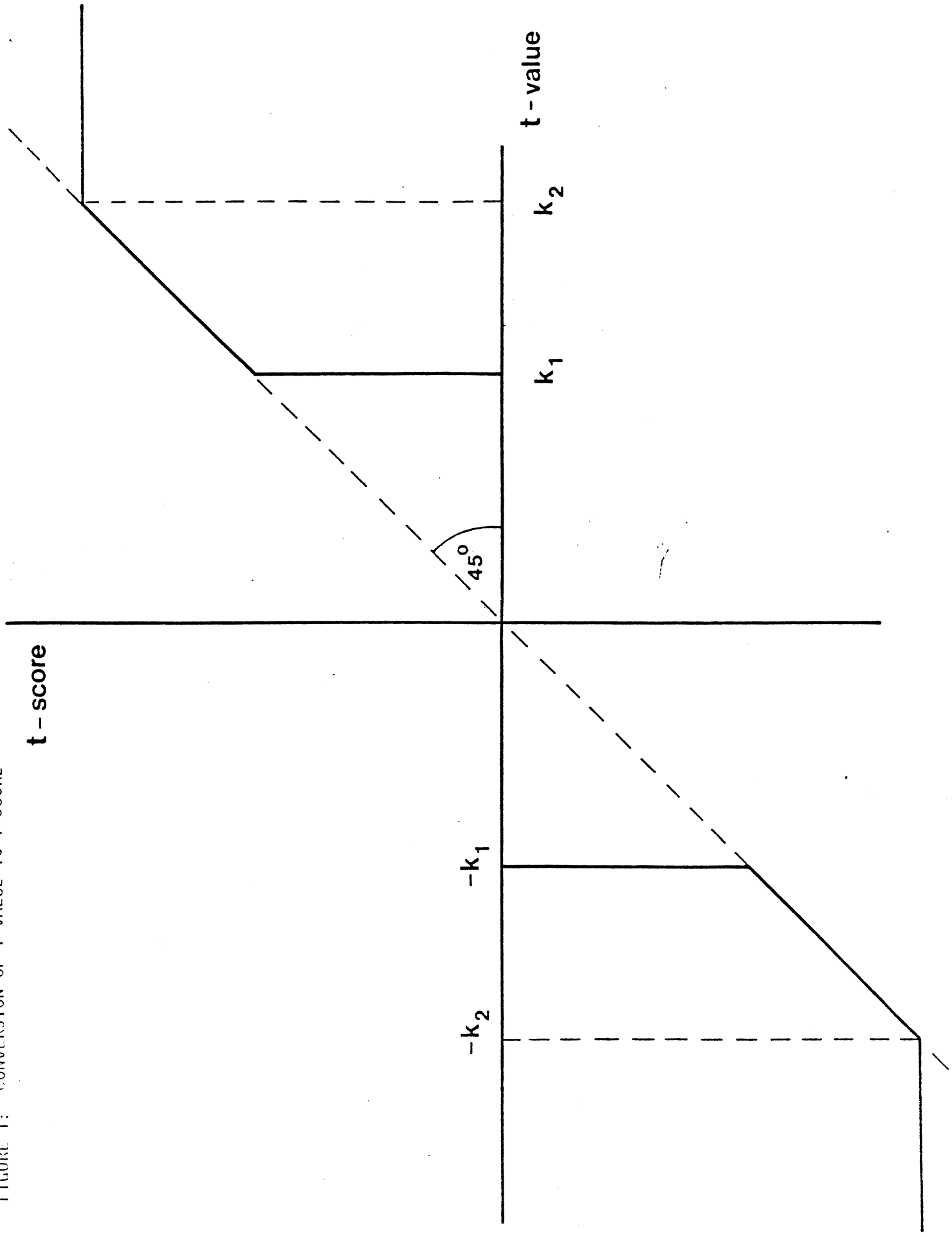

FIGURE 1: CONVERSION OF T-VALUE TO T-SCORE

FIGURE 2: OPERATING CHARACTERISTICS FOR DISTINCTNESS CRITERIA

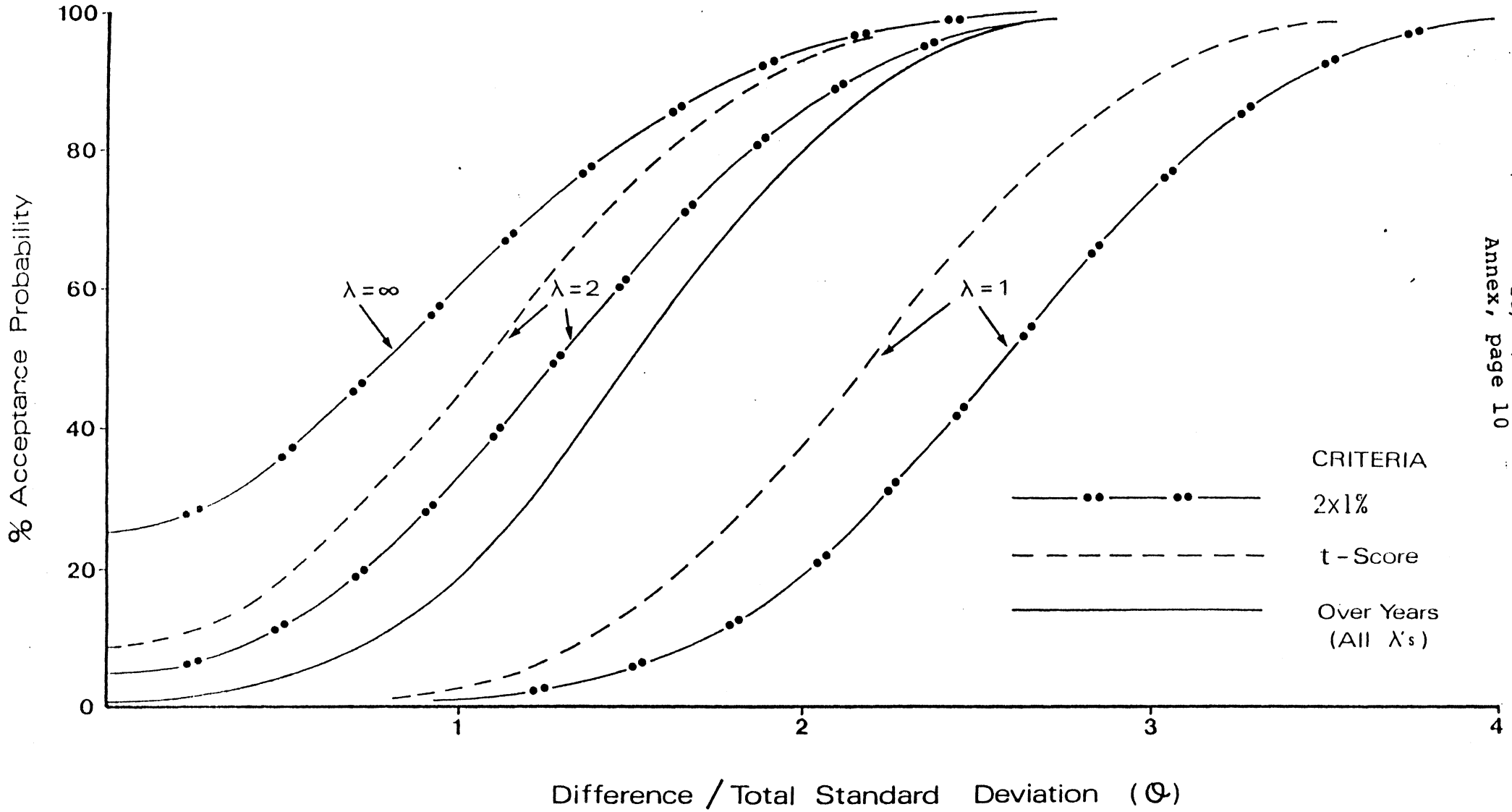

TABLE 1

Definitions of measured characters

Character Number	Definition
4	Angle of growth in year of sowing (deg)
5	Height of pulled up leaves measured in the spring (cm)
8	Date of ear emergence (days from 1 March)
10	Natural plant height at date of ear emergence (cm)
11	Plant width at date of ear emergence (cm)
14	Length of flag leaf at ear emergence (cm)
15	Width of flag leaf at ear emergence (mm)
17	Stem length 30 days after ear emergence (cm)
*19	Number of heads/plant estimated on 0-9 scale
*20	Height of plant in aftermath (cm)
24	Ear length (cm)

*All plants of each variety are cut down at a defined time relative to their recorded date of ear emergence. Characters 19 and 20 are measured on the plant regrowth 8 weeks after cutting.

TABLE 2

Statistics derived from specimen trial data

(Early Perennial Ryegrass (Diploid) trials Crossnacreevy, UK, 1979-81)

Character	Mean squares/plot						Tester's risks (%)		
	Years (Y)	Varieties (V)	Years x Varieties (VxY)	Plot error (E)	$F_1 = V/(VxY)$	$\lambda = \sqrt{(VxY)/E}$	2x1%	t-score	over-years
4	570.34	214.03	28.41	15.29	7.5	1.36	0.5	1.3	1.0
5	2678.61	351.99	25.50	4.44	13.8	2.41	9.2	14.1	1.0
8	32317.11	351.85	18.84	5.21	18.7	1.90	3.9	7.2	1.0
10	940.04	777.98	44.72	12.79	17.4	1.87	3.6	6.8	1.0
11	2461.98	430.37	19.18	13.16	22.4	1.21	0.2	0.5	1.0
14	30.79	61.95	3.83	1.35	16.2	1.68	2.1	4.4	1.0
15	0.02	2.30	0.18	0.11	13.0	1.25	0.2	0.7	1.0
17	1364.30	533.30	32.91	16.06	16.2	1.41	0.6	1.6	1.0
19	10.96	3.91	0.41	0.17	9.6	1.53	1.2	2.7	1.0
20	3327.77	396.55	19.55	12.96	20.3	1.23	0.2	0.6	1.0
24	412.41	38.84	3.13	1.60	12.4	1.40	0.6	1.6	1.0

TABLE 3

Variety pair AvB

Character 20 (λ = 1.23)

	Year			
Variety	1979	1980	1981	Mean
A	33.9	41.3	36.1	37.1
B	26.6	36.1	31.0	31.2
SE	1.25	1.69	1.43	1.04
t-values	4.13**	2.18*	2.52*	3.98‡

2x1%	2.59
t-score	8.01‡
Over years	3.98‡
F_2	0.23

* Significant at 5% level

** Significant at 1% level

‡ Distinctness criterion achieved

TABLE 4

Variety pair CvD

Character 11 (λ = 1.21)

Variety	Year 1979	1980	1981	Mean
C	58.3	63.1	56.5	59.3
D	52.3	58.7	52.3	54.4
SE	1.40	1.50	1.53	1.04
t-values	3.03**	2.07*	1.94	3.33‡

2x1%	2.59
t-score	8.10‡
Over years	3.33‡
F_2	0.15

* Significant at 5% level

** Significant at 1% level

‡ Distinctness criterion achieved

TABLE 5

Variety pair EvF

Character 5 (λ = 2.41)

	Year			
Variety	1979	1980	1981	Mean
E	27.1	35.4	37.4	33.3
F	26.6	31.9	32.4	30.3
SE	0.70	0.86	0.99	1.19
t-values	0.51	2.88**	3.57**	1.78

2x1%	5.18 ‡
t-score	6.19 ‡
Over years	1.78
F_2	0.61

* Significant at 5% level

** Significant at 1% level

‡ Distinctness criterion achieved

TABLE 6

Variety pair EvF

Character 17 (λ = 1.41)

	Year			
Variety	1979	1980	1981	Mean
E	85.2	87.0	88.9	87.0
F	87.6	76.8	76.7	80.4
SE	1.37	1.72	1.78	1.34
t-values	-1.24	4.19**	4.85**	3.49‡

2x1%	5.18‡
t-score	6.63‡
Over-years	3.49‡
F_2	6.54**

* Significant at 5% level

** Significant at 1% level

‡ Distinctness criterion achieved

TABLE 7

t-values of comparisons between varieties X and Y over all characters for 3 years

(t-values positive if X is larger than Y)

Character	Year 1981	1982	1983	t-score	Over years
4	2.40*	4.12***	0.11	5.77	*
5	0.09	2.04*	0.31	2.04	NS
8	-0.25	-3.78***	-1.26	-3.37	NS
10	-0.72	0.59	-1.95	0.00	NS
11	-0.54	0.46	1.01	0.00	NS
14	-1.96	-3.20**	-1.26	-3.20	*
15	1.40	0.31	2.28*	2.28	NS
17	-0.67	-0.29	-0.68	0.00	NS
19	0.34	-0.49	0.80	0.00	NS
20	1.93	0.56	2.00*	2.00	NS
24	-1.83	-0.73	-0.77	0.00	NS

Mahalanobis generalized distance, D^2 = 36.43***

* P<0.05

** P<0.01

*** P<0.001

[End of Annex and of document]