E

TWC/30/19
ORIGINAL: English
DATE: June 14, 2012

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
Geneva

TECHNICAL WORKING PARTY ON AUTOMATION AND COMPUTER PROGRAMS

Thirtieth Session
Chisinau, Republic of Moldova, June 26 to 29, 2012

CONSEQUENCES OF DECISIONS FOR DUS EXAMINATION WHEN USING STATISTICAL METHODS FOR VISUALLY OBSERVED CHARACTERISTICS

Document prepared by experts from Denmark and Poland

BACKGROUND

1. The Technical Committee (TC), at its forty-eighth session, held in Geneva from March 26 to 28, 2012, considered the revision of document TGP/8 "Trial Design and Techniques Used in the Examination of Distinctness, Uniformity and Stability" on the basis of document TC/48/19 Rev. The TC noted that new drafts of relevant sections would need to be prepared by April 26, 2012, in order that the sections could be included in the draft to be considered by the TWPs at their sessions in 2012 (see document TC/48/22 "Report on Conclusions" paragraph 49).

2. The TC, at its forty-eighth session, agreed that the section "Statistical methods for visually observed characteristics" of document TGP/8 should be redrafted with assistance from DUS experts in Denmark in order to focus on guidance for DUS examiners and should replace detailed statistical models with a general reference to suitable statistical methods. The TC agreed that the examples based on sugar beet should be replaced by a crop for which there are Test Guidelines and that the example for wheat should be replaced by a realistic example, such as could be found in Hemp or Spinach. The TC also agreed that the TWC should explore the consequences of the decisions for DUS examination, because the method is a test for differences in the distribution (both location and dispersion). It also agreed that the consequences of excluding certain varieties from the test, where there were insufficient numbers in some cells, should be further investigated (see document TC/48/22 "Report on Conclusions" paragraph 61).

3. A proposed text for a new section of document TGP/8 "Statistical Methods for Visually Observed Characteristics" has been prepared by an expert from Denmark as document TWC/30/29 "Revision of document TGP/8: Part II: Techniques used in DUS examination: New section – Statistical methods for visually observed characteristics".

4. In addition to that document, the Annex to this document contains supplementary information concerning consequences of the decisions for DUS examination as background information for consideration when discussing document TWC/30/29, at its thirtieth session, held in Chisinau, Republic of Moldova, from June 26 to 29, 2012.

[Annex follows]

INTRODUCTION

5. The methods that have been suggested for testing for distinctness in visually observed characteristics are based on the distribution of the data. This applies to methods that are based on the multinomial distribution, i.e.:

- The generalized linear mixed model for nominal characteristics using the generalised logit as link function
- The generalised linear mixed model for ordinal characteristics using the cumulative logit as link function
- The χ^2-test used for both nominal and ordinal characteristics
- The analysis of each characteristic using the generalized linear mixed model using the logit as link and assuming each characteristic to be binomial distributed
- The analysis of each characteristic using the present COY-D method for each note after an appropriate transformation

PROBLEMS

Uniformity

6. As an example we consider some artificial data for a characteristic such as intensity of anthocyanin coloration on coleoptiles for varieties in winter wheat are recorded on an ordinal scale (table 1).

Table 1 True percentage of individual plants with each note for a hypothetical characteristic recorded on the ordinal scale

Variety	Note					
	1 very weak	2 weak	3 medium	4 strong	5 very strong	Total
1	80.0	16.0	3.8	0.1	0.1	100
2	2.0	8.0	80.0	8.0	2.0	100
3	0.1	1.9	8.0	80.0	10.0	100
4	60.0	20.0	14.0	5.9	0.1	100
5	5.0	15.0	60.0	15.0	5.0	100
6	3.0	7.0	10.0	60.0	20.0	100

7. In the example here the data are constructed such that variety 1, 2 and 3 are more uniform than variety 4, 5 and 6. From the data is seen that variety 1, 2 and 3 are expected to be judged uniform and distinct. Variety 1 may be considered to be not distinct from variety 4, and that variety 4 to be less uniform than variety 1. Similarly, variety 2 and 5 may be considered to be not distinct and variety 5 to be less uniform than variety 2 and similarly variety 3 and 6 may be considered to be non distinct and variety 6 to be less uniform than variety 3.

8. If 100 observations were sampled from each of these varieties in two years (with some interaction between variety and year) and the data were analysed using a generalised mixed model varieties 1-3 are expected to be distinct from each other whereas the variety pairs 1-4, 2-5, 3-6 may should not be considered distinct, but may very well be so. A simulation study (1000 simulations) and the analysis of each simulation (6 varieties $\times$ 2 years $\times$ 100 plants) showed that the variety pair 1-4 became significant in more than 50% of the cases (table 2). Variety pair 2-5 and 3-6 was only significant in a few cases which both were less than the expected number. However, if the same distribution was assumed for a nominal characteristic all three pairs (1-4, 2-5 and 3-6) became significant in about 70 % of the cases. Using a χ^2-test, which are the same for both ordinal and nominal scaled characteristics those three pairs (1-4, 2-5 and 3-6) became significant in about 95 % of the cases. Also the methods of analysing each note separately are identically for both ordinal and nominal scaled characteristics. When each note were analysed separately (either assuming Binomial distributed data or normal distributed data (after arc-sinus-sqrt transformation) characteristics those three pairs (1-4, 2-5 and 3-6) became significant in about 80-90 % of the cases. If the tests were corrected for multiple tests (here 5 tests using Bonferroni's method) the relative number of significant pairs were reduced to about 50-70 percent (table 2).

Table 2 Percent of significant (α=0.05) differences between selected variety pairs for 1000 simulations

Analysis method	Variety pairs					
	1-2	1-3	2-3	1-4	2-5	3-6
GLIMM ordinal	100.0	100.0	99.9	54.6	1.4	3.8
GLIMM nominal	99.2	99.6	99.0	72.0	70.1	65.7
χ^2 test for independence	100.0	100.0	100.0	94.6	94.4	95.9
Binomial Uncorrected	99.2	97.6	100.0	83.1	87.7	90.7
Binomial Corrected[2]	98.6	91.6	100,0	50.1	61.9	69.5
Normal Uncorrected[1]	100.0	100.0	100.0	89.0	89.3	88.4
Normal Corrected[12]	100.0	100.0	100.0	64.0	57.9	57.7

[1)] After that transformation of relative figures using the arc-sin-square-root transformation
[2)] Corrected for multiple tests (one test for each of five notes using Bonferroni's method)

Distribution "variability" depends on where the variety are located on the scale and how the characteristic is constructed

9. Assume that the notes (ordinal) can be regarded to be the result of an underlying unknown continuous variable and that the recorded notes depend on some borders (threshold) on the unknown continuous variable. Assume that the unknown continuous variable runs from about 1 to about 100 and that the notes 1-5 are recorded as follows:

- The note 1 is recorded if the value is less than 10
- The note 2 is recorded if the value is between 10 and 20
- The note 3 is recorded if the value is between 20 and 35
- The note 4 is recorded if the value is between 35 and 60
- The note 5 is recorded if the value is larger than 60

10. In practice we do not know the thresholds, but they are defined indirectly by the definition of the notes.

11. The value on this unknown continuous variable is assumed to be normally distributed with a variety specific mean, μ_v and a variety specific standard deviation, σ_v. As an example we consider 7 varieties with different means and standard deviations (table 3).

Table 3 Assumed means and standard deviation on the continuous scale for 6 varieties

Variety	A	B	C	D	E	F	G
mean, μ_v	5	20	27.5	80	5	20	80
standard deviation, σ_v	4	4	4	4	8	8	8

12. From this we can calculate the distribution of notes for each of the 7 varieties (table 4). The table show that the apparent distribution over the notes depends not just on the standard deviation on the unknown continuous variable. Additionally in table 4 another measure of variation (in form of the so-called coefficient of concentration) is given. More details about it are given in APPENDIX 1. As an example variety A and C seems to be more uniform than variety B. The reason for that is mainly that the mean value of variety B is located just at the border between two notes and therefore most of the observations fall in the two notes on each side of the border whereas the mean value variety A and C is located half way between two borders and therefore most of the observations fall in the note defined by those two borders. Variety D, seem to be much more uniform than variety A and both are located about half way between two borders. The reason that variety D looks more uniform than variety A is mainly that variety D belongs to a note that covers a larger range on the unknown continuous variable than variety A.

Table 4 True percentage of individual plants with each note

Variety	Note						Std. Dev. on Note[a]	Coefficient of concentration, h[b]
	1	2	3	4	5	Total		
A	89.44	10.56	0.01	0.00	0.00	100	0.31	0.24
B	0.62	49.38	49.99	0.01	0.00	100	0.52	0.63
C	0.00	3.04	93.92	3.04	0.00	100	0.25	0.15
D	0.00	0.00	0.00	0.00	100.00	100	0.00	0.00
E	73.40	23.56	3.03	0.01	0.00	100	0.52	0.51
F	10.56	39.44	46.96	3.04	0.00	100	0.72	0.77
G	0.00	0.00	0.00	0.62	99.38	100	0.08	0.02

[a] Approximate as it assume interval scaled. Based on 100 observations per variety
[b] For calculation see Appendix 1

13. Variety A, B and D all seem more uniform than E, F and G, respectively. This is as expected as they have the comparable mean value on the unknown continuous variable but different standard deviation.

14. It should be noted that variety G seems more uniform than variety A, B and C even variety G has a larger standard deviation on the unknown continuous variable than variety A, B and C. The reason is mainly that variety G is located in the centre of a note that covers a larger range on the unknown continuous variable whereas the varieties A, B and C are located in notes that have a shorter range on the unknown continuous variable – an for variety B also at the border between two notes.

15. The two measure of uniformity ranked the varieties the same way except that variety B and E had the same value when using standard deviation while variety B were judged to be more uniform than variety E when using the coefficient of concentration.

16. In order to further illustrate this dependence between standard deviation and the mean of the notes, the expected value of mean note and mean standard deviation was calculated for the each whole number on the continuous underlying (latent) variable. This is done here – even the condition for calculation both mean and standard deviation are not fulfilled – as approximate way to show that a measure of homogeneity will depend not just on the variety, but also where it is located on this continuous scale. Both the expected mean value and the standard deviation were calculated under the assumption that 100 plants were recorded (visually accessed). The results are shown in figure 1.

17. The results clearly show that standard deviation under the assumption clearly depends on the mean value of the note and especially how far the mean value is from a threshold value and the width of the note on the underlying continuous variable, meaning that the standard deviation is expected to depend indirectly on how the notes are defined. The standard deviation on the note also depends on the standard deviation on the underlying scale – especially where the threshold on the underlying scale is relatively close.

18. In order to see if such relationship exists for real data the same measurements of standard deviation, coefficient of concentration and mean scores were calculated for some characteristics for wheat (Table 5).

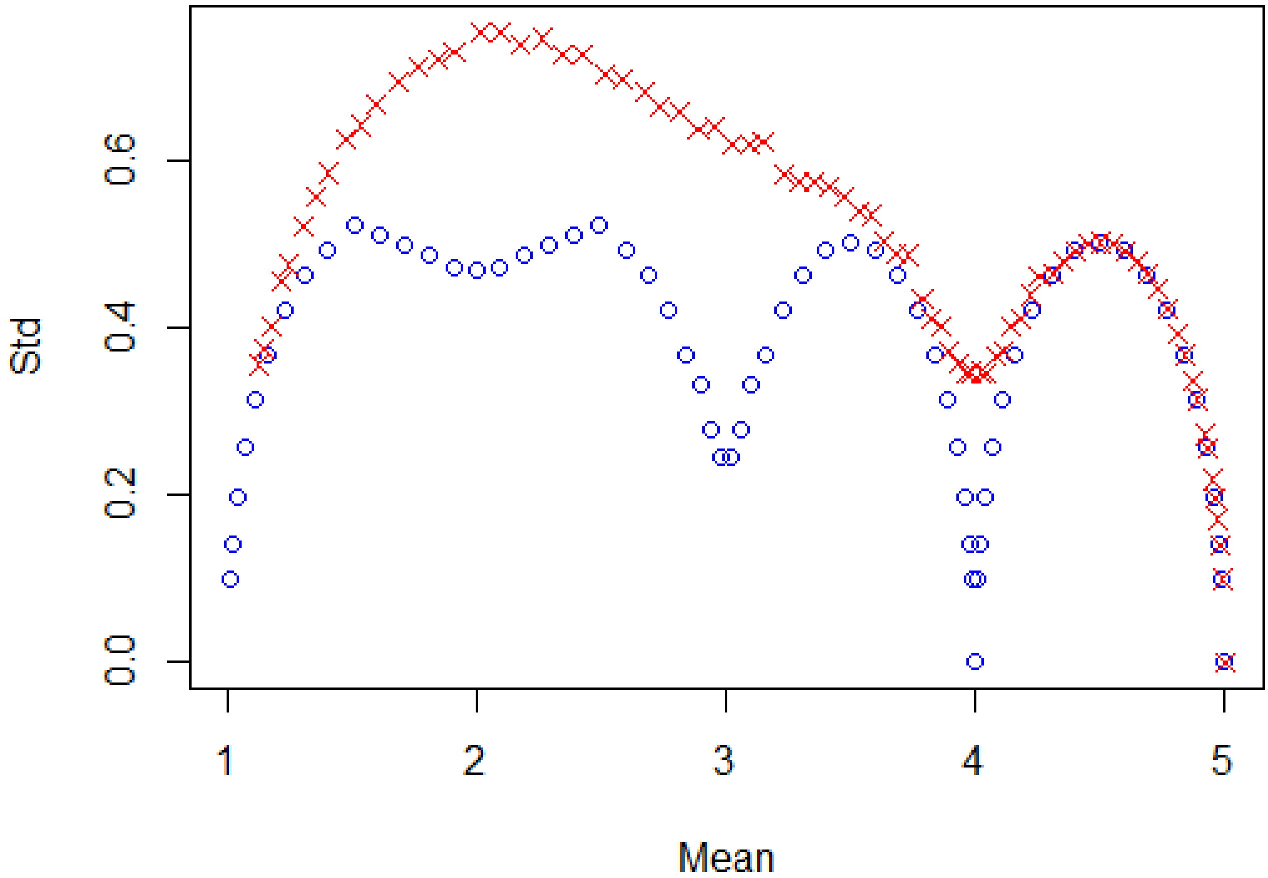


Figure 1 Relation between the standard deviation and mean of notes using the threshold stated above (Red crosses: Std. on the underlying continuous variable is 8. Blue circles: Std. on the underlying continuous variable is 4.)

Table 5 List of characteristics shown in figure 2 together with applied symbol and average standard deviation within varieties

UPOV no	Description	Symbol in figure 2 to 4	Average standard deviation[a]	Average coefficient of concentration	Applied notes
12	Ear: Density	⊠ (cyan)	0.33	0.18	2, 3, 4 ,5 ,6 ,7 ,8
15	Awns of scours at tip of ear: Length	● (grey)	0.26	0.20	3, 4 ,5 ,6 ,7
17	Apical rachis segment: Hairiness of convex surface	◆ (red)	076	0.61	1, 2, 3, 4 ,5 ,6 ,7, 8, 9
18	Lower glume: Shoulder width	● (green)	0.41	0.26	3, 4, 5, 6, 7
19	Lower glume: Shoulder shape	● (blue)	0.59	0.35	3, 4, 5, 6, 7
20	Lower glume: Beak length	○ (cyan)	0.35	0.20	1, 2, 3, 4, 5, 6, 7
21	Lower glume: Beak shape	□ (magenta)	0.56	0.25	1, 3, 5, 7
23	Lower lemma: Beak shape	△ (grey)	1.25	0.64	1, 3, 5, 7, 9

[a] Approximate as it assume interval scaled. Based on 100 observations per variety

19. Figure 2 shows that such relationship exists although the relationship is not clear for all characteristics. The clearest relations were seen for 12, 15, 18, 20 and 21 while the least clear relations were seen for characteristic 17 and 23. There seem to be a tendency that the clearest relations were found for the characteristics where the variation within variety was small (Table 5) while the least relations were found for characteristics where the variation within variety was large. For the characteristics where a clear relationship was found the smallest standard deviations was found when the mean note for the variety was close to one of the recorded values.

20. Similar results are found when using the coefficient of concentration (Figure 3), although the two measures are not strongly correlated for all characteristics (Figure 4).

21. The measure of heterogeneity for a variety depends much on the mean note (APPENDIX 2). A possible method for heterogeneity for such characteristics could be to judge if any of the plants are considered as an off-type – either directly when accessing the characteristic or based on figures such as those in appendix 2.

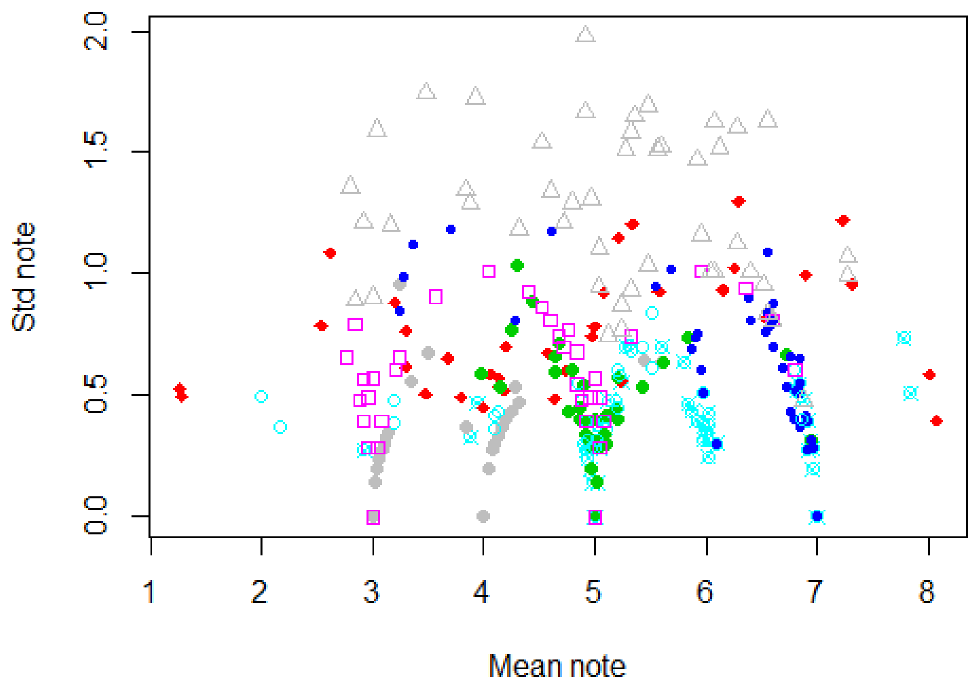


Figure 2 Relation between standard deviations and means for 8 characteristics of wheat (see Table 5 for a list of the characteristics)

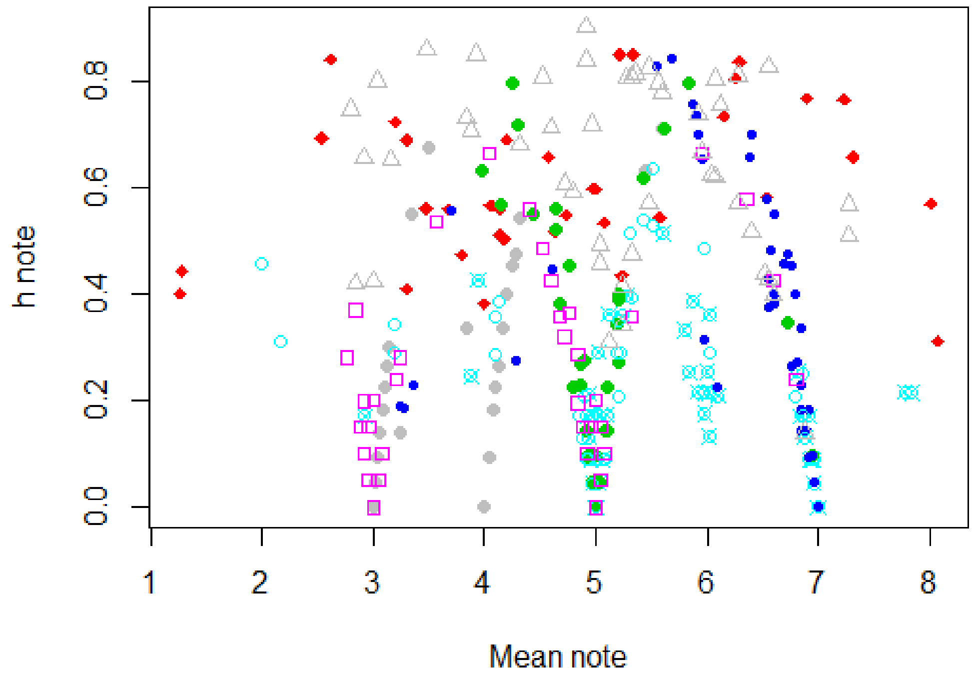


Figure 3 Relation between the coefficients of concentration, h, and means for 8 characteristics of wheat (see Table 5 for a list of the characteristics)

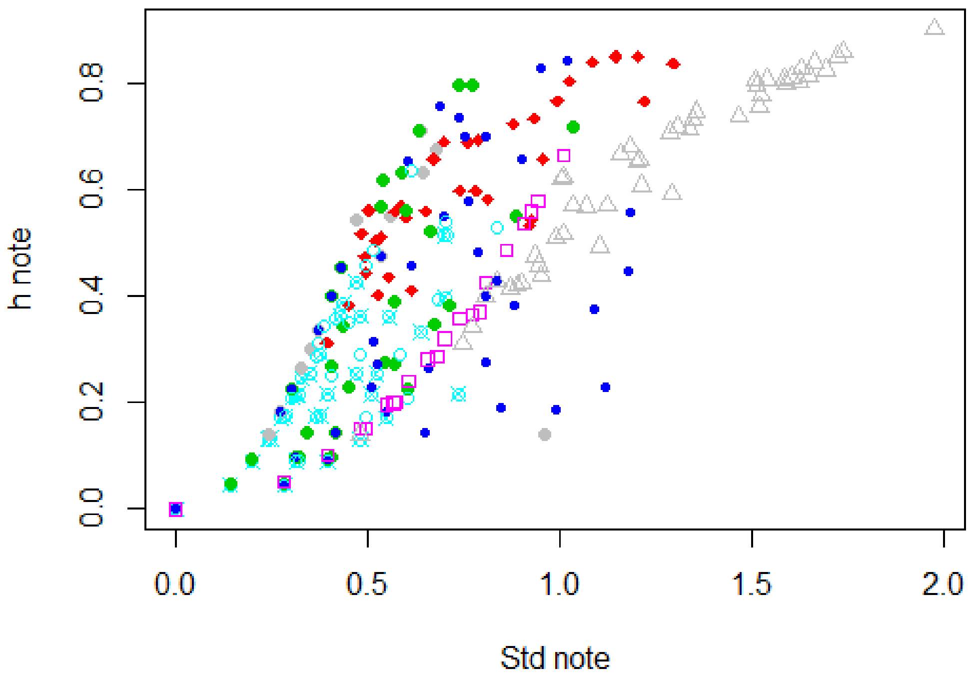


Figure 4 Relation between the coefficients of concentration, h, and means for 8 characteristics of wheat (see Table 5 for a list of the characteristics)

Discussion

22. The above examples clearly show that the uniformity for visually accessed characteristics in these examples depended on the mean or more correctly on where it is located on the underlying scale and where the thresholds are located. However, the results depend very much the assumption that the notes are formed as a result of an underlying continuous variable.

23. For ordered data it is expected that the standard deviation or the coefficient on the underlying variable will be a good measure of heterogeneity, but this is unknown. Unfortunately, the standard deviation (or the coefficient of concentration) on the note is not directly related to the standard deviation on the underlying variable, because the standard deviation and other measures of heterogeneity depend much on where the mean of the variety on the underlying variable is located relative to how the notes are defined. The two measures of uniformity used here showed similar relation with the mean note.

24. The most unfavourable (for variety) situation when the variety mean value is very close to the note threshold can be partly overcome by amalgamation of two categories with the largest observations before calculation any measure of variation such as for example coefficient of concentration. After amalgamation, two varieties with the same dispersion but with different location (with respect to the threshold) of the mean value will receive approximately the same measure of uniformity. As an example this has been done for the data in Table 4. The results are shown in Table 6. Variety *B* had large values for both the standard deviation and the coefficient of concentrations because its mean value was located right at the border between to notes. After merging this variety had smaller values and thus be not be rejected as non-uniform just because it happened to be close to the border between two notes. However, variety *C*, which measure of uniformity should be comparable to that of variety A, seemed to be much more heterogenic than variety *A* after merging.

Table 6 Measures of uniformity for artificial varieties with notes based on the parameters shown in Table 3 and distribution of notes shown in Table 4 before and after merging the two most frequent notes

Variety	True Std. Dev. on continuous variable	Std. Dev. on Note. Recorded	Coefficient of concentration, h. Recorded	Std. Dev. on Note. Original[a]	Coefficient of concentration, h. Merged
A	4	0.31	0.24	0.010	0.0003
B	4	0.52	0.63	0.080	0.0167
C	4	0.25	0.15	0.173	0.0786
D	4	0.00	0.00	0.000	0.0000
E	8	0.52	0.51	0.173	0.0786
F	8	0.72	0.77	0.363	0.3219
G	8	0.08	0.02	0.000	0.0000

a) After merging the notes were renumbered (1, 2, 3,…) before calculating the standard deviation.

25. For nominal scaled characteristics it is expected that the uniformity of the varieties also will depend on the note and on how the note are defined.

26. As we do not know the underlying scale and where the thresholds are defined indirectly the above examples show that it may be difficult to decide how to define uniformity for visually accessed characteristics.

Appendix 1 Coefficient of concentration

27. The - so called - coefficient of concentration h_i (probably the better name for it is the coefficient of diffuseness) is calculated according to the formula (1) and can be treated as an alternative measure of uniformity, see also TWC/13/3

$$h_i = \frac{k}{k-1}\left(1-\frac{\sum_{j=1}^{k} x_{ij}^2}{\left(\sum_{j=1}^{k} x_{ij}\right)^2}\right) \qquad (1)$$

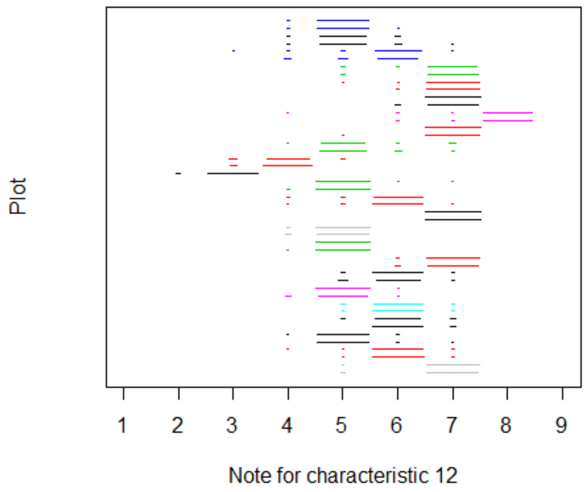
where k stands for the number of “effective” categories, x_{ij} is the observation (fraction, number of plants) for i-th variety in j-th note (category). The term “effective category” denotes category with at least one observation different from zero for at least one variety.

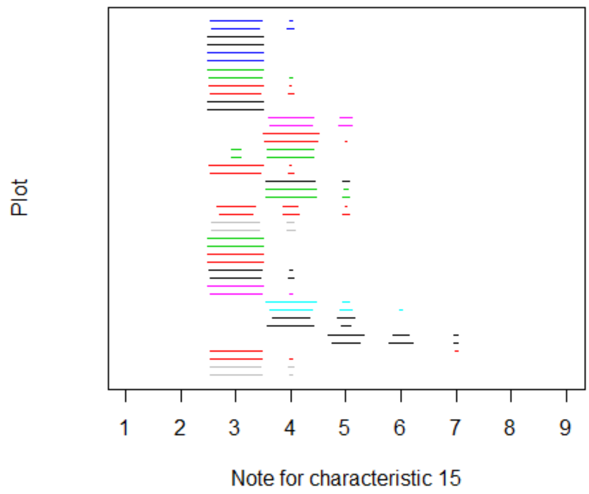
28. The main advantage of this coefficient is that it takes values from the range from 0 (perfect uniformity – all observations received the same note) to 1 (the same numbers (fractions) of observations in all notes).
As crop experts know from their experience which variety is more uniform than the other, so – at least within the same trial – they can compare coefficient of concentration of new variety with those of known varieties to have some information on degree of uniformity of new variety.

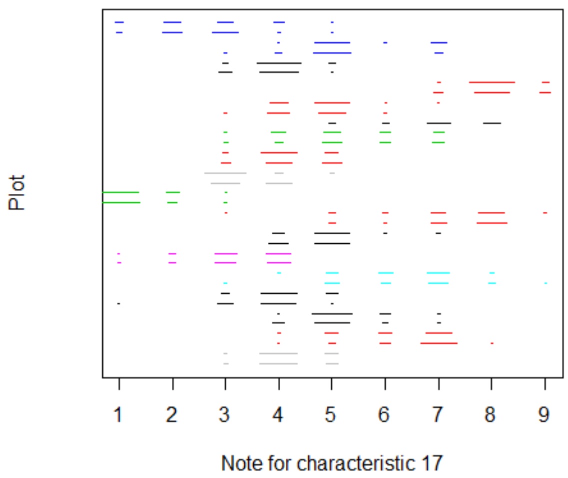
Appendix 2 Distribution of notes for each characteristic

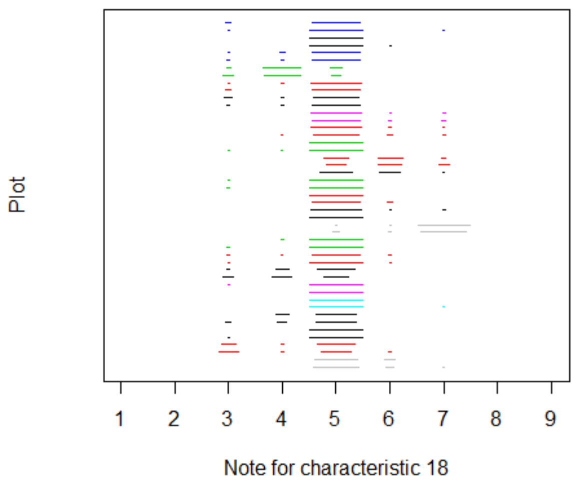
29. In the figures to follow the length of the lines indicates the relative number of observation (out of 50) for each plot that had the actual note. The colour of the line indicates the variety (so if two neighbouring lines have the same colour they belong to the same variety).

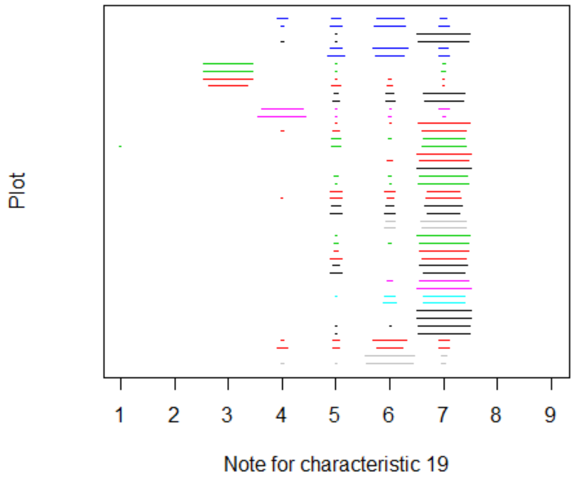
30. So as an example the bottom 2 lines of the figure for characteristic 12 show that these two plots comes from the same variety – as they both have the same colour (grey). In both plots most plants had note 7, but a few plants had note 5. The next two lines also belong to the same variety (red lines) and most of the plats had note 6 with a few plants in both replicates had note 5 and 7 and in one of the replicates a single plant had note 4. This single plant with note 4 may be considered as an off-type.

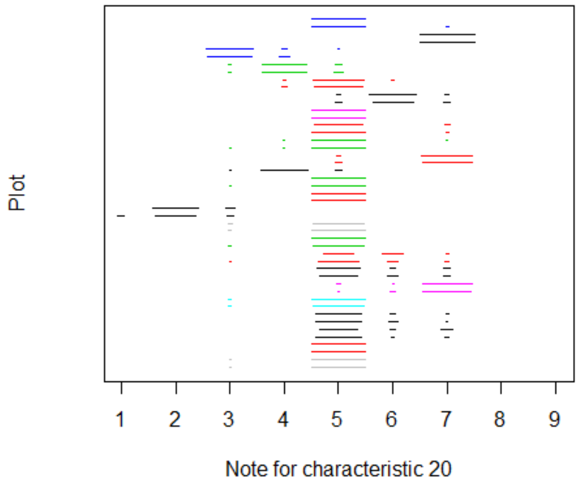


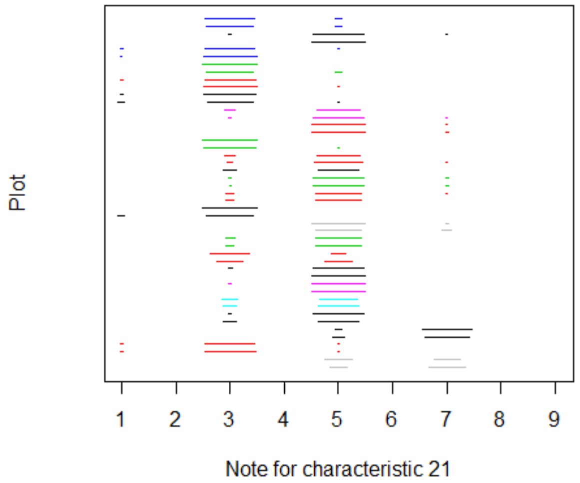


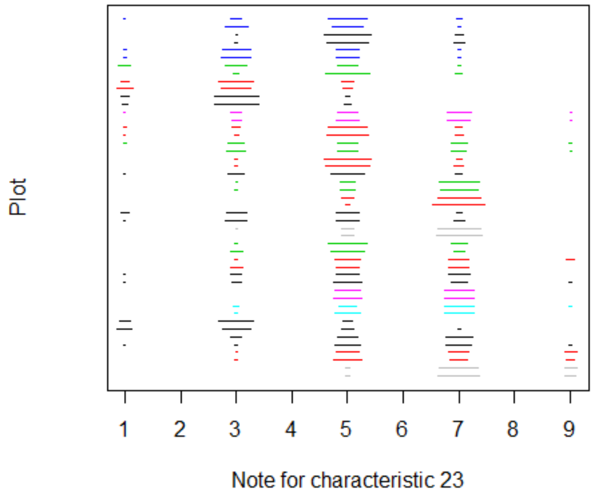












[End of Annex and of document]