E

UPOV

TC-EDC/Jan13/19
ORIGINAL: English
DATE: November 27, 2012

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
Geneva

ENLARGED EDITORIAL COMMITTEE

Geneva, January 9 and 10, 2013

REVISION OF DOCUMENT TGP/8: PART II: TECHNIQUES USED IN DUS EXAMINATION,
NEW SECTION: STATISTICAL METHODS FOR VISUALLY OBSERVED CHARACTERISTICS

Document prepared by the Office of the Union

BACKGROUND

1. The Technical Committee (TC), at its forty-eighth session, held in Geneva from March 26 to 28, 2012, considered the proposal for a New Section: "Statistical methods for visually observed characteristics" to be introduced in document TGP/8: Part II: Techniques Used in DUS Examination on the basis of document TC/48/19 Rev. "Revision of document TGP/8: Trial Design and Techniques Used in the Examination of Distinctness, Uniformity and Stability", Annex X, as prepared by an expert from Denmark. The TC agreed that the section "Statistical methods for visually observed characteristics" 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 the conclusions" paragraph 61).

2. The TC agreed with the workplan for the development of document TGP/8 presented in Annex XV to document TC/48/19 Rev., which indicated that New Section "Statistical methods for visually observed characteristics", would be considered by the TWPs in 2012. 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 Technical Working Parties (TWPs) at their sessions in 2012 (see document TC/48/22 "Report on the conclusions" paragraphs 49 and 78).

3. The following abbreviations are used in this document:

TC:	Technical Committee
TC-EDC:	Enlarged Editorial Committee
TWA:	Technical Working Party for Agricultural Crops
TWC:	Technical Working Party on Automation and Computer Programs
TWF:	Technical Working Party for Fruit Crops
TWO:	Technical Working Party for Ornamental Plants and Forest Trees
TWPs:	Technical Working Parties
TWV:	Technical Working Party for Vegetables

4. The structure of this document is as follows:

COMMENTS BY THE TECHNICAL WORKING PARTIES IN 2012

The TWA, TWV, TWC, TWF and TWO considered documents TWA/41/29, TWV/46/29, TWC/30/19 and TWC/30/29, TWF/43/29, TWO/45/29, respectively, and commented as follows:

General	The TWA noted that the presented method was an alternative to the Chi-square test for independence in the contingency table. The TWA proposed that the new Section for TGP/8 be developed in closer relation to TGP/8/1 Section 5 “Pearson’s chi-square test applied to contingency tables”. The TWA agreed that the example of Sugar Beet was not appropriate and the example on Carrot needed to be reconsidered. The TWA suggested to consider the development of a new Section with the same example as in TGP/8/1 Section 5 (see document TWA/41/34 “Report”, paragraphs 41 and 42).	TWA
	The TWV considered that the method presented in the Annex to document TWV/46/29 was a useful alternative to the Chi-square test for independence in the contingency table and agreed to suggest that more examples and data be provided to further develop the document (see document TWV/46/41 “Report”, paragraph 42).	TWV
	The TWC considered documents: TWC/30/19 “Consequences of Decisions for DUS Examination of Distinctness, Uniformity and Stability” and TWC/30/29 and received a presentation by an expert from Denmark (see document TWC/30/41 “Report”, paragraphs 54 and 57).	TWC
	The TWC noted the changes introduced in document TWC/30/29 and agreed that new examples should be requested from Italy (Beetroot) and other countries for preparation of a new draft of the document for the TWPs sessions in 2013 (see document TWC/30/41 “Report”, paragraph 55).	TWC
Pages 3 and 4	The TWC requested the drafter to check if variety Q in the table 2, page 3, was variety T and to provide more explanations on the first paragraph of page 4 of document TWC/30/29. The TWC also requested that statistics F3 and F4 be described (see document TWC/30/41 “Report”, paragraph 56).	TWC
	The TWC agreed the following editorial changes to document TWC/30/29 (see document TWC/30/41 “Report”, paragraph 58): • Heading of Annex to read “TWC” • Page 4, first paragraph to change “form” to read “from” • Page 4, first paragraph to read “significantly” (in two places) • Page 4, first paragraph to read “P-value” and “always” • Page 7 to read “varieties E and H have	TWC

4. Annex I to this document contains proposed text by the drafter (Mr. Kristian Kristensen, Denmark) for New Section: “Statistical methods for visually observed characteristics”, on the basis of comments by the TWPs at their sessions in 2012. The amendments to the text considered by the TWPs at their sessions in 2012 are indicated by highlighting and strikethrough for deletions and highlighting and underlining for additions.

5. Annex II to this document presents a copy of supplementary information concerning consequences of the decisions for DUS examination as background information for consideration when document TWC/30/29 was discussed (see document TWC/30/19 “Consequences of Decisions for DUS Examination of Distinctness, Uniformity and Stability”), prepared by experts from Denmark and Poland considered by the TWC, at its thirtieth session, in 2012.

[Annex I follows]

ANNEX I

PROPOSED TEXT FOR: TGP/8/1: PART II: STATISTICAL METHODS FOR VISUALLY OBSERVED CHARACTERISTICS

THE COMBINED OVER-YEARS METHOD FOR NOMINAL CHARACTERISTICS

Summary of requirements for application of the method

The method is appropriate to use for assessing distinctness of varieties where:

- The characteristic is nominal and recorded for individual plants (usually recorded visually)
- There are some differences between plants
- The observations are made over at least two years or growing cycles on a single location
- There should be at least 20 degrees of freedom for estimating the random variety-by-year interaction term.
- The expected number of plants for each combination of variety and note should be at least one – and for most of the combinations the number should be at least 5.

Summary

The method can be considered as an alternative to the χ^2-test for independence in a contingency table. The χ^2-test only takes the variation caused by random sampling into account and may thus be too liberal if additional sources of variation are present. The combined over-years method for nominal characteristics takes other sources of variation into account by including a random variety-by-year interaction term (as for the COYD method described in TGP/8/1 Part II: 3). The inclusion of the random effect is expected to decrease the number of distinct pairs of varieties compared to the χ^2-test for independence, but to better ensure that the decisions are consistent over coming years. The method is based on a generalisation of the traditional analyses of variance and regression methods for normally distributed data, which are called "generalized linear mixed models". A detailed description of the method – using other examples of data may be found in Agresti (2002) or Kristensen (2011).

The combined over-years method for nominal characteristics involves

- Calculating the number of plants for each note for each variety in each of the two or three years of trials, which results in a 3-way table (see the example)
- Analyse the data using appropriate software
- Compare each candidate to the reference varieties and the other candidates at the appropriate level of significance to see which varieties the candidate is distinct from
- Check if the variety-by-year interaction term for distinct pairs is considerably larger than the average for all variety pairs

Example

For demonstration a subset of varieties from a DUS experiment with Sugar beet was chosen. The notes for hypocotyl colour (Table 1) were analysed. Because some varieties had notes with zero plants in both years, there were difficulties in meeting the requirements mentioned above. Therefore, the varieties *M*, *N*, *O*, *Q*, *R*, *S* and *V* were excluded from the analyses shown here.

The estimated percent of plants in each note for each variety are shown in table 2.

Treating varieties *A* and *B* as candidates and the remaining varieties *C, D, …, U,* as reference varieties, the F-values and the P-values for testing the hypothesis of no difference between candidate and reference varieties were calculated. The F-values and the P-values are shown in Table 3. The F_3-values and their significances are also shown in Table 3.

Using the 1% level of significance as a decision rule for comparing the candidates with the reference varieties, we found that candidate *A* was distinct from 7 of the other varieties, while candidate *B* was distinct from 5 of the other varieties. The largest F_3-values were found for the variety pairs *B*-*K* and *A*-*K*. This seemed to be caused mainly by variety *K*, which had many green and no red hypocotyls in year 1, but few green and many red hypocotyls in year 2.

Table 1. Number of individuals with each note for hypocotyl colors for some varieties in Sugar beet

Variety	Color							
	1 Green		2 White		3-5 Red[1]		7 Orange	
	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
A	30	21	9	1	15	25	46	53
B	5	9	9	5	48	46	38	40
C	0	3	17	12	31	35	52	50
D	1	0	7	8	71	77	21	15
E	0	3	5	0	80	72	20	25
F	30	28	0	4	30	30	40	38
G	33	25	12	2	16	24	39	49
H	72	76	2	4	3	2	23	18
I	3	2	4	2	37	29	56	67
J	82	82	2	0	7	5	9	13
K	52	7	16	33	0	44	32	16
L	50	37	17	9	5	12	28	42
M	0	0	12	2	58	56	30	42
N	0	0	9	8	74	69	17	23
O	0	0	12	10	58	65	30	25
P	25	22	0	10	17	11	58	57
Q	0	0	0	10	65	64	35	26
R	0	0	0	0	75	55	25	45
S	0	0	6	1	53	61	41	38
T	83	92	5	1	3	1	9	6
U	54	30	12	13	3	4	31	53
V	0	0	6	18	71	63	23	19

1) Sum of three different reddish colours (pink, red and dark red)

Table 2. Estimated percent of plants for each note of each variety

Variety	Color			
	1 Green	2 White	3-5 Red	7 Orange
A	25.8	3.9	19.8	50.5
B	7.0	6.8	47.2	39.1
C	1.5	14.3	33.0	51.1
D	0.5	7.5	74.2	17.8
E	1.5	1.8	74.7	22.0
F	29.1	1.7	30.1	39.2
G	29.5	5.6	20.1	44.8
H	74.1	2.9	2.5	20.5
I	2.5	2.9	33.0	61.6
J	82.2	0.9	6.0	11.0
K	27.7	29.3	14.0	29.0
L	44.0	12.7	8.0	35.2
P	23.9	3.4	14.1	58.7
TQ	88.0	2.5	2.0	7.5
U	41.7	12.8	3.5	42.0

Table 3. Differences and F_3 values together with P-values for relevant pairs of varieties

Variety	Candidate A				Candidate B			
	F	$P_{dif.}$	F_3	P_{F3}	F	$P_{dif.}$	F_3	P_{F3}
A	-	-	-	-	-2.34	0.1157	0.50	0.6855
B	2.34	0.1157	0.50	0.6855	-	-	-	-
C	5.70	0.0062	0.57	0.5829	2.06	0.1432	0.02	0.9826
D	6.29	0.0033	0.50	0.6485	2.05	0.1404	0.42	0.7800
E	5.40	0.0063	0.41	0.6601	1.35	0.2866	0.19	0.8542
F	0.52	0.6757	1.20	0.2671	3.20	0.0522	0.50	0.7097
G	0.16	0.9224	0.01	0.9976	2.79	0.0786	0.46	0.7701
H	6.91	0.0036	0.94	0.4998	14.33	<.0001	0.15	0.9024
I	5.44	0.0073	0.24	0.7018	2.27	0.1143	0.24	0.9500
J	10.36	0.0004	0.19	0.8365	17.65	<.0001	0.18	0.9506
K	2.19	0.1361	3.17	0.0405	4.54	0.0189	4.31	0.0071
L	2.02	0.1621	0.11	0.9719	6.55	0.0051	0.64	0.7790
P	0.21	0.8896	1.79	0.0934	2.67	0.0847	0.92	0.4270
T	13.62	<.0001	0.65	0.7695	21.42	<.0001	0.05	0.9946
U	2.34	0.1202	0.52	0.7387	7.38	0.0027	1.18	0.8181

In order to get an indication of whether the varieties left in the present analysis the analyses may be performed using another method (Laplace's maximum likelihood) instead of the present method (a kind of residual maximum likelihood). However it should be noted that using maximum likelihood may give too many significant results, but this method is much to prefer compared to a χ^2-test for independence in a contingency table. As an example variety *M* would then be estimated to have M 0, 6.2, 57.6 and 36.2 per cent plants with note green, white, red and orange, respectively Variety *O* would – using this method – be significantldifferent form variety *A* (P=0.0014) but not significantly different from variety *B* (P=0.7224). From this it is judged that variety *A* and *M* would be distinct as the P-values is well below 0.01 and that variety *B* and *M* are not significant as the P-values for proposed method allways will be larger .

In order to examine whether one or more varieties have a different variety by year interaction than for the altenative maximum likelihood method. The F_4 -values for each variety in the analysis of the hypocotyl colours are shown in Figure 1. The largest F_4-value was found for variety *K*. The value seemed to be extremely large and an explanation for the unusual result should be sought.

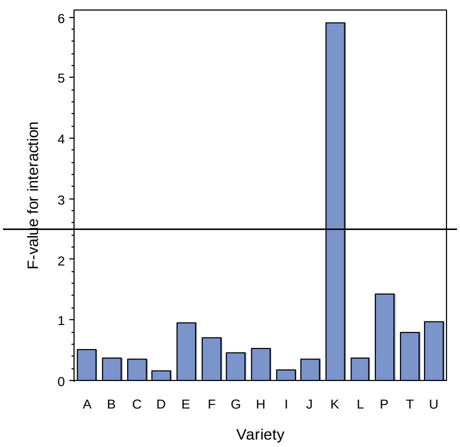


Figure 1: F_4-values for each variety's contribution to the interaction for nominal characteristic hypocotyl color

THE COMBINED OVER-YEARS METHOD FOR ORDINAL CHARACTERISTICS

Summary of requirements for application of the method

- The method is appropriate to use for assessing distinctness of varieties where:
- The characteristic is ordinal and recorded for individual plants (usually recorded visually)
- There are some differences between plants
- The observations are made over at least two years or growing cycles on a single location
- There should be at least 20 degrees of freedom for estimating the random variety-by-year interaction term.
- The distribution of the characteristic should be unimodal, i.e. notes with large number of plants should occur next to each other, zeros at one or both ends of the scale should not cause problems as long as most varieties have plants that fall in different notes
- The total number of plants for each variety should not be too low, at least 5 times the number of notes the variety covers

Summary

The method can be considered as an alternative to the χ^2-test for independence in a contingency table. The χ^2-test only takes the variation caused by random sampling into account and may thus be too liberal if additional sources of variation are present. Also the χ^2-test does not take the ordering of the notes into account. The combined over-years method for ordinal characteristics takes other sources of variation into account by including a random variety-by-year interaction term (as for the COYD method described in TGP/8/1 Part II: 3).It takes the ordering of notes into account by using a cumulative function over the ordered notes. The inclusion of the random effect is expected to decrease the number of distinct pairs of varieties compared to the χ^2-test for independence, but to better ensure that the decisions are consistent over coming years. Taking the ordering of notes into account is expected to increase the power of the test and thus to increase the number of distinct pairs.

The method is based on a generalisation of the traditional analyses of variance and regression methods for normally distributed data, which are called “generalized linear mixed models”. A general description of the method may be found in Agresti (2002) and a more specific description – using other examples of data may be found in Kristensen (2011).

The combined over-years method for ~~nominal~~ ordinal characteristics involves

- Calculating the number of plants for each note for each variety in each of the two or three years of trials, which results in a 3-way table (see the example)
- Analyse the data using appropriate software
- Compare each candidate to the reference varieties and the other candidates at the appropriate level of significance to see which varieties the candidate is distinct from
- Check if the variety-by-year interaction term for distinct pairs is considerably larger than the average for all variety pairs

~~Example~~

~~For demonstration a subset of varieties from a DUS experiment with carrots in France was chosen. The notes for root tip (when fully developed) (Table 4) were analysed (Characteristic 13 of TG/49/8). In most cases 60 plants were recorded in each year.~~

The estimated percent of plants in each note for each variety are shown in Table 5.

Table 4. Number of individual plants with each note for note on root tip for some varieties in carrots

Variety	Note					
	1 blunt		2 slightly pointed		3 strongly pointed	
	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
A	15	50	31	9	14	1
B	26	52	4	8	0	0
C	30	43	29	17	1	0
D	55	53	5	7	0	0
E	43	54	15	6	2	0
F	0	1	3	24	57	35
G	43	52	16	8	1	0
H	16	29	30	28	13	2
I	39	55	17	5	4	0

Table 5. Estimated percent of plants for each note of each variety

Variety	Note		
	1 blunt	2 slightly pointed	3 strongly pointed
A	52.4	42.3	5.3
B	86.1	12.9	1.0
C	62.8	33.7	3.5
D	90.1	9.2	0.7
E	82.6	16.1	1.3
F	1.3	16.4	82.3
G	80.5	18.0	1.5
H	35.3	54.6	10.1
I	81.0	17.6	1.4

Treating varieties *A* and *B* as candidates and the remaining varieties *C, D,…, I* as reference varieties, the F-values and the P-values for testing the hypothesis of no difference between candidate and reference varieties were calculated. The F-values and the P-values are shown in Table 6. The F_3-values and their significances are also shown in Table 6.

For the data shown here candidate *A* and *B* could both be separated from 1 of the other varieties (variety *F*) when using a 1% level of significance. The F_3 values were not significantly larger than 1 for any of the tested variety pairs shown in table 6. The largest F_3 was found for the variety pair *A-D*. The second largest F_3 was found for the variety pair *A-B*.

Table 6. Differences and F_3 values together with P-values for relevant pairs of varieties

Variety	Candidate A				Candidate B			
	Difference	$P_{Difference}$	F_3	P_{F3}	Difference	$P_{Difference}$	F_3	P_{F3}
A	-	-	-	-	1.73	0.0485	3.90	0.0836
B	-1.73	0.0485	3.90	0.0836	-	-	-	-
C	-0.43	0.5593	2.30	0.1675	1.30	0.1158	0.21	0.6591
D	-2.11	0.0214	4.97	0.0563	-0.38	0.6373	0.06	0.8060
E	-1.46	0.0764	1.46	0.2610	-0.27	0.7342	0.59	0.4655
F	4.42	0.0003	0.18	0.6846	6.15	<.0001	2.42	0.1586
G	-1.33	0.1007	2.11	0.1848	0.41	0.6050	0.28	0.6139
H	0.70	0.3434	1.56	0.2477	2.43	0.0109	0.53	0.4868
I	-1.36	0.0966	0.71	0.4226	0.38	0.6340	1.28	0.2909

The F_4 values for each variety in the analysis of root tip on carrots are shown in Figure 2. It is seen that only three varieties have a value larger than 1. The largest F_4 is found for variety *A*. It is also seen that variety *E* and *H* has a very low interaction with year indicating that their response to year is very close to the mean reaction for all varieties.

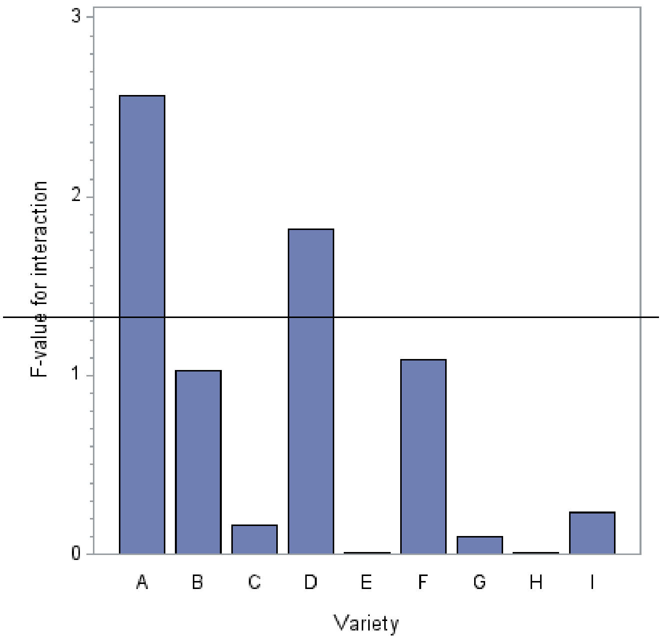


Figure 2 F_4-values for each variety's contribution to the interaction for ordinal characteristic toot tip of carrots

Example

For demonstration a subset of varieties from a DUS experiment with Meadow fescue *(Festuca pratensis)* in Finland was chosen. The notes for Plant: growth habit at inflorescence emergence (Characteristic 9 of TG/39/8) in 2010, 2011 and 2012 were analysed (Table 4). In most cases 40-60 plants were recorded in each year. This characteristic is rather sensitive to the growing conditions. This is apparent from table 4 where it is seen that the note 1 was recorded only in 2012 while note 7 was recorded only in 2010. Also it is seen that the most common note (over all varieties) in the three years was note, 5, 3 and 3, respectively in 2010, 2011 and 2012. The applied analysis method takes this into account by calculating an additive effect of each year (as for the COYD method for normal distributed data).

The estimated percent of plants in each note for each variety are shown in Table 5.

Table 4. Number of individual plants with each note for each variety and year for the characteristic Plant: growth habit at inflorescence emergence in Meadow fescue *(Festuca pratensis)*

Variety	Note																				
	1			2			3			4			5			6			7		
	2010	2011	2012	2010	2011	2012	2010	2011	2012	2010	2011	2012	2010	2011	2012	2010	2011	2012	2010	2011	2012
A	0	0	2	0	2	20	4	27	23	1	23	5	32	2	8	4	0	1	0	0	0
B	0	0	0	0	1	20	1	12	21	9	5	11	29	0	5	8	0	0	0	0	0
C	0	0	0	0	4	24	3	21	21	1	21	7	30	7	6	8	1	1	0	0	0
D	0	0	2	0	6	17	7	35	23	6	11	14	31	1	3	3	0	0	0	0	0
E	0	0	1	1	9	22	9	30	28	13	12	6	31	1	1	0	0	0	0	0	0
F	0	0	0	0	1	11	0	13	14	6	22	15	27	14	18	10	4	1	0	0	0
G	0	0	0	0	3	29	8	34	25	10	18	4	25	3	1	4	0	0	0	0	0
H	0	0	5	0	6	28	7	48	21	19	6	4	19	0	1	1	0	0	0	0	0
I	0	0	1	0	2	20	5	29	21	6	23	8	29	5	9	6	0	0	0	0	0
J	0	0	0	0	0	15	1	35	27	0	16	12	35	5	6	4	0	0	2	0	0
K	0	0	0	0	0	16	2	24	14	4	17	13	29	17	13	9	0	2	2	0	0
L	0	0	3	0	3	20	4	34	26	7	17	8	28	5	3	2	0	0	0	0	0
M	0	0	0	0	1	18	5	24	22	7	27	13	30	7	6	5	0	0	2	0	0
N	0	0	0	0	2	10	3	18	24	2	15	9	25	16	14	11	1	1	1	0	0
O	0	0	0	0	5	19	9	39	29	9	8	10	23	2	1	3	0	0	0	0	0
P	0	0	2	0	9	23	13	30	32	7	4	3	19	0	0	2	0	0	0	0	0
Q	0	0	1	0	4	24	9	27	24	10	19	8	28	5	2	3	0	0	0	0	0
R	0	0	0	0	3	24	2	30	26	6	21	6	35	6	1	5	0	0	0	0	0
S	0	0	1	0	5	16	6	25	27	14	19	11	26	8	4	2	0	0	0	0	0
T	0	0	0	0	6	19	3	36	24	4	5	7	18	3	7	5	0	0	0	0	0
U	0	0	2	0	7	17	11	41	31	15	11	8	30	0	0	0	0	0	0	0	0
V	0	0	3	0	15	32	11	33	18	13	6	5	30	3	0	4	0	1	0	0	0
W	0	0	0	0	7	22	4	28	30	6	16	6	37	5	2	6	0	0	1	0	0
X	0	0	1	0	5	19	2	24	17	4	17	15	40	6	7	2	0	0	0	0	0
Y	0	0	1	0	3	12	2	8	24	4	6	5	24	0	13	6	0	0	0	0	0
Z	0	0	0	0	1	14	1	25	17	2	16	15	26	10	13	10	0	0	0	0	0
1	0	0	2	0	6	24	5	38	24	8	9	8	34	2	2	0	0	0	0	0	0
2	0	0	0	1	4	20	5	29	26	5	16	11	37	5	3	3	0	0	0	0	0
3	0	0	2	0	10	24	7	28	27	8	12	4	30	1	0	0	0	0	0	0	0
4	0	0	1	0	9	17	7	31	28	6	10	9	30	2	2	2	0	0	0	0	0
5	0	0	0	0	3	14	1	24	26	9	22	16	36	8	4	5	0	0	0	0	0

Table 5. Estimated percent of plants for each note of each variety

Variety	Note						
	1	2	3	4	5	6	7
A	0.2	5.7	34.8	33.7	24.5	1.1	0.1
B	0.2	5.9	35.4	33.5	23.9	1.0	0.0
C	0.1	4.8	31.2	34.4	28.1	1.3	0.1
D	0.2	8.2	41.8	30.8	18.2	0.7	0.0
E	0.4	12.4	48.7	25.7	12.4	0.5	0.0
F	0.0	1.7	14.6	28.9	51.0	3.6	0.2
G	0.3	10.3	45.8	28.2	14.9	0.6	0.0
H	0.6	17.0	52.3	20.9	8.9	0.3	0.0
I	0.2	5.6	34.1	33.9	25.1	1.1	0.1
J	0.1	4.3	29.2	34.6	30.3	1.4	0.1
K	0.1	2.5	19.6	32.5	42.8	2.5	0.1
L	0.2	7.8	40.8	31.4	19.1	0.8	0.0
M	0.1	4.6	30.2	34.5	29.1	1.3	0.1
N	0.1	2.2	18.1	31.6	45.1	2.8	0.1
O	0.3	10.1	45.5	28.4	15.1	0.6	0.0
P	0.5	16.0	51.8	21.8	9.5	0.3	0.0
Q	0.3	8.8	43.1	30.0	17.1	0.7	0.0
R	0.2	6.7	37.8	32.7	21.7	0.9	0.0
S	0.2	7.0	38.8	32.3	20.8	0.8	0.0
T	0.2	7.9	41.0	31.2	18.8	0.7	0.0
U	0.4	12.1	48.4	25.9	12.7	0.5	0.0
V	0.5	16.5	52.1	21.4	9.2	0.3	0.0
W	0.2	7.1	38.9	32.2	20.7	0.8	0.0
X	0.1	5.2	32.6	34.2	26.6	1.2	0.1
Y	0.1	4.4	29.7	34.6	29.7	1.4	0.1
Z	0.1	2.7	21.3	33.3	40.3	2.2	0.1
1	0.3	10.6	46.2	27.8	14.5	0.5	0.0
2	0.2	6.7	37.8	32.7	21.7	0.9	0.0
3	0.4	12.6	49.0	25.4	12.2	0.4	0.0
4	0.3	9.3	44.1	29.4	16.3	0.6	0.0
5	0.1	4.4	29.7	34.6	29.7	1.4	0.1

The candidates were variety *A* and *B* and the remaining varieties *C, D,…, 5* were reference varieties, a measure of the differences and the P-values for testing the hypothesis of no difference between candidate and reference varieties were calculated. The differences and the *P*-values are shown in Table 6. An F_3-value is calculated in a similar way as for COY-D for normally distributed characteristics and is used in order to ensure that the pair did not became distinct because of a very large difference in only of the years without being different in other years (TGP/8/1 Draft 13 Section 3.6.3). Therefore, a significant difference between two varieties with a high F_3-value should be examined carefully before the final decision is taken.The F_3-values and their significances are also shown in Table 6.

For the data shown here candidate *A* could be separated from 11 of the reference varieties when using a 1% level of significance while candidate B could be separated form 10 of the reference varieties. The two candidates could not be separated from each other. The largest F_3-*value,* 5.43, was found for variety pair *B-S* (the approximate threshold for the F_4 values to be significant is 4.98). This means that the interaction for this pair should have been considered if this pair had been distinct on this characteristic.

Table 6. Differences and F_3 values together with P-values for relevant pairs of varieties

Variety	Candidate A				Candidate B			
	Difference	$P_{Difference}$	F_3	P_{F3}	Difference	$P_{Difference}$	F_3	P_{F3}
A	-	-	-	-	0.03	0.9011	0.22	0.4051
B	-0.03	0.9011	0.21	0.6566	-	-	-	-
C	0.19	0.4507	0.02	0.8782	0.22	0.4051	0.09	0.7694
D	-0.39	0.1243	0.04	0.8522	-0.35	0.1856	0.07	0.7947
E	-0.84	0.0011	0.73	0.4154	-0.81	0.0030	1.73	0.2215
F	1.26	<.0001	0.56	0.4743	1.29	<.0001	1.46	0.2584
G	-0.63	0.0125	1.66	0.2298	-0.60	0.0255	3.06	0.1144
H	-1.22	<.0001	1.17	0.3080	-1.19	<.0001	2.37	0.1579
I	0.03	0.8922	0.29	0.6041	0.07	0.8004	0.99	0.3448
J	0.30	0.2267	1.13	0.3146	0.34	0.2081	0.37	0.5600
K	0.88	0.0007	0.00	0.9669	0.91	0.0010	0.25	0.6274
L	-0.33	0.1879	0.52	0.4895	-0.30	0.2651	1.39	0.2681
M	0.24	0.3255	0.82	0.3878	0.28	0.2949	1.87	0.2047
N	0.99	0.0002	0.00	0.9734	1.02	0.0003	0.18	0.6805
O	-0.61	0.0162	0.27	0.6151	-0.58	0.0317	0.96	0.3525
P	-1.15	<.0001	0.24	0.6350	-1.11	0.0001	0.90	0.3664
Q	-0.47	0.0630	2.59	0.1421	-0.43	0.1039	4.28	0.0685
R	-0.17	0.5056	0.06	0.8115	-0.13	0.6174	0.50	0.4984
S	-0.22	0.3813	3.50	0.0943	-0.18	0.4858	5.43	0.0448
T	-0.34	0.1848	0.82	0.3879	-0.31	0.2578	0.20	0.6650
U	-0.82	0.0013	1.04	0.3352	-0.79	0.0035	2.18	0.1735
V	-1.18	<.0001	0.03	0.8674	-1.15	<.0001	0.08	0.7799
W	-0.23	0.3621	0.17	0.6870	-0.19	0.4653	0.00	0.9662
X	0.12	0.6441	0.00	0.9863	0.15	0.5764	0.23	0.6444
Y	0.27	0.3246	0.19	0.6753	0.30	0.2936	0.00	0.9791
Z	0.77	0.0032	0.64	0.4435	0.80	0.0038	0.12	0.7404
1	-0.66	0.0093	0.00	0.9861	-0.63	0.0196	0.23	0.6443
2	-0.17	0.5049	0.15	0.7116	-0.13	0.6165	0.71	0.4219
3	-0.87	0.0009	0.07	0.8017	-0.83	0.0026	0.52	0.4907
4	-0.53	0.0393	0.03	0.8714	-0.49	0.0684	0.09	0.7760
5	0.27	0.2712	0.31	0.5938	0.31	0.2471	1.03	0.3376

In order to examine whether one or more varieties have a different variety by year interaction than the main part of the varieties, the actual contribution to the interaction was calculated for each variety and compared to the average contribution from all varieties. This was done using an *F*- value, F_4.

The F_4 values for each variety in the analysis are shown in Figure 2. The largest F_4-value, 2.78, was found for variety *S* (the approximate threshold for the F_4-values to be significant is 4.98). This value was not significant larger than 1. The F_4-value is calculated as the quotients between the each varieties contribution to the overall interaction and the average interaction over all varieties. As the contribution for the actual variety enters in both the numerator and denominator of the F_4-value this test is approximate.

It is also seen that some varieties, e.g. *I, K, N, X, 1, 2, 3* and *5* have a very low interaction with year indicating that their response to year is very close to the mean reaction for all varieties.

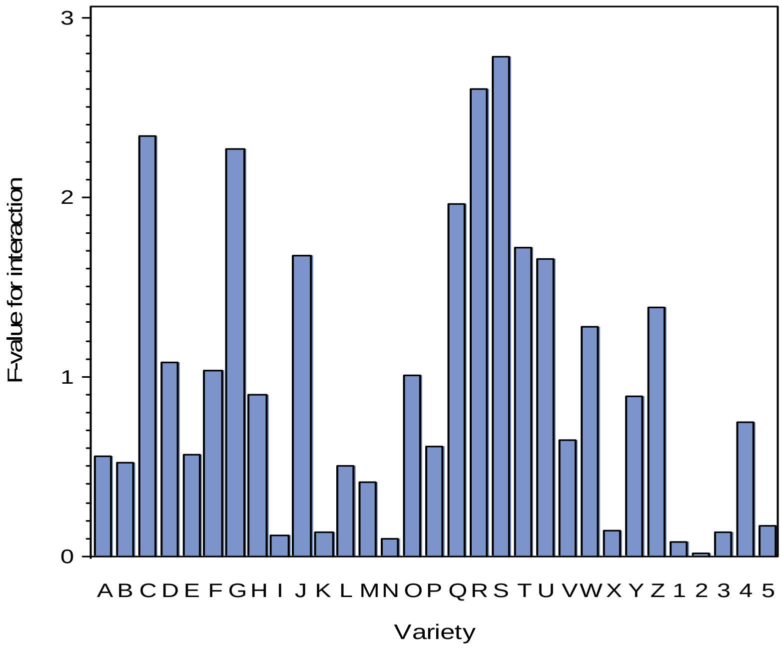


Figure ~~3~~ 2 F_4-values for each variety's contribution to the interaction for ordinal characteristic growth habit

THE COMBINED OVER-YEARS METHOD FOR NOMINAL CHARACTERISTICS

Summary of requirements for application of the method

The method is appropriate to use for assessing distinctness of varieties where:

- The characteristic is nominal and recorded for individual plants (usually recorded visually)
- There are some differences between plants
- The observations are made over at least two years or growing cycles on a single location
- There should be at least 20 degrees of freedom for estimating the random variety-by-year interaction term.
- The expected number of plants for each combination of variety and note should be at least one – and for most of the combinations the number should be at least 5.

Summary

The method can be considered as an alternative to the χ^2-test for independence in a contingency table. The χ^2-test only takes the variation caused by random sampling into account and may thus be too liberal if additional sources of variation are present. The combined over-years method for nominal characteristics takes other sources of variation into account by including a random variety-by-year interaction term (as for the COYD method described in TGP/8/1 Part II: 3). The inclusion of the random effect is expected to decrease the number of distinct pairs of varieties compared to the χ^2-test for independence, but to better ensure that the decisions are consistent over coming years. The method is based on a generalisation of the traditional analyses of variance and regression methods for normally distributed data, which are called "generalized linear mixed models". A detailed description of the method – using other examples of data may be found in Agresti (2002) or Kristensen (2011).

The combined over-years method for nominal characteristics involves

- Calculating the number of plants for each note for each variety in each of the two or three years of trials, which results in a 3-way table (see the example)
- Analyse the data using appropriate software
- Compare each candidate to the reference varieties and the other candidates at the appropriate level of significance to see which varieties the candidate is distinct from
- Check if the variety-by-year interaction term for distinct pairs is considerably larger than the average for all variety pairs

EXAMPLE

No example shown at present.

THE COMBINED OVER-YEARS METHOD FOR BINOMIAL CHARACTERISTICS

Summary of requirements for application of the method

The method is appropriate to use for assessing distinctness of varieties where:
- The characteristic is recorded for individual plants (usually recorded visually) using a scale with only 2 levels (such as present/absent or similar)
- There are some differences between plants
- The observations are made over at least two years or growing cycles on a single location
- There should be at least 20 degrees of freedom for estimating the random variety-by-year interaction term.
- The expected number of plants for each combination of variety and note should be at least one – and for most of the combinations the number should be at least 5.

Summary

The method can be considered as an alternative to the χ^2-test for independence in a contingency table. The χ^2-test only takes the variation caused by random sampling into account and may thus be too liberal if additional sources of variation are present. The combined over-years method for ~~ordinal~~ binomial characteristics take other sources of variation into account by including a random variety-by-year interaction term (as for the COYD method described in TGP/8/1 Part II: 3). The inclusion of the random effect is expected to decrease the number of distinct pairs of varieties compared to the χ^2-test for independence, but to better ensure that the decisions are consistent over coming years.

The method is based on generalisation of the traditional analyses of variance and regression methods for normally distributed data, which are called “generalized linear mixed models”.

The combined over-years method for binomial characteristics involves
- Calculating the number of plants for each note for each variety in each of the two or three years of trials, which results in a 3-way table
- Analyse the data using appropriate software
- Compare each candidate to the reference varieties and the other candidates at the appropriate level of significance to see which varieties the candidate is distinct from
- Check if the variety-by-year interaction term for distinct pairs is considerably larger than the average for all variety pairs

Example

The proportion of plants with cyanid glucoside (Characteristic 4 in TG/38/7) was measured for some white clover varieties in Northern Ireland in each of 3 years. The variable was recorded as absent or present. In this example only 20 varieties are used and variety 1 and 2 are considered as candidates, while the remaining varieties are considered as references. The data are shown in Table 7.

Table 7. Number of plants without and with cyanid glucoside in 20 white clover varieties in each of 3 years

	Year 1		Year 2		Year 3	
Variety	Absent	Present	Absent	Present	Absent	Present
1	31	29	22	38	17	43
2	40	20	42	18	41	19
3	50	10	52	8	55	5
4	42	18	40	20	34	26
5	37	23	42	18	37	23
6	51	9	49	11	52	8
7	30	30	25	35	26	34
8	37	23	31	29	30	30
9	27	33	27	33	25	35
10	48	12	47	13	43	17
11	40	20	40	20	32	28
12	18	42	13	47	12	48
13	10	50	12	48	5	55
14	41	19	46	14	45	15
15	58	2	55	5	58	2
16	7	53	10	50	11	49
17	25	35	22	38	20	40
18	48	12	54	6	52	8
19	20	40	20	40	23	37
20	57	3	54	6	55	5

The analysis showed that for these data there was no interaction between variety and year, which means that the variance component for year by variety was estimated to be zero and thus all variation in the data could be explained by sampling variation. The F-test for comparing the varieties was 36.67 with a P-value less than 0.01%, so there were clearly some differences among the varieties.

More specifically the analysis showed that candidate variety 1 was significantly different from 12 of the reference varieties at the 1% level (Table 8) whereas candidate variety 2 was significantly different from 11 of the reference varieties. Also the two candidate varieties were significantly different at the 1% level (Table 8).

As there was no interaction between variety and year, all F_3 and F_4 values are estimated to be zero for these data. Therefore, they are not shown here.

Table 8. Estimated percent of plants with cyanid glucoside for each variety and comparison of each variety with the candidate varieties 1 and 2 using F-tests

	Estimated	Candidate 1		Candidate 2	
Variety	percent	F	P	F	P
1	61.1			30.45	<.0001
2	31.6	30.45	<.0001		
3	12.7	77.01	<.0001	17.58	0.0002
4	35.5	23.05	<.0001	0.61	0.4395
5	35.5	23.05	<.0001	0.61	0.4395
6	15.5	70.09	<.0001	12.54	0.0011
7	55.0	1.38	0.2473	19.58	<.0001
8	45.5	8.69	0.0054	7.27	0.0104
9	56.1	0.93	0.3414	21.39	<.0001
10	23.3	49.59	<.0001	3.12	0.0853
11	37.8	19.27	<.0001	1.48	0.2309
12	76.1	9.28	0.0042	66.21	<.0001
13	85.0	24.61	<.0001	90.68	<.0001
14	26.6	41.43	<.0001	1.09	0.3034
15	5.0	82.34	<.0001	33.21	<.0001
16	84.5	23.44	<.0001	89.25	<.0001
17	62.8	0.11	0.7463	33.81	<.0001
18	14.4	72.95	<.0001	14.45	0.0005
19	65.0	0.58	0.4492	38.53	<.0001
20	7.8	84.99	<.0001	28.18	<.0001

COMMON TO ALL THREE METHODS

Software

The procedure *GLIMMIX* of *SAS* (SAS Institute Inc., 2010) can be used to estimate the parameters of the generalised linear mixed model, and the data-step facilities (and/or the procedure *IML*) of the same package can be used for the remaining calculations. However, similar facilities may be found in other statistical packages, thus the *glmer*() function of the package *lme4* of R can do the binomial analysis provided that there are more than one observations for each combination of variety and year.

Final note

In the case where there are only two notes, the methods for nominal and ordinal scaled characteristics both become identical as they reduce to the same binomial method: meaning that both methods can be applied to binomially distributed data.

References and literature

Agresti, A., 2002, Categorical data analysis, 2nd edition. Wiley & Sons, Inc. 710 pp.

Kristensen, K. 2011 Analyses of visually accessed data from DUS trials using a combined over years analysis for testing distinctness. Biuletyn Oceny Odmian (Cultivar Testing Bulletin) 33, 49-62.

SAS Institute Inc. 2010, SAS/STAT® 9.22 User's Guide. Cary, NC: SAS Institute Inc.8460 pp. (online access: http://support.sas.com/documentation/cdl/en/statug/63347/PDF/default/statug.pdf, accessed 15th November 2010)

[Annex II follows]

ANNEX II

CONSEQUENCES OF DECISIONS FOR DUS EXAMINATION OF DISTINCTNESS, UNIFORMITY AND STABILITY

INTRODUCTION

1. 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

2. 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

3. 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.

4. 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)

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

5. 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

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

7. 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

8. 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

9. 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.

10. 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.

11. 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.

12. 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.

13. 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.

14. 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).

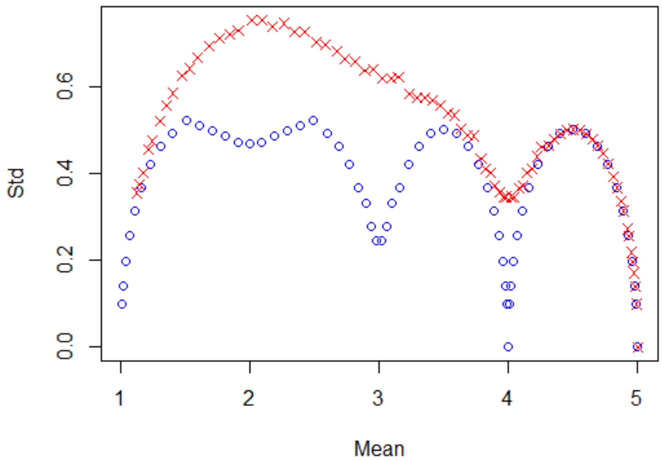


Figure 1 Relation between the standard deviation and mean of notes using the threshold stated above (Red crosses: Std. on the underlying continuous variable is 8. Blue circles: Std. on the underlying continuous variable is 4.)

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	⊠	0.33	0.18	2, 3, 4 ,5 ,6 ,7 ,8
15	Awns of scours at tip of ear: Length	●	0.26	0.20	3, 4 ,5 ,6 ,7
17	Apical rachis segment: Hairiness of convex surface	◆	076	0.61	1, 2, 3, 4 ,5 ,6 ,7, 8, 9
18	Lower glume: Shoulder width	●	0.41	0.26	3, 4, 5, 6, 7
19	Lower glume: Shoulder shape	●	0.59	0.35	3, 4, 5, 6, 7
20	Lower glume: Beak length	○	0.35	0.20	1, 2, 3, 4, 5, 6, 7
21	Lower glume: Beak shape	□	0.56	0.25	1, 3, 5, 7
23	Lower lemma: Beak shape	△	1.25	0.64	1, 3, 5, 7, 9

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

15. 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.

16. 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).

17. 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.

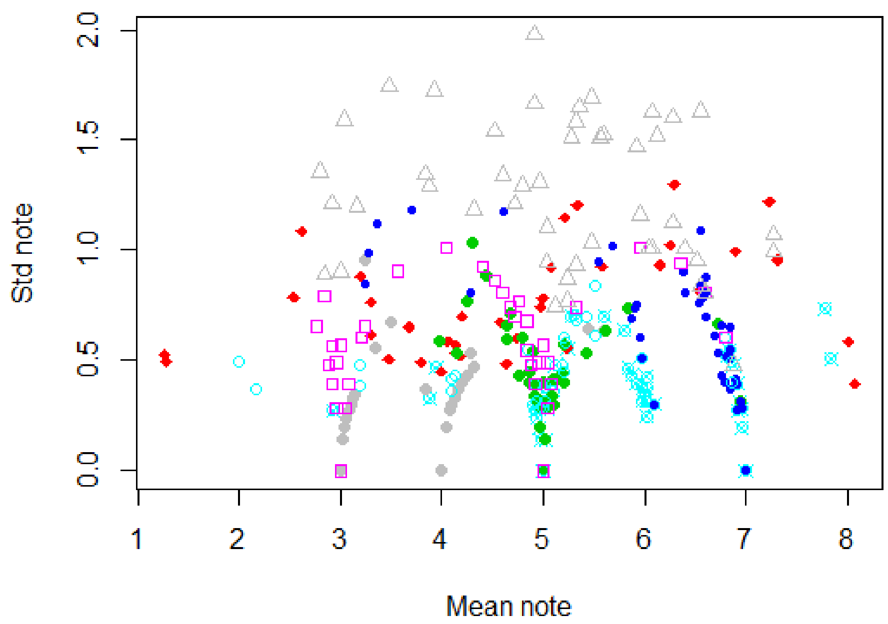


Figure 2 Relation between standard deviations and means for 8 characteristics of wheat (see Table 5 for a list of the characteristics)

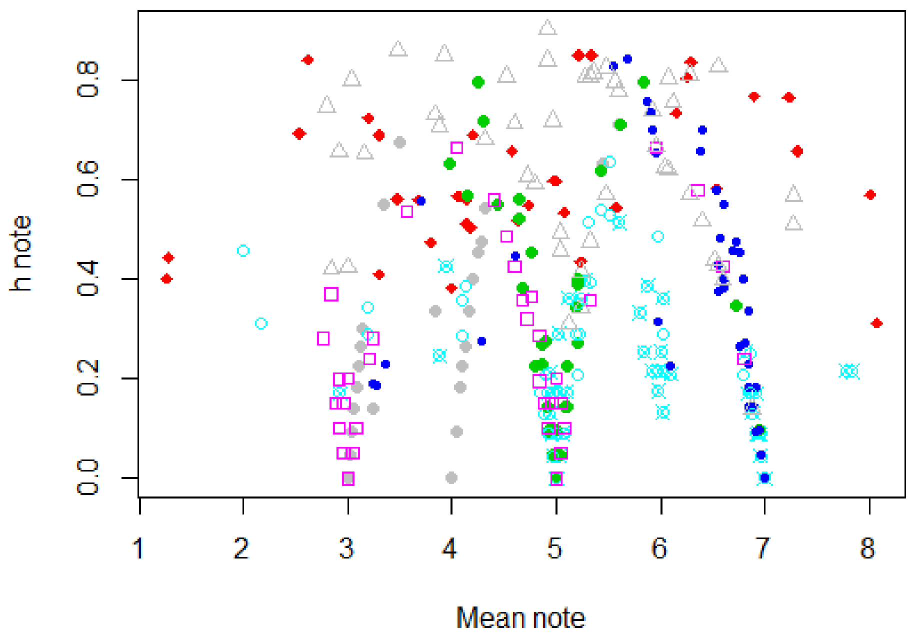


Figure 3 Relation between the coefficients of concentration, h, and means for 8 characteristics of wheat (see Table 5 for a list of the characteristics)

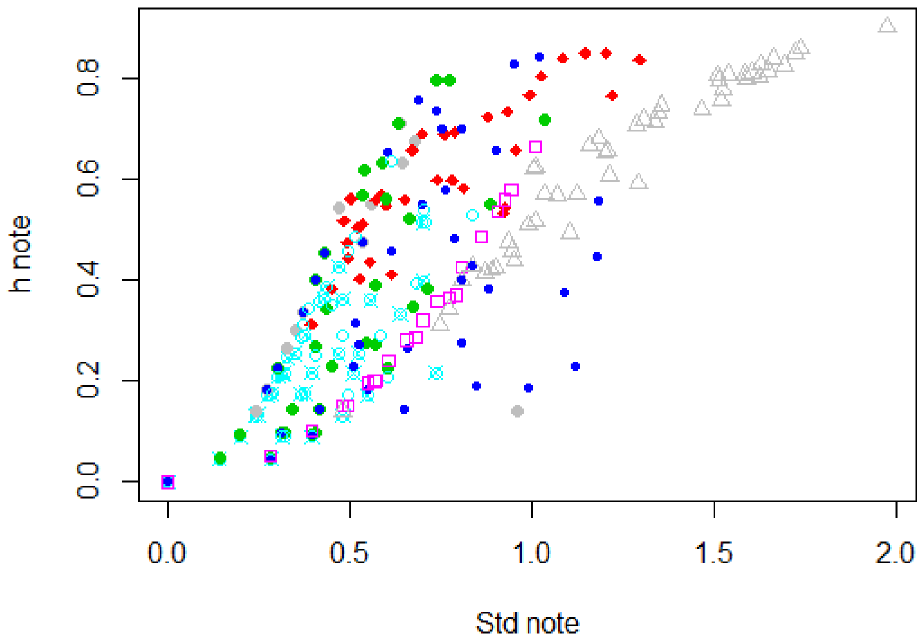


Figure 4 Relation between the coefficients of concentration, h, and means for 8 characteristics of wheat (see Table 5 for a list of the characteristics)

<u>Discussion</u>

18. 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.

19. 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.

20. 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.

21. 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.

22. 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

23. 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

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

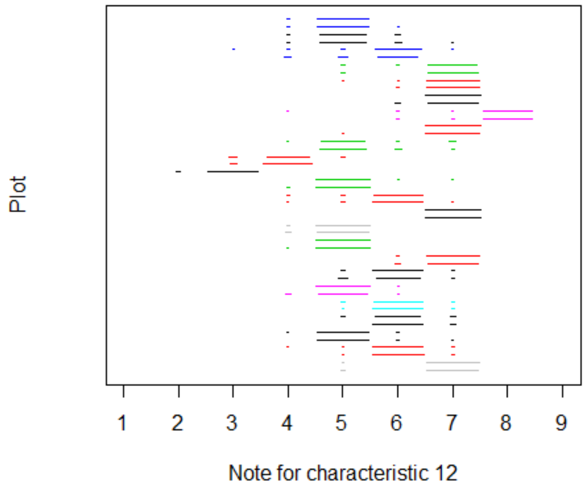
where k stands for the number of “effective” categories, x_{ij} is the observation (fraction, number of plants) for i-th variety in j-th note (category). The term “effective category” denotes category with at least one observation different from zero for at least one variety.

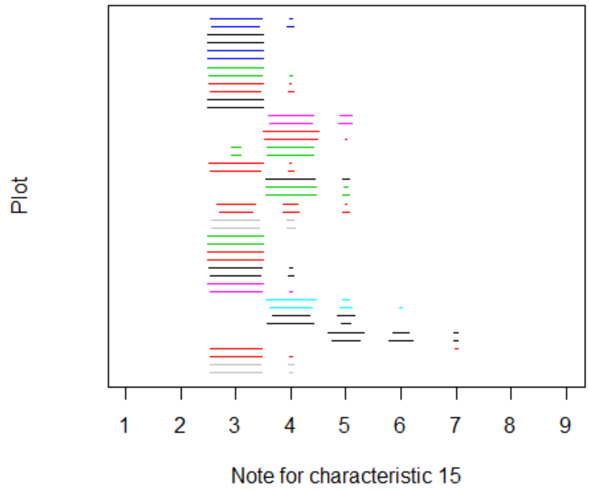
24. The main advantage of this coefficient is that it takes values from the range from 0 (perfect uniformity – all observations received the same note) to 1 (the same numbers (fractions) of observations in all notes). As crop experts know from their experience which variety is more uniform than the other, 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.

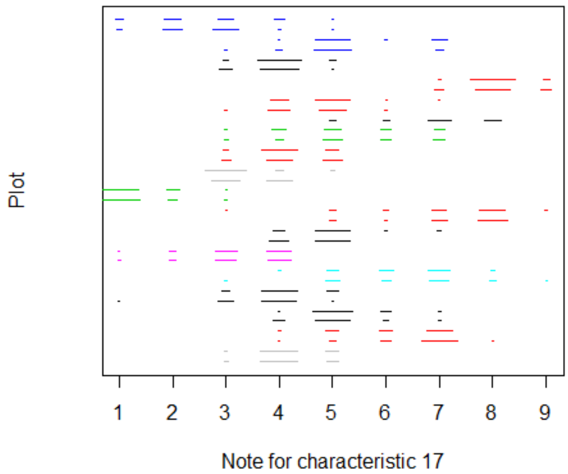
Appendix 2 Distribution of notes for each characteristic

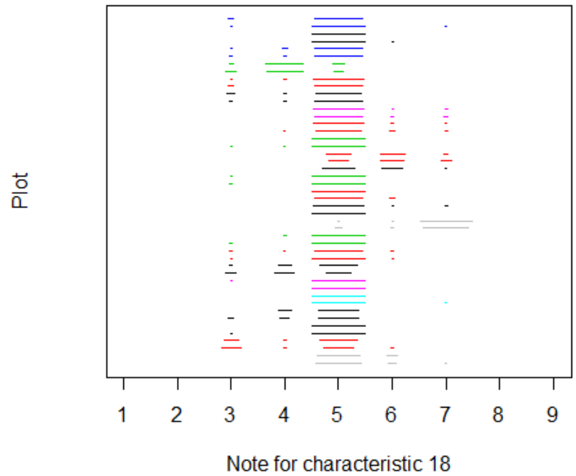
25. In the figures to follow the length of the lines indicates the relative number of observation (out of 50) for each plot that had the actual note. The colour of the line indicates the variety (so if two neighbouring lines have the same colour they belong to the same variety).

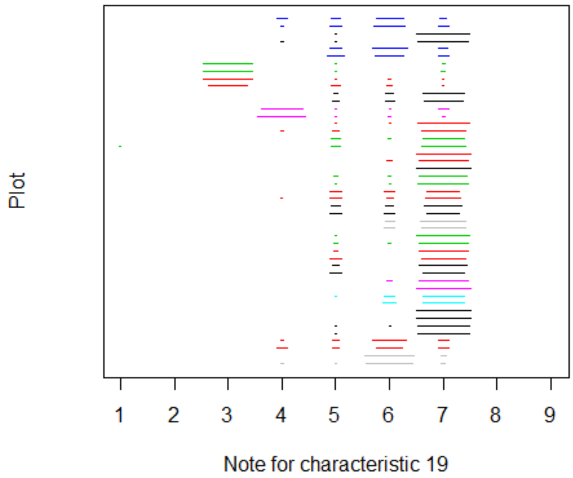
26. So as an example the bottom 2 lines of the figure for characteristic 12 show that these two plots comes from the same variety – as they both have the same colour (grey). In both plots most plants had note 7, but a few plants had note 5. The next two lines also belong to the same variety (red lines) 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.

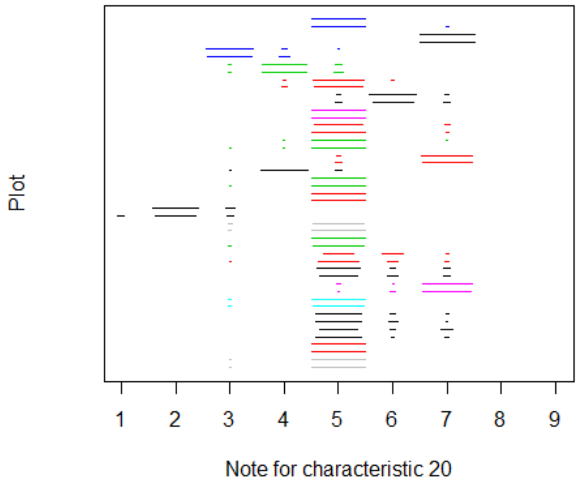


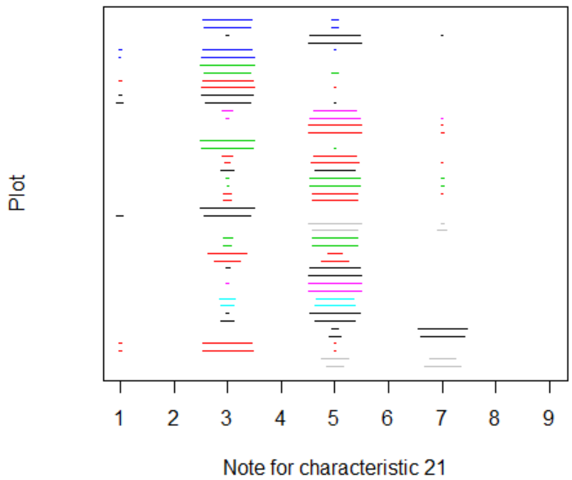


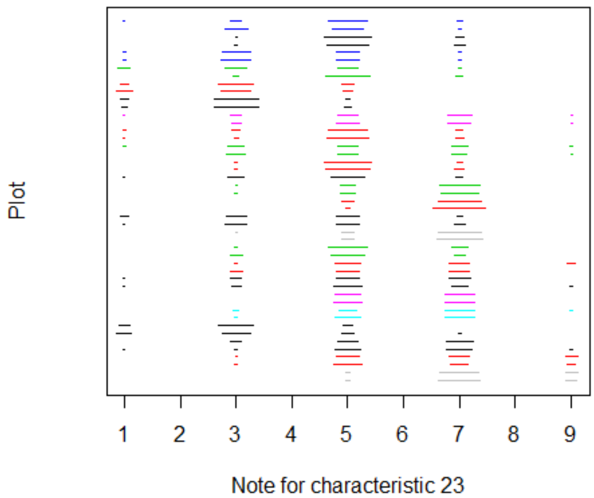












[End of Annex II and of document]