E

TWC/26/20
ORIGINAL: English
DATE: August 21, 2008

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

TECHNICAL WORKING PARTY ON AUTOMATION AND COMPUTER PROGRAMS

Twenty-Sixth Session
Jeju, Republic of Korea, September 2 to 5, 2008

CORRELATION BETWEEN DIFFERENT TYPES OF DISTANCE/SIMILARITY MEASURES ON A SET OF WINTER OILSEED RAPE CHARACTERISTICS OF DIFFERENT TYPES (NOMINAL TO RATIO SCALE)

Document prepared by experts from Germany

Introduction

1. The statistical analysis described in this document is based on a sub-set of morphological data arising from a project which was supported by the Community Plant Variety Office of the European Community (CPVO). The aims of the project were a study of the management of winter oilseed rape reference collections and the identification of appropriate statistical procedures for the analysis of morphological data. The data comprised two parts: the structure provided by the examination offices (dataset 1) and the consolidated data developed by the United Kingdom (dataset 2).

Available morphological data

2. Dataset 1 consists of notes and measurements per country and year. Characteristics are described in Tables 1 and 2 in the Annex to this document. The number of characteristics (notes and measurements) with data from different countries per year is summarized in Table 3. As a number of the characteristics were available both as measured continuous data as well as discrete notes, it was possible to use the continuous data and to derive discrete notes.

3. Dataset 2 consists of notes and measurements consolidated for countries and years by experts from the United Kingdom (see Section 2). The description of characteristics is the same as for dataset 1. As a number of the characteristics were available both as measured continuous data as well as discrete notes, it was possible to use the continuous data and derive discrete notes.

Common definitions, requirements and possibilities of interpretations

Distance measures

4. A distance measure defines a distance between any two objects (a set of characteristics per variety). Important properties of a distance measure are the invariance of scale and the invariance of translation. That means invariance of translation regarding two types of linear transformations such as the following:

$$z_{ij} = a_j * x_{ij} \text{ and } z_{ij} = b_j + x_{ij}$$

with $i = 1,\ldots,v$ and $j=1,..,n$
where v is the number variables (characteristics) and
n is the number of varieties.

5. If a distance measure is not scale or translation invariant, then characteristics have to be standardized (transformed to the same scale). Well known distance measures of quantitative characteristics (measurements) are the Euclidian Distance and the Cityblock Distance. An alternative expression for distance measure is dissimilarity index.

Similarity measures

6. A similarity measure is a relation between a function of two objects (a set of characteristics per variety) and a number which fulfilled a set of conditions. Definitions of special similarity measures are included in chapter 4.2. Gower's index is a well known similarity measure.

Correlation measures

7. Correlation is a relation between two or more statistical variables. There can be positive or negative correlations. An example of correlation measures is included in the section "Correlation measures".

Normalization

8. Since variables with large variances tend to have a greater effect on distance or similarity measures than those with small variances, it is recommended to standardize the variables (characteristics). In the SAS software package the DISTANCE procedure provides a convenient way to standardize each variable with its own method depending on the scale level of the variable before measures are computed. Standardization is not required if there is only one level of measurement, otherwise it is mandatory. Standardization depends on the scale level (nominal, ordinal, interval, ratio). Scale levels are defined in the following section.

Definition of scale levels

9. Types of characteristics in relation to their scale levels are explained in document TGP/8 Trial Design and Techniques Used in the Examination of Distinctness, Uniformity and Stability, chapter 4).

10. The most relevant definitions are the following:

Nominal scale:

Nominal scaled qualitative data are qualitative data without any logical order of the discrete categories. Characteristics with only two categories (notes) (dichotomous characteristic e.g. absent (1)/present(9)) are a special form of nominal scales. The nominal scale is the lowest classification of the scales.

Ordinal scale:

Ordinal scaled data are qualitative data for which discrete categories can be arranged in an ascending or descending order. They result from visually assessed quantitative characteristics.

The distances between the discrete categories of an ordinal scale are not exactly equal. Therefore, an ordinal scale does not fulfil the condition to calculate arithmetic mean values, which is the equality of intervals throughout the scale. The ordinal scale is higher classified than the nominal scale.

Interval scale:

An interval scale is a quantitative scale without a defined absolute zero point. There is always a constant non-zero distance between two adjacent expressions. Interval scale data may be distributed continuously or discretely. The interval scale is higher classified than the ordinal scale.

Ratio scale:

A ratio scale is a quantitative scale with a defined absolute zero point. There is always a constant non-zero distance between two adjacent expressions. Ratio scaled data may be continuous or discrete. The ratio scale is the highest classification of the scales.

<u>Evaluated methods for distance and similarity measures</u>

Distance measures

11. The following notation is used:

v	number of variables or the dimensionality
x_j	data for observation x on the i^{th} variable (characteristic), where i=1 to v
y_j	data for observation y on the i^{th} variable (characteristic), where i=1 to v
w_j	weight for the i^{th} variable. w_i=0 when either x_i or y_i is missing
W	the sum of total weights
$\bar{x}$	mean for observation x .

$$\bar{x} = \frac{\sum_{i=1}^{v} (w_i * x_i)}{\sum_{i=1}^{v} w_i}$$

$\bar{y}$ mean for observation y .

$$\bar{y} = \frac{\sum_{i=1}^{v} (w_j * y_j)}{\sum_{i=1}^{v} w_j}$$

d(x,y) the distance or dissimilarity between observations x and y
s(x,y) the similarity between observations x and y

In normal cases weights are equal to 1/v.

Minkowski metric

12. The Minkowski metric is a general metric which defines distances in the following way:

$$d(x,y) = \sqrt[p]{\sum_{i=1}^{v} |x_i - y_i|^p}$$

The Minkowski metric cannot be computed without defining the parameter p. So the Cityblock and the Euclidian distance result from the Minkowski metric.

Cityblock (p=1)

13. The Cityblock metric is a special case of the Minkowski metric with p=1.

$$d(x,y) = \sum_{i=1}^{v} |x_i - y_i|$$

Euclidian distance (p=2)

14. The Euclidian metric is a special case of the Minkowski metric with p=2.

$$d(x,y) = \sqrt[2]{\sum_{i=1}^{v} |x_i - y_i|^2}$$

The Euclidian distance is invariant to translation but not scale invariant.

Maximum distance (Chebychev)

15. The Chebychev distance is defined in the following way:

$$d(x,y) = \max\left\{ |x_i - y_i|, \ 1 \le i \le v \right\}$$

The Chebychev distance is translation invariant but not scale invariant.

Similarity measures

16. The following similarity measures were considered:

(a) Cosinus similarity measure
(b) Dice
(c) Jaccard (Tanimoto)
(d) M coefficient
(e) RR coefficient
(f) Kulczinski coefficient
(g) Gower's index

17. A basic requirement for the application of a specific method is the type of data to be analyzed. Dataset 1 and 2 both contain nominal, ordinal, interval and ratio scaled data. Among the methods listed above, the Gower's index is the only one which allows the consideration of nominal scaled data in combination with ordinal, interval and ratio scaled data (see section "Data evaluation within locations"). Therefore, only the Gower's index is presented in further detail here.

Gower's index

18. The Gower's index is a combination of distances depending on the scale level in the following general form:

$$s(x,y) = \frac{\sum_{i=1}^{v} w_i * \delta^i_{x,y} * d^i_{x,y}}{\sum_{i=1}^{v} w_i \delta^i_{x,y}}$$

To compute $\delta^i_{x,y}$: for nominal, ordinal, interval or ratio characteristics,

$$\delta^i_{x,y} = 1;$$

To compute $d^i_{x,y}$: for nominal characteristics:

$$d^i_{x,y} = 1, if\ x_i = y_i$$

$$d^i_{x,y} = 0, if\ x_i \neq y_i$$

For ordinal (where data are replaced by corresponding rank scores), interval or ratio characteristics

$$d^i_{x,y} = 1 - \left| x_i - y_i \right|$$

Correlation measures

Pearson

19. The Pearson correlation coefficient can be used to assess the linear relation between two variables (s and t).

$$r(s,t) = \frac{\sum_{j=1}^{n}\left(s_j - \bar{s}\right)*\left(t_j - \bar{t}\right)}{\sqrt{\sum_{j=1}^{n}\left(s_j - \bar{s}\right)^2 * \sum_{j=1}^{n}\left(t_j - \bar{t}\right)^2}}$$

Selection of appropriate methods

20. A basic requirement for the application of a specific method is the type of data to be analyzed. It is important if the data set contains only one type of data or a combination of different types of data. In particular it is important if the dataset contains nominally scaled data.

	Nominal		Ordinal	Interval	Ratio	Combination nominal/ ordinal/ interval/ratio
	two categories (notes)	>two categories (notes)				
Cityblock			**X**	**X**	**X**	
Euclidian			**X**	**X**	**X**	
Chebychev			**X**	**X**	**X**	
Cosinus	**X**					
Dice	**X**					
Jaccard	**X**					
M coefficient	**X**					
RR coefficient	**X**					
Kulczinski coefficient	**X**					
Gower's index	**X**	**X**	**X**	**X**	**X**	**X**

Data evaluation within locations

21. The following evaluations were based on the subset of data within dataset 1 which were provided by Germany (DE). The dataset consist of nominal, ordinal, interval and ratio scaled characteristics.

Influence of the year

22. The years 2003, 2004 and 2005 were compared for the German dataset which was part of dataset 1. For each year, the similarity measures between all variety pairs were evaluated using the “Gower’s index”. The results are summarized in the following table:

Sample 1	Sample 1	Similarity Measure	Correlation coefficient
DE2003	DE2004	Gower’s index	0.81926 (P<0.0001)
DE2003	DE2005	Gower’s index	0.82339 (P<0.0001)
DE2004	DE2005	Gower’s index	0.84790 (P<0.0001)

The correlation coefficients vary between 0.82 and 0.85. This means that the influence of the years in the German dataset is very low.

Influence of the distance/similarity measure

23. The aim of the evaluations is to identify the most appropriate distance/similarity measure for variety comparisons. To compare different distance/similarity measures the German part of dataset 1 was modified as follows:

The characteristics:

- b1 (Seed: erucic acid; 1=absent, 9=present)
- b6 (Leaf: lobes; 1=absent, 9=present)
- b13 (Production of pollen; 1=absent, 9=present)

were defined in the evaluation procedure as ordinal characteristics instead of nominal. This is possible because of the absence of more than two categories.

The characteristic:

- b10 (Flower: Color of petals; 1=white, 2=cream, 3=yellow, 4=orange-yellow)

was dropped from the dataset. It is impossible and from the theoretical point of view, forbidden to handle this nominal scaled characteristic with four categories (colors) as ordinal, interval or ratio scaled characteristic.

24. The distance measures “Cityblock”, “Euclidian” and “Chebychev” and the similarity index “Gower” were calculated. The results are summarized in the following table:

Sample	Measure 1	Measure 2	Correlation Coefficient
DE2003	Cityblock	Euclid	0.94925 (P<0.001)
		Chebychev	0.81139 (P<0.001)
		Gower	-0.95598 (P<0.001)
	Euclid	Chebychev	0.94786 (P<0.001)
		Gower	-0.85326 (P<0.001)
	Chebychev	Gower	-0.67777 (P<0.001)
DE2004	Cityblock	Euclid	0.94779 (P<0.001)
		Chebychev	0.84356 (P<0.001)
		Gower	-0.94456 (P<0.001)
	Euclid	Chebychev	0.96479 (P<0.001)
		Gower	-0.82675 (P<0.001)
	Chebychev	Gower	-0.68652 (P<0.001)
DE2005	Cityblock	Euclid	0.94967 (P<0.001)
		Chebychev	0.84734 (P<0.001)
		Gower	-0.94590 (P<0.001)
	Euclid	Chebychev	0.96476 (P<0.001)
		Gower	-0.82953 (P<0.001)
	Chebychev	Gower	-0.68807 (P<0.001)

25. The correlation coefficients vary from 0.67 to 0.97. That means the influence of the distance or similarity measure in the German dataset is high. The best correlations are between “Euclidian” and “Chebychev” in 2004 (0.96479) and between “Cityblock” and “Gower” in 2003 (0.95598).

26. From the theoretical point of view, “Gower’s index” is the best for the structure of dataset 1. The best correlated measure to “Gower’s index” is the “Cityblock distance”.

27. The other parts of the dataset 1 (France, the United Kingdom, and Denmark) were not evaluated because of the differences between some characteristics (non-harmonized dataset 1).

28. Differences are between expressions of characteristics:

- m9 (Time of flowering) - the starting point to count the days was not the same in all countries

Additionally there are some unclear data as described in the following table:

Country	year	Characteristic	variety	expression	Remark
United Kingdom	2005	b10	Cheops	“1&3”	1 or 2 or 3
United Kingdom	2004	b6	Grenat	5	1 or 9 allowed
United Kingdom	2004	b13	Grenat	6	1 or 9 allowed
Denmark/ United Kingdom	all	b20	all	all	the United Kingdom: b20 and b21
Denmark	2003	m11, m12	PO8861	0	0 instead of missing
France	2005	b10	Manitou	9	1, 2, 3, 4 allowed

Evaluation using data from all location

29. The following evaluations were based on dataset 2 which contains the consolidated data from all countries. The dataset consist of nominal, ordinal, interval and ratio scaled characteristics. In order to allow the calculation of distances according to “Cityblock”, “Euclidian” and “Chebychev” and of the "Gower's" similarity index, the dataset was modified in the same way as described in section 6.2.

30. The results are summarized in the following table:

Sample	Measure 1	Measure 2	Correlation Coefficient
Consolidated dataset 2	Cityblock	Euclid	0.95687 (P<0.001)
		Chebychev	0.87801 (P<0.001)
		Gower	-0.92994 (P<0.001)
	Euclid	Chebychev	0.97336 (P<0.001)
		Gower	-0.81894 (P<0.001)
	Chebychev	Gower	-0.70844 (P<0.001)

31. The correlation coefficients vary from 0.71 to 0.97. This means that the influence of the distance or similarity measure in the consolidated dataset 2 is also not low. The best correlation is between “Euclidian” and “Chebychev” (0.97336) or between “Cityblock” and “Euclid” (0.95687).

32. From the theoretical point of view “Gower’s index” is the best for the structure of the dataset 2. The best correlated measure to “Gower’s index” is the “Cityblock distance” (0.92994).

Conclusions

(a) Main efforts are to be made on harmonization of protocols, and harmonization of notations between experts that register the measures;

(b) Statistical computations, as shown above, need to be selected according to the type of scale of the characteristics;

(c) When some characteristics have a great influence on the synthetic value obtained over all characteristics, or when there are different types of scales in a dataset, one has to consider using either the whole dataset, or to drop some characteristics, or to compute subsets per type of characteristic;

(d) The “Gower’s index” is the most appropriate procedure for the structure of dataset 1 and 2 because it is the only one which allows a combination of the present data types;

(e) It is not allowed to use nominal scaled characteristics like characteristic b10 (Flower: color of petals; 1=white, 2=cream, 3=yellow, 4=orange-yellow) with more than two categories for evaluation of “Cityblock distance”;

(f) For comparison of different distance measurements dichotomous characteristics (b1, b6, b13) can be handled as ordinal characteristics. Nominal characteristics with more than two categories (b10) have to be dropped;

(g) The best correlated measure to “Gower’s index” on the basis of dataset 1 and 2 is the “Cityblock distance”.

[Annex follows]

ANNEX

Table 1: Table of characteristics (all characteristics)

CPVO-number	UPOV number	Characteristic	Type of characteristic	Scale level	Record/ states of expressions	Variable names
1	1	Seed:erucic acid	QL	nominal	1 absent	b1
					9 present	
2	2	Cotyledon: length	QN	ordinal	3 short	b2
					5 medium	
					7 long	
2	2	Cotyledon: length	QN	ratio	measured	m2
3	3	Cotyledon: width	QN	ordinal	3 narrow	b3
					5 medium	
					7 broad	
3	3	Cotyledon: width	QN	ratio	measured	m3
4	4	Leaf: green colour	QN	ordinal	3 light	b4
					5 medium	
					7 dark	
5	-	Leaf: glaucosity	QL	nominal	1 absent	b5
					9 present	
6	5	Leaf: lobes	QL	nominal	1 absent	b6
					9 present	
7	6	Leaf: number of lobes	QN	ordinal	3 few	b7
					5 medium	
					7 many	
7	6	Leaf: number of lobes	QN	ratio	counted	m7
8	7	Leaf: dentation of margin	QN	ordinal	3 weak	b8
					5 medium	
					7 strong	
9	11	Time of flowering	QN	ordinal	1 very early	b9
					3 early	
					5 medium	
					7 late	
					9 very late	
9	11	Time of flowering	QN	interval	counted	m9
10	12	Flower: colour of petals	QL	nominal	1 white	b10
					3 cream	
					5 yellow	
					7 orange-yellow	
11	13	Flower: length of petals	QL	ordinal	1 very short	b11
					3 short	
					5 medium	
					7 long	
11	13	Flower: length of petals	QL	ratio	measured	m11
12	14	Flower: width of petals	QL	nominal	3 narrow	b12
					5 medium	
					7 broad	

Table 1: Table of characteristics (all characteristics)

CPVO-number	UPOV number	Characteristic	Type of characteristic	Scale level	Record/ states of expressions	Variable names
12	14	Flower: width of petals	QL	ratio	measured	m12
13	15	Production of pollen	QL	nominal	1 absent	b13
					9 present	
14	16	Plant: height	QN	ordinal	1 very low	b14
					3 low	
					5 medium	
					7 tall	
					9 very tall	
14	16	Plant: height	QN	ratio	measured	m14
15	17	Plant: total length including side branches	QN	ordinal	1 very short	b15
					3 short	
					5 medium	
					7 long	
					9 very long	
15	17	Plant: total length including side branches	QN	ratio	measured	m15
16	18	Siliqua: length (between peduncle and beak)	QN	ordinal	1 very short	b16
					3 short	
					5 medium	
					7 long	
16	18	Siliqua: length (between peduncle and beak)	QN	ratio	measured	m16
17	-	Siliqua: width	QN	ordinal	3 narrow	b17
					5 medium	
					7 broad	
					9 very broad	
17	-	Siliqua: width	QN	ratio	measured	m17
18	19	Siliqua: length of beak	QN	ordinal	1 very short	b18
					3 short	
					5 medium	
					7 long	
					9 very long	
18	19	Siliqua: length of beak	QN	ratio	measured	m18
19	20	Siliqua: length of peduncle	QN	ordinal	3 short	b19
					5 medium	
					7 long	
19	20	Siliqua: length of	QN	ratio	measured	m19
20	21	Tendency to form inflorescences	QN	ordinal	1 absent or very weak	b20
					3 weak	
					5 medium	
					7 strong	
					9 very strong	

Table 2: Table of characteristics (optimized combination)

CPVO-number	UPOV-number	Characteristic	Type of characteristic	Scale level	Record/ states of expressions	Variable names
1	1	Seed:erucic acid	QL	nominal	1 absent	b1
					9 present	
2	2	Cotyledon: length	QN	ratio	measured	m2
3	3	Cotyledon: width	QN	ratio	measured	m3
4	4	Leaf: green colour	QN	ordinal	3 light	b4
					5 medium	
					7 dark	
5	-	Leaf: glaucosity	QL	nominal	1 absent	b5
					9 present	
6	5	Leaf: lobes	QL	nominal	1 absent	b6
					9 present	
7	6	Leaf: number of lobes	QN	ratio	counted	m7
8	7	Leaf: dentation of margin	QN	ordinal	3 weak	b8
					5 medium	
					7 strong	
9	11	Time of flowering	QN	interval	counted	m9
10	12	Flower: colour of petals	QL	nominal	1 white	b10
					3 cream	
					5 yellow	
					7 orange-yellow	
11	13	Flower: length of petals	QL	ratio	measured	m11
12	14	Flower: width of petals	QL	ratio	measured	m12
13	15	Production of pollen	QL	nominal	1 absent	b13
					9present	
14	16	Plant: height	QN	ratio	measured	m14
15	17	Plant: total length including side branches	QN	ratio	measured	m15
16	18	Siliqua: length (between peduncle and beak)	QN	ratio	measured	m16
17	-	Siliqua: width	QN	ratio	measured	m17
18	19	Siliqua: length of beak	QN	ratio	measured	m18
19	20	Siliqua: length of peduncle	QN	ratio	measured	m19
20	21	Tendency to form inflorescences	QN	ordinal	1 absent or very weak	b20
					3 weak	
					5 medium	
					7 strong	
					9 very strong	

Table 3: Number of characteristics per year and country with results (notes or measurements)

Dataset	Country	2003		2004		2005		consolidated	
		N	M	N	M	N	M	N	M
1	FR	13	2	12	3	19	9	-	-
1	the United	19	11	20	12	20	12	-	-
1	DK	0	12	0	12	0	12	-	-
1	DE	20	12	20	12	20	12	-	-
2	consolidated	-	-	-	-	-	-	20	12

Legend: N – Notes M - Measurements or counts

Table 4: Summaries of descriptive statistic (optimized dataset, DE2003)

Parameter / Char.	b1	b4	b5	b6	b8	b10	b13	b19	b20		m2	m3	m7	m9	m11	m12	m14	m15	m16	m17	m18	m19
Number of varieties	244	244	244	244	244	244	244	238	242		244	244	244	238	236	236	225	239	238	238	238	238
Number of missings	166	166	166	166	166	166	166	172	168		166	166	166	172	174	174	185	171	172	172	172	172
Min	1	3	9	1	2	2	1	3	1		13,32	21,13	1	22,1	12,93	6,9	90	97,5	51,76	3,4	9,52	15,09
Max	9	7	9	9	6	3	9	9	9		19,53	31,23	9,27	60,38	18,82	12,66	141	151,63	88,05	5,49	17,29	29,85
Mode	1	5	9	9	4	3	9	5	7													
Median	1	5	9	9	4	3	9	5	7													
Mean	1,6	5,0	9,0	8,8	3,7	3,0	9,0	5,3	6,2		16,5	26,9	6,0	26,8	16,9	10,1	108,2	127,0	70,5	4,6	13,7	21,2
Standard deviation											1,2	1,7	1,2	3,1	0,8	0,8	9,4	8,5	6,1	0,3	1,4	2,7
Coeff. of variation											7,4	6,5	20,0	11,7	5,0	7,5	8,7	6,7	8,6	6,2	10,4	12,9

Table 5: Summaries of descriptive statistic (optimized dataset, DE2004)

Parameter / Char.	b1	b4	b5	b6	b8	b10	b13	b19	b20		m2	m3	m7	m9	m11	m12	m14	m15	m16	m17	m18	m19
Number of varieties	260	260	260	260	260	260	260	257	258		260	260	255	257	256	256	258	257	257	257	257	257
Number of missings	150	150	150	150	150	150	150	153	152		150	150	155	153	154	154	152	153	153	153	153	153
Min	1	4	9	1	2	2	1	3	1		12,29	20,88	1	14,51	13,56	7,87	98	109,57	48,18	3,54	8,05	17,33
Max	9	7	9	9	7	3	9	9	9		20,16	31,74	9,15	64,87	19,16	13	180	208,97	80,6	5,32	18,8	33,33
Mode	1	5	9	9	3	3	9	5	7													
Median	1	5	9	9	4	3	9	5	5													
Mean	1,6	5,1	9,0	8,8	3,6	3,0	9,0	5,4	4,9		15,5	26,4	6,1	20,5	16,9	10,6	137,8	158,8	68,2	4,7	14,3	23,4
Standard deviation											1,2	1,9	1,0	4,0	0,9	0,8	12,8	13,2	5,5	0,3	1,6	2,8
Coeff. of variation											7,5	7,2	17,1	19,5	5,4	7,7	9,3	8,3	8,0	6,5	11,0	11,9

Table 6: Summaries of descriptive statistic (optimized dataset, DE2005)

Parameter / Char.	b1	b4	b5	b6	b8	b10	b13	b19	b20		m2	m3	m7	m9	m11	m12	m14	m15	m16	m17	m18	m19
Number of varieties	264	264	264	264	264	263	264	262	263		263	263	259	263	263	263	263	263	262	262	262	262
Number of missings	146	146	146	146	146	147	146	148	147		147	147	151	147	147	147	147	147	148	148	148	148
Min	1	4	9	1	2	2	1	3	1		13,7	21,36	1	14,88	13,14	7,4	107	119,35	40,22	3,62	8,67	15,16
Max	9	7	9	9	6	3	9	9	9		21,38	32,37	9,95	67,26	18,5	13,35	183	202,02	77,68	5,56	17,53	32,75
Mode	1	5	9	9	3	3	9	5	1													
Median	1	5	9	9	3	3	9	5	4													
Mean	1,5	5,0	9,0	8,8	3,3	3,0	9,0	5,3	4,0		16,8	26,4	6,2	22,2	16,3	10,3	133,9	149,5	61,4	4,6	13,7	22,2
Standard deviation											1,3	1,8	1,2	4,6	0,9	0,8	10,9	13,3	5,3	0,3	1,5	3,0
Coeff. of variation											7,6	7,0	19,6	20,5	5,8	7,8	8,1	8,9	8,6	6,2	10,6	13,6

Table 7: Summaries of descriptive

Parameter / Char.	b1	b4	b5	b6	b8	b10	b13	b20		M2	M3	M7	M9	M11	M12	M14	M15	M16	M17	M18	M19
Number of varieties	335	335	335	335	335	334	335	329		335	335	335	335	335	335	335	335	335	335	335	335
number of missings	0	0	0	0	0	1	0	6		0	0	0	0	0	0	0	0	0	0	0	0
Minimum	1	4	9	1	2	2	1	1		11,3	20,6	1,0	17,3	12,7	7,5	96,6	111,1	43,6	3,4	7,8	14,4
Maximum	9	7	9	9	7	5	9	9		18,8	30,3	8,8	46,3	18,6	12,8	183,8	187,5	90,4	5,3	17,8	31,1
Mode	1	5	9	9	4	3	9	1													
Median	1	5	9	9	4	3	9	4													
Mean	1,5	5,2	9,0	8,9	4,2	3,1	9,0	4,2		15,5	25,5	6,1	27,2	16,5	10,2	133,6	148,9	67,5	4,5	14,0	21,7
Standard deviation										1,3	1,7	1,0	3,7	0,8	0,7	12,2	10,7	6,1	0,4	1,5	2,5
coeff. of variation										8,1	6,8	16,5	13,6	5,0	7,0	9,1	7,2	9,1	7,9	10,5	11,5

[End of Annex and of document]