E

TWC/18/9
ORIGINAL: English
DATE: May 18, 2000

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

TECHNICAL WORKING PARTY ON AUTOMATION AND COMPUTER PROGRAMS

Eighteenth Session
Kyiv, June 12 to 15, 2000

TYPES OF CHARACTERISTICS AND THEIR SCALE LEVELS

Document prepared by experts from Germany

TYPES OF CHARACTERISTICS AND THEIR SCALE LEVELS

Uwe Meyer, Bundessortenamt, Hannover

1. Introduction

For revision of UPOV Test Guidelines or establishing of new ones and for understanding of the relations between the different parts of work of the crop experts during the DUS test it is necessary to have an answer for each of the following questions:

1. What is a characteristic?
2. What is a scale level of a characteristic?
3. What is the influence of the scale level on :
    - planning of the trial,
    - recording of the data,
    - determination of distinctness and uniformity and
    - description of the varieties?

2. Categories of characteristics as outlined in document TC/36/6

The following definitions for categories of characteristics are given in the General Introductions (TC/36/6, chapter 5.2.3):

- "Truly qualitative characteristics" are those that show discrete discontinuous states with no arbitrary limit on their number (for instance, number of whorls: one (1), two (2), three (3)). These are qualitative characteristics with clear-cut (discrete) discontinuous states of expression, each state being self-explanatory and independently meaningful. Each state is clearly different from the others and as a rule the characteristics are not influenced by environment.

- "Quantitative characteristics" are those that can be recorded on a one-dimensional scale and show continuous variation from one extreme to the other. They are divided into a number of states of expression for the purpose of description. The division is made only for description and not for distinctness purposes. The Test Guidelines do not specify the difference needed for distinctness. The states of expression should, however, be meaningful for DUS assessment.

- "Pseudo-qualitative characteristics" are characteristics that do not fit the definition of truly qualitative characteristics, but are treated as qualitative when it is more reasonable to disregard continuous variation for practical purposes and the states created are meaningful and sufficiently different from each other (e.g. shape: ovate (1), elliptic (2), round (3), obovate (4), or expression: absent or very weakly expressed (1) weakly expressed (2), strongly expressed (3)).

3. Remarks to the current definitions

This way of classification makes no clear separation between characteristic, scale level for assessed data and transformation of these data into a variety description.

Example for an adjustment:

The ‘Length of plants’ is usually measured in cm and the characteristic is defined as a quantitative one. In some cases there is the possibility to decide that it is sufficient to assess the data not by measurements but visually by another scale to have smaller requirements for assessment in order to save time and money. Such other scale may be the following:

Characteristic	Note	Expression
Length of plants	1	very short
	2	very short to short
	3	short
	4	short to medium
	5	medium
	6	medium to long
	7	long
	8	long to very long
	9	very long

If we assess the data visually then we are using a qualitative scale. If we measure the plants then we are using a quantitative scale for assessment.

From the crop expert point of view the characteristic ‘length of plants’ is in both cases of the quantitative type. We only have changed the scale for assessment but the ‘original characteristic’ is of the same type (here quantitative). Other words for ‘original characteristic’ are ‘latent characteristic’ or ‘underlying characteristic’.

4. Other views on characteristics and scale levels

At first it is necessary to distinguish between characteristics and scale levels.

In the following table we distinguish between three points of view to look at the characteristics or the data. The characteristics as expressed in the trial are view point 1, the data for evaluation of characteristics are view point 2 and the data for variety description are view point 3. The information level decreases from view point 1 to 3.

view point	description of the view point	Information level
1	characteristics as expressed in trial	high
2	data for evaluation of characteristics	medium
3	data for variety description	low

Types of data obtained from the trials can be considered in different ways. They can be divided into quantitative (metric) and qualitative (categorical) data in dependence on the type of the scale of assessed data. Here we are only looking from view point 2 and 3.

4.1. Characteristics as expressed in trial

In the UPOV routine work often we are using the term ‘characteristic’ and what we mean is the sum of our current knowledge about properties of varieties. The crop expert distinguishes between qualitative and quantitative characteristics on the basis of the underlying characteristic and not of the scale level of data recorded in the trial. The types of characteristics are defined according to what the crop expert can see or measure or to his experience.

4.2. Quantitative data for the evaluation of trials and for variety description

“Quantitative data” are data which are recorded by measuring or counting. The assessed data are distributed on a metric scale (see 5.1). Quantitative data can be divided into continuous and discrete data.

Continuous data result from measurements. They can take every value out of the defined range.

Discrete quantitative data result from counting.

From the statistical point of view characteristics with quantitative data are usually denoted as quantitative characteristics.

Examples:

Quantitative data	**Example**	**Example number**
- continuous	Plant length in cm	1
- discrete	Number of inflorescences	2

For description of the stages of expressions see Annex 1.

The continuous quantitative data of the characteristic ‘Plant length’ are measured in cm and there is a continuous scale with defined units of assessment. It depends only on the costs and the necessity to get any value in cm or mm. Changing of measure e.g. from cm into mm is only a question of precision and not a change of scale.

The discrete quantitative data of the characteristic ‘Number of inflorescences’ are assessed by counting (1, 2, 3, 4, and so on). The distances between the neighbouring units of assessment are constant and for this example equal to 1. There are no real values between two neighbouring units but it is allowed to compute an average which is between those units.

Additionally it is possible to distinguish between ratio scaled and interval scaled quantitative data (see 5.1.1. and 5.1.2.).

Remarks:

Weighing is a special form of measuring.

In some cases visually assessed characteristics may be handled as quantitative. The questions are:

How precise are our assessments and how robust are the statistical procedures when using that kind of data?

4.3. Qualitative data for the evaluation of the trials and for variety description

"Qualitative data" are data which can be arranged in discrete qualitative different categories. Qualitative data can be divided into ordinal and nominal scaled data (see 4.3.1. and 4.3.2.).

From the statistical point of view characteristics with qualitative data are usually denoted as qualitative characteristics.

Remark:

Usually qualitative characteristics are assessed visually.

4.3.1. Ordinal scaled qualitative data

"Ordinal scaled" data are qualitative data whose discrete categories can be brought in an ascending or descending order (see 5.2.).

Examples:

Qualitative data	**Example**	**Example number**
- ordinal	Intensity of anthocyan	3

For description of the stages of expressions see Annex 1.

From the statistical point of view characteristics with ordinal scaled qualitative data are usually denoted as ordinal qualitative characteristics.

4.3.2. Nominal scaled qualitative data

"Nominal scaled" qualitative data are qualitative data without a logical order of the discrete categories (see 5.3.).

Examples:

Qualitative data	Example	Example number
- nominal	Sex of plant	4
- nominal with two stages	Leaf blade: variegation	5

For description of the stages of expressions see Annex 1.

From the statistical point of view characteristics with nominal scaled qualitative data are usually denoted as nominal qualitative characteristics.

5. Definition of scale levels

5.1. Metric or cardinal scale

A "metric scale" consists of two types of scales – the ratio scale and the interval scale. A synonym for metric scale is cardinal scale.

5.1.1. Ratio scale

A "ratio scale" is a metric scale with a defined absolute zero point. There is always a constant distance (different from zero) between two adjacent expressions. Ratio scaled data may be continuous or discrete.

The ratio scale is the highest classification of the scales (Annex 2). That means that ratio scaled data include the highest information about the characteristic and it is possible to use many statistical procedures (chapter 8).

The examples 1 and 2 are examples for characteristics with ratio scaled data.

The definition of an absolute zero point makes it possible to define additional constant ratios. This is also a requirement for the construction of index numbers (e.g. ratio length to width). An index is the combination of at least two characteristics. In UPOV terms this special case is defined as combined characteristic.

5.1.2. Interval scale

An "interval scale" is a metric scale without a defined absolute zero point. There is always a constant distance (different from zero) between two adjacent expressions. Interval scaled data may be distributed continuously or discrete.

The interval scale is higher classified than the ordinal scale but lower than the ratio scale (Annex 2). That means that interval scaled data include more information about the characteristic than ordinal scaled data and it is possible to use more statistical procedures. In comparison to ratio scaled data interval scaled data include less information about the characteristic and therefore less procedures can be used (chapter 8).

The interval scale is theoretically the minimum scale level to calculate arithmetic mean values.

An example for a characteristic with continuous interval scaled data is the relative measurement “Temperature in °C”. It is probably impossible to find an example for this kind of scale in the Technical Guidelines. But there are examples for characteristics with discrete interval scaled data in many guidelines (e.g. Time of beginning of flowering as date, see example 6).

5.2. Ordinal scale

An “ordinal scale” consists of numbers which correspond to the states of expression of the characteristic (notes). It is possible to rank all the expressions from a lower to a higher level or vice versa. The discrete categories are only denoted by numbers. The distances between the discrete categories are not exactly known and therefore not necessarily equal. It is not possible to change the order of the expressions. But it is not important which kind of numbers is used for the description of categories. The triple 1, 2 and 3 has in this case the same information as 3, 5 and 7 or 1, 2, and 5.

A simple ordinal scale does not fulfil the assumptions necessary to calculate arithmetic mean values. The requirement is the equality of intervals throughout the scale.

The ordinal scale is higher classified than the nominal scale but lower than the interval scale (Annex 2). That means that ordinal scaled data include more information about the characteristic than nominal scaled data. It is possible to use more statistical procedures than for nominal scaled data but less than for interval scaled data (chapter 8).

5.3. Nominal scale

A “nominal scale” consists of numbers or letters which correspond to the states of expressions of the characteristic (notes). It is only possible to decide that two expressions are equal or not. Although numbers are used for denotation there is no ranking in the expressions and so it is possible to change the arbitrary order of the expressions. A special case is a characteristic with only two categories (alternative characteristic).

The nominal scale is the lowest classification of the scales (Annex 2). Nominal scaled data include the lowest degree of information about the characteristic. Only few statistical procedures are applicable for evaluations (chapter 8).

Additionally there are data without clear structure. They may be partly ordinal and partly nominal scaled. In this case the scale has to be considered as a nominal scale (e.g. Flower bud: shape; see example 8).

6. Scale levels for variety description

In the General Introduction and in all Technical Guidelines there are only characteristics with qualitative scaled data used to establish variety descriptions. All quantitative characteristics have to be transformed from measurements (quantitative scaled data) into notes from 1 to 9 if possible (ordinal scaled data) for a better presentation and comparison.

7. Relation between the UPOV categories of characteristics and scale levels

There is no contradiction between the definition of quantitative characteristics in the General Introduction (TC/36/8) and the use of that term in this paper under consideration of scale levels.

7.1. Relation between 'truly qualitative' and 'nominal scaled' characteristics

Truly quantitative characteristics have to be considered as nominal scaled characteristics. An example for this type of characteristic is "Sex of plant" (example 4).

The characteristic "Number of whorls" given in the General Introduction as example for a truly qualitative characteristic would not be classified in that way under consideration of scale levels. The number of whorls is recorded by counting and as explained in chapter 4.2. this kind of assessment provides discrete ratio scaled data (example 7).

"Number of whorls" may appear for the crop expert as a qualitative characteristic because there is only a limited number of states of expression and normally there is no variation within varieties (except off-types). Because of uniformity of the varieties it is sufficient to take one record per variety in order to establish distinctness and to produce a description. Only for the assessment of uniformity it is necessary to look on the whole sample for off-types. Due to the variation between and within varieties and the environmentally independence it is obvious to use a minimum distance of one.

Nevertheless, the underlying characteristic is quantitative and therefore we propose to replace the example in the General Introduction.

Remark:

In the General Introduction (TC/36/8 - April 20, 2000, chap. 10.3.7.3) it is recommended as a rule to use consecutive numbers beginning with one for the states. This requirement has to be considered as a very useful practical rule. It is not necessary to fix any rule depending on the scale level. Therefore the "exception" for the characteristic "Ploidy" causes no problems for any evaluation (par. 185).

7.2. Relation between 'pseudo-qualitative' and 'ordinal scaled' characteristics

In the definition for pseudo-qualitative characteristics it is assumed that the 'underlying' characteristic is quantitative (continuous variation). From the statistical point of view there is no information about an 'underlying' characteristic (latent variable) for the definition of ordinal scaled characteristics necessary . Only for special applications (e.g. use of threshold models) it may be useful to assume that the 'underlying' characteristic has definite properties but it is not a general requirement.

The examples in the General Introduction for pseudo-qualitative characteristics is suitable as example for ordinal scaled characteristics.

7.3. Relation between 'pseudo-qualitative' and quantitative characteristics

Pseudo-qualitative characteristics are defined by UPOV as quantitative characteristics which are handled like qualitative characteristics. From the viewpoint of scale levels this means that the underlying characteristic is of quantitative type but a qualitative (ordinal) scale is used for assessment. Further evaluations have to be done based on the ordinal data.

8. Relation between scale levels and recommended procedures

The different scale levels of data makes it possible to use different statistical procedures. Depending on the scale level and some other conditions (e.g. number of degrees of freedom, unimodality) it is possible to recommend well known statistical procedures. Some examples for distinctness procedures are listed in the following table:

<table>
<tr><th rowspan="2">type of scale</th><th rowspan="2">discrete/ continuous</th><th colspan="4">condition for distinctness</th></tr>
<tr><th>df<20</th><th>df>=20</th><th>Unimodality</th><th>other</th></tr>
<tr><td rowspan="2">ratio</td><td>continuous</td><td rowspan="4">long term LSD or 2*1%</td><td rowspan="4">COY-D</td><td></td><td></td></tr>
<tr><td>discrete</td><td></td><td></td></tr>
<tr><td rowspan="2">interval</td><td>continuous</td><td></td><td></td></tr>
<tr><td>discrete</td><td></td><td></td></tr>
<tr><td>ordinal</td><td>discrete</td><td colspan="2">minimum distance>=1</td><td>threshold model</td><td></td></tr>
<tr><td>nominal</td><td>discrete</td><td colspan="2">minimum distance=1</td><td></td><td></td></tr>
</table>

An analogous table can be constructed for uniformity or for correlation.

9. Conclusions

The UPOV categories of characteristics (qualitative, quantitative, pseudo-qualitative) make no clear separation between the scale levels of the underlying characteristics, the recorded data and the data transformed for the variety description. But in many cases different scale levels are present in different levels of observation or evaluation. Therefore, the UPOV terms for the categories may be misleading for specialists from different sciences (e.g. agriculture, biology, statistics, mathematics). The presented paper shall demonstrate, that it is possible to adjust the UPOV definitions to the commonly used terms.

The TWC should discuss the proposed terminology and the corresponding definitions in order to develop a harmonised proposal which then has to be discussed in all TWP´s. At the end it seems necessary to adjust some parts of the text of the General Introduction and some TGP´s to the harmonised commonly understandable definitions of types of characteristics and data. All information about scale levels of characteristics and data should be given in TGP/8.

[Two Annexes follow]

ANNEX I

Annex 1: Description of the stages of expression of characteristics

Example number	Name of characteristic	Unit of assessment	Description	Scale level
1	Length of plant	cm	assessment in cm without digits after decimal point	ratio scaled continuous quantitative data
2	Number of inflorescences	counts	1, 2, 3, ...	ratio scaled discrete quantitative data
3	Intensity of anthocyan	1	very low	ordinal scaled qualitative data
		3	low	
		5	medium	
		7	high	
		9	very high	
4	Sex of plant	1	dioecious female	nominal scaled qualitative data
		2	dioecious male	
		3	monoecious unisexual	
		4	monoecious hermaphrodite	
5	Leaf blade: variegation	1	absent	nominal scaled qualitative data
		2	present	
6	Time of beginning of flowering	date	e.g. May 21st 1999	interval scaled discrete quantitative data
7	Number of whorls	1	one whorl	ratio scaled discrete quantitative data
		2	two whorls	
		3	three whorls	
8	Flower bud: shape	1	narrow elliptic	nominal scaled qualitative data
		2	elliptic	
		3	round	
		4	asymmetric	

ANNEX II

Annex 2: Type of scale and scale levels

Type of scale		Description		Data recording by	Scale Level
Quantitative (metric)	Ratio	Constant distances with exact zero point	Continuous	Absolute measurements	High
			Discrete	Counting	
	Interval	Constant distances without exact zero point	Continuous	Relative measurements ↑	
			Discrete	Date	
Qualitative (categorial)	Ordinal	Ordered expressions with varying distances	Discrete	Visually assessed notes ↑	
	Nominal	No order, no distances	Discrete	Visually assessed notes	Low

[End of document]