E

TWC/17/6
ORIGINAL: English
DATE: May 27, 1999

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

# TECHNICAL WORKING PARTY ON AUTOMATION AND COMPUTER PROGRAMS

## Seventeenth Session
## Helsinki, June 29 to July 2, 1999

HANDLING OF VISUALLY ASSESSED CHARACTERISTICS

*Document prepared by experts from Germany*

n:\orgupov\shared\document\twc\twc17\twc-17-06.doc

# HANDLING OF VISUALLY ASSESSED CHARACTERISTICS

## Introduction

Since 1987 a series of papers on the handling of visually assessed characteristics was presented in the TWC in addition to the "General Introduction to the Guidelines for the Conduct of Tests for Distinctness, Homogeneity and Stability" (TG/1/2) and the "Harmonization of States of Expression and Notes of Characteristics" (TC/26/4 Rev.).

The majority of characteristics in DUS testing is visually assessed and only a small number is assessed by measurements. With the presented TWC documents (Table 1) the applicability of statstical methods for visually observed characteristics should be promoted for example to prove the empirically determined minimum distances by statistical methods. A summary of these documents is given in the following.

UPOV has decided that for the description of the varieties all results should as a rule be presented in a 1 to 9 scale. Visually assesssed characteristics are generally recorded in this scale whereas measurements and counts have still to be transformed. Only the papers dealing with visually observed characteristics were considered here. Papers about the transformation of measurements and counts into a 1 to 9 scale were not included in this summary.

Table 1: Chronological list of papers on visually observed characteristics

| **Document** | **Date** | **Title** | **Author** | **Country** | **Contens***) |
|---|---|---|---|---|---|
| TG/1/2 | 14.11.1979 | Revised General Introduction to the Guidelines for the Conduct of Tests for Distinctness, Homogeneity and Stability of New Varieties of Plants | | UPOV | D/U |
| TWC/V/4 | 17.02.1987 | Logical Order of States of Expression in Test Guidelines | Laidig | DE | D |
| TWC/VI/3 | 23.03.1988 | The Use of Nonparametric Statistics in the Testing of Distinctness, Homogeneity and Stability | Baltjes, van der Heijden | NL | D/U |
| TC/26/4 | 12.10.1990 | Harmonization of States of Expression and Notes of Characteristics | | UPOV | D/U |
| TWC/10/8 | 19.05.1992 | Handling of Visually Observed Characteristics | Laidig | DE | D |
| TWC/11/12 | 17.06.1993 | Handling of Visually Observed Characteristics | Laidig | DE | D |
| TWC/13/9 | 19.04.1995 | Homogeneity Criterion for Visually Assessed Characteristics in Turnip Rape | Talbot | UK | U |

| | | | | | |
|---|---|---|---|---|---|
| TWC/13/14 | 18.05.1995 | A Biometrical Evaluation of Visually Observed Characteristics in French Beans | Laidig | DE | D |
| TWC/14/12 | 04.06.1996 | Threshold Models for Visually Observed Data | van Eeuwijk, Jansen | NL | D/U |
| TWC/15/14 | 13.05.1997 | Analysing visually observed data in Two Grass Species | van Eeuwijk | NL | D/U |
| TWC/15/-- | 1997 | A Study on Visually Observed Characters | Veress | H | D |

*) dealing with distinctness (D) and/or uniformity (U)

**Type of Scales**

A good introduction into the types of scales is presented in the documents TWC/VI/3 and TWC/10/8. The different scales and corresponding examples are summarized in table 2.

Table 2: Description of scales

| **Type of scale** | DESCRIPTION | **Discrete/ Non-discrete** | **Example** |
|---|---|---|---|
| Nominal scale (qualitative) | independent and exchangeable classes, different size of classes | Discrete | Colour of ear: White=1, coloured=2 |
| Ordinal scale or Rank scale (qualitative) | independent and non-exchangeable classes, different size of classes | Discrete | Attitude: erect=1, semi-erect=3, intermediate=5, semi-prostrate=7, prostrate=9 |
| Interval scale (quantitative) | Classes of fixed size and without exact zeropoint | Discrete | Date |
| | | Non-discrete | Relative measurements (Temperature in C) |
| Ratio scale (quantitative) | Classes of fixed size and with exact zeropoint | Discrete | Counts (number of plants) |
| | | Non-discrete | Absolute measure-ments (plant height) |

**Logical Order of States of Expression in Test Guidelines (TWC/V/4)**

Qualitative characteristics whose states can be arranged in a logical order (Order of ranks) should receive the notes in that way. But this is not always clear, especially when the characteristic can be arranged in several ways depending on the sorting criterion. That arrangement should be used which produces the best scale for distinctness assessment. The concrete problems are explained by examples (shape of fruit, shape of ear).

**The Use of Nonparametric Statistics in the Testing of Distinctness, Homogeneity and Stability (TWC/VI/3)**

Nonparametric or so-called distribution-free methods are suitable for distinguishing varieties by visually observed characteristics. Nonparametric methods have no important requirements on the collected data (no normal distribution, no equivalence of variances). In comparison to parametric methods nonparametric methods do not have always a loss of power. They are not as sensitive to outliers as parametric tests. Different statistical methods are appropriate depending on the type of scale (see Table 3). The paper gives a general summary on the various statistical methods.

Table 3: Statistical methods depending on the type of scale

| **Type of scale** | **One sample** | **k independent samples** |
|---|---|---|
| Nominal scale | Chi-square-Test | Chi-square-Test |
| Ordinal scale | Median | Median-Test |
| Interval scale/ Ratio scale | Mean, Standard deviation | ANOVA, t-Test |

**Handling of Visually Observed Characteristics (TWC/10/8)**

Problems of types of scales were discussed especially with regard to sampling sizes in DUS testing. Concrete problems were demonstrated on the basis of a winter wheat data set. The study deals with those visually observed characteristics which have an ordinal scale with 9 classes. Characteristics whose observations concentrate on 2 or 3 scores were not included. All characteristics had only one observation per variety and year. If there is more than one observation per variety experience shows that parametrical methods are often applicable supplying valid results (at least 5 observations per plot). This is the case for visually observed characteristics on single plants of cross pollinated crops.

Conclusions:

- Descriptive statistics provide an intuitive understanding of the data to the DUS experts additionally to his experience.
- Attention should be paid to strongly correlated characteristics and it should be checked if some of them could be dropped in order to save time and costs.
- ANOVA can be applied for exploratory use if the data have a 1 to 9 scale with a non degenerated sampling distribution. F-values for years allow to identify characteristics which are highly influenced by year effects
- Large F-values for varieties indicate characteristics with high discriminative power and therefore they are important for establishing distinctness. On the other hand it should be checked if characteristics with low F-values for varieties are dispensable.
- Comparison of LSD-values derived by ANOVA with the minimum distance derived by expert knowledge provides a helpful check.

- The study should be considered as a first step to find out how to handle visually observed characteristics. More experience is required by extending studies to other species (i.e. ornamentals) and including characteristics with nominal scales.
- The methods available should be used by experts drafting new technical guidelines or revising existing ones.

**Handling of Visually Observed Characteristics (TWC/11/12)**

In this study document TWC/10/8 was continued for Pelargonium as an ornamental example. The Pelargonium varieties were vegetatively propagated as normal, and they were tested under glasshouse conditions for one growing period (10 plants per variety). Measurements (mean of 10 single plants) and ordinal scaled characteristics (one observation per plot) were included in the data analyses as well.

The statistical methods discribed in document TWC/10/8 were applied (descriptive statistics, correlations, analysis of variances, LSD). Additionally two way frequency tables were used in order to get better information about the association between two characteristics than correlation coefficients provide.

Conclusions:

- The use of simple summary statistics together with histograms can aid in checking characteristics if the applied scaling is appropriate.
- Strongly correlated characteristics should not be used for distinctness any longer.
- The ANOVA results of visually assessed characteristics confirm that the applied minimum distance corresponds to a COY-D LSD with significance level below 1%.
- The minimum distance established by expert knowledge is rather small for most measured characteristics. The distances applied have a significance level above 10% when compared with COY-D LSD values.
- Studies of this kind are suitable to help statisticians to better understand the crop specific problems in DUS testing.

**Homogeneity Criterion for Visually Assessed Characteristics in Turnip Rape (TWC/13/9)**

Turnip rape is a crosspollinating crop. Relative uniformity standards have to be applied. The question is how to meet this requirement of the UPOV guidelines (TG/1/2, paragraph 32) for visually assessed characteristics. Three methods for solving the problem are discussed:

- Chi-square-test for two-way-tables (1 degree of freedom)
- ANOVA for the frequencies (with and without transformation respectively) and t-test
- Using a linear logistic model for binary data und construction of contrasts

The advantages and disadvantages of all methods are discussed. The second method is preferred in UK because of the simpler approach.

**A Biometrical Evaluation of Visually Observed Characteristics in French Beans (TWC/13/14)**

The known methods from TWC/10/8 and TWC/11/12 should be proven for applicability to an example with French beans. Then following questions were important for the next revision of the technical guidelines:

- Are the numbers of states of expression, as specified in the test guidelines for individual characteristics, still appropriate?
- Which characteristics are strongly correlated, and if so, which of them can possibly be dropped?
- Which characteristics have a low discriminative power and should better not be included in the guidelines?
- How large should the minimum distance of a visually observed characteristic be in order to establish distinctness (2 or 3 notes from a 1 to 9 scale)?

Recommendations are given to the following problems:

- Combination of two characteristics to only one
- Checking of scaling for 6 characteristics
- Dropping of characteristics (low discrimination power) for 7 characteristics
- Checking of minimum distances for 3 characteristics

Questions and opinions (special characteristics in French beans) received from Slupia Wielka (PL) are answered and commented.

**Threshold Models for Visually Observed Data (TWC/14/12)**

The aim of this paper was the description of a class of models especially developed for ordinal data (threshold models). These models allow the same kind of questions to be answered for ordinal data as linear models do for measured characteristics. Threshold models are more general than linear models as they provide a means to model expectation and variance simultaneously, thus effectively combining the examination of distinctness and homogeneity.

At first basic theoretical problems were discussed for general linear models and then threshold models were introduced. Threshold models assume that there is an underlying variable in the background, which cannot be observed. It is assumed that this variable is normally distributed (or logistic). The range of the variable is divided into intervals which are restricted by so-called "thresholds". The categories of the ordinal variable are assigned to these intervals. The models estimate the thresholds and calculate indirectly probabilities for the categories of the ordinal data.

The applicability of threshold models was demonstrated by two examples from TWC/13/3. Estimated thresholds were given and interpreted. The main effect for varieties was proven by the WALD-Statistik. Means and dispersion parameters including a t-test to the reference variety were presented. Data from 3 years were included in the second example. The interaction was considered first as fix and then as random (mixed model).

The document is useful for the presentation of the methods to the members of the TWC working group. Examples are not directly from the DUS test. Some comments about literature and software complete the paper.

**Analysing Visually Observed Data in Two Grass Species (TWC 15/14)**

The application of threshold models for the assessment of distinctness and uniformity was shown with two data sets obtained from routine DUS testing in *Dactylis* and *Festuca*. At first the theory and the assumptions of the method were briefly summarized. The author refers especially to the uni-modality of the underlying variable and to their examination. Categories in the examples had to be fused to achieve uni-modality. Because of the low number of remained categories only distinctness testing was possible. At least 4 or 5 non-empty categories are necessary for uniformity testing.

Results obtained by applying threshold models were compared to those of ANOVA on logit transformed cumulative frequencies in *Dactylis*, as there were only two remaining categories. In contrast to the COY-D analysis each pairwise comparison of varieties has an individual standard error in threshold models (!). The results of COY-D analysis (logit transformed data) and those of threshold models are in good agreement in the *Dactylis* example. Both methods indicate the presence of interaction between varieties and years. It is proposed to reduce this interaction by the "Modified joint regression method". Tables of variance and deviance show clearly parallels.

Interpretation of results (especially of tables 3 and 10) are not always easy to understand.

**A Study on Visually Observed Characters (TWC/15/--)**

In Hungary distinctness, uniformity and stability are mainly assessed with visually observed characteristics and the variety descriptions are based on these characteristics. Generally only one or two observations (one per year) are available. Different observations from year to year cause difficulties for the preparation of variety descriptions.

A proposal was given to practically solve these problems. Using the maximum differences between years and knowing the minimum distances for these characteristics 32% of the data are greater than or equal to the minimum distances. The same value for standardized variables is 13%.

U. Meyer
Bundessortenamt
Hannover

[End of Document]