E

TWA/38/3 Add.
ORIGINAL: English
DATE: September 1, 2009

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

TECHNICAL WORKING PARTY FOR AGRICULTURAL CROPS

Thirty-Eighth Session
Seoul, August 31 to September 4, 2009

ADDENDUM TO TGP DOCUMENTS

Document prepared by the Office of the Union

1. At its twenty-seventh session, held in Alexandria, Virginia, United States of America, from June 16 to 19, 2009, the Technical Working Party on Automation and Computer Programs (TWC) discussed document TGP/8/1 Draft 13. With respect to Section 5.3 of Part II it was agreed that Mr. Hulse (Australia) would prepare a new draft of the section for circulation by the Office to the TWC by July 17, 2009, with a request for comments to be provided by July 31, 2009. On the basis of comments received, Mr. Hulse would prepare a document by August 3, 2009, to be presented to the Technical Working Party for Agricultural Crops (TWA) and subsequent Technical Working Party sessions in 2009.

2. The Office issued Circular E-1036 asking for comments on Sections 5 and 6 of Part II of TGP/8/1 Draft 13, as agreed by the TWC. On July 23, 2009, Mr. Kristian Kristensen (Denmark) sent a reply with comments, which did not reach Mr. Hulse. The annex to this document contains the comments sent by Mr. Kristian Kristensen.

[Annex follows]

ANNEX

COMMENTS ON SECTIONS 5 AND 6, PART II OF TGP/8/1 DRAFT 13
SENT BY MR. KRISTIAN KRISTENSEN

Section 5.2.2.: I suggest that "relevant criteria" to be replaced with "statistical assumptions".

Comment [c4] in section 5.3.1. I am not sure how this can be moved to Part I because this note does not apply to all statistical methods but only to those were the random variation is determined by the underlying distribution (usually pure Poisson, binomial or multinomial), i.e. here to the tests in contingency tables (either based on Chi-square tests or Fisher exact test). So this note does not apply to analysis of variance methods or use of LSD-values when data from individual replicates or years are analysed assuming data to be normally distributed.

Section 5.3.8. Here it is stated that: "Hence the distribution of scores in different classes observed for this reference variety is considered to be the expected distribution". I do not think that the values in this column can be called expected values as they are samples from a large population and may thus deviate from the expected distribution. So I think that this section should be reformulated.

Section 5.3.9. As a consequence of my comments to section 5.3.8 the value in the table and the calculations are wrong. I suggest deleting the whole section.

Section 5.3.11. I suggest the text to read "For comparing reference variety 1 with the first generation of the candidate we get:" Alternatively the aggregated value of both candidate varieties could be compared with the reference variety.

Section 5.3.11. As a consequence of my comments to section 5.3.8 the formula for the Chi-square should read: (34-23.51)^2/23.51+(6-14.82)^2/14.82+(6-7.67)^2/7.67
+(12-22.49)^2/22.49+(23-14.18)^2/14.18+(9-7.33)^2/7.33=21.05
(using the formulas in section 2 of TWC/27/14).

Section 5.3.12. If kept I suggest writing: "When comparing two varieties the number of degrees of freedom are one less the number of classes."

Section 5.3.14. Using the method given above (and described in TWC/27/14) I do not get the revised values.

Section 6.1.6 to section 6.1.11. I have just realised that the description of the test seem to be in error. The reason is that the description seems to focus on the probability of getting the actual outcome if the distribution is the same for both varieties (section 6.1.6). I think that the probability should be calculated as the probability of getting a result that is at least as extreme as the actual one (which is similar to the probability that is calculated when using e.g. t-tests for normally distributed data). For the example in table 1 I think that there are in total 4 outcomes that are at least as extreme as the actual one - and have the same marginal’s. These are the following:

4 9
8 3,

3 10
9 2,

2 11
10 1 and,

1 12
11 0

The first one is the actual one. The probabilities of each of those 4 outcomes are 0.04362, 0.00582, 0.00032 and 0.0000048. Summing those 4 values I get 0.04977 which rounded to 2 decimal points are 0.05 and thus different from the result stated in section 6.1.8. Note that I have here assumed a one-tailed test. (If a two tailed test is required there are more outcomes that are as extreme as the actual one and the p-value gets larger.)
For the example in section 6.1.10 I think that there are 2 outcomes that there are as extreme as the actual one:

1 9
11 3 and

0 10
12 2

These have probabilities 0.001346 and 0.000034 which sums to 0.001380.
This viewpoint is supported by several programs that can be found on the internet, e.g. *http://faculty.vassar.edu/lowry/fisher.html*.

Section 6.1.10: The fraction line is missing in the expression for p.

Kristian Kristensen
Aarhus University
Faculty of Agricultural Sciences
Department of Genetics and Biotechnology

July 22, 2009

[End of Annex and of document]