E

TWC/25/14
ORIGINAL: English
DATE: August 14, 2007

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

TECHNICAL WORKING PARTY ON AUTOMATION AND COMPUTER PROGRAMS

Twenty-Fifth Session
Sibiu, Romania, September 3 to 6, 2007

A RATIONALE FOR EXCLUDING VARIETIES OF COMMON KNOWLEDGE FROM THE SECOND GROWING CYCLE WHEN COYD IS USED

Document prepared by an expert from the United Kingdom

A RATIONALE FOR EXCLUDING VARIETIES OF COMMON KNOWLEDGE FROM THE SECOND GROWING CYCLE WHEN COYD IS USED

Mr. Adrian M.I. Roberts
Biomathematics & Statistics Scotland (BioSS)
James Clerk Maxwell Building
The King's Buildings
Edinburgh EH9 3JZ
United Kingdom

Email: adrian@bioss.ac.uk

INTRODUCTION

1. In cases where the duration of the DUS test is two independent growing cycles, results are reviewed after the first year of testing, in order to exclude varieties of common knowledge that are clearly distinct from the candidates (see document TGP/9 “Examining Distinctness”). When COYD is used to assess distinctness in a characteristic, no formal mechanism has yet been described to inform such early decisions on distinctness.

2. In document TWC/24/9, a possible approach was described. This document reiterates the description of the approach and extends the work on an example application. This is still “work in progress”.

Objective

3. The aim of this approach is to estimate the probability that a candidate will be distinct on the 2-year COYD criterion from a particular variety of common knowledge after the first year of test. In order to judge the variability associated with measurements in a particular characteristic we need to have past data. If the probability is suitably large, the candidate is declared distinct from a particular variety and does not need to be compared in the second year.

Mathematical details

4. Two varieties, A and B, are tested in two years, labelled 1 and 2. Mean measurements, x_{ij}, are made in the character of interest for each variety, i, and year, j. Let the difference d_j in year j, be given by:

$$d_j = x_{Aj} - x_{Bj}$$

and assume that it is normally distributed. Also let the COYD difference after two years be $D=(d_1+ d_2)/2$. The COYD criterion says that variety A and variety B should be considered distinct if

$$\left| \frac{d_1 + d_2}{2 s_{12}} \right| \geq t_{1-p/2,\nu_{12}}$$

where $t_{1-p/2,\nu_{12}}$ denotes the 1-p/2 quantile of the student t-distribution with ν_{12} degrees of freedom and s_{12} is the square root of the residual variance for the two year COYD analysis of variance (with year and variety effects removed).

5. We wish to estimate the probability p_D that A and B will be considered distinct after two years of tests, given the first year result, d_1, and the historical data, $\underline{x}$, i.e.

$$\Pr\left(\left|\frac{d_1+d_2}{2s_{12}}\right| \geq t_{1-p/2,\nu_{12}} \middle| d_1, \underline{x}\right) = p_D. \qquad (1)$$

Conditional on d_1 being known, under the null hypothesis of no difference between varieties, d_2 has a Normal distribution with mean d_1 and standard deviation sqrt(2)* σ_{12}. If we assume that the true value, σ, which in fact is to be estimated from the historical data, were also known and that s_{12}^2 divided by σ^2/ν_{12}, is chi-squared distributed with degrees of freedom ν_{12}, then equation (1) could then be estimated from a non-central t distribution (D*sqrt(2)/s_{12} has non-centrality parameter sqrt(2)* d_1/ s_{12}).

6. In truth, exact knowledge of σ is an approximating assumption: nevertheless it is instructive to calculate the threshold values for d_1 under this assumption for given target values for p_D.

Considerations

7. A choice of level of p_D needs to be made. If a value of 0.5 (50%) is used, then the resulting threshold will result in even odds of making an incorrect early decision. Instead a larger value of p_D is required, such as 99% or 99.9%

8. The long-term COYD LSD (based on a 2-year test) is the same as the threshold produced when p_D is 50%.

9. In this first step of development, two approximations have been used to simplify the mathematics and to facilitate discussion of the approach. It is intended to investigate robustness to deviations from the normality assumption, and how to allow for imperfect knowledge of σ_{12}. The implicit assumption that σ is applicable to all years and all varieties also needs to be verified. An example application given below explores some of these aspects.

10. A Bayesian approach to this problem could allow the use of information from sources other than the first year's experiment, e.g. from the Technical Questionnaire. It should be possible to represent the expert's judgement on the quality of this information.

11. An alternative approach to the one described might use the ideas of sequential testing.

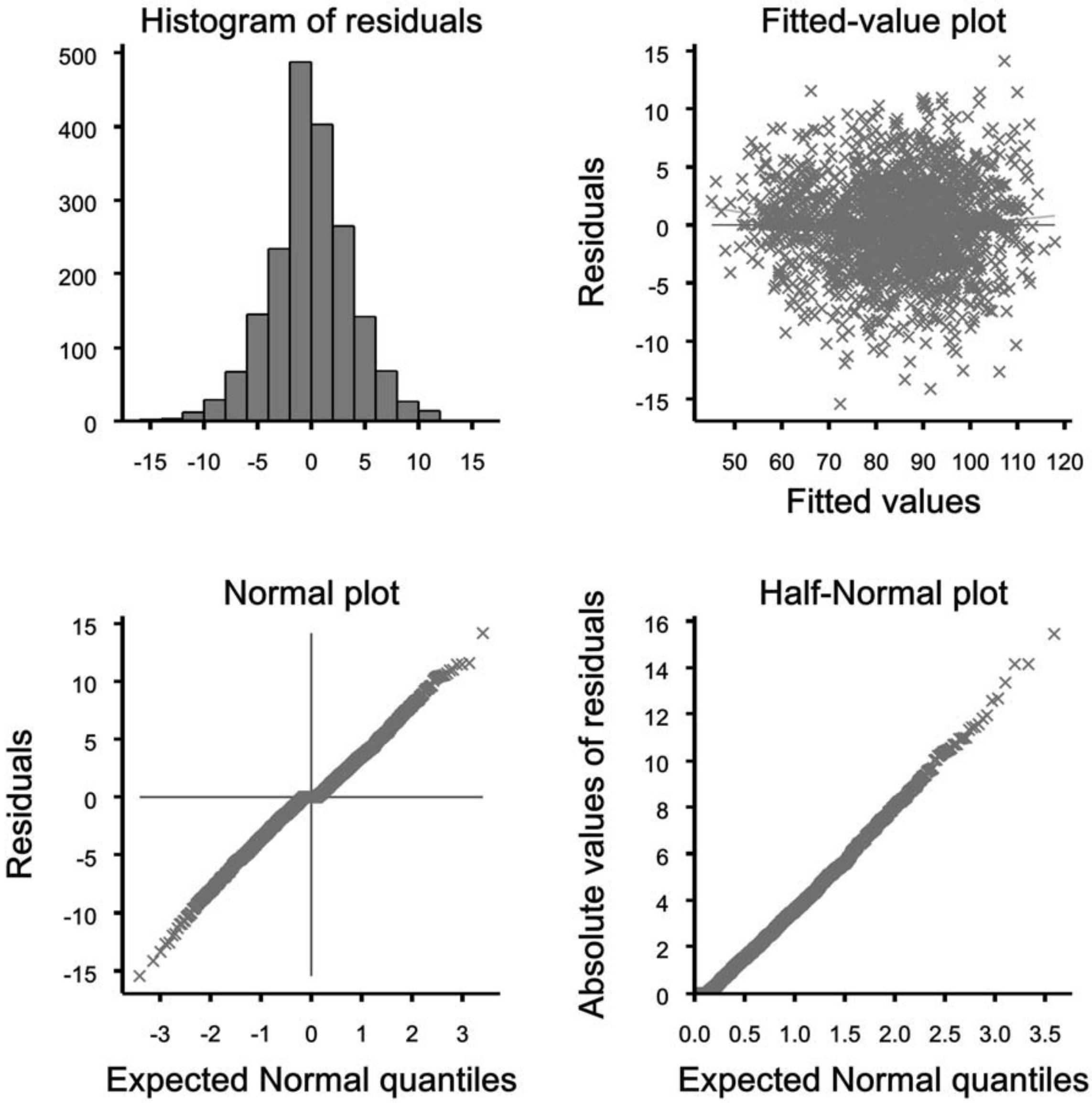
An example application

12. The assessment of characteristic "stipule length", for field pea, is used as an example. The historical data x is based on UK semi-leafless pea data from 1995-2004. COYD is used with a probability level of 2%. It is assumed that the number of degrees of freedom in the current 2-year test is large so that D is approximately normally distributed in this example.

13. The long-term 2% LSD for a 2-year test based on the 10 years of historical data is 10.64 mm (note that the data ranges from 45.0 mm to 121.5 mm). For comparison, the long-term 2% LSD for a single year test is 15.04 mm. The table below gives the required approximate thresholds for the first year difference d_1 to obtain a p_D probability of being distinct after the second year of tests.

p_D	d_1 threshold
99.9%	±20.63
99%	±18.16
98%	±17.28
95%	±15.95
90%	±14.78
80%	±13.36
50%	±10.64

14. Normality of the means is a key assumption of this method, being more important than for the COYD test itself. The figures below show residual diagnostics for a mixed model analysis of the stipule length data (year is treated as a random effect and variety as a fixed effect). The assumption of normality seems reasonable.



15. Next the assumption that σ is applicable to all years is investigated. First, the data set was reduced to those varieties present in all 10 years – there were 15 varieties. The table below shows the individual variances of the varieties and years along with an associated F-test p-value. The F-test is for the comparison of the individual variance with the residual variance for the data set with the particular variety or year removed. The degrees of freedom for the individual variances were low, particularly for variety variances which had just 9 degrees of freedom. The years seem to have fairly consistent variances. There is just one odd variety, H, for which there seems to be evidence for a lower than average variance (overall p-value based on the deviance test was 0.24).

Year	Variance	p-value
1995	13.58	0.37
1996	10.02	0.79
1997	10.58	0.75
1998	12.59	0.59
1999	15.74	0.34
2000	12.98	0.56
2001	10.58	0.75
2002	16.98	0.26
2003	11.04	0.72
2004	15.27	0.38

Variety	Variance	p-value
A	15.45	0.38
B	12.65	0.55
C	5.92	0.935
D	9.34	0.77
E	17.94	0.25
F	24.69	0.069
G	10.93	0.66
H	3.52	0.989
I	17.96	0.25
J	14.15	0.45
K	19.49	0.19
L	10.93	0.66
M	11.25	0.64
N	18.49	0.23
O	8.5	0.82

16. It is possible to test more formally whether σ is applicable to all years and varieties by fitting a mixed model and applying particular variance structures to the (random) year-by-variety interaction term. For more information on variances structures, see e.g. the GenStat Statistics manual. The structure of the interaction term is defined using a direct product formulation, for example $diag(year) \otimes id(variety)$. Here, *diag* refers to a diagonal matrix structure (so in this case, one variance parameter for each year) and *id* refers to an identity matrix structure (so just one variance parameter over all varieties). These more complex models can be compared to the simplest model ($id(year) \otimes id(variety)$ – just one variance parameter) using deviances. For this data set, no evidence was found for overall differences in variance for year-to-year (p-value 0.96) or variety-to-variety (p-value 0.24).

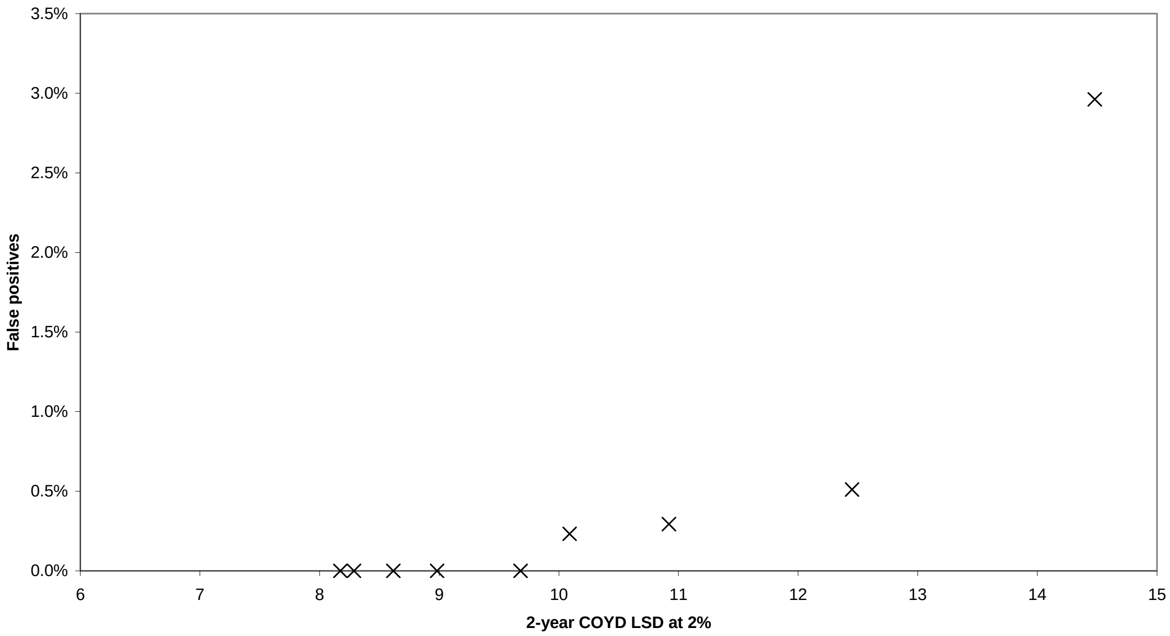
17. However, this analysis is unsatisfactory because the data set used with only varieties present in all years represented a small fraction of the whole data set. There are only 15 varieties of the total 591 and 150 observations of a total of 1900. The varieties represent only two groups out of the total of 16. However, it is more difficult to make a sensible comparison of the yearly residual variances in the full, unbalanced data set.

18. The tolerances are next applied to the historical data set to evaluate how well they perform. COYD analyses were carried out on each of the 9 consecutive pairs of years (1995-6, 1996-7 etc.), having first identified candidate varieties and varieties of common knowledge. In the table below, predicted decisions based on the first year are compared to COYD decisions after two years (labelled D and ND for distinct and non-distinct). As p_D decreases so more varieties of common knowledge are predicted to be distinct from the candidates after the first year and so potentially could be eliminated from the second year of trials. However this comes at a cost as more comparisons that are actually non-distinct (on the basis of COYD) after two years are predicted to be distinct after the first year (false positives). This proportion needs to be very small.

	Frequency of correct predictions of COYD decision (no. of decisions)	
p_D	*ND COYD (16751)*	*D COYD (7167)*
99.9%	99.3%	18.5%
99%	98.3%	28.6%
98%	97.8%	32.8%
95%	96.7%	39.7%
90%	95.4%	46.5%
80%	93.1%	54.9%
50%	84.9%	73.3%

19. The proportion of false positives varies considerably from year-to-year according to the size of the COYD residual error. This is illustrated below. For some pairs of years, with smaller LSDs, there are no false positives. However for at least one pair of years, 2002-3, the rate of false positives is unacceptably high. The other pairs of years with non-zero rates are 1998-9, 1990-1 and 2003-4.

Frequency of false positives for first year tolerance with p_D=99.9%



20. An example of a false positive is given below, based on data from 2002-3:

Variety	First Year	Second Year	Overall
Reference X	94.45	80.60	87.53
Candidate Y	117.65	85.45	101.55
Difference	23.2	4.85	14.02
Tolerance (p_D=99.9% in first year, COYD 2% overall)	20.63	n/a	14.48

21. This problem is due to the variability of the COYD LSD threshold from year-to-year. It is clearly important to accommodate this feature, although a larger first year threshold will be the result.

22. It should be noted that, after the first year, some varieties of common knowledge will have been eliminated because they are sufficiently different from candidates. This may have biased the analysis above. In particular, the proportion of false negatives may have been over-estimated.

Conclusions & future work

23. The approach is worth consideration: in practice, first-year tolerances for elimination of varieties of common knowledge are widely used, either explicitly or implicitly. These are based on expert judgement. This approach formalises the idea.

24. Work is required to accommodate properly variability in the COYD thresholds that naturally occur from year-to-year.

25. The methodology developed so far is based on assumptions of normality. This will be fine for many characteristics, but consideration is required as to how to extend the concept to other characteristics that do not meet this criterion.

26. In principle, the method could be applied in other circumstances, such as to produce tolerances for use after two years of tests when the duration of the DUS test is three independent growing cycles.

[End of document]