E

TWC/28/30
ORIGINAL: English
DATE: June 21, 2010

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

TECHNICAL WORKING PARTY ON AUTOMATION AND COMPUTER PROGRAMS

Twenty-Eighth Session
Angers, France, June 29 to July 2, 2010

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

Document prepared by experts 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 & Mr. Ian M. Nevison
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 cycle of testing, in order to exclude varieties of common knowledge which 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/25/14, a possible approach was described for measured characteristics. However this approach makes some strong assumptions regarding the properties of the data. This document reiterates the description of the approach, explores the effect of deviations from these assumptions and suggests possible solutions that may be employed if necessary.

Overview

3. The aim of this approach is to identify after the first cycle of testing which varieties of common knowledge are so different from the candidate that they do not need to be compared in the second cycle. To enable this, we estimate the probability that a candidate would be distinct on the 2-cycle COYD criterion from a particular variety of common knowledge, given the results from the first cycle of testing. If the probability is suitably large, the candidate is declared distinct from that variety and does not need to be compared in the second growing cycle. The method is applied characteristic by characteristic. In order to judge the variability associated with measurements in a particular characteristic we need to have past data. The approach might be used in combination with processes such as GAIA.

Mathematical details

4. We make a number of assumptions about the data in order to set up the method in a relatively straightforward way. The data that we are considering are the variety means in each year (or growing cycle). We assume that these are independent and normally distributed, and that their variances are the same in each cycle and for each variety and equal to σ^2. We assume that the historical data is large (so many degrees of freedom) and so would be the (potential) two-cycle data set used for COYD.

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

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

based on the assumptions that d_j is i.i.d for different cycles, being normally distributed with mean θ and variance $2\sigma^2$.

6. Also let the COYD difference after two cycles, D, be given by $D = \frac{d_1 + d_2}{2}$. The COYD criterion says that variety A and variety B should be considered distinct if

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

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 cycle COYD analysis of variance (with growing cycle and variety effects removed).

7. We wish to estimate the probability p_D that A and B will be considered distinct after two cycles of tests, given the first cycle result, d_1, and the historical data, $\underline{\mathbf{X}}$, i.e.

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

8. Using equation (1), we can then develop tables for threshold values of d_1, given various values of p_D.

9. Conditional on d_1 and σ (for now) being known, the prediction distribution of d_2 is normal with mean d_1 and variance $2\sigma^2$. For proof, consider the distribution of d_1-d_2 [see http://en.wikipedia.org/wiki/Prediction_interval, also c.f. Bayesian predictive theory for a normal model with known variance and a flat prior].

10. Thus $D = \frac{d_1 + d_2}{2}$ has a normal distribution with mean mean d_1 and variance σ^2. Also $\nu_{12} \frac{s_{12}^2}{\sigma^2}$ is chi-squared distributed with ν_{12} degrees of freedom. So $\frac{d_1+d_2}{2s_{12}} = \frac{D}{s_{12}}$ has a non-central t-distribution with ν_{12} degrees of freedom and non-centrality parameter $\frac{d_1}{\sigma}$.

11. If ν_{12} is very large then $\frac{D}{s_{12}}$ is approximately normal with mean $\frac{d_1}{\sigma}$ and variance 1. It is then straightforward to calculate tables. Since the lower probability is very small, the threshold for d_1 is given by $\sigma(\Phi^{-1}(1 - p/2) + \Phi^{-1}(p_D))$.

Considerations

12. 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 90%.

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

14. A number of assumptions have been made:

a) Both the historical data set and two-cycle data sets are sufficiently large. This seems a reasonable restriction on the use of the approach given its main application is to help manage large reference collections.

b) That the errors associated with variety by cycle means are normally distributed.

c) That the variance for variety by cycle is constant and not dependent on cycle on variety.

15. We demonstrate below how to assess the extent to which the data for a characteristic deviate from assumptions (b) and (c) using a real data example. We examine how robust the approach is to different levels of deviation and consider solutions for cases where validity of the approach would be affected. In particular for assumption (b), we look at skewness and levels of kurtosis.

16. Skewness is the degree of asymmetry of a distribution (in this case of the errors associated with variety by cycle means) –see http://en.wikipedia.org/wiki/Skewness for more details. An index can be calculated to measure the skewness of a sample from a distribution. This is zero for a symmetric distribution, such as the normal, and other values indicate a positive or negative skew to the distribution.

17. Kurtosis is the degree which the distribution is peaked or flat compared to the normal - http://en.wikipedia.org/wiki/Kurtosis. Again an index can be calculated for samples, so that a normal distribution gives a value of (about) zero. Positive values indicate a distribution with a sharper peak and longer, fatter tails, whilst negative values indicate a more rounded peak and shorter thinner tails.

An example application

18. To demonstrate the approach, we refer to a 10 cycle dataset for field pea from UK field trials from 1995 to 2004 (using the semi-leafless group). COYD is used with a probability level of 2%.

19. We show the thresholds for a particular characteristic, stipule length. The long-term 2% LSD for a 2-cycle test based on the 10 cycles 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

cycle test is 15.04 mm. The table below gives the required thresholds for the first cycle difference d_1 to obtain a p_D probability of being distinct after the second cycle of tests.

p_D	d_1 threshold
99.9%	±24.77
99%	±21.27
98%	±20.03
95%	±18.16
90%	±16.50
80%	±14.48
50%	±10.64

20. The sample skewness and kurtosis indices are given in the table below with their standard errors for the range of quantitative characteristics measured. Note that characteristic 46 is currently visually assessed and distinctness decisions are made on the basis of a two-state difference rather than COYD. However, we include it for comparison. By comparison with the standard errors, we can see that characteristic 71 is both skew and has positive kurtosis (possibly because it is skew). This is perhaps not surprising since it is a count. Characteristic 46 also has positive kurtosis. Other characteristics have lesser degrees of skewness and kurtosis. The question naturally arises: how robust is the proposed approach to different levels of skewness and kurtosis?

Characteristic		Skewness	SE	Kurtosis	SE
01	Plant height at flowering	-0.18	0.056	1.7	0.11
03	Petiole length	-0.05	0.056	0.7	0.11
07	Peduncle length	-0.08	0.056	0.7	0.11
10	Days to first flowering	-0.31	0.056	2.7	0.11
12	Days to 80% flowering	0.14	0.056	2.5	0.11
25	Standard width	-0.07	0.056	0.9	0.11
41	Stipule length	-0.09	0.056	0.7	0.11
44	Stipule width	0.01	0.056	0.9	0.11
46	Intensity of foliage colour	0.64	0.069	6.3	0.14
71	Number of nodes up to the first fertile node	-1.29	0.056	10.7	0.11
74	Pod length	-0.06	0.056	1.2	0.11
75	Pod width	0.01	0.056	1.2	0.11
76	Ovule number	0.06	0.056	1.7	0.11
80	100 seed weight	0.21	0.058	1.6	0.12

21. We can study whether the assumption that the variance for variety-by-cycle is constant using mixed models (we use the GenStat REML procedure). Different variance structures can be fitted to the random cycle-by-variety term (for more information see the GenStat Statistics Guide). The structure of the interaction term is defined using a direct product formulation, for example *diag(cycle)* ⊗ *id(variety)*. Here, *diag* refers to a diagonal matrix structure (so in this case, one variance parameter for each cycle) 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(cycle)* ⊗ *id(variety)* – just one variance parameter) using deviances.

22. Here we examine characteristic 41 and fit mixed models to a reduced data set consisting of only those varieties with four or more cycles of data. This gave a data set of 228 varieties

with 1320 rows of data. We fitted four different models, all with variety as a fixed effect and cycle as a random effect, and obtained the following deviances:

Model	*Deviance*	*Degrees of freedom*
id(cycle) ⊗ id(variety)	4779.07	1090
diag(cycle) ⊗ id(variety)	4712.93	1081
id(cycle) ⊗ diag(variety)	4477.50	863
diag(cycle) ⊗ diag(variety)	4430.61	854

23. The inclusion of one variance parameter for each variety (*id(cycle) ⊗ diag(variety))* was not a statistically significant improvement to the model as judged by the reduction in deviance compared to the simplest model (*id(cycle) ⊗ id(variety))*. In contrast the inclusion of one variance parameter for each cycle (*diag(cycle) ⊗ id(variety))* resulted in a highly statistically significant (P<0.001) reduction in the deviance compared to the simplest model. While the (*diag(cycle) ⊗ diag(variety)) model* gave a statistically significant reduction in deviance compared to the (*diag(cycle) ⊗ id(variety))* model, it was evident from comparing the change in deviance to the change in degrees of freedom that statistical significance was due to the extent of data rather than the magnitude of the improvement in fit. Thus only the inclusion of one variance parameter for each cycle would appear to be justified in the case examined.

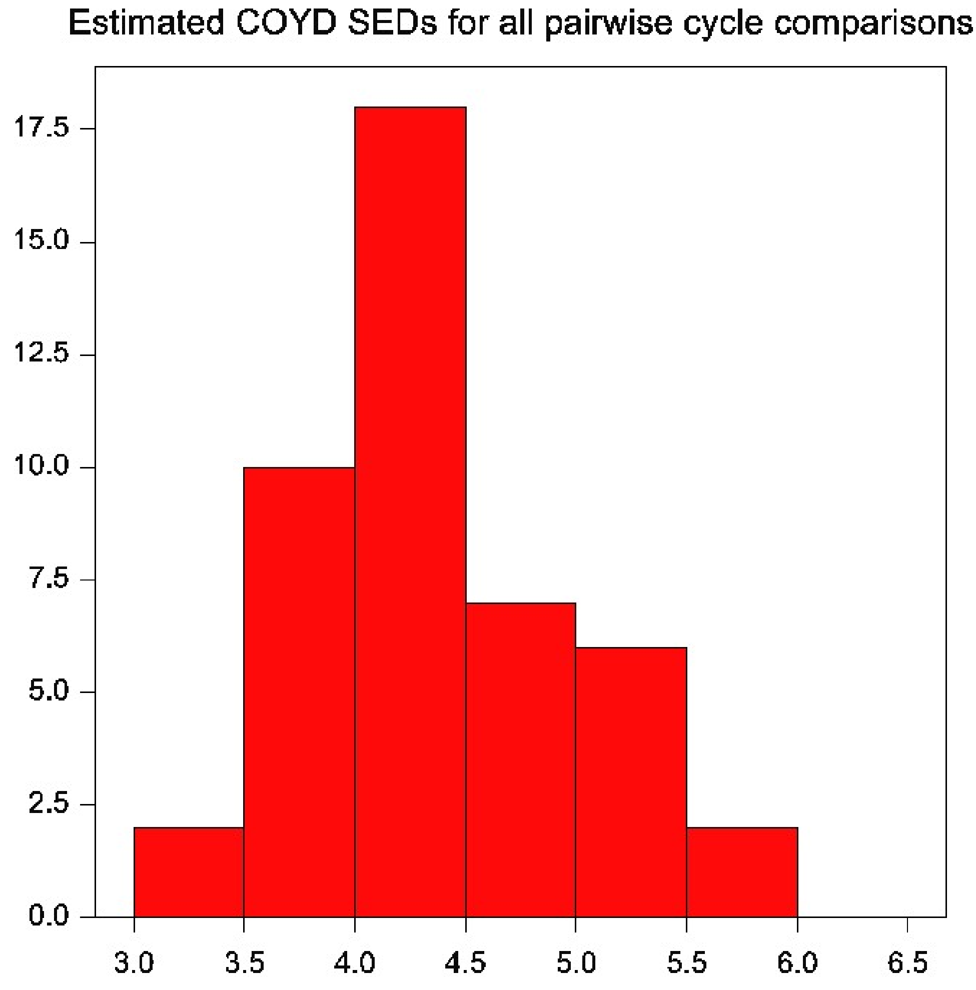
24. The residual variances are shown below. The residual variance associated with seven out of the ten cycles was less than the residual variance from the *id(cycle) ⊗ id(variety)* model.

Residual variance model	Estimate	SE
id(cycle) ⊗ id(variety)	19.36	0.83
diag(cycle) ⊗ id(variety) - cycle 1	15.61	3.10
diag(cycle) ⊗ id(variety) - cycle 2	18.50	3.20
diag(cycle) ⊗ id(variety) - cycle 3	11.76	1.96
diag(cycle) ⊗ id(variety) - cycle 4	20.56	2.89
diag(cycle) ⊗ id(variety) - cycle 5	18.99	2.59
diag(cycle) ⊗ id(variety) - cycle 6	17.24	2.28
diag(cycle) ⊗ id(variety) - cycle 7	8.44	1.42
diag(cycle) ⊗ id(variety) - cycle 8	18.48	2.55
diag(cycle) ⊗ id(variety) - cycle 9	40.13	5.40
diag(cycle) ⊗ id(variety) - cycle 10	26.43	3.78

Based on the above parameter estimates SEDs for COYD variety comparisons in each pairwise (=45) combination of the ten cycles were computed and are shown in Figure 1. The two-cycle COYD SED based on the *id(cycle) ⊗ id(variety)* residual variance model was 4.40. 60% of the pairwise SEDs were less than this, which is to be expected given the ten cycle estimates shown above were skew distributed.

25. For other characteristics in our data set sizeable cycle effects on the variance were found. In contrast, variety effects on the variance were much smaller.

Figure 1:



Assessment of the robustness of the proposed approach to skewness and kurtosis

26. Here we examine how robust the proposed one-cycle approach is to skewness and kurtosis. We have not yet looked at non-uniform variance – this may be the subject of future work.

27. We examine the effect of kurtosis by simulating data with errors from Student's t-distribution. The student t-distribution is symmetric with the degree of kurtosis depending on the degrees of freedom – see the table below.

Degrees of freedom	Kurtosis
5	6.00
6	3.00
7	2.00
8	1.50
9	1.20
10	1.00
11	0.86
12	0.75
13	0.67
14	0.60
15	0.55
16	0.50
17	0.46
18	0.43
19	0.40
20	0.38

28. We have run simulations using errors with student t-distributions with degrees of freedom from 1 to 20, as well as the normal distribution. We investigate what decisions are made after one-cycle (using the new approach) and two cycles (using COYD at 2%), assuming the true difference between the two varieties is 0, 1/5, 1/4, 1/3, 1/2, 1, 2, 3, 4, 5 times the COYD criterion. We have also looked at p_D values of 50%, 80%, 90%, 95%, 99% and 99.9% - but here we will show results for 90% only. For each case we ran 100,000 simulations.

29. Figure 2 shows how the proportion of pairs that exceed the COYD criterion after two cycles depends on the true difference, shown as a multiple of the COYD criterion (x-axis) and the degrees of freedom (panels). The degrees of freedom has an influence on results: where the true difference is zero, the lower the degrees of freedom is, the further the proportion of distinctness decisions increases above the target 2% level (see also Figure 4). Also fewer distinctness decisions are made when the true difference is above the COYD criterion as the degrees of freedom decrease (see also Figure 5).

30. Figure 3 shows how the proportion of pairs that exceed the one-cycle threshold depends on the true difference, shown as a multiple of the COYD criterion (x-axis) and the degrees of freedom (panels). The proportion of pairs exceeding the one-cycle threshold is greatest for 1 degree of freedom when the true differences are low and smallest for 1 degree of freedom when the true differences are high. Again, where the true difference is zero, the lower the degrees of freedom is, the further the proportion of distinctness decisions increases above the target 2% level (see also Figure 4). Also, fewer distinctness decisions are made when the true difference is above the COYD criterion as the degrees of freedom reduce (see also Figure 5).

31. Figures 4 and 5 show how the proportions of COYD and one-cycle distinctness decisions compare, given true differences of zero (Fig. 4) and three times the mean COYD criterion (Fig. 5). Whilst the proportions are fairly similar for a zero true difference (the one-cycle is slightly higher), for three times the COYD criterion the one-cycle method is less effective than COYD. This is perhaps the best construction; we are willing to identify fewer varieties as distinct on the basis of one cycle. The graphs also show that the effect of kurtosis can be severe.

32. From these results it seems that it would be best to use the method only where the level of kurtosis is quite low. Of greatest concern is the rate of false positives, i.e. declaring varieties as different when in fact the true difference between their characteristic scores is zero. With a p_D tolerance level of 90%, then t-distributed data with 6 degrees of freedom gives a false positive rate of 4.4%. In comparison, normally distributed data has a rate of 1.1%. This could be considered acceptable. All of the characteristics in the example have kurtosis levels equivalent to 6 degrees of freedom or more except for intensity of foliage colour and the number of nodes. Later in this document we consider what should be done about such characteristics.

Figure 2: Kurtosis simulations

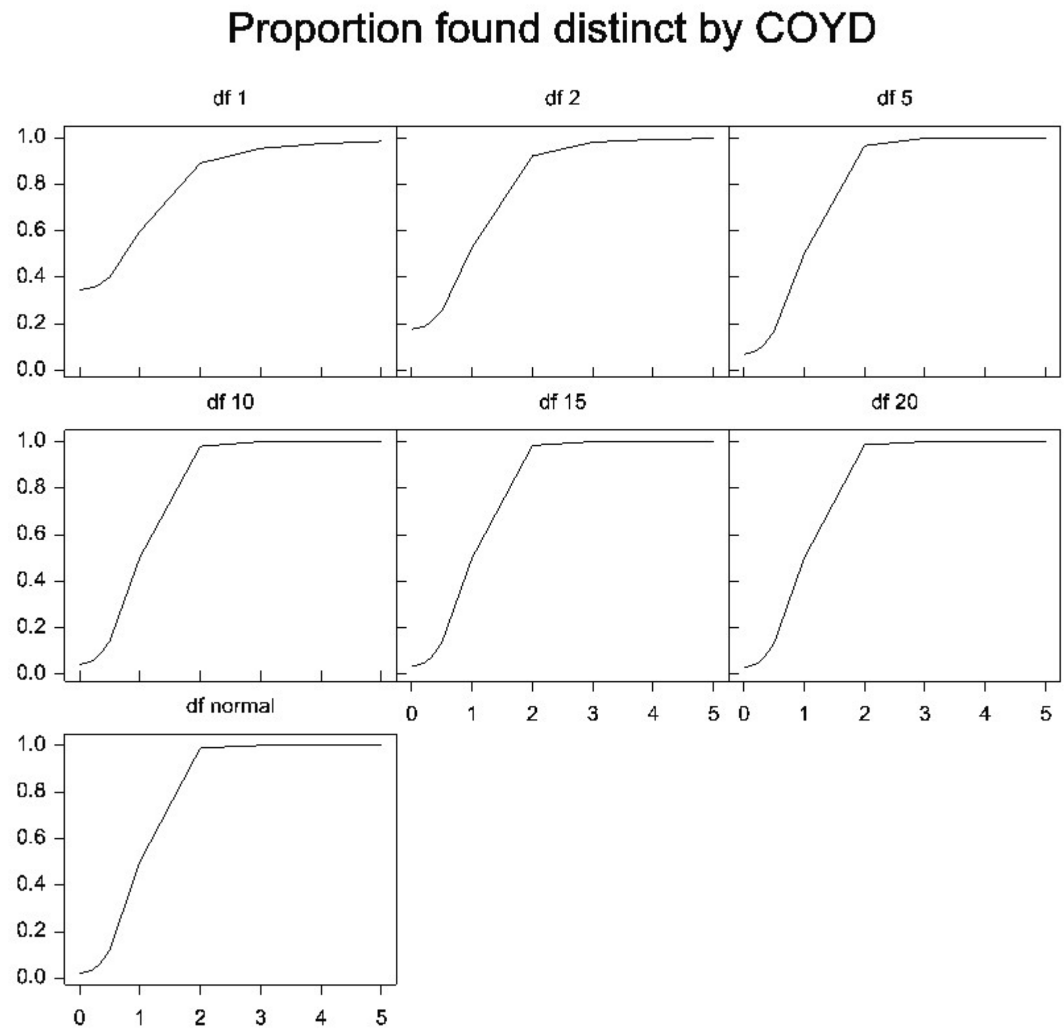


Figure 3: Kurtosis simulations

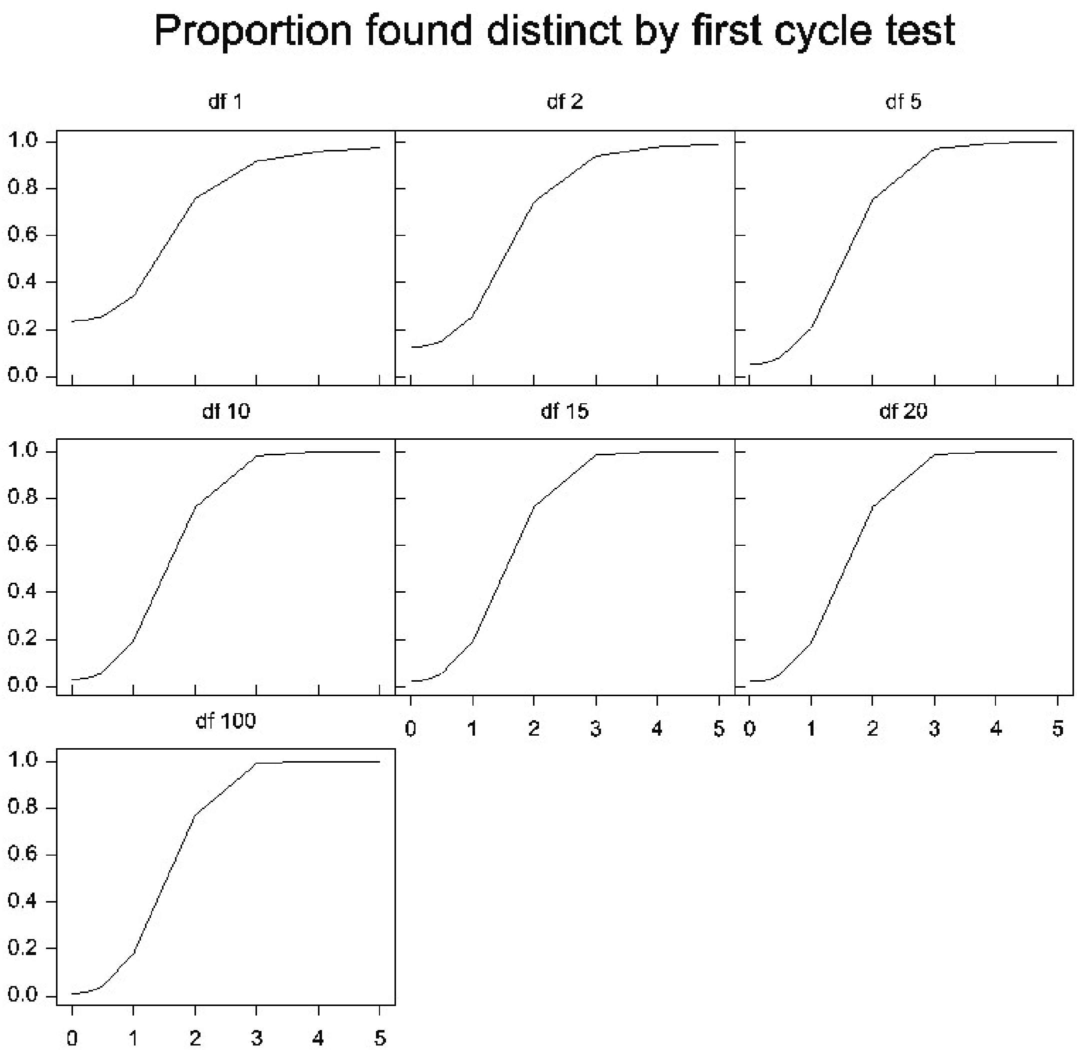


Figure 4: Kurtosis simulations

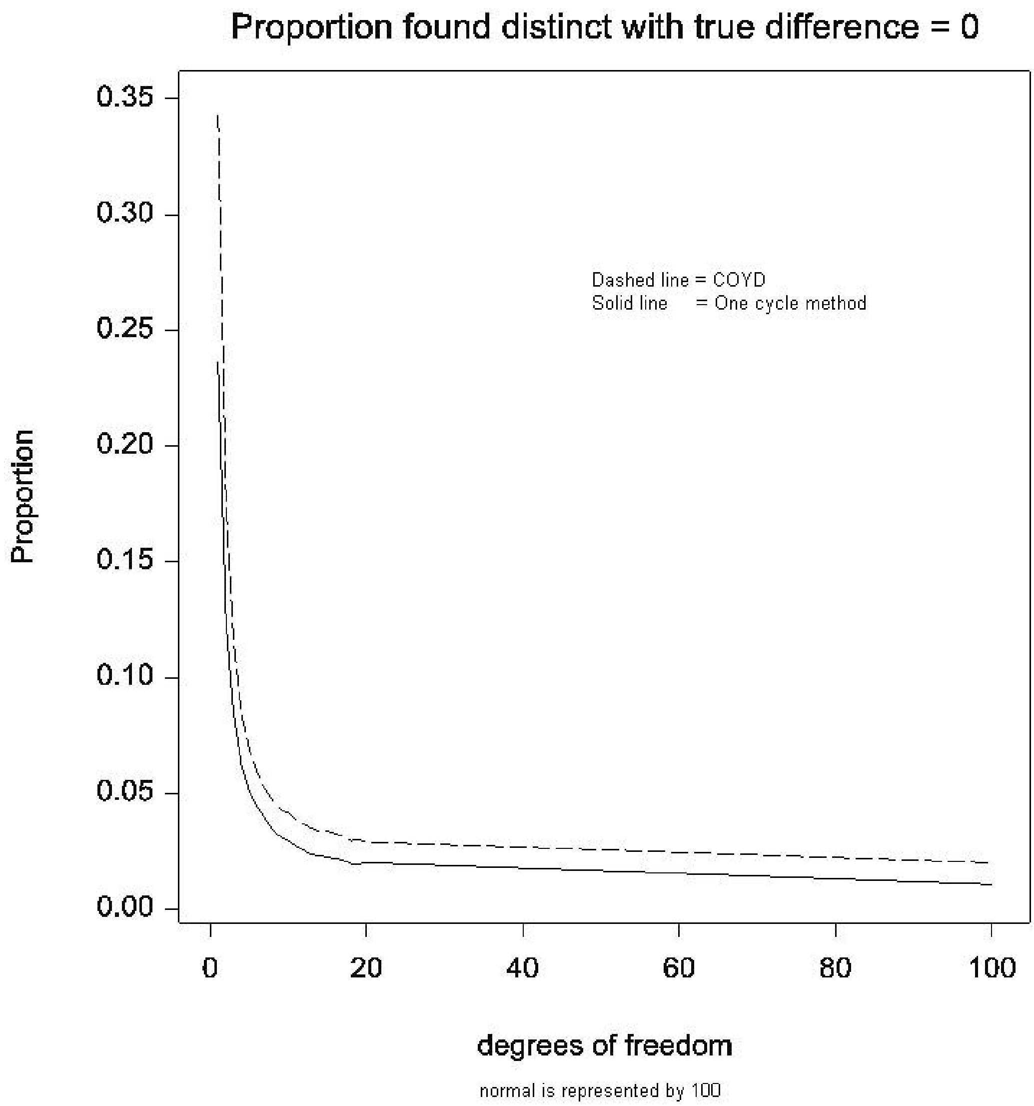
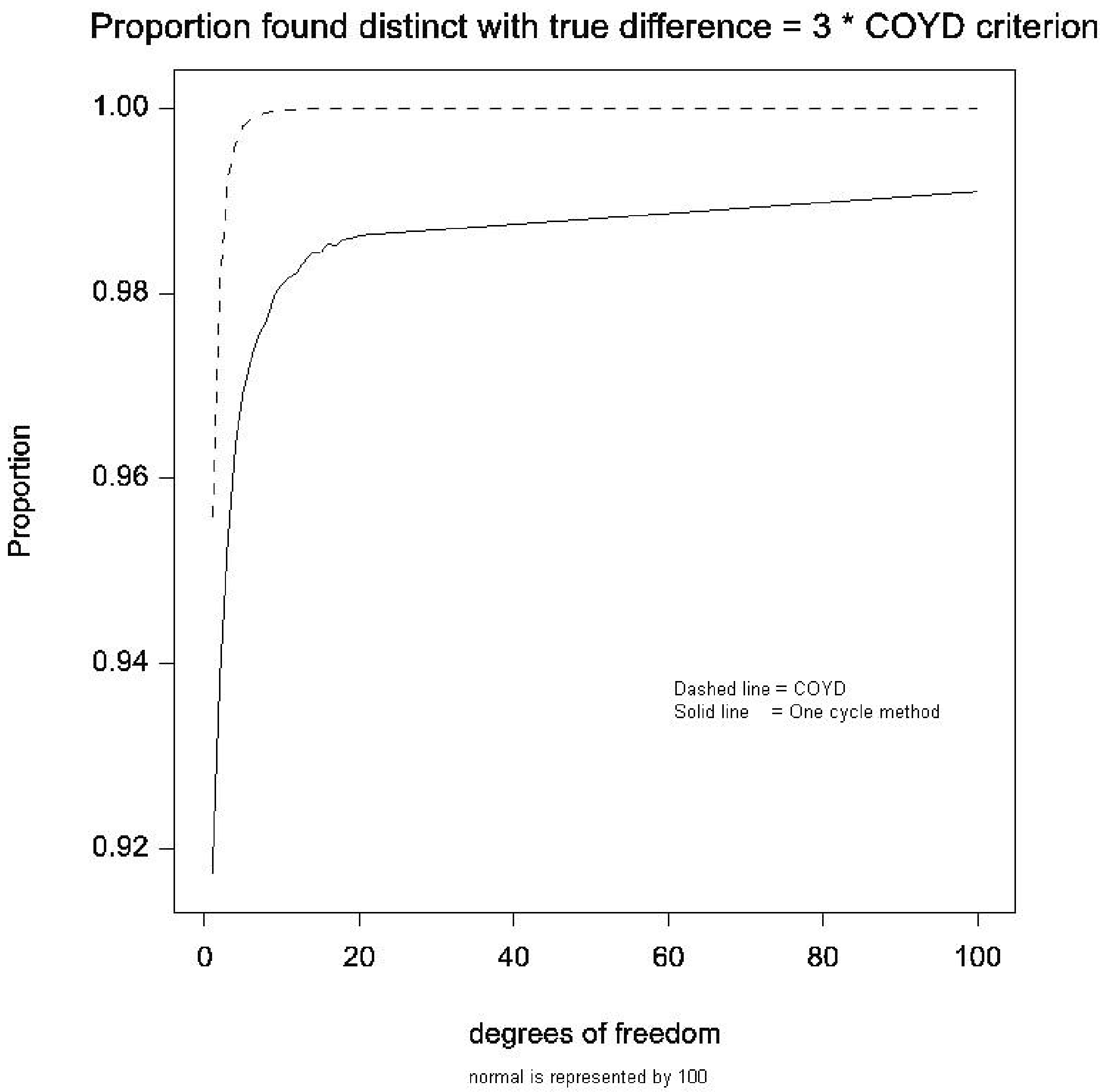


Figure 5: Kurtosis simulations



33. We examine the effect of skewness by simulating data with errors from a skew-normal distribution. This distribution can be denoted by SN (ξ, ω^2, α) where ξ ω and α are the location, scale and shape parameters respectively. For further details see: http://en.wikipedia.org/wiki/Skew_normal_distribution and http://azzalini.stat.unipd.it/SN/Intro/intro.html .
While data can be simulated from the skew normal distribution with varying degrees of skewness in the range [0,1), both skewness and kurtosis are functions of only the skewness parameter α. Consequently, specifying a value for the parameter α to give a pre-selected population skewness will also fix the population kurtosis. The kurtosis is only zero when the skewness is also zero. Therefore it has been impossible to isolate out through simulation the effect of skewness from that of kurtosis. However, given that kurtosis is usually present in skew-distributed data, this should not prove to be a substantial limitation.

34. We have run simulations using errors from skew-normal distributions SN (ξ, ω^2, α) where the location, scale and shape parameters ξ, ω and α respectively took the values shown in the table below. The table shows the theoretical mean, variance, skewness and kurtosis for each

distribution from which errors were simulated. Thus errors were simulated from distributions with the same mean and variance as the standard normal distribution but with skewness varying from 0 (standard normal) to 0.9. We investigate what decisions are made after one-cycle (using the new approach) and two cycles (using COYD at 2%), assuming the true difference between the two varieties is 0, 1/5, 1/4, 1/3, 1/2, 1, 2, 3, 4, 5 times the COYD criterion. We have looked at a p_D value of 90%. For each case we ran 30,000 simulations. (It should be noted that changing the sign of the shape parameter α merely changes the sign of the skewness.)

ξ	ω	α	Mean	Variance	Skewness	Kurtosis
0.0000	1.0000	0.0000	0.00	1.00	0.0	0.000
0.6153	1.1742	0.8711	0.00	1.00	0.1	0.041
0.7753	1.2653	1.1988	0.00	1.00	0.2	0.102
0.8875	1.3370	1.4992	0.00	1.00	0.3	0.176
0.9768	1.3979	1.8141	0.00	1.00	0.4	0.258
1.0522	1.4516	2.1738	0.00	1.00	0.5	0.347
1.1181	1.5001	2.6187	0.00	1.00	0.6	0.443
1.1771	1.5445	3.2260	0.00	1.00	0.7	0.544
1.2307	1.5857	4.1902	0.00	1.00	0.8	0.650
1.2799	1.6243	6.2979	0.00	1.00	0.9	0.760

35. Figure 6 shows how the proportion of pairs that exceed the COYD criterion after two cycles depends on the true difference, shown as a multiple of the COYD criterion (x-axis) and the skewness (panels). It is evident that changes in the coefficient of skewness within the range [0,0.9] have very little impact on the proportion of pairs exceeding the COYD criterion after two cycles and this pattern holds for true differences over a range of multiples of the COYD criterion.

36. Figure 7 shows how the proportion of pairs that exceed the one-cycle threshold depends on the true difference, shown as a multiple of the COYD criterion (x-axis) and the skewness (panels). It is evident that changes in the coefficient of skewness within the range [0, 0.9] have little impact on the proportion of pairs exceeding the one-cycle threshold and this pattern holds over a range of multiples of the COYD criterion.

37. Figures 8 & 9 show how the proportions of COYD and one-cycle distinctness decisions compare, given true differences of zero and three times the COYD criterion respectively. As to be expected when the true difference is zero (Fig. 8), the proportion found distinct is lower under the one-cycle method than COYD. However, the proportion found distinct under the one-cycle method when the true difference is zero does appear to increase with increasing skewness. The extent of variation in the underlying trend in the one-cycle method proportion as skewness increases appears at first sight to be more serious than it actually is. In fact, in part it is an artefact of the choice of scale on the Y-axis. Nevertheless, the variation in this trend could be reduced by increasing the number of simulations but there would be limited benefit in doing so as the trend is already clear and increasing the number of simulations for the skew-normal distribution would be very computer-intensive. In Figure 9, in which the true difference equals three times the COYD criterion, the proportion found distinct is lower under the one-cycle method than COYD. The proportion found distinct increases slightly under the one-cycle method as skewness increases but in all cases already exceeds 0.99. The graphs show that the effect of skewness is minimal up to the level that we have simulated.

38. From these results it seems that changes in skewness within the range [-0.9, 0.9] have a minimal impact on decision making. With the exception of characteristic 71 all characteristics had observed skewness within this range in our example data set. Larger levels of skewness may be encountered in practice and might require some action.

Figure 6: Skewness simulations

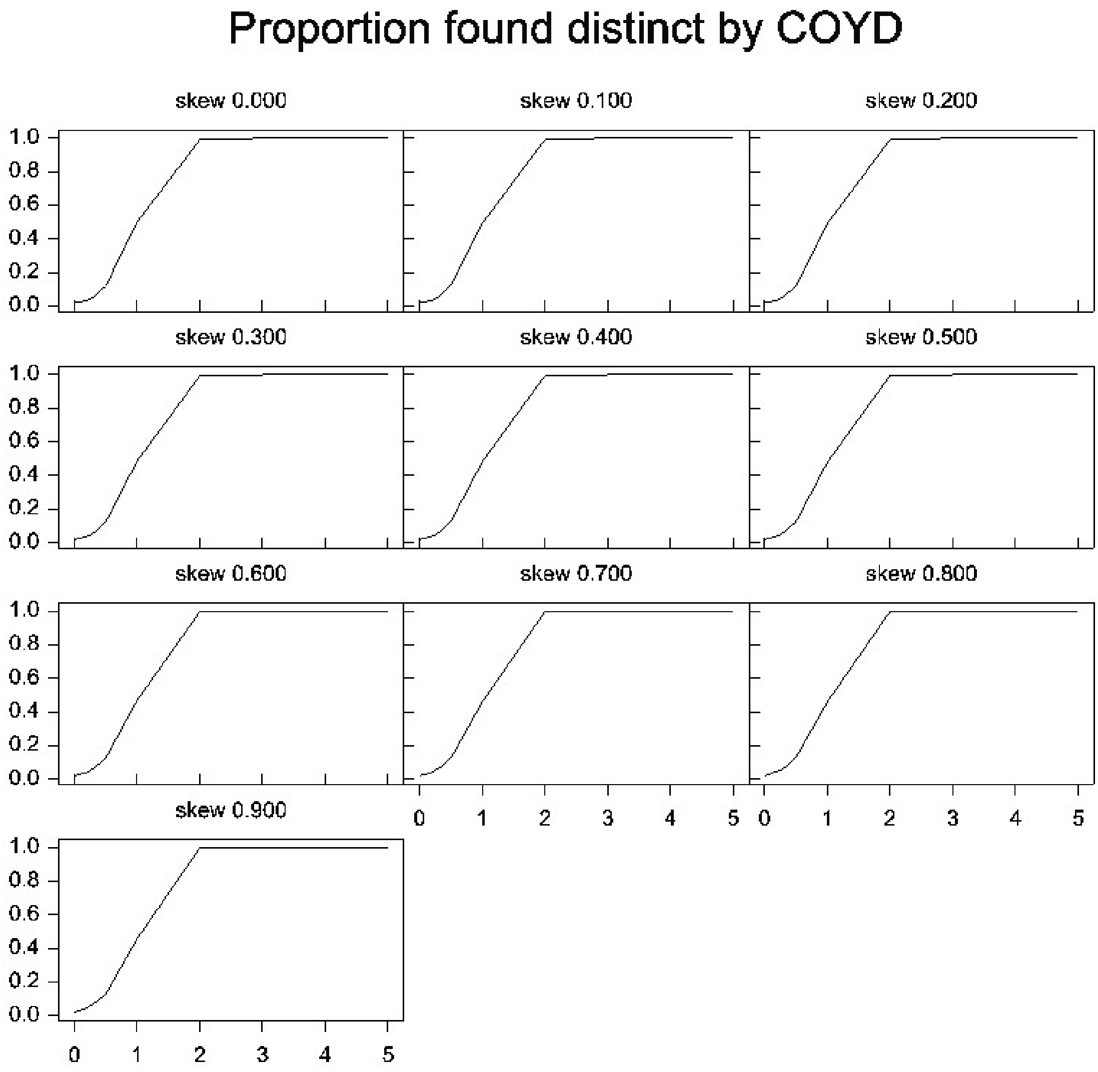


Figure 7: Skewness simulations

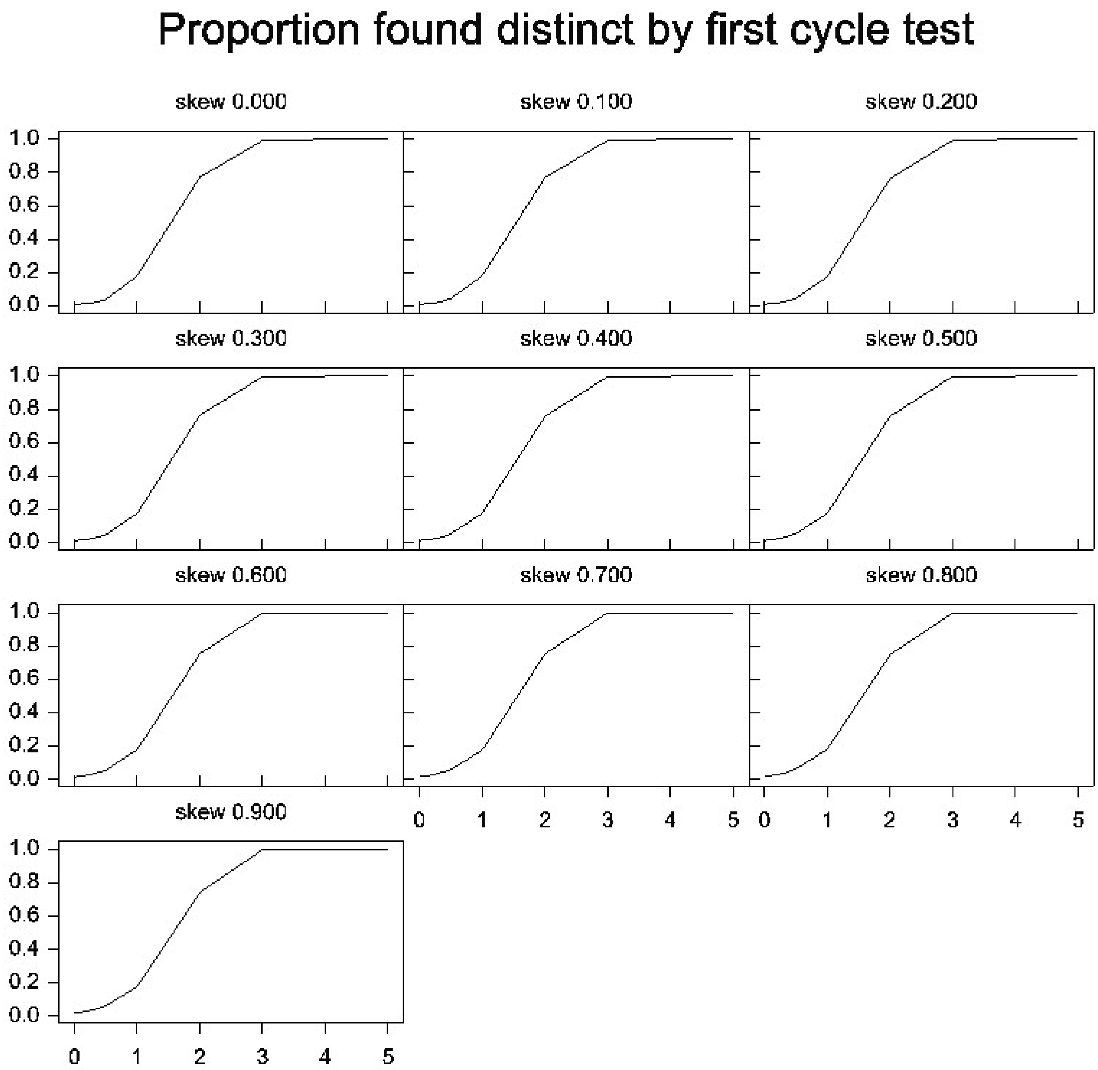


Figure 8: Skewness simulations

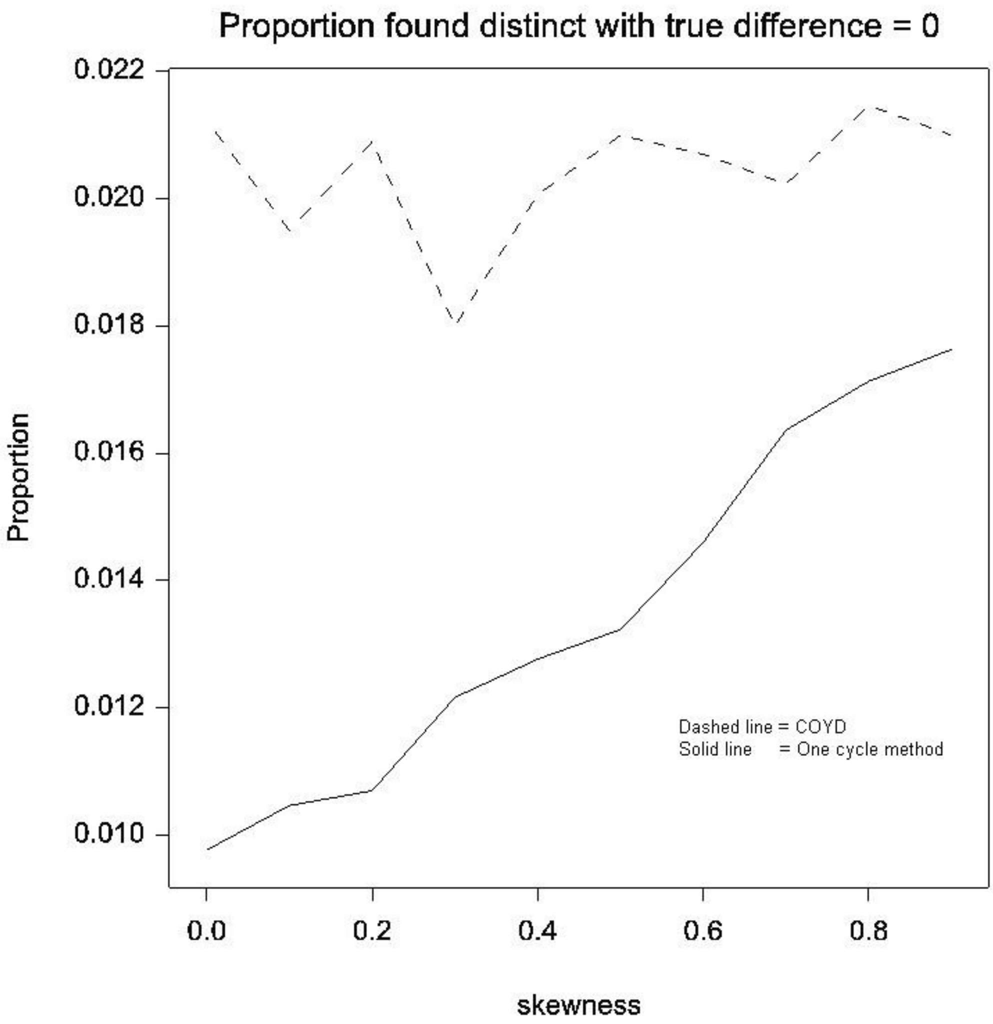
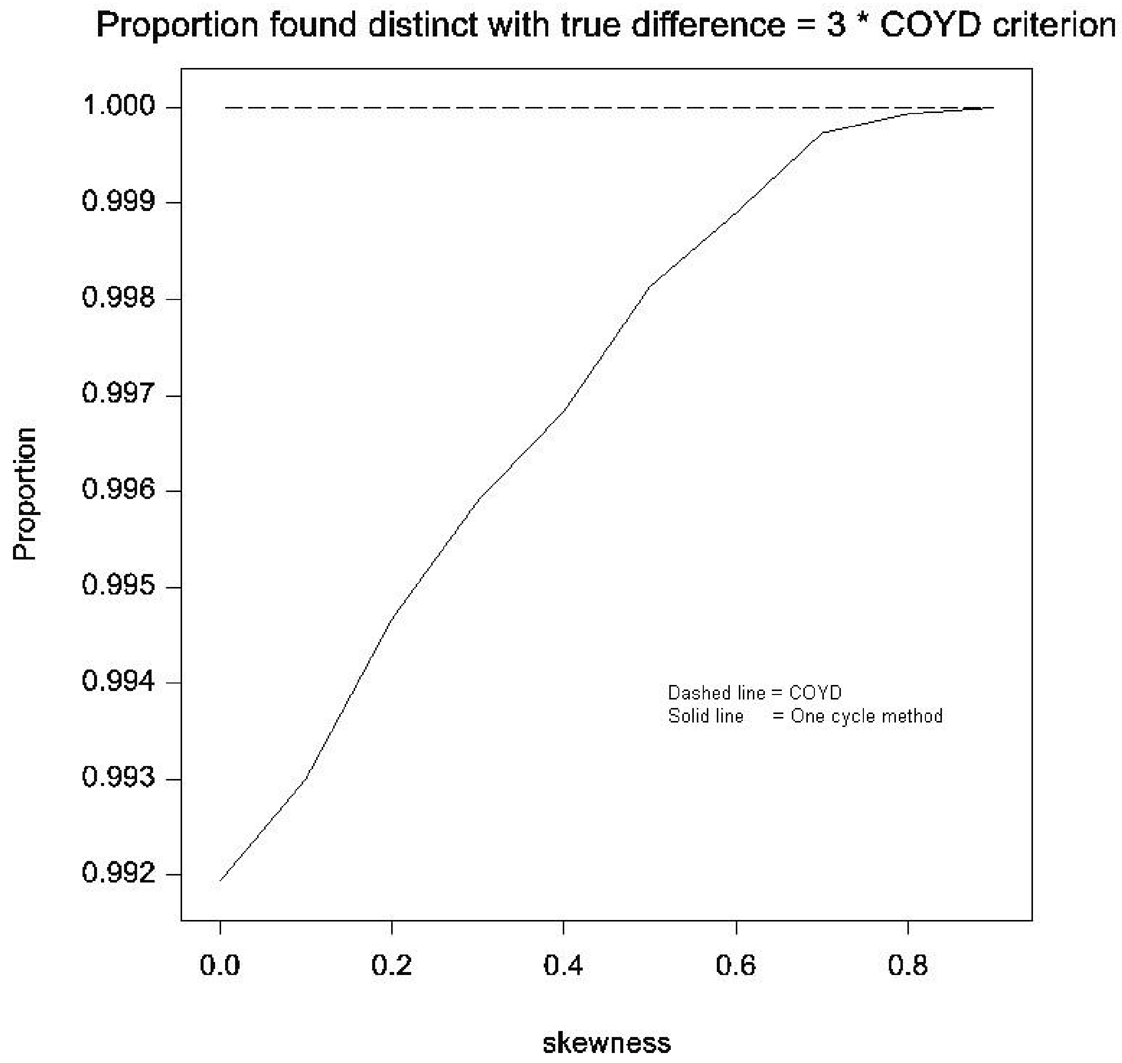


Figure 9: Skewness simulations



Dealing with excessive deviation from the assumptions

39. Here we consider possible ways of dealing with characteristics that have important degrees of skewness or kurtosis, or where the assumption about uniform variance is not true. Of course, where no solution proves successful then it may be best to ignore the characteristic for the purpose of first-cycle comparisons.

40. The second-largest absolute level of skewness and the largest positive skewness observed in our data set was for characteristic 46. However, it should be remembered that this characteristic is currently visually assessed and distinctness decisions are based on a two-state difference rather than by COYD. In statistics, natural logarithm transformations are commonly applied to positively skewed data and therefore the skewness and kurtosis were assessed after log transformation of this characteristic. The skewness was reduced by approximately a third but kurtosis by only approximately a fifth.

Characteristic	Scale	Skewness	S.E.	Kurtosis	S.E.
46	Original	0.64	0.07	6.26	0.14
46	ln(x)	0.41	0.07	5.19	0.14
71	Original	-1.29	0.06	10.69	0.11
71	ln(max[x]+1-x)	-0.02	0.06	10.01	0.11

41. In contrast, characteristic 71 exhibited negative skewness. Therefore, in this case, the slightly more complex transformation of ln (max(x) + 1 – x) was evaluated where x denotes responses for characteristic 71. This handles the fact that skewness is negative rather than positive and avoids attempting to log-transform a non-positive number. The transformation was very successful in reducing skewness to a negligible level for this characteristic but had minimal impact on kurtosis.

42. As in our example, kurtosis may accompany skewness and it is possible that there may be a transformation not yet identified which may help to reduce both skewness and kurtosis.

43. One potential approach to dealing with kurtosis in the absence of a suitable transformation is to extend the proposed methodology to deal with data that has t-distributed errors. We may consider the formulation for this in the future.

44. The variance was found not to be constant between cycles in many characteristics in our example. At this stage, we have not completed our studies as to how problematic this might be. If this heterogeneity is an issue and transformation does not provide a cure, it may be possible to extend the methodology to allow for this extra variation. This is something that we intend to investigate in the future.

Conclusions & future work

45. A method has been introduced that allows the prediction of COYD decisions based on only one cycle of results where both historic and current data sets are large. The approach presented here is preferable to applying a 1-cycle long-term LSD.

46. A number of assumptions are made; we have investigated the consequences of departures from these. Only serious degrees of skewness or kurtosis are likely to be an issue, and it should be possible to deal with characteristics with such problems by a suitable transformation or simply by not using those characteristics for first cycle decisions. In our example, variance heterogeneity was a more common issue. More work is required to investigate the importance of variance heterogeneity and possible modification of the method to allow for it.

[End of document]