E

TWC/17/11 Rev.
ORIGINAL: English
DATE: June 22, 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

REDUCTION OF HERBAGE DUS TRIAL SIZES BY CYCLIC PLANTING OF THE REFERENCE COLLECTION AND ANALYSIS BY COMPENSATED DATA

Document prepared by experts from the United Kingdom

Reduction of Herbage DUS trial sizes by cyclic planting of the reference collection and analysis by compensated data.

Summary

A system has been devised by which further candidate varieties can be accommodated in DUS herbage trials without increasing the trial size while maintaining a distinctness test of similar stringency to that of the present system.

The system comprises allocating the control varieties to three cycles, one of which is omitted cyclically from trial each year. Candidate varieties are included in trial for the three years of their test period plus a fourth year. If they are granted NL/PBR, they join the reference collection of control varieties, are allocated to a cycle and are omitted from trial every third year accordingly.

Distinctness is assessed by applying an adaptation of COYD to the incomplete table of variety (candidate and control) character means in the three year test period. Where data is missing for a variety, it is compensated for by use of two years' data from before the test period. Modified Joint Regression Analysis (MJRA) is used for all characters.

Uniformity is assessed by applying COYU to the incomplete table of variety (candidate and control) character standard deviations in the three year test period.

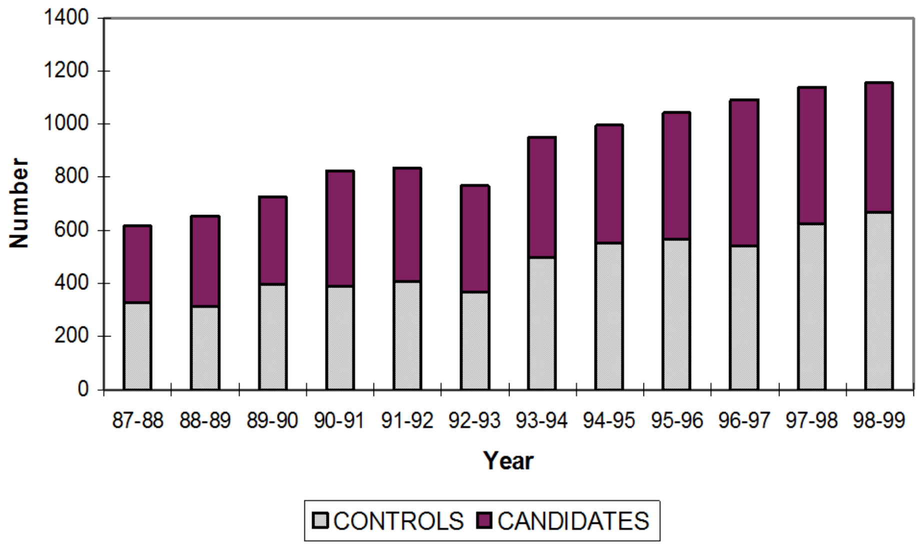
The system is compared to the present system and is found to be in close agreement, being only slightly less stringent in distinctness testing and slightly more stringent in uniformity testing. The overall effect of adoption the system on the DUS variety pass rate is expected to be minimal.

Details of fitting the MJRA model and a worked example of the analysis are given in the Appendices, as are a case study of disagreement over distinctness between the new and the existing system and details of the computer programs used

1. Introduction and statement of the problem

For breeders to receive plant breeders rights on a newly developed variety this candidate variety must be found to be distinct on one or more characters from all other (known) varieties. Plants of the new variety must also be as uniform as possible and the variety must be stable. Guidelines for assessing distinctness, uniformity and stability (DUS) are given by the International Union for the Protection of New Varieties of Plants (UPOV). Currently these guidelines recommend that DUS decisions on cross-fertilised species are based on the Combined-Over-Years-Distinctness (COYD) and Combined-Over-Years-Uniformity (COYU) criteria applied to data from candidate varieties grown together with established varieties (the reference collection) as controls for a three year test period. As this is a continuing process with each accepted candidate variety becoming a control for later candidates, the number of varieties which must be accommodated in the trials is ever increasing. In UK grass and clover DUS trials the number of varieties has nearly doubled since 1988 [Fig. 1]. The stage has now been reached where further varieties cannot be accepted into the present trialling system without exceeding reasonable limits of size both with regard to space and manpower. As a result, an alteration to the UK Herbage DUS system has been adopted which will enable further candidate varieties to be accommodated in the trials without increasing their size while providing a distinctness test of similar stringency as at present.

Figure 1. Numbers of Candidate and Control varieties in UK DUS Herbage Trials 1988-1999



2. Possible Solutions

Several methods of keeping trial sizes within reasonable limits were considered.
The possibility of reducing the number of plants per variety was examined. This could be achieved by reducing the number of replicates in a trial from 6 to 4, or by reducing the number of plants per plot from 10 to 7. However both would cause a weakening of the uniformity test which relies upon a good assessment of the distribution of the plants making up a variety and would also cause some weakening in the distinctness test.

Another solution which has been adopted elsewhere for vegetable varieties, is to eliminate large numbers of control varieties by selecting a few control varieties, known as marker varieties, to each represent several other control varieties. In this process the marker varieties can be selected by means of a cluster analysis of all control varieties, with the marker varieties being chosen to represent groups of other varieties. However clustering can only be considered to be an approximate process of bringing together similar varieties and could lead to some candidates not being compared with their most similar control varieties.

3. The Preferred Solution - Cyclic Controls

The preferred solution requires that each control variety belongs to a ‘cycle’ and is omitted from tests in one year out of three on a cyclical basis (Fig. 2). Thus varieties belonging to Cycle 1 in Fig. 2 will not be planted in 2000, 2003 or 2006, whereas those in Cycle 3 will not be planted in 2002, 2005 or 2008. This therefore removes one third of the control varieties, amounting to around 200 varieties from the current level of 1150 candidate and control varieties in UK grass and clover trials.

Figure 2. **Data patterns for the test period 2004 to 2006**

					TEST PERIOD				
TRIAL YEARS	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
Candidate Varieties					X	X	X	*	
Control Varieties – Use of data									
Cycle 1		X	X		X	X		*	*
Cycle 2	O		X	X		X	X		*
Cycle 3	O	X		X	X		X	*	
New Control Varieties – Assimilation into matrix									
Final DUS tested in 2002 (Cycle 2)	O	O	X^F	X		X	X		*
Final DUS tested in 2003 (Cycle 3)		O	X	X^F	X		X	*	
Final DUS tested in 2004 (Cycle 1)			X	X	X^F	X		*	*
Final DUS tested in 2005 (Cycle 2)				O	X	X^F	X		*

X Indicates data retrieved using maximum of 4 years for distinctness testing and within the (boxed) test period for uniformity testing

O Indicates data present but not retrieved

F Indicates final DUS test year of new control varieties

* Indicates future inclusion in trial

☐ (within box) Indicates the data used for uniformity testing

As before, each candidate variety is planted in trial and has its data recorded in each year of a three year test period (2004 to 2006 in Fig. 2), after which it is DUS tested. Because of the possible lag between final DUS testing and being granted NL/PBR, candidate varieties are kept in trial for a fourth year after the three year test period. If granted NL/PBR, they will then become part of the reference collection and will enter the compensated data matrix. This means that all newly accepted varieties are initially present for four consecutive years in trials and that all varieties entering the same year follow the same cycle of omissions in future years. Thus candidate varieties that were final DUS tested in 2002 in Fig. 2 are in trial for a fourth year in 2003 and so join the Cycle 2 controls. Candidate varieties final DUS tested in 2003, 2004 and 2005, would join Cycles 3, 1 and 2 respectively.

To minimise the risks of bias in the initial allocation of control varieties to cycles, they are ranked by date of ear emergence and then alternately allocated to a cycle. Other than this initial allocation, the choice of controls following each cycle is determined by the candidate varieties entered for trial in earlier years and by which control varieties the breeders choose to withdraw. Although an exactly equal number of controls belonging to each cycle is not essential, it is likely to be necessary to balance the numbers belonging to each cycle in the future. This should be done by transferring control varieties between the cycles by planting them in years when they should be omitted.

3.1 Distinctness testing by data compensation

In previous years, distinctness testing of herbage in the UK has been based on applying COYD to a complete variety (candidate and control) by test period years matrix of character means. With cyclic planting, this matrix is incomplete for the control varieties. For distinctness testing, where data on a control variety is missing, data held in computer files from earlier years is used to compensate for the loss of data. Due to lack of overlap years with the candidates, the value of back data is not as good as data from the test period. To maintain the present stringency of testing, it has been found that two years of past data must be included when one year of current data is removed from a control variety. Thus for the 2004 to 2006 test period illustrated in Fig. 2, controls in Cycle 1 would have data from 2001 and 2002 retrieved, those in Cycle 2 data from 2002 and 2003 and those in Cycle 3 data from 2001 and 2003. Even where more years of past data are available (marked by an O in Fig. 2), to avoid reducing the stringency of the distinctness test, only the two most recent years are used to compensate for the missing current year. Hence, while data from 2000 and before are available for varieties in Cycles 2 and 3, such data is not retrieved for the 2004 to 2006 test period.

Sometimes data on a control variety will be available for a year when its cycle suggests it would not be present in the trial. Such cases are where candidate varieties join the reference collection and become controls, or where a control variety is needed for a special test with a problem variety. In this case the control variety would have full data available during the test period and so no historical data would be retrieved for the distinctness testing. Thus successful candidate varieties final DUS tested in 2005 would have full data available during the test period from years 2004 and 2006 and so no historical data would be retrieved. However successful candidate varieties final DUS tested in 2002, 2003 and 2004, would have only two years of data available during the test period from years 2004 and 2006 and so would have two years of historical data retrieved from 2002 and 2003 to compensate for the missing year.

3.2 Uniformity testing

As with distinctness testing, uniformity testing of herbage in the UK in the past has been based on applying COYU to a complete variety (candidate and control) by test period years matrix of within variety standard deviations. With cyclic planting, as may be seen from the boxed year by variety combinations in Fig. 2, this matrix is incomplete for the control varieties. COYU is applied to this matrix and no attempt is made to compensate for the incomplete data. This is because COYU consists of pooling over years the within variety standard deviations for all available control varieties while taking into account variety means to provide a uniformity standard against which to compare the standard deviations of the candidate varieties. Consequently, it is not possible to make a correction for standard deviations beyond the range of years for the candidates. As a result, only uniformity data from the control varieties within the test period are used to set the uniformity standard for the candidates.

4. Method of Analysis for Distinctness Testing

Over year data arising from DUS trials are strongly influenced by year effects. Thus in a late year the range of the dominant character, heading date, is very much reduced compared with its range in an early year. Other characters are affected similarly. This suggests that the variety effects should be modified by a constant having the value of unity for a medium year and taking values above and below unity for other years depending on their earliness. This constant therefore reflects the slope of the variety means on a character in a single year against standard variety values determined over several years. It is therefore assumed for distinctness purposes that for n_v varieties in n_y years the data arising from the proposed trial scheme will follow the modified joint regression analysis (MJRA) model:

$$c_{ij} = \mu + y_j + \beta_j v_i + \varepsilon_{ij} \qquad (1)$$

where c_{ij} is the value on a character for variety i in year j, $i = 1,\ldots,n_v$ and $j = 1,\ldots, n_y$

v_i is the effect of the ith variety with $\Sigma\, v_i = 0$

y_j is the effect of the jth year with $\Sigma\, y_j = 0$

β_j is the slope of variety means in year j against variety means over all years and is referred to as the sensitivity of year j.

ε_{ij} is a random error associated with variety i in year j

This model was originally proposed by Digby,P (1979) to allow for varying slopes of the means of one variety versus means over all varieties and has been adapted to allow for varying slopes of variety means in one year versus means over all years.

Equation (1) is non linear and so cannot be fitted directly to the data. Instead it is fitted iteratively. This provides estimates of the variety effects, comparisons of which determine distinctness. Details of the model fitting and variety comparisons are given in Appendix A. An example of fitting the model to data is given in Appendix B.

5. Comparison of the Cyclic Controls and the Complete Data Approaches

Before adopting the approach of using cyclic controls with compensated data for distinctness decisions and uncompensated data for uniformity decisions, it was necessary to compare it with the complete data approach of using COYD and COYU on complete variety by test period years matrices. The distinctness and uniformity decisions that would have been made using the cyclic controls approach on the 105 Perennial Ryegrass Diploid and Tetraploid candidate varieties that were final DUS tested in 1997 and in 1998 were compared with those that were made with the complete data approach. Although all control varieties were planted, the cyclic controls approach was simulated by allocating control varieties to the cycles and replacing their data with missing data symbols in the computer files where appropriate. For distinctness testing, data from previous years were retrieved from computer files and used to compensate for this ‘missing’ data.

5.1 Comparison of uniformity decisions

The following table compares the uniformity decisions by the two approaches.

		Cyclic controls		
		Uniform	Not Uniform	Total
Complete data	Uniform	89	2	91
	Not Uniform	0	14	14
	Total	89	16	105

It shows nearly identical results with 103 out of 105 decisions being the same.

5.2 Comparison of distinctness decisions

The following table compares the distinctness decisions by the two approaches.

		Cyclic controls		
		Distinct	Not Distinct	Total
Complete data	Distinct	77	4	81
	Not Distinct	4	20	24
	Total	81	24	105

It shows that the same number of varieties were found to be distinct by both approaches. However there are some differences in the varieties passed by the criteria. Of the 81 varieties passed by each approach, only 77 were in common, resulting in 8 varieties being classified differently by the two approaches. These cases of disagreement are examined below.

5.3 Stringency of distinctness testing

Distinctness testing in the cyclic controls with compensated data approach and the complete data approach both involve comparing variety effects with LSD's. Consequently the stringency of distinctness testing using two approaches can be assessed by the ratio of their LSD’s. This is

$$\frac{LSD\,compensated\ data}{LSD\,complete\,data} = \frac{t_1\sqrt{RMS_1(v_{ii} + v_{jj} - 2v_{ij})}}{t_2\sqrt{RMS_2(1/3 + 1/3)}} = \sqrt{1.5(v_{ii} + v_{jj} - 2v_{ij})}$$

assuming that $t_1 = t_2$ and $RMS_1 = RMS_2$.

Stringency factors for a range of characters derived from several trials in two years were generally around 0.95 and ranged between 0.89 and 1.04. They were found to differ among the characters and depend on the test year from which the control variety was omitted. This indicates that the LSD's tend to be smaller using the cyclic controls with compensated data approach to determine distinctness, i.e. it is slightly less stringent than using the complete data approach. This suggests some over-compensation in using two years' past data instead of one year's. However, with only complete years to manipulate it is not possible to obtain LSD ratios closer to unity.

5.4 Study of the disparity between the approaches

Further study has been made of the 8 candidates that were differently classified for distinctness by the two approaches. Marginal differences in the means by the two approaches explain the discrepancy in the distinctness decisions for 3 out of the 8 candidates. For the other 5 candidates the differences in the means by the two approaches are large. These differences are caused by the control variety performing differently in the year of the test period in which it is 'omitted' compared to how it performed in the years that are used to compensate. Data on one of the 5 candidates is presented in Appendix C as a case study to illustrate the reasons for the disparity between the approaches.

6. The Way Forward

A small reduction in stringency of distinctness testing and a slight increase in the stringency of uniformity testing are expected from the cyclic controls approach relative to the complete data approach. This is due to slight overcompensation in using two years past data to compensate for one year's missing from the test period in the distinctness testing, and the reduction in the information on the controls used to compile the uniformity standard in the uniformity testing. The altered stringencies would be expected to cause a slight increase in the number of candidate varieties found to be distinct and a slight reduction in the number of varieties found to be uniform in the cyclic controls approach. All in all, the likely changes in the distinctness and uniformity stringencies suggest that there will be little change in the overall DUS pass rate.

The results of the comparison of the distinctness and uniformity decisions arising from the cyclic controls approach with those from the complete data approach has been presented to the testing authority and the breeders' authority in the UK. They have accepted the need to change to the new approach and have agreed to it. The diploid perennial ryegrass trials that have been sown in 1999 are sown according to the cyclic controls approach and, in following years, the other trials are expected to follow. Adoption of the cyclic controls approach should eventually translate into availability on the ground of up to 200 'slots' for new reference varieties within an absolute limit of 1,100 reference and candidate varieties. These 'slots' should allow us to assimilate more Common Catalogue additions into the reference collection in future.

Dr M. Camlin & Mr B. Waters
D.A.N.I.
Plant Testing Station
50 Houston Road
Crossnacreevy
Belfast
BT6 9SH
N. Ireland

Dr S. Watson
D.A.N.I.
Newforge Lane
Belfast
BT9 5 PX
N. Ireland

Dr S.T.C. Weatherup
The Open University
c/o D.A.N.I.
Newforge Lane
Belfast
BT9 5 PX
N. Ireland

May 1999

References

Digby,P (1979) Modified joint regression for incomplete variety x environment data. Journal of Agricultural Science 93, Cambridge, 81-86.

Payne, R.W., Arnold, G. M., and Morgan, G. W. (editors) (1993). Genstat 5 Procedure Library Manual Release 3[1]. Numerical Algorithms Group, Oxford, 126-127.

Appendix

A ANALYSIS DETAILS

A.1. Fitting the MJRA model

The MJRA model has been given as:

$$c_{ij} = \mu + y_j + \beta_j v_i + \varepsilon_{ij} \qquad (1)$$

Equation (1) is non linear and hence cannot be fitted directly to the data. Therefore to make progress it is necessary to solve it iteratively as follows:

Step 1: Since year slopes β_j are likely to be around 1 assume $\beta_j = 1$, ($j = 1,\dots, n_y$) as first estimates of β_j so that (1) becomes the linear equation

$$c_{ij} = \mu + y_j + v_i + \varepsilon_{ij}$$

and solve to get the estimates $\hat{y}_j\,(j = 1,\dots,n_y)$ and $\hat{v}_i\,(i = 1,\dots,n_v)$ using regression.

Step 2: Substitute the $\hat{v}_i\,(i = 1,\dots,n_v)$ values derived at step 1 into (1) to obtain the linear equation

$$c_{ij} = \mu + y_j + \beta_j \hat{v}_i + \varepsilon_{ij}$$

and solve to get the new estimates $\hat{y}_j\,(j = 1,\dots,n_y)$ and $\hat{\beta}_j\,(j = 1,\dots,n_y)$ using regression.

Step 3: Substitute the $\hat{\beta}_j\,(j = 1,\dots,n_y)$ values derived at step 2 into equation (1) to obtain the linear equation

$$c_{ij} = \mu + y_j + \hat{\beta}_j v_i + \varepsilon_{ij}$$

and solve to get the new estimates $\hat{y}_j\,(j = 1,\dots,n_y)$ and $\hat{v}_i\,(i = 1,\dots,n_v)$ using regression.

Repeat from step 2 until the ratio of residual sum of squares between 2 cycles is greater than a defined constant e.g. 0.999.

That each cycle should end on an odd step means that estimates are available of both the variety effects and of their variances and covariances. The RMS degrees of freedom at each step are in accordance with the parameters being estimated. Thus, if n is the total number of observations, on odd steps the d.f. are n - 1 - (n_y - 1) - (n_v - 1). On even steps the d.f. are n - 1 - 2(n_y - 1). The d.f. for estimating the sensitivities β_j are n_y - 1 and not n_y because, although they are not constrained to sum to zero, they have a weighted mean of 1.0. Once the cycles have converged, the RMS is recalculated using the d.f. n - 1 - 2(n_y - 1) - (n_v - 1). This gives identical results to the MJRA analysis produced by Genstat.

A.2. Variety comparisons

Having obtained convergence between the above cycles, use is made of the resulting variance-covariance matrix to compare the estimated effects of particular varieties. This variance-covariance matrix of effects is given by:

$$(X'X)^{-1}\sigma^2 \qquad (2)$$

where X is the design matrix at an odd step number, X' is its transpose and σ^2 is the residual mean square. Using terms from the variance-covariance matrix, the variance between the effects of two varieties *i* and *k* is

$$\text{var}(i,k) = (v_{ii} + v_{kk} - 2v_{ik})\sigma^2 \qquad (3)$$

where $v_{ii}\sigma^2$ and $v_{kk}\sigma^2$ are the variety effect variances and $v_{ik}\sigma^2$ is the covariance of the variety pair. The variance between the effects of two varieties *i* and *k* is used to calculate an LSD which is used compare the variety effects to determine distinctness.

B Worked Example of Distinctness Testing

Consider the following matrix of within year variety means c_{ij}. Variety A represents candidate varieties and varieties B, C and D represent the three cycles of control varieties.

<u>Example data</u>

	Year					
Variety	1	2	3	4	5	6
A	-	-	-	6	2	3
B	-	6	4	-	6	7
C	7	10	-	8	11	-
D	11	-	14	10	-	17

Step 1

We first take first estimates of β_j as $\hat{\beta}_j = 1$, j = 1,…,n_y and solve the equation $c_{ij} = \mu + y_j + v_i + \varepsilon_{ij}$ for μ, $(y_1,\ldots,y_6)$, $(v_1,\ldots,v_4)$. In matrix terms this is

$$\mathbf{c = Xb}$$

or

$$\begin{vmatrix} 6 \\ 2 \\ 3 \\ 6 \\ 4 \\ 6 \\ 7 \\ 7 \\ 10 \\ 8 \\ 11 \\ 11 \\ 14 \\ 10 \\ 17 \\ 0 \\ 0 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{vmatrix} \times \begin{vmatrix} \mu \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{vmatrix}$$

The last 2 rows of the design matrix X impose the constraints $\sum y_j = \sum v_i = 0$
Regression analysis gives estimates $\hat{\mu}$, $(\hat{y}_1,\ldots,\hat{y}_6),(\hat{v}_1,\ldots,\hat{v}_4)$ as 7.71, (-2.37, 0.79, -0.16, -0.61, 0.50, 1.86), (-4.63, -2.70, 1.72, 5.6) respectively with residual mean square (RMS) of 5.496 with 6 df.

Step 2

<u>Example data</u>

	Year					
Variety	1	2	3	4	5	6
A	-	-	-	6	2	3
B	-	6	4	-	6	7
C	7	10	-	8	11	-
D	11	-	14	10	-	17

We now solve $c_{ij} = \mu + y_j + \beta_j \hat{v}_i + \varepsilon_{ij}$ for μ, (y_1,…,y_6), (β_1,…, β_6) and substitute the values $\hat{v}_i\,(i = 1,\ldots,n_v)$ obtained at step 1 viz (-4.63, -2.70, 1.72, 5.6). In matrix terms this is

c = Zd

or

c		Z														d
6		1	0	0	0	1	0	0	0	0	0	-4.63	0	0		μ
2		1	0	0	0	0	1	0	0	0	0	0	-4.63	0		y_1
3		1	0	0	0	0	0	1	0	0	0	0	0	-4.63		y_2
6		1	0	1	0	0	0	0	0	-2.7	0	0	0	0		y_3
4		1	0	0	1	0	0	0	0	0	-2.7	0	0	0		y_4
6		1	0	0	0	0	1	0	0	0	0	0	-2.7	0		y_5
7		1	0	0	0	0	0	1	0	0	0	0	0	-2.7		y_6
7	=	1	1	0	0	0	0	0	1.72	0	0	0	0	0	x	β_1
10		1	0	1	0	0	0	0	0	1.72	0	0	0	0		β_2
8		1	0	0	0	1	0	0	0	0	0	1.72	0	0		β_3
11		1	0	0	0	0	1	0	0	0	0	0	1.72	0		β_4
11		1	1	0	0	0	0	0	5.61	0	0	0	0	0		β_5
14		1	0	0	1	0	0	0	0	0	5.61	0	0	0		β_6
10		1	0	0	0	1	0	0	0	0	0	5.61	0	0		
17		1	0	0	0	0	0	1	0	0	0	0	0	5.61		
0		0	1	1	1	1	1	1	0	0	0	0	0	0		

Regression analysis gives estimates of $\hat{\mu}$, $(\hat{y}_1,\ldots,\hat{y}_6),(\hat{\beta}_1,\ldots,\hat{\beta}_6)$ of 7.87, (-2.64, 0.57, -0.62, -0.22, 1.02, 1.88), (1.03, 0.91, 1.20, 0.38, 1.37, 1.32) respectively with RMS = 0.572 with 4 df.

Step 3

<u>Example data</u>

Variety	Year 1	2	3	4	5	6
A	-	-	-	6	2	3
B	-	6	4	-	6	7
C	7	10	-	8	11	-
D	11	-	14	10	-	17

We now solve $c_{ij} = \mu + y_j + \hat{\beta}_j v_i + \varepsilon_{ij}$ for μ, ($y_1,\ldots,y_6$), ($v_1,\ldots,v_4$) and substitute the values $\hat{\beta}_j\,(j = 1,\ldots,n_y)$ obtained at step 2 viz (1.03, 0.91, 1.20, 0.38, 1.37, 1.32). In matrix terms this is

c = Xb

or

$$
\begin{vmatrix} 6 \\ 2 \\ 3 \\ 6 \\ 4 \\ 6 \\ 7 \\ 7 \\ 10 \\ 8 \\ 11 \\ 11 \\ 14 \\ 10 \\ 17 \\ 0 \\ 0 \end{vmatrix}
=
\begin{vmatrix}
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0.38 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 1.37 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 1.32 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0.91 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1.20 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1.37 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1.32 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1.03 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0.91 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0.38 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1.37 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.03 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1.20 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0.38 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1.32 \\
0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1
\end{vmatrix}
\text{ x }
\begin{vmatrix} \mu \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{vmatrix}
$$

Regression analysis gives estimates $\hat{\mu}$, $(\hat{y}_1,\ldots,\hat{y}_6),(\hat{v}_1,\ldots,\hat{v}_4)$ as 7.80, (-2.55, 0.48, -0.87, -0.09, 1.12, 1.91), (-5.05, -2.24, 1.62, 5.67), respectively with RMS = 0.0952 on 6 df.

Repeat from the previous step until the ratio of RMS values from 2 cycles is greater than a defined constant close to 1 eg 0.999. In this case 13 iterations were required for convergence giving a final residual mean square (RMS) of 0.1255 with 1 df. The final estimates $\hat{\mu},(\hat{y}_1,...,\hat{y}_6),(\hat{\beta}_1,...,\hat{\beta}_6),(\hat{v}_1,...,\hat{v}_4)$ were 7.862, (-2.12, 0.55, -1.20, -0.12, 1.16, 1.73), (0.91, 1.14, 1.26, 0.36, 1.39, 1.28), (-5.09, -2.12, 1.38, 5.81), from which the following table of means is derived:

	Year						Means
Variety	1	2	3	4	5	6	
A	-	-	-	6	2	3	2.78 = 7.86 + -5.09
B	-	6	4	-	6	7	5.76
C	7	10	-	8	11	-	9.24
D	11	-	14	10	-	17	13.67
Means	5.74	8.42	6.66	7.75	8.92	9.03	
Sensitivities	0.91	1.14	1.26	0.36	1.37	1.39	

The part of the matrix $(X'X)^{-1}$ relevant to the varieties is:

Varieties	A	B	C	D
A	0.2943			
B	0.0081	0.2094		
C	-0.0211	0.0145	0.3009	
D	-0.0313	0.0180	-0.0444	0.3077

From this using formula (2) it can be seen, for example, that the variance of the difference between varieties A and D is:

$$(0.2943 + 0.3077 + 2 \times 0.0313) \times 0.1255 = 0.08341$$

Using the t value at 1% level with 1 df of 63.66, the 1% LSD between varieties A and D is 18.39. This is compared with their actual difference of 10.89. Thus these varieties are not significantly different at the 1% level.

Proceeding in this way, the full table of 1% LSD values between all variety pairs is:

Variety	A	B	C
B	15.75		
C	18.00	15.64	
D	18.39	15.64	18.83

C Case study illustrating reasons for the disparity between the approaches

Candidate variety A and control variety H were classified as not distinct using the complete data approach on the “Height at Ear Emergence” character but distinct by the cyclic controls approach. The following table illustrates the reason for this. In rows 1 and 2 are the Height at Ear Emergence means of the two varieties in each year of the test period (96, 97, 98) and the variety H means in ’93 and ’94. These latter are used in the cyclic controls approach to compensate for the bracketed 1998 control H mean which is regarded as missing. For simplicity the sensitivities are assumed to have a slope of 1.

	Year						Complete data mean	Cyclic control mean
				Test Period				
	'93	'94	'95	'96	'97	'98		
1. *CANDIDATE MEAN (A)*	-	-	-	33.15	35.97	32.66		
2. *CONTROL MEAN (H)*	45.53	46.12	-	37.29	37.87	(34.37)		
3. *YEAR MEAN (YM)*	39.34	45.57	32.97	38.68	39.92	39.58		
4. *CORR' CANDIDATE MEAN (= A - YM)*	-	-	-	-5.53	-3.95	-6.92	-5.47	-5.47
5. *CORR' CONTROL MEAN (= H - YM)*	6.19	0.55	-	-1.39	-2.05	(-5.21)	-2.88	0.83
				Differences			2.59	6.30

Year means over a large number of varieties derived by a fitted constants method are given for each year (row 3). These are subtracted from both the candidate and control variety means to correct for the influence of years (rows 4 and 5).

In both approaches the overall corrected mean for variety A is calculated as the mean of the numbers in row 4 and columns '96, '97 and '98. The overall corrected mean for variety H is calculated from row 5, averaging the numbers in columns '96, '97 and '98 in the complete data approach and the numbers in columns '93, '94, '96 and '97 in the cyclic controls approach. Thus the large negative corrected control H mean for '98 is included in the average in the complete data approach and is replaced by the two positive corrected control means for '93 and '94 in the cyclic controls approach. The resulting overall corrected mean for variety H is much larger using the cyclic controls approach than the complete data approach, as is the difference between the overall corrected means for varieties A and H. Hence the disparity in the decisions by the two approaches, which can be summarised as being due to the values for the control variety in 1995 to 1997 being lower than would be expected on its performance in 1993 and 1994.

D Software used in the analysis

A statistical analysis program, COYD9, has been developed to enable the compensated data to be retrieved and analysed for distinctness by the MJRA technique, and the results presented in reports suitable for presentation to decision making groups. Uniformity testing continues to be based on the data within the test period and uses the BIOSS program COYU9. Both programs are available as part of the DUST9 (MSDOS based) and DUSTNT (Windows NT and 95) versions of the DUST software.

[End of document]