E

TWC/17/14
ORIGINAL: English
DATE: August 20, 1999

UPOV

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

APPLICATION OF A THRESHOLD MODEL ON A NUMBER OF UPOV CHARACTERISTICS

Document prepared by the experts from the Netherlands

Application of a threshold model on a number of UPOV characteristics

1. Introduction

1. In the past several documents have been presented about visually observed data. TWC/17/6 gives a clear overview. The 2 documents about threshold models are TWC/14/12 and TWC/15/14. The first handles the theory and shows an application on the basis of a small example. The second compares a threshold model analysis with an analysis of variance on the basis of 2 bigger data sets. This document demonstrates the application of threshold models on a great number of characteristics in different species. The basis for these models is that the observations are recorded in an ordinal (or rank) scale, see TWC/10/08, and that the distribution of the underlying variable is unimodal.

2. By showing the results visually there is a better insight in the results of the analysis. Besides a better understanding this can also lead to a reconsideration of the notes to be used.

2. Available data

3. The same data as in document TWC/15/14 (from Geves in France) are used, namely the character 'alternativité' (in French) or 'the tendency to form inflorescenses in the year of sowing' (in English) for the species Cocksfoot and Tall fescue.

4. From Tystofte in Denmark data of characteristics of the following species are available:

Species	*Characteristic*	*UPOV no.*
Peas	Stipule: maximum density of flecking	34
Ryegrass	Plant: growth habit in autumn	2
	Leaf: color	5
Spring rape	Leaf: green color	4
	Leaf: lobes	5
	Leaf: dentation of margin	7
	Flower: color of petals	12
	Tendency to form inflorescenses in year of sowing	22
Timothy	Plant: growth habit in 2nd year before elongation	5

3. Elementary explanation of a threshold model

5. We explain the threshold model on the basis of the data of the characteristic 'Tendency to form inflorescenses in the year of sowing' of the species Tall fescue. The records of 180 plants (20 plants in 3 replicates in 3 years) are tabulated in Table 7 of TWC document 15/14 Rev. At three timepoints (with 1 month intervals) each plant is observed and gets a note of 2, 4, 6 or 8 in the following way:

8 = the plant has 3 inflorescenses at the first visit
6 = the plant has 3 inflorescenses at the second visit, but not in the first
4 = the plant has 3 inflorescenses at the third visit, but not in the first and second
2 = the plant doesn't flower after the three visits

6. A threshold model assumes that there is an underlying variable which (although it can not be observed) determines in which category an individual plant will lie. Furthermore it is assumed that this underlying variable follows a distribution like the normal or the logistic and that the number of plants in each category give information about the location and the variation of this variable (characteristic).

7. Let's first assume that all varieties have the same variation and that the underlying variable follows a logistic distribution. Then performing a threshold model on the data of the first 10 varieties of Tall fescue the thresholds (cutpoints) and the means of each variety are estimated such that the observed and fitted percentages are as close as possible, see Table 1:

Table 1: Observed and fitted percentages of the number of plants of Tall fescue

	% Observed				% Fitted				Difference			
class	2	4	6	8	2	4	6	8	2	4	6	8
variety												
426960	63	5	26	6	62	7	25	6	+1	-2	+1	0
426950	78	8	11	2	79	5	14	3	-1	+3	-3	-1
423810	48	9	38	5	50	8	33	9	-2	+1	+5	-4
423710	30	7	39	24	27	7	45	21	+3	0	-6	+3
419770	43	8	42	7	45	8	36	11	-2	0	+6	-4
416800	31	5	43	21	29	7	45	20	+2	-2	-2	+1
416530	34	5	46	14	34	8	42	16	0	-3	+4	-2
416520	44	8	41	7	46	8	36	10	-2	0	+5	-3
411190	29	9	39	23	28	7	45	20	+1	+2	-6	+3
407030	31	11	36	23	30	7	44	19	+1	+4	-8	+4

8. Figure 1 is a visualisation of these fitted percentages in the form of 10 logistic distributions only different in location. The estimated thresholds are plotted as well.

9. Testing whether there are significant differences (P=0.05) in distinctness of the characteristic gives the following results:

*** Homogeneous groups, P=0.05

variety 426950	-0.8400	a . . . .
variety 426960	0.0000	. b . . .
variety 423810	0.4936	. b c . .
variety 416520	0.6565	. . c d .
variety 419770	0.6929	. . c d .
variety 416530	1.1501	. . . d e
variety 407030	1.3338	. . . . e
variety 416800	1.3947	. . . . e
variety 411190	1.4298	. . . . e
variety 423710	1.4752	. . . . e

10. Instead of showing the complete distribution we can show the range of the distribution between the 2.5% and 97.5% point of the distribution by drawing a line between these points with a cross in the middle indicating the mean of the distribution. This is shown in Figure 2. The advantage of displaying the results in this way is the possibility to give a concise

overview of a great number of varieties. The display for all 85 varieties of Table 7 of TWC/15/14 is shown in Figure 3.

11. Apart from testing the distinctness between varieties threshold models give also information about the used classes (scores or notes). As you can see from Figure 1 and 2, category 4 is very narrow indicating a note not so much scored. So the difference between plants with notes 2 and 6 is much smaller than the difference between notes 4 and 8.

12. Looking at the same characteristic for Cocksfoot, see Figure 4, gives a similar picture. Class 4 is not very informative, so that maybe the 3rd visit can be skipped. The figure even suggests to skip the 2nd visit, so only to score notes 2 and 8. For testing distinctness this would be okay but for testing uniformity a characteristic with only 2 notes isn't very informative. Therefore it can be better to choose the timepoints more close by in stead of visiting at timepoints with 1 month intervals.

4. Characteristics of other species

4.1 PEAS TG/7/9 STIPULE: MAXIMUM DENSITY OF FLECKING UPOV no 34

13. The notes of this characteristic can be:

	Examples
1 = very sparse	*Progreta, Resco*
3 = sparse	*Allround, Finale*
5 = medium	*Mars, Sentinel*
7 = dense	*Avola, Roi de Carouby*
9 = very dense	

In the dataset notes 1, 2, 3, 4, 5, 6 and 7 were present.

14. In each of the years 1995, 1996 and 1997 two replicates were available. In 1995 ±20 plants were observed per replicate and variety and in 1996 and 1997 ±10 plants. Not every variety was present each year. The minimum number of plants per variety (over replicates and years) was 13 and the maximum number 120.

15. The results of the analysis with a threshold model are shown in Figure 5. From this picture it is clear that the not officially notes 2, 4 and 6 are much less present than the notes 1, 3, 5 and 7.

4.2 RYEGRASS TG/4/7 PLANT: GROWTH HABIT IN AUTUMN UPOV no 2

16. The notes of this characteristic can be:

	Examples
1 = erect	
3 = semi-erect	*Trani (Lp), Matador (Lm)*
5 = medium	*Talbot (Lp)*
7 = semi-prostrate	*Barclay (Lp), Wilo (Lm)*
9 = prostrate	

In the dataset notes 3, 5, 7 and 9 were present.

17. In each of the years 1995 and 1997 three replicates were available. In each of the 2 years ±20 plants were observed per replicate and variety. The varieties in 1995 were not present in 1997 and vice versa. The minimum number of plants per variety (over replicates and years) was 53 and the maximum number 61.

18. The results of the analysis for each of the years are shown in Figures 6 and 7.

4.3 RYEGRASS TG/4/7 LEAF: COLOR UPOV no 5

19. The notes of this characteristic can be:

	Examples
1 = very light green	
3 = light green	*Callan (Lp)*
5 = medium green	*Melina(Lp), Lemtal (Lm)*
7 = dark green	*Condesa (Lp), Elving (Lm)*
9 = very dark green	

In the dataset notes 3, 5, 7 and 9 were present.

20. In each of the years 1995 and 1997 three replicates were available. In each of the 2 years ±20 plants were observed per replicate and variety. The varieties in 1995 were not present in 1997 and vice versa. The minimum number of plants per variety (over replicates and years) was 52 and the maximum number 61.

21. The results of the analysis for each of the years are shown in Figures 8 and 9.

4.6 RAPE SEED TG/36/5 LEAF: GREEN COLOR UPOV no 4

22. The notes of this characteristic can be:

	Examples
3 = light	*Linetta; Anton*
5 = medium	*Drakkar, Jaguar; Akela*
7 = dark	*Logo, Orly; Gaspard*

23. In the dataset notes 2, 3, 4, 5, 6, 7 and 8 were present.

24. In each of the years 1995, 1996 and 1997 three replicates were available. In all years most of the time 1 plant was observed per replicate and variety. In 1995 and 1996 only 2 replicates were used and in 1997 all 3. Not every variety was present each year. The minimum number of plants per variety (over replicates and years) was 2 and the maximum number 17.

25. The results of the analysis are shown in Figure 10. Because of the assumed constant variability over varieties even for varieties with 2 observations a distribution can be plotted.

4.7 RAPE SEED TG/36/5 LEAF: LOBES UPOV no 5

26. The notes of this characteristic can be:

	Examples
1 = absent	*Arista, Orly; Akela*
9 = present	*Drakkar; Falcon, Samourai*

27. In the dataset notes 1, 3, 5 and 9 were present.

28. In each of the years 1995, 1996 and 1997 three replicates were available. In all years most of the time 1 plant was observed per replicate and variety. In 1995 and 1996 only 2 replicates were used and in 1997 all 3. Not every variety was present each year. The minimum number of plants per variety (over replicates and years) was 2 and the maximum number 17.

29. The results of the analysis are shown in Figure 11. From the graph it is obvious that the notes of the same variety are almost nearly the same. A threshold model isn't appropriate here.

4.8 RAPE SEED TG/36/5 LEAF: DENTATION OF MARGIN UPOV no 7

30. The notes of this characteristic can be:

	Examples
3 = weak	*Orly; Arvor*
5 = medium	*Drakkar; Diadem, Tapidor*
7 = strong	*Briol; Stego*

31. In the dataset notes 2, 3, 4, 5, 6, 7, 8 and 9 were present.

32. In each of the years 1995, 1996 and 1997 three replicates were available. In all years most of the time 1 plant was observed per replicate and variety. In 1995 and 1996 only 2 replicates were used and in 1997 all 3. Not every variety was present each year. The minimum number of plants per variety (over replicates and years) was 2 and the maximum number 17.

33. The results of the analysis are shown in Figure 12. From the graph it is clear that the notes of the same variety are almost nearly the same.

4.9 RAPE SEED TG/36/5 FLOWER: COLOR OF PETALS UPOV no 12

34. The notes of this characteristic can be:

	Examples
1 = white	*-; -*
2 = cream	*-; Hobson*
3 = yellow	*Lisonne; Balcon, Samourai*
4 = orange-yellow	*-;Pasha*

35. In the dataset notes 1, 4, 5, 6 and 7 were present.

36. In each of the years 1995, 1996 and 1997 three replicates were available. In all years most of the time 1 plant was observed per replicate and variety. In 1995 and 1996 only 2

replicates were used and in 1997 all 3. Not every variety was present each year. The minimum number of plants per variety (over replicates and years) was 2 and the maximum number 17.

37. The results of the analysis are shown in Figure 13. From the graph it is clear that most of the varieties have always note 5.

4.10 RAPE SEED TG/36/5 TENDENCY TO FORM INFLORESCENSES IN YEAR OF SOWING FOR LATE SUMMER SOWN TRIALS UPOV no 22

38. The notes of this characteristic can be:

	Examples
1 = absent or very weak	*Petranova;* -
3 = weak	*Kardinal;* -
5 = medium	
7 = strong	*Lisonne;* -
9 = very strong	*Drakkar;* -

39. In the dataset notes 1, 2, 3, 4, 5, 6, 7, 8 and 9 were present.

40. In each of the years 1995, 1996 and 1997 three replicates were available. In all years most of the time 1 plant was observed per replicate and variety. In 1995 and 1996 only 2 replicates were used and in 1997 all 3. Not every variety was present each year. The minimum number of plants per variety (over replicates and years) was 2 and the maximum number 17.

41. The results of the analysis are shown in Figure 14. From the graph it is clear that some of the varieties have always note 1 or 9. For most varieties with notes between 1 and 9 there is some variation.

4.11 TIMOTHY TG/34/6 PLANT: GROWTH HABIT IN 2nd YEAR BEFORE ELONGATION UPOV no 5

42. The notes of this characteristic can be:

	Examples
1 = erect	
3 = semi-erect	*Phlewiola (P.p)*
5 = medium	
7 = semi-prostrate	*Castella (P.p.)*
9 = prostrate	

43. In the dataset notes 1, 2, 3, 4 and 5 were present.

44. In each of the years 1995, 1996 and 1997 three replicates were available. In all years ±20 plants were observed per replicate and variety. Not every variety was present each year. The minimum number of plants per variety (over replicates and years) was 50 and the maximum number 179.

45. The results of the analysis are shown in Figure 15.

5. Extending the model with unequal variances for the varieties

46. The former analyses are all based on comparisons between varieties assuming constant variances for all varieties. This was done because of 3 reasons:

47. First, at this moment procedure DORDINAL within the statistical package Genstat can only make plots for equal variances. For a matter of fact all analyses are performed with Genstat.

48. Secondly, extending the model with variances for each of the varieties separately can only be done using procedure CLASS. This procedure uses directive REML with as many covariates as the number of varieties. Since at this moment REML can't handle more than 20 covariates datasets with more than 20 varieties can't be analyzed in one single run.

49. Third some characteristics like the characteristics of Rape seed had so little observations that it should be impossible to estimate the variability.

50. Nevertheless when extending the model with variances for each of the varieties separately the results of this model are comparable with the results of analyses of measured characteristics. The analysis estimates means and log standard deviations as COYD and COYU do for measured characteristics.

6. Discussion and conclusion

51. Not all visually scored data can be analysed by a threshold model. The notes should be recorded on an ordinal scale and the distribution of the underlying variable should be unimodal.

52. To estimate variances for each variety separately a lot of individual plants per variety must be observed.

53. The results of threshold models are analogous with those of measured characteristics. Means and standard deviations for each variety are estimated.

54. The results, especially the location of the cutpoints, can be used to reconsider the choice of the notes.

55. At this moment there is no software that performs threshold model analyses automatically, particularly for testing uniformity.

7. Acknowledgements

56. Thanks are due to Sylvain Gregoire and Kristian Kristensen for kindly making available the data on Tall fescue and Cocksfoot resp. on Peas, Spring rape seed, Ryegrass and Timothy.

8. Reference

Keen, A. and B. Engel (1997), Analysis of a mixed model for ordinal data by iterative re-weighted REML, *Statistica Neerlandica,* **50,** 129-144

[Figures 1-15 follow]