E

TWC/29/23
ORIGINAL: English
DATE: May 24, 2011

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

TECHNICAL WORKING PARTY ON AUTOMATION AND COMPUTER PROGRAMS

Twenty-Ninth Session
Geneva, June 7 to 10, 2011

COMPARISON OF UNIFORMITY DECISIONS BASED ON COYU AND MODIFIED BENNETT'S METHODS USING REAL AND SIMULATED DATA

Document prepared by experts from Poland

COMPARISON OF UNIFORMITY DECISIONS BASED ON COYU, BENNETT'S AND MODIFIED BENNETT'S METHODS USING REAL AND SIMULATED DATA

[1]BOGNA ZAWIEJA, [1,2]WIESŁAW PILARCZYK, [2]BOGNA KOWALCZYK

[1]Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Wojska Polskiego 28, 60-637 Poznań, Poland
[2]The Research Centre for Cultivar Testing, 63-022 Słupia Wielka, Poland
e-mail: bogna13@up.poznan.pl; wpilar@up.poznan.pl; b.kowalczyk@coboru.pl

Summary

1. In Poland uniformity of new varieties for quantitative characteristics is usually checked using the COYU method after collecting results from two or three years of trials. There are some other possibilities of testing uniformity as indicated, for example, in papers by Zawieja and Pilarczyk (2005, 2006) and by Zawieja, Pilarczyk and Kowalczyk (2009).

2. In documents TWC/23/9 "A Comparison of COYU and a Method Based on Bennett's Test for Coefficients of Variation", TWC/24/7 "Further Comparison of Decisions on Uniformity of Rye Varieties Based on COYU Approach and on Bennett's Test", and TWC/25/8 "Comparison of COYU and a Method Based on Bennett's Test for Coefficients of Variation", the conclusions concerning uniformity of rye varieties based on the UNIF (COYU) approach and on the Bennett's test were compared. The conclusions were generally similar, but in some cases differences appeared.

3. During the discussions at the twenty-fourth session of the TWC, held in Nairobi, Kenya, from June 19 t0 22, 2006, it was proposed to make additional comparisons of these two methods in order to investigate if there was a relationship between the degree of correlation between level of expression of characteristic and log transformed values of standard deviations and decisions concerning uniformity supported by the two mentioned methods. It was also suggested to apply McNemar's (McNemar, 1947) test instead of a test of independence. This problem was initially discussed at the twenty-fifth session of the TWC (see document TWC/25/8) and - in conclusion - it was also suggested to compare these two methods of testing uniformity using results of another species. In document TWC/27/10 these problems were addressed again with the use of DUS data for oilseed rape varieties. There were some differences between decisions concerning uniformity for these two methods, but they were statistically indistinct (when tested at 0.01 significance level). Because there were only six candidate varieties, during discussions it was suggested to use larger set of candidates using simulated data.

4. The problem was also discussed at the Ninth Working Seminar on Statistical Methods in Variety Testing (Dolsk, Poland, June 2010). It was suggested by Johannes Forkman from Sweden to modify the Bennett's approach by replacement the 2Z used for testing equality of coefficients of variation with the F-Fisher statistic as described by Forkman (2009). In this document the results of comparisons of COYU, Bennett's test and modified Bennett's test are presented. The real data for reference set varieties are taken from oil-seed rape trials performed in Poland whereas the data for candidate varieties are generated (simulated).

Introduction

5. In the case of uniformity of cross-pollinated varieties, the General Introduction explains that

> "6.4.2 Cross-Pollinated Varieties
>
> Cross-pollinated varieties, including mainly cross-pollinated and synthetic varieties, generally exhibit wider variations within the variety than vegetatively propagated or self-pollinated varieties and inbred lines of hybrid varieties, and it is more difficult to determine off-types. Therefore, relative tolerance limits, for the range of variation, are set by comparison with comparable varieties, or types, already known. This means that the candidate variety should not be significantly less uniform than the comparable varieties.
>
> […]
>
> 6.4.2.2 Measured Characteristics
>
> 6.4.2.2.1 For measured characteristics, the acceptable level of variation for the variety should not significantly exceed the level of variation found in comparable

varieties already known. UPOV has proposed several statistical methods for dealing with uniformity in measured quantitative characteristics. One method, which takes into account variations between years, is the Combined Over Years Uniformity (COYU) method.

6.4.2.2.2 For more details on the handling of uniformity in measured quantitative characteristics, see document TGP/10, "Examining Uniformity."

6. In the COYU method, the log transformed and adjusted by moving average method, values of standard deviations of new varieties are compared with similar (averaged) values calculated for varieties treated as standards. Such comparisons are made for all relevant measured characteristics in DUS trials. If new values for the variety do not exceed significantly the average values of reference varieties for all characteristics under consideration, the new variety is accepted and in the next cycles it can be included in the set of reference[1] varieties.

7. Because standard deviations sometimes depend on the levels of expression of the characteristic under consideration some additional procedures have been elaborated to remove these influences. The COYU method is a slightly sophisticated method; a possible alternative is the application of a measure of uniformity based on coefficient of variation. Such an approach was described in documents TWC/23/9, TWC/24/7 and TWC25/8. Equality of coefficients of variation of the new (candidate) variety and of the varieties belonging to the reference set can also be tested using the Bennett test, which is much simpler than COYU.

Data

8. The data from DUS trials on oilseed rape performed at experimental station Słupia Wielka in the period 2006-2008 form the basis of investigations. Only data for varieties already registered are used.

9. Because the aim of this research was comparison of decisions concerning uniformity supported by COYU and by modified Bennett's test, and – additionally – the comparison of decisions supported by Bennett's test and modified Bennett's test, there was no necessity of use of all characteristics observed. So one characteristic – the plant height – was chosen. For every of analysed periods, namely 2006-2007, 2007-2008 and 2006-2008, the data for candidate varieties were generated using method as follows:

1) the minimum and maximum value of real variety mean and standard deviation were calculated x_{min}, x_{max}, s_{min}, s_{max};

2) starting from (rounded) x_{min}, the values for "candidate" varieties were formed using formula

$$x_i = x_{min} + (i\text{-}1)d, \quad i=1,2,3,....$$

where values x_i were generated as far as x_{max} was reached.

3) every value x_i was associated with all values of standard deviations generated as follows

$$s_j = s_{min} + (j\text{-}1)s, \quad j = 1,2,3,....$$

[1] The term reference varieties here refers to established varieties which have been included in the growing trial and which have comparable expression of the characteristics under investigation.

where the s_j were generated so far as s_{max} was reached.
The values of d and s were chosen in a way that guarantee the reasonable number of "candidate" varieties.

10. For the period 2006-2007, there were 66 established varieties (forming so-called reference set) and 187 candidate (simulated) varieties. Similarly for the period 2007-2008, there were 57 established and 272 simulated varieties and finally, for the period 2006-2008, 72 and 238 such varieties. Uniformity of every "candidate" variety was tested using the methods described bellow.

The Method

11. Each candidate variety was tested using COYU (combined over year uniformity) method, Bennett's and modified Bennett's tests. The method similar to that described by Zawieja, Pilarczyk and Kowalczyk (2009) was used to compare decisions concerning uniformity. The COYU method uses average values of within-plot standard deviations as a measure of uniformity. These values are next ln (natural logarithm) transformed, and "adjusted" using moving average approach. Adjusted values are compared with similar values received for the reference set varieties. All details of COYU approach can be found in a paper by Talbot (2000).

12. In a Bennett's approach the coefficients of variations are used as a measure of uniformity. Equality of coefficient of variation of candidate variety and a subset of coefficients of variation of reference set varieties is a criterion of acceptance of candidate variety as uniform. It can be applied when all coefficients of variation are not higher than 0.3 (Forkman, 2006 Iglewicz and Meyers, 1970). In our case this condition was always fulfilled. The subset of reference set varieties was formed in similar way as in COYU approach, namely varieties with closest mean values were taken. More details on Bennett's test are given in a paper by Zawieja and Pilarczyk (2006).

13. The decisions concerning uniformity of candidate varieties supported by the two considered methods are compared in such way, that two by two contingency table (Table 1) is formed at first of the form

Table 1. Two by two contingency table for decisions on uniformity of candidate varieties

Method		Bennett's test	
COYU approach	decision	uniform	not uniform
	uniform	n_{11}	n_{12}
	not uniform	n_{21}	n_{22}

14. The COYU and Bennett's methods were applied at the same significance level. The $n_{11}+n_{22}$ denotes the number of unanimous decisions why $n_{12}+n_{21}$ denotes the number of contradictory decisions. There are several methods for testing concordance of decisions with use such data. In a paper by Zawieja and Pilarczyk (2006) the Fisher exact test was used to find out if there is an association between decisions, why in a paper by Zawieja and Pilarczyk (2007) the McNemar test was used to verify if the hypothesis that probabilities of contradictory decisions $p_{12} = p_{21}$ can be accepted or not. Here the "odds ratio" *OR* (Rudas, 1998, Uebersax 2006) is applied as a measure of association between decisions. Odds ratio is calculated as

$$OR = \frac{n_{11} \bullet n_{22}}{n_{12} \bullet n_{21}}$$

15. Large value of *OR* indicates association between methods. The statistical significance of lack of association can be tested using statistics Z_0 of the form

$$Z_0 = \frac{\ln(OR)}{\sigma_{\ln(OR)}}$$

where $\sigma_{\ln(OR)} = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$. The Z_0 statistics has asymptotic normal distribution.

16. Coefficient *OR* can be easily transformed to the Yule coefficient of association *Q* (Yule and Kendall, 1966), using formula

$$Q = \frac{OR - 1}{OR + 1}$$

This coefficient is interpreted similarly to the coefficient of correlation. $Q = 0$ means lack of association between methods, value close to 1 means high agreement.

17. To have additional characterisation of association, the probability *p* of agreement (concordance) was also calculated according to the formula

$$p = \frac{n_{11} + n_{22}}{n}$$

where *n* denotes the total number of candidate varieties.

The Results

18. The full comparison of COYU and Bennett's test – using the same data – is given in the TWC/28/26 document. Therefore in this document only the results of comparisons of COYU with modified Bennett's test and Bennett's with modified Bennett's test are given.

19. All three methods were applied for three sets of data generated in above described method (data for candidate varieties). The data for reference varieties were taken from real experiments performed at the experimental station Słupia Wielka. The COYU analysis was performed with use of DUST package of Weatherup (1992). For Bennett's and modified Bennett's tests the EXCEL spreadsheet was utilized.

20. The results of comparison of COYU and modified Bennett's test for two years data concerning period 2006-2007 are given in Table 2 (testing at significance level $\alpha = 0.002$ and 0.02). Similar results for the period 2007-2008 are given in Table 3.

Table 2. Decisions on uniformity of candidate varieties (for data from the period 2006-2007)

Significance level		$\alpha = 0.002$		$\alpha = 0.02$			
Method		Modified Bennett		Method		Modified Bennett	
COYU approach	decision	uniform	Not uniform	COYU approach	decision	uniform	not uniform
	uniform	162	25		uniform	142	36
	not uniform	0	0		not uniform	0	9

When testing was performed at the level $\alpha = 0.002$, the probability of concordance was $p = 86.6\%$, but when the level of significance 0.02 was used the probability of concordance between methods equals to 80.7%.

For the period 2007-2008, the probability of concordance was equal to 77.9% (when testing was performed at 0.002 level) and 74.6% (testing at 0.02 level).

Table 3. Decisions on uniformity of candidate varieties (for data from the period 2007-2008)

Significance level		$\alpha = 0.002$		$\alpha = 0.02$			
Method		Modified Bennett		Method		Modified Bennett	
COYU approach	decision	uniform	not uniform	COYU approach	decision	uniform	not uniform
	uniform	212	60		uniform	184	69
	not uniform	0	0		not uniform	0	19

21. The results for the 2007-2008 period are presented in Table 4 (also for $\alpha = 0.002$ and α=0.02). When testing was performed at $\alpha = 0.002$ level, probability of concordance was p = 69.33%. For testing performed at $\alpha = 0.02$ level, the value p = 69.75 %was obtained.

Table 4. Decisions on uniformity of candidate varieties (for data from the period 2006-2008)

Significance level		$\alpha = 0.002$		$\alpha = 0.02$			
Method		Modified Bennett		Method		Modified Bennett	
COYU approach	decision	uniform	not uniform	COYU approach	decision	uniform	not uniform
	uniform	162	73		uniform	141	26
	not uniform	0	3		not uniform	0	25

22. The results of the comparison of original Bennett's method (with use of $2Z$ statistic) and modified Bennett's method test (the method that uses F statistic) are presented in the Tables 4, 5 and 6.

Table 5. Decisions on uniformity of candidate varieties (for data from the period 2006-2007)

Significance level		$\alpha = 0.002$		$\alpha = 0.02$			
Method		Bennett		Method		Bennett	
Modified Bennett	decision	uniform	not uniform	Modified Bennett	decision	uniform	not uniform
	uniform	162	0		uniform	142	0
	not uniform	25	0		not uniform	45	0

When testing was performed at the level $\alpha = 0.002$, the probability of concordance was $p = 86.6\%$. But for tests at 0.02 level the probability of concordant decisions between methods equals to 75.9%.

23. The results for the 2007-2008 period are presented in Table 6. The probability of concordance was equal to 77.9% (when testing performed at 0.002 level) and 72.1% (for

testing at 0.02). At this significance level the other measures of concordance were OR =12.4 Q=0.85 Z =3.888.

Table 6. Decisions on uniformity of candidate varieties (for data from the period 2007-2008)

Significance level		$\alpha = 0.002$		$\alpha = 0.02$			
Method		Bennett		Method		Bennett	
Modified Bennett	decision	uniform	not uniform	Modified Bennett	decision	uniform	not uniform
	uniform	212	0		uniform	181	3
	not uniform	60	0		not uniform	73	15

24. The results for the three years period (2006-2008) are presented in Table 7. When testing was performed at α = 0.002 level, the probability of concordance was p = 76.05%. When testing was performed at α = 0.02 level, the probability of concordance was p = 76.89%. The other measures of association (for testing at 0.02 level) are equal respectively $OR = 39.808, Z = 5.96$ $Q = 0.95$.

Table 7. Decisions on uniformity of candidate varieties (for data from the period 2006-2008)

Significance level		$\alpha = 0.002$		$\alpha = 0.02$			
Method		Bennett		Method		Bennett	
Modified Bennett	decision	uniform	not uniform	Modified Bennett	decision	uniform	not uniform
	uniform	162	0		uniform	138	3
	not uniform	57	19		not uniform	52	45

Comments and Conclusions

25. In papers by Zawieja and Pilarczyk (2005, 2006) it has been shown that the COYU method and the Bennett's test applied to real data concerning winter rye varieties did not differ statistically. It was observed that the Bennett's method was slightly more tolerant than COYU method but that statistically (at α = 0.01 level) these two methods gave the same decisions.

26. In paper by Zawieja and others (2009), using real oilseed rape data, it has been shown that again these two method did not differ statistically but for oilseed rape the method the COYU was slightly more tolerant. In all previous investigation there were very limited numbers of candidate varieties.

27. The results presented last year, see TWC/28/26 (with use mixture of real and simulated data) showed that in some cases these two methods of testing varietal uniformity did not differ (results for years 2006-2007). In some other cases there existed meaningful differences in decisions. The Bennett's test rejected more candidate varieties. Detailed inspection of analyzed data indicated that in all such cases the Bennett's test rejected varieties with small mean values and high standard deviations (with large coefficients of variation). COYU method was - for part of such varieties - more tolerant.

28. The Bennett's approach with replacement $2Z$ statistic with the F statistics used for testing uniformity of candidate varieties is more restrictive (less varieties accepted as uniform) than COYU. The Bennett's method with F used for testing uniformity of candidate varieties is more restrictive than original Bennett's method. Both versions of Bennett's tests

(with and without F) reject usually varieties with small mean values and large standard deviations.

References

Forkman J., 2006, Statistical inference for the coefficient of variation in normally distributed data. Research Report 2006:2, Centre of Biostochatics Swedish University of Agricultural Sciences.

Iglewicz B., Meyers R. H., 1970, Comparison of approximations of the percentage points of the sample coefficient of variation. Technometrics 12:pp 166-169.

Kristensen K., Roberts A., 2009, Potential approaches to improving COYU. UPOV Geneva. TWC/27/15: str. 1-8.

McNemar Q., 1947, Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 12: pp. 153-157.

Rudas T., 1998, Odds Ratios in the Analysis of Contingency Tables. Thousand Oaks, CA: Sage Publ.

Shafer N. J., Sulivan J. A., 1986, A simulation study of test for the equality of the coefficient of variation. Communications in Statistics – Simulation and Computation 15: pp. 681-698.

Talbot M., 2000, The Combined-Over-Years Distinctness and Uniformity criteria. UPOV, TWC/18/10, Geneva

Uebersax J., 2006, Odds Ratio and Yule's Q. http://www.john-uebersax.com/stat/odds.htm.

Weatherup S.T.C., 1992, Distinctness, Uniformity and Stability trial (DUST) analysis system. User manual. Department of Agriculture for Northern Ireland Biometrics Division, Belfast BT9 5PX.

Yule G. U., 1912, On the methods of measuring association between two attributes. Journal of the Royal Statistical Society 75: pp. 579–652.

Yule G. U., Kendall M. G., 1966, Wstęp do teorii statystyki. PWN.

Zawieja B., Pilarczyk W., 2005, The comparison of traditional UPOV uniformity criterion and new approach based on Bennett's test for coefficients of variation. Colloquium Biometricum 35: str. 155-163.

Zawieja B., Pilarczyk W., 2006, The comparison of decisions on uniformity of rye varieties based on COYU approach and Bennett's test. Colloquium Biometricum 36: str. 225-233.

Zawieja B., Pilarczyk W., 2007, Further comparison of decisions concerning uniformity of rye varieties based on COYU approach and on Bennett's test. Colloquium Biometricum 37:str. 71-76

Zawieja B., Pilarczyk W., Kowalczyk B., 2009, The comparison of uniformity decisions based on COYU and Bennett's method – oilseed rape data. Colloquium Biometricum 39: pp. 170-176.

Zawieja B., Pilarczyk W., Kowalczyk B., 2010, Comparison of uniformity decisions based on COYU and Bennett's methods using real and simulated data, TWC/28/28.

[End of document]