Colloquium Biometricum 41
2011, 19–26

ON AN ALTERNATIVE METHOD OF TESTING VARIETAL UNIFORMITY IN DUS TRIALS

Bogna Zawieja[1], Wiesław Pilarczyk[1,2], Bogna Kowalczyk[2]

[1]Department of Mathematical and Statistical Methods
Poznań University of Live Sciences
Wojska Polskiego 28, 60–637 Poznań, Poland
[2]The Research Centre for Cultivar Testing, 63–022 Słupia Wielka, Poland
e–mail: bogna13@au.poznan.pl

Summary

In this paper frequencies of decisions concerning uniformity of new varieties taken using the COYU (combined over years uniformity) method and the Bennett method were compared. The Bennett method was applied in two versions, namely in its traditional form as described by Bennett [1976] and in a new version in which the chi–square statistic was replaced by F–Fisher statistics as described by Forkman (2009). All the three methods were applied earlier by the same authors, see Zawieja and others (2009, 2010), to real data from DUS (distinctness, uniformity and stability) trials on rye and oil–seed rape. In all analyzed previously data sets the number of candidate varieties were very limited. To avoid this shortcoming, in this paper these methods were applied to partly simulated data (data for reference varieties were real – taken from oil–seed rape trials – while data for candidate varieties were simulated). Some differences between decisions appeared but – in general – the decisions were statistically equivalent.

Key words and phrases: Bennett's method, coefficient of variation, COYU method, oilseed rape, simulation, test *F*, variety uniformity

Classification AMS 2010: 62K99, 62P10

1. Introduction

In all UPOV (International Union for the Protection of New Varieties of Plants) members countries, each new variety before listing in National List, have to be tested on account its distinctness, uniformity and stability (DUS). For this purpose special experiments are performed. These are usually planned in the randomized complete block design in 2–6 replicates, in one location. Decisions concerning DUS of varieties are usually taken after 2–3 years of testing. Observations are made for single plant (or parts of plant – stem, leaf, flower). There are three types of characteristics distinguished: quantitative – e.g. leaf long and width, qualitative – e.g. color of flowers, pseudo–qualitative – e.g. shape of leaf. In single trial, the number of measurements (observed plants) for quantitative characteristics is usually between 30 and 60. The number of characteristics – dependent on species – can be as high as 60–90.

In the DUS trials, for quantitative characteristics, decisions concerning distinctness are based on the mean values of characteristics (significant difference from any other variety for one characteristic fulfills this requirement). Decisions concerning uniformity are based on standard deviations for between–plants variation (variety must be uniform for all characteristics observed in DUS trial). Stability usually is not tested. By assumption each variety satisfactorily uniform is deemed to be also stable. If necessary, different generations of seeds are sown to check this requirement.

In this study three methods for checking uniformity of varieties of oil–seed rape are compared. The first of them is a method COYU, which is officially recommended for using in the member states of UPOV. In this method the standard deviation of each new variety is compared with the average standard deviation calculated over all known varieties. The second is a Bennett method. In this method the hypothesis of equality coefficients of variation of known varieties and new variety is tested with the use of the Bennett's test statistic. This method was proposed, as a test of uniformity, in the papers by Zawieja at al. (2009, 2010). In the third method, again the equality of coefficients of variation is tested, but this time with application of the F –Fisher test as described by Forkman (2009).

2. Data

To compare the results of application of three methods of testing uniformity (COYU, Bennett's, and F –test) the simulated data but with original data as a starting point were used. The data for known (established) varieties of oil–seed rape were taken from results of official DUS experiments performed in the

period 2006–2008 by the Research Center for Cultivar Testing at Słupia Wielka. All experiments were established in randomized complete block design with two replicates. 30 randomly chosen plants from each plot were measured giving in total 60 measurements for each variety. For every of analyzed periods, namely 2006–2007, 2007–2008 and 2006–2008, the data for candidate varieties were generated using the method as it is described in the paper by Zawieja et. al. (2010). In the period 2006–2007, there were 66 established varieties (forming so–called reference set) and 187 candidate (simulated) varieties. Similarly in the period 2007–2008, there were 57 established and 272 simulated varieties and finally, in the period 2006–2008, 72 and 238 such varieties.

3. Methods

Officially adopted (within UPOV) method of testing uniformity is known as COYU (combined over years uniformity) method. Before starting any statistical interpretation of uniformity, some basic statistical descriptors are calculated. These are $\bar{x}_i$ – the mean value for *i*–th variety (*i*=1,2, *v*), s_i^2 – the standard deviation for between plants variation for *i*–th variety. Both $\bar{x}_i$ and s_i^2 are calculated independently using data of DUS trials conducted across two or three years. These values supplemented by the number of measurements n_i for *i*–th variety (within years) and the number of years are sufficient to apply COYU method which is based on comparison of (transformed) standard deviation of any candidate variety with the mean value of standard deviations of reference set varieties. The threshold value UC of *j*–th characteristic, is calculated as

$$UC_j = \bar{s}_d + t_{p;w-1}^{tab}\sqrt{s^2\left(\frac{1}{l}+\frac{1}{lw}\right)}, \tag{3.1}$$

where $\bar{s}_d$ is the average of corrected standard deviations calculated over all varieties assigned to the reference collection (the set of varieties the new variety is compared with), s^2 is the sample variance among corrected standard deviations (of reference collection varieties) after removing the effects of years. Next, l stands for the number of years of trialling (usually 2 or 3), *w* is the size of reference collection, t_p means the one–side t–Student's distribution critical value at probability p and degrees of freedom associated with s^2 (see Talbot,

2000). Usually the value of $p = 0.001$ or $p = 0.002$ is accepted but other values are also admitted.

If (possibly adjusted) standard deviation of particular candidate variety is smaller than the UC_j value (threshold) for all considered characteristics, the variety is declared uniform. So, if for just one characteristic, the standard deviation is larger than the threshold, the variety is treated as non–uniform and as a consequence can not be registered.

In a Bennett's approach, the appropriate statistic for hypothesis $H_0: \ \zeta_1 = ... = \zeta_v (= \zeta, \text{say})$ is given by the formula

$$2Z = (n - v)\log\left(\frac{\sum_i y_i}{n - v}\right) - \sum_i (n_i - 1)\log\left(\frac{y_i}{n_i - 1}\right). \tag{3.2}$$

This statistic is approximately distributed as χ^2 with (v –1) degrees of freedom. In this formula, n_i denotes the number of measurements for i–th variety, y_i is calculated as

$$y_i = \frac{(n_i - 1)z_i^2}{1 + \frac{(n_i - 1)}{n} z_i^2}\ , \tag{3.3}$$

where z_i denotes the empirical coefficient of variation and where ψ_i is the transformed value of the theoretical coefficient of variation ζ_i, namely $\psi_i = \zeta_i^2 / \left(1 + \zeta_i^2\right)$.

In our first approach presented last year at the Ninth Working Seminar on Statistical Methods in Variety Testing (Dolsk, June 2010), the Bennett's test was used for two purposes: for testing of uniformity of reference set varieties, and for testing whether a candidate variety is sufficiently uniform. During the discussion after presentation, it was suggested by Johannes Forkman to replace the Bennett's test for testing uniformity of candidate variety t by the F statistic of the form (Forkman, 2009):

$$F = \frac{y_t / (n_t - 1)}{\sum_i y_i / \sum_i (n_i - 1)}, \tag{3.4}$$

It has an approximate *F* distribution with $n_t - 1$ and $\sum_i (n_i - 1)$ degrees of freedom.

Uniformity of every "candidate" variety was tested using the methods given bellow. Each variety was tested using COYU (combined over year uniformity) method and the Bennett's test. The method similar to that described by Zawieja at al. (2009) was used to compare decisions concerning uniformity. The Bennett's method can be applied when all coefficients of variation are not higher than 0.3 (Forkman 2009, Iglewicz and Meyers 1970). In our case this condition was always fulfilled. The decisions concerning uniformity of candidate varieties supported by the two methods are compared using two–way contingency table. Conclusions using the COYU and Bennett's methods were drawn at the same significance level. The n_{11}+n_{22} denote the number of unanimous decisions while n_{12}+n_{21} denotes the number of contradictory decisions. In order to compare these methods, the approach described by Zawieja et al. (2010) was used. The "odds ratio" *OR* (Rudas 1998, Uebersax 2006) was applied as a measure of association between decisions. Odds ratio is calculated as $OR = (n_{11}n_{22})/(n_{12}n_{21})$. Large value of *OR* indicates association between methods. The statistical significance of lack of association can be tested using statistic Z_0 of the form $Z_0 = \ln(OR)/\sigma_{\ln(OR)}$, where $\sigma_{\ln(OR)} = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$. The Z_0 statistic has an asymptotic normal distribution. The coefficient *OR* can be easily transformed to the Yule coefficient of association *Q* (Yule and Kendall, 1966), using formula $Q = (OR - 1)/(OR + 1)$. This coefficient is interpreted similarly to interpretation of the coefficient of correlation. $Q = 0$ means lack of association between methods, value close to 1 means high agreement. To have additional characterization of association, the probability *p* of concordance was also calculated according to the formula $p = (n_{11} + n_{22})/n$, where *n* denotes the total number of candidate varieties.

4. Results

The COYU method and the corrected Bennett's test (Shafer and Sulivan, 1986) were applied for three sets of generated data (data for candidate varieties). The data for reference varieties were taken from real experiments performed at the experimental station in Słupia Wielka. The COYU analysis was performed with the use of DUST package of Weatherup (1992). For Bennett's test the

EXCEL spreadsheet was utilized. The results for two years data concerning the period 2006–2007 are given in Table 1.

Table 1. Decisions on uniformity of candidate varieties (for data from the period 2006–2007)

Significance level		$\alpha = 0.002$		$\alpha = 0.02$			
Method		F test		Method		F test	
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%.

The results for the years 2007–2008 are presented in Table 2. 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 2. Decisions on uniformity of candidate varieties (for data from the period 2007–2008)

Significance level		$\alpha = 0.002$		$\alpha = 0.02$			
Method		F test		Method		F test	
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

The results for the three years period (2006–2008) are presented in Table 3. 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 3. Decisions on uniformity of candidate varieties (for data from the period 2006–2008)

Significance level		$\alpha = 0.002$		$\alpha = 0.02$			
Method		F test		Method		F test	
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

The results of the comparison of original Bennett's method (with use of $2Z$ statistic according to (3.2)) with the F test (the method that uses F statistic according to (3.4)) are presented in the Tables 4, 5 and 6.

Table 4. 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	
F test	decision	Uniform	not uniform	F test	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 (Table 4).

The results for the 2007–2008 period are presented in Table 5. 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 5. 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	
F test	decision	uniform	not uniform	F test	decision	uniform	not uniform
	uniform	212	0		uniform	181	3
	not uniform	60	0		not uniform	73	15

The results for the three years period (2006–2008) are presented in Table 6. When testing was performed at $\alpha = 0.002$ level, the probability of concordance was $p = 76.05\%$. When testing was performed at $\alpha = 0.02$ 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 6. 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	
F test	decision	uniform	not uniform	F test	decision	uniform	not uniform
	uniform	162	0		uniform	138	3
	not uniform	57	19		not uniform	52	45

5. Conclusions

1) The Bennett's approach with replacement $2Z$ statistic by the F statistics used for testing uniformity of candidate varieties is more restrictive (less varieties accepted as uniform) than COYU;
2) The Bennett's method with F used for testing uniformity of candidate varieties is more restrictive than original Bennett's method;
3) Both versions of Bennett's tests (with and without F) reject usually varieties with small mean values and large standard deviations;
4) The (rather sophisticated) COYU method can be replaced by much simpler Bennett's test without serious changes in decisions concerning uniformity of candidate varieties.

References

Bennett B.M. (1976). On an approximate test for homogeneity of coefficients of variation, in: *Contributions to applied statistics* (ed. W.I. Ziegler). Birkhäuser Verlag, 169–171.

Forkman J. (2009). Estimator and Tests for Common Coefficients of Variation in Normal Distributions. *Communications in Statistics Theory and Methods* 38, 233–251.

Iglewicz B., Meyers R. H. (1970). Comparison of approximations of the percentage points of the sample coefficient of variation. *Technometrics* 12, 166–169.

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, 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., Kendall M. G. (1966). *Wstęp do teorii statystyki*. PWN.

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, 170–176.

Zawieja B., Pilarczyk W., Kowalczyk B. (2010). Comparison of uniformity decisions based on COYU and Bennett's methods – simulated data. *Colloquium Biometricum* 40, 53–61.