Bulgarian Journal of Agricultural Science, 20 (No 5) 2014, 1040-1045
Agricultural Academy

DISTINCTNESS, UNIFORMITY AND STABILITY TRIALS WITH NEW VARIETY OF PLANT (OILSEED RAPE), FINDING OUTLIERS

B. ZAWIEJA
Department of Mathematical and Statistical Methods, Poznan University of Life Sciences, 60-637 Poznan, Poland

Abstract

ZAWIEJA, B., 2014. Distinctness, uniformity and stability trials with new variety of plant (oilseed rape), finding outliers. *Bulg. J. Agric. Sci.,* 20: 1040-1045

Each new variety reported by breeders to registration, has to be tested with regards to its distinctness, uniformity and stability (DUS). It can only be registered if it differs from other existing varieties in at least one characteristic. Unfortunately if the samples include outliers, then often the wrong decision concerning population is taken. Thus such observations should be eliminated from samples. In this study, a few methods for detecting outliers were compared based on the trials conducted in the Research Center for Cultivar Testing in Słupia Wielka in Poland with new varieties oilseed rape. The comparison of statistical method was executed for the one of aspect of DUS (distinctness, uniformity and stability) tests, namely for distinctiveness of the new varieties. As a criterion of choice the best method the numbers of distinguished pairs of varieties was used. The largest number of varietals pairs was distinguished after application the M-estimate method, thus this method was declared as the best.

Key words: Biweight Tukey's method; DUS trials; Dixon test; ESD test; M-estimate; Oilseed rape, outliers

Introduction

The breeding new line of each species is a very long and costly. Moreover before it will be introduced on market one can be registered and placed in the national list of varieties. The new varieties are tested in respect of fulfilled established conditions relate to distinctness, uniformity and stability (DUS). Distinctness means that new variety must be distinguishable from every another variety (know and new) for at least one characteristic. Degree of uniformity of new variety must be not worse that uniformity of all varieties used for comparison for all considered (observed) characteristics. Stability means that the plants of given variety not change, as regards researched characteristics, from generation to generation. Decisions are usually taken after two or three years of trialling. In one trial, apart new varieties, all known variety, which are similar to new, are tested. There are even a few hundreds varieties in one experiment. It should be added that the statistical analyses are used only then if decision can not be taken on basis visual observation. In order to conduct these analyzes, the samples have to be drawn at random (for each variety). Unfortunately, often among experimental data, can be outliers. The occurrence these observations sometimes due completely change the statistical inference and it can lead to the wrong decisions about population, so they are eliminated from the sample in the DUS testing. You can mention a few reasons for such observations. For example: observer's error (measurement or recording), called the gross error, the result of abnormal data in the sample (e.g. varieties very different from the others in terms of observed characteristics). Such data is called outliers or contaminated data. If mismatched observations are gross errors than elimination is good procedure but if there are so called contaminated than elimination this observation can results to the loss of important information about the tested population. Thus it is very important to correctly designate outliers.

In this paper, several tests for detecting outliers were compared for one of aspects of DUS test namely for distinctness. For this purpose the results of trial for winter oilseed rape from 2007 was used. Because distinctness is assessed on the basis multiple comparisons among pairs, the analysis of variance was performed before and after eliminating outliers. Next received results were compared with results received

E-mail: bogna13@up.poznan.pl

from the method recommended by UPOV (International Union for the Protection of New Varieties of Plant).

Material and Methods

The data from DUS trial for oilseed rape varieties, conducted in 2007 at the Research Center for Cultivar Testing in Słupia Wielka in Poland, were used for the analysis. Five measurable characteristics were observed during this experiment: PH– Plant: height (at full flowering), SL– Siliqua: length (between peduncle and beak), SLB – Siliqua: length of beak, SLP – Siliqua: length of peduncle, SW– Siliqua: width. The number of characteristics is consistent with the marking characteristics used in UPOV. The experiment was conducted in a randomized complete block design with two replicates. The 230 varieties were sown for this trial, known as well as candidate one and the sample size was 30 randomized plants from each plot.

In the case assessment of distinctness the following mathematical model of observations is used

$$y_{ijk} = \mu + \tau_i + \beta_j + \eta_{ij} + e_{ijk}, \quad i = 1,2,...,t, \; j = 1,2,...,b, \; k = 1,2,...,m, \tag{1}$$

where y_{ijk} means the measurement related with i-th variety and executed on k-th plant in j-th block, $\boldsymbol{\mu}$ is the grand effect, $\boldsymbol{\tau}_i$ is fixed effect of i-th variety, $\boldsymbol{\beta}_j$ means random effect of j-th block, $\boldsymbol{\eta}_{ij}$ is experimental random error (plots), e_{ijk} means random error of measurement and b – number of blocks (replications), t – number of variety, m – number of measurements on one plot. If on the basis of conducted analysis of variances, significant differences between varieties are showed, then multiple comparison tests (Fisher's or Tukey's) are used to examine which treatments are different. In this paper Tukey's test was used.

The following methods were used to search for outliers:

1) The Dixon method (1950). This type of test is based on the ratio of ranges. Because ranges are ordered statistics, there is no need to assume normality of data. Depending on the number of suspected outliers, different ratios are used to identify potential outliers. Namely in the first the test statistic are as follows

$$D_6 = \max\left(\frac{y_{(2)} - y_{(1)}}{y_{(n)} - y_{(1)}}; \frac{y_{(n)} - y_{(n-1)}}{y_{(n)} - y_{(1)}}\right) \tag{2}$$

and the second when the first or last observation were omitted respectively

$$D_7 = \max\left(\frac{y_{(2)} - y_{(1)}}{y_{(n-1)} - y_{(1)}}; \frac{y_{(n)} - y_{(n-1)}}{y_{(n)} - y_{(2)}}\right), \tag{3}$$

where $y_{(1)} \le y_{(2)} \le .... \le y_{(n)}$ are the ordered observations of one variety and plot (in the presented example $n = m$). The critical values these statistics are in the statistical tables for example Zieliński (1990). It is known that method (3) is less susceptible to masking effects (Penny and Jolliffe, 2001) than method (2).

2) The ESD method – Extreme Studentized Deviate (McMilan, 1971; Stefansky, 1972; Tietjen et al., 1973; Prescott, 1975). The test statistic in this method is calculated as

$$B_{ESD} = \max(B_n; B_1), \tag{4}$$

where $B_n = \frac{y_{(n)} - \bar{y}}{S}$, $B_1 = \frac{\bar{y} - y_{(1)}}{S}$, $\bar{y}$ is the mean and S is the standard deviation from the sample. Test (4) can also be used in other forms, where the numerator and denominator of the statistic (4) is calculated based on independent samples. The critical values for this test are in the appropriate statistical tables. Moreover for $n > 50$ and $\alpha < 0.2$ the critical values can be calculated using the following formula

$$b(\alpha, n) = z\sqrt{\frac{2(n-1)}{2n - 5 + z^2 + (3 + z^2 + 2z^4)[(6(2n-5))]^{-1}}},$$

where z is quantile of the rank $1 - \frac{\alpha}{2n}$ of normal distribution $N(0,1)$. It should be noted that the masking effect, which appears in this method, is higher than in the Dixon method.

3) M-estimate. This method was proposed by Huber (1973) and is a robust least squares estimator. It can be obtain by minimizing the function

$$Q(\tilde{\theta}) = \sum_{l=1}^{N} \rho\left(\frac{r_l}{\sigma}\right), \tag{5}$$

where $\mathbf{r} = \mathbf{y} - E(\mathbf{y}) = \mathbf{y} - \mathbf{X}\theta$ is the vector of the rest received by the least squares method in the following model

$$\mathbf{y} = \mathbf{X\theta} + \mathbf{r},$$

where $\mathbf{y}$ is the vector of observations, X is the matrix of design, $\boldsymbol{\theta}$ is the vector of unknown parameters, $\rho(s)$ is the function connected with distribution (where $s = \frac{r_l}{\sigma}$ is a response variable under consideration), σ is a common variance. The $\rho(s)$ is selected in this way to the received estimator, form minimization (5), rises less rapidly than traditional estimators of least squares method (LSE) and the maximum likelihood method (MLE). If σ is known and ρ is differentiable, then the minimization of function (5) is executed by solving the set of equations $\sum \psi(s) = 0$, where $\psi(s) = \rho'(s)$. The M-estimator has asymptotic normal distribution for the large sample (Maronna et al., 2006, p. 45).

In the theory of M-estimator the weight function $w(s,c)=\psi(s)/s$ is defined. These functions can take different form. The weight function named Tukey's biweight (Rocke, 1983) was in the following

$$w(s,c)=\begin{cases}\left(1-\left(\frac{s}{c}\right)^2\right)^2 & \text{if } |s|<c\\ 0 & \text{otherwise}\end{cases},$$

then

$$\rho(s,c)=\begin{cases}1-\left(1-\left(\frac{s}{c}\right)^2\right)^3 & \text{if } |s|<c\\ 1 & \text{otherwise}\end{cases}.$$

In this paper the c equal 4.68 (in the purpose of getting 0.95 of asymptotic efficiency, Maronna et al., 2006) was used to estimate location parameter and $c=1$ to estimating scale parameter (then the solution σ estimator is approximately consistent - Hampel at al., 1986). It is easy to see that Tukey weight function $w(s,c)$ takes the value zero for $|s|\geq c$ which in reality means the exclusion of these observations from the analysis.

In model (1) $E(y_{ijk})=\mu+\tau_i$ and then $r_{ijk}=y_i-E(y_{ijk})$.

Thanks to the above estimator, there is no need to remove outliers form the sample because, as was mentioned before, it is robust. Nevertheless, the M-estimator may be used to find outliers and then eliminate them in the sample for this research.

Finally it should be mentioned the method used in UPOV for the DUS tests. In this method outliers are removed by using the tolerance interval (Bobrowski, 1986). Firstly the differences between two extreme observations on each side of an ordered sample are calculated. Next if $y_{(n)}-y_{(n-1)}<y_{(2)}-y_{(1)}$ then the largest observation is abandoned otherwise smallest. For a reduced sample the parameters $\bar{y}$ (mean) and s (standard deviation) are calculated and the tolerance interval is determined as follows

$$(\bar{y}-t_{a/2,n-2}s\sqrt{n(n-2)};\bar{y}+t_{a/2,n-2}s\sqrt{n(n-2)})$$

where $t_{a/2,n-2}$ is critical value of the t distribution with $n-1$ degree of freedom and α significance level. If the remover observation does not belong to the designated interval then it is considered to be outliers.

In case of all method detecting of outliers, greater or lesser extent, appears the masking and swapping effects. Masking means that too little outliers is detected and swapping means that too much outliers is detected.

Results

The number of observations declared as outliers after using the methods described above (jointly for all examined varieties) are presented in Table 1. Since the number of detected outliers after application Bobrowski tests, at the significance level 0.01 (used in DUS tests), was much larger then the number such observations detected after application other methods, at the same level of significance, thus for these methods significance level 0.05 was used. It is easy to see that these numbers are different than the result of masking and swapping effects, which to varying degrees are used methods.

The existence of these effects was shortly presented based on the measurements SLP characteristics, for four selected varieties, namely Bristol (second replication), Gara (first replication), Bazyl (second replication), and Dogger (second replication) for a 0.05 level of significance. Observations of the considered characteristic for the chosen varieties were shown in Figure 1.

In the purpose of better presentation of selected varieties the descriptive statistics for these varieties was given in Table 2.

In the case of the Bristol variety it can be noticed that two extreme observations (on the right and left) differ form the other by about 0.2 cm, whereas the distances between the remaining observations are smaller or equal to 0.1 cm. When the DUS method was used the observation, which had a value of 2.5 cm (marked as a triangle in the figure) was recognized as an outlier. In turn methods (2), (3) and (4) did not mark this observation as an outlier. The masking effect on observation 2.5 cm by observation 1.5cm (marked as rhomb) appeared. Using the M-estimator, two observations 2.3 cm were executed (circled with discontinuous line in the figure) turned out to be outliers as well as observation 2.5 cm.

Table 1
Numbers of detected outliers after application of different tests sequentially, the M-estimator and procedures used in the DUS test

Characteristics	Significance level	D_6	D_7	B_{ESD}	*M*-estimator	DUS test
PH - Plant: height	0.05	30	41	27	24	32
SL- Siliqua: length	0.05	45	48	30	9	32
SLB - Siliqua: length of beak	0.05	138	142	45	44	61
SLP - Siliqua: length of peduncle	0.05	93	113	60	222	76
SW- Siliqua: width	0.05	50	68	42	108	73

Table 2
Descriptive statistics for chose variety and 19 characteristic

Variety	Means	Median	Standard deviation	minimum	maximum
Bristol	1.86	1.80	0.22	1.5	2.5
Gara	2.33	2.20	0.44	1.7	3.2
Bazyl	2.22	2.20	0.25	1.2	2.5
Dogger	2.32	2.30	0.34	1.9	2.7

Table 3
ANOVA of observed characteristics. SLP - Siliqua: length of peduncle

Method	Souce	d.f.	SS	MS	F	pr > F
	varieties	229	1147.4249	5.0106	22.73	<.0001
	blocks	1	3.3301	3.3301	15.11	0.0001
All observations	error of plots	197	43.42243	0.2204		
	error	12412	440.2953	0.0355		
	total	12839	1632.3502			
	varieties	229	1146.6499	5.0072	23.51	<.0001
	blocks	1	4.0472	4.0472	19.00	<.0001
ESD	error of plots	197	41.9539	0.2130		
	error	12352	371.5486	0.0301		
	total	12779	1561.5073			
	varieties	229	1139.7140	4.9769	29.26	<.0001
	blocks	1	3.8518	3.8518	22.65	0.0001
M-Estimator	error of plots	197	33.5064	0.1701		
	error	12190	311.7941	0.0256		
	total	12617	1488.1859			
	varieties	229	1143.070661	4.991575	23.80	<.0001
	blocks	1	3.548430	3.548430	16.92	<.0001
Bobrowski	error of plots	197	41.309055	0.209691		
	error	12339	379.425429	0.030750		
	total	12766	1565.526868			

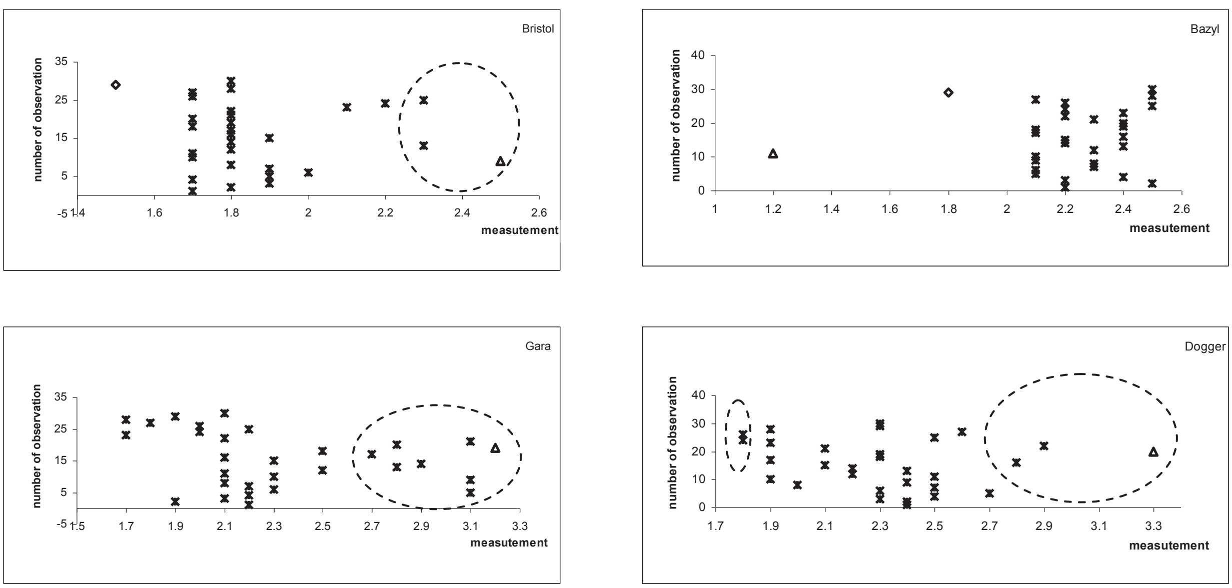

Fig. 1. Observations of considered characteristic for chosen varieties.
Abscissa axis – value to the characteristic, ordinate – the number of observation

Among the measurements of the variety Gara observation 3.2 (marked as a triangle) was detected as an outlier by the method used in the DUS test. However, a swapping effect occurred, caused by the three observations 3.1 cm in this method. The other methods did not accept this observation as an outlier. The M-estimator indicated 8 outliers (marked on the diagram with broken line).

There was one outlier (1.2 cm) for the Bazyl variety sample by all used methods; moreover both versions of the Dixon method detected one more outlier 1.8 cm. This could be the result of the swapping effect occurrence in these methods (by seven observations 2.1 cm). There was only one outlier detected using the M-estimator. Almost all of the methods, except the ESD method, marked observation 3.3 as an outlier for the Dogger. Data analysis allows us to conclude that there can be a swapping effect in these methods; this means that there were no outliers in this sample. The M-estimator shows the existence of 5 outliers (marked with a broken line).

Because the number of detected outliers after applying methods (2), (3) and (4) is most often quite similar, thus in Tables only the results of the method (4) were presented. Moreover analyzes of variance received for two chosen characteristics were shown. There were characteristic SPL in the Table 3 and SW in the Table 4. Since the data was not orthogonal the III type of the sum of squared was used in this paper (some of the varieties were sown in one replication).

It is easy to see the calculated F statistics has changed a bit when the outliers were removed. In the case of characteristic SPL the *p*-value was always less then 0.0001, but in the case of characteristic SW rejection of outliers resulted in a decrease the *p*-value.

In order to compare methods for searching of outliers, in Table 5 the results of multiple comparisons of varieties (Tukey test was used with a significance level of 0.05) was presents. The number of distinguished pairs, like the sum of the squares changed. Namely, the smaller the sum of the squares for error was, the more varieties were distinguished (smaller differences are detected between means).

Removing only the most extreme observations allowed distinguishing a larger number of pairs of varieties. Applications of the M-estimate resulted, in some characteristics, rejecting a very large number of outliers (Table 1). This allowed obtaining a greater number of pairs of varieties that differed significantly from each other. However, for characteristic SL (Siliqua: length) the M-estimate detected a relatively small number of outliers, which resulted in a distinction between a smaller numbers of varieties than the ESD method.

Table 4
ANOVA of observed characteristics. SW - Siliqua: width

Method	Souce	d.f.	SS	MS	F	pr > F
	varieties	229	2425.0139	10.5896	1.41	0.0068
	blocks	1	2.7717	2.7717	0.37	0.5445
All observations	error of plots	197	1481.2588	7.5191		
	error	12412	76704.7104	6.1799		
	total	12839	80615.1808			
	varieties	229	1255.6709	5.4833	9.27	<.0001
	blocks	1	0.9502	0.9502	1.61	0.2064
ESD	error of plots	197	116.4813	0.5913		
	error	12370	988.7096	0.0799		
	total	12797	2361.81200			
	varieties	229	1222.4786	5.3383	11.25	<.0001
	blocks	1	0.7861	0.7861	1.66	0.1996
M-Estimator	error of plots	197	93.4740	0.4745		
	error	12304	932.4624	0.0758		
	total	12731	2248.0497			
	varieties	229	1281.8770	5.597718	9.87	<.0001
	blocks	1	0.9642	0.964173	1.7	0.1939
Bobrowski	error of plots	197	111.7734	0.567378		
	error	12336	1157.9100	0.093864		
	total	12763	2551.6090			

Table 5
Numbers pairs of varieties differing at significance level 0.05 (used Tukey's studentized multiple comparisons test)

	Method of detected of outliers			
Characteristic	All observations	ESD	M	Bobr
PH - Plant: height	22965	23016	23013	23024
SL - Siliqua: length	17217	17574	17513	17199
SLB - Siliqua: length of beak	19417	21694	21741	21633
SLP - Siliqua: length of peduncle	17479	18076	18664	17986
SW - Siliqua: width	239	13067	13240	12212

Discussion and Conclusion

Since in the literature this approach has not been found thus the discussion with results given in other authors was been omitted.

In the experiment under the consideration the number of observations for each variety was 60. It can be concluded that the distribution of the studied characteristics was asymptotically normal. However, detailed analysis showed that for SL characteristics distribution only 52 (from 230) varieties was consistent with the normal distribution (at a significance level 0.002). Distributions for characteristic SW, distribution of most varieties were compatible with normal (only for 12 varieties there was no agreement). Looking at the graphs in Figure 1 it can be seen that the estimator M smoothens "tails" which is consistent with the recommended use of this estimator for the heavy-tailed distributions. So the obvious conclusion is that, the appropriate method for removing outliers on the basis the distribution of a given characteristic should be chosen. Perhaps, another method should be used for each characteristic. It should be noted that in the DUS testing particular attention was paid to considering as many reported varieties as possible separately (even the smallest differences among varieties should be detected). The reason is that developing a new variety is a long (often 10 years) and very expensive process.

As has already been mentioned, removing outliers is necessary in DUS testing because there presence in samples may have an effect on the inference from ANOVA and multiple comparisons of varieties. It would consider whether it would make sense to apply different outlier detection methods for different traits (for example for characteristic SL a number of different pairs is greater in the ESD method than M-estimator). The application of the M-estimator method, for the removal of outliers from the sample, usually can distinguish the largest number of pairs of varieties. The M-estimator, as indicated above, should be used for characteristics which have the heavy tailed distribution. It would be good to expand robust methods for a mixed model, which is used in varietal experiments. These methods would facilitate the analyzing of experimental results. In further studies, we will use robust methods in DUS experiments.

References

Barnett, V. and T. Lewis, 1994. Outliers in Statistical Data. *John Wiley & Sons*, New York.

Bobrowski, D., 1986. Probabilistic in technical applications. (in polish: Probabilistyka w zastosowaniach technicznych), *PWN* Warszawa (Pl).

Dixon, W. J., 1950. Analysis of extreme values. *Ann. Math. Stat.*, **21** (4): 488-506.

Hampel, F. R., E. M. Ranchetti, P. J. Rousseeuw and W. A. Stahel, 1986. Robust statistics: The approach based on influence functions. *John Wiley & Sons*. New York.

Huber, P. J., 1973. Robust regression: asymptotic, conjectures, and Monte Carlo. *Annals of Statistics,* **1**: 799-821.

Maronna, R. A., D. R. Matrin and V. J. Yohai, 2006. Roust statistics: Theory and Methods. *John Wiley & Sons*, New York.

Mcmilan, R. G., 1971. Tests for one or two outliers in normal samples with unknown variance. *Technometrics,* **13** (1): 87-100.

Penny, K. I. and I. T. Jolliffe, 2001. A comparison of multivariate outlier detection methods for clinical laboratory safety data. *The Statistician,* **50** (3): 295-308.

Prescott, P., 1975. An appropriate test for outliers in linear model. *Technometrics,* **17**: 129-132.

Rocke, D., 1983. Robust statistical analysis of inter laboratory studies. *Biometrika,* **70** (2): 421-431.

Rosner, B., 1983. Percentage point for generalized ESD many-outliers procedure. *Technometrics,* **25** (2): 165-172.

Stefansky, W., 1972. Rejecting outliers in factorial designs. *Technometrics,* **14**: 469-479.

Tietjen, G. L, R. H Moore and R. J. Backman, 1973. Testing for a single outlier in a simple linear regression. *Technometrics,* **15**: 717-721.

Zieliński, R., 1990. Statistical tables (in polish: Tablice statystyczne). *PWN*. Warszawa (Pl).

Received January, 29, 2014; accepted for printing July, 2, 2014.