TESTING OF SITE-SPECIFIC YIELD IN DIFFERENT HARVEST PASSES

Vojislav Simonovic, Dragan Markovic, Ivana Markovic

This paper analyzes the differences in mass yield of moist rapeseed grain for six different passes of combine harvester along the plot. The Mann-Whitney U test and Z-statistic are used for analysis needs. For harvest pass 1 and pass 4 the Z-value is only 0,211, whereas the statistical significance level to confirm this finding is 0,8333, therefore there is not statistically significant difference between yield levels in mentioned passes. It is established that there are another three pairs of similar harvest passes, of which two are adjacent, while all other passes exhibit statistical difference with respect to the yield. The approximate value of the effect size *r* indicator is applied for all passes, where a statistically significant difference turned out to exist between them. The highest difference is noted between pass 2 and pass 3, the effect size *r* value amounting to 0,464, which can be considered a large effect size according to Cohen's criterion. This paper suggests extending the current routine implementation of yield analysis to an automated post-processing system.

Keywords: Mann-Whitney U test; median; pass; rapeseed

Testiranje lokacijske specifičnosti prinosa u različitim žetvenim prohodima

U ovom radu analizirana je razlika u masenom prinosu vlažnog zrna uljane repice za 6 različitih prohoda žitnog kombajna duž parcele. Za potrebe analize korišten je Mann-Whitney U test i Z statistike. Za prvi i četvrti prohod je Z vrijednost svega 0,211, a razina statističke značajnost kojim se ovo tvrdi 0,8333, te nema statistički značajne razlike između vrijednosti prinosa u ovim prohodima. Utvrđeno je da postoji još tri para sličnih prohoda, od kojih su dva susjedna, dok su svi ostali prohodi statistički različiti u odnosu na prinos. Približna vrijednost statističkog pokazatelja effect size *r* primijenjena je za sve prohode kod kojih se ispostavilo da se statistički značajno razlikuju. Najveća razlika uočena je između drugog i trećeg prohoda za koje vrijednost effect size *r* iznosi 0,464 što bi se po Cohenov kriteriju smatralo srednjim utjecajem. Ovaj rad sugerira proširenje trenutne rutine analize prinosa ka automatiziranoj naknadnoj analizi.

Ključne riječi: Mann-Whitney U test; medijana; prohod; uljana repica

1 Introduction

The main goal of rapeseed production is to achieve high seed yield per unit area, which will, with high oil content in the seed, achieve a high yield of oil. Hybrids or rapeseed varieties differ in yield seeds, the oil content in the seed, and the like. Seed yield is a complex trait that is determined by the components yield. The basic components of the yield of rapeseed are: the number of plants per unit area (m^2), number of pods by plant, number of seeds per pods and weight of 1000 seeds [1÷3]. Seed yield and oil content in the seed, except for certain genetic yield potential of hybrids or varieties, significantly affect the soil-climatic conditions of cultivation, cultural practices, as well as their mutual interaction [4, 5]. The weather conditions during the growing season have a large impact on yield, however, any deviation from the optimal conditions for the growth and development of rapeseed can contribute to reduction in seed yield and oil [6]. Density has the greatest effect on seed yield and yield components [2]. The goal of precision agriculture is to recognize and make use of the plot site specificity. In order to examine the differences in moist grain mass yield between the observed passes, especially those adjacent, the Mann-Whitney U test will be used, one of the most powerful nonparametric tests. This test is a nonparametric alternative to the t-test of independent samples and tests the null hypothesis that two independent samples are selected from the same population. So, the research question reads: Do the measured moist grain mass yields in two harvest passes differ, which would consequently prove the plot site specificity, or are they statistically similar, so they can be approximated by each other? Instead of comparing mean values of the two groups, as the t-test does, the Mann-Whitney U test compares their medians. It turns the obtained values of continuous variable into the ranks for both groups and thereafter it calculates if there are significant differences between the ranks of those groups. As the result is turned into the ranks, real distribution is not of importance. Yield monitoring in combine harvesters is a cornerstone of precision agriculture. It relies on measurement of the grain flow through the harvesting equipment. Typical mechanisms that have been implemented to monitor grain flow through a combine can be grouped into volumetric flow sensors, mass flow sensors, and indirect measurement devices. Among them, impact-type mass flow sensors are widely used in many state-of-the-art yield monitors [7]. They consist of an impact plate and a force transducer that converts the net time-averaged impact force into a voltage signal. This type of structure is so simple that impact-type sensors can be easily mounted on combine harvesters and risk of causing an obstruction of the normal threshing process, even when the sensors are damaged, is minimized [8,9]. The goal of this paper was to evaluate the differences between the yield of rapeseed neighboring passes during harvest.

2 Material and method

Rapeseed combine harvester Claas Lexion 450 used in this investigation was fitted with a header, 6m wide and a grain mass flow sensor positioned on the top of the clean grain auger. The sensor measures the impact force with which the grain expelled from the paddle elevator strikes against the impact plate, Fig. 1. Using this force, as well as known header width, speed of motion and grain

auger speed, the moist grain mass yield is calculated. The effect of combine vibrations was eliminated by previous sensor calibration. In general all the theoretical statements have to be experimentally verified. This paper presents the results of investigations carried out throughout the harvest of rapeseed (*Brasica napus*), hybrid PR46W15, on plot near Belgrade, Serbia. Note that harvest time, June 2014, was ideal for research activities of this concept, because yield level in experimental plot was above average and markedly non-uniform due to high precipitation amounts, exellent cultural practices during vegetative growth and especially denser assembly plants per m^2.
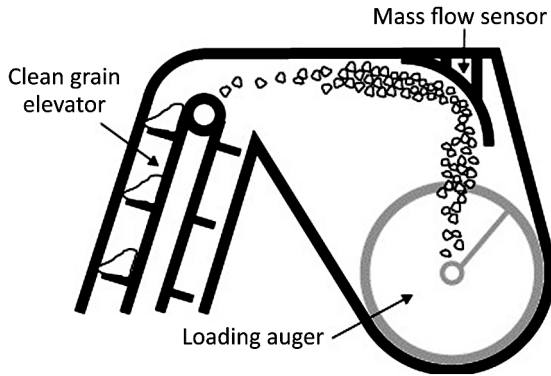
Figure 1 Shema of sensor position and picture of impact plate mass flow sensor with module behind of **g**ape between clean grain elevator and loading auger (view from left side of grain tank)

The system for measuring rapeseed yield is adjusted to consecutively register yield at 2-second intervals. This was a constant measuring time interval. The only parameter that changed was the distance travelled during that time, which was dependent on the combine harvester speed of motion. It was also registered for each 2-second time interval. During this interval, a number of grain contingents carried by the grain auger paddles were discarded and directed to the impact plate of the mass flow sensor.

Mass flow monitoring started 10 seconds after the adapter with a cutterbar was lowered for working position, and finished 10 seconds on lifting the cutterbar. Practically, there was a time shift for mass flow monitoring, consequently the yield, actually amounting to 10 seconds and representing transport time delay, i.e. the time needed for crop grain to travel through combine technological devices from the time moment of cutting to the time moment of grain striking against the impact plate of the mass flow sensor. Various factors such as combine separator design and settings and monitoring systems can affect the data gathering process so that the time shift should be adjusted. Without this adjustment, the grain flow and moisture values cannot be properly coordinated with location and area information to deliver data that accurately represent that location.

If the results for samples from two populations and two passes, respectively, are designated with X and Y, respectively, the null hypothesis is that X and Y have identical distributions, i.e.

$$H_0 : F_y(z) = F_x(z), \tag{1}$$

For each z, or

$$H_0 : P\{X > Y\} = \{Y > X\} = 0{,}5. \tag{2}$$

The testing method is based on combining the results for both groups and their ranking in an ascending sequence, saving the origin of observation. Let m be the number of observations in a smaller-size sample and n in a larger-size sample. The U-statistic, used for testing, is defined by the number of cases, so that the result for a larger-size sample comes before the result for a smaller-size sample in the process of ranking.

If observations from sample 1 and sample 2, respectively, are designated with $X_1, X_2, ..., X_m$ and $Y_1, Y_2, ..., Y_n$ and if the indicator of functions is introduced:

$$\begin{aligned} D_{ij} &= 1 \text{ if } Y_j < X_i, \text{ for } i = 1, 2, ..., m \\ D_{ij} &= 0 \text{ if } Y_j < X_i, \text{ for } j = 1, 2, ..., n \end{aligned} \tag{3}$$

U-statistic can be expressed as follows:

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} D_{ij}, \tag{4}$$

which takes the value from the set $u = \{0, 1, 2, ..., m+n\}$.

For small-size samples, i.e., when m and $n < 10$, a special table with significance levels for obtained U-statistic values is applied, whereas for large-size samples, like the one used here for moist grain mass yield, the approximation of U-statistic distribution by normal distribution [10], i.e., the statistic used, is:

$$Z = \frac{U - \dfrac{mn}{2}}{\sqrt{\dfrac{mn(m+n+1)}{12}}}. \tag{5}$$

For analysis, program procedures in the program package IBM SPSS Statistic 21 were used in the paper.

3   Results and discussion

Moist grain mass yields along each of the six passes are displayed in Fig. 2. Numbers are specified to point

towards the north, i.e., from down to up. A certain optimal level of moisture content for seeds and stalks was available to decreasing seed loss and damage during harvesting with combine harvester [11]. It was shown that

moisture content of seeds at 11 to 15 % was reasonable [5]. Our results indicated that moisture content of seeds was about 13 %, however the Mann-Whitney test was done with the moist grain yield.
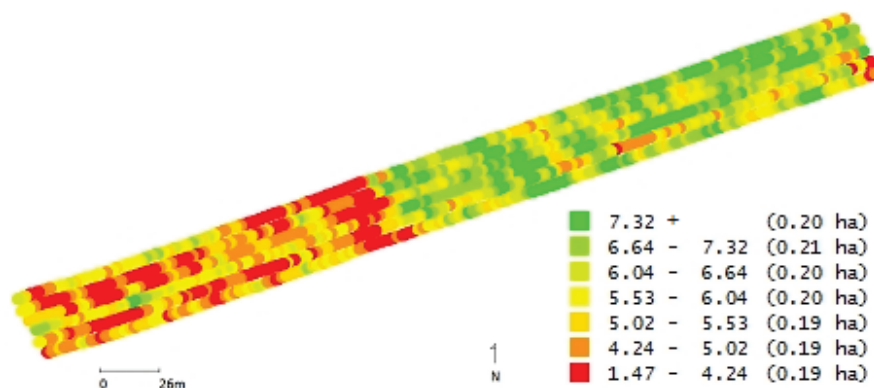

7.32 +	(0.20 ha)
6.64 − 7.32	(0.21 ha)
6.04 − 6.64	(0.20 ha)
5.53 − 6.04	(0.20 ha)
5.02 − 5.53	(0.19 ha)
4.24 − 5.02	(0.19 ha)
1.47 − 4.24	(0.19 ha)

0   26m

Figure 2 Representation of site-specific moist grain yield, tons per hectare, software SMS Advance by AG Leader Technology

Tab. 1 shows medians for all six analyzed passes, with more detailed analyses, i.e., comparisons to follow.

Table 1 Mean value, number of samples, standard deviation and median of moist grain mass yield for 6 different consecutive passes in harvesting rapeseed

Pass Num	Mean	N	Std. Deviation	Median
1	5,72717	163	1,094960	5,60000
2	5,18522	165	1,227784	5,20300
3	6,47254	154	1,310511	6,28600
4	5,67810	152	1,381609	5,56750
5	5,78701	150	1,567272	5,95750
6	5,97590	144	1,639878	6,16250

Apart from information about atypical points, the boxplot helps to visualize the form of mass yield distribution of different passes, Fig. 3. It indicates the variability of results within each pass and enables visual survey of differences between groups. This capability will be used to analyze yield variability per pass on the same plot.
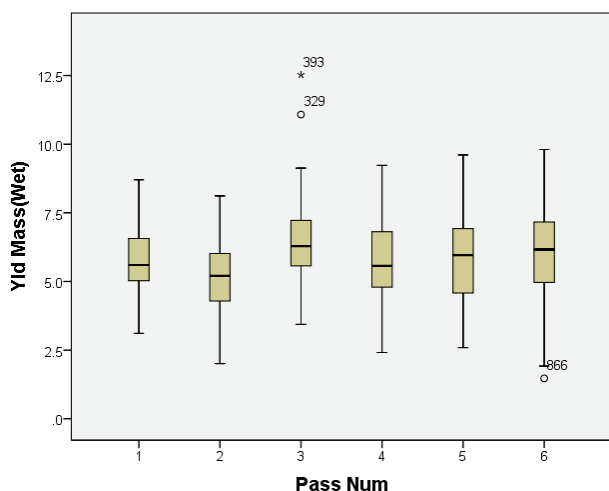

Figure 3 Boxplot for six-pass yields

To estimate central tendency or location of some phenomenon, arithmetic mean is most often used, however it is known for its extreme sensitivity to the presence of atypical or extreme values. The median, on

the other hand, is a measure of central tendency (site) very little sensitive to the presence of extreme values. This is why the median is employed for analysis, with mass yield distribution that is not normal, so this necessitates the use of nonparametric test where the median is just used. Yet, note that the median is a very restrictive measure, because the position mean value is involved. Namely, its value was determined based on a single or a small number of values for mass yield, which means loss of information.

Results obtained for the analysis of pass 1 and pass 2 are given in Tab. 2 and Tab. 3.

Table 2 Ranks for mass yield in passes 1 and 2

	Pass No.	N	Mean Rank	Sum of Ranks
Yield Mass (Wet)	Pass 1	163	184,47	30 068,00
	Pass 2	165	144,78	23 888,00
	Total	328		

Table 3 Results of Mann-Whitney U test for mass yield in passes 1 and 2

Grouping Variable: Pass No.	Yield Mass (Wet)
Mann-Whitney U	10 193,000
Z	−3,790
Asymp. Sig. (2-tailed)	0,000

For the case of comparing pass 1 and pass 2, Z-statistic equals −3,790 with significance level $p = 0,000$, which leads to the conclusion that the null hypothesis is rejected, i.e., there is a significant difference in average mass yield of moist grain for pass 1 and pass 2, Tab. 3. This fact, on one hand, confirms site-specificity of the plot itself on its observed part and, on the other hand, suggests impossibility of approximating yield registered in one pass by another pass. As for average rank values, Tab. 2 shows that average rank for pass 1 is 184,47 and for pass 2 it is 144,78. This difference indicates the direction of moist grain mass yield level. Since in calculating the rank, the lowest value was assigned 1, it is clear that yield values for pass 1, on average, received a higher rank.

Mutual values of Z_{ij} for all six analyzed passes are shown in Tab. 4. Values of statistical significance are

given in parentheses. It is evident from the Table that among adjacent passes there is no statistically significant difference between passes 4 and 5 as well as 5 and 6, respectively. As for other passes, passes 1 and 4 as well as 1 and 5, respectively, are similar for the results of moist grain mass yield. All this is supported by the small Z values and values of statistical significance >0,05.

Table 4 Values of Z_{ij} and statistical significance

Z_{ij} (p)	1	2	3	4	5	6
1		−3,790 (,000)	−5,064 (,000)	−0,211 (,833)	−0,606 (,544)	−2,171 (,030)
2	−3,790 (,000)		−7,967 (,000)	−3,167 (,002)	−3,665 (,000)	−4,935 (,000)
3	−5,064 (,000)	−7,967 (,000)		−4,668 (,000)	−3,428 (,001)	−2,028 (,043)
4	−0,211 (,833)	−3,167 (,002)	−4,668 (,000)		−0,733 (,464)	−2,041 (,041)
5	−0,606 (,544)	−3,665 (,000)	−3,428 (,001)	−0,733 (,464)		−1,088 (,276)
6	−2,171 (,030)	−4,935 (,000)	−2,028 (,043)	−2,041 (,041)	−1,088 (,276)	

The values of Z given in the results are not used to calculate the effect size statistical indicators, but it is possible to calculate approximate value of the effect size r:

$$r = \frac{z}{\sqrt{N}},$$

where N is the total number of cases (observations) that for the case of measuring moist grain mass yield occurred at 2 seconds each during combine harvester travel across the plot and along the corresponding pass. In statistics, an effect size is a measure of the strength of the relationship between two variables in a statistical population, or a sample-based estimate of that quantity. An effect size calculated from data is a descriptive statistic that conveys the estimated magnitude of a relationship without making any statement about whether the apparent relationship in the data reflects a true relationship in the population. In that way, effect sizes complement inferential statistics such as p-values [12].

For the case of comparing yields in pass 1 and pass 2 ($z_{12} = -3,790$ amd $N = 328$), r_{12} amounts to 0,21. This would be considered a small effect size according to Cohen's criterion [13]. In his landmark book on statistical power, Cohen drew on his general impression of the range of effect sizes found in research in order to create examples of power analysis for detecting smaller and larger effects. For Cohen's effect size of 0,1 might be a "small" effect, around 0,3 a "medium" effect and 0,5 to infinity, a "large" effect. Ever since, these values have been widely cited as standards for assessing the magnitude of the effects found in intervention research despite Cohen's own cautions about their inappropriateness for such general use. Cohen was attempting, in an unsystematic way, to describe the distribution of effect sizes one might find if one piled up all the effect sizes on all the different outcome measures for all the different interventions [14].

Tab. 5 displays the cross-value of effect size r only for the passes with previously observed statistically significant difference, i.e., with the p value <0,05. The highest difference is noted between passes 2 and 3, because the value of effect size r equals 0,446. This value suggests the existence of large difference in the yield between these passes.

Table 5 Approximate value of effect size r_{ij} for passes with statistically significant difference in yield

r_{ij}	1	2	3	4	5	6
1		0,210	0,284			0,124
2	0,210		0,446	0,178	0,206	0,281
3	0,284	0,446		0,267	0,197	0,117
4		0,178	0,267			0,119
5		0,206	0,197			
6	0,124	0,281	0,117	0,119		

For the case of comparing yields between adjacent passes in rapeseed harvest the effect size r was relevant. In the study case, large difference in yield between adjacent combine passes 2 and 3 amounted to 1,29 t/h.

4 Conclusion

After the entire above presented analysis based on the Mann-Whitney U test principles, it is inferred in this paper that Z-statistic applied to the combine harvester passes in harvesting rapeseed can be used as a good indicator of soil site specificity for grain mass yield. The procedure and principles of the analysis of yield variability and soil site specificity applied to rapeseed can be applied to any crop. If this analysis establishes that the value of probability level (p) is < 0,05, it is then deduced that the results are statistically significant, i.e., there is a significant difference in mass yields between the observed passes. Thus, we have proved only site specificity of the plot, where passes were observed and analyzed.

Furthermore, the analysis of difference levels between passes can follow, and for this purpose the effect size r is applicable, as well as Cohen's criterion. For analyzed passes the difference in yields was mostly small, except for large difference, according to Cohen, observed in yields analysis for passes 2 and 3. If this analysis still establishes that probability level (p) is neither small nor equals 0,05, it is then deduced that results (Z) are not statistically significant, i.e., there is no significant difference in mass yields between the observed passes, therefore the observed and analyzed passes can be approximated to one another. This enables to complement the gap of missing data on mass yield in one part of the pass by the adjacent pass data. Missing data is possible in different cases when the operator stops the harvester and reverses it, when sensors are damaged or during discharging the grain tank. However, it should be examined which of the two adjacent passes approximates better the observed part, i.e., one of its parts with no available screening data. Among the observed passes, the most similar, in respect to moist grain mass yield, are passes 1 and 4, with Z-statistic value of −0,211 and high statistical significance of 0,833.

5   References

[1]   Özer, H.; Oral, E.; Doğrum Ü. Relationship between yield and yield components on currently improved spring rapeseed cultivars. // Turkish Journal of Agriculture and Forestry. 23(1999), pp. 603-607.

[2]   Diepenbrock, W. Yield analysis of winter oilseed rape (Brassica napus L.). // Field Crops Research. 67(2006), pp. 35-49. DOI: 10.1016/S0378-4290(00)00082-4

[3]   Pospišil, A.; Pospišil, M. Ratarstvo – praktikum. Zagreb, Sveučilište u Zagrebu, Agronomski fakultet, 2013.

[4]   Gunasekere, C. P.; Martin, L. D.; Siddiquw, K.H.M.; Walton, G.H. Genotype by environment interactions of Indian mustard (Brassica juncea L.) and canola (B. Napus L.) in Mediterranean-type enviroments: 1. Crop growth and seed yield. // European Journal of Agronomy. 25, 1(2006), pp. 1-12. DOI: 10.1016/j.eja.2005.08.002

[5]   Marjanović-Jeromela, A.; Nagl, N.; Gvozdanović-Varga, J.; Hristov, N.; Kondić-Špika, A.; Vasić, M.; Marinković, R. Genotype by environment interactions for seed yield per plant in rapessed using AMMI model. // Pesquisa Agropecuria Brasileira. 46, 2(2011), pp.174-181.

[6]   Pospišil, M.; Pospišil, A.; Butorac, J.; Gunjača, J; Brčić, M. Utjecaj roka sjetve na prinos uljane repice. // Zbornik radova 45. Hrvatski i 5. Medjunarodni simpozij agronoma. Poljoprivredni fakultet Sveučilišta Josipa Jurja Strossmayera u Osijeku, 15-10. veljače, Opatija, 2010, pp-888-891.

[7]   Ryan, R.; Harry, D.; Jim, P.; Wonmo, K. A dynamic grain flow model for a mass flow yield sensor on a combine. // Precision Agriculture. 12, 5(2011), pp. 732-749. DOI: 10.1007/s11119-010-9215-0

[8]   Zhou, J.; Cong, B.; Liu, C. Elimination of vibration noise from an impact-type grain mass flow sensor. // Precision Agriculture, (2014). DOI 10.1007/s11119-014-9359-4

[9]   Simbahan, G. C.; Dobermann, A.; Ping, J. L. Screening yield monitor data improves grain yield maps. // Agronomy Journal, 96, 4(2004), pp. 1091-1102. DOI: 10.2134/agronj2004.1091

[10] Garson, G. D. Significance Testing: Parametric & Nonparametric. // Asheboro, NC: Statistical Associates Publishers, (2012), pp. 91-95.

[11] Mohammad, R. A.; Iraj, B.; Mir, H. P. Evaluation of a rice reaper used for rapeseed harvesting. //American-Eurasian Journal of Agricultural Environment Science, 2(2007), pp. 388-394.

[12] Ellis, P. D. The Essential Guide to Effect Sizes: An Introduction to Statistical Power. // Meta-Analysis and the Interpretation of Research Results. / United Kingdom: Cambridge University Press. 2010. DOI: 10.1017/CBO9780511761676

[13] Cohen, J.W. Statistical Power Analysis for the Behavioral Sciences (second ed.). // Hillsdale, Lawrence Erlbaum Associates, New Jersey, 1988.

[14] Lipsey, M. W.; Puzio, K.; Yun, C.; Hebert, M. A.; Steinka-Fry, K.; Cole, M.W.; Roberts, M.; Anthony, K.S.; Busick, M. D. Translating the Statistical Representation of the Effects of Education Interventions Into More Readily Interpretable Forms. // United States: U.S. Dept of Education, National Center for Special Education Research,Institute of Education Sciences, NCSER pp. 2013-3000. 2012.

Authors' addresses

Vojislav Simonovic, Research and Teaching Assistant
University of Belgrade
Faculty of Mechanical Engineering
Kraljice Marije 16
11120 Belgrade, Serbia
E-mail: vojislav@simonovic.rs

Dr Dragan Markovic, Full Professor
University of Belgrade
Faculty of Mechanical Engineering
Kraljice Marije 16
11120 Belgrade, Serbia
E-mail: dmarkovic@mas.bg.ac.rs

Ivana Markovic, Teaching Assistant
University of Belgrade
Faculty of Mechanical Engineering
Kraljice Marije 16
11120 Belgrade, Serbia
E-mail: imarkovic@mas.bg.ac.rs