doi: 10.5937/SelSem1902023B

# SIGNIFICANCE OF FIELD TRIALS DATA CLEANING PROCESS FOR MAKING MORE RELIABLE BREEDER DECISIONS

Milosav Babić[1]*, Petar Čanak[1], Bojana Vujošević[1], Vojka Babić[2], Dušan Stanisavljević[1]

## Abstract

Field trials supposed to enable selection of the most successful genotypes which is critical because of the existence of Genotype by Environment interaction. To assess this interaction we are forced to conduct field trials in several environments and/or years. When we asses grain yield of maize hybrids during the breeding process, it is always based on multi-environment small plot field trial (MESPT). That is why this part of the breeding process is most demanding in terms of technical, financial and labor requirements. In this paper, one possible systematic approach to assessing multi-environment field trials conduction is described. The main goal of the described approach is to provide the best possible results with the use of reasonable resources. As the results of trials cannot be directly interpreted without previous statistical processing, quality of raw data as input for biometrical (statistical) analysis is essential for obtaining a relevant and objective measure of genotype relative value in terms of productivity and adaptability (reliability) of new advanced maize hybrids. There are many definitions of data quality but data is generally considered high quality if it is fit for intended uses in operations, decision making and planning. The main aim of this paper is to underline the importance of the data cleaning process in MESPT.

***Key words:*** field inspection, maize breeding, plot scoring, proper results, raw data cleaning

## Introduction

Given that grain yield represents the most important and the most complex traits, modern maize breeding cannot be imagined without field trials. Here we assume that selected genotypes have no other critical weaknesses (stalk, tolerance to abiotic and biotic stresses). Assessing Multi Environment Small Plot Trial (MESPT) in terms of data quality and proper, as simple as possible, data analysis is of crucial importance for breeding program success. It is not complicated to understand the impact of inappropriate and inaccurate data, as misleading for final decisions after statistical processing (garbage in – garbage out). It is generally known that a phenotype is formed on the basis of the capacity of its genotype affected by environmental factors (Babic et al., 2011). From a breeding point of view, the most important thing for MESPT data point is to represent genotype performance as much as possible. Having it in mind, regular field inspection and plot representativeness scoring are very important, so non-representative plots can be excluded from data processing (i.e. considered as missing data that can be assessed using appropriate calculation).

Many statistical analyses try to find a pattern in a data series, based on a hypothesis or assumption about the nature of the data. 'Cleaning' is the process of removing those data points which are either (a) disconnected with the effect of assumption which we are trying to isolate, due to some other factor which applies only to those particular data points, (b)

Original Scientific Paper (Originalni naučni rad)

[1] Babić M, Čanak P, Vujošević B, Stanisavljević D, Institutte for Field and Vegetable Crops, Maksima Gorkog 30, Novi Sad, Serbia

[2] Babić V, Maize Research Institute „Zemun Polje", Slobodana Bajića 1, 11185 Belgrade, Serbia

*e-mail: milosav.babic@gmail.com

erroneous, i.e. some external error is reflected in that particular data point, either due to a mistake during data collection, reporting etc.

According to Sunil (2018), outliers can drastically change the results of the data analysis and statistical modelling. There are numerous unfavourable impacts of outliers in the data set:

It increases the error variance and reduces the power of statistical tests; If the outliers are non-randomly distributed, they can decrease normality; They can bias or influence estimates that may be of substantive interest; They can also impact the basic assumption of Regression, ANOVA and other statistical model assumptions.

If we accept a definition of data quality as a dataset that fits for intended uses in operations, decision making and planning, the main aim of this paper is to underline the importance of data cleaning process in MESPT.

## Material and methods

For illustration of data cleaning process a block of 15 trials planted according CRB design in two replication were used. Each trial contained 24 genotypes, including 4 check hybrids. Collecting of harvest data was done in 2018 from six locations. Box plot was used as main tool for detecting possible outliers. Separate box plot was created for each row and column. As outlier threshold range between Q1-IQR*2.2 and Q3+IQR*2.2 was used, values out of this range marked as suspected outliers (underperforming-red, over performing-green).

## Result and discussion

***Detecting outliers and data cleaning.*** The points to be cleaned are generally extreme outliers. 'Outliers' are those points which stand out for not following a pattern which is generally visible in the data. One way of detecting outliers is to plot the data points (if possible) and visually inspect the resultant plot for points which lie far outside the general distribution. Another way is to run the analysis on the entire dataset, and then eliminating those points which do not meet mathematical 'control limits' for variability from a trend, and then repeating the analysis on the remaining data (Vakili and Schmitt, 2014).

The importance of having clean and reliable data in any statistical analysis cannot be stressed enough. Often, in real-world applications, the analyst may get fascinated by the complexity or beauty of the statistical method being applied, while the data itself may be unreliable and lead to results which suggest courses of action without a solid basis. A good statistician/researcher (personal opinion) spends more than 75% of his/her time on collecting and cleaning data, and developing hypothesis which covers as many external explainable factors as possible, and only up to 25% on the actual mathematical manipulation of the data and deriving results.

There are seven key purposes data cleaning serve in delivering useful end-user data:

- Eliminate errors
- Eliminate redundancy
- Increase data reliability
- Deliver accuracy
- Ensure consistency
- Assure completeness
- Provide feedback for improvement

In order to provide systematic approach for decision making, first step is to standardize the process. Certain protocols have to be made for each step. It should enable more reliable decision making process as well as possibilities for process improvement. In terms of breeding process, necessary steps are: error monitoring, accuracy validation, analysis standardization and decision making. Each step should be covered by appropriate protocol to avoid mistakes that will influence final decision.

Error monitoring in the context of MESPT requires regular field visits at certain stages of growth and development to conduct observations, collection of relevant information and scoring results for each plot. There are supposed to be at least two field observations with plot scoring. One at initial stages of growth (4-6 leaves) and one at a late stage (end of kernel maturity, waxy stage), before harvesting. Usually, first scoring has five classes. The number of classes can be reduced to three (unacceptable, questionable, good plot). In order to make subsequent scoring easier and more accurate, it is better if we retain five classes for the first scoring:

– Unacceptable plot (to be excluded) - 1
– Most probably unacceptable plot - 2
– Questionable plot - 3
– Most probably good plot - 4
– Good plot - 5

Final plot scoring has only two classes:
– Unacceptable plot (to be excluded) - 1
– Good plot - 5

The system of plot quality classification can be adjusted (scoring classes, class numbering, number of scoring) as long as final plot scoring assumes only two classes. It is extremely important that once MESPT plot is classified as unacceptable "To be excluded", this classification should not be changed during subsequent plot scoring, regardless of how many plot scoring we have. Based on field observations most of data cleaning should be done once we have harvesting results.

The main aim of data cleaning (removing outliers) is to provide as reliable as possible data on hybrid performance. Influence of factors such as different densities, weed spots, standing water, mechanical damage of plants during different field operations, errors in the application of fertilizers, pesticides etc., on the particular plot should be avoided as much as possible (i.e. such influenced plot data should be excluded from further analysis). For complex databases, special software's are being developed depending on the type of data to be cleaned. Statistical outliers are data points that are far removed and numerically distant from the rest of the points. Outliers occur frequently in many statistical analyses and it is important to understand them and their occurrence in the right context of the study to be able to deal with them.

Accuracy validation usually means the application of some statistical tool that can indicate data inconsistency. Most frequently it underlies unusually low or unusually high values comparing to expectation. This does not mean automatic suppression of such data, but rather checking our notes from field observations, and based on it, eventual new data suppression. For the purpose of MET, standard deviation and/or box plot (interquartile range) are most frequently used as the univariate test. The interquartile

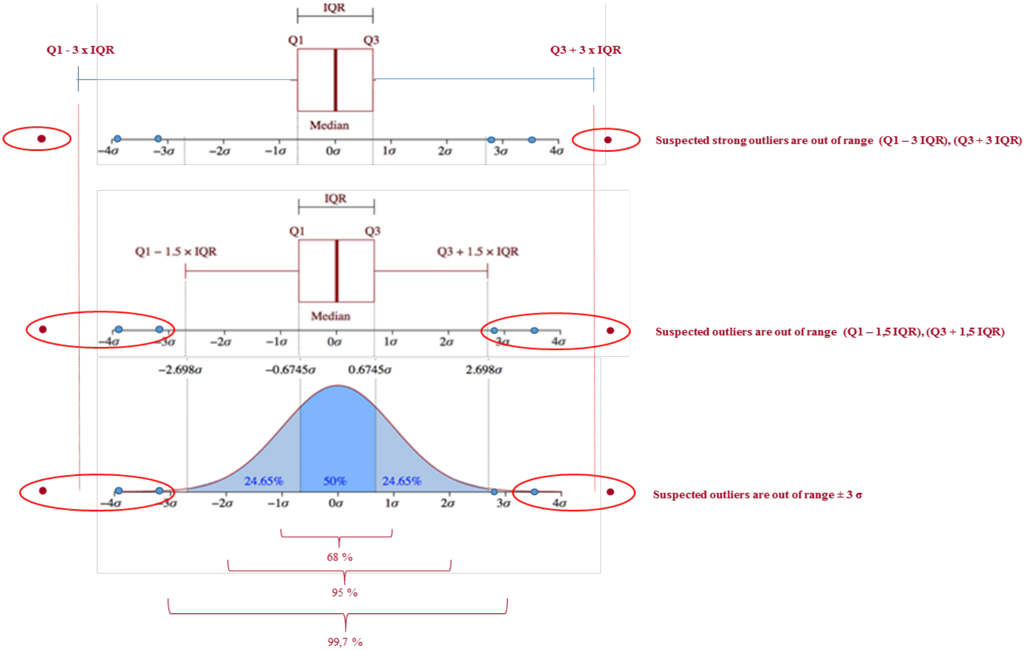


*Figure 1. Illustration of application of Standard deviation and Box plot in detecting possible outliers in a case of normal distribution*

*Grafikon 1. Ilustracija primene standardne devijacije i box-plota u detekciji ekstrema u slučaju normalne distribucije* range is probably more suitable as not all data is Gaussian distributed. Galarnyk (2019) gives a nice clarification for understanding Boxplot as a method of detecting possible outliers in the dataset.

As univariate tests for detecting possible outliers, standard deviation and box plot tests are most widely used. On graph 1, relationships of these tests are given in case of normal distribution. Reason for more frequent use of box plot is that this test is less demanding in terms of distribution. Most frequently as a lower and upper limit for detecting outliers are used Q1-IQR*1.5 (lower limit) and Q3+IQR*1.5 (upper limit). A detailed guide on how to create a boxplot in Excel is given at https://support.office.com/en-us/article/Create-a-box-plot-10204530-8cdf-40fe-a711-2eb9785e510f. In a case of using boxplot for detecting outliers, the lower and upper limit is being used instead of minimum and maximum, respectively. Depending on data and research goal lower limit can be set up between Q1-IQR*1.5 and Q1-IQR*3. At the same time the upper limit can be set up between Q3+IQR*1.5 and Q3+IQR*3. Figure 3 and 4 give suspected outliers out of range Q1-IQR*2.2 and Q3+IQR*2.2. Once created, boxplot tells about outliers and what their values are. It also tells if the dataset is symmetrical, how tightly data are grouped, and if and how data are skewed (Galarnyk, 2019).

A consequence of data cleaning is unbalanced data set. There are two ways of dealing with unbalanced data set, either to remove data points that are unbalanced or to apply one of many procedures recommended to calculate and replace the missing value. From the decision- making process, the second option is frequently applied, as from various reasons some MESPT data points have been lost and there is a strong need for included genotypes to be analyzed. (Yan et al., 2011). It is widely accepted that even up to 30% of data points can be replaced with calculated value without significant loss in quality (Woyann et al., 2017). There is no need to underline, the fewer data points replaced, the better. Usually, in practice, not more than 15% of missing or removed data is being replaced with calculated values. The most simple way to obtain a balanced data set is to replace missing data with the environment, row or column mean. One possible approach is filling the missing cells with values estimated from fitted multiplicative or mixed linear models (Arciegas-Alarcon et al., 2011; Kumat et al., 2012). Some researchers prefer to use either the mixed linear model based on the statistical method of Restricted Maximum Likelihood/Best Linear Unbiased Prediction (REML/BLUP) or Bayesian approaches. For a description of these methodologies see, e.g. (Fritsche-Neto et al., 2010.; Crossa et al., 2011.; Josse et al., 2014.; Omar et al., 2015).

Ray Sunil (2016) gives a comprehensive guide to data exploration that clarify strategies of detecting and dealing with outliers (missing values). He also mentions two types of outliers – Artificial (ERROR) and Natural (NOT ERROR). According to the author, there could be several reasons for artificial outlier:

***Data entry errors:*** Human errors such as errors caused during data collection, recording, or entry can cause outliers in data.

***Measurement error:*** It is the most common source of outliers. This is caused when the measurement instrument used turns out to be faulty. The weights measured on the faulty machine can lead to outliers.

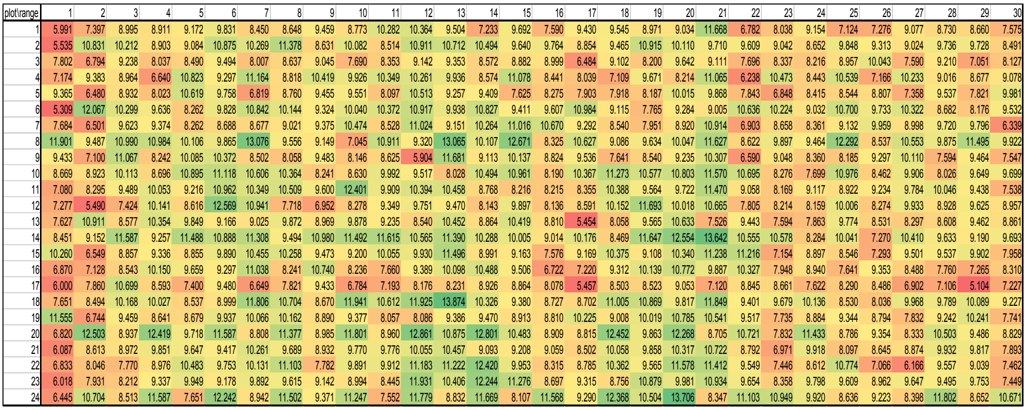
***Experimental error:*** Another cause of outliers is experimental error.

***Intentional outlier:*** This is commonly found in self-reported measures that involve sensitive data. For example, Teens would typically under - report the amount of alcohol that they consume. Only a fraction of them would report actual value. Here actual values might look like outliers because the rest of the teens are under - reporting the consumption.

***Data processing error:*** Whenever we perform data mining, we extract data from multiple sources. It is possible that some manipulation or extraction errors may lead to outliers in the dataset.

***Sampling error:*** Can occur if we include late hybrid in a trial that encompass early hybrids. It is very likely that value for this genotype will appear as an outlier.

Yan (2013) stated that the following strategies may overcome the lack of balance in the data, but none of them is simple and fully effec-

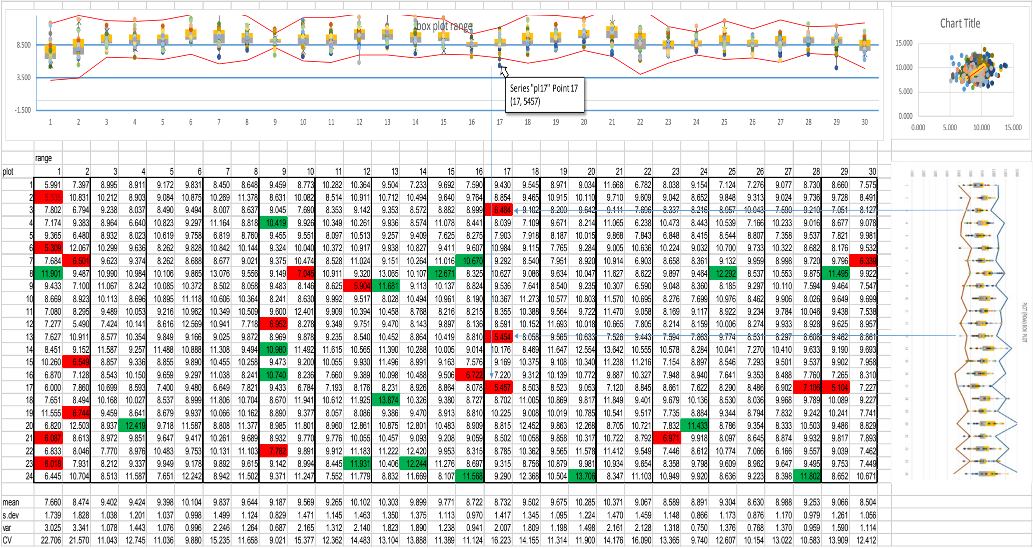


| plot\range | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 5.991 | 7.397 | 8.995 | 8.911 | 9.172 | 9.831 | 8.450 | 8.648 | 9.459 | 8.773 | 10.282 | 10.364 | 9.504 | 7.233 | 9.692 | 7.590 | 9.430 | 9.545 | 8.971 | 9.034 | 11.668 | 6.782 | 8.038 | 9.154 | 7.124 | 7.276 | 9.077 | 8.730 | 8.660 | 7.575 |
| 2 | 5.535 | 10.831 | 10.212 | 8.903 | 9.084 | 10.875 | 10.269 | 11.378 | 8.631 | 10.082 | 8.514 | 10.911 | 10.712 | 10.494 | 9.640 | 9.764 | 8.854 | 9.465 | 10.915 | 10.110 | 9.710 | 9.609 | 9.042 | 8.652 | 9.848 | 9.313 | 9.024 | 9.736 | 9.728 | 8.491 |
| 3 | 7.802 | 6.794 | 9.238 | 8.037 | 8.490 | 9.494 | 8.007 | 8.637 | 9.045 | 7.690 | 8.353 | 9.142 | 9.353 | 8.572 | 8.882 | 8.999 | 6.484 | 9.102 | 8.200 | 9.642 | 9.111 | 7.696 | 8.337 | 8.216 | 8.957 | 10.043 | 7.590 | 9.210 | 7.051 | 8.127 |
| 4 | 7.174 | 9.383 | 8.964 | 6.640 | 10.823 | 9.297 | 11.164 | 8.818 | 10.419 | 9.926 | 10.349 | 10.261 | 9.936 | 8.574 | 11.078 | 8.441 | 8.039 | 7.109 | 9.671 | 8.214 | 11.065 | 6.238 | 10.473 | 8.443 | 10.539 | 7.166 | 10.233 | 9.016 | 8.677 | 9.078 |
| 5 | 9.365 | 6.480 | 8.932 | 8.023 | 10.619 | 9.758 | 6.819 | 8.760 | 9.455 | 9.551 | 8.097 | 10.513 | 9.257 | 9.409 | 7.625 | 8.275 | 7.903 | 7.918 | 8.187 | 10.015 | 9.868 | 7.843 | 6.848 | 8.415 | 8.544 | 8.807 | 7.358 | 9.537 | 7.821 | 9.981 |
| 6 | 5.309 | 12.067 | 10.299 | 9.636 | 8.262 | 9.828 | 10.842 | 10.144 | 9.324 | 10.040 | 10.372 | 10.917 | 9.938 | 10.827 | 9.411 | 9.607 | 10.984 | 9.115 | 7.765 | 9.284 | 9.005 | 10.636 | 10.224 | 9.032 | 10.700 | 9.733 | 10.322 | 8.682 | 8.176 | 9.532 |
| 7 | 7.684 | 6.501 | 9.623 | 9.374 | 8.262 | 8.688 | 8.677 | 9.021 | 9.375 | 10.474 | 8.528 | 11.024 | 9.151 | 10.264 | 11.016 | 10.670 | 9.292 | 8.540 | 7.951 | 8.920 | 10.914 | 6.903 | 8.658 | 8.361 | 9.132 | 9.959 | 8.998 | 9.720 | 9.796 | 6.339 |
| 8 | 11.901 | 9.487 | 10.990 | 10.984 | 10.106 | 9.865 | 13.076 | 9.556 | 9.149 | 7.045 | 10.911 | 9.320 | 13.065 | 10.107 | 12.671 | 8.325 | 10.627 | 9.086 | 9.634 | 10.047 | 11.627 | 8.622 | 9.897 | 9.464 | 12.292 | 8.537 | 10.553 | 9.875 | 11.495 | 9.922 |
| 9 | 9.433 | 7.100 | 11.067 | 8.242 | 10.085 | 10.372 | 8.502 | 8.058 | 9.483 | 8.146 | 8.625 | 5.904 | 11.681 | 9.113 | 10.137 | 8.824 | 9.536 | 7.641 | 8.540 | 9.235 | 10.307 | 6.590 | 9.048 | 8.360 | 8.185 | 9.297 | 10.110 | 7.594 | 9.464 | 7.547 |
| 10 | 8.669 | 8.923 | 10.113 | 8.696 | 10.895 | 11.118 | 10.606 | 10.364 | 8.241 | 8.630 | 9.992 | 9.517 | 8.028 | 10.494 | 10.961 | 8.190 | 10.367 | 11.273 | 10.577 | 10.803 | 11.570 | 10.695 | 8.276 | 7.699 | 10.976 | 8.462 | 9.906 | 8.026 | 9.649 | 9.699 |
| 11 | 7.080 | 8.295 | 9.489 | 10.053 | 9.216 | 10.962 | 10.349 | 10.509 | 9.600 | 12.401 | 9.909 | 10.394 | 10.458 | 8.768 | 8.216 | 8.215 | 8.355 | 10.388 | 9.564 | 9.722 | 11.470 | 9.058 | 8.169 | 9.117 | 8.922 | 9.234 | 9.784 | 10.046 | 9.438 | 7.538 |
| 12 | 7.277 | 5.490 | 7.424 | 10.141 | 8.616 | 12.569 | 10.941 | 7.718 | 6.952 | 8.278 | 9.349 | 9.751 | 9.470 | 8.143 | 9.897 | 8.136 | 8.591 | 10.152 | 11.693 | 10.018 | 10.665 | 7.805 | 8.214 | 8.159 | 10.006 | 8.274 | 9.933 | 8.928 | 9.625 | 8.957 |
| 13 | 7.627 | 10.911 | 8.577 | 10.354 | 9.849 | 9.166 | 9.025 | 9.872 | 8.969 | 9.878 | 9.235 | 8.540 | 10.452 | 8.864 | 10.419 | 8.810 | 5.454 | 8.058 | 9.565 | 10.633 | 7.526 | 9.443 | 7.594 | 7.863 | 9.774 | 8.531 | 8.297 | 8.608 | 9.462 | 8.861 |
| 14 | 8.451 | 9.152 | 11.587 | 9.257 | 11.488 | 10.888 | 11.308 | 9.494 | 10.980 | 11.492 | 11.615 | 10.565 | 11.390 | 10.288 | 10.005 | 9.014 | 10.176 | 8.469 | 11.647 | 12.554 | 13.642 | 10.555 | 10.578 | 8.284 | 10.041 | 7.270 | 10.410 | 9.633 | 9.190 | 9.693 |
| 15 | 10.260 | 6.549 | 8.857 | 9.336 | 8.855 | 9.890 | 10.455 | 10.258 | 9.473 | 9.200 | 10.055 | 9.930 | 11.496 | 8.991 | 9.163 | 7.576 | 9.169 | 10.375 | 9.108 | 10.340 | 11.238 | 11.216 | 7.154 | 8.897 | 8.546 | 7.293 | 9.501 | 9.537 | 9.902 | 7.958 |
| 16 | 6.870 | 7.128 | 8.543 | 10.150 | 9.659 | 9.297 | 11.038 | 8.241 | 10.740 | 8.236 | 7.660 | 9.389 | 10.098 | 10.488 | 9.506 | 6.722 | 7.220 | 9.312 | 10.139 | 10.772 | 9.887 | 10.327 | 7.948 | 8.940 | 7.641 | 9.353 | 8.488 | 7.760 | 7.265 | 8.310 |
| 17 | 6.000 | 7.860 | 10.699 | 8.593 | 7.400 | 9.480 | 6.649 | 7.821 | 9.433 | 6.784 | 7.193 | 8.176 | 8.231 | 8.926 | 8.864 | 8.078 | 5.457 | 8.503 | 8.523 | 9.053 | 7.120 | 8.845 | 8.661 | 7.622 | 8.290 | 8.486 | 6.902 | 7.106 | 5.104 | 7.227 |
| 18 | 7.651 | 8.494 | 10.168 | 10.027 | 8.537 | 8.999 | 11.806 | 10.704 | 8.670 | 11.941 | 10.612 | 11.925 | 13.874 | 10.326 | 9.380 | 8.727 | 8.702 | 11.005 | 10.869 | 9.817 | 11.849 | 9.401 | 9.679 | 10.136 | 8.530 | 8.036 | 9.968 | 9.789 | 10.089 | 9.227 |
| 19 | 11.555 | 6.744 | 9.459 | 8.641 | 8.679 | 9.937 | 10.066 | 10.162 | 8.890 | 9.377 | 8.057 | 8.086 | 9.386 | 9.470 | 8.913 | 8.810 | 10.225 | 9.008 | 10.019 | 10.785 | 10.541 | 9.517 | 7.735 | 8.884 | 9.344 | 8.794 | 7.832 | 9.242 | 10.241 | 7.741 |
| 20 | 6.820 | 12.503 | 8.937 | 12.419 | 9.718 | 11.587 | 8.808 | 11.377 | 8.985 | 11.801 | 8.960 | 12.861 | 10.875 | 12.801 | 10.483 | 8.909 | 8.815 | 12.452 | 9.863 | 12.268 | 8.705 | 10.721 | 7.832 | 11.433 | 8.786 | 9.354 | 8.333 | 10.503 | 9.486 | 8.829 |
| 21 | 6.087 | 8.613 | 8.972 | 9.851 | 9.647 | 9.417 | 10.261 | 9.689 | 8.932 | 9.770 | 9.776 | 10.055 | 10.457 | 9.093 | 9.208 | 9.059 | 8.502 | 10.058 | 9.858 | 10.317 | 10.722 | 8.792 | 6.971 | 9.918 | 8.097 | 8.645 | 8.874 | 9.932 | 9.817 | 7.893 |
| 22 | 6.833 | 8.046 | 7.770 | 8.976 | 10.483 | 9.753 | 10.131 | 11.103 | 7.782 | 9.891 | 9.912 | 11.183 | 11.222 | 12.420 | 9.953 | 8.315 | 8.785 | 10.362 | 9.565 | 11.578 | 11.412 | 9.549 | 7.446 | 8.612 | 10.774 | 7.066 | 6.166 | 9.557 | 9.039 | 7.462 |
| 23 | 6.018 | 7.931 | 8.212 | 9.337 | 9.949 | 9.178 | 9.892 | 9.615 | 9.142 | 8.994 | 8.445 | 11.931 | 10.406 | 12.244 | 11.276 | 8.697 | 9.315 | 8.756 | 10.879 | 9.981 | 10.934 | 9.654 | 8.358 | 9.798 | 9.609 | 8.962 | 9.647 | 9.495 | 9.753 | 7.449 |
| 24 | 6.445 | 10.704 | 8.513 | 11.587 | 7.651 | 12.242 | 8.942 | 11.502 | 9.371 | 11.247 | 7.552 | 11.779 | 8.832 | 11.669 | 8.107 | 11.568 | 9.290 | 12.368 | 10.504 | 13.706 | 8.347 | 11.103 | 10.949 | 9.920 | 8.636 | 9.223 | 8.398 | 11.802 | 8.652 | 10.671 |

***Figure 2.** Illustration of yield data Heat map for one location of Multi environment small plot field trial*
***Grafikon 2.** Ilustracija Heat map tabele, višelokacijskog sortnog mikro ogleda*

tive. The first strategy (removing missing data points from the analysis) does not make use of all the available information; the second one (replacing missing value with environment/row/column mean) may have problems when too many values are missing, and the third one (estimation of missing values using statistical tools that enable the analysis of such data) involves multiple steps and complicated procedures. Arciniegas-Alarcón et al. (2016) give a wide range of literature regarding calculation of the missing data.

The first step in data analysis is to create a Heat map of our data in order to check for any systematic variation in our raw data. Table (Heat map) illustrated in Figure 2, give us no indication of such variation, even indicating wide range of grain yield recorded (From 5.309 to 13.874 t ha$^{-1}$). Next step supposed to exclude all plots scored as unacceptable during the scoring process, regardless of the reason. This usually works well for low values recorded, as it is based on environmental factors affecting some particular plot (low density, standing



| | range | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| plot | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| 1 | 5.991 | 7.397 | 8.995 | 8.911 | 9.172 | 9.831 | 8.450 | 8.648 | 9.459 | 8.773 | 10.282 | 10.364 | 9.504 | 7.233 | 9.692 | 7.590 | 9.430 | 9.545 | 8.971 | 9.034 | 11.668 | 6.782 | 8.038 | 9.154 | 7.124 | 7.276 | 9.077 | 8.730 | 8.660 | 7.575 |
| 2 | 5.535 | 10.831 | 10.212 | 8.903 | 9.084 | 10.875 | 10.269 | 11.378 | 8.631 | 10.082 | 8.514 | 10.911 | 10.712 | 10.494 | 9.640 | 9.764 | 8.854 | 9.465 | 10.915 | 10.110 | 9.710 | 9.609 | 9.042 | 8.652 | 9.848 | 9.313 | 9.024 | 9.736 | 9.728 | 8.491 |
| 3 | 7.802 | 6.794 | 9.238 | 8.037 | 8.490 | 9.494 | 8.007 | 8.637 | 9.045 | 7.690 | 8.353 | 9.142 | 9.353 | 8.572 | 8.882 | 8.999 | 6.484 | 9.102 | 8.200 | 9.642 | 9.111 | 7.696 | 8.337 | 8.216 | 8.957 | 10.043 | 7.590 | 9.210 | 7.051 | 8.127 |
| 4 | 7.174 | 9.383 | 8.964 | 6.640 | 10.823 | 9.297 | 11.164 | 8.818 | 10.419 | 9.926 | 10.349 | 10.261 | 9.936 | 8.574 | 11.078 | 8.441 | 8.039 | 7.109 | 9.671 | 8.214 | 11.065 | 6.238 | 10.473 | 8.443 | 10.539 | 7.166 | 10.233 | 9.016 | 8.677 | 9.078 |
| 5 | 9.365 | 6.480 | 8.932 | 8.023 | 10.619 | 9.758 | 6.819 | 8.760 | 9.455 | 9.551 | 8.097 | 10.513 | 9.257 | 9.409 | 7.625 | 8.275 | 7.903 | 7.918 | 8.187 | 10.015 | 9.868 | 7.843 | 6.848 | 8.415 | 8.544 | 8.807 | 7.358 | 9.537 | 7.821 | 9.981 |
| 6 | 5.309 | 12.067 | 10.299 | 9.636 | 8.262 | 9.828 | 10.842 | 10.144 | 9.324 | 10.040 | 10.372 | 10.917 | 9.938 | 10.827 | 9.411 | 9.607 | 10.984 | 9.115 | 7.765 | 9.284 | 9.005 | 10.636 | 10.224 | 9.032 | 10.700 | 9.733 | 10.322 | 8.682 | 8.176 | 9.532 |
| 7 | 7.684 | 6.501 | 9.623 | 9.374 | 8.262 | 8.688 | 8.677 | 9.021 | 9.375 | 10.474 | 8.528 | 11.024 | 9.151 | 10.264 | 11.016 | 10.670 | 9.292 | 8.540 | 7.951 | 8.920 | 10.914 | 6.903 | 8.658 | 8.361 | 9.132 | 9.959 | 8.998 | 9.720 | 9.796 | 6.339 |
| 8 | 11.901 | 9.487 | 10.990 | 10.984 | 10.106 | 9.865 | 13.076 | 9.556 | 9.149 | 7.045 | 10.911 | 9.320 | 13.065 | 10.107 | 12.671 | 8.325 | 10.627 | 9.086 | 9.634 | 10.047 | 11.627 | 8.622 | 9.897 | 9.464 | 12.292 | 8.537 | 10.553 | 9.875 | 11.495 | 9.922 |
| 9 | 9.433 | 7.100 | 11.067 | 8.242 | 10.085 | 10.372 | 8.502 | 8.058 | 9.483 | 8.146 | 8.625 | 5.904 | 11.681 | 9.113 | 10.137 | 8.824 | 9.536 | 7.641 | 8.540 | 9.235 | 10.307 | 6.590 | 9.048 | 8.360 | 8.185 | 9.297 | 10.110 | 7.594 | 9.464 | 7.547 |
| 10 | 8.669 | 8.923 | 10.113 | 8.696 | 10.895 | 11.118 | 10.606 | 10.364 | 8.241 | 8.630 | 9.992 | 9.517 | 8.028 | 10.494 | 10.961 | 8.190 | 10.367 | 11.273 | 10.577 | 10.803 | 11.570 | 10.695 | 8.276 | 7.699 | 10.976 | 8.462 | 9.906 | 8.026 | 9.649 | 9.699 |
| 11 | 7.080 | 8.295 | 9.489 | 10.053 | 9.216 | 10.962 | 10.349 | 10.509 | 9.600 | 12.401 | 9.909 | 10.394 | 10.458 | 8.768 | 8.216 | 8.215 | 8.355 | 10.388 | 9.564 | 9.722 | 11.470 | 9.058 | 8.169 | 9.117 | 8.922 | 9.234 | 9.784 | 10.046 | 9.438 | 7.538 |
| 12 | 7.277 | 5.490 | 7.424 | 10.141 | 8.616 | 12.569 | 10.941 | 7.718 | 6.952 | 8.278 | 9.349 | 9.751 | 9.470 | 8.143 | 9.897 | 8.136 | 8.591 | 10.152 | 11.693 | 10.018 | 10.665 | 7.805 | 8.214 | 8.159 | 10.006 | 8.274 | 9.933 | 8.928 | 9.625 | 8.957 |
| 13 | 7.627 | 10.911 | 8.577 | 10.354 | 9.849 | 9.166 | 9.025 | 9.872 | 8.969 | 9.878 | 9.235 | 8.540 | 10.452 | 8.864 | 10.419 | 8.810 | 5.454 | 8.058 | 9.565 | 10.633 | 7.526 | 9.443 | 7.594 | 7.863 | 9.774 | 8.531 | 8.297 | 8.608 | 9.462 | 8.861 |
| 14 | 8.451 | 9.152 | 11.587 | 9.257 | 11.488 | 10.888 | 11.308 | 9.494 | 10.980 | 11.492 | 11.615 | 10.565 | 11.390 | 10.288 | 10.005 | 9.014 | 10.176 | 8.469 | 11.647 | 12.554 | 13.642 | 10.555 | 10.578 | 8.284 | 10.041 | 7.270 | 10.410 | 9.633 | 9.190 | 9.693 |
| 15 | 10.260 | 6.549 | 8.857 | 9.336 | 8.855 | 9.890 | 10.455 | 10.258 | 9.473 | 9.200 | 10.055 | 9.930 | 11.496 | 8.991 | 9.163 | 7.576 | 9.169 | 10.375 | 9.108 | 10.340 | 11.238 | 11.216 | 7.154 | 8.897 | 8.546 | 7.293 | 9.501 | 9.537 | 9.902 | 7.958 |
| 16 | 6.870 | 7.128 | 8.543 | 10.150 | 9.659 | 9.297 | 11.038 | 8.241 | 10.740 | 8.236 | 7.660 | 9.389 | 10.098 | 10.488 | 9.506 | 6.722 | 7.220 | 9.312 | 10.139 | 10.772 | 9.887 | 10.327 | 7.948 | 8.940 | 7.641 | 9.353 | 8.488 | 7.760 | 7.265 | 8.310 |
| 17 | 6.000 | 7.860 | 10.699 | 8.593 | 7.400 | 9.480 | 6.649 | 7.821 | 9.433 | 6.784 | 7.193 | 8.176 | 8.231 | 8.926 | 8.864 | 8.078 | 5.457 | 8.503 | 8.523 | 9.053 | 7.120 | 8.845 | 8.661 | 7.622 | 8.290 | 8.486 | 6.902 | 7.106 | 5.104 | 7.227 |
| 18 | 7.651 | 8.494 | 10.168 | 10.027 | 8.537 | 8.999 | 11.806 | 10.704 | 8.670 | 11.941 | 10.612 | 11.925 | 13.874 | 10.326 | 9.380 | 8.727 | 8.702 | 11.005 | 10.869 | 9.817 | 11.849 | 9.401 | 9.679 | 10.136 | 8.530 | 8.036 | 9.968 | 9.789 | 10.089 | 9.227 |
| 19 | 11.555 | 6.744 | 9.459 | 8.641 | 8.679 | 9.937 | 10.066 | 10.162 | 8.890 | 9.377 | 8.057 | 8.086 | 9.386 | 9.470 | 8.913 | 8.810 | 10.225 | 9.008 | 10.019 | 10.785 | 10.541 | 9.517 | 7.735 | 8.884 | 9.344 | 8.794 | 7.832 | 9.242 | 10.241 | 7.741 |
| 20 | 6.820 | 12.503 | 8.937 | 12.419 | 9.718 | 11.587 | 8.808 | 11.377 | 8.985 | 11.801 | 8.960 | 12.861 | 10.875 | 12.801 | 10.483 | 8.909 | 8.815 | 12.452 | 9.863 | 12.268 | 8.705 | 10.721 | 7.832 | 11.433 | 8.786 | 9.354 | 8.333 | 10.503 | 9.486 | 8.829 |
| 21 | 6.087 | 8.613 | 8.972 | 9.851 | 9.647 | 9.417 | 10.261 | 9.689 | 8.932 | 9.770 | 9.776 | 10.055 | 10.457 | 9.093 | 9.208 | 9.059 | 8.502 | 10.058 | 9.858 | 10.317 | 10.722 | 8.792 | 6.971 | 9.918 | 8.097 | 8.645 | 8.874 | 9.932 | 9.817 | 7.893 |
| 22 | 6.833 | 8.046 | 7.770 | 8.976 | 10.483 | 9.753 | 10.131 | 11.103 | 7.782 | 9.891 | 9.912 | 11.183 | 11.222 | 12.420 | 9.953 | 8.315 | 8.785 | 10.362 | 9.565 | 11.578 | 11.412 | 9.549 | 7.446 | 8.612 | 10.774 | 7.066 | 6.166 | 9.557 | 9.039 | 7.462 |
| 23 | 6.018 | 7.931 | 8.212 | 9.337 | 9.949 | 9.178 | 9.892 | 9.615 | 9.142 | 8.994 | 8.445 | 11.931 | 10.406 | 12.244 | 11.276 | 8.697 | 9.315 | 8.756 | 10.879 | 9.981 | 10.934 | 9.654 | 8.358 | 9.798 | 9.609 | 8.962 | 9.647 | 9.495 | 9.753 | 7.449 |
| 24 | 6.445 | 10.704 | 8.513 | 11.587 | 7.651 | 12.242 | 8.942 | 11.502 | 9.371 | 11.247 | 7.552 | 11.779 | 8.832 | 11.669 | 8.107 | 11.568 | 9.290 | 12.368 | 10.504 | 13.706 | 8.347 | 11.103 | 10.949 | 9.920 | 8.636 | 9.223 | 8.398 | 11.802 | 8.652 | 10.671 |
| mean | 7.660 | 8.474 | 9.402 | 9.424 | 9.398 | 10.104 | 9.837 | 9.644 | 9.187 | 9.569 | 9.265 | 10.102 | 10.303 | 9.899 | 9.771 | 8.722 | 8.732 | 9.502 | 9.675 | 10.285 | 10.371 | 9.067 | 8.589 | 8.891 | 9.304 | 8.630 | 8.988 | 9.253 | 9.066 | 8.504 |
| s.dev | 1.739 | 1.828 | 1.038 | 1.201 | 1.037 | 0.998 | 1.499 | 1.124 | 0.829 | 1.471 | 1.145 | 1.463 | 1.350 | 1.375 | 1.113 | 0.970 | 1.417 | 1.345 | 1.095 | 1.224 | 1.470 | 1.459 | 1.148 | 0.866 | 1.173 | 0.876 | 1.170 | 0.979 | 1.261 | 1.056 |
| var | 3.025 | 3.341 | 1.078 | 1.443 | 1.076 | 0.996 | 2.246 | 1.264 | 0.687 | 2.165 | 1.312 | 2.140 | 1.823 | 1.890 | 1.238 | 0.941 | 2.007 | 1.809 | 1.198 | 1.498 | 2.161 | 2.128 | 1.318 | 0.750 | 1.376 | 0.768 | 1.370 | 0.959 | 1.590 | 1.114 |
| CV | 22.706 | 21.570 | 11.043 | 12.745 | 11.036 | 9.880 | 15.235 | 11.658 | 9.021 | 15.377 | 12.362 | 14.483 | 13.104 | 13.888 | 11.389 | 11.124 | 16.223 | 14.155 | 11.314 | 11.900 | 14.176 | 16.090 | 13.365 | 9.740 | 12.607 | 10.154 | 13.022 | 10.583 | 13.909 | 12.412 |

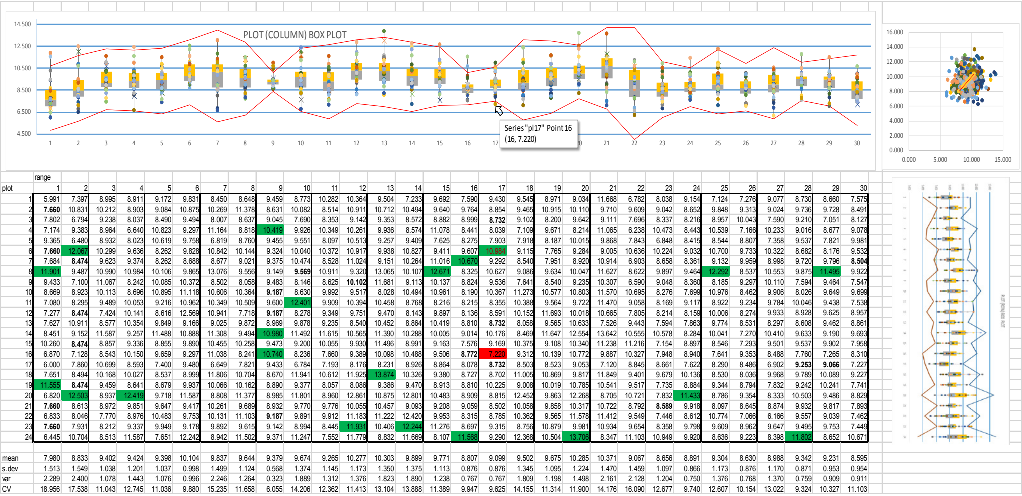
***Figure 3.** Illustration of possible outliers test conducted by application of Excel algorithm for boxplot creation (before replacing values of discarded plots with column mean)*
***Grafikon 3.** Prikaz detekcije mogućih ekstrema primenom Excel algoritma za kreiranje boxplota (pre zamene vrednosti izbačenih parcelica sa prosekom kolone)*

water, weed oasis, plots mechanically damaged during field operations, etc.). This helps to avoid I type error - discarding genotypes that actually should be advanced, which is very important as usually once discarded genotype will not be tested in the next level of field yield test trials. Still remains a possibility of II type error – Advancing of a genotype that actually should be discarded, unless in breeding notes the reason for high performing plot can be assigned to poor neighboring plot in terms of extremely low density, missing neighboring row/rows, etc. The third step would be replacing excluded plot value, most frequently with the environment (row or column mean).

Depending on data and available software, more complex algorithms can be used for missing data value calculation. The fourth step would be statistical testing for possible outliers. Box plot is enough simple and can be efficiently used by the application of Excel algorithm (Figure 3 and 4). The table illustrated in Figure 3, contains raw harvest yield data for one location with maize hybrids FAO 300-400 being tested (before replacing values of excluded plots with column mean). For each row and column separate boxplot was created. Suspected low outliers are colored red, and suspected high outliers are colored green. Column boxplots are given above the table, while row boxplots are placed on right side of the table. At top right corner is placed bivariate test for outliers (regression model).

Replacing of excluded plots based on breeder notes with column means lead to decreased variance, standard deviation and coefficient of variation, with column mean value increased. There remained only one suspected outlier that was not indicated during raw data testing. This outlier plot also was not scored for suppression during field plot scoring visits. If we look at suspected over - performing outliers, we can see that situation here largely remained unchanged (figure 4).

It is important that, when we speak of yield MESPT, this testing does not give information on plots to be excluded, but rather gives information on plots to pay attention to. In practice, it means to recheck breeding notes collected during field observations, and eventually exclude some additional plots if breeding notes give us reason to do so. This is particularly important for unusually high values (colored green), as inbreeding, we actually are in search of “NATURAL” outliers (genotypes showing exceptional performance).



| | range | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| plot | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| 1 | 5.991 | 7.397 | 8.995 | 8.911 | 9.172 | 9.831 | 8.450 | 8.648 | 9.459 | 8.773 | 10.282 | 10.364 | 9.504 | 7.233 | 9.692 | 7.590 | 9.430 | 9.545 | 8.971 | 9.034 | 11.668 | 6.782 | 8.038 | 9.154 | 7.124 | 7.276 | 9.077 | 8.730 | 8.660 | 7.575 |
| 2 | **7.660** | 10.831 | 10.212 | 8.903 | 9.084 | 10.875 | 10.269 | 11.378 | 8.631 | 10.082 | 8.514 | 10.911 | 10.712 | 10.494 | 9.640 | 9.764 | 8.854 | 9.465 | 10.915 | 10.110 | 9.710 | 9.609 | 9.042 | 8.652 | 9.848 | 9.313 | 9.024 | 9.736 | 9.728 | 8.491 |
| 3 | 7.802 | 6.794 | 9.238 | 8.037 | 8.490 | 9.494 | 8.007 | 8.637 | 9.045 | 7.690 | 8.353 | 9.142 | 9.353 | 8.572 | 8.882 | 8.999 | **8.732** | 9.102 | 8.200 | 9.642 | 9.111 | 7.696 | 8.337 | 8.216 | 8.957 | 10.043 | 7.590 | 9.210 | 7.051 | 8.127 |
| 4 | 7.174 | 9.383 | 8.964 | 6.640 | 10.823 | 9.297 | 11.164 | 8.818 | 10.419 | 9.926 | 10.349 | 10.261 | 9.936 | 8.574 | 11.078 | 8.441 | 8.039 | 7.109 | 9.671 | 8.214 | 11.065 | 6.238 | 10.473 | 8.443 | 10.539 | 7.166 | 10.233 | 9.016 | 8.677 | 9.078 |
| 5 | 9.365 | 6.480 | 8.932 | 8.023 | 10.619 | 9.758 | 6.819 | 8.760 | 9.455 | 9.551 | 8.097 | 10.513 | 9.257 | 9.409 | 7.625 | 8.275 | 7.903 | 7.918 | 8.187 | 10.015 | 9.868 | 7.843 | 6.848 | 8.415 | 8.544 | 8.807 | 7.358 | 9.537 | 7.821 | 9.981 |
| 6 | **7.660** | 12.067 | 10.299 | 9.636 | 8.262 | 9.828 | 10.842 | 10.144 | 9.324 | 10.040 | 10.372 | 10.917 | 9.938 | 10.827 | 9.411 | 9.607 | 10.984 | 9.115 | 7.765 | 9.284 | 9.005 | 10.636 | 10.224 | 9.032 | 10.700 | 9.733 | 10.322 | 8.682 | 8.176 | 9.532 |
| 7 | 7.684 | **8.474** | 9.623 | 9.374 | 8.262 | 8.688 | 8.677 | 9.021 | 9.375 | 10.474 | 8.528 | 11.024 | 9.151 | 10.264 | 11.016 | 10.670 | 9.292 | 8.540 | 7.951 | 8.920 | 10.914 | 6.903 | 8.658 | 8.361 | 9.132 | 9.959 | 8.998 | 9.720 | 9.796 | **8.504** |
| 8 | 11.901 | 9.487 | 10.990 | 10.984 | 10.106 | 9.865 | 13.076 | 9.556 | 9.149 | **9.569** | 10.911 | 9.320 | 13.065 | 10.107 | 12.671 | 8.325 | 10.627 | 9.086 | 9.634 | 10.047 | 11.627 | 8.622 | 9.897 | 9.464 | 12.292 | 8.537 | 10.553 | 9.875 | 11.495 | 9.922 |
| 9 | 9.433 | 7.100 | 11.067 | 8.242 | 10.085 | 10.372 | 8.502 | 8.058 | 9.483 | 8.146 | 8.625 | **10.102** | 11.681 | 9.113 | 10.137 | 8.824 | 9.536 | 7.641 | 8.540 | 9.235 | 10.307 | 6.590 | 9.048 | 8.360 | 8.185 | 9.297 | 10.110 | 7.594 | 9.464 | 7.547 |
| 10 | 8.669 | 8.923 | 10.113 | 8.696 | 10.895 | 11.118 | 10.606 | 10.364 | **9.187** | 8.630 | 9.992 | 9.517 | 8.028 | 10.494 | 10.961 | 8.190 | 10.367 | 11.273 | 10.577 | 10.803 | 11.570 | 10.695 | 8.276 | 7.699 | 10.976 | 8.462 | 9.906 | 8.026 | 9.649 | 9.699 |
| 11 | 7.080 | 8.295 | 9.489 | 10.053 | 9.216 | 10.962 | 10.349 | 10.509 | 9.600 | 12.401 | 9.909 | 10.394 | 10.458 | 8.768 | 8.216 | 8.215 | 8.355 | 10.388 | 9.564 | 9.722 | 11.470 | 9.058 | 8.169 | 9.117 | 8.922 | 9.234 | 9.784 | 10.046 | 9.438 | 7.538 |
| 12 | 7.277 | **8.474** | 7.424 | 10.141 | 8.616 | 12.569 | 10.941 | 7.718 | **9.187** | 8.278 | 9.349 | 9.751 | 9.470 | 8.143 | 9.897 | 8.136 | 8.591 | 10.152 | 11.693 | 10.018 | 10.665 | 7.805 | 8.214 | 8.159 | 10.006 | 8.274 | 9.933 | 8.928 | 9.625 | 8.957 |
| 13 | 7.627 | 10.911 | 8.577 | 10.354 | 9.849 | 9.166 | 9.025 | 9.872 | 8.969 | 9.878 | 9.235 | 8.540 | 10.452 | 8.864 | 10.419 | 8.810 | **8.732** | 8.058 | 9.565 | 10.633 | 7.526 | 9.443 | 7.594 | 7.863 | 9.774 | 8.531 | 8.297 | 8.608 | 9.462 | 8.861 |
| 14 | 8.451 | 9.152 | 11.587 | 9.257 | 11.488 | 10.888 | 11.308 | 9.494 | 10.980 | 11.492 | 11.615 | 10.565 | 11.390 | 10.288 | 10.005 | 9.014 | 10.176 | 8.469 | 11.647 | 12.554 | 13.642 | 10.555 | 10.578 | 8.284 | 10.041 | 7.270 | 10.410 | 9.633 | 9.190 | 9.693 |
| 15 | 10.260 | **8.474** | 8.857 | 9.336 | 8.855 | 9.890 | 10.455 | 10.258 | 9.473 | 9.200 | 10.055 | 9.930 | 11.496 | 8.991 | 9.163 | 7.576 | 9.169 | 10.375 | 9.108 | 10.340 | 11.238 | 11.216 | 7.154 | 8.897 | 8.546 | 7.293 | 9.501 | 9.537 | 9.902 | 7.958 |
| 16 | 6.870 | 7.128 | 8.543 | 10.150 | 9.659 | 9.297 | 11.038 | 8.241 | 10.740 | 8.236 | 7.660 | 9.389 | 10.098 | 10.488 | 9.506 | **8.772** | 7.220 | 9.312 | 10.139 | 10.772 | 9.887 | 10.327 | 7.948 | 8.940 | 7.641 | 9.353 | 8.488 | 7.760 | 7.265 | 8.310 |
| 17 | 6.000 | 7.860 | 10.699 | 8.593 | 7.400 | 9.480 | 6.649 | 7.821 | 9.433 | 6.784 | 7.193 | 8.176 | 8.231 | 8.926 | 8.864 | 8.078 | **8.732** | 8.503 | 8.523 | 9.053 | 7.120 | 8.845 | 8.661 | 7.622 | 8.290 | 8.486 | 6.902 | **9.253** | **9.066** | 7.227 |
| 18 | 7.651 | 8.494 | 10.168 | 10.027 | 8.537 | 8.999 | 11.806 | 10.704 | 8.670 | 11.941 | 10.612 | 11.925 | 13.874 | 10.326 | 9.380 | 8.727 | 8.702 | 11.005 | 10.869 | 9.817 | 11.849 | 9.401 | 9.679 | 10.136 | 8.530 | 8.036 | 9.968 | 9.789 | 10.089 | 9.227 |
| 19 | 11.555 | **8.474** | 9.459 | 8.641 | 8.679 | 9.937 | 10.066 | 10.162 | 8.890 | 9.377 | 8.057 | 8.086 | 9.386 | 9.470 | 8.913 | 8.810 | 10.225 | 9.008 | 10.019 | 10.785 | 10.541 | 9.517 | 7.735 | 8.884 | 9.344 | 8.794 | 7.832 | 9.242 | 10.241 | 7.741 |
| 20 | 6.820 | 12.503 | 8.937 | 12.419 | 9.718 | 11.587 | 8.808 | 11.377 | 8.985 | 11.801 | 8.960 | 12.861 | 10.875 | 12.801 | 10.483 | 8.909 | 8.815 | 12.452 | 9.863 | 12.268 | 8.705 | 10.721 | 7.832 | 11.433 | 8.786 | 9.354 | 8.333 | 10.503 | 9.486 | 8.829 |
| 21 | **7.660** | 8.613 | 8.972 | 9.851 | 9.647 | 9.417 | 10.261 | 9.689 | 8.932 | 9.770 | 9.776 | 10.055 | 10.457 | 9.093 | 9.208 | 9.059 | 8.502 | 10.058 | 9.858 | 10.317 | 10.722 | 8.792 | **8.589** | 9.918 | 8.097 | 8.645 | 8.874 | 9.932 | 9.817 | 7.893 |
| 22 | 6.833 | 8.046 | 7.770 | 8.976 | 10.483 | 9.753 | 10.131 | 11.103 | **9.187** | 9.891 | 9.912 | 11.183 | 11.222 | 12.420 | 9.953 | 8.315 | 8.785 | 10.362 | 9.565 | 11.578 | 11.412 | 9.549 | 7.446 | 8.612 | 10.774 | 7.066 | 6.166 | 9.557 | 9.039 | 7.462 |
| 23 | **7.660** | 7.931 | 8.212 | 9.337 | 9.949 | 9.178 | 9.892 | 9.615 | 9.142 | 8.994 | 8.445 | 11.931 | 10.406 | 12.244 | 11.276 | 8.697 | 9.315 | 8.756 | 10.879 | 9.981 | 10.934 | 9.654 | 8.358 | 9.798 | 9.609 | 8.962 | 9.647 | 9.495 | 9.753 | 7.449 |
| 24 | 6.445 | 10.704 | 8.513 | 11.587 | 7.651 | 12.242 | 8.942 | 11.502 | 9.371 | 11.247 | 7.552 | 11.779 | 8.832 | 11.669 | 8.107 | 11.568 | 9.290 | 12.368 | 10.504 | 13.706 | 8.347 | 11.103 | 10.949 | 9.920 | 8.636 | 9.223 | 8.398 | 11.802 | 8.652 | 10.671 |
| mean | 7.980 | 8.833 | 9.402 | 9.424 | 9.398 | 10.104 | 9.837 | 9.644 | 9.379 | 9.674 | 9.265 | 10.277 | 10.303 | 9.899 | 9.771 | 8.807 | 9.099 | 9.502 | 9.675 | 10.285 | 10.371 | 9.067 | 8.656 | 8.891 | 9.304 | 8.630 | 8.988 | 9.342 | 9.231 | 8.595 |
| s.dev | 1.513 | 1.549 | 1.038 | 1.201 | 1.037 | 0.998 | 1.499 | 1.124 | 0.568 | 1.374 | 1.145 | 1.173 | 1.350 | 1.375 | 1.113 | 0.876 | 0.876 | 1.345 | 1.095 | 1.224 | 1.470 | 1.459 | 1.097 | 0.866 | 1.173 | 0.876 | 1.170 | 0.871 | 0.953 | 0.954 |
| var | 2.289 | 2.400 | 1.078 | 1.443 | 1.076 | 0.996 | 2.246 | 1.264 | 0.323 | 1.889 | 1.312 | 1.376 | 1.823 | 1.890 | 1.238 | 0.767 | 0.767 | 1.809 | 1.198 | 1.498 | 2.161 | 2.128 | 1.204 | 0.750 | 1.376 | 0.768 | 1.370 | 0.759 | 0.909 | 0.911 |
| CV | 18.956 | 17.538 | 11.043 | 12.745 | 11.036 | 9.880 | 15.235 | 11.658 | 6.055 | 14.206 | 12.362 | 11.413 | 13.104 | 13.888 | 11.389 | 9.947 | 9.625 | 14.155 | 11.314 | 11.900 | 14.176 | 16.090 | 12.677 | 9.740 | 12.607 | 10.154 | 13.022 | 9.324 | 10.327 | 11.103 |

***Figure 4.*** *Illustration of possible outliers test conducted by application of Excel algorithm for boxplot creation (after replacing values of discarded plots with column mean)*

***Grafikon 4.*** *Prikaz detekcije mogućih ekstrema primenom Excel algoritma za kreiranje boxplota (posle zamene vrednosti izbačenih parcelica sa prosekom kolone)*

Once we are confident in our data, we can proceed with their statistical processing.

## Conclusion

In order to obtain proper results of statistical processing of multi - environment small plot field trials during the breeding process, it is necessary to have “good data” as input. Only analysis of such data gives us a solid basis for decision making.

Before data processing, data should be “cleaned” which means that all data that can be misleading should be excluded from the analysis and replaced with calculated values.

Data cleaning leads to increase of mean value and, at the same time, a decrease of variance, standard deviation and coefficient of variation.

In a case of assessing yield by multi - environment small plot field trials, breeder notes are essential for good “cleaning process”.

Only based on breeder notes artificial outliers (errors) should be excluded from the analysis.

## Acknowledgement

This paper is a result of the research within the Project TR 31723, supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

## References

Arciniegas-Alarcón S, García-Peña M, Dias CTS (2011): Data imputation in trials with genotype×environment interaction. Interciencia, 36: 444-449.

Arciniegas-Alarcón S, García-Peña M, Krzanowski W. (2016): Missing value imputation in multi-environment trials: Reconsidering the Krzanowski method. Crop Breeding and Applied Biotechnology, 16: 77-85.

Babić M, Babić V, Delić N, Anđelković V, Prodanović S (2011): The comparison of stability parameters according to the Finlay-Wilkinson, Eberhart-Russell and AMMI model. Selekcija i Semenarstvo, 17(2): 35-40.

Crossa J, Perez-Elizalde S, Jarquin D, Cotes JM, Viele K, Liu G, Cornelius PL (2011): Bayesian estimation of the additive main effects and multiplicative interaction model. Crop Science, 51: 1458-1469.

Fritsche-Neto R, Gonçalves MC, Vencovsky R and Souza Junior CL (2010): Prediction of genotypic values of maize hybrids in unbalanced experiments. Crop Breeding and Applied Biotechnology, 10: 32-39.

Josse J, van Eeuwijk F, Piepho HP, Denis JB (2014): Another look at Bayesian analysis of AMMI models for genotype-environment data. Journal of Agricultural, Biological, and Environmental Statistics, 19: 240-257.

Kumar A, Verulkar SB, Mandal NP, Variar M, Shukla VD, Dwivedi JL, Singh BN, Singh ON, Swain P, Mall AK, Robin S, Chandrababu R, Jain A, Haefele SM, Piepho HP, Raman A (2012): High-yielding, drought-tolerant, stable rice genotypes for the shallow rainfed lowland droughtprone ecosystem. Field Crops Research, 133: 37-47.

Michael Galarnyk (2019): Understanding Boxplots. Available in https://www.kdnuggets.com/2019/11/understanding-boxplots.html (14. October, 2019).

Omer SO, Abdalla AWH, Mohammed MH, Singh M (2015): Bayesian estimation of genotype-by-environment interaction in sorghum variety trials. Communications in Biometry and Crop Science, 10: 82–95.

Ray SuniL (2016): A comprehesive Guide to Data Exploration. https://www.analyticsvidhya.com/blog/2016/01/guide-data-exploration/ (02. October, 2019).

Vakili K, Schmitt E (2014): Finding multivariate outliers with FastPCS. Computational Statistics and Data Analysis, 69:55–66.

Woyann LG, Benin G, Storck L, Trevizan DM, Menaguzzi C, Marchioro VS, Madureira A (2017): Estimation of missing values affects important aspects of GGE biplot Analysis. Crop Science, Vol. 57: 1-13.

Yan W (2013): Biplot analysis of incomplete two-way data. Crop Science, 53: 48-57.

Yan W, Pageau D, Frégeau-Reid J, Durand J (2011): Assessing the representativeness and repeatability of test locations for genotype evaluation. Crop Science, 51: 1603-1610.

# ZNAČAJ PROCESA ČIŠĆENJA PODATAKA SORTNOG OGLEDA ZA DONOŠENJE KVALITETNIH ODLUKA U OPLEMENJIVANJU

Milosav Babić, Petar Čanak, Bojana Vujošević,
Vojka Babić, Dušan Stanisavljević

## Sažetak

Svrha poljskih ogleda u oplemenjivanju biljaka je da omogući odabir najuspešnijeg genotipa, što nije uvek jednostavan zadatak prevashodno zbog postojanja interakcija genotipa i spoljašnje sredine. Upravo zbog postojanja interakcija sortni ogledi se izvode u brojnim lokacijama i godinama, kako bi se dobila pouzdana procena vrednosti genotipa. U toku oplemenjivačkog procesa, procena prinosa, recimo, hibrida kukuruza, je zasnovana isključivo na rezultatima višelokacijskih sortnih mikro ogleda (VSMO). Ovaj deo oplemenjivačkog procesa je stoga najzahtevniji sa tehničkog i finansijskog aspekta ali i sa stanovišta angažovanja obučene radne snage i specifične opreme. U ovom radu prikazan je jedan od mogućih sistematskih pristupa u proceni višelokacijskog sortnog ogleda. Glavni cilj predstavljenog pristupa je da obezbedi najbolji mogući rezultat uz angažovanje razumnih resursa. Kako rezultati poljskog ogleda ne mogu biti direktno interpretirani bez prethodne obrade podataka, kvalitet ulaznih-sirovih podataka je od krucijalne važnosti za dobijanje relevantnih i objektivnih procena relativne vrednosti novostvorenih genotipova (hibrida kukuruza) u smislu njihove produktivnosti i stabilnosti. Postoje brojne definicije kvaliteta podataka, ali se podaci generalno mogu smatrati visoko kvalitetnim ako su odgovarajući za planirane statističke obrade, donošenje odluka i planova. Cilj izloženog rada je da naglasi značaj pročišćavanja/čišćenja podataka pre procesa statističke obrade podataka sortnih mikro ogleda.

***Ključne reči:*** zapisi sa polja, oplemenjivanje kukuruza, ocena elementarne parcele, čišćenje podataka

Primljen: 18.11.2019.
Prihvaćen: 1.12.2019.