Journal of Al Rafidain University College **Issue No. 46/ 2020**
The (15th & the 2nd International) Conference of Statistical Applications **ISSN (1681- 6870)**

Detection and Treatment of Outliers in Experimental Design: Real Data for Completely Randomized Design

Assist. Prof. Dr. Fayyadh Abdulla Ali

Department of statistic, Wasit university.

E-mail: fa311057@yahoo.com

Abstract

The presence of outliers values in the data leads to errors in statistical analysis due to the use of traditional methods of calculation, so it is necessary to switch to new methods that deal with these outlier values so as to ensure the accuracy of the calculations to the proper statistical analysis, and in this research resorted to the method adjusted boxplot to detect outlier values and then deleted and re-statistical analysis data have been used for a realistic agricultural experiment to completely randomized design in the College of Agriculture Wasit for 2017 has shown the result of statistical analysis that there is a difference in the results before and after deleting outlier values.

Keyword: IQR, outliers, boxplot, adjusted boxplot, informal methods, med couple (MC)

كشف ومعالجة الشوارد والقيم المتطرفة في التصاميم الزراعية: بيانات حقيقية لتصميم تام التعشية

أ.م.د. فياض عبد الله علي
جامعة واسط / قسم الإحصاء

المستخلص: ان وجود القيم الشاردة في البيانات يؤدي الى اخطاء في التحليل الاحصائي نتيجة استخدام الطرق التقليدية في الحسابات لذا لابد من التحول الى طرق جديدة تتعامل مع هذه القيم المتطرفة او الشاردة بحيث نضمن دقة الحسابات وصولا للتحليل الاحصائي السليم، وفي هذا البحث تم اللجوء الى طريقة adjusted boxplot لكشف القيم الشاردة ومن ثم حذفها واعادة التحليل الاحصائي وقد استخدمت بيانات لتجربة زراعية واقعية لتصميم تام التعشية في كلية زراعة واسط لعام 2017 وقد بينت نتيجة التحليل الاحصائي ان هناك اختلافا في النتائج قبل وبعد حف القيم الشاردة .

الكلمة الرئيسية: تصميم تام التعشية، القيم المتطرفة ، boxplot، boxplot المعدل ، الطرق غير الرسمية ، med couple (MC)

ISSN (1681- 6870)

1. Introduction

The problem of extreme values is a problem that many researchers have been interested in. In (1978) john studied the problems which arise from presence of outliers in results of factorial experiment[4]. Later in the same year John and Draper[5] investigated the problem of outliers in two-way table and provided a statistic Q_k which is the differences between sum squares residuals from the original data and sum squares revised residuals from fitted basic model after deleting k-observations. In 1992 Ben - Yohai [2]studied M- estimates and their test for a one-factor experiment in randomized block design. In 2001 Bhar and Gupta[6] modified, Qk-statistic(Cook-statistic) and AP- statistic in application to experimental design . and newly, in 2019 the researchers Stefan Mandic-Rajcevic and Claudio Colosio [8] conducted a study aimed to propose and test methods to identify outliers and quantify its influence on the overall exposure assessment, and Validation of the approach using biological monitoring on a sample of agricultural workers in Italian vineyards. Inhyeok Bae and Un Ji [1] introduced in 2019 a study aimed at developing a generalized algorithm for statistical processes to remove outliers and random errors from water level data obtained by ultrasonic sensors in actual flow conditions and soften water level data to reduce dispersion caused by water waves are classified as random errors

From a time that we started of employing and exploiting information in the collected data to help us to understand the world we lived in, there is a concern over outlier(s) of observations or the unrepresentative in the data set. Outlier(s) which found in a data set is (are) defined as observation (or sub-set of observations) that appears to be inconsistent with the remainder of that set of data.
Occurrence of outlier(s) is (are) very common in varies fields involving collection of data and outlier(s) arises from distributions with heavy tail or is simply by bad data point due to error. When the outlier(s) presented in the data, the whole set up of the experiment is disturbed. many statistics to detect influential point or a single outlier (in regression analysis now are available), where these statistics were developed under a assumption that data are generated from a kind of linear model (the design matrix is a full rank)[3]. However, in case of experimental design, the matrix is not full rank(less than full rank) and the interesting is in estimation parameters of some linear functions, as treatment effects. Estimation to these functions may be affected in the existence of outliers. Attention must be focused on the influence of outliers on estimation of this subset of parameters.

2. Research problem

The presence of outliers or extreme values in the data leads to errors in statistical analysis due to the use of traditional methods of calculation, Incorrect analysis leads to a false conclusion so it is necessary to switch to new methods dealing with these outliers or extremes values so as to ensure the accuracy of the calculations to the proper statistical analysis, and conclusions, so one should pay careful attention to which approach to use in the analysis

3. Outliers and methods of detection

In many cases, when collecting data for a research or experiment, there are many views that differ significantly from the main body of data. The values of these observations are higher or

lower than the rest of the observations (given our previous knowledge of the nature of the data). These are often called extreme or outliers values that arise for a variety reasons
Outliers, or extreme values, can be defined as that observations numerically distant from the rest of the data, can reduce data quality and cause erroneous judgment [8].
The question is how to determine that these observations are far from the original data set.
There are two methods for detection outliers [7]: Formal test (discordancy) and informal (labeling) test.
Most formal tests use test statistics for hypothesis testing. Which is assumed some well-behaving distribution, and test what if the extreme value is an outlier, i.e., weather it deviates from the assumed distribution. Selection the test depends on type of data distribution, and type of target outliers. In spite of powerful of formal tests under well-behaving statistical assumptions such as assumption of distribution, most distributions of real data may not follow specific distributions or unknown such as the normal, exponential or gamma. Another restriction is that of swamping or susceptible to masking problems which is defined by Acuna and Rodriguez (2004) as follows:

Masking effect: when one outlier masks a second outlier if the second outlier can be regarded as an outlier by itself only, but not in presence of the first outlier. Thus, after the first outlier is detected the second instance is appeared as an outlier.
Swamping effect: one outlier swamps the second if the second can be regarded as an outlier only under presence of the first one. In other words, after deleting of the first outlier the second observation becomes not outlier observation.
Many studies take into account these problems have been conducted by Bendre and Kale (1987). Davies and Gather (1993), Iglewicz and Hoaglin (1993), Barnett and Lewis (1994)
On the other hand however, most outlier informal (labeling) tests generate criterion or an interval to detect outlier instead of. testing hypothesis, and any observations exceed the criterion or interval is regarded as an outlier. Different scale and location parameters are generally used in each in formal method to define feasible interval or criterion to detect the outlier. There are two causes for employ an outlier informal method. One is to detect possible outliers as a checking device prior conducting a formal test. The other is to detection the outlier values far from the lump of the data regardless of the distribution.
Whereas the formal tests call for test statistics based on distribution a hypothesis and assumptions to determine whether the extreme value is a real of the distribution. Most outlier informal methods display the interval use the scale and location parameters of the data. Although the informal method is generally simple to use, several observations out the interval may turn out to be falsely specified outliers after a formal test when the outliers are defined as only observations that veer from the assuming distribution. However, if the object of finding outlier is not introductory step to detection the extreme values violate the assumptions of distribution of the main statistical analyses such as ANOVA, the t-test and regression. But mainly to detect the extreme values far from the lump of the data regardless the distribution, the outlier informal methods may be usable. In addition, for a big data set that is statistic problematic, e.g., when it is hard to identify the distribution of data or change it into appropriate distribution such the normal distribution, informal methods can be applied to detect outliers.
In this paper we focuses on outlier informal methods

ISSN (1681- 6870)

4. Outlier informal method

There are seven informal methods[7] to detect outliers, all adopt the principle of excluding values that exceed two limits (minimum and maximum) but we will choose the method of being the most recent and appropriate for the data under study.

4.1. Tukey's method (boxplot)

Tukey's (1977) method, building a boxplot, is a known simple graphical tool to show information on continued univariate data, such the median, lower quartile and upper quartile , lower extremist, upper extremist of a data. It is less sensitive for extreme values than other methods (standard deviation (sd) method, z-score and the modified z-score) that use sample mean and standard variance because it uses quartiles which are resistant the extreme values. The steps of the method defined as follows:

1. The IQR (Inter Quartile Range) is the range between the lower (Q3) and upper (Q1) Quartiles.
2. Inner fencing are located at a distance 1.5 IQR beneath Q_1 and over Q_3 $[Q_1 - 1.5\, IQR, Q_3 + 1.5 IQR]$.
3. Outer fencing are located at a distance 3 IQR below Q_1 and over Q_3 [Q_1-3 IQR, Q_3+3 IQR].
4. A value among the inside and outside fences is a probable outlier. An extreme value beyond the outer fences is a possible outlier. There is no statistical grounds for the reason that Tukey uses 1.5 and 3 concerning the IQR to create inner and outer fences.

While other methods are restricted to mound-shaped and symmetric data such normal distribution, Tukey's method is viable to skewed or non mound-shaped data since it no need distributional assumptions and it does not count on a mean or standard deviation.

4.2. Adjusted boxplot

The criterion boxplot is one of the most common nonparametric tools to detect outliers in univariate data. For Gaussian or any symmetric distributions, the opportunity of data occurring outside of the criterion boxplot fence is not exceed 0.7%[10].Though the boxplot perhaps applicable to symmetric or skewed data, when the data more skewed, we may be detect more observations as outliers. This conclusion come from the fact that this method is depend on robust measures such lower and upper quartiles and the IQR regardless the skewness of the data.

Vanderviere and Huber (2004) introduced an adjusted boxplot considering the med couple (MC), a robust measure of skewness for a skewed distribution.

To defeat this problem, a medcouple (MC) that is robust to resist outliers and has sensitivity to detect skewness was introduced to construct new robust skewed boxplot fence[10].

Let $Y_n = y_1, y_2, \ldots, y_n$ is a set of data independently sampled from a continuous variable distribution and it is arranged such as $y_1 \le y_2 \le \cdots \le y_n$, the MC is defined as :

$MC(y_1, y_2, \ldots, y_n) = \frac{(y_j - med_k) - (med_k - y_i)}{y_j - y_i}$, where med_k is median of Y_n , and I and j satisfying $y_i \le med_k \le y_j$, and $y_i \ne y_j$

The adjusted boxplot has intervals [4]:

$$L = Q_1 - 1.5 \times exp(-4MC) \times IQR\ ,\ \ U = Q_3 + 1.5 \times exp(3MC) \times IQR \quad if\ MC \ge 0 \ \ldots (1)$$

ISSN (1681- 6870)

And

$$L = Q_1 - 1.5 \times exp(-3MC) \times IQR \,, \quad U = Q_3 + 1.5 \times exp(4MC) \times IQR \quad if\ MC < 0 \ldots (2)$$

where L , U represent the lower and upper fence of the interval respectively. The observations that fall outside this interval are considered as an outliers.
MC value ranges is between -1 and 1. If MC equal 0, the data is symmetric and the adjusted boxplot be Tukey's box plot. If MC greater than0, the data has a right skewed distribution, while if MC less than 0, the data has a left skewed distribution.
Vanderviere and Huber (2004) were computed the mean percentage of outliers beyond the upper and lower fence of the adjusted Boxplot and Tukey's Boxplot method, for various distributions and several sample sizes. In simulation experiment , less observations, especially in the right tail, were categorized as outliers compared to Tukey's method when the data are skewed to the right. In case of a mildly right-skewed distribution, the lower fence of the interval might move to the right and addional observations in the left side will be categorized as outliers compared with Tukey's method. This difference comes from decreasing in lower fence and increasing in the upper fence from Q1 and Q3, respectively.

5. Field experiment

Field experiment were carried in Wasit district during the 2017 season to study the influence of genetic inbreed line of maize. The seeds were planted (two seeds) on a 75 cm cedar and 25 cm in diameter.
After germination, Superphosphate fertilizer was added at an average of 200 kg /ha at tillage and urea fertilizer was added at a rate of 200 kg / ha. Weeding operations and Incineration were carried out as needed. Ten random plants were selected and reaped at maturity.

6. Statistical analysis

The table (1) shows the original data on Average grain yield (g / plant) of eight inbred lines of maize in nine replications.

Table 1 : Average grain yield (g / plant) of eight inbred lines of maize in nine replications.

Replications	Inp-6 (Locale)	Pio-17 (Yugoslavia)	Syn-9 (French)	Zm-17 (Yugoslavia)	Pio-3 (Yugoslavia)	S-10 (Australian)	MGW-1 (Yugoslavia)	Ast-B (Australian)
1	**230.00**	**205.00**	250.00	284.00	231.00	300.00	190.00	305.00
2	240.32	**212.33**	**260.89**	**230.00**	220.66	**244.74**	**157.82**	322.00
3	**171.03**	**247.86**	240.19	**250.67**	**205.19**	291.23	**173.56**	291.42
4	242.66	**170.67**	**188.33**	262.89	**291.33**	**328.86**	181.67	358.31
5	250.19	180.83	**258.80**	**345.16**	**258.66**	**262.20**	210.00	**364.55**
6	**292.35**	181.14	244.81	**305.70**	231.23	319.05	**219.00**	**416.05**
7	**290.06**	183.47	251.41	**349.21**	**249.72**	305.59	197.89	**368.97**
8	276.82	184.59	**285.52**	269.14	**204.81**	**330.08**	**213.60**	291.29
9	**294.51**	**175.02**	**207.17**	281.76	240.33	**322.82**	**241.17**	**264.63**

Below are given the analysis of variance with the original data (table 2)

ISSN (1681- 6870)

Table (2) ANOVA with origin data

Source of variations	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	147287.218	7	21041.031	18.032	.000
Within Groups	74679.817	64	1166.872		
Total	221967.035	71			

Table (3) Multiple Comparisons (L.S.D) with origin data

(I) genetic	(J)genetic	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Inp-6	Pio-17	60.78111*	16.10295	.000	28.6118	92.9504
	Syn-9	11.20222	16.10295	**.489**	-20.9671-	43.3716
	Zm-17	-32.28778*	16.10295	.049	-64.4571-	-.1184-
	Pio-3	17.22333	16.10295	**.289**	-14.9460-	49.3927
	S-10	-46.29222*	16.10295	.005	-78.4616-	-14.1229-
	MGW-1	55.91444*	16.10295	.001	23.7451	88.0838
	Ast-B	-77.14222*	16.10295	.000	-109.3116-	-44.9729-
Pio-17	Inp-6	-60.78111*	16.10295	.000	-92.9504-	-28.6118-
	Syn-9	-49.57889*	16.10295	.003	-81.7482-	-17.4096-
	Zm-17	-93.06889*	16.10295	.000	-125.2382-	-60.8996-
	Pio-3	-43.55778*	16.10295	.009	-75.7271-	-11.3884-
	S-10	-107.07333*	16.10295	.000	-139.2427-	-74.9040-
	MGW-1	-4.86667-	16.10295	**.763**	-37.0360-	27.3027
	Ast-B	-137.92333*	16.10295	.000	-170.0927-	-105.7540-
Syn-9	Inp-6	-11.20222-	16.10295	**.489**	-43.3716-	20.9671
	Pio-17	49.57889*	16.10295	.003	17.4096	81.7482
	Zm-17	-43.49000*	16.10295	.009	-75.6593-	-11.3207-
	Pio-3	6.02111	16.10295	**.710**	-26.1482-	38.1904
	S-10	-57.49444*	16.10295	.001	-89.6638-	-25.3251-
	MGW-1	44.71222*	16.10295	.007	12.5429	76.8816
	Ast-B	-88.34444*	16.10295	.000	-120.5138-	-56.1751-
Zm-17	Inp-6	32.28778*	16.10295	.049	.1184	64.4571
	Pio-17	93.06889*	16.10295	.000	60.8996	125.2382
	Syn-9	43.49000*	16.10295	.009	11.3207	75.6593
	Pio-3	49.51111*	16.10295	.003	17.3418	81.6804
	S-10	-14.00444-	16.10295	**.388**	-46.1738-	18.1649
	MGW-1	88.20222*	16.10295	.000	56.0329	120.3716
	Ast-B	-44.85444*	16.10295	.007	-77.0238-	-12.6851-
Pio-3	Inp-6	-17.22333-	16.10295	**.289**	-49.3927-	14.9460
	Pio-17	43.55778*	16.10295	.009	11.3884	75.7271
	Syn-9	-6.02111-	16.10295	**.710**	-38.1904-	26.1482
	Zm-17	-49.51111*	16.10295	.003	-81.6804-	-17.3418-

 ISSN (1681- 6870)

	S-10	-63.51556*	16.10295	.000	-95.6849-	-31.3462-
	MGW-1	38.69111*	16.10295	.019	6.5218	70.8604
	Ast-B	-94.36556*	16.10295	.000	-126.5349-	-62.1962-
S-10	Inp-6	46.29222*	16.10295	.005	14.1229	78.4616
	Pio-17	107.07333*	16.10295	.000	74.9040	139.2427
	Syn-9	57.49444*	16.10295	.001	25.3251	89.6638
	Zm-17	14.00444	16.10295	**.388**	-18.1649-	46.1738
	Pio-3	63.51556*	16.10295	.000	31.3462	95.6849
	MGW-1	102.20667*	16.10295	.000	70.0373	134.3760
	Ast-B	-30.85000-	16.10295	**.060**	-63.0193-	1.3193
MGW-1	Inp-6	-55.91444*	16.10295	.001	-88.0838-	-23.7451-
	Pio-17	4.86667	16.10295	**.763**	-27.3027-	37.0360
	Syn-9	-44.71222*	16.10295	.007	-76.8816-	-12.5429-
	Zm-17	-88.20222*	16.10295	.000	-120.3716-	-56.0329-
	Pio-3	-38.69111*	16.10295	.019	-70.8604-	-6.5218-
	S-10	-102.20667*	16.10295	.000	-134.3760-	-70.0373-
	Ast-B	-133.05667*	16.10295	.000	-165.2260-	-100.8873-
Ast-B	Inp-6	77.14222*	16.10295	.000	44.9729	109.3116
	Pio-17	137.92333*	16.10295	.000	105.7540	170.0927
	Syn-9	88.34444*	16.10295	.000	56.1751	120.5138
	Zm-17	44.85444*	16.10295	.007	12.6851	77.0238
	Pio-3	94.36556*	16.10295	.000	62.1962	126.5349
	S-10	30.85000	16.10295	**.060**	-1.3193-	63.0193
	MGW-1	133.05667*	16.10295	.000	100.8873	165.2260
*. The mean difference is significant at the 0.05 level.						

To ensure that the data is free from outliers, we use the boxplot and adjusted boxplot method described above and then we extract the upper and lower values , table(4) and table(5) show the results for boxplot and table(6) for adjusted boxplot:

Table(4): Inner fences for average grain yield (g / plant) of eight inbred lines of maize(boxplot)

Inner fences	Inp-6 (Locale)	Pio-17 (Yugoslavia)	Syn-9 (French)	Zm-17 (Yugoslavia)	Pio-3 (Yugoslavia)	S-10 (Australian)	MGW-1 (Yugoslavia)	Ast-B (Australian)
lower	231.7413	176.9986	216.2841	254.7728	207.6315	268.8378	175.271	290.7792
upper	287.5887	199.3714	256.0934	299.2272	248.7985	322.4972	213.0165	363.5333

Table(5): Outer fences for average grain yield (g / plant) of eight inbred lines of maize(boxplot)

Outer fences	Inp-6 (Locale)	Pio-17 (Yugoslavia)	Syn-9 (French)	Zm-17 (Yugoslavia)	Pio-3 (Yugoslavia)	S-10 (Australian)	MGW-1 (Yugoslavia)	Ast-B (Australian)
lower	70.07	106.1975	90.8425	114.075	94.1125	112.1975	64.25	76.32
upper	449.26	270.1725	381.535	439.925	362.3175	479.1375	324.0375	577.9925

ISSN (1681- 6870)

Note from the boundaries of the two tables (4,5) above that there are values between the inner and outer fences is a possible outlier, they are bolded in a table (1)

Table (6): lower and upper fence for average grain yield (g / plant) of eight inbred lines of maize(adjusted boxplot)

223.895	169.9041	203.7807	240.6723	202.9145	261.0531	169.6373	279.8322
721.5964	208.5372	263.2548	304.5463	1243.023	593.4213	317.5383	681.663

Note from the boundaries of the table(6) above that there are observations fall outside this interval which considered as an outliers (they are underlined in table(1))
Due to the fact that boxplot method did not specify the extreme values Categorically, we will use the adjusted boxplot method in determining outliers.
The tables (7,8) show the results of statistical analysis after deleting the outlier values

Table (7): ANOVA after deleting outliers

Source of variation	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	139333.558	7	19904.794	26.364	.000
Within Groups	40014.611	53	754.993		
Total	179348.169	60			

Table (8): Multiple Comparisons (L.S.D) after deleting outliers

(I) genetic	(J) genetic	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Inp-6	Pio-17	86.59857*	14.68714	.000	57.1399	116.0573
	Syn-9	24.80571	14.68714	**<u>.097</u>**	-4.6530-	54.2644
	Zm-17	-6.13476-	15.28686	**<u>.690</u>**	-36.7963-	24.5268
	Pio-3	32.57190*	13.84717	.022	4.7980	60.3458
	S-10	-37.92018*	14.22077	.010	-66.4434-	-9.3969-
	MGW-1	66.19732*	14.22077	.000	37.6741	94.7206
	Ast-B	-61.79921*	13.84717	.000	-89.5731-	-34.0253-
Pio-17	Inp-6	-86.59857*	14.68714	.000	-116.0573-	-57.1399-
	Syn-9	-61.79286*	14.68714	.000	-91.2515-	-32.3342-
	Zm-17	-92.73333*	15.28686	.000	-123.3949-	-62.0718-
	Pio-3	-54.02667*	13.84717	.000	-81.8006-	-26.2528-
	S-10	-124.51875*	14.22077	.000	-153.0420-	-95.9955-
	MGW-1	-20.40125-	14.22077	**<u>.157</u>**	-48.9245-	8.1220
	Ast-B	-148.39778*	13.84717	.000	-176.1717-	-120.6239-

Syn-9	Inp-6	-24.80571-	14.68714	**.097**	-54.2644-	4.6530
	Pio-17	61.79286*	14.68714	.000	32.3342	91.2515
	Zm-17	-30.94048*	15.28686	.048	-61.6020-	-.2789-
	Pio-3	7.76619	13.84717	**.577**	-20.0077-	35.5401
	S-10	-62.72589*	14.22077	.000	-91.2491-	-34.2026-
	MGW-1	41.39161*	14.22077	.005	12.8684	69.9149
	Ast-B	-86.60492*	13.84717	.000	-114.3788-	-58.8310-
Zm-17	Inp-6	6.13476	15.28686	**.690**	-24.5268-	36.7963
	Pio-17	92.73333*	15.28686	.000	62.0718	123.3949
	Syn-9	30.94048*	15.28686	.048	.2789	61.6020
	Pio-3	38.70667*	14.48172	.010	9.6600	67.7533
	S-10	-31.78542*	14.83935	.037	-61.5494-	-2.0215-
	MGW-1	72.33208*	14.83935	.000	42.5681	102.0960
	Ast-B	-55.66444*	14.48172	.000	-84.7111-	-26.6178-
Pio-3	Inp-6	-32.57190*	13.84717	.022	-60.3458-	-4.7980-
	Pio-17	54.02667*	13.84717	.000	26.2528	81.8006
	Syn-9	-7.76619-	13.84717	**.577**	-35.5401-	20.0077
	Zm-17	-38.70667*	14.48172	.010	-67.7533-	-9.6600-
	S-10	-70.49208*	13.35149	.000	-97.2718-	-43.7124-
	MGW-1	33.62542*	13.35149	.015	6.8457	60.4051
	Ast-B	-94.37111*	12.95284	.000	-120.3512-	-68.3910-
S-10	Inp-6	37.92018*	14.22077	.010	9.3969	66.4434
	Pio-17	124.51875*	14.22077	.000	95.9955	153.0420
	Syn-9	62.72589*	14.22077	.000	34.2026	91.2491
	Zm-17	31.78542*	14.83935	.037	2.0215	61.5494
	Pio-3	70.49208*	13.35149	.000	43.7124	97.2718
	MGW-1	104.11750*	13.73856	.000	76.5614	131.6736
	Ast-B	-23.87903-	13.35149	**.079**	-50.6587-	2.9007
MGW-1	Inp-6	-66.19732*	14.22077	.000	-94.7206-	-37.6741-
	Pio-17	20.40125	14.22077	**.157**	-8.1220-	48.9245
	Syn-9	-41.39161*	14.22077	.005	-69.9149-	-12.8684-
	Zm-17	-72.33208*	14.83935	.000	-102.0960-	-42.5681-
	Pio-3	-33.62542*	13.35149	.015	-60.4051-	-6.8457-
	S-10	-104.11750*	13.73856	.000	-131.6736-	-76.5614-
	Ast-B	-127.99653*	13.35149	.000	-154.7762-	-101.2168-
Ast-B	Inp-6	61.79921*	13.84717	.000	34.0253	89.5731
	Pio-17	148.39778*	13.84717	.000	120.6239	176.1717
	Syn-9	86.60492*	13.84717	.000	58.8310	114.3788

ISSN (1681- 6870)

	Zm-17	55.66444*	14.48172	.000	26.6178	84.7111
	Pio-3	94.37111*	12.95284	.000	68.3910	120.3512
	S-10	23.87903	13.35149	**.079**	-2.9007-	50.6587
	MGW-1	127.99653*	13.35149	.000	101.2168	154.7762
*. The mean difference is significant at the 0.05 level.						

It is noted through statistical analysis before and after deletion of outliers that there is a difference in the results, for example, the value of F (tables (2,7)) before the deletion is less than after deletion and thus significantly increased differences between types of genetics.
It is also noted that the results of the comparisons (L.S.D) differ between the averages of the different genetics before and after the removal of outliers (genetics inp-6, Zm-17,Pio-3,S-10).

7. Conclusions

- Adjusted boxplot method is better than boxplot method because it clearly defines the outliers while boxplot method defines the observations among the inside and outside fences as probable outliers.
- - The need to conduct a process of data examination through the method of adjusted boxplot (for ease and accuracy) before statistical analysis in order to get rid of outliers that cause error in the analysis and therefore error in conclusions.

References

[1] Bae ,Inhyeok and Un Ji,"Outlier Detection and Smoothing Process for Water Level Data Measured by Ultrasonic Sensor in Stream Flows”, Water **2019**, 11, 951 http://www.mdpi.com/journal/water

[2] Ben, M. G. and Yohai, V. J. “Robust analysis of variance for a randomized block design”. *Commun. Statist – Theory. Meth.*, **21**, 1779 – 1798.,1992

[3] Bhar ,Lalmohan, V. K. Gupta and Rajender Parsad, “Detection of outliers in designed experiments in presence of masking, Statistics and Applications”, Volume 11, Nos. 1&2, pp. 147-160, 2013 .

[4] Hubert,M. and E. Vandervieren, “An Adjusted Boxplot for Skewed Distributions”, Preprint submitted to Elsevier,30 October 2007.

[5] John, J. A. , “Outliers in Factorial experiments”. *Appl. Stat*., **27**, 111 – 119, 1978

[6] John, J. A. and Draper, N. R., “On testing for two outliers or one outlier in two-way tables”. *Technometrics*, **20**, 69 – 78,1978.

[7] Selinski ,Silvia and Claudia Becker ,” Outlier detection in experimental data using a modified Hampel identifier”,Technical Report, No. 40, Collaborative Research Center 'Reduction of Complexity in Multivariate Data Structures' (SFB 475), University of Dortmund,2001

[8] Seo, Songwon, M.S. ,”A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets”, University of Pittsburgh, 2006

[9] Stefan, Mandic-Rajcevic and Claudio Colosio” Methods for the Identification of Outliers and Their Influence on Exposure Assessment in Agricultural Pesticide Applicators: A Proposed Approach and Validation Using Biological Monitoring”, *Toxics* , *7*, 37; doi:10.3390/toxics7030037.,2019

[10] Zhao, chao and Jinyan yang, **“**A Robust Skewed Boxplot for Detecting Outliers in Rainfall Observations in Real-Time Flood Forecasting**,** Advances in Meteorology”, , Article ID 1795673, 7 pages https://doi.org/10.1155/2019/1795673,2019