International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 13s,2025
https://theaspd.com/index.php`

Effect of Outlier on Analysis of Variance

M. Shafiqur Rahman[1*], Syfun Nahar[2]

[1,2] Mathematics, Statistics and Computer Science, SNPS, University of Papua New Guinea, Port Moresby, Papua New Guinea
[1]srahman@upng.ac.pg; [2]syfun.nahar@upng.ac.pg

Abstract–This article discusses the effect of outliers on the analysis of variance. It is observed that outliers significantly increase the variation within groups, which leads to a significant change to F-statistic and resulting in a Type II error, where the null hypothesis is accepted incorrectly. Outliers significantly affect the P-values derived from ANOVA, potentially leading to misleading significance levels. Outliers violate the basic assumptions of ANOVA that the residuals are normally and independently distributed with mean 0 and constant variance, making the ANOVA results unreliable. One practical example is given.
Keywords– ANOVA, hypothesis test, outlier, P-Value, type II error.

I. INTRODUCTION

An outlier is an observation in a data set that is unusually small or large compared to the rest of the data in the data set. Statistical results can be affected by outliers which lead to wrong decisions. Hawkins [1] defined outlier as an observation in a data set that deviates so much from other observations in the data set and created suspicion that it was generated by a different mechanism. Outliers appear due to several reasons: mechanical faults, changes in system behaviour, fraudulent behaviour, human error, instrumentation error, or simply through natural deviation from a standard situation. Because of these, outlier detection has applications in areas such as fraud detection, network intrusion, and data cleaning. Outliers are usually removed to improve accuracy of the estimators. Barnett and Lewis [2], Hodge and Austin [3], and Markou and Singh [4,5] discussed outliers and presented various outlier detection techniques. Penny and Jolliffe [6] compared six statistical techniques for outlier detection. Let us consider a sample data set of 'n' observations of a variable y, let $\bar{y}$ be the sample mean and let 's' be the sample standard deviation. One observation is an outlier if it lies outside the interval ($\bar{y} - cs$, $\bar{y} + cs$), where the value of the arbitrary constant c is usually taken as 2 or 3. The justification of these values relies on the fact that assuming normal distribution one expects to have 95.45% chance that all data values will be within the interval ($\bar{y} - 2s$, $\bar{y} + 2s$) and 99.73% chance that all data values will be within the interval ($\bar{y} - 3s$, $\bar{y} + 3s$). The issue with the above criteria is that it assumes normal distribution of the data, something that does not usually exist. Moreover, the mean and standard deviation are extremely sensitive to outliers. Tukey [7] proposed Boxplot for exploratory data analysis where outliers are displayed. There are two types of outliers: mild outliers and extreme outliers. An observation y is an extreme outlier if it lies outside of the interval (Q_1-3×IQR, Q_3+3×IQR). Note that the centre of the interval is (Q_1+Q_3)/2, where IQR=Q_3-Q_1 is the Interquartile Range and can be considered a robust estimator of the variability which can replace 's'. On the other hand, (Q_1+Q_3)/2 is a robust estimator of the centre that can be used instead of $\bar{y}$. An observation y is a mild outlier if it lies outside of the interval (Q_1-1.5×IQR, Q_3+1.5×IQR). Ziller [8] showed that omitting an outlier leads to a decrease of the error variance. If the linear model is assumed, omitting an outlier leads to a decrease of the variance of the estimator and therefore to an increase of the accuracy of prediction. In this article an attempt has been made to study the effect of outliers on the analysis of variance by considering a one-way classification with and without outliers.

I taught a course entitled Experimental Design and Variance-Covariance Analysis in Semester 2 of 2025 at the University of Papua New Guinea. The following is an assignment question for this course. Four groups of salespeople for a magazine sales agency took four different sales training programs. Because there were dropouts during the training programs, the number of trainees varied from group to group. At the end of the training programs each salesperson selected randomly and assigned a sales area from a group of sales areas those have equivalent sales potentials. Table 1 lists the number of sales made by each person in each of the four groups of salespeople during the first week after completing the training program. Do the data present sufficient evidence to show a difference in the mean achievements for the four training programs at 5% level?

Table 1: Original data on number of sales

Group 1	Group 2	Group 3	Group 4

74	75	60	94
87	69	78	89
73	76	67	80
79	81	62	88
81	72	83	
69	79	76	
	90		
77.17	**77.43**	**70.83**	**87.75**

Here the average of all observations is $\bar{y}$ =77.48, standard deviation of all observations is s=8.79, then $\bar{y}$ -2s = 59.89, $\bar{y}$ +2s = 95.07, $\bar{y}$ -3s = 51.09, $\bar{y}$ +3s = 103.86, Q_1=72, Q_3 =83, IQR =11, Q_1-1.5IQR=55.5, Q_1-3IQR=39, Q_3+1.5IQR=99.5, and Q_3+3IQR=116. Therefore, there is no mild or extreme outlier in the data set.Most of the students solved the assignment question correctly. One student made a mistake while inserting data and wrote one observation 881 instead of 81 in Group1 which is unusually large compared to other data values and solved the problem and came up with wrong conclusion. Another student made a mistake while inserting data and wrote one observation 18 instead of 81 in Group1 which is unusually small compared to other data values and solved the problem and came up with similar wrong conclusion. This motivated me to check how it happened. We found 881 and 18 both are extreme outliers. The objective of this study was to investigate the effect of outliers on Analysis of variance and how to tackle these issues.

II. CONTRIBUTION AND METHODS

Halldestam [9] concluded that the parameter estimates used in the one-way analysis of variance are not robust against outliers. He claimed that one single observation may cause the estimate to deviate exceptionally far from the true value. One-way ANOVA cannot be robust as outliers affect the type-I error probability. Scariano and Davenport [10], Hoaglin et.al. [11], Huber [12], Krishnaiah [13] studied the effect of outliers and concluded that outliers affect the type-I error. In this article we have shown that outlier affects type-II error in ANOVA of one-way classification which is our contribution. We also proposed two methods to tackle these issues.Here a practical example is considered where an outlier is detected using traditional methods and observed how it is affecting the ANOVA test results. Then the outlier is removed and conducted the ANOVA test for the reduced data set. Kruskal Wallis test was also applied to test the same hypothesis in presence of outlier. We observed similar results in both cases.

III. RESULTS AND DISCUSSIONS

Outliers affect the group/treatment means, the F-statistic, the P-values and violate the basic assumptions of ANOVA. Statistical results can strongly be affected by outliers. Therefore, such outliers are sometimes omitted to perform a more robust analysis. One-way classification refers to the comparison of several treatment means. Suppose there are k independent random samples of sizes $n_1, n_2, \ldots, n_k$ from k normal populations with unknown means $\mu_1, \mu_2, \ldots,$ μ_k and with a common unknown variance σ^2. We test the hypothesis that $\mu_1 = \mu_2 = \ldots = \mu_k$. Let y_{ij} be the j^{th} observation in the i^{th} sample. So, the classification scheme is given in Table 2.

Table 2: Layout of one-way classification

Treatments	1	2	...	i	...	*k*
Observations	y_{11}	y_{21}	...	y_{i1}	...	y_{k1}
	y_{12}	y_{22}	...	y_{i2}	...	y_{k2}
	.	.	...	.	...	.
	y_{1n_1}	y_{2n_2}	...	y_{in_i}	...	y_{kn_k}
Mean	$\bar{y}_{1.}$	$\bar{y}_{2.}$	...	$\bar{y}_{i.}$	...	$\bar{y}_{k.}$

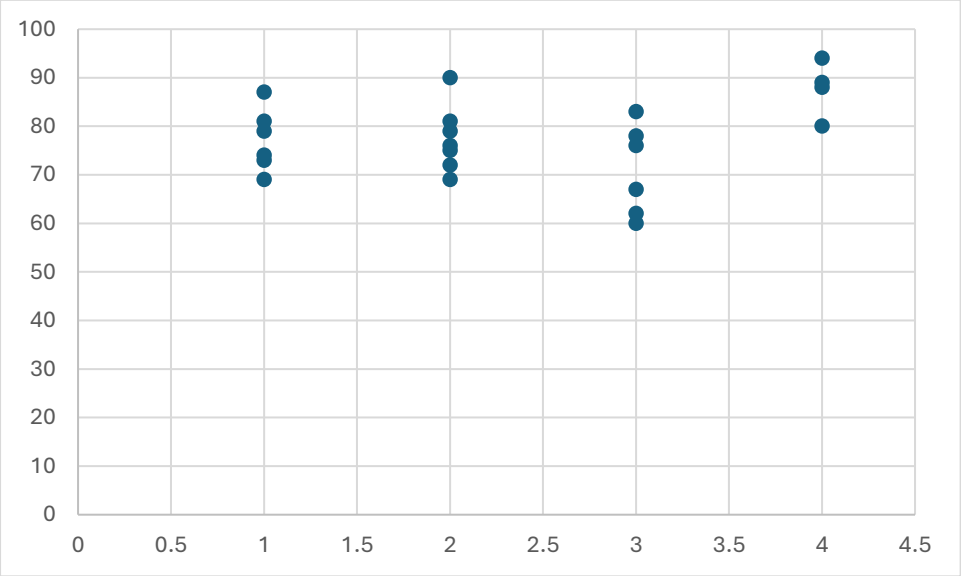
Here $\bar{y}_{i.}$ is the mean of all observations in the i^{th} treatment. Let $\bar{y}_{..}$ be the grand mean of all observations in the whole data set, and $\sum n_i = n$.

We may consider that these k treatments are the only treatments in which we are interested. Let the observations y_{ij} follow the linear model $y_{ij} = \mu_i + u_{ij}$, i= 1,2, . . .,k and j=1,2, . . .,n_i. This model can also be written as $y_{ij} = \mu + (\mu_i - \mu) + u_{ij} = \mu + \alpha_i + u_{ij}$, where μ is the general mean, α_i is the differential effect due to the i^{th} treatment and u_{ij} is the random error component. Let us assume that u_{ij} are distributed normally and independently with mean 0 and constant variance σ^2. To test the hypothesis H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$ against H_1: Not all means are equal, we need to construct the following ANOVA Table 3.

Table 3: ANOVA Table for one-way classification

Source of variation	df	Sum of squares (SS)	MSS	F ratio
Treatment	k-1	$\sum n_i(\bar{y}_{i.} - \bar{y})^2$	$\frac{\sum n_i(\bar{y}_{i.} - \bar{y})^2}{k-1}$	$\frac{(n-k)\sum n_i(\bar{y}_{i.} - \bar{y})^2}{(k-1)\sum\sum(y_{ij} - \bar{y}_{i.})^2}$
Error	n-k	$\sum\sum(y_{ij} - \bar{y}_{i.})^2$	$\frac{\sum\sum(y_{ij} - \bar{y}_{i.})^2}{n-k}$	
Total	n-1	$\sum\sum(y_{ij} - \bar{y})^2$		

The graphical representation of the data given in Table 1 is given Graph 1.



Graph 1: Scatter plot for the original data

To test the significant differences between the mean achievements for the four training programs, we need to construct the following ANOVA Table 4.

Table 4: ANOVA for original data

Source	SS	df	MS	F	P-value	F crit
Between	674.442	3	224.814	4.158	0.020	3.127
Within	1027.298	19	54.068			
Total	1701.740	22				

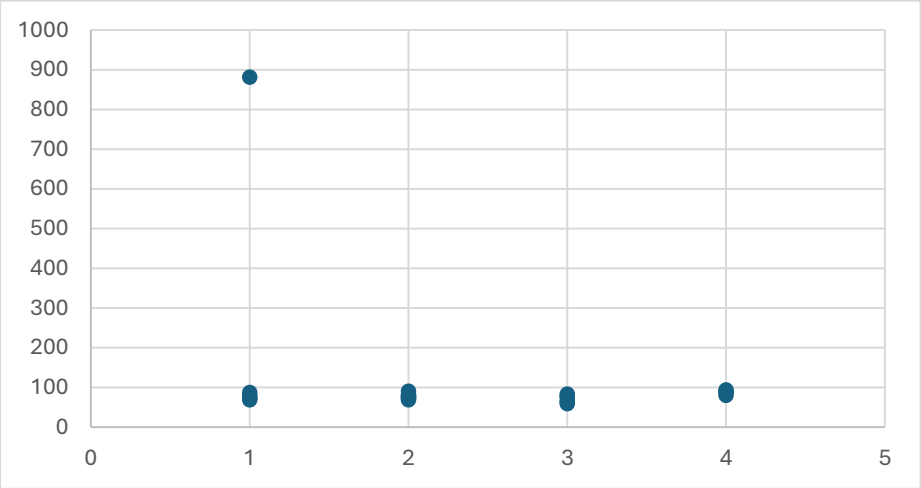
H_0: There are no significant differences in the mean achievements for the four training programs.
H_1: There are significant differences in the mean achievements for the four training programs.
Test statistic F=4.158

P-value = 0.02
Decision Rule: Reject H_0 if P-value < 0.05.
Conclusion: As P-value = 0.02 < 0.05, reject H_0. That is there are significant differences in the mean achievements for the four training programs at 5% level.
This test can also be done in another way.
Decision Rule: Reject H_0 if F >3.127.
Conclusion: As F =4.158 > 3.127, reject H_0. That is there are significant differences in the mean achievements for the four training programs at 5% level.Most students typed data and solved this assignment question correctly. One student, while typing data made a mistake and wrote one observation 881 instead of 81 in Group1 which is unusually large compared to other data values. Table 5 is for the data with an unusually large observation.

Table 5: Data with an unusually large observation

Group 1	Group 2	Group 3	Group 4
74	75	60	94
87	69	78	89
73	76	67	80
79	81	62	88
881	72	83	
69	79	76	
	90		
210.50	77.43	71.00	87.75

Here the average of all observations is $\bar{y}$ =112.26, standard deviation of all observations s=167.81, $\bar{y}$ +2s = 447.88, $\bar{y}$ +3s = 615.69, Q_1=72, Q_3 =87, IQR =15, Q_1-1.5IQR=49.5, Q_1-3IQR=27, Q_3+1.5IQR=109.5, and Q_3+3IQR=132. Therefore, observation 881is an extreme outlier.



Graph 2: Scatter plot for an unusually large data value

Table 6: ANOVA table with large outlier

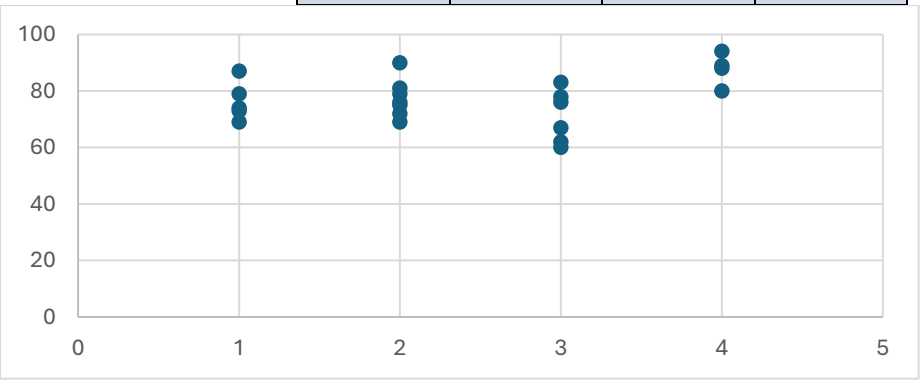
Source	SS	df	MS	F	P-value	F crit
Between	79016.47	3	26338.82	0.926	0.447	3.127
Within	540494.00	19	28447.05			
Total	619510.47	22				

The test results for the data set with an extreme outlier are as follows:
H_0: There are no significant differences in the mean achievements for the four training programs.
H_1: There are significant differences in the mean achievements for the four training programs.
Test statistic F=0.926
P-value = 0.447

Decision Rule: Reject H_0 if P-value < 0.05.
Conclusion: As P-value = 0.447 > 0.05, do not reject H_0. That is there are no significant differences in the mean achievements for the four training programs at 5% level.
This can also be done in another way.
Decision Rule: Reject H_0 if F >3.127.
Conclusion: As F =0.926 < 3.127, do not reject H_0. That is there are no significant differences in the mean achievements for the four training programs at 5% level.
If we look at the group means, there are substantial differences between the group means. But the test results are opposite. One outlier affected the test results and leads to a wrong decision. Here we have accepted null hypothesis but looking at the treatment means alternative hypothesis should be the right choice and hence leading to a Type II error. We remove the outlier and the reduced data is presented in Table 7.

Table 7: Reduced data by removing the large outlier

Group 1	Group 2	Group 3	Group 4
74	75	60	94
87	69	78	89
73	76	67	80
79	81	62	88
69	72	83	
	79	76	
	90		
76.40	77.43	71.00	87.75



Graph 3: Scatter plot for the reduced data

Table 8: ANOVA table for the reduced data

Source	SS	df	MS	F	P-value	F crit
Between	679.108	3	226.369	4.036	0.023	3.160
Within	1009.664	18	56.092			
Total	1688.772	21				

The test results for the data set after dropping the outlier are as follows:
H_0: There are no significant differences in the mean achievements for the four training programs.
H_1: There are significant differences in the mean achievements for the four training programs.
Test statistic F=4.036
P-value = 0.023
Decision Rule: Reject H_0 if P-value < 0.05.
Conclusion: As P-value = 0.023 < 0.05, reject H_0. There are significant differences in the mean achievements for the four training programs at 5% level.
This can also be done in another way.
Decision Rule: Reject H_0 if F >3.160.
Conclusion: As F =4.036 > 3.160, reject H_0. That is there are significant differences in the mean achievements for the four training programs at 5% level.
Instead of removing the unusual observation or outlier, Kruskal Walli's Test can be applied. Replacing the observations by their corresponding ranks the following Table 9 is obtained.

Table 9: Kruskal Walli's test statistic value calculation

	Group 1	Group 2	Group 3	Group 4	Total
	8	9	1	22	
	18	4.5	12	20	
	7	10.5	3	15	
	13.5	16	2	19	
	23	6	17		
	4.5	13.5	10.5		
		21			
R_i	74	80.5	45.5	76	276
n_i	6	7	6	4	23
$\frac{R_i^2}{n_i}$	912.67	925.75	345.04	1444.00	3627.46

The test results of Kruskal Wallis' test are as follows:
H_0: There are no significant differences in the mean achievements for the four training programs.
H_1: There are significant differences in the mean achievements for the four training programs.

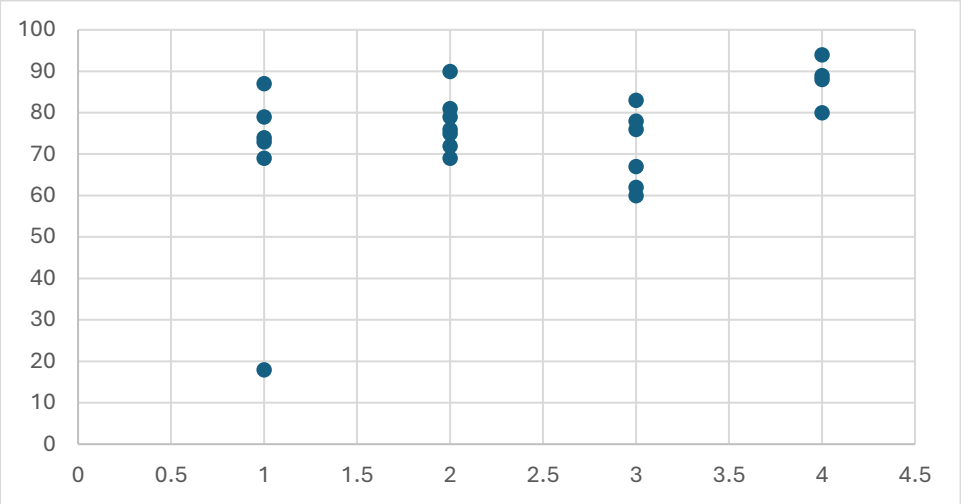
The test statistic $T = \frac{12}{N(N+1)} \sum \frac{R_j^2}{n_j} - 3(N+1) = \frac{12 \times 3627.46}{23 \times 24} - 3 \times 24 = 6.858$

The critical value at 5% level is 5.991.
Decision Rule: Reject H_0 if $T > 5.991$.
Conclusion: As $T = 6.858 > 5.991$, reject H_0. There are significant differences in the mean achievements for the four training programs at 5% level.
Another student, while typing data made a mistake and wrote one observation 18 instead of 81 in Group1 which is unusually small compared to other data values. Table 10 is for the data with an incorrect observation.

Table 10: Data with an unusually small observation

Group 1	Group 2	Group 3	Group 4
74	75	60	94
87	69	78	89
73	76	67	80
79	81	62	88
18	72	83	
69	79	76	
	90		
66.67	77.43	71.00	87.75

Here the average of all observations is $\bar{y}$ =74.74, standard deviation of all observations is s=15.16, then $\bar{y}$ -2s = 44.42, $\bar{y}$ -3s = 29.27, $\bar{y}$ +2s = 105.05, $\bar{y}$ +3s = 120.21, Q_1=69, Q_3 =83, IQR =14, Q_1-1.5IQR=48, Q_1-3IQR=27, Q_3+1.5IQR=104, and Q_3+3IQR=125. Therefore, observation 18 is an extreme outlier.



Graph 4: Scatter plot for the data with unusually small observation

Table 11: ANOVA table with a small outlier

Source	SS	df	MS	F	P-value	F crit
Between	1202.637	3	400.879	1.977	0.152	3.127
Within	3851.798	19	202.726			
Total	5054.435	22				

The test results for the data set with a small outlier are as follows:
H_0: There are no significant differences in the mean achievements for the four training programs.
H_1: There are significant differences in the mean achievements for the four training programs.
Test statistic F=1.977
P-value = 0.152
Decision Rule: Reject H_0 if P-value < 0.05.
Conclusion: As P-value = 0.152 > 0.05, do not reject H_0. That is there are no significant differences in the mean achievements for the four training programs at 5% level.
This can also be done in another way.
Decision Rule: Reject H_0 if F >3.127.
Conclusion: As F =1.977 < 3.127, do not reject H_0. That is there are no significant differences in the mean achievements for the four training programs at 5% level.
If we look at the group means, there are significant differences between the means. But the test results are opposite. One small outlier affected the test results and leads to a wrong decision and Type II error. If we remove the small outlier, then the reduced data will be like Table 7 for which the test results are given before.
Instead of removing the unusually small observation or outlier, Kruskal Walli's Test can be applied. Replacing the observations by their corresponding ranks the following Table 12 is obtained.

Table 12: Kruskal Walli's test statistic value for data with small outlier

	Group 1	Group 2	Group 3	Group 4	Total
	9	10	2	23	
	19	5.5	13	21	
	8	11.5	4	16	
	14.5	17	3	20	
	1	7	18		
	5.5	14.5	11.5		
		22			
R_i	57	87.5	51.5	80	276

n_i	6	7	6	4	23
$\frac{R_i^2}{n_i}$	541.50	1093.75	442.042	1600	3677.29

The test results of Kruskal Wallis' test are as follows:
H_0: There are no significant differences in the mean achievements for the four training programs.
H_1: There are significant differences in the mean achievements for the four training programs.

The test statistic $T = \frac{12}{N(N+1)} \sum \frac{R_j^2}{n_j} - 3(N+1)$

$= \frac{12 \times 3677.29}{23 \times 24} - 3 \times 24$

$= 7.941$

The critical value at 5% level is 5.991.
Decision Rule: Reject H_0 if $T > 5.991$.
Conclusion: As $T = 7.941 > 5.991$, reject H_0. There are significant differences in the mean achievements for the four training programs at 5% level.
Therefore, the test results of the original data, the test results of the reduced data after eliminating outliers and the test results of Kruskal Wallis' test are similar.

IV. CONCLUSION

Outliers significantly affect ANOVA estimates, potentially leading to inflated type-II error rates and distorted parameter estimates. In ANOVA, outliers can increase the variance, making the data appear more spread out than it is, which can lead to incorrect conclusions about the data's reliability. Previous studies found that outliers affect type-I error but here we found that outliers affect type-II error. In ANOVA outliers can be handled in two ways: (i) remove the outlier and conduct the usual ANOVA, or, (ii) conduct Kruskal Wallis' test without removing outlier.

REFERENCES

[1]. Hawkins, D., 1980, *Identification of Outliers*, Chapman and Hall, London, UK

[2]. Barnett, V. and Lewis, T., 1994, *Outliers in Statistical Data,* John Wiley, New York

[3]. Hodge, V. J. and Austin, J., 2004, *A survey of Outlier Detection Methodologies, Artificial Intelligence Review*, Kluwer Academic Publishers, 22: 85-126

[4]. Markou, M. and Singh, S., 2003, Novelty Detection: A Review, Part I: Statistical Approaches, *Signal Processing*, 83, 2481-2497

[5]. Markou, M. and Singh, S., 2003, Novelty Detection: A Review, Part II: Neural Network Based Approaches, *Signal Processing*, 83, 2499-2521

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

[7]. Tukey, J.W., 1977, *Exploratory Data Analysis,* Addison-Wesley, Boston, USA

[8]. Ziller, M., 2003, Quantifying the relative accuracy of data-based prediction, *Proc. ASIM* 2003, *9-12.* June, Almaty, Kasachstan, 486-489.

[9]. Halldestam, M., 2016, ANOVA-The effect of Outliers, Bachelor's thesis, Department of Statistics, Uppsala University, Sweden

[10]. Scariano, S. M., and Davenport, J. M., 1987, The effects of violations of independence assumptions in the one-way ANOVA. *The American Statistician, Vol. 41, No. 2 (May)*, 123-129.

[11]. Hoaglin, D. C., Mosteller, F., and Tukey, J. W., 1983, *Understanding Robust and Exploratory Data Analysis.* John Wiley.

[12]. Huber, P. J., 1981, *Robust Statistics.* New York: John Wiley.

[13]. Krishnaiah, P. R., 1980, *Analysis of variance.* Amsterdam; New York: North-Holland Pub. Co.