IASJ المجلات الاكاديمية العراقية

Tikrit Journal of Pure Science

ISSN: 1813 – 1662 (Print) --- E-ISSN: 2415 – 1726 (Online)

Journal Homepage: http://tjps.tu.edu.iq/index.php/j

Detecting Outliers in the Simple Linear Regression for Children Affected with Leukemia in Mosul City

Shaymaa Riyadh Thanoon[1] , Mohammed Nafe Abd Alrazzaq[2]

[1]*department Basic Sciences, College Of Nursing, Mosul University, Mosul, Iraq.*

[2]*directorarate Of education, Mosul, Iraq.*

https://doi.org/10.25130/tjps.v27i3.47

ARTICLE INFO.

Article history:

-Received: 12 / 2 / 2022

-Accepted: 14 / 4 / 2022

-Available online: / / 2022

Keywords: Outliers, median, mean absolute deviation about the median, regression analysis.

Corresponding Author:

Name: Shaymaa Riyadh Thanoon

E-mail:

shaymaa.riadh@uomosul.edu.iq

mohammednafe@yahoo.com

Tel:

ABSTRACT

In this study, the problem of outlier detection in "linear regression" analysis is studied using the "median" and "mean" "absolute deviation" about the median. "The mean and standard deviation" are heavily affected by outliers Hence, the outlier detection techniques based on these measures may not correctly identify all outliers in a dataset. However, "the mean absolute deviation about the median", in combination with the median is sufficiently robust in the presence of outliers and provides a better alternative. The conceptualized method was tested using leukemia patients data and the results indicate that the new method performed better than the methods based on the mean/ standard deviation combination. It is recommended that the median and "mean absolute deviation" about the median be used in detecting outliers in regression analysis due to their inherent potential for increasing the" goodness-of-fit of the "linear regression mode".

1. Introduction

In the beginning, we must give a simple overview of this disease, its causes and methods of treatment. It is a genetic disorder in the blood cells. It is described as low hemoglobin level, and a lower number of red blood cells than the normal range. There are other names for it, which are Mediterranean anemia. As for the types of leukemia.The type of leukemia depends on the number of genetic mutations, and on the affected part; Where the mutation occurs in one of the parts of hemoglobin alpha or beta this disease is that leukemia occurs due to a genetic mutation in the DNA of the cells that make up hemoglobin, and this mutation is genetically transmitted from parents to children; This causes the occurrence of genetic mutations to disrupt the production of normal hemoglobin, and thus the low levels of hemoglobin, and the high rate of damage The presence of outliers in bivariate data can significantly alter the conclusions drawn from linear regression analysis. Their presence in a dataset increases the risk of making a wrong decision. Regression analysis provides a way of examining the presence of linear relationships between a "dependent variable and independent variables".

Outliers could be encountered in practice due to several reasons that could be categorized into two groups –"those arising from errors in the data and those arising from the inherent variability in the process yielding the data" [1]. It is therefore important to identify outliers and establish the root cause of the outlying observations through the provision of background information [2];. The detection of outliers is still an area of ongoing research. This study is thus aimed at providing an alternative method for outlier detection using "the mean absolute deviation about the median in linear regression analysis". The objectives were to develop the technique for detecting outliers in "linear regression analysis using the median and the mean absolute deviation about the median" and to evaluate the technique in comparison with the use of the standardized sores based on the mean/ "standard deviation".

Leukemia is a genetic disorder of the blood that is prevalent in Asia, Middle East and other parts of the world. There had been a raft of approaches in the epidemiology, treatment and prognosis of the disease which have dramatically improved on the care of affected patients [3]. It has been an upward trajectory

in the development of medical and non-medical responses to the disease, and these changes are redefining the clinical management of leukemia .

The" mean absolute deviation about the median", MAD(md) is a direct measure of the dispersion of a random variable from its" median". If the "mean absolute deviation" is the preferred measure of variability, then the median is a better measure of central tendency especially for asymmetric distributions [4]. In such instances, the median is a better representation of the center of the distribution than the mean. The MAD(md) is different from the "median absolute deviation" MAD, and it is a more representative measure of the variability of sample observations.

For asymmetrical distributions, the MAD(md) provides a more meaningful dispersion measure related to the center of the distribution since it is sufficiently impervious to outliers more than the "standard deviation". The MAD(md) tends to increase with the size of the sample though not proportionately and not rapidly as the range [5].

In the next section a review of some of the important works in the field is undertaken. Section 3 outlines the materials and methods used in the study, while section 4 contain the results and discussion. Finally, the conclusion in provided in section 5.

Outliers and how they can be detected has been an area of vigorous research. There are several methods available in the literature, and some use graphical tools like boxplots [6]. For the case of multivariate data, the box-plotted Mahalanobis distances have been used to identify outlying observations, and the removal of these outliers can increase the descriptive classification accuracy of statistical models. An overview of outlier detection tools for univariate, low-dimensional, and high-dimensional data was presented [7]. In the field of medical research, [8] outlined a data-driven approach for detecting anomalous patient management decisions using past patient cases stored in electronic health records. Outlier-based alerting in medical practice could significantly improve the true alert rates. We must have a pause through detecting Outliers in the Simple Linear Regression for Children Affected with Leukemia In Mosul City.

The concept of data preprocessing has been employed as an outlier detection tool in medical data [9]. Distance-based and cluster-based outlier detection algorithms have also been developed for detecting and removing outliers. used robust regression analysis for structural health monitoring data which was capable of being unaffected by outliers. [10] used a modified Grubb's method for outlier detection based on the median and median absolute deviation applied in multiple measuring points parameters. indicated that the modified Grubb's method was more efficient and robust than the original Grubb's approach.

The MAD(md) has been presented as an alternative to the variance when there are departures from the normality assumption [10]. Approximate confidence intervals for the mean absolute deviation about the median in one-sample and two-sample designs have been determined. Simulation results have also shown that the confidence intervals had coverage probabilities close to the true confidence interval in mild leptokurtic and mild skewed distributions.

There are some test of hypothesis procedures based on the mean absolute deviation about the median and other alternative measures of scale [12] It is also claimed that these alternative procedures could produce superior Type I and Type II error probabilities. The mean absolute deviation about the median is a sufficiently robust measure of the scale and shape of a probability distribution [13]. The median is increasingly being used in various areas of research and recently, it has been applied to correlation analysis via the MAD(md) correlation. An elegant and computationally easier estimator of the mean absolute deviation about the median (MAD(md)) was developed by [14]. estimated of the MAD(md) bypasses the use of the absolute operator and provides a straightforward way of computing the statistic in both grouped and ungrouped data.

Rosner's and Grubb's test was used to detect outliers in linear regression analysis [1]. An observation could be an outlier in either the response or explanatory variable, yet be influential in the dataset. In simple linear regression analysis, the standardized score for the dependent and independent variable have been used to detect outliers in medical data [15]. It is suggested that the use of only the residuals and standardized residuals for outlier detection could be insufficient in identifying all the outliers in a bivariate data.

Much research has been done on establishing statistical relationships between variables related to leukemia. [16] investigated the contributions of known modifiers to the prediction of the clinical severity of beta-leukemia using the patient's age at first transfusion. A positive correlation between bone mass and hemoglobin levels of leukemia patients was shown in a study by [17], established a negative correlation between hemoglobin levels and ferritin levels of leukemia patients. [18],showed that hemoglobin levels and ferritin levels in leukemia patients were negatively correlated. In the study on maxillofacial anomalies in leukemia patients, a positive correlation between the frequency of anomalies and the ferritin levels while a negative correlation was observed between the frequency and the hemoglobin levels.

2. Materials and Methods

Regression analysis is describe used to the nature of the relationship between variables," In simple regression, there are two variables – the independent or explanatory variable, and the dependent or response variable. The independent variable (X) is

used to predict the dependent variable (Y). In multiple regression", there are two or more independent variables that are used to predict the dependent or response variable.

A "simple linear regression model" is given as

$Y = \beta_0 + \beta_1 X + \varepsilon .. (1)$

where

β_0 is called the intercept." It is the value of the dependent variable "(Y) for which X=0.

β_1 is the slope or gradient. "It is the amount of change in Y for a unit change in X.

ε is the error variable",

The method of least squares is the most effective way of estimating the regression parameters, and the estimates ($\hat{\beta}_0$ and $\hat{\beta}_1$) are "obtain by minimizing the error sum of squares "[19]

$\hat{\beta}_1 = \frac{\sum x_i y_i - (\sum x_i \sum y_i)/n}{\sum {x_i}^2 - (\sum x_i)^2/n} = \frac{n\sum x_i y_i - (\sum x_i \sum y_i)}{n\sum {x_i}^2 - (\sum x_i)^2} .. (2)$

and

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} .. (3)$

Thus, the" estimated least squares regression equation" is

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i .. (4)$

Generally, the "mean absolute deviation about the median" ($MAD(md)$), the mean absolute deviation about mean ($MAD(\mu)$) and the standard deviation (σ) for any distribution can be related via a corollary of Lyapounov's inequality [20]

$MAD(md) \leq MAD(\mu) \leq \sigma .. (5)$

The MAD is thus the least among the three measures of spread for any dataset. The MAD can be expressed elegantly in terms of the sums above and below the sample median, thus bypassing the absolute operator [14].a theorem from [14] is now presented without proof.

Theorem 1:" Let $x_1, x_2, \cdots, x_n$ be a random sample of size n with median" $md(x)$. Let b represent the sum of all observations below the median, a, the sum of all observations above the median and T_n, the "sum of all the observations. Define the indicator function"

$$I_n = \begin{cases} 1, & if\ n\ is\ odd \\ 0, & if\ n\ is\ even \end{cases}$$

Then the "mean absolute deviation about the median", $MAD(md)$ is given as

$MAD(md) = \frac{1}{n}(a - b) = \frac{1}{n}(T_n - 2b - I_n md(x)) .. (6)$

The methods of outliers detection in linear regression analysis to be considered in this study are based on the residuals, standardized scores of the dependent and independent variables based on the mean/ standard deviation and the method based on the standardized score of the variables using the median/ MAD(md) combination. These procedures are outlined below.

Given a linear regression model estimated from data, the residuals are given as $e_i = y_i - \hat{y}_i$, which is a measure of the variability in the prediction. The residuals are assumed to be distributed normally with 0 mean and constant variance. If the residual is standardized, then it becomes

$d_i = \frac{e_i}{\sqrt{MSE}}, i = 1,2, \cdots, n .. (7)$

where MSE is the mean squared error and is given by

$MSE = \frac{\sum e_i^2}{n-1}$

Values of d_i lying outside ± 3 will then be classified as outliers and the corresponding pairs (x_i, y_i) of the sample observations are excluded from the subsequent regression analysis done on the data.

The second method of outlier detection is via the use of the standardized score based on the mean/ standard deviation for both the dependent and independent variable as applied by [15]. The standardized score is obtained by subtracting the mean ($\bar{y}$ or $\bar{x}$) from each observation and dividing by the standard deviation (s_y or s_x), and are given by equations (8) and (9) below.

$\frac{y_i - \bar{y}}{s_y}, i = 1,2, \cdots, n .. (8)$

$\frac{x_i - \bar{x}}{s_x}, i = 1,2, \cdots, n .. (9)$

Values of the standardized scores lying outside ± 3 are also classified as outliers and the corresponding pairs excluded from the subsequent regression analysis.

Finally, the third method of outlier detection in this study will be the standardized scores of the median/MAD(md) for both the dependent and independent variables. This is given as

$\frac{y_i - median(y)}{MAD(md(y))}, i = 1,2, \cdots, n .. (10)$

$\frac{x_i - median(x)}{MAD(md(x))}, i = 1,2, \cdots, n .. (11)$

Similar to the standardized score based on the mean and standard deviation, observations having standardized scores lying outside ± 3 are excluded pairwise from the bivariate data. The test is carried both on the dependent and independent variables to detect the outlying observations.

The MAD(md) standardized scores yields a confidence interval of $md(y) \pm 3MAD(md)$. It uses the median and the MAD(md) as the measure of location and spread respectively rather than the mean and standard deviation. There is an inherent justification for the choice of the median and the MAD(md). Outliers heavily affect the mean and standard deviation; hence any outlier technique relying on these two statistics may be unable to detect some observations which may be outliers in reality. On the other hand, the median is sufficiently robust to outliers and the MAD(md), which is a generally lower measure than the standard deviation, provides a more stable outlier detection approach.

3. Results and Discussion

Data on the age in years (X) and serum hemoglobin levels in g/dl (Y) of 108 leukemia patients were used in the study the data is as shown in table 1.

Table 1: leukemia patients samples

N0.	Age years	Hb Level	NO.	Age years	Hb Level	NO.	Age years	Hb Level
1	12	8	37	8	6	72	2	6
2	1	8	38	1	7	73	1	6
3	22	7	39	14	7	74	1	7
4	26	7	40	14	7	75	2	7
5	10	6	41	15	8	76	16	7
6	16	6	42	11	8	77	12	8
7	16	7	43	15	11	78	15	8
8	11	7	44	17	11	79	10	11
9	13	7	45	14	4	80	3	11
10	11	8	46	1	4	81	2	4
11	9	8	47	1	7	82	1	4
12	7	11	48	2	7	83	11	8
13	8	11	49	3	7	84	1	8
14	10	4	50	3	8	85	2	7
15	8	4	51	2	8	86	2	7
16	11	8	52	3	11	87	2	6
17	5	8	53	12	8	88	16	6
18	7	7	54	1	8	89	16	7
19	14	7	55	2	7	90	2	7
20	15	6	56	2	7	91	10	7
21	16	6	57	2	6	92	12	8
22	16	7	58	1	6	93	2	8
23	11	7	59	1	7	94	3	11
24	10	7	60	5	7	95	3	11
25	7	8	61	10	7	96	2	4
26	4	8	62	2	8	97	1	4
27	6	11	63	9	8	98	16	7
28	11	11	64	5	11	99	2	7
29	22	4	65	8	11	100	13	7
30	1	4	66	14	4	101	12	8
31	12	8	67	17	4	102	2	8
32	1	8	68	3	8	103	2	18
34	9	7	69	7	8	104	13	19
35	16	7	70	10	7	105	3	23
36	8	6	71	11	7	106	3	25
						107	1	3
						108	13	6

Clinically, the range of hemoglobin levels in leukemia patients lie between 3g/dl and 11g/dl. Lower hemoglobin levels indicate increasing severity of the disease, and such patients may require regular blood transfusions. On the other end of the spectrum, higher levels of hemoglobin could indicate a non-leukemia patient or a possibly wrong hemoglobin measurement for the patient. A linear regression model of the serum hemoglobin measurement against the age of the leukemia patients was fitted to the original data. The method of standardized residuals was then used to identify the outliers in the dataset and another regression analysis was done with the outliers deleted from the data. This procedure is also replicated using the standardized scores of the dependent (Y) and independent (X) variables and the standardized scores based on the median/MAD(md) combination for both the independent and dependent variables. In Figure 1, the scatterplot of the original data is presented. The descriptive summary of the original data and for the cases where the outliers have been removed, are presented in Table 2. And the result of the tests that were used to find outliers shown in table 3.

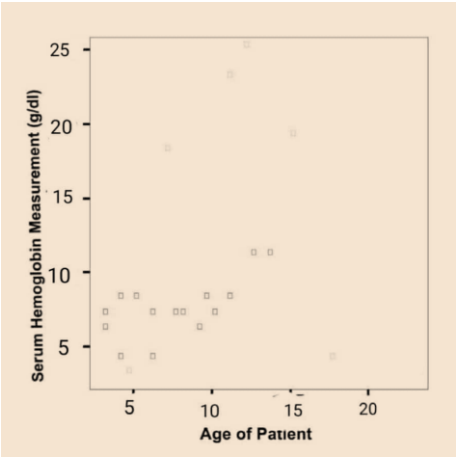

Fig. 1: Scatterplot of serum hemoglobin levels against the age of leukemia patients

The scatterplot of the dataset as shown in Figure 1 indicates a possible linear relationship and the presence of some outliers. Six observations are somewhat outliers and this could point to the possibility that they are outliers. However, a formal statistical technique is required to establish outlying observations. The linear regression model is then fitted on the original data and the three outlier detection techniques are applied on the data with the aim of improving the goodness-of-fit of the regression model.

Table 2: Descriptive statistics of the original leukemia patients data and with the outliers removed using three different methods.

	Original data	Standardized Residuals	Mean/Standard deviation Standardized Score	Median/MAD(md) Standardized Score
Mean	7.722	7.356	7.346	7.275
Median	7.0	7.0	7.0	7.0
Minimum	3	3	3	3
Maximum	25	19	18	11
Range	22	16	15	8
Standard deviation	3.3288	2.2292	2.1801	1.8992
MAD(md)	1.8148	1.4135	1.4038	1.2941

Table 3: the results of the tests that were used to find outliers

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	3.3415412	1.074684	3.10932	0.00240	1.21087	5.47220	1.21087	5.47220
Age	0.1693926	0.039931	4.24208	4.75314E	0.09022	0.24856	0.09022	0.24856

Furthermore, the estimates of the regression model, goodness-of-fit measures and the predicted intervals are presented in Table 4.

Table 4: Summary output of the dataset with outliers identified through the standardized residuals, standardized scores due to the mean and standardized scores due to the median.

	Original data	Standardized Residuals	Standardized Scores (mean/ standard deviation)	Median/MAD(md) Standardized Score
Outliers detected	-	4	4	6
Intercept	3.3415	2.6011	3.3944	2.545
Slope	0.1694	0,1868	0.1562	0.1883
R^2	0.1451	0.3384	0.2369	0.4242
R^2_{adj}	0.1371	0.332	0.2294	0.4184
Predicted Interval	-	-	(-2,18)	(2,13)

The median/MAD(md) method yielded more compact data than the original data and the two other methods. This can be seen from the ranges in Table 1 and the predicted intervals in Table 4. The median remained unchanged across all the methods and in the original data. This reflects the inherent property of the median in terms of being highly unaffected by outliers. The median may shift drastically only if a large number of outliers are present in the data. As seen in Table 2, the MAD(md) had a much smaller value than the standard deviation in all instances. In addition, there was a progressive convergence in the values of both the MAD(md) and standard deviation when the median/MAD(md) approach is used to detect and remove outliers from the data. This also means that these three methods will yield similar results in the presence of mild variability in the data since the mean will be close to the median and the standard deviation close to the MAD(md).

A progressive improvement in the" goodness-of-fit of the linear regression model" was observed starting with the original data, then the standardized scores based on the mean/ standard deviation, the standardized score based on the residuals, and finally the new method based on the median/MAD(md). As seen from Table 4, the percentage of variability in the data that is explained by the regression model rose from 13.71% when the original data was used, up to 41.84% when the outliers were detected by the median/MAD(md) method were removed. This shows a remarkable improvement in the goodness-of-fit of the regression model. It also indicated that the median/MAD(md) method provided the best approach for detecting outlier more than the other methods used in the study.

4. Conclusions and Recommendations

In this paper, the problem of outlier detection in linear regression analysis is studied using the median and mean absolute deviation about the median Data on the age in years (X) and serum hemoglobin levels in g/dl (Y) of 108 leukemia patients were used in the study the data is as shown in table 2. Clinically, the range of hemoglobin levels in leukemia patients lie between 3g/dl and 11g/dl. Lower hemoglobin levels

indicate increasing severity of the disease, and such patients may require regular blood transfusions. On the other end of the spectrum, higher levels of hemoglobin could indicate a non-leukemia patient or a possibly wrong hemoglobin measurement for the patient. A progressive improvement in the goodness-of-fit of the linear regression model was observed starting with the original data, then the standardized scores based on the mean/ standard deviation, the standardized score based on the residuals, and finally the new method based on the median/MAD(md). It also indicated that the median/MAD(md) method provided the best approach for detecting outlier more than the other methods used in the study. After studying the research, we reached some important results, including:

One of these conclusions is the use of one of the methods for determining and testing the divergence of anomalous values in any statistical analysis, In addition to these results, there are some suggestions that can be worked on in later research, where one of the methods of testing the divergence for abnormal values can be used in any statistical analysis, and this proposed method can also be used to detect other diseases.

References

[1] Ogu, A. I., Inyama, S. C., and Achugamonu, P. C. **2013**. "Methods of detecting outliers in a regression analysis model". *West African Journal of Industrial and Academic Research*, 7(1), 105-112.

[2] Dervillis, N., Worden, K. and Cross, E. J. **2015**. "On robust regression analysis as a means of exploring environmental and operational conditions for SHM data". *Journal of Sound and Vibration*, 347, 279-296.

[3] Cohen, A. R., Galanellor, R., Pennel, C. J., Cunningham, M. J., and Vichinsky, E. **2004**. Leukemia. ASH Program Book (1), 14-34.

[4] Pham-Gia, T. and Hung, T. L. **2001**. The mean and median absolute deviations. *Mathematical and Computer Modelling*, 34, 921-936.

[5] Gupta, S. C. **2011**. Fundamentals of Statistics. Himalaya Publishing House. India.

Hameed, M., Raziq, F., and Mir, A. **2020**." Correlation of serum ferritin with hemoglobin A2 level in beta-leukemia traits". *Journal of Ayub Medical College, Abottabad*, 32(4), 476-480.

[6] Laurikkala, J., Juhola, M., Kentala, E., Lavrac, N., Miksch, S. and Kavsek, B. **2000**. "Informal identification of outliers in medical data". *Fifth International Workshop on Intelligent Data Analysis in Medicine and Pharmacology*, 1, 20-24.

[7] Rousseeuw, P. and Hubert, M. **2011**. "Robust statistics for outlier detection". *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1), 73-79.

[8] Hauskrecht, M., Batal, I., Valko, M., Visweswaran, S., Cooper, G. F. and Clermont, G. **2013**. Outlier detection for patient monitoring and alerting. *Journal of Biomedical Informatics*, 46(1), 47-55.

[9] Christy, A., Gandhi, G. M. and Vaithyasubramanian, S. **2015**." Cluster based outlier detection algorithm for healthcare data". *Procedia Computer Science*, 50, 209-215.

[10] Qi, M., Fu, Z., and Chen, F. **2015**. "Outlier detection method of multiple measuring points of parameters in power plant units". *Applied Thermal Engineering*, 85, 297-303.

[11] Bonnet, D. G. and Seier, E. **2003**. Confidence Intervals for Mean Absolute Deviations. *The American Statistician*, 57(4), 233-236.

[12] Boos, D. D. and Brownie, C. **2004**. "Comparing Variances and other Measures of Dispersion". *Statistical Science*, 19(4), 571-578.

[13] Elamir, E. A. (2012). Mean absolute deviation about median as a tool for explanatory data analysis. *Proceedings of the World Congress on Engineering*, 1.

[14] Sanni, O. O. M., Ikoba, N. A., and Adegboye, O. S. **2021**. Modified methods of computing some descriptive statistics for grouped data. *Benin Journal of Statistics*, 4, 113-133.

[15] Stephen, R. S. and Senthamarai, K. K. **2017**. "Detection of outliers in regression model for medical data". *International Journal for Medical Research & Health Sciences*, 6(7), 50-56.

[16] Danjou, F., Anni, F., Perseu, L., Satta, S., Dessi, C., Lai, M. E., Fortina, P., Devoto, M., and Glanello, R. **2012**. "Genetic modifiers of β-leukemia and clinical severity as assessed by age at first transfusion". *Haematological*, 97(7), 989.

[17] Nakavachara, P., Petchkul, J., Jeerawongpanich, K., Kiattissakthree, P., Manpayak, T., Netsakulnee, P., Chaichanwattanakul, K., Pooliam, J., Srichairatanakool, S., and Viprakasit, V. **2018**. Prevalence of low bone mass among adolescents with non-transfusion-dependent hem1oglobin E/ β-leukemia and its relationship with anemia severity. *Pediatric Blood & Cancer*, 65(1), e26744.

[18] Bayati, S., Keikhaei, B., Bahadoram, M., Mahmoudian-Sani, M., Vanaeshani, M., and Behbahani, F. **2021**. "Radiographic features of the maxillofacial anomalies in beta-leukemia major: with new view". *World Journal of Plastic Surgery*, 10(3), 78.

[19] Bluman, A. G.**2012**. Elementary Statistics: A Step by Step Approach. McGraw-Hill, New York.

[20] Choulakian, V. and Abou-Samra, G. **2020**. "Mean Absolute Deviations about the Mean, the Cut Norm and Taxicab Correspondence Analysis". *Open Journal of Statistics*, 10, 97-112.

الكشف عن القيم الشاذة في الانحدار الخطي البسيط للأطفال المصابين بسرطان الدم في مدينة الموصل

شيماء رياض ذنون[1] ، محمد نافع عبد الرزاق[2]

[1]كلية التمريض ، جامعة الموصل ، الموصل ، العراق

[2]مديرية تربية نينوى ، الموصل ، العراق

الملخص

في هذا البحث تمت دراسة مشكلة الكشف الخارجي عن القيم الشاذة في تحليل الانحدار الخطي باستخدام الوسيط ومتوسط الانحراف المطلق عن الوسيط يتأثر المتوسط و المعياري بشدة بالقيم الشاذة ,وبالتالي تقنيات الكشف عن القيم الشاذة المعتمدة على هذه القياسات قد لا تحدد جميع القيم الشاذة بشكل صحيح ومع ذلك فان متوسط الانحراف المطلق عن الوسيط بالاقتران مع الوسيط قوي بما فيه الكفاية في وجود القيم الشاذة ويوفر بديلا افضل. تم اختبار الطريقة باستخدام بيانات مرضى سرطان الدم وتشير النتائج الى ان الطريقة الجديدة تؤدي بشكل افضل من الطرق التي تعتمد على مزيج الانحراف المعياري /المتوسط.

يوصى باستخدام الوسيط ومتوسط الانحراف المطلق عن الوسيط في الكشف عن القيم الشاذة في تحليل الانحدار نظرا لقدرتها الكامنة على زيادة ملاءمة نموذج الانحدار الخطي .