اختيار أفضل طريقة لتقدير المعلمات في نموذج الانحدار مع التطبيق

تواريخ البحث

تاريخ تقديم البحث : 2023/11/25

تاريخ قبول البحث :2023/12/26

تاريخ رفع البحث على الموقع:

2024/6/15

الاستاذ المساعد الدكتورسميرة محمد صالح

قسم الاحصاء / جامعة السليمانية

Samira.muhamad@univsul.edu.iq

المستخلص :

تعتمد الأساليب الاحصائية في أغلب الاحيان على فرضية التوزيع الاحتمالي لبيانات نموذج الانحدار طبيعي، ولكن في الحالة التطبيقية على الأغلب تكون هذه البيانات لها توزيعات أخرى بسبب القيم المتطرفة، وذلك لاستخدام أساليب غير حساسة للقيم المتطرفة. واعطاء تقديرات فعالة مثل طرق التقدير الحصينة. في هذا العمل تم استخدام بعض الطرق (الكلاسيكية والحصينة) .وقد تم استخدام خمس طرق لتقدير المعلمات لنموذج الانحدارالمتعدد (OLS, WLS,M-estimator, LTS and LMS)، وهما طريقتان من الطرق الكلاسيكية، وثلاث طرق حصينة، حيث تم سحب عينة عشوائية بسيطة على (299) مريضاً في (مستشفى شار) و الذين كانوا مصابين بضغط الدم ممثلة بالمتغير التابع (y=BP)والمتغيرات المستقلة (B.s,B.u,Cr,Cho,Tri) .بهدف المقارنة بين هذه الطرق لايجاد أفضل طريقة للتقدير باستخدام مقاييس المقارنة وكفاءة تلك التقديرات مثل (MSE) متوسط مربع الخطأ و (R2) معامل التحديد و كل من AIC,AICc,,BIC, و RMSE لقياس كفاءة النموذج ، باستخدام البرنامج الاحصائي (JASP ،SPSS26، systa12) أظهرت النتائج أن طريقة WLS هي الأفضل وتأتي بعدها بالمرتبة الثانية طريقة LTM من بين جميع الطرق الحصينة، أما الطرق الكلاسيكية (OLS) فقد أثبت فشلها في تقدير المقدرين ذوي الكفاءة.

الكلمات الافتتاحية: طريقة التقدير، مقدرالحصين ، دالة التأثير، طريقة مربعات صغرى الموزونة ، معاييرالمفاضلة .

Choosing Best Method of Estimation Parameters in Regression Model With an Application

Samira Muhamad Salh

University of sulamani

Samira.muhamad@univsul.edu.iq

Administration & Economics College , Department of Banking & finance

University of Basrah

Abstract :

Mostly the statistical methods depend on hypothesis that probability distribution for data of regression model is normal, but in applied state at most to be this data have other distributions because of the extreme values, in order to use methods are not sensitive to extreme values. And give efficient estimations such as robust estimation Methods.in this work used some (classical and robust) methods which are (OLS, WLS, M-estimator, LTS and LMS). In this work five methods have been estimated the parameters of robust multiple regression model .Two methods of classical methods , and three methods are robust methods , It has been drawn simple random sample (299) patients in (shar hospital) infected with blood pressure represented by dependent variables (BP) and independent variables (B.S, B.U,BMI,CR,CHO, and Tri.) . To aim compare between these methods in order to find the best method for estimation by using measures reflect the quality and efficiency of those estimates , such as mean square error (MSE) and determination factor (R^2 ,AIC,AICc,,BIC,MSE, and RMSE) to measure the model efficiency ، by using statistical program (JASP,SPSS26 and systa12). The results show that the method of WLS is the best and secondly method is LTM among of all robust method ، Either the classical (OLS) methods has demonstrated its failure to estimate efficient estimators.

Keywords: **Estimation Method, Robustness, Influence function, WLS method, Best criteria.**

Aim of research:

1. **Comparison between several methods.**
2. **Choosing best models**

1- Introduction:

It is known as that the estimation in statistical methods depends on a number of important assumptions.To obtain an accurate regression model, the probability distribution of the data is the most important hypothesis that is under study, which is often distributed normally.Sometimes the distributed data takes a different pattern and may not be represented by a particular pattern of distributions Which leads to (outliers) distributions, and the reason may sometimes be due to the presence of extreme values not checking the least squares hypotheses, and then they will lose their good properties, and so we do Searching for alternative, immunized ways to address this problem and that are insensitive to the presence of extreme values and give us efficient estimators . In this work, some types of classical estimation methods will be used .Ordinary least squares (OLS) and the weighted least squares method (WLS) ,compare it with the some robust estimation methods such as M-estimation method, Least median squares method (LMS), Least trimmed squares method (LTS) to find the best estimation of the parameters of the multiple regression model. [7,17,22]

4. Methods of Estimation:

The main purpose of the estimation methods is to obtain the estimations of the parameters of the model it must have good specifications that make it a model that can be used and used Estimation methods are different according to different ideas and methods to achieve two important purposes.The first is to contain the capabilities are on good specifications, and the other purpose is to show the method in an easy-to-implement manner. There are several methods for estimating the parameters of the regression model, and these methods give good results when there are regression hypotheses, but when one of these hypotheses is violated or when there are extreme values, these the methods will lose their advantages and will negatively affect the results of the assessment. [19,24]

4.1.1- ordinary least square method (OLS method) [1,5,9,12]:

This method is one of the most widely used methods for estimating the parameters of the linear regression model. The least squares estimate of the regression parameters in this method . The goal is predicting the unknown parameters by minimize ($\sum_{i=1}^{n} \varepsilon_i^2$). The intention is to select the parameters that the residuals are "small." While there are various methods with which to minimize the residuals. The model can be thought of as a similar to the slope-intercept form of a line with an error (residual) term.

The simple linear regression model is given by:

$$Y_i = f(X_i, \beta) + u_i \qquad (1)$$

Where: yi the response (dependent) variables with (n) of sample size.

β_j The parameters of the models.

X_{ij} Explanatory (independent) variables with $i = 1, 2, \ldots, n$ $j = 1, 2, \ldots, m$

U_i The error term.

We can express it in matrix form as: yi = Xβ + U (2)

For estimate the parameters of the model, there are certain hypotheses that must be met. These assumptions are:

1. The relationship between the variables involved is a linear relationship and this means that the variable y is determined as a linear structure of independent variables, X and random variable u.

2. The variance of the random error are independent that is means:

$$\text{cov}(U_i, U_j) = E(U_i, U_j) = 0 \qquad \forall \; i \neq j \qquad (3)$$
$$i, j = 1, 2, \ldots, n$$

3. The normal error with mean zero and variance σ2. $U_i \sim N(0, \sigma_u^{\,2})$

4. No correlated between (U_i) and (X_i), so the covariance is zero, i.e.: $E(U_i X_i) = 0$

5. No multicollinearity between the values of the explanatory variables. $E(X_i X_j) = 0$

Also some properties needed in this method and these assumptions are:

The estimated parameters in this method are linear in terms of the response variables, this property called Linearity. $\hat{\beta}_{ols} = (X'X)^{-1} X'Y$

2. Second un-biasness, That is, the expected value of the estimated parameters is equal to its real value: $E(\hat{\beta}_{ols}) = \beta$

3. The minimum variance of the estimated parameters: $\text{var}(\hat{\beta}_{ols}) = (X'X)^{-1} \sigma^2$

4.1.2-Weighted least square method (OLS method)[8,14,22]:

The ordinary least squares method assumes that there is a continuous variance in errors, so we will resort to using the weighted least squares method when the assumptions of squares are violated. The

normal minimum continuous change in errors (which is called heterogeneity) and the model under consideration

$$Y = X\beta + U$$

$$\hat{\beta}_{WLS} = \left(X'W^{-1}X\right)^{-1} X'W^{-1}Y \qquad ...(4)$$

The equation it gives us the best-unbiased linear estimator of the parameter (β) in a case of heterogeneity It is called the estimation method in the weighted least squares method .this method generally depend on the weight.

$$\text{var}-\text{cov}(\hat{\beta}_{wls}) = \left[\sigma_u^2 (X'W^{-1}X)^{-1}\right] \qquad ...(5)$$

The equation (5) give us a variance, and the var-covarince of parameters ($\hat{\beta}_{WLS}$) ,Which contains parameters of model when the variance is heterogeneity.

4.2--Robust Estimation:

The classical methods in estimating the parameters of the model is inaccurate in the analysis Data when there is a defect in one of the regression hypotheses, the presence of outliers, or the error distribution the random distribution is not the normal distribution that is suitable for the method adopted in the estimation. One or more outliers will lead to a defect in the properties of the least squares estimators, even if the estimator is the robust is the one that maintains the desired properties of abilities when certain assumptions are violated. We will look at some of the following estimation methods: [2,7,18]

4.2.1.:The General Class of M-estimators:

An estimator is often chosen as a member of a general class of estimators that is optimal in some sense or fulfills a set of good properties .Huber (1964, 1967) proposed a class of M-estimators that naturally generalize the MLE. An M-estimator is given by the solution $\hat{\beta}_M$ of the minimization $\min \sum_{i=1}^{n} \rho(x_i, \beta)$ or, alternatively, by the solution for θ of $\min \sum_{i=1}^{n} \psi(x_i, \beta)$ for suitable ρ and ψ functions, where:

$$\psi(x, \beta) = \frac{\partial \rho(x;\beta)}{\partial \beta}$$

In general, $\psi(x,\beta)$ needs not be the derivative of some ρ -function with respect to the parameter of interest, therefore (2.73) is more general and is often ref erred as the proper definition of an M-estimator. M-estimators include the so-called weighted MLE (WMLE) defined as the solution for β .The weights can depend on the observations only, on a quantity that depends itself on the observations and the parameters or, more generally, directly on the score function.In the linear regression model $y_i = x_i'\beta + \varepsilon_i$, $\text{var}(x_i) = 1$. The score function has a similar expression $s(y,x,\beta) = r.x$ but is proportional to $r = y - x'\beta$, the residual, and x, the covariate function itself. In the univariate case, popular choices are Huber.s weights $w_h(r;\beta,c) = \dfrac{\psi_h(r,\beta,c)}{r} = \min\{1;\dfrac{c}{r}\}$ i.e. The weight is equal to one for all (small) values of r satisfying |r| < c and $\dfrac{c}{r}$ otherwise. Note that in a regression model with known scale, the Huber estimator is an M-estimator associated with $w_h(r,x;\beta,c) = \psi_h(r,x,\beta,c)x$ a bounded function of r (or the response y).The **ρ** and **Ψ**functions of the MLE and Huber proposal are depicted in Figure 1,left and middle panels.

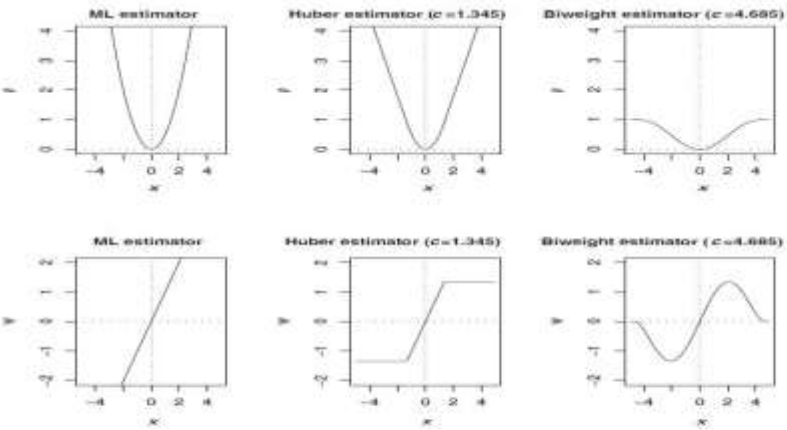

Figure 1

This Figure displays the ρ and ψ - function for the MLE, Huber and bi-weight estimators. For the MLE, the ψ -function (score) and corresponding $\rho = \log(f)$ are unbounded .in contrast, the Huber estimator has a score function bounded by c , and ρ function is quadratic in the middle and linear in the tails. Finally the bi-weight ψ -function score is re-descending with the Corresponding ρ -function

being symmetric and constant for $|r| \succ c$ .as a result the MLE are not robust, the Huber estimator is robust. [1,11,16,19]

There are some methods for estimating the coefficients of the linear model, in this work we use two of this method, which are:

1. Least median squares method (LMS)[4,11,20]:

The median least squares method is proposed by Rousseeuw to provide robust estimators for the parameters in a linear regression model, by replacing the sum of squares with a residual in (ols) and replacing them with a median of squares with a residual .The following estimators (LMS) can be obtained if optimization is achieved :

Suppose that :

$X' = (x_{i1,} x_{i2}, \ldots., x_{ip})$ **Where** $i = 1, 2, 3, \ldots, n$, $j = 1, 2, 3, \ldots, p$

$Y_i = '(1, 2, 3, \ldots, y_n)$, $i = 1, 2, 3, \ldots, n$

Where: X', Y_i **are two real values vectors with** $\frac{n}{2} \geq p$

P: numbers of variables.

$X = |x_{ij}|$ **is full rank matrix with (** $n \times p$ **) dimensions .**

Let $\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$, β **vectors of parameters in the model.**

And by least median squares (LMS) it is possible to find (β_{LMS} **) according to the following formula:**

$$\beta_{LMS} = \min_{b} \underset{i=1}{\overset{n}{med}} (Y_i - X'\beta)^2 \qquad \ldots(6)$$

Where: $e_i^2 = (Y_i - X'\beta)^2$

In addition, can write the models as follow: $\beta_{LMS} = \min_{b} \underset{i=1}{\overset{n}{med}} (e_i^2)$

β_{LMS} **: is the middle of the shortest half in a partial sample and is called the median least squares estimator. To find an estimator in the case of a single variable of size n , one of the following partial samples must be taken (n-h+1) as follow :**

$\{X_{(1)}, X_{(2)}, \ldots, X_{(h)}\}$
$\{X_{(2)}, X_{(3)}, \ldots, X_{(h+1)}\}$
$\{X_{(3)}, X_{(4)}, \ldots, X_{(h+2)}\}$

$\{X_{(n-h+1)}, X_{(n-h+2)}, \ldots, X_{(n)}\}$ ...(7)

Then find the shortest half in a partial sample, and this is done by finding the smallest differences as follows:

$X_{(h)} - X_{(1)}$
$X_{(h+1)} - X_{(2)}$
$X_{(h+2)} - X_{(3)}$

$X_{(n)} - X \ (n-h+1)$ ...(8)

Where: $h = \frac{n}{2} + 1$ **and** X_i **is order value.**

And the sub-sample that represent corresponds to the smallest difference in the equation (@) and contains (h) values is called the shortest half because it has the shortest range among all possible sub samples with (h) elements.

And the estimator (β_{LMS} **) is equal to the midpoint of this sub-sample that corresponds to the smallest difference, and in the case of several equal short halves, the average of its middle is taken .**

2. Least trimmed squares method (LTS):

It is a statistical method for estimating the unknown coefficients of the linear regression model, and it is considered a robust and alternative to the classical methods.This method was also suggested by (Rousseeuw), because it is characterized by statistical efficiency and better positional stability than the median least squares **method. Let us assume that the linear regression model for a sample (** Y_i, X_i **) and Response variable** $Y_i \in R$ **and a vector of explanatory variables** $X \in R^p, \beta \in R^p$.

Where: p :numbers of parameters .

By using (LTS) can obtain estimate β_{LMS} **as follow:**

$b_{LTS} = \min \sum_{i=1}^{h} \left(e^2\right)_{in}$ **, h: is constant values the range of h is** $\left(\frac{n}{2} \prec h \prec n\right)$ **. And square residual ordered is** $\left(e^2\right)_{in}$ **.**

$$\left(e^2\right)_{1:n} \le \left(e^2\right)_{2:n} \le \dots \le \left(e^2\right)_{n:n}$$

The researcher suggested an estimator (β_{LMS} **) for the location of the univariate variable of size (n) and taking into account (n-h+1) of the following partial samples:**

$$\{X_{(1)}, X_{(2)}, \dots, X_{(h)}\}$$
$$\{X_{(2)}, X_{(3)}, \dots, X_{(h+1)}\}$$
$$\{X_{(3)}, X_{(4)}, \dots, X_{(h+2)}\}$$

$$\{X_{(n-h+1)}, X_{(n-h+2)}, \dots, X_{(n)}\} \quad \text{....(9)}$$

Each sample contains h elements, each sub-sample is called the contiguous half, and the average is calculated for each partial sample as follows:

$$\bar{X}_{(1)} = \frac{1}{h}\sum_{i=1}^{h} X_{(i)}$$
$$\bar{X}_{(2)} = \frac{1}{h}\sum_{i=2}^{h+1} X_{(i)}$$

$$\bar{X}_{(n-h+1)} = \frac{1}{h}\sum_{i=n-h+1}^{n} X_{(i)} \quad \text{...(10)}$$

And calculate the sum of squares for each sub-sample:

$$SQ_{(1)} = \sum_{i=1}^{h}\left(X_{(i)} - \bar{X}_{(1)}\right)^2$$
$$SQ_{(2)} = \sum_{i=2}^{h+1}\left(X_{(i)} - \bar{X}_{(2)}\right)^2$$

$$SQ_{(n-h+1)} = \sum_{i=n-h+1}^{n}\left(X_{(i)} - \bar{X}_{(n-h+1)}\right)^2 \quad \text{...(11)}$$

The β_{LMS} **It is the mean that corresponds to the least squared sum of last equation. [11,13,20]**

3. Choosing the best estimate of the parameters of the regression model:

After we estimated the parameters of the model for each method using the classical and robust estimation methods, we choosing the best method from among these methods, this was done by means

of some criteria of comparison, including. In this work has many criteria are used to comparison between models such as (R2, AIC, AICc, BIC, RMSE, and MSE).and it is calculated according to the following formula [3,6,10,15]**:**

criteria		Formula	Descriptions
1	R2	$R^2 = \frac{MSE}{MST}$	MSE :Mean square error MST :Mean square total
2	AIC	$AIC = -2\log L + 2(k+p)$	Akaike Information Criterion
3	AICc	$AICc = AIC + \frac{2k^2 + 2k}{N - K - 1}$	Correct Akaike Information Criterion
	BIC	$AIC = -2\log L + 2\log n(k+p)$	Bayesian Information Criterion
4	MSE	$MSE = \frac{SSE}{d.f}$	MSE :Mean square error
5	RMSE	$RMSE = \sqrt{MSE}$	Root square of MSE :Mean square error

2.1- Data Description and analysis:

In this section, we discuss practical part and all results will be presented to the Applied of the study by using Statistical package (JASP, SYSTAT12 and SPSS26) program. Which is characterized by ease of handling and accuracy of its results and contains many tools that help researchers in the process of analyzing and arriving at accurate results.

2.1.1: Data Description:

In this part, results will be presented to the applied side of the study by using a by using Statistical package (JASP, SYSTAT12 and SPSS26) program. which is characterized by ease of handling and accuracy of its results and contains many tools that help researchers in the process of analyzing and arriving at accurate results. Sample taken from the (central laboratory in the sulamani) data as an (299) patients with (In this study a sample was taken from (299) patients suffering from hyperactive blood pressure, with six independent variables which are {(B.S) Blood sugars, (B.U) Blood Urea, Certain, (BMI) Body mass index, (Chol.)Cholesterol, and (Tri.) triglyceride},with (y=response) variable is the hyperactive blood pressure (BP)The goal of this study is to determine the effect of these variables on the hyperactive blood pressure (y).

Result and analysis:

At the first by using (JASP , SYSTA13 and SPSS26 program package), we show that the result of descriptive statistics for all variables entered in study.

Table (1): Represents the descriptive statistics for all variable's

	PB	B.SUGEAR	B.UERA	BMI	Certain	chol	tri
mean	34.30	257.70	7.08	103.83	59.76	4.97	13.94
S.Deviation	7.87	64.29	1.90	15.50	8.92	0.98	2.75
N	299	299	299	299	299	299	299

And then to determine whether there is all independent variables are affected the Dependent variable or not, multiple regression analysis according to the (ALL) methods.

Table (2): Represents the Coefficients of OLS method

model	Beta	Std.Error	t	Sig	Lower bound	Upper bound
constant	102.414	3.759	27.244	0.000	95.006	109.822
B.sugaer	-0.003	0.005	-0.729	0.467	-0.012	0.006
B.urea	-0.141	0.161	-0.878	0.381	-0.459	0.176
BMI	-0.010	0.017	-0.590	0.555	-0.044	0.024
Creatin	-0.822	0.034	-23.94	0.000	-0.889	-0.754
Cholesterol	-0.569	0.471	-1.208	-0.288	-1.496	0.359
Trigl.	-0.950	0.177	-5.363	0.000	-1.299	-0.601

Table (3): Represents the Coefficients of WLS method

model	Beta	Std.Error	t	Sig	Lower bound	Upper bound
constant	83.450	2.505	33.307	0.000	95.006	109.822
B.sugaer	-0.000	0.002	0.115	0.908	-0.012	0.006
B.urea	-0.287	0.101	-3.544	0.004	-0.459	0.176
BMI	0.007	0.012	0.592	0.555	-0.044	0.024
Creatin	-0.840	0.017	-48.96	0.000	-0.889	-0.754
Cholesterol	0.959	0.382	2.512	0.013	-1.496	0.359
Trigl.	-0.280	0.136	-2.058	0.041	-1.299	-0.601

Table (4): Represents the Coefficients of (M-Estimate-Huber) method

model	Beta	Std.Error	t	Sig	Lower bound	Upper bound
constant	91.147	5.753	15.843	0.0001	79.811	102.483
B.sugaer	0.001	0.006	0.166	0.867	-0.012	0.013
B.urea	-0.093	0.216	-0.430	0.667	-0.519	0.332
BMI	-0.002	0.022	-0.090	0.927	-0.044	0.041
Creatin	-0.915	0.047	-19.468	0.0001	-1.007	-0.823
Cholesterol	0.136	0.643	0.211	0.832	-1.131	1.402
Trigl.	-0.079	0.276	-0.286	0.774	-0.623	0.465

Table (5): Represents the Coefficients of (M-Estimate-Humble) method

model	Beta	Std.Error	t	Sig	Lower bound	Upper bound
constant	92.256	7.231	12.758	0.0001	77.997	106.515
B.sugaer	0.001	0.008	0.125	0.900	-0.015	0.016
B.urea	0.029	0.268	0.108	0.914	-0.499	0.558
BMI	0.001	0.027	0.037	0.970	-0.052	0.054
Creatin	-0.956	0.059	-16.203	0.0001	-1.072	-0.840
Cholesterol	0.076	0.782	0.097	0.922	-1.467	1.618
Trigl.	0.010	0.344	0.029	0.976	-0.669	0.688

Table (6): Represents the Coefficients of (M-Estimate- Bi –square) method

model	Beta	Std.Error	t	Sig	Lower bound	Upper bound
constant	93.696	7.869	11.906	0.0001	78.169	109.223
B.sugaer	-0.001	0.009	-0.111	0.911	-0.018	0.016
B.urea	0.100	0.295	0.338	0.735	-0.481	0.682
BMI	0.002	0.029	0.068	0.945	-0.055	0.059
Creatin	-0.969	0.064	-15.140	0.0001	-1.096	-0.842
Cholesterol	0.032	0.855	0.037	0.970	-1.655	1.719
Trigl.	-0.029	0.378	-0.076	0.939	-0.776	0.717

Table (7): Represents the Coefficients of (LMS) method

model	Beta	Std.Error	t	Sig	Lower bound	Upper bound
constant	93.766	0.890	105.355	0.0001	92.009	95.524
B.sugaer	-0.001	0.002	-0.500	0.617	-0.003	0.001
B.urea	0.114	0.034	3.352	0.0009	0.048	0.181
BMI	0.003	0.013	0.230	0.818	-0.003	0.009
Creatin	-0.968	0.007	-138.286	0.0001	-0.982	-0.954
Cholesterol	93.766	0.890	105.355	0.0001	92.009	95.524
Trigl.	-0.001	0.011	-0.090	0.928	-0.003	0.001

Table (8): Represents the Coefficients of (LTS) method

model	Beta	Std.Error	t	Sig	Lower bound	Upper bound
constant	88.833	3.582	24.799	0.0001	81.762	95.904
B.sugaer	0.001	0.004	0.250	0.802	-0.006	0.009
B.urea	-0.360	0.133	-2.703	0.007	-0.422	0.102
BMI	-0.011	0.013	-0.846	0.398	-0.037	0.015
Certain	-0.912	0.028	-32.571	0.0001	-0.968	-0.856
Cholesterol	0.175	0.525	0.333	0.739	-0.862	1.211
Trigl.	0.103	0.188	0.547	0.584	-0.269	0.474

Table (9): Represents the Best models selection criteria

	Classical Method		Robust Method				
Method / Criteria	**OLS**	**WLS**	**M-Estimate Huber**	**M-Estimate Humble**	**M-Estimate Bi-square**	**LTS**	**LMS**
MSE	15.85	1.28	28.24	37.24	42.04	24.50	18.51
R^2	0.75	0.86	0.63	0.61	0.61	0.87	0.89
AIC	644.82	85.11	1010.88	1093.99	1129.85	968.40	884.57
AICc	645.10	85.40	1011.17	1094.28	1130.14	968.69	884.86
BIC	666.20	88.59	27.51	26.96	26.72	27.80	28.36
RMSE	3.98	1.130	6.11	6.48	5.31	4.94	4.30

Table (10): Represents the Collinearity test

Tolerance	0.798	0.793	0.986	0.742	0.322	0.291
VIF	**1.253**	**1.353**	**1.014**	**1.348**	**3.105**	**3.438**

Table (11): Represents the Test of Heteroscedasticity

Test	Chi –square	Sig.
White test	**6.437**	**0.011**
Breusch- pagan	**5.045**	**0.025**
F-Test	**7.649**	**0.010**
Notice that all of the testes indicated that there is a problem with Heteroscedasticity.		

Result and discussion:

After applying All the methods that is used in this work on the sample data size (n=299) for the cases under the hyperactive blood pressure. By taking the hypothesis that the (BP: y) depends on the expletory variables such as {X1=B.S, X2=B.urea, X3=Cr., X4=BMI, X5= Cholesterol and X6=TRI} and comparing their results, the following important points are concluded below:

1. Table (1) shows all the measurements of descriptive statistics. Since the Descriptive statistics gave us an overview for the work that we have done.
2. From table (2) both of the variables (x4,x6) have highly significant effects on the response variable. In addition, the intercept has a highly significant effect. However, the other four variables (x1,x2,x3,x5) appears to have no significant effects.
3. The output from Table (3) shows that the results of fitting a multiple linear regression model by using (WLS) to describe the relationship between (BP) and six independent variables. Since the Significance level (P-value) in the table is less than 0.01, there is a statistically significant relationship between the variables (x2,x4) but x5 is nearly significant at the 99% confidence level. If we put the confidence interval at 95% then the variables (x2, x4, x5 and x6) will be significance.
4. Tables (4,5,6) shows the result analysis of Robust M-estimate methods by using (Huber ,humble ,bi-square) Weights. These results represent that the only variable that has effect on the dependent variable and highly significant is X_4= Creatin and its values is less than 0.01.

5. Table (7,8) shows that the Robust method of (LTS ,LMS) respectively, in table (7) both of X_2 , X_4 and X_5 are significant since their values are less than 0.01. In table (8) the most significant variables are X_2 and X_4.

6. In Table (9) we depend on the criteria methods such as (R^2, AIC, AICc, BIC, MSE and RMSE) to detect the best method, where WLS approach is the best method by comparing it with OLS method where four method of criteria are shows us that WLS is better than OLS. As well as in the Robust methods LMS is the best method since Three of the criteria methods is least by comparing it with the other methods at the end WLS is better than LMS because Four methods of criteria in WLS is smaller than LMS.

7. Table (10) represents the value of VIF which is (1.06, 1.32, 1.50, 1.42, 1.02, 1.24) and this indicates that there is no correlation among the (explanatory variables).

8. WLS is the best model among all the other models but in table (11) we can see that the tests shows us that there are Heteroscedasticity problem. Where WLS was able to remove the effect of heteroscedasticity in the data.

Reference:

[1].Abed, Mohamed.S.,(2013)" comparing some robust methods with OLS in multiple regression with application "MSc mathematics /college of science /university of Duhok .

[2].Al-Amleh ,M.A., AND Al-Athari,F.M,(2015) "Robust methods in regression analysis :comparison and improvement ,zarqa university ,Jordan .

[3] Akossou, A. Y. J. and Palm, R. (2013) 'Impact of data structure on the estimators r-square and adjusted r-square in linear regression', Int. J. Math. Comput, 20, pp. 84–93.

[4] Arnold J. Stromberg , (1991),"Computing the exact value of the least median of squares estimate in multiple linear regression" University of Minnesota
Technical Report #561.

[5]. Burton, A. L. (2020) 'THE OLS LINEAR REGRESSION FRAMEWORK'.

[6]. Burnham, Kenneth P. and David R. Anderson. 1998. Model Selection and Inference: A Practical Information-Theoretical Approach. New York: Springer-Verlag.

[7].chen,C., (2002). "robust regression and outlier detection with the ROBUSTREG procedure" .in proceeding of the twenty –seven Annual SAS user group international conference. paper 265-27

[8] . Farahani, H. A., Rahiminezhad, A. and Same, L. (2010) 'A Comparison of Partial Least Squares (PLS) and Ordinary Least Squares (OLS) regressions in predicting of couples mental health based on their communicational patterns', Procedia-Social and Behavioral Sciences. Elsevier, 5, pp. 1459–1463.

[9] .Güneri, Ö. İ. and Göktaş, A. (2017) 'A Comparion of Ordinary Least Squares Regression and Least Squares Ratio via Generated Data', 7(2), pp. 60–70. doi: 10.5923/j.ajms.20170702.02.

[10] .H. Bozdogan.(2000) 'Akaike's information criterion and recent developments in information complexity,' J. Math. Psychol., vol. 44, no. 1, pp. 62–91.

[11]. Huber, P. J. & Ronchetti, E. M. (2009), Robust Statistics, (2nd ed.), John Wiley & Sons, Inc., Hoboken, New Jersey, Publication.

[12]Hutcheson, G. D. (2011) 'Ordinary least-squares regression', L. Moutinho and GD Hutcheson, The SAGE dictionary of quantitative management research, pp. 224–228.

[13] Jurgen A. Doornik ,(2011),"Robust Estimation Using Least Trimmed Squares" Institute for Economic Modelling, Oxford Martin School, and Economics Department, University of Oxford, UK.

[14] Montgomery, D., Peck, E. and Vining, G. (2012)." Introduction to Linear Regression Analysis", 5th edn. Hoboken, New York: John Wiley & Sons.

[15]. Moore.D.F. (2016)."APPLIED SURVIVAL Analysis using R "Piscataway, NJ, Springer international publishing

[16].nahar,J,and Purwani,S.,2017"application of robust M-estimator regression in handling data outlie" .4th ICRIEMS proceeding ,pp.53-60.

[17]. Omer, A. W. (2010), Fitting Non-Linear Regression Models to Thalassemia patients, (Master's Thesis), Collage of Administration and Economics University of Sulaimani.

[18] Ronchetti, Elvezio M., (2006), The Historical Development of Robust Statistics, University of Geneva, Switzerland, ICOTS-7, 2006: Ronchetti, Elvezio.Ronchetti@metri.unige.ch.

[19].susanti,Y. and Pratiwi,H.,2014.m-estimation,s-estimation ,and MM-estimation in robust regression .international journal of pure and applied mathematics ,91(3),pp.349-360.

[20] Vanessa Berenguer-Rico_, Søren Johanseny & Bent Nielsenz , (2019) , "Models where the Least Trimmed Squares and Least Median of Squares estimators are maximum likelihood " .

[21] Weisberg, S. (2005). Applied Linear Regression, 3rd edn. Hoboken, NJ: John Wiley & Sons.

[22]. West, B. T., Welch, K. B. & Gałecki, A. T. (2015), Linear Mixed Models a Practical Guide Using Statistical Software, (2nd ed.), © 2015 by Taylor & Francis Group, LLC.

[23].Yeliz. Mert .Kantar ,(2015)," Generalized Least Squares and Weighted Least Squares Estimation Methods For Distributional Parameters "Statistical Journal ,Volume 13, Number 3, November 2015, 263–282.

[24].yulita ,T.,Notodiputro,K.A. and sadik ,K.,2018,M-Estimation use Bi-Square,Hample,Huber and welsch Weight functions in robust regression .